Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Direct Discriminative Bag Mapping for Multi-Instance Learning

Jia Wu,[†] Shirui Pan,[†] Peng Zhang,[†] Xingquan Zhu[⋆]
[†] Quantum Computation & Intelligent Systems Centre, University of Technology Sydney, Australia
[⋆] Dept. of Computer & Electrical Engineering and Computer Science, Florida Atlantic University, USA
jia.wu@student.uts.edu.au; {shirui.pan, peng.zhang}@uts.edu.au; xzhu3@fau.edu;

Abstract

Multi-instance learning (MIL) is useful for tackling labeling ambiguity in learning tasks, by allowing a bag of instances to share one label. Recently, bag mapping methods, which transform a bag to a single instance in a new space via instance selection, have drawn significant attentions. To date, most existing works are developed based on the original space, *i.e.,* utilizing all instances for bag mapping, and instance selection is indirectly tied to the MIL objective. As a result, it is hard to guarantee the distinguish capacity of the selected instances in the new bag mapping space for MIL. In this paper, we propose a direct discriminative mapping approach for multi-instance learning (MILDM), which identifies instances to directly distinguish bags in the new mapping space. Experiments and comparisons on real-world learning tasks demonstrate the algorithm performance.

Introduction

In multi-instance learning (MIL), the learning object is a bag which contains a number of instances. The label (positive or negative) is assigned to the bag, but not individual instances. Such a setting makes MIL effective to accommodate labeling ambiguity in numerous real-world applications (Wu et al. 2014b).

To support MIL, one common approach is to convert the multi-instance learning (bag learning) to a traditional supervised learning (single-instance learning) problem. For example, one can propagate the bag label to instances inside the bag, so a propositional classifier can be learned for further classification (Foulds and Frank 2008). However, label transmission for positive bag may case some negative instances being assigned with wrong labels. Another transmission strategy is to use one instance to represent the related bag (*i.e., bag representation*), based on the statistic properties in the bag. In (Dong 2006) three different types of summarization approaches ("arithmetic mean", "geometric mean", and "minimax") have been proposed for bag representation. Wu et al. (2014a) also proposed another bag representation method based on the distribution of the negative bags. Although this type of single-instance representation algorithm works well in some prediction problems, it discards most of the instance information in the bags.

Alternatively, bag mapping approaches based on instance selection have been proposed (Chen, Bi, and Wang 2006) to map each bag to a new feature space constituted by an hidden instance set, *i.e.,* intermediate instance pool (IIP), constructed from training bag of instances. Following the IIP based bag mapping strategy, Fu et al. (2011) proposed another bag mapping method, which selects a subset of instances for bag-level feature computation based on the distribution of negative instances. For both methods, it is hard to guarantee the distinguish capability of selected instances in the new bag mapping space. In this paper, we propose a direct discriminative mapping approach for multi-instance learning (MILDM), which identifies instances to make the bags being directly distinguished in the new mapping space.

MIL with Discriminative Bag Mapping

Preliminaries and Optimization Problem

In multi-instance learning, a bag B_i consists of a set of instances, with $x_{i,j}$ denoting the jth instance in B_i. The class label of B_i is denoted by $y_i = \mathcal{Y}$, with $\mathcal{Y} = \{-1, +1\}$. The collection of all bags can be denoted by $\mathcal{B}$. In our proposed bag mapping framework, each bag B_i will be transformed to B_i^{ϕ}, a single instance in a new feature space by using the discriminative instance pool (DIP, denoted by $\mathcal{P}$). Given $\mathcal{B}$ with n bags, and the instance set $\mathcal{X}$ collected from all bags in $\mathcal{B}$, Our objective is to find a subset $\mathcal{P} \subseteq \mathcal{X}$ using an instance selection matrix $\mathcal{I}_{\mathcal{X}}$ (a diagonal matrix, $diag(\mathcal{I}_{\mathcal{X}}) = \mathbf{d}(\mathcal{X})$), where $\mathbf{d}(\mathcal{X})$ is a indicator vector, if $\mathbf{x}_i \in \mathcal{P}$, $\mathbf{d}(\mathcal{X})_i = 1$, otherwise 0. Accordingly, we define $\mathcal{J}(\mathcal{P})$ as an instance evaluation function to measure the $\mathcal{P}$ as follows:

