LAMDA
Learning And Mining from DatA
http://lamda.nju.edu.cn

南京大學

Multi-Instance Learning Revisited
（多示例学习回顾）

Zhi-Hua Zhou

http://cs.nju.edu.cn/zhouzh/
Email: zhouzh@nju.edu.cn

LAMDA Group
National Key Laboratory for Novel Software Technology,
Nanjing University, China

The talk involves some joint work with my students :

Min-Ling Zhang (张敏灵)

Yu-Feng Li (李宇峰)

Jun-Ming Xu (眭俊明)

Yu-Yin Sun (孙雨音)

Xiao-Bing Xue (薛晓冰)

… …

And my collaborators:

James Kwok

Ivor Tsang

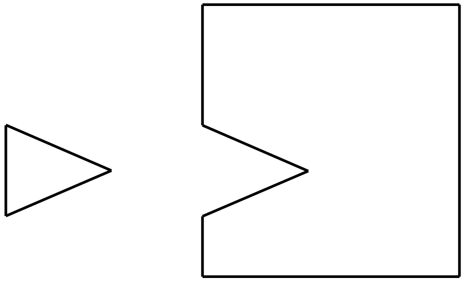
… …

Origin

The Motivating Problem

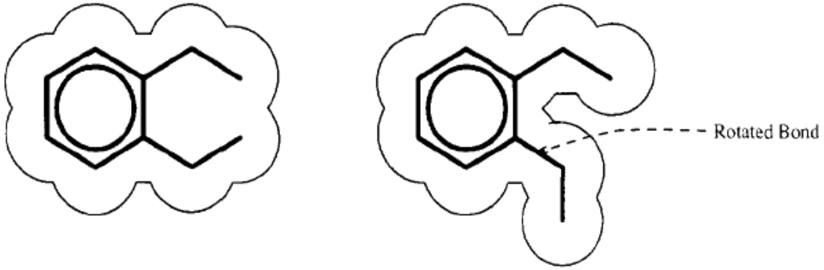
Originated from the research on drug activity prediction
[Dietterich et al. AIJ97]

- Drugs are small molecules working by binding to the target area
- For molecules qualified to make the drug, one of its shapes could tightly bind to the target area



The Motivating Problem (con't)

- A molecule may have many alternative shapes



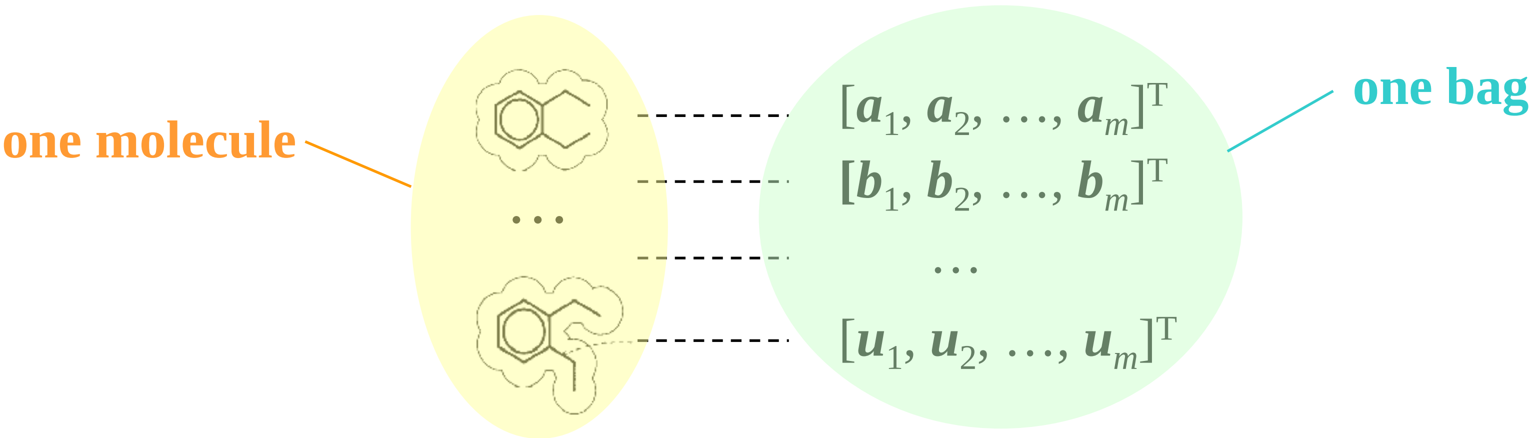
Reprinted from [Dietterich et al., AIJ97]

The difficulty:

Biochemists know that whether a molecule is qualified or not, but do not know which shape responses for the qualification

To Represent the Molecule

Each shape can be represented by a feature vector, i.e., an instance



Thus, a molecule is a **bag** of instances

- A bag is positive if it contains at least one positive instance; otherwise it is negative
- The labels of the training bags are known
- The labels of the instances in the training bags are unknown

Formal Definition of MIL

Given a data set $\{(X_1, y_1), \cdots, (X_i, y_i), \cdots, (X_N, y_N)\}$, where $X_i = \{\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{ij}, \cdots, \boldsymbol{x}_{i,n_i}\} \subseteq \mathcal{X}$ is called a bag and $y_i \in \mathcal{Y} = \{-1, +1\}$ is the label of X_i, the goal is to predict the labels of unseen bags

X_i is a positive bag (thus $y_i = +1$) if there exists $g \in \{1, \cdots, n_i\}$, $\boldsymbol{x}_{ig}$ is positive. Yet the value of the index g is unknown

$\boldsymbol{x}_{ij} \in \mathcal{X}$ is an instance $[x_{ij1}, \cdots, x_{ijl}, \cdots, x_{ijd}]'$,
x_{ijl} is the value of $\boldsymbol{x}_{ij}$ at the l-th attribute

Why Multi-Instance Learning?

Can we solve this problem with traditional supervised learning?

Possible solution 1: Assign the bag labels to instances

+

$[a_1, a_2, \ldots, a_m]^T$ +

$[b_1, b_2, \ldots, b_m]^T$ +

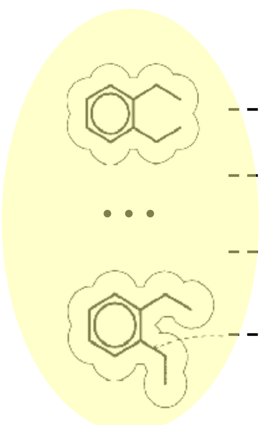
… …

$[u_1, u_2, \ldots, u_m]^T$ +

Maybe most instances in the positive bag are actually negative

Why Multi-Instance Learning? (con't)

Possible solution 2: Concatenate the instances in a bag



$[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m]^{\mathrm{T}}$

$[\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m]^{\mathrm{T}}$

…

$[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m]^{\mathrm{T}}$

$\Rightarrow [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_m, \ldots, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]^{\mathrm{T}}$

?

$[\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_m, \ldots, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^{\mathrm{T}}$

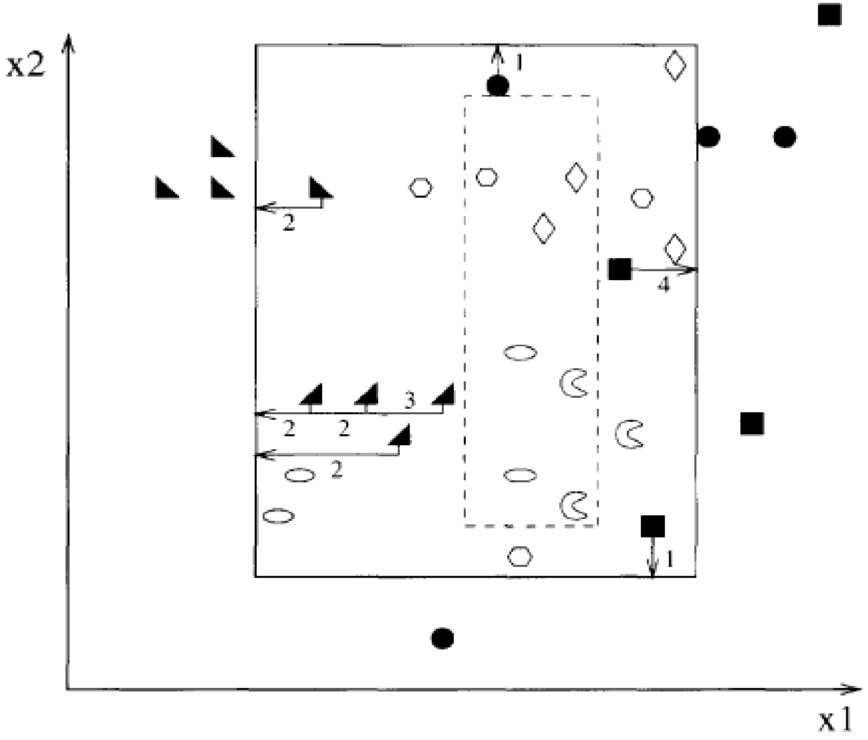
Which order should be used?

