Multiple Instance Learning: Semisupervised Object Detection

Corring

March 26, 2015
Where is ?

...Imagine you don’t know Waldo!
An unsupervised problem hiding in a detection problem

Which ‘thing’ in the training data is the object to detect?

- Training samples are typically images
- Goal is to learn associations from parts of the image: *windowing*
- But the problem remains with strategies like this: *which window(s) to associate with the fact that the sample is positive?*

Reserve “instance” for a general window, a part of the data that needs a label that may not necessarily coincide with supervised training data
The Multiple Instance Classification Problem

Suppose the domain is an $n-$fold product of feature spaces

$$D = \times_{j=1}^{n} X_j$$

with samples (or bags)

$$x_i = (x_{i1}, \ldots, x_{in}).$$

Then we say that the $x_{ij}$ are instances and one may attempt to solve the multiple instance problem: given a collection of pairs of samples $x_i \in D$ with labels $y_i \in \mathcal{L}$ (and information about the spaces $X_j$) return a collection of functions

$$f_j : X_j \rightarrow \mathcal{L}$$

from the feature spaces to the label space.
Structure: The Usual Structure

Assumptions (or M.O):

- Bags are iid; effect: ML/MAP algorithms are almost workable.
- $X = X_j$ and the instances are drawn iid given a bag; effect: only need to learn one concept. MAP/ML gets easier.
- Want to learn $f : X \rightarrow \mathcal{L}$ or (possibly simpler)

$$f_{MI} : X \times \ldots \times X \rightarrow \mathcal{L}$$

via (eg. an OR gate).
MI-SVM problem can be written as a mixed integer program [1]

$$\min_{w,b,y} \frac{1}{2}\|w\|^2 + C \sum_{i,j} \xi_{iI(j)}$$

The constraints enforce the idea that an instance should reflect knowledge about the sample.
MI-SVM Equation

The MI-SVM problem can be written as a mixed integer program [1]

$$\min_{w,b,y} \frac{1}{2}||w||^2 + C \sum_{i,j} \xi_{iI(j)}$$

such that

$$y_{iI(j)} \left( w^T f_{iI(j)} + b \right) - 1 + \xi_{iI(j)} \geq 0,$$

The constraints enforce the idea that an instance should reflect knowledge about the sample.
The MI-SVM Equation

MI-SVM problem can be written as a mixed integer program [1]

\[
\min_{w,b,y} \frac{1}{2} \|w\|^2 + C \sum_{i,j} \xi_{iI(j)} \\
\text{such that } y_{iI(j)} (w^T f_{iI(j)} + b) - 1 + \xi_{iI(j)} \geq 0, \\
\exists j \in I : y_{iI(j)} = 1 \forall I : t_i = 1.
\]

The constraints enforce the idea that an instance should reflect knowledge about the sample.
The MI-SVM Picture

- Instances (windows) represented by circles
- Image class / Assigned instance class
- Instances in the same image are connected by link
- Still want structure of SVM, but want to exploit “commonalities” among positive images
- MI-SVM handles this unsupervised problem
The mixed integer program for the MI-SVM can’t be solved directly, however

- The problem above is a mixed integer program (NP-Hard!)
- We need to make assumptions or add structure in a way to improve on a base heuristic algorithm:
  - Initialize the instance labels somehow
  - Run an SVM with some level of tolerance
  - Use the margins to rank the instances and assign new labels based on these
  - Simple algorithm, lots of parameters and somewhat hard to predict on smaller datasets
Figure: The samples consist of 15 instances; drawn from 45 prototypes with noise. There are 3 positive prototypes. For positive samples, at least 1 (but fewer than 3) of the instances are from these positive prototypes. This video shows the iterations of the MI-SVM gradually converging to a good state.
**Figure:** This is the ROC curve from the running example. This synthetic case is a bit simplistic.
Figure: In this case there were 25 instances with only 2 positive instances in a given positive bag; I also moved the prototypes closer together. Need strategies to mitigate the issues encountered here.
Questions?
An Alternative: Diverse Density

Under the distribution assumptions outlined above AND assuming that the bags are conditionally independent given the target point $t$ we can take a maximum a posteriori approach (with a uniform prior on $x$—begin explicit) we get:

$$\arg \max_x \prod_i Pr(B_i^+ | x = t) \prod_j Pr(B_j^- | x = t)$$

and using Bayes rule (and the prior again) we get

$$c = \arg \max_x \prod_i Pr(x = t | B_i^+) \prod_j Pr(x = t | B_j^-)$$

now we just need a model of $Pr(x = t | B_i^+)$ in terms of instances. [2]
Two sided Noisey Learning $\leq^T_p$ MI Learning

Independence, Chernoff, Ignoring multiple instances, 1 sided $< 2$ sided. [3]
Also Blum (98)
Structure: Alternatives


P. Long and L. Tan, “Pac-learning axis aligned rectangles with respect to product distributions from multiple-instance examples,” 1996.