Multiple Instance Learning with Manifold Bags

Boris Babenko[1] Nakul Verma[1] Piotr Dollar[2] Serge Belongie[1]

[1] Computer Science and Engineering
University of California, San Diego
{bbabenko,naverma,sjb}@cs.ucsd.edu

[2] Computer Science
California Institute of Technology
{pdollar}@caltech.edu

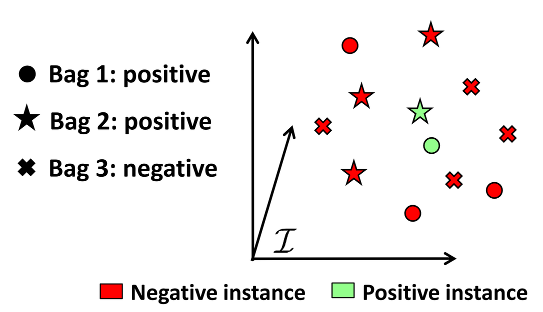
ABSTRACT

- Multiple Instance Learning is a relaxed form of supervised learning
  - » Learner receives labeled **bags** rather than labeled instances
  - » Reduces burden of collecting labeled data
- Existing analysis assumes bags have a **finite size**
- For many applications, bags are modeled better as **manifolds** in feature space; thus existing analysis is not appropriate
- In this setting we show:
  - » geometric structure of manifold bags affects PAC learnability
  - » a MIL algorithm that learns from finite sized bags can be trained with manifold bags
  - » a simple heuristic algorithm for reducing memory requirements

MULTIPLE INSTANCE LEARNING (MIL)

DEFINITION

MIL: relaxed form of supervised learning
- » (**set of examples**, label) pairs provided
- » MIL lingo: set of examples = **bag** of instances
- » Bag labeled positive if at least one instance in bag is positive



EXISTING ANALYSIS

Data model (bottom up)
- » Draw r instances and their labels from fixed distribution $\mathcal{D}_\mathcal{I}$
- » Create bag from instances, determine its label (max of instance labels)
- » Return bag & bag label to learner

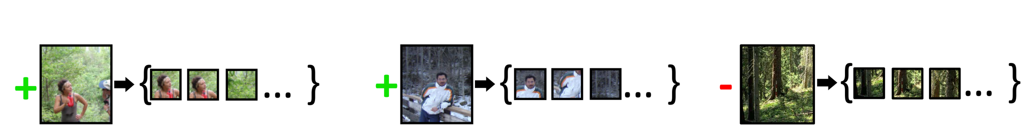
Blum & Kalai (1998)
- » If: access to noise tolerant instance learner, instances drawn independently
- » Then: bag sample complexity **linear** in r

Sabato & Tishby (2009)
- » If: can minimize empirical error on bags
- » Then: bag sample complexity **logarithmic** in r

APPLICATIONS

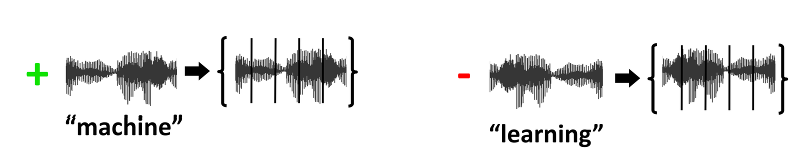
Object Detection (images)

Instance: image patch
Instance Label: is face?
Bag: whole image
Bag Label: contains face?



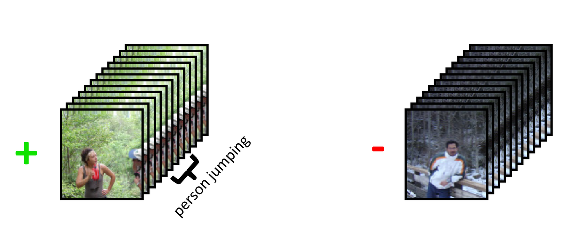
Phoneme Detection (audio)

Instance: audio clip
Instance Label: is 'sh'?
Bag: whole audio
Bag Label: contains 'sh'?



Event Detection (video)

Instance: video clip
Instance Label: is 'jump'?
Bag: whole video
Bag Label: contains 'jump'?