Moreover, different bags may contain different number of instances

Learning And Mining from DatA
http://lamda.nju.edu.cn

Three APR (Axis-Parallel Rectangle) algorithms

Iterated-discrim APR



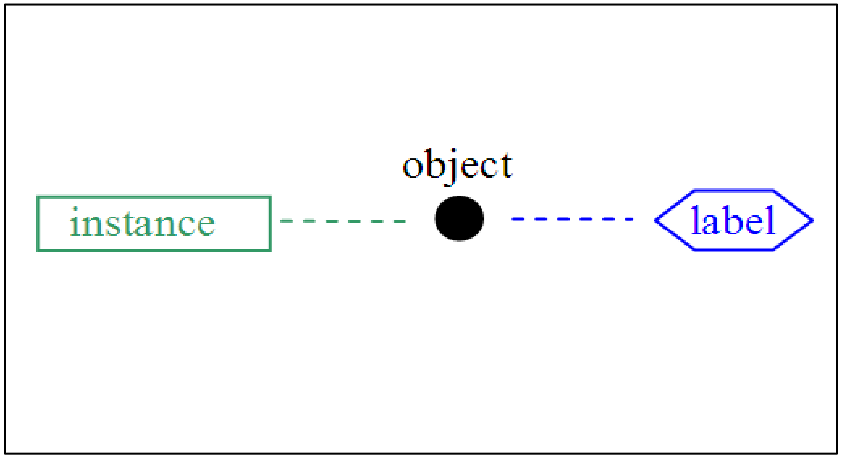
- Iterated-discrim APR achieves the best performance on the MUSK data (92.4% / 89.2%)
- BP neural network and C4.5 decision tree cannot work well

    BP (75.0% / 67.7%)

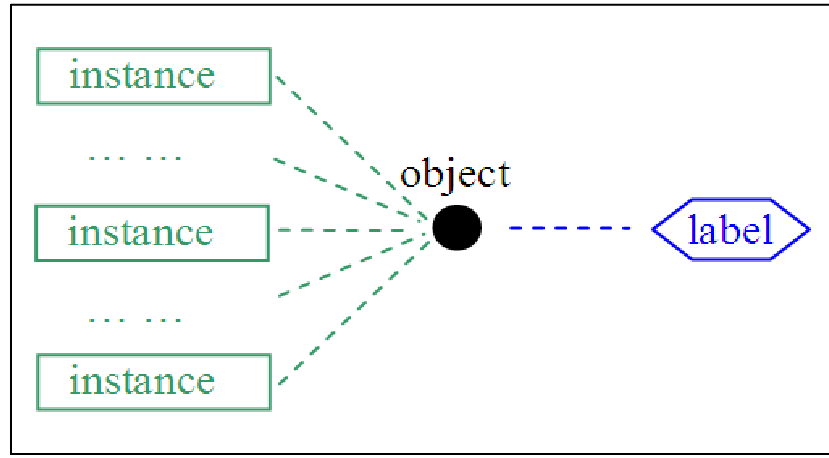
    C4.5 (68.5% / 58.8%)

Figure reprinted from [Dietterich et al., AIJ97]

Comparing with Traditional Supervised Learning



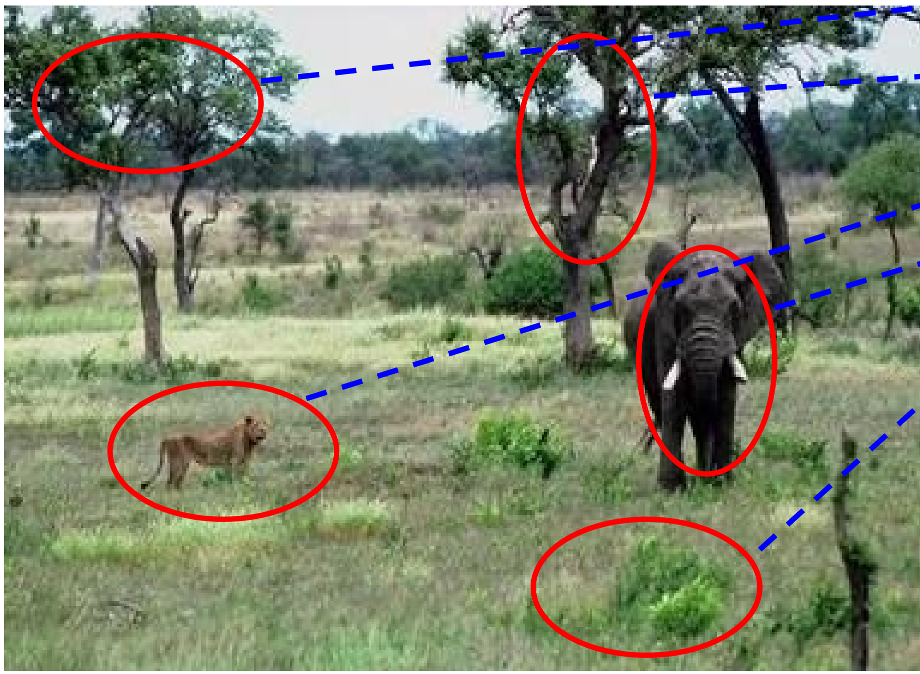
Traditional (single-instance) supervised learning



Multi-instance learning

Why MIL is Appealing ?

Many tasks can be modeled as an MIL task



$[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m]^{\mathrm{T}}$

$[\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_m]^{\mathrm{T}}$

…

…

$[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m]^{\mathrm{T}}$

image $\Rightarrow$ bag

regions in the image $\Rightarrow$ instances in the bag

Applications

MIL Applications

- ✓ Drug prediction [Dietterich et al., AIJ97]
- ✓ Image categorization [Maron & Ratan, ICML'98; Chen & Wang, JMLR04; Chen et al., PAMI06]
- ✓ Image retrieval [Zhang et al., ICDE'02; Zhou et al., AJCAI'05]
- ✓ Text categorization [Andrews et al., NIPS'02; Settles et al., NIPS'07]
- ✓ Computer security [Ruffo, Thesis00]
- ✓ Web mining [Zhou et al., APIN05]
- ✓ Face detection [Viola et al., NIPS'05; Zhang & Viola, NIPS'07]
- ✓ Computer-aided medical diagnosis [Fung et al., NIPS'06]
- ✓ … …

Application: Image Categorization

Learning And Mining from DatA
http://lamda.nju.edu.cn

Figure 6: Results for the waterfall concept using the **single blob with neighbors** concept with +10fp. Top row: Initial training set–5 positive and 5 negative examples. Second Row: Additional false positives. Last three rows: Top 30 matches retrieved from the large test set. The red squares indicate where the closest instance to the learned concept is located.

Application: Image Retrieval

Figure 6. A sample run with 3 rounds of training: retrieving cars

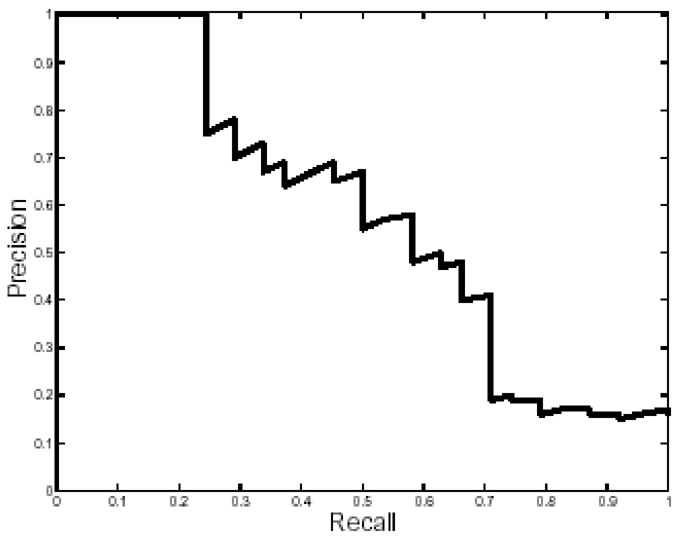


Figure 7. Precision-recall curve for Figure 6

Application: Stock Selection

Goal: To choose stocks perform well for fundamental reasons

Positive bag: 100 stocks with the highest return in every month

Negative bag: the bottom 5 stocks in every month

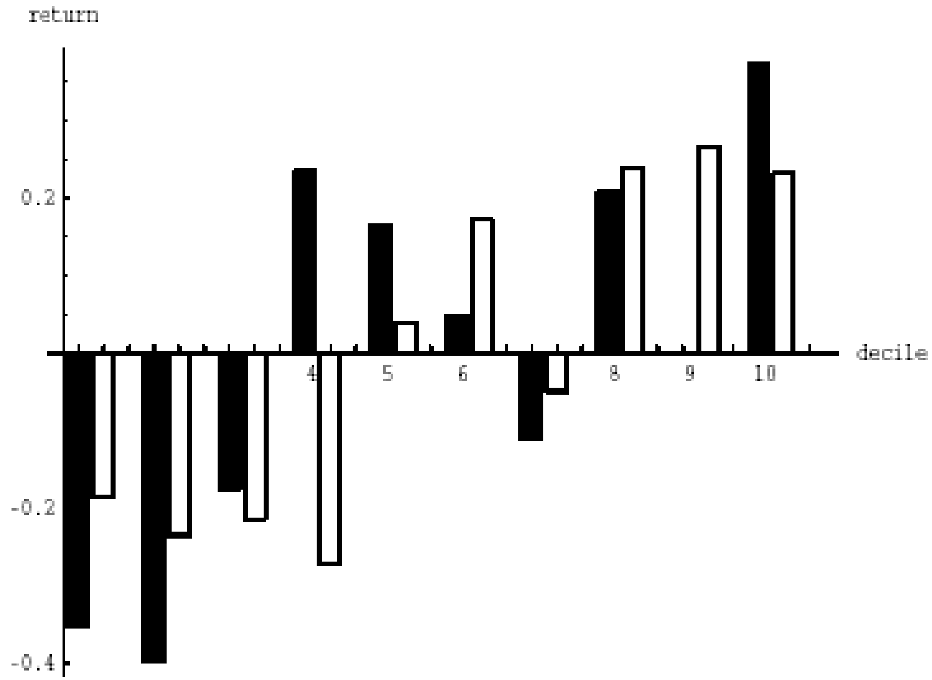


Figure 6: Black bars show Diverse Density's average return on a decile, and the white bars show GMO's predictor's return.