$$\mathcal{P}_* = \arg\max_{\mathcal{P} \subseteq \mathcal{X}} \mathcal{J}(\mathcal{P}) \quad s.t.\ |\mathcal{P}| = m \tag{1}$$

where $|\cdot|$ denotes the cardinality of the instance set, and m is the number of instances to be selected from $\mathcal{X}$ (*i.e.,* size of DIP). The objective function in Eq. (1) represents that instances selected for MIL $\mathcal{P}_*$ should have maximum discriminative power in the new mapping space.

DIP Evaluation Criteria

To obtain the DIP with maximum discriminative power, we impose that the optimal DIP should have the properties: (a)

Bag Mapping Must-Link. Because each bag B_i is associated with a class label (positive or negative), the selected DIP should ensure that bags B_i^ϕ with the same label are similar to each other in the mapping space. (b) *Bag Mapping Cannot-Link*. For bags with different class labels in the mapping space, they should represent the disparity between them. Accordingly, DIP evaluation criteria could be measured as

$$\begin{aligned} \mathcal{J}(\mathcal{P}) &= \frac{1}{2}\sum\nolimits_{i,j} \left\|\mathcal{I}_\mathcal{X} B_i^\phi - \mathcal{I}_\mathcal{X} B_j^\phi\right\|^2 Q_{i,j} \\ &= tr(\mathcal{I}_\mathcal{X}^\top \mathcal{X}_\phi (D-Q) \mathcal{X}_\phi^\top \mathcal{I}_\mathcal{X}) \\ &= tr(\mathcal{I}_\mathcal{X}^\top \mathcal{X}_\phi L \mathcal{X}_\phi^\top \mathcal{I}_\mathcal{X}) \\ &= \sum\nolimits_{\mathbf{x}_k^\phi \in \mathcal{P}} \boldsymbol{\phi}_k^\top L \boldsymbol{\phi}_k \end{aligned} \tag{2}$$

where $S_{i,j}$ embeds class label information between two bags with $Q_{i,j} = \{-1/|A|, y_i y_j = 1; 1/|B|, y_i y_j = -1\}$, where $A = \{(i,j)|y_i y_j = 1\}$ denotes the *bag mapping must-link* pairwise bag constraint sets with $B = \{(i,j)|y_i y_j = -1\}$ denoting the *bag mapping cannot-link* pairwise sets. Besides, we adopt the bag to instance method in (Chen, Bi, and Wang 2006; Fu, Robles-Kelly, and Zhou 2011) for bag representation. Specifically, for a DIP $\mathcal{P}$ with m instances, B_i can be mapped to a single instance $B_i^\phi = [s(B_i, \mathbf{x}_1^\phi), \cdots, s(B_i, \mathbf{x}_m^\phi)]$, with $s(B_i, \mathbf{x}_k^\phi)$ denoting the similarity between the bag B_i and the kth instance $\mathbf{x}_k^\phi$. $\mathcal{X}_\phi = [B_1^\phi, \cdots, B_n^\phi] = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_m]^\top$. L is a Laplacian matrix generalized from Q. By doing so, in order to find the optimal instance set $\mathcal{P}$ which maximizes the criterion defined in Eq. (1), we can calculate the score of each instance (*i.e.,* $\boldsymbol{\phi}_k^\top L \boldsymbol{\phi}_k$) in $\mathcal{X}$ and then collect top-m instance to form the final discriminative instance pool (DIP).

