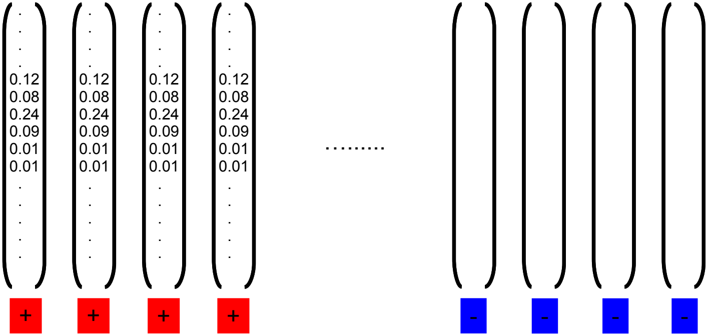
Support Vector Machines for Multiple-Instance Learning

S. Andrews, I. Tsochantaridis and T. Hofmann

presented by Y. Chai

June 7, 2011

Supervised Learning



- Given: a training set of labeled patterns, $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times Y$, i.i.d.
- Target: a classifier, i.e., a function to map patterns to labels $f : \mathbb{R}^d \to Y$

Multiple Instance Learning (MIL)

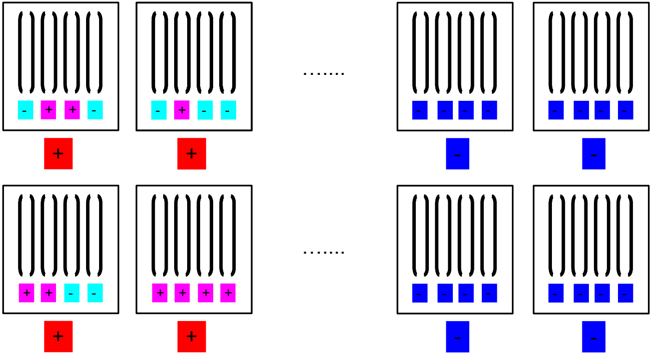
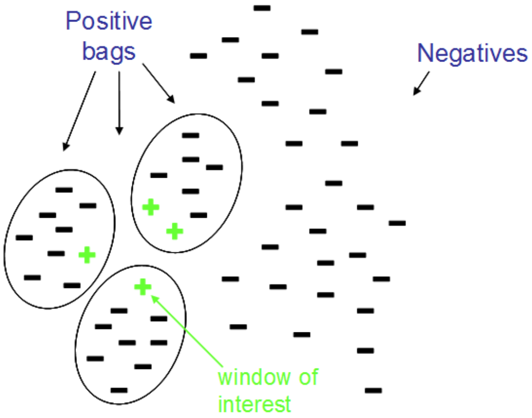


Figure: Plus(red), minus(blue). Hidden plus(purple), hidden minus(cyan)

Multiple Instance Learning (MIL)



Multiple Instance Learning (MIL)

Given:

- a training set of **patterns** $\mathbf{x}_1, \cdots, \mathbf{x}_n$, grouped into **bags** $\mathbf{B}_1, \cdots, \mathbf{B}_m$, with $\mathbf{B}_I = \{\mathbf{x}_i : i \in I\}$ for $I \subset \{1, \cdots, n\}$.
- each of the bags has a label $\mathbf{Y}_I$.
- if $\mathbf{Y}_I = -1$, all patterns in the particular bag have -1.

$$\sum_{i \in I} \frac{y_i + 1}{2} = 0$$

- if $\mathbf{Y}_I = 1$, **at least** one pattern in the particular bag has 1.

$$\sum_{i \in I} \frac{y_i + 1}{2} \leq 1$$

Multiple Instance Learning (MIL)

Some application areas:

- drug design: molecule conformations
- image retrieval
- information retrieval from documents

Multiple Instance Learning (MIL)

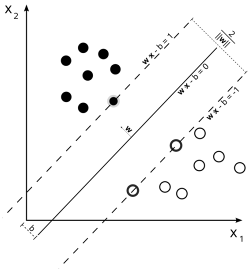
Some application areas:

- drug design: molecule conformations
- image retrieval
- information retrieval from documents

The authors introduced two MIL formulations:

- Maximum Pattern Margin Formulation (mi-SVM)
- Maximum Bag Margin Formulation (MI-SVM)

Support-Vector-Machine (SVM)



SVM with ℓ^2-regularizer and hinge loss:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \xi_i$$

s.t. $\forall i : y_i(\langle \mathbf{w}, \mathbf{x} \rangle + b) \geq 1 - \xi_i$, $\xi_i \geq 0$, $y_i \in \{-1, 1\}$

Maximum Pattern Margin Formulation (mi-SVM)

We look for a discriminant, where:

- all patterns from negative bags are on the negative halfspace.
- at least one pattern from every positive bag is on the positive halfspace.
- maximized the margin with respect to the entire set with some patterns from every positive bag turning to negative.