Application: Webpage Recommendation

Fig. 1. The web index page is regarded as a bag, while its linked pages are regarded as the instances in the bag

A web index page linking to m pages, i.e. a bag containing m instances, can be represented as $\{[t_{11}, t_{12}, \ldots, t_{1n}], [t_{21}, t_{22}, \ldots, t_{2n}], \ldots, [t_{m1}, t_{m2}, \ldots, t_{mn}]\}$

The label of the bag is positive if the web index page interested the user; otherwise the label is negative

Application: Face Detection

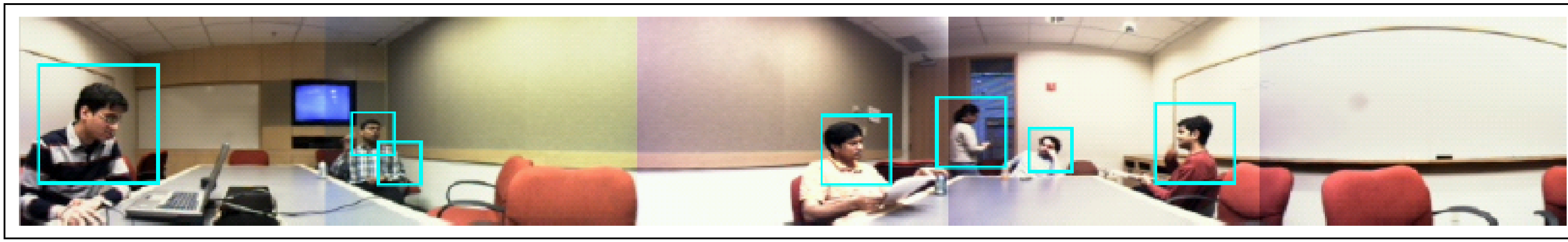


Figure 4: One example from the testing dataset and overlaid results.

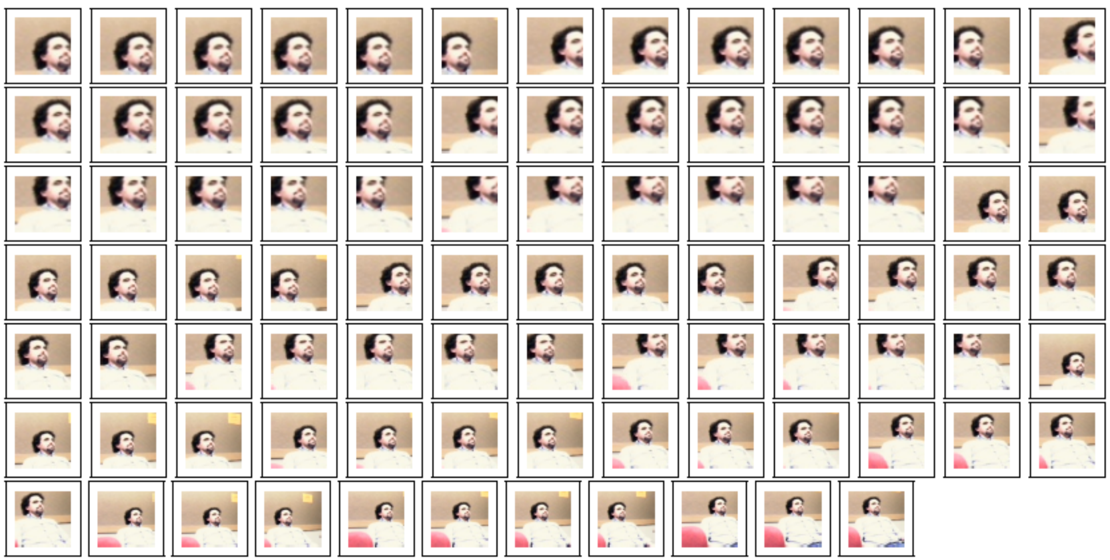


Figure 2: Some of the subwindows in one positive bag.

By incorporating MIL, the detection rate of the famous Viola-Jones face detector improves by nearly 15% (at a 10% false positive rate)

Learnability

Learnability - Earlier Results

- [Long & Tan, COLT'96; MLJ98]
    - If the instances in the bags are independently drawn from product distribution, then the APR is PAC-learnable
    - A polynomial-time theoretical algorithm
- [Auer et al., JCSS98]
    - If the instances in the bags are not independent, then APR learning under MIL is NP-hard
    - A theoretical algorithm that does not require product distribution but with smaller sample complexity than that of Long and Tan's algorithm. Later transformed to MULTINST [Auer, ICML'97]
- [Blum & Kalai, MLJ98]
    - A reduction from the problem of PAC-learning under MIL to PAC-learning with one-sided random classification noise
    - A theoretical algorithm with smaller sample complexity than that of Auer et al.'s algorithm

Learnability - Summary of Main Results

	Sample	Time	Assumption
Long & Tan	$\widetilde{O}(\frac{d^2 r^6}{\varepsilon^{10}})$	$\widetilde{O}(\frac{d^5 r^{12}}{\varepsilon^{20}})$	Instances in bag are independent, and drawn from production distribution
Auer et al.	$\widetilde{O}(\frac{d^2 r^2}{\varepsilon^2})$	$\widetilde{O}(\frac{d^3 r^2}{\varepsilon^2})$	Instances in bag are independent
Blum & Kalai	$\widetilde{O}(\frac{d^2 r}{\varepsilon^2})$	$\widetilde{O}(\frac{d^3 r^2}{\varepsilon^2})$	Instances in bag are independent

Learnability - Important Insights

The heterogeneous case: MIL is hard [Auer et al., JCSS98]

- PAC-learning disjunctions of r APR over R^d is as hard as learning DNF formulas with r clauses over d variables
- A polynomial-time algorithm exists only if $\mathcal{RP} = \mathcal{NP}$

The homogeneous case: If instances are independent, MIL is easy [Blum & Kalai, MLJ98]

- There is a polynomial-time reduction from MIL to the problem of classification with random label noise

<u>Heterogeneous</u>: Each instance in the bag is classified by a different rule

<u>Homogeneous</u>: All instances are classified by the same rule

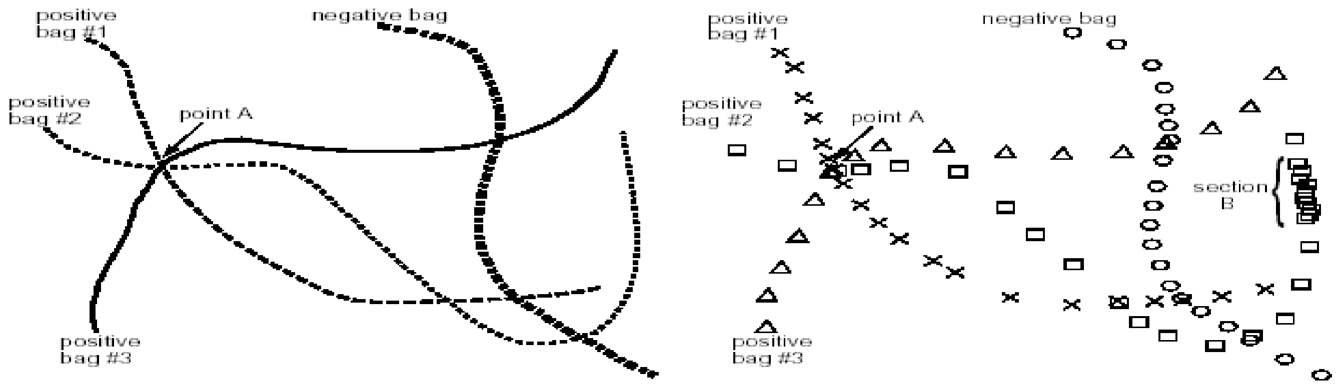
Learnability - A Recent Result

The homogeneous case, when instances are statistically dependent [Sabato & Tishby, COLT'08] :

- At least for $f = \mathrm{OR}$, MIL is PAC-learnable for arbitrary distribution over bags
- VC-dim grows logarithmically with r

Algorithms

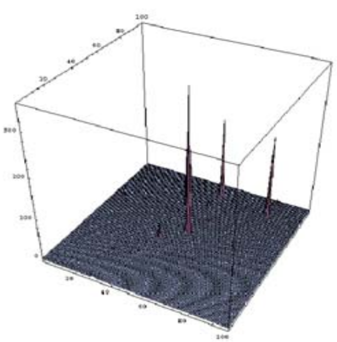
Representative Algorithms – Diverse Density



(a) The different shapes that a molecule can take on are represented as a path. The intersection point of positive paths is where they took on the same shape.

