Pattern Recognition Letters 37 (2014) 107–114

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Constrained instance clustering in multi-instance multi-label learning

Yuanli Pei *, Xiaoli Z. Fern

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:
Available online 11 July 2013

Keywords:
MIML
Instance clustering
Constrained clustering
Bag constraints
Spectral clustering

ABSTRACT

In multi-instance multi-label (MIML) learning, datasets are given in the form of bags, each of which contains multiple instances and is associated with multiple labels. This paper considers a novel instance clustering problem in MIML learning, where the bag labels are used as background knowledge to help group instances into clusters. The goal is to recover the class labels or to find the subclasses within each class. Prior work on constraint-based clustering focuses on pairwise constraints and cannot fully utilize the bag-level label information. We propose to encode the bag-label knowledge into soft bag constraints that can be easily incorporated into any optimization based clustering algorithm. As a specific example, we demonstrate how the bag constraints can be incorporated into a popular spectral clustering algorithm. Empirical results on both synthetic and real-world datasets show that the proposed method achieves promising performance compared to state-of-the-art methods that use pairwise constraints.

1. Introduction

The multi-instance multi-label (MIML) learning framework (Zhou and Zhang, 2007) has been successfully applied in a variety of applications including computer vision (Feng and Xu, 2010; Xue et al., 2011; Zha et al., 2008) and audio analysis (Xu et al., 2011). In MIML, datasets are given in the form of bags and each bag contains multiple instances. It is assumed that there exists a class structure such that each instance in the bag belongs to one of the classes. However, the instance class labels are not directly observed. Instead, the class labels are only provided at the bag level, which is the union of all instance labels within the bags. The goal of MIML learning is then to build a classifier to predict the labels for an unseen bag (Zhang and Zhou, 2008; Zhou and Zhang, 2007) or to annotate the label of each instance within the bag (Briggs et al., 2012).

In this paper, we consider a novel instance clustering problem within the MIML framework, where the goal is to group instances from all bags into clusters. In particular, we seek to find a cluster structure that corresponds to or refines the existing class structure. That is, we assume that each class contains one or more subclasses and our goal is to find such subclasses via clustering. In our motivating application, we want to understand the structure of bird song within each species. Here a bag corresponds to the spectrogram of a 10-s field recording of multiple birds, and each instance corresponds to a segment in the spectrogram capturing a single bird utterance (a syllable). The labels of a bag are the set of species (one or more) present in the recording. Birds from a single species may vocalize in different modes. For instance, the sound made by a woodpecker has at least two distinct modes: pecking and calling. We are interested in finding such distinct modes within each species by applying clustering techniques to instances. Ideally we would perform clustering on instances of the same species to learn such modes. However, this is impractical because the labels are only provided at the bag level and we do not have accurate instance-level species labels. Therefore, we cast this problem as an instance clustering problem with bag-level class labels as side information.

Existing literature on clustering with side information primarily focuses on pairwise Must-Link (ML) and Cannot-Link (CL) constraints (Ji and Xu, 2006; Kamvar et al., 2003; Kulis et al., 2009; Wang et al., 2009; Wang and Davidson, 2010; Yu and Shi, 2001, 2004). Note that one could potentially generate ML and CL constraints based on the bag-level labels, but they incorporate only limited information for MIML datasets (as will be discussed in Section 4.3) and are not effective for our problem. Another closely related topic is MIML instance annotation (Briggs et al., 2012; Yang et al., 2010; Zha et al., 2008), where an instance classifier is learned from MIML data that predicts the class label of each instance. The key difference between MIML instance annotation and our work is that we are interested in finding the refinement of class structure for the instances, whereas instance annotation only focuses on recovering the class labels of instances based on the bag-level labels.

In this paper, we propose to incorporate the bag-level side information in the form of *bag constraints*. Our approach defines two similarity measures between bags based on *class* labels and *cluster* labels respectively. By requiring the two similarities to

* Corresponding author. Tel.: +1 (541) 737 3617; fax: +1 (541) 737 1300.
E-mail addresses: peiy@eecs.oregonstate.edu (Y. Pei), xfern@eecs.oregonstate.edu (X.Z. Fern).
URL: http://web.engr.oregonstate.edu/~xfern (X.Z. Fern).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.07.002

order pairs of bags consistently, we encode bag-level label knowledge into soft constraints, which can be easily incorporated into traditional clustering objectives as a penalty term. In particularly, we incorporate such constraints into a popular spectral clustering algorithm and validate the effectiveness of the resulting method on both synthetic and real-world datasets. Experiments show that our method produces good clustering results compared to spectral clustering methods with pairwise constraints.