Maximum Pattern Margin Formulation (mi-SVM)

We look for a discriminant, where:

- all patterns from negative bags are on the negative halfspace.
- at least one pattern from every positive bag is on the positive halfspace.
- maximized the margin with respect to the entire set with some patterns from every positive bag turning to negative.

$$\min_{y_i} \min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \xi_i$$

s.t. $\forall i : y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i$, $\xi_i \geq 0$, $y_i \in \{-1, 1\}$, and the conditions for MIL hold.

Maximum Pattern Margin Formulation (mi-SVM)

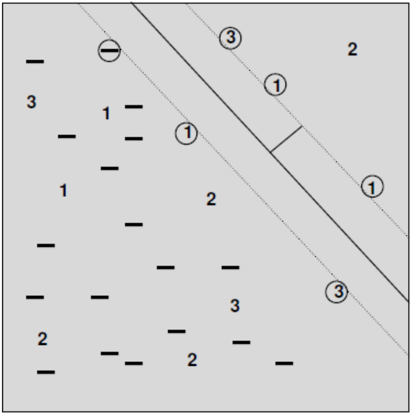


Figure: '-': patterns from negative bags. 'i': patterns from the i-th bag, which is positive.

Maximum Bag Margin Formulation (MI-SVM)

We look for a discriminant, where:

- all patterns from negative bags are on the negative halfspace.
- for positive bag, only the pattern with the largest margin has an impact on training. Rest is ignored.
- maximized the bag margin instead of pattern margin.

Maximum Bag Margin Formulation (MI-SVM)

We look for a discriminant, where:

- all patterns from negative bags are on the negative halfspace.
- for positive bag, only the pattern with the largest margin has an impact on training. Rest is ignored.
- maximized the bag margin instead of pattern margin.

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_I \xi_I$$

s.t. $\forall I : Y_I \max_{i \in I}(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_I$, $\xi_I \geq 0$

Maximum Bag Margin Formulation (MI-SVM)

We look for a discriminant, where:

- all patterns from negative bags are on the negative halfspace.
- for positive bag, only the pattern with the largest margin has an impact on training. Rest is ignored.
- maximized the bag margin instead of pattern margin.

$$\min_{s} \min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{I} \xi_I$$

$$\text{s.t. } \forall I \quad \begin{array}{lcll} Y_I = -1 & \bigwedge & -\langle \mathbf{w}, \mathbf{x}_i \rangle - b \geq 1 - \xi_I, & \forall i \in I \text{ and } \xi_I \geq 0 \\ Y_I = 1 & \bigwedge & \langle \mathbf{w}, \mathbf{x}_{s(I)} \rangle + b \geq 1 - \xi_I, & \text{and } \xi_I \geq 0 \end{array}$$

Maximum Bag Margin Formulation (MI-SVM)

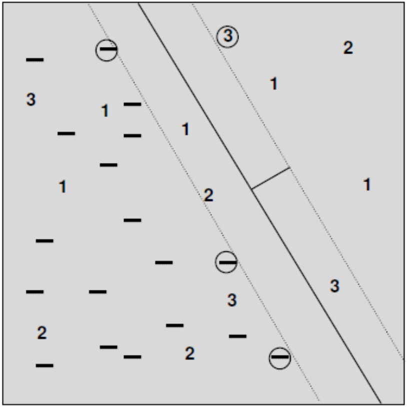


Figure: '-': patterns from negative bags. 'i': patterns from the i-th bag, which is positive.

Optimization Heuristics

For mi-SVM:

$$\min_{y_i} \min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \xi_i$$

s.t. $\forall i : y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i$, $\xi_i \geq 0$, $y_i \in \{-1, 1\}$, and the conditions for MIL hold.

```
initialize y_i = Y_I  for  i ∈ I
REPEAT
    compute SVM solution w, b for data set with imputed labels
    compute outputs f_i = ⟨w, x_i⟩ + b for all x_i in positive bags
    set y_i = sgn(f_i) for every i ∈ I, Y_I = 1
    FOR (every positive bag B_I)
        IF (∑_{i∈I}(1 + y_i)/2 == 0)
            compute i* = arg max_{i∈I} f_i
            set y_{i*} = 1
        END
    END
WHILE (imputed labels have changed)
OUTPUT (w, b)
```

Optimization Heuristics

For MI-SVM:

$$\min_{s} \min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{I} \xi_I$$

$$\text{s.t. } \forall I \quad \begin{array}{lll} Y_I = -1 & \wedge & -\langle \mathbf{w}, \mathbf{x}_i \rangle - b \geq 1 - \xi_I, \quad \forall i \in I \text{ and } \xi_I \geq 0 \\ Y_I = 1 & \wedge & \langle \mathbf{w}, \mathbf{x}_{s(I)} \rangle + b \geq 1 - \xi_I, \quad \text{and } \xi_I \geq 0 \end{array}$$

initialize $\mathbf{x}_I = \sum_{i \in I} \mathbf{x}_i / |I|$ for every positive bag B_I
REPEAT
  compute QP solution $\mathbf{w}, b$ for data set with
    positive examples $\{\mathbf{x}_I : Y_I = 1\}$
  compute outputs $f_i = \langle \mathbf{w}, \mathbf{x}_i \rangle + b$ for all $\mathbf{x}_i$ in positive bags
  set $\mathbf{x}_I = \mathbf{x}_{s(I)}$, $s(I) = \arg\max_{i \in I} f_i$ for every I, $Y_I = 1$
WHILE (selector variables $s(I)$ have changed)
OUTPUT $(\mathbf{w}, b)$