(b) Samples taken along the paths. Section B is a high density area, but point A is a high Diverse Density area.

Figure 1: A motivating example for Diverse Density

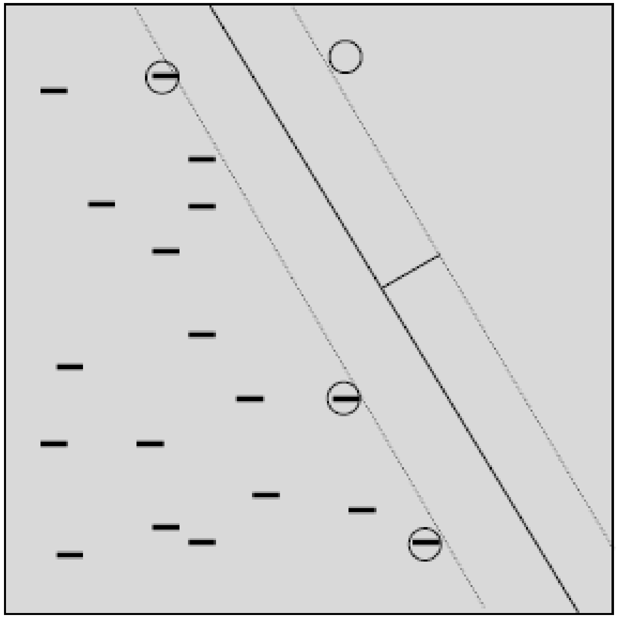


To search for the point with the maximal diverse density by gradient search

every instance in positive bags is used as a start point for search

Figures reprinted from [Maron & Lozano-Pérez, NIPS'97]

Representative Algorithms – MI-SVM



To search for the maximal margin hyperplane

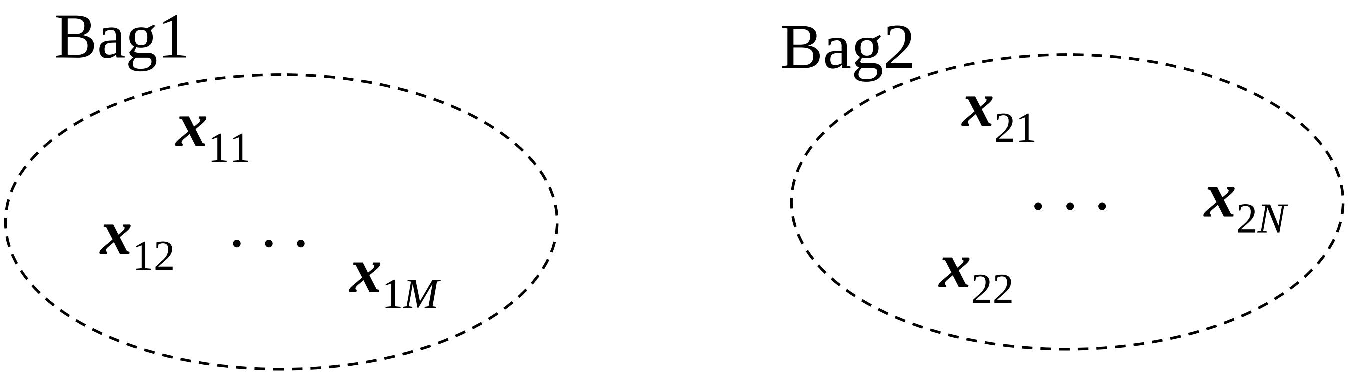
the margin of a "positive bag" is the margin of its "most positive" instance

$$\textit{MI-SVM} \qquad \min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_I \xi_I$$

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

Figure reprinted from [Andrews et al., NIPS'02]

Representative Algorithms - MI-Kernel



Based on ***set kernel*** :

$$\mathbf{K}(\text{Bag1}, \text{Bag2}) = \sum_{i=1}^{M} \sum_{j=1}^{N} \boldsymbol{k}(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2j})$$

k is Gaussian RBF kernel

[Gärtner et al., ICML'02]

Many MIL Algorithms

- ✓ Diverse Density [Maron & Lozano-Perez, NIPS'97], EM-DD [Zhou & Goldman, NIPS'01]
- ✓ K-Nearest Neighbor: Citation-kNN [Wang & Zucker, ICML'00]
- ✓ Decision trees: RELIC [Ruffo, Thesis00], MITI [Blockeel et al., ICML'05]
- ✓ Neural networks: BP-MIP [Zhou & Zhang, ICIIT'02], RBF-MIP [Zhang & Zhou, NPL06]
- ✓ Rule learning algorithm: RIPPER-MI [Chevaleyre & Zucker, CanadianAI'01]
- ✓ Ensemble methods: MI-Ensemble [Zhou & Zhang, ECML'03], MI-Boosting [Xu & Frank, PAKDD'04] , MIL-Boosting [Auer & Ortner, ECML'04]
- ✓ Logistic regression algorithm: MI-LR [Ray & Craven, ICML'05]
- ✓ … …

Kernel/SVM Methods

- ✓ MI-Kernel [Gärtner et al., ICML'02]
- ✓ mi-SVM and MI-SVM [Andrews et al., NIPS'02]
- ✓ DD-SVM [Chen & Wang, JMLR04]
- ✓ CCCP SVM [Cheung & Kwok, ICML'06]
- ✓ marginalized Kernel [Kwok & Cheung, IJCAI'07]
- ✓ MissSVM [Zhou & Xu, ICML'07]
- ✓ PPMM Kernel [Wang et al., ICML'08]
- ✓ … …

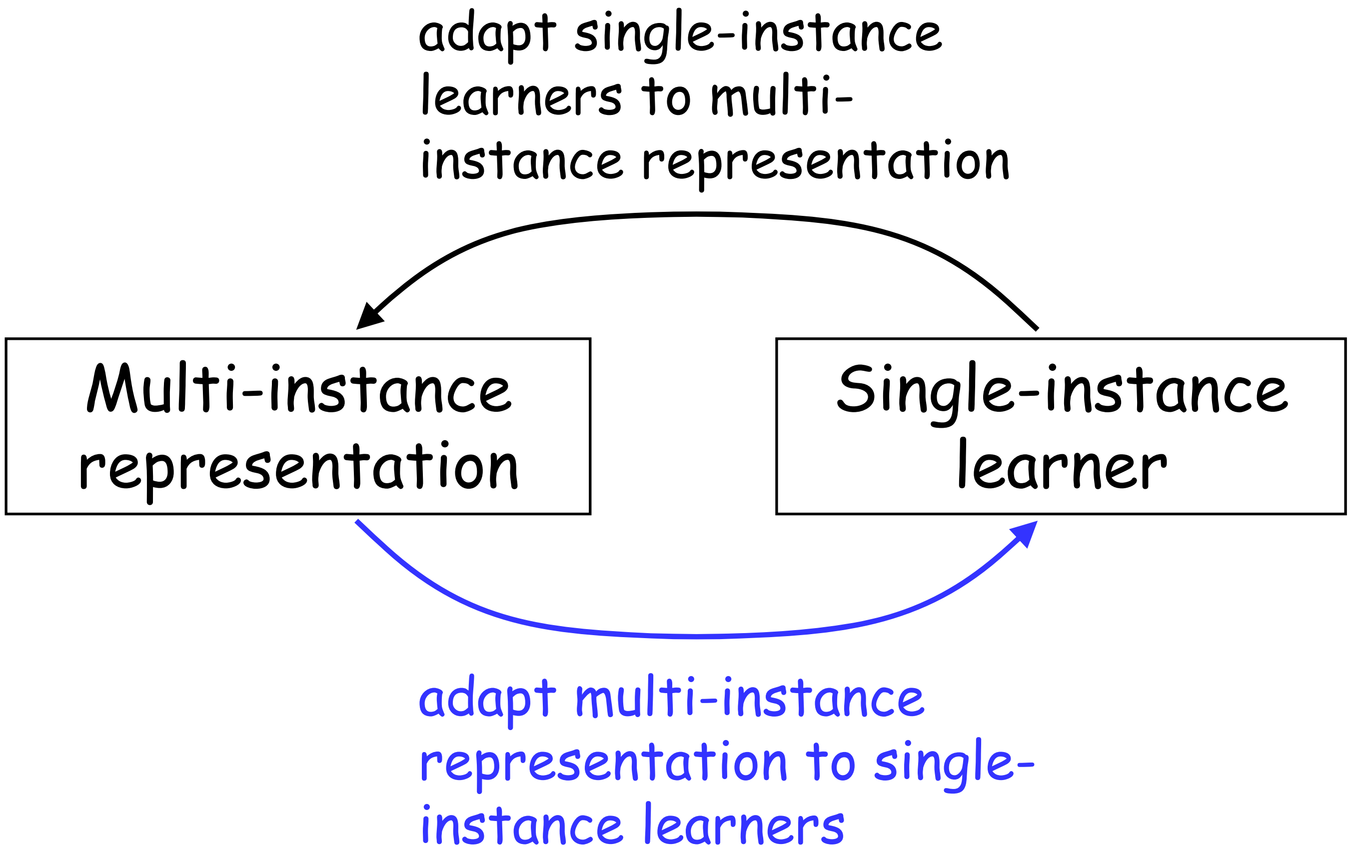
Adapting SIL Algorithms to MIL

A general routine :