2. Problem statement

In our problem, the data consists of M bags $\{\mathbb{B}_1, \ldots, \mathbb{B}_M\}$, where each bag $\mathbb{B}_i$ contains n_i instances, i.e., $\mathbb{B}_i = \{x_{i1}, \ldots, x_{in_i}\}$ with $x_{iq} \in \mathcal{R}^d$. As prior knowledge, each $\mathbb{B}_i$ is associated with a set of class labels, denoted by $\mathbb{Y}_i \subseteq \{1, \ldots, C\}$, where C is the total number of distinct classes. Denote $\mathcal{X} = \bigcup_{m=1}^{M} \mathbb{B}_m$ and let $N = \sum_{m=1}^{M} n_m$ be the total number of instances[1] in $\mathcal{X}$, our goal is to partition the N instances in $\mathcal{X}$ into K disjoint clusters that respect the class boundaries. That is, if x_p and x_q belong to the same cluster, they must belong to the same class, while the converse is true only if $K = C$, in which case we wish to recover the classes perfectly by clustering. In the case of $K > C$, some classes may contain multiple clusters that correspond to subclasses of the existing classes.

3. Bag constraints for MIML instance clustering

In our setup, the desired cluster labels are closely related to the class labels. To capture this relationship, we introduce two different representations for each pair of bags using their class-label set and cluster-label set respectively, and require these two representations to induce similarities that behave similarly in terms of their ranking orders. That is, if a pair of bags $\mathbb{B}_i$ and $\mathbb{B}_j$ is more similar to each other than another pair $\mathbb{B}_r$ and $\mathbb{B}_s$ according to their class labels, the similarity should maintain the same order when measured using cluster labels. This will allow us to find a clustering solution that implicitly respects the class labels.

More formally, we use (i,j) to represent a pair of bags $\mathbb{B}_i$ and $\mathbb{B}_j$. Let $\Omega_L(i,j)$ be the *class-label similarity* between $\mathbb{B}_i$ and $\mathbb{B}_j$, and let $\Omega_A(i,j)$ be their *cluster-label similarity*.[2] Conceivably, a good clustering result is such that a large value of $\Omega_L(i,j)$ corresponds to a large value of $\Omega_A(i,j)$. For example, for a pair of bags $\mathbb{B}_i$ and $\mathbb{B}_j$ with a certain number of class labels, the more class labels they share, the larger the value $\Omega_L(i,j)$ will be and correspondingly we expect the value $\Omega_A(i,j)$ to be larger.

Using the above defined notation, we introduce the bag constraints as follows:

$$[\Omega_L(i,j) - \Omega_L(r,s)][\Omega_A(i,j) - \Omega_A(r,s)] \geqslant 0, \quad \forall i,j,r,s \in \{1, \ldots, M\} \tag{1}$$

The first term on the left hand side of the above inequality compares the difference of class-label similarities between (i,j) and (r,s). The second term computes the corresponding difference of the cluster-label similarities. By requiring the nonnegativity of the product, the inequality requires the two similarities to consistently order any pairs of bags. In this way, the bag constraints indirectly enforces the consistency between class labels and cluster labels for all bags.

The above bag constraints can be easily incorporated into any optimization based clustering algorithm. Let f_A be the objective to be maximized by a clustering algorithm, the bag constraints can be incorporated as

$$\max_A \quad f_A + \frac{\alpha}{2M^2} \sum_{(i,j)} \sum_{(r,s)} [\Omega_L(i,j) - \Omega_L(r,s)][\Omega_A(i,j) - \Omega_A(r,s)] \tag{2}$$

where M is the total number of bags, $2M^2$ is introduced as a normalizer to make α invariant to different number of bags, and the parameter α controls the trade-off between the bag constraints and the original clustering objective.

4. Incorporate bag constraints to spectral clustering

In this section, we incorporate the bag constraints into spectral clustering by modifying the *Normalized LinkRatio* objective. We show that this leads to a standard spectral clustering problem with a modified similarity matrix.

