Joint Clustering and Classification for Multiple Instance Learning

Karan Sikka
ksikka@ucsd.edu
Ritwik Giri
rgiri@ucsd.edu
Marian Bartlett
mbartlett@ucsd.edu

University of California San Diego
La Jolla, California
USA

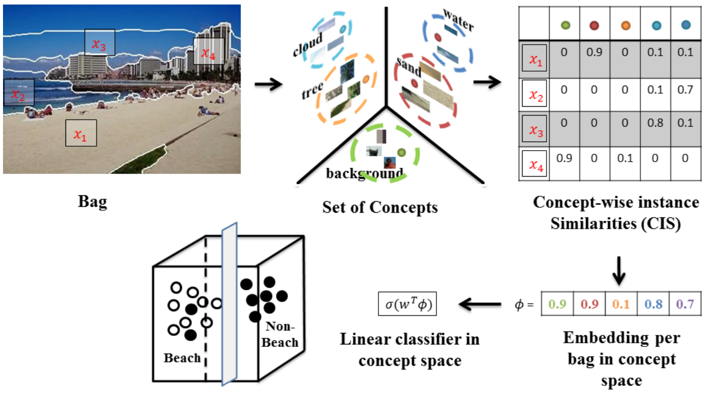

Figure 1: Illustrates the underlying idea in Embedding Space based MIL approaches. A beach scene, segmented into regions, is represented as a bag of instances, where each instance represents the corresponding region. A set of concepts is then used to calculate a similarity between each instance and the concept, referred to as concept-wise instance similarity (CIS). The *max* MIL assumption is used to embed each bag into the concept space using the CIS. Classification can then be performed in the embedding space using standard classifiers (best viewed in color). This work proposes to learn both the concept space and the classifier in a joint fashion.

1 Overview

The **Multiple Instance Learning** (MIL) has been extensively used to solve weakly supervised visual classification problems, where visual data is represented as a bag of instances. Previous MIL algorithms mostly extended standard supervised learning algorithms to the MIL settings by defining the classification score of a bag as the maximum score of each of its instances. Although these algorithms (referred to as Instance Space (IS) based methods) are popularly used, they do not account for the possibility that the instances may have multiple intermediate concepts. On the other hand, Embedding-space (ES) based MIL approaches are able to tackle this issue by defining a set of concepts, and then embedding each bag into a concept space, followed by training a standard classifier in the embedding space.

Contribution. In previous ES based approaches, the concepts were discovered separately from the classifier, and thus were not optimized for the final classification task. This work proposes a novel algorithm to estimate concepts and classifier parameters by jointly optimizing a classification loss. This approach discovers a set of discriminative concepts, which yield superior classification performance. The proposed algorithm is referred to as **Joint Clustering Classification for MIL** (JC^{2}MIL) because the discovered concepts induce clusters of data instances. Moreover, we show that the proposed algorithm achieves state-of-the-art results on several MIL datasets, by discovering fewer number of concepts compared to previous ES-based methods.

2 Methodology

Our target is to jointly learn (1) a set of concepts that are used to embed each bag into a concept space, and (2) a classifier that combines the embedding to produce a classification score. We achieve this by posing the problem as joint minimization of the classification loss with respect to both the set of concepts and the classifier parameters. The model makes two assumptions (1) the probability of a concept lying in a bag is maximum over the probability of each of its instances (similar to IS based methods), and (2) the similarity between k^{th} concept and an instance is defined using the rbf kernel (similar to [2, 3, 5] . The classification loss includes the mean negative log-likelihood and a regularization term, and is written as:

$$\mathcal{L}(B) = -\frac{1}{N}\sum_i (y_i \log p_i + (1-y_i)\log(1-p_i)) + \frac{\lambda}{2} w^T w \quad (1)$$

$$\{\mathcal{C}^*, w^*\} = \arg\min_{\mathcal{C},w} \mathcal{L}(B) \quad (2)$$

where C and w are the concept and classifier parameters, and p_i is the classifier score for each bag. The above optimization is solved using co-ordinate descent.

Method	Corel-2000
MI-SVM [1]	54.6: [53.1 63.1]
MILES [3]	68.7: [67.3 70.1]
DD-SVM [2]	67.5: [66.1 68.9]
k-means-SVM [4]	52.3: [51.6 52.9]
DMIL [5]	70.2: [68.3 72.1]
JC^{2}MIL (Ours)	**73.2**: [71.2 74.8]

Table 1: Evaluation (multiclass % accuracy) of Instance Space and Embedding Space based MIL algorithms on the Corel-2000 dataset [2] along with 95% confidence interval.

3 Evaluation

Experiments: The algorithm was evaluated on five MIL datasets- Tiger, Fox, Elephant, Corel-2000 and UCSB Breast Cancer dataset. The results for the Corel-2000 dataset are shown in Table 1 and extensive results on other datasets can be found in the paper.

Results: The proposed algorithm achieved state-of-the-art results on the MIL datasets. In addition, we showed that, for the Corel-2000 dataset, JC^{2}MIL outperformed its unsupervised variant (similar to BoW) by discovering a (relatively) small number of concepts. The best performance of JC^{2}MIL on Corel-2000 dataset was 73.2% with a concept size of 20, while the best performance for the unsupervised counterpart was 70.1% with a concept size of 3000. Thus joint learning of concepts and classifier leads to excellent results with a smaller number of discriminative concepts.

[1] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann. Support vector machines for multiple-instance learning. In *Advances in neural information processing systems*, pages 561–568, 2002.

[2] Yixin Chen and James Z Wang. Image categorization by learning and reasoning with regions. *The Journal of Machine Learning Research*, 5:913–939, 2004.

[3] Yixin Chen, Jinbo Bi, and James Ze Wang. Miles: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1931–1947, 2006.

[4] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *European Conference on Computer Vision Workshops*, volume 1, pages 1–2. Springer, 2004.

[5] A. Shrivastava, J.K. Pillai, V.M. Patel, and R. Chellappa. Dictionary-based multiple instance learning. In *IEEE International Conference on Image Processing*, pages 160–164, Oct 2014.