Supervised learning algorithms can be adapted to multi-instance learning, by shifting their focuses from the discrimination on instances to the discrimination on bags

[Zhou & Zhang, ECML'03; JCST06]

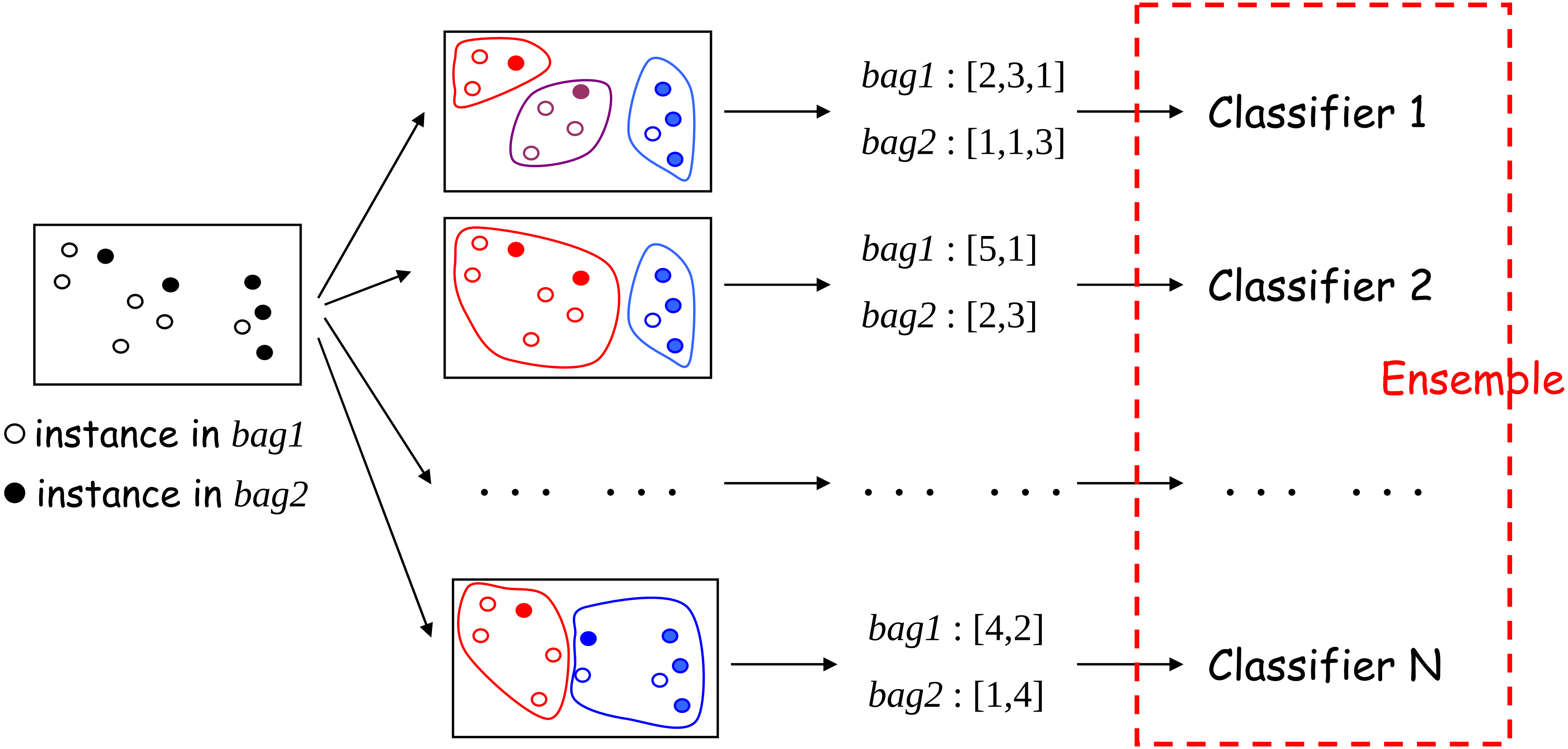
A New General Solution



Example: The CCE Algorithm

CCE (Constructive Clustering based Ensemble) [Zhou & Zhang, KAIS07]

The key: Re-represent bags as single-instances



It is worth noting that:

Most existing MIL algorithms assume i.i.d. instances

A Recent Study

[Zhou & Xu, ICML'07] disclosed that if instances in bags were assumed as i.i.d. samples, MIL is just a special case of semi-supervised learning

The definition of MIL [Dietterich et al., AIJ97] implies that:

- Negative bags contain only negative instances

  Thus, We can regard instances from negative bags as labeled negative examples

- Positive bags can contain positive as well as negative instances

  Thus, we can regard instances from positive bags as unlabeled examples with <u>positive constraints</u>

A Reformulation of MIL Task - [Zhou & Xu, ICML'07]

An semi-supervised learning task:

Definition 1 *Given a set of labeled negative examples* $\{(\boldsymbol{x}_1, -1), (\boldsymbol{x}_2, -1), \cdots, (\boldsymbol{x}_{T_L}, -1)\}$ *and a set of unlabeled instances* $\{\boldsymbol{x}_{T_L+1}, \cdots, \boldsymbol{x}_T\}$*, to learn a function* $F^s : \mathcal{X} \to \{-1, +1\}$ *subject to: For* $i = q+1, \cdots, m$*, at least one instance in* $\{\boldsymbol{x}_{s_i}, \cdots, \boldsymbol{x}_{e_i}\}$ *is positive.*

The task can be solved by semi-supervised SVM algorithm:

MissSVM (Multi-instance learning by semi-supervised SVM)

- Our main focus is not the proposal of a new algorithm (although we do propose a new algorithm)
- Instead of designing elaborate method, we try to use typical and simple SSL technique

The MissSVM Algorithm - [Zhou & Xu, ICML'07]

Learning And Mining from DatA
http://lamda.nju.edu.cn

The optimization problem for popular semi-supervised support vector machine:

$$\min_f \frac{1}{2}||f||^2_{\mathcal{H}} + \lambda \sum_{t=1}^{T_L} H_1(y_t f(\boldsymbol{x}_t)) + \delta \sum_{t=T_L+1}^{T} D(f(\boldsymbol{x}_t))$$

where $H_1(z) = \max\{0, 1-z\}$ is hinge loss

$$D(z) = \min\{H_1(z), H_1(-z)\}$$ [Bennett & Demiriz, NIPS'98]

Considering the positive constraints, the term should be added:

$$\sum_{i=q+1}^{m} H_1\left(\max_{t=s_i,\cdots,e_i} f(\boldsymbol{x}_t)\right)$$

The MissSVM Algorithm (con't)

Thus, the optimization problem can be written as:

$$\min_{f\in\mathcal{H},\eta,\theta,\varepsilon,\xi} \frac{1}{2}||f||^2_{\mathcal{H}} + \lambda\eta'\mathbf{1} + \gamma\theta'\mathbf{1} + \delta\min(\boldsymbol{\varepsilon},\boldsymbol{\xi})'\mathbf{1}$$

$$\text{s.t.}\begin{cases} (-1)f(\boldsymbol{x}_t) + \eta_t \geqslant 1,\ \eta_t \geqslant 0,\ t = 1, 2, \cdots, T_L; \\ \max\limits_{t=s_i,\cdots,e_i} f(\boldsymbol{x}_t) + \theta_{i-q} \geqslant 1,\ \theta_{i-q} \geqslant 0, \\ \qquad\qquad\qquad\qquad i = q+1, \cdots, m; \\ f(\boldsymbol{x}_t) + \varepsilon_{t-T_L} \geqslant 1,\ \varepsilon_{t-T_L} \geqslant 0, \\ \qquad\qquad\qquad\qquad t = T_L+1, \cdots, T; \\ (-1)f(\boldsymbol{x}_t) + \xi_{t-T_L} \geqslant 1,\ \xi_{t-T_L} \geqslant 0, \\ \qquad\qquad\qquad\qquad t = T_L+1, \cdots, T. \end{cases}$$

- $\eta = [\eta_1, \cdots, \eta_{T_L}]'$ - slack variables for errors on instances from negative bags
- $\theta = [\theta_1, \cdots, \theta_p]'$ - slack variables for errors on positive bags
- $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_{T_U}]'$ and $\xi = [\xi_1, \ldots, \xi_{T_U}]'$ - slack variables for errors on instances from positive bags
- λ, γ and δ - parameters