OBSERVATIONS

- Top down process: draw entire bag from a **bag distribution**, then get instances
- Instances of a bag lie on a **manifold**
- Potentially infinite number of instances per bag -- existing analysis inappropriate
- Expect sample complexity to scale with **manifold** parameters (curvature, dimension, volume, etc)

MANIFOLD BAGS

FORMULATION

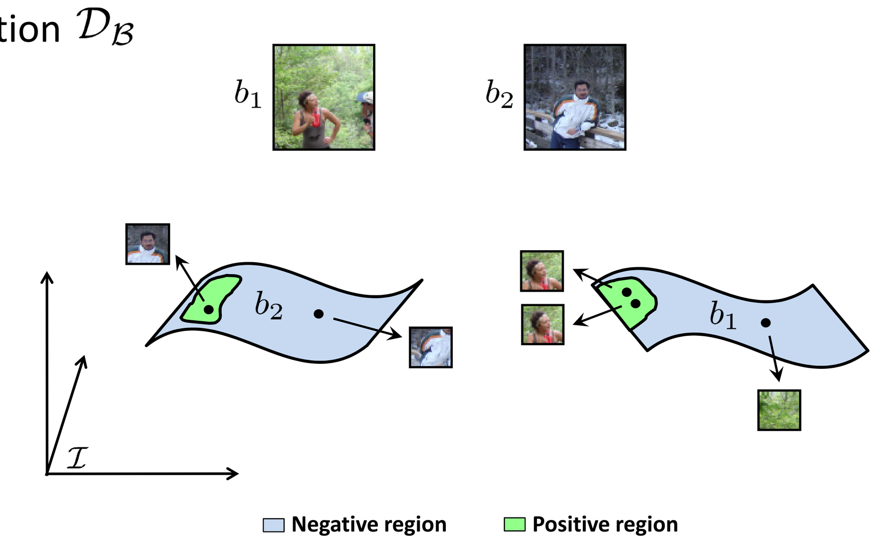
- Manifold bag b drawn from **bag** distribution $\mathcal{D}_\mathcal{B}$
- Instance hypotheses:

$$h \in \mathcal{H}, \quad h : \mathcal{I} \to \{0,1\}$$

- Corresponding bag hypotheses:

$$\bar{h} \in \overline{\mathcal{H}}, \quad \bar{h} : \mathcal{B} \to \{0,1\}$$

$$\bar{h}(b) \stackrel{\text{def}}{=} \max_{x \in b} h(x)$$



GENERALIZATION BOUNDS

VC DIMENSION

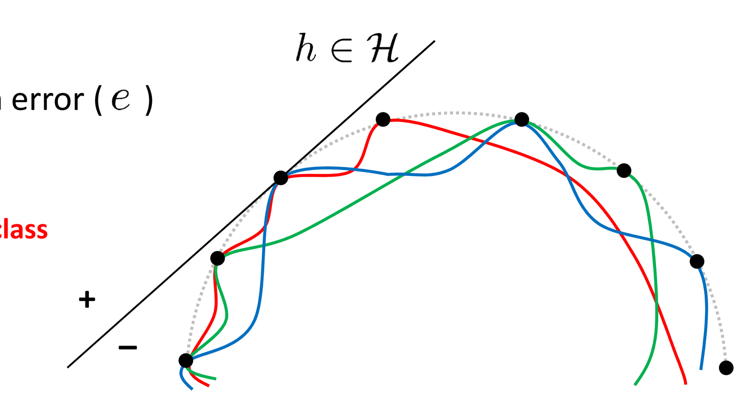
- A way of relating **empirical** error ($\hat{e}$) to **generalization** error (e)
- Standard bound:

$$e \le \hat{e} + O\left(\sqrt{\frac{VC(\overline{\mathcal{H}})}{m}}\right)$$ ← VC Dimension of bag hypothesis class

- Want to relate $VC(\mathcal{H})$ to $VC(\overline{\mathcal{H}})$
- For **finite** sized bags (Sabato & Tishby 2009):

