Multiple Instance Learning for Soft Bags via Top Instances

Weixin Li, Nuno Vasconcelos
University of California, San Diego, California, United States.

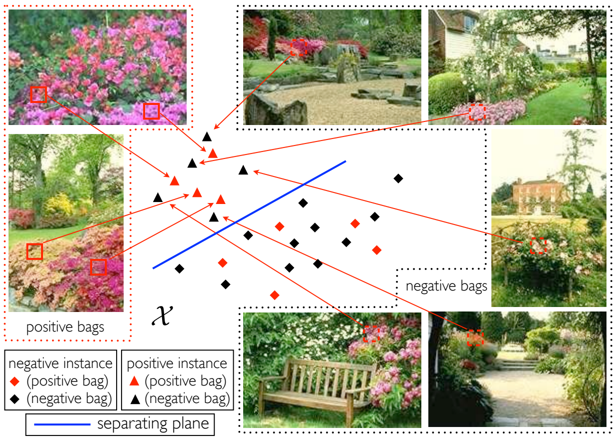

Figure 1: Noisy weak annotation of "flower" in Corel5k, where negative bags are far more than positive bags in training data.

Multiple instance learning (MIL) [2] is a family of learning algorithms suitable for problems involving substantial amounts of labeling noise. Examples, denoted as *instances* in MIL, are grouped into *bags*, and a label is attached to each bag. A bag that contains at least one positive example is considered a positive bag, otherwise it is negative. A classifier is finally designed to classify bags, rather than individual examples. While a substantial step towards the development of systems that can learn from very weakly labeled data, the current formulation of MIL fails to account for a very characteristic property of this type of data: *negative* bags can also have very noisy instance composition.

In fact, for some of the most popular applications of MIL, *e.g.*, semantic annotation of images or video, the labeling of negatives is as noisy as that of positives. Figure 1 illustrates the problem for the popular Corel5k dataset [1]. Image patches are considered instances of visual concepts and the goal is to classify images, which are bags of such patches, with concept labels. The figure shows examples of positive and negative bags for the concept "flower". Note that many of the negative images include regions of flowers. This labeling noise is common in weakly supervised learning, where human annotators are asked to label data with a few keywords. The absence of the "flower" label does not mean that there are no flowers in the image, just that the labeler did not think of "flower" as one of its *predominant* concepts. In result, negative bags frequently contain positive instances. This nullifies the core MIL assumption of clean negative bags.

To address this common issue, we consider a more general definition of MIL, which softens its constraint on negative bags. Under this formulation, both positive and negative bags are *soft* (*i.e.*, they can contain both positive and negative instances, differing only in their *composition* with regards to the two instance types). A soft bag B is a set of instances, or examples, $x \in \mathcal{X} \subseteq \mathbb{R}^D$. Instances are sampled independently from two distributions $p^+_{\mathcal{X}}(x)$ (denoted positive source) and $p^-_{\mathcal{X}}(x)$ (denoted negative source). The positive source is the distribution of the target concept (*e.g.*, image patches of "flower"), the negative source the distribution of background clutter (*e.g.*, image patches of everything else).

Definition 1. *A soft bag $B = \{x_i\}_{i=1}^{N_B}$ is a set of N_B instances, where $N_B \geqslant N$ and $N \in \mathbb{Z}_{++}$ is a lower bound on bag size, sampled as follows*
- sample N_B^+ from a probability mass function $p(N_B^+ = i)$ with $0 \leqslant i \leqslant N_B$;
- sample N_B^+ i.i.d. instances $\{x_i\}_{i=1}^{N_B^+}$ from the positive source $p^+_{\mathcal{X}}(x)$;
- sample $N_B^- = N_B - N_B^+$ i.i.d. instances $\{x_i\}_{i=N_B^+ +1}^{N_B}$ from the negative source.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