To take home

- the multiple-instance-learning model
- the ideas behind the mi-SVM and MI-SVM formulation
- optimization heuristics towards mi-SVM and MI-SVM

To take home

- the multiple-instance-learning model
- the ideas behind the mi-SVM and MI-SVM formulation
- optimization heuristics towards mi-SVM and MI-SVM

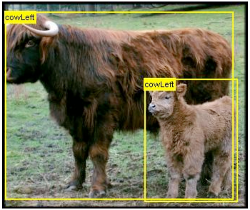
Oh God, this presentation is toooo short!!!

Felzenszwalb et al. [CVPR08]

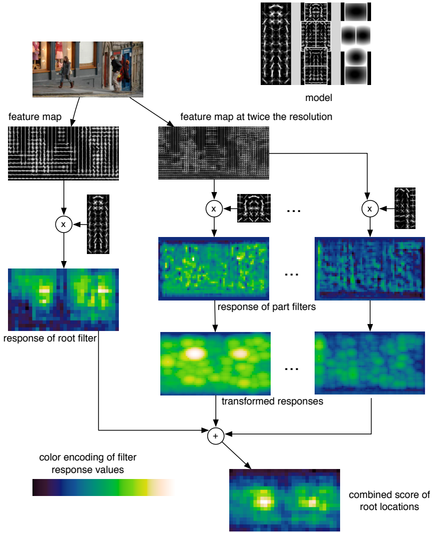
A Discriminatively Trained, Multiscale, Deformable Part Model

(images are taken from Felzenszwalb et al.'s slides)

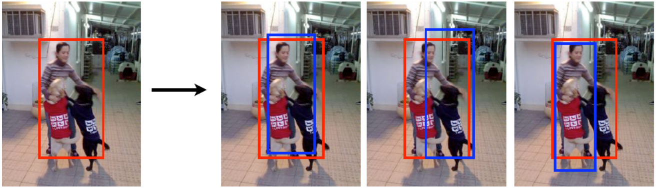
- PASCAL Challenge (object detection)
- objects are provided with labeled bounding boxes
- won the PASCAL VOC "Lifetime Achievement" prize



Felzenszwalb et al. [CVPR08]



Felzenszwalb et al. [CVPR08]



Training algorithm (where MIL comes into play):

- root filter initialization
- root filter update (in a MIL fashion)
- part initialization
- model update (in a MIL fashion)

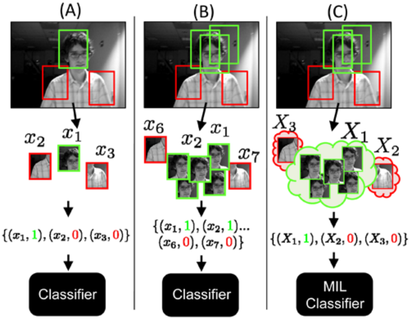
Babenko et al. [CVPR09]

Visual Tracking with Online Multiple Instance Learning

- tracking by detection
- discriminative classifier in an online manner to separate the object from the background
- MIL involved in the training

Babenko et al. [CVPR09]

Visual Tracking with Online Multiple Instance Learning
(image is taken from Babenko et al.'s project page)



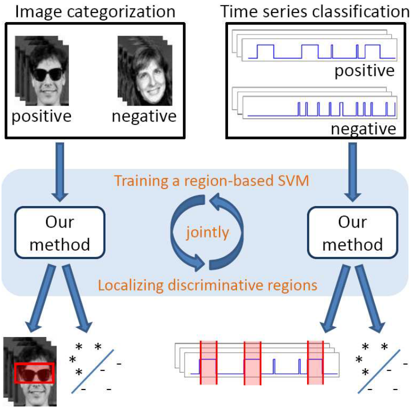
Nguyen et al. [ICCV09]

Weakly supervised discriminative localization and classification: a joint learning process

- localization-classification SVM
- training set consists of weakly supervised images with only labels given
- subwindows of each image correspond to patterns with images being the bags
- algorithm tries to jointly segment and classify the images
- Lampert, Blaschko and Hofmann [ECCV2008] have made this method computationally possible

Nguyen et al. [ICCV09]

Weakly supervised discriminative localization and classification: a joint learning process
(image is taken from Nguyen et al.'s paper)



Conclusion continued

We have seen:

- high applicability of MIL in many computer vision areas, e.g. detection, tracking and segmentation
- hot topic in the recent years
- besides MI-SVM, MIL can be also applied to boosting, e.g. Galleguillos et al. [ECCV08] and Viola et al. [NIPS07]