The MissSVM Algorithm (con't)

Let $\mathbf{K}$ denote a $T{\times}T$ kernel matrix and let $\mathbf{k}_t$ denote the t-th column:

$$\min_{\boldsymbol{\alpha},\eta,\theta,\varepsilon,\xi,b} \frac{1}{2}\boldsymbol{\alpha}'\mathbf{K}\boldsymbol{\alpha} + \lambda\eta'\mathbf{1} + \gamma\theta'\mathbf{1} + \delta\min(\boldsymbol{\varepsilon},\boldsymbol{\xi})'\mathbf{1}$$

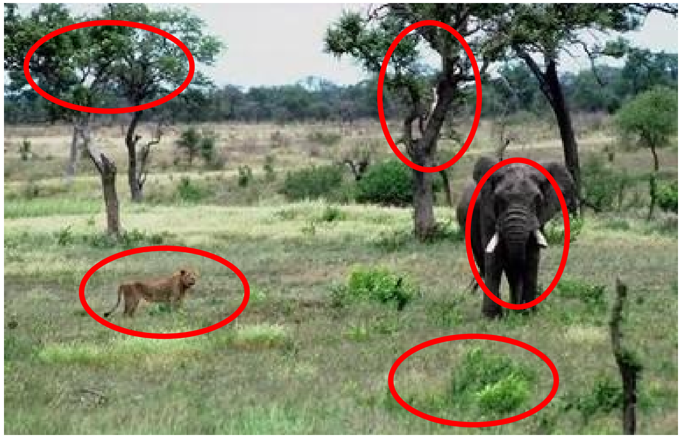
$$\text{s.t.} \begin{cases} (-1)(\mathbf{k}'_t\boldsymbol{\alpha}+b)+\eta_t \geqslant 1,\ \ \eta_t \geqslant 0, \\ \qquad\qquad\qquad\qquad t=1,2,\cdots,T_L; \\ \max\limits_{t=s_i,\cdots,e_i}(\mathbf{k}'_t\boldsymbol{\alpha}+b)+\theta_{i-q} \geqslant 1,\ \ \theta_{i-q}\geqslant 0, \\ \qquad\qquad\qquad\qquad i=q+1,\cdots,m; \\ (\mathbf{k}'_t\boldsymbol{\alpha}+b)+\varepsilon_{t-T_L} \geqslant 1,\ \ \varepsilon_{t-T_L}\geqslant 0, \\ \qquad\qquad\qquad\qquad t=T_L+1,\cdots,T; \\ (-1)(\mathbf{k}'_t\boldsymbol{\alpha}+b)+\xi_{t-T_L} \geqslant 1,\ \ \xi_{t-T_L}\geqslant 0, \\ \qquad\qquad\qquad\qquad t=T_L+1,\cdots,T. \end{cases}$$

After replacing the gradients of the non-smooth *min* and *max* by their subgradients, CCCP (Constrained Convex-Concave Procedure) [Smola et al., AISTATS'05] can be used to solve this optimization problem

Shouldn't Assume I.I.D. Instances

From the aspect of the nature of MIL task:

It is often not reasonable to assume independent instances since the instances were generally extracted from the same object



From the aspect of MIL techniques:

If i.i.d. instances were assumed, MIL problems can be solved by adapting SSL techniques, and therefore we do not need to study dedicated MIL techniques

Can we do MIL without assuming i.i.d. instances ?

An Inspirational Example

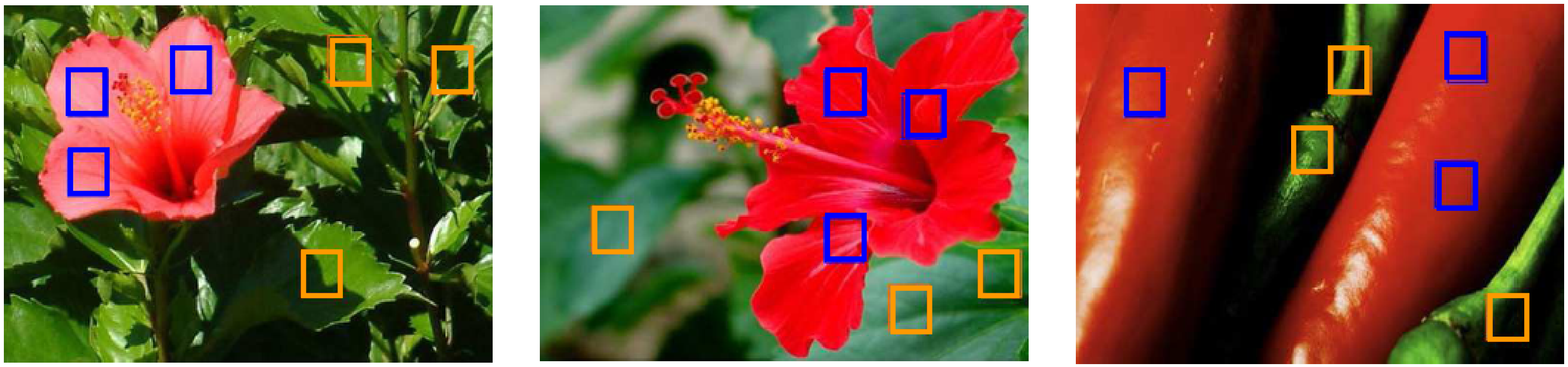


Figure 1. Example images with six marked patches each corresponding to an instance

An Inspirational Example (con't)

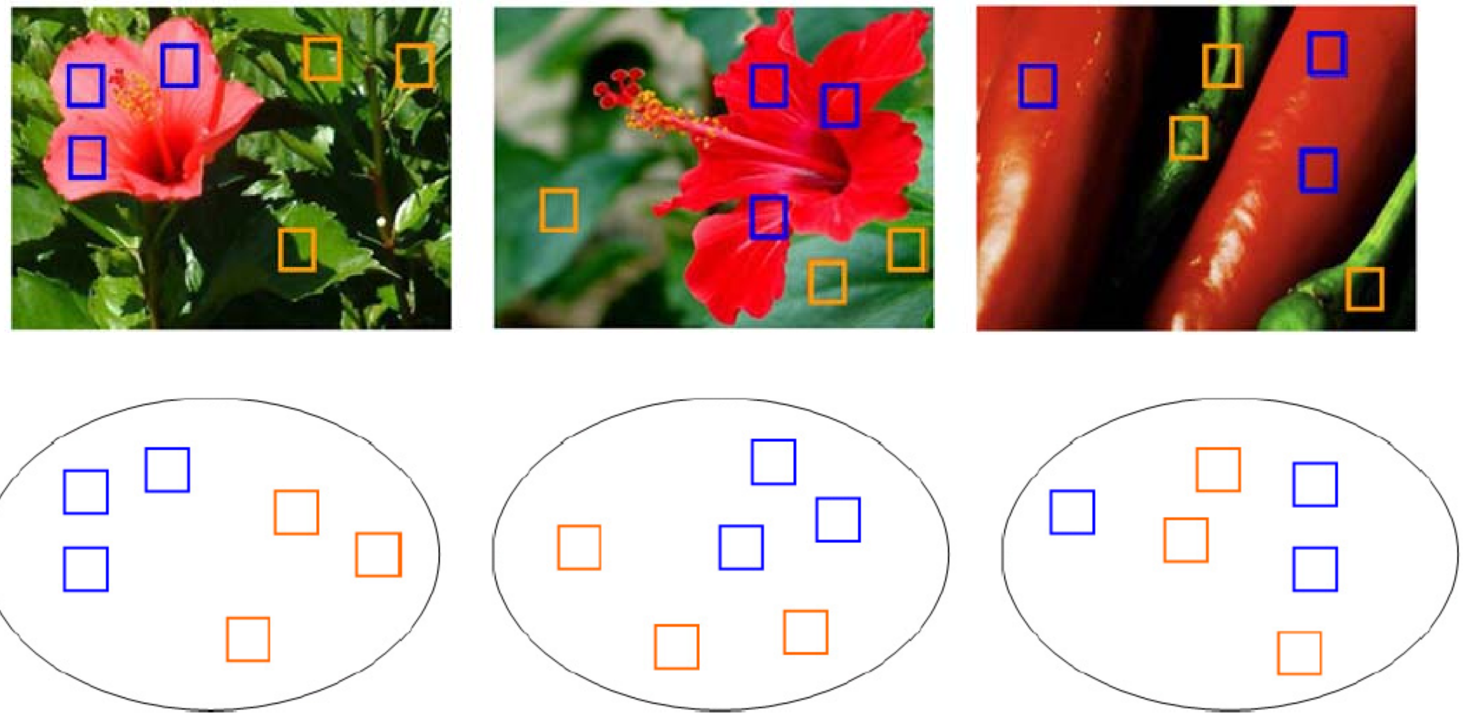


Figure 2. If we do not consider the relations among the instances, the three bags are similar to each other since they have identical number of very similar instances

Figure reprinted from [Zhou et al., ICML'09]

An Inspirational Example (con't)