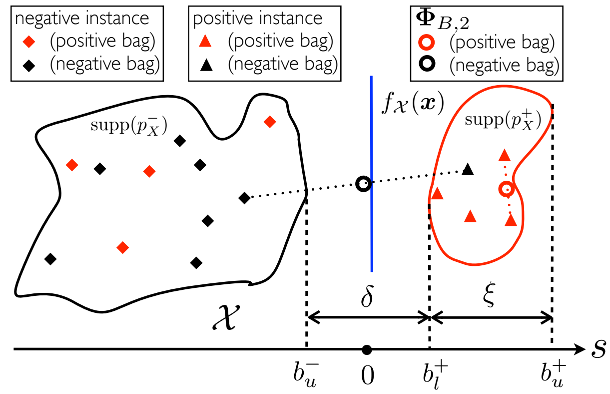

Figure 2: MIL for soft bags. The two contours delimit the support sets of the two sources. Positive and negative instances are identified by symbol shape, their bag ownerships by color (see insets). The top 2 instances are used to represent each of the two bags (one positive and one negative). These are the instances connected by a dotted line.

Definition 2. *Let $0 < \mu \leqslant N$ be a lower bound on the number of positive examples per positive bag. A soft bag B is μ-positive (label $+1$) if $N_B^+ \geqslant \mu$ and μ-negative (label -1) otherwise.*

Definition 3. *The soft bag classification problem is denoted separable if the regions of support of $p^+_{\mathcal{X}}(x)$ and $p^-_{\mathcal{X}}(x)$ are linearly separable, i.e., if there exists a linear prediction rule $f_{\mathcal{X}} : \mathcal{X} \mapsto \mathbb{R}$ such that $\forall\, x^+ \in supp(p^+_{\mathcal{X}})$, $x^- \in supp(p^-_{\mathcal{X}})$,*

$$f_{\mathcal{X}}(x^+) \geqslant 0 \geqslant f_{\mathcal{X}}(x^-). \tag{1}$$

$f_{\mathcal{X}}$ is denoted a separator of the soft bag problem.

The goal is to predict if a query bag B_q is μ-positive or μ-negative, *i.e.*, to learn a prediction rule $f : \mathcal{X}^{N_B} \mapsto \mathcal{Y}$, from a training dataset $\mathcal{D} = \{(B_i, y_i)\}$. Figure 2 illustrates a separable soft bag classification problem with $N = 8$. Denoting the positive (negative) bag by B_p (B_n), the figure depicts a scenario where $N_n^+ = 1$, and $N_p^+ = 4$.

A soft bag B is represented by a compound feature $\Phi_{f,k}$ of its k ($1 \leqslant k \leqslant N_B$) "most positive" instances (*top instances* of B)

$$\Phi_{f,k}(B) = \frac{1}{k}\sum\nolimits_{x_i \in \Omega^*_{f,k}(B)} x_i,\ \Omega^*_{f,k}(B) = \underset{\Omega \subseteq B, |\Omega| = k}{\operatorname{argmax}} \sum\nolimits_{x_i \in \Omega} f(x_i). \tag{2}$$

The application of the linear separator $f(x)$ to a bag B produces a bag score

$$s_{f,k}(B) = f\left(\frac{1}{k}\sum_{x_i \in \Omega^*_{f,k}(B)} x_i\right) = \frac{1}{k}\sum_{x_i \in \Omega^*_{f,k}(B)} f(x_i). \tag{3}$$

Given the separability between a positive bag B_p and a negative bag B_n as

$$\Delta s_{f,k}(B_p, B_n) = s_{f,k}(B_p) - s_{f,k}(B_n), \tag{4}$$

theoretical analysis shows that, both the expected and absolute separability can be guaranteed under reasonable assumptions. We propose a max-margin classifier that can be efficiently trained and effectively solve the soft bag classification problem. Our solution is shown to outperform conventional MIL methods in soft bag problems. Please refer to the paper for more details.

[1] Gustavo Carneiro, Antoni B Chan, Pedro J Moreno, and Nuno Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. *IEEE TPAMI*, 29(3):394–410, 2007.

[2] T.G Dietterich, R.H Lathrop, and T Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71, 1997.