After each $B_i \in \mathcal{B}$ is mapped to B_i^ϕ based on the optimal DIP, any generic single-instance learner can be applied for multi-instance learning.

Experiments

We carry out experiments on three real-world learning tasks: (a) content-based image annotation with 100 positive ("elephant" images) and 100 negative example images (Foulds and Frank 2008); (b) text categorization with DBLP data set being used to select papers published in Artificial Intelligence (AI: IJCAI, AAAI ,NIPS, UAI, COLT, ACL, KR, ICML, ECML and IJCNN) and Computer Vision (CV: ICCV, CVPR, ECCV, ICPR, ICIP, ACM Multimedia and ICME) fields to form an MIL learning task with 200 positive (AI) and 200 negative bags (Wu et al. 2013); and (c) train bound challenge predicting whether a train is eastbound or westbound, with a train (bag) containing a variable number of cars (instances). This dataset contains 10 eastbound trains as positive examples and another 10 westbound negative trains. After all bags are represented as instances, we use $k-$NN algorithm to train classification model (other approaches could also be applied, but because we are mainly focusing on bag representation issue, we omit detailed results from other classifiers).

For comparison purposes, we implement following baseline approaches. The first set of methods (*Instance Selection based Bag Representation*) directly use instances in the bags for representation, examples include MIWapper (Foulds and Frank 2008), MISM (Dong 2006) and MILIR (Wu et al. 2014a). The second set of methods IMBR *Instance Mapping based Bag Representation* use the intermediate instance pool (IIP) to map each bag to an instance in a new space. Both MILES (Chen, Bi, and Wang 2006) and MILIS (Fu, Robles-Kelly, and Zhou 2011) belong to the second group. Table 1 shows the classification accuracy of all comparison algorithms, in which the size of DIP is set to the same in Fu et al. (2011) (Our proposed method is denoted by MILDM). Overall, the instance mapping based approaches performs better than instance selection methods. For example, compared with other baselines, MILIS obtains the best performance 81.5% and 70.0% on image and train bound challenge data set, respectively. By using discriminative instance pool for bag mapping, MILDM consistently outperforms all baselines in all cases.

Table 1: Comparison of MILDM *w.r.t.* other approaches: %.

Data	**MILDM**	MILES	MILIS	MIWapper	MISM	MILIR
Elephant	**85.5**	77.0	81.5	77.5	76.0	70.5
DBLP	**72.5**	68.5	58.5	66.25	56.5	55.5
EastWest	**85.0**	65.0	70.0	45.0	65.0	55.0

Conclusion

In this paper, we proposed a direct discriminative bag mapping approach for multi-instance learning. Our method builds a discriminative instance pool to ensure bags in the new mapping space being directly distinguished from each other for MIL. Experiments demonstrate the effectiveness of our approach.

References

Chen, Y.; Bi, J.; and Wang, J. 2006. Miles: Multiple-instance learning via embedded instance selection. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(12):1931–1947.

Dong, L. 2006. A comparison of multi-instance learning algorithms. Technical report, University of Waikato.

Foulds, J., and Frank, E. 2008. Revisiting multiple-instance learning via embedded instance selection. In *AI*, 300–310.

Fu, Z.; Robles-Kelly, A.; and Zhou, J. 2011. Milis: Multiple instance learning with instance selection. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(5):958–977.

Wu, J.; Hong, Z.; Pan, S.; Zhu, X.; Cai, Z.; and Zhang, C. 2013. Multi-graph-view learning for graph classification. In *ICDM*, 1550–4786.

Wu, J.; Hong, Z.; Pan, S.; Zhu, X.; Zhang, C.; and Cai, Z. 2014a. Multi-graph learning with positive and unlabeled bags. In *SDM*, 217–225.

Wu, J.; Zhu, X.; Zhang, C.; and Yu, P. 2014b. Bag constrained structure pattern mining for multi-graph classification. *IEEE Trans. Knowl. Data Eng.* 26(10):2382–2396.