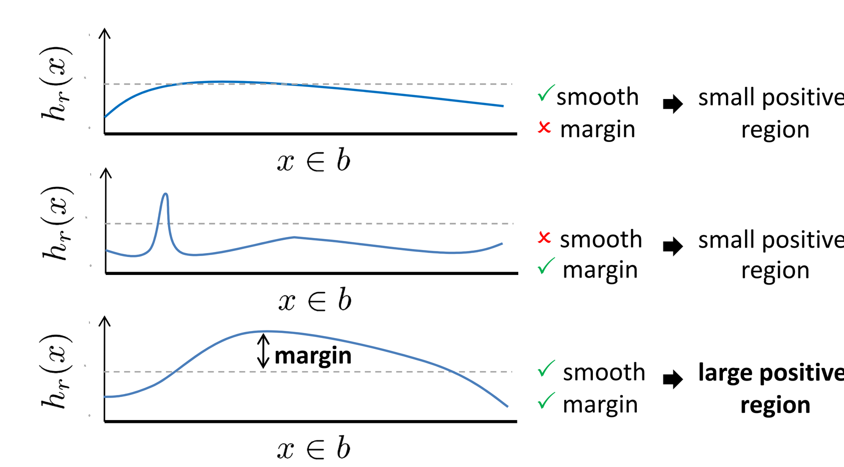
$$VC(\overline{\mathcal{H}}) \le VC(\mathcal{H}) \log(r)$$

- Turns out $VC(\overline{\mathcal{H}})$ is unbounded even for arbitrarily smooth bags



TAMING THE RICHNESS

- Bag hypothesis class too powerful: for positive bag, need to only classify one instance as positive
- Infinitely many instances -> too much flexibility for bag hypothesis
- Would like to ensure a **non-negligible portion** of positive bags is labeled **positive**
- Solution:
  - » Switch to real-valued hypothesis class $h_r \in \mathcal{H}_r : \mathcal{I} \to [0,1]$
  - » h_r must be Lipschitz smooth w.r.t. $\mathcal{I}$
  - » h_r must label bags with a **margin**



FAT SHATTERING DIMENSION

- Fat shattering dimension relates empirical error at **margin** γ to generalization error
- Standard bound:

$$e \le \hat{e}_\gamma + O\left(\sqrt{\frac{\mathrm{F}_{\gamma/8}(\overline{\mathcal{H}}_r)}{m}}\right)$$ ← Fat shattering dimension of bag hypothesis class

$$e \le \hat{e}_\gamma + O\left(\sqrt{\frac{n^2 \mathrm{F}_{\gamma/16}(\mathcal{H}_r)}{m} \log^2\left(\frac{Vm}{\gamma^2 \kappa^n}\right)}\right)$$

- Unlike VC, we **can** relate $\mathrm{F}_\gamma(\overline{\mathcal{H}}_r)$ to $\mathrm{F}_\gamma(\mathcal{H}_r)$
- Key quantities: empirical error at margin γ ($\hat{e}_\gamma$), number of training bags (m), manifold bag dimension (n), manifold bag volume (V), smoothness (κ)

QUERYING INSTANCES

- In practice, learner can only access small number of instances (ρ)
- Same bound holds with increased failure probability δ

$$\rho \ge \Omega\left(\left(V/\kappa^n\right)\left(n + \ln\left(\frac{mV}{\kappa^n \delta}\right)\right)\right)$$

TAKE-HOME MESSAGE

- Increasing m reduces **complexity term**
- Increasing ρ reduces **failure probability**
  - » Seems to contradict previous results (smaller bag size r is better)
  - » Important difference between r and ρ !
  - » If ρ is small, may only get negative instances from a positive bag
- Increasing m requires **extra labels**, increasing ρ does not