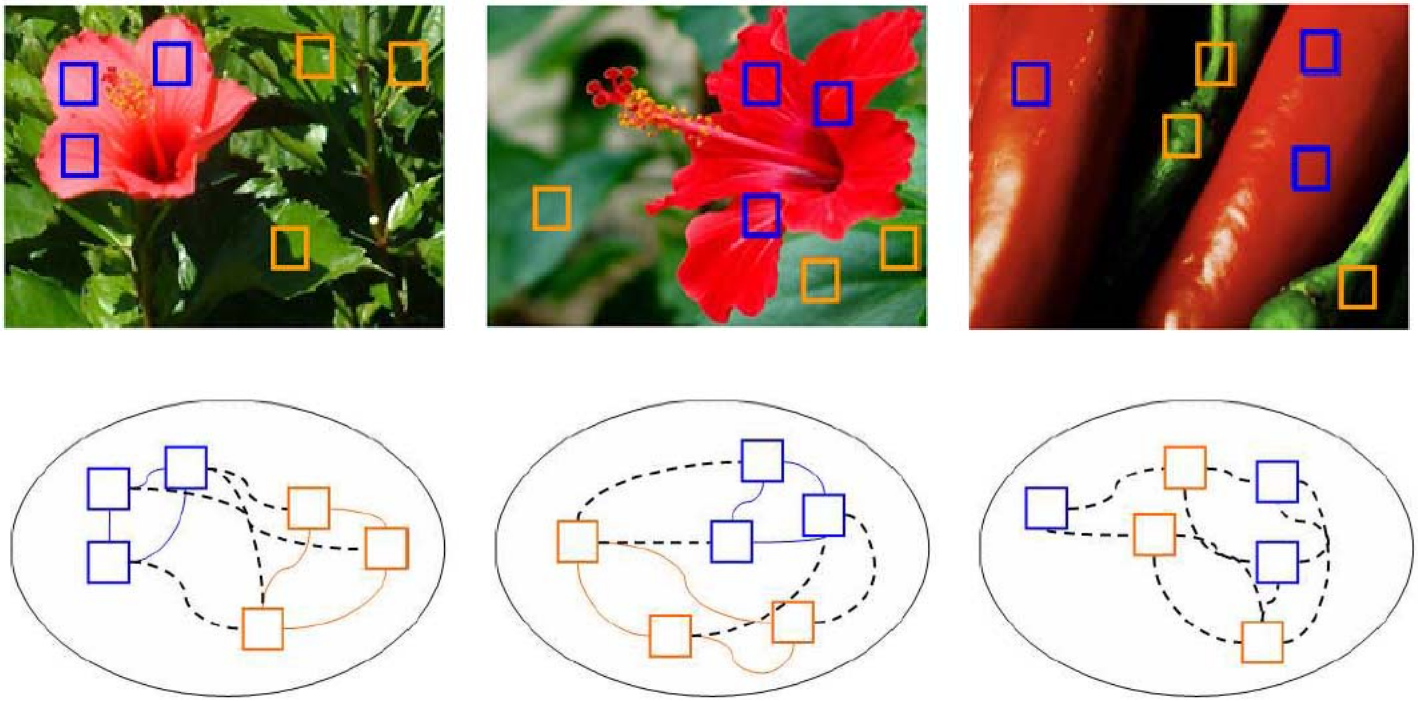
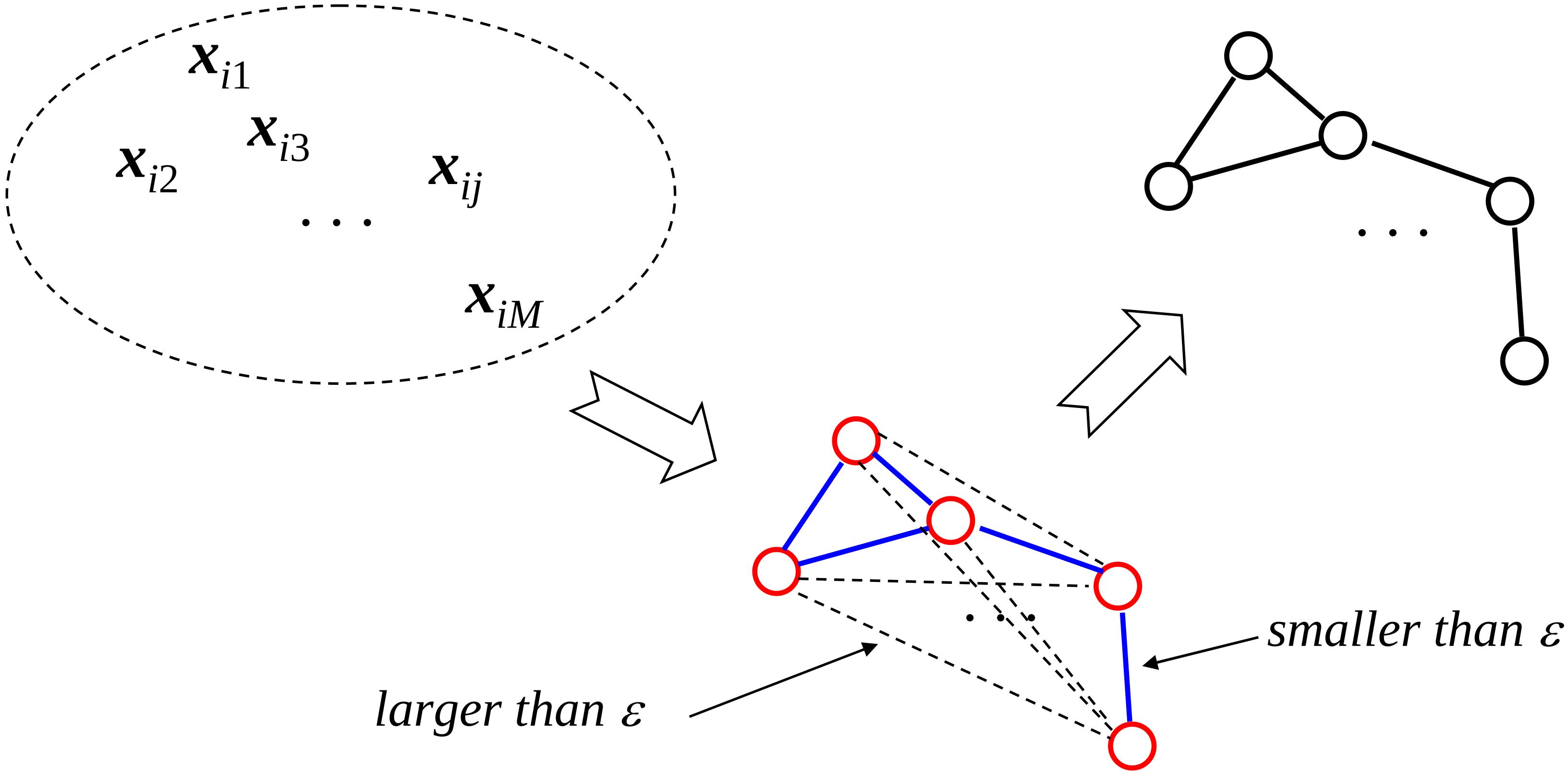


Figure 3. If we consider the relations among the instances, the first two bags are more similar than the third bag. Here, the solid lines highlight the high affinity among similar instances

The MIGraph Method - [Zhou et al., ICML'09]

Learning And Mining from DatA
http://lamda.nju.edu.cn

1) Map each bag into a graph (e.g., ε-graph)



The MIGraph Method (con't)