4.1. Preliminaries on spectral clustering

We first briefly review the spectral clustering. Let $A = [a_1, \ldots, a_K]$ be a *partition matrix*, where each column a_k is a binary assignment vector for cluster $\mathbb{X}_k$, with $a_{qk} = 1$ if instance x_q is assigned to cluster $\mathbb{X}_k$ and 0 otherwise. Let W be the symmetric *similarity matrix* of instances. Define the *degree matrix* $D = \text{Diag}(W\mathbf{1}_N)$, where $\text{Diag}(\cdot)$ forms a diagonal matrix with elements of the input vector as the diagonal elements, $\mathbf{1}_N$ denotes a N-dimensional vector of all 1's, and N is the total number of vertices. The K-way spectral clustering with *Normalized LinkRatio* objective is defined as (Yu and Shi, 2003)

$$\max_A \quad \frac{1}{K} \sum_{k=1}^{K} \frac{a_k^T W a_k}{a_k^T D a_k} \tag{3}$$

$$\text{s.t.} \quad A \in \{0,1\}^{N \times K}, \quad A\mathbf{1}_K = \mathbf{1}_N \tag{4}$$

Rewrite the objective as $\frac{1}{K}\sum_{k=1}^{K} \frac{a_i^T W a_i}{a_i^T D a_i} = \sum_{k=1}^{K} \frac{a_k^T D^{1/2} D^{-1/2} W D^{-1/2} D^{1/2} a_k}{a_k^T D a_k}$. Define $z_k = \frac{D^{1/2} a_k}{\|D^{1/2} a_k^T\|}$, and $Z = [z_1, \ldots, z_K]$. Ignoring the discrete constraint for Z at this stage, one can formulate a new clustering problem with respect to variable Z as

$$\max_Z \quad \text{tr}(Z^T D^{-1/2} W D^{-1/2} Z) \tag{5}$$

$$\text{s.t.} \quad Z^T Z = I \tag{6}$$

where the constraint (6) comes from the definition of Z. The solution of Z for this new problem is the eigenvectors associated with the K largest eigenvalues of $D^{-1/2} W D^{-1/2}$ (Chung, 1997). Correspondingly, a discrete solution A of the original problem can be obtained by taking a rounding procedure from Z (e.g., using Kmeans or the approach proposed in Yu and Shi (2003)).

4.2. Spectral clustering with bag constraints

To incorporate the bag constraints, we need to define the two similarity functions in Eq. (1), the class-label similarity function $\Omega_L(\cdot)$ and the cluster-label similarity $\Omega_A(\cdot)$. Ideally, $\Omega_L(\cdot)$ should satisfy the following conditions: (1) In the case where class label information between two bags $\mathbb{B}_i$ and $\mathbb{B}_j$ is unambiguous, i.e., they do not share class label or they both belong to the same single class, $\Omega_L(i,j)$ should achieve minimum or maximum values; (2) In the ambiguous case where bags $\mathbb{B}_i$ and $\mathbb{B}_j$ have multiple labels and $\mathbb{Y}_i \cap \mathbb{Y}_j \neq \phi$, the smaller the quantity $\frac{|\mathbb{Y}_i \cap \mathbb{Y}_j|}{|\mathbb{Y}_i \cup \mathbb{Y}_j|}$ ($|\mathbb{Y}_i|$ is the number of distinct classes in $\mathbb{Y}_i$) is, i.e., the smaller the relative "common-label" set is, the smaller $\Omega_L(i,j)$ should be.

Based on the above considerations, we define the following class-label similarity function. Let y_i be the $C \times 1$ binary class indicator vector for bag $\mathbb{B}_i$, with elements $y_{ic} = 1/|\mathbb{Y}_i|$ if $c \in \mathbb{Y}_i$, and $y_{ic} = 0$ otherwise. Denote $Y = [y_1, \ldots, y_M]$, where $y_m = \mathbf{0}$ for any bag $\mathbb{B}_m$ that is not labeled. The *class-label similarity* between (i,j) is defined as

[1] In this paper, we assume that all instances are distinct.

[2] At this point, we do not specify the function forms of $\Omega_L(\cdot,\cdot)$ and $\Omega_A(\cdot,\cdot)$, since they can be problem-specified. However, this does not prevent us from viewing them as geometrical similarities.