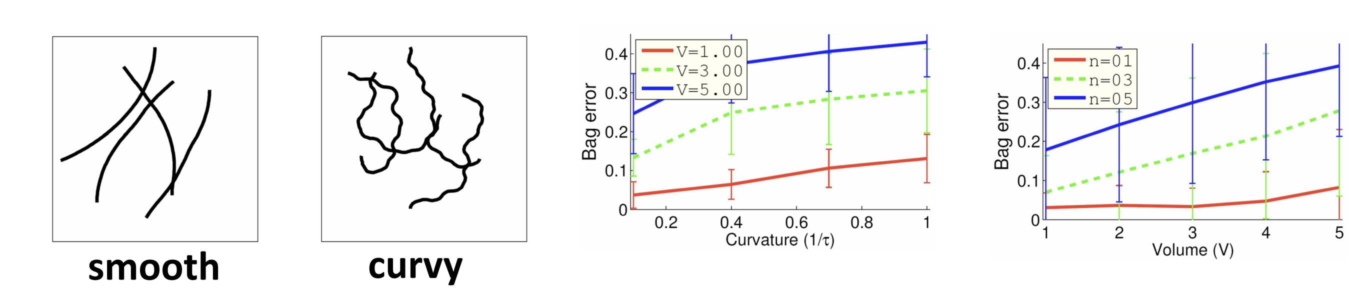
TRAINING WITH MANY INSTANCES

- Problem: want many instances/bag, but have computational limits
- Solution: Iterative Querying Heuristic (IQH)
  - » Grab small number of instances/bag, run standard MIL algorithm
  - » Query more instances from each bag, only keep the ones that get high score from current classifier

- At each iteration, train with small # of instances

EXPERIMENTS

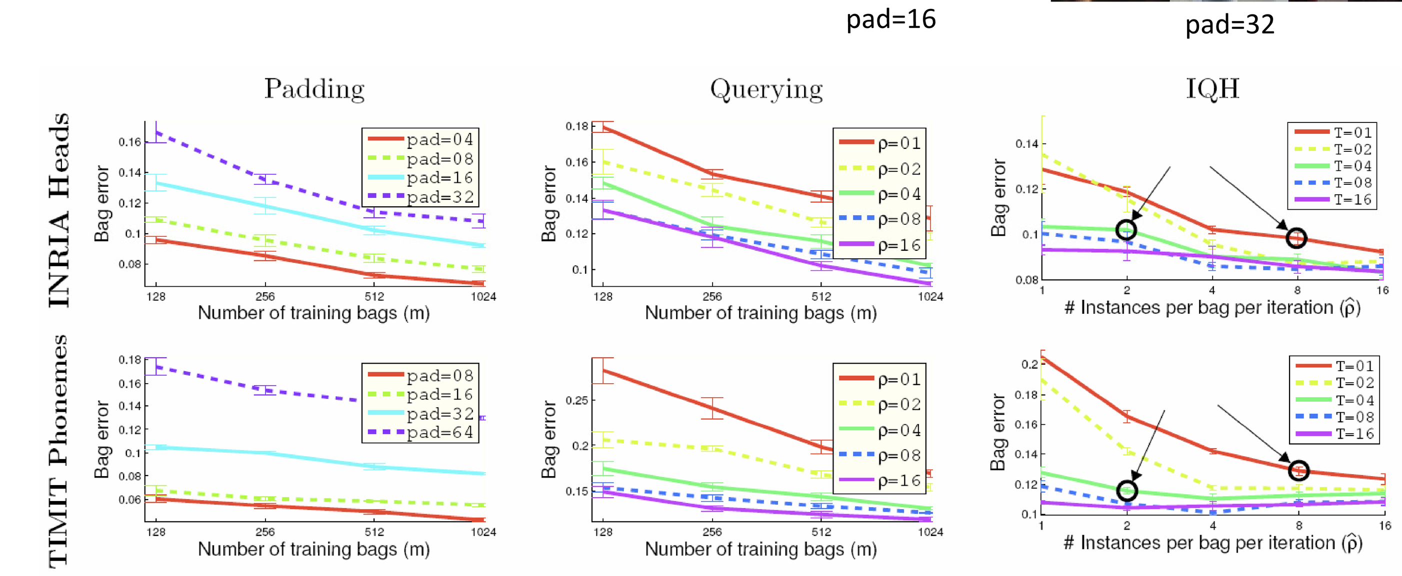
SYNTHETIC DATA



- Error scales with curvature and volume

REAL DATA

- INRIA Heads (Dalal et al. '05)
- TIMIT Phonemes (Garofolo et al., '93)



- Error scales with number of training bags, volume, number of queried instances and number of IQH iterations.