2) A graph kernel to distinguish positive/negative graphs

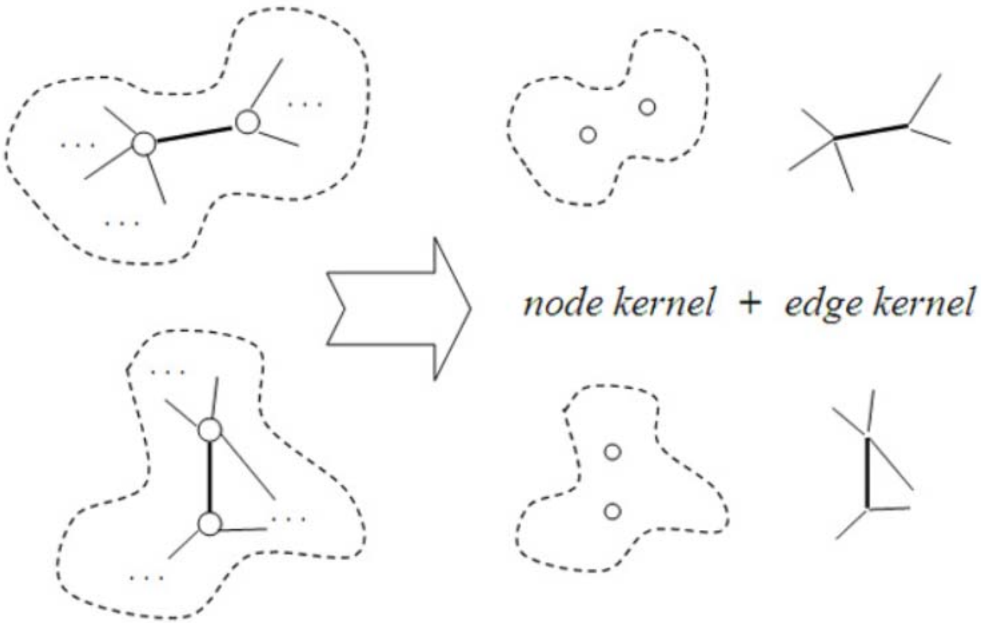
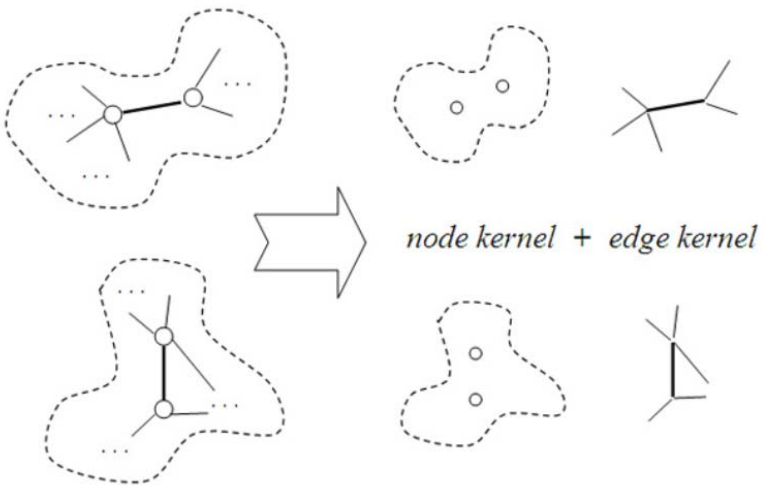


Figure 4. Illustration of the graph kernel in MIGraph

The MIGraph Method (con't)



MI-Kernel [Gärtner et al., ICML'02]

$$k_G(X_i, X_j) = \sum_{a=1}^{n_i} \sum_{b=1}^{n_j} k_{node}(\boldsymbol{x}_{ia}, \boldsymbol{x}_{jb}) + \sum_{a=1}^{m_i} \sum_{b=1}^{m_j} k_{edge}(\mathbf{e}_{ia}, \mathbf{e}_{jb})$$

We define the edge connecting $\boldsymbol{x}_{iu}$ and $\boldsymbol{x}_{iv}$ as $[d_u, p_u, d_v, p_v]'$

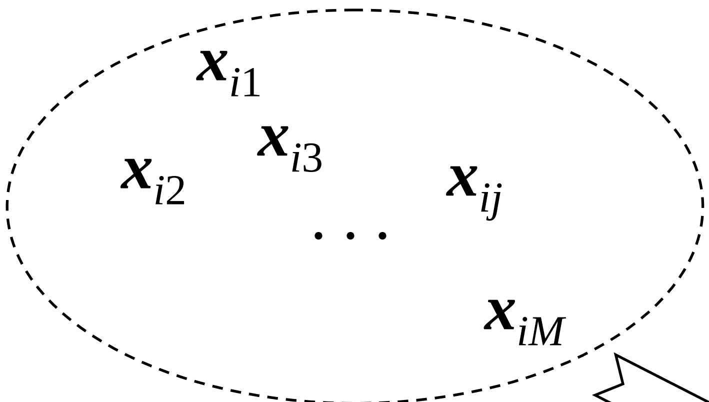
d_u is the degree of $\boldsymbol{x}_{iu}$ $\qquad p_u = w_{uv} / \sum w_{u,*}$

We define both k_{node} and k_{edge} as Gaussian RBF kernel

Complexity of MIGraph: $O(n_i n_j + m_i m_j)$

The miGraph Method - [Zhou et al., ICML'09]

1) Instead of constructing a graph explicitly, we consider an affinity matrix for each bag



$$\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & d_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ d_{M1} & d_{M2} & \cdots & d_{MM} \end{bmatrix}$$

smaller than ε, set to 1

larger than ε, set to 0

$$\begin{bmatrix} 1 & 1 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

The miGraph Method (con't)

2) A kernel to distinguish positive/negative graphs

$$k_g(X_i, X_j) = \frac{\sum_{a=1}^{n_i} \sum_{b=1}^{n_j} \boxed{W_{ia}} W_{jb} \boxed{k(\boldsymbol{x}_{ia}, \boldsymbol{x}_{jb})}}{\sum_{a=1}^{n_i} W_{ia} \sum_{b=1}^{n_j} W_{jb}}$$

The node kernel as similar as in MIGraph, realized with Gaussian RBF kernel

$$W_{ia} = 1 / \underline{\sum_{u=1}^{n_i} w_{au}^i}$$

$$\begin{bmatrix} 1 & 1 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Complexity of miGraph: $O(n_i n_j)$

Four Kinds of Tasks

1. Benchmark tasks
2. Image categorization
3. Text categorization
4. Multi-instance regression

We have re-implemented MI-Kernel since the comparison With MI-Kernel will clearly show whether it is helpful to treat instances as non-i.i.d. samples

The performance of MI-Kernel in our implementation is better than that reported in [Gärtner et al., ICML'02]

The Implication

When designing MIL algorithms, we should not treat instances in bags as i.i.d. samples !!

A new way to the design of MIL algorithms. Many things can be done along this way, e.g.,

- ✓ Better graph/kernel
- ✓ Extend to other MIL methods/settings
- ✓ Alternative non-i.i.d. strategies
- ✓ … …

Can we identify the "key instances" ?

Locating ROI in CBIR

Learning And Mining from DatA

http://lamda.nju.edu.cn

Return relevant images with marked ROI

Potential usage: medical, military, etc.

Fig. 3. The ROI located by Diverse Density and C*k*NN-ROI. Each row shows five pairs of example images on the target concepts *castle*, *firework*, *mountain*, *sunset*, and *waterfall*, respectively. In each pair the first image is obtained with Diverse Density while the second one is obtained with C*k*NN-ROI.

DD can be used in scene classification, but can hardly in CBIR

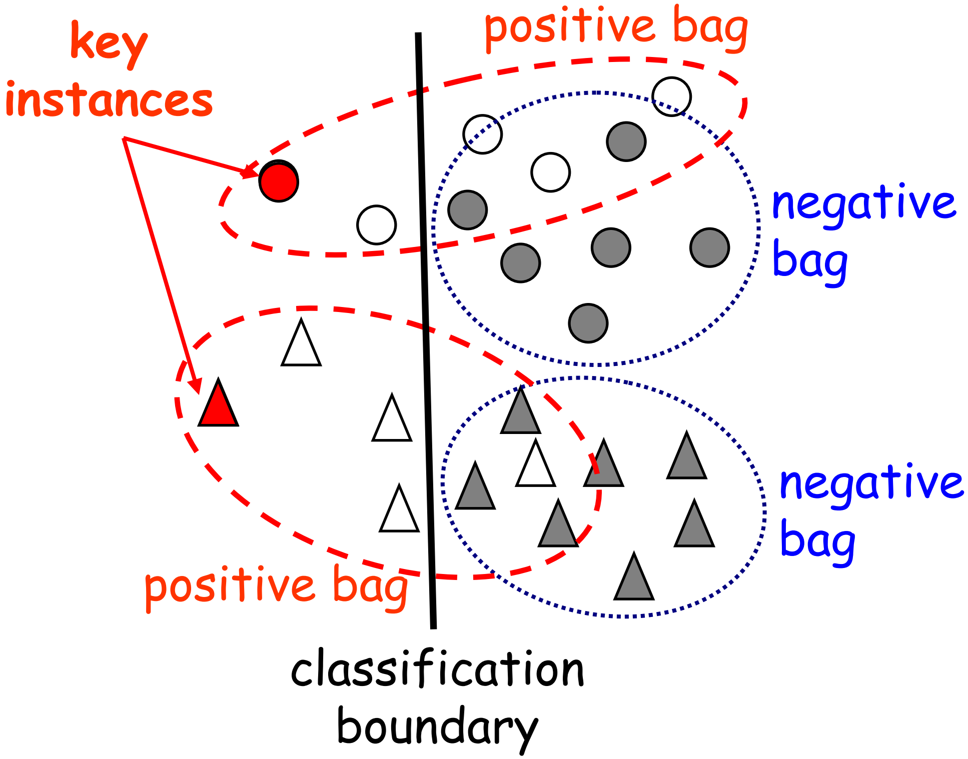
Time per query:

- DD :
  14,475 seconds
  (≈ 4 hours)
- CkNN-ROI:
  0.32 second

The first attempt of locating ROI in CBIR

Locating ROI in CBIR (con't)

MI-SVM [Andrews et al., NIP'02] can identify key instances



The MI-SVM formulation is non-convex

KI-SVM [Li et al., ECML'09]:

A Convex relaxation method

- Inst-KI-SVM
- Bag-KI-SVM

Improving efficiency by using *cutting-plane* and *label generation*

Performance better than DD, CkNN-ROI and MI-SVM

Efficiency better than DD and MI-SVM

Multi-Instance Learning … more details

Learning And Mining from DatA
http://lamda.nju.edu.cn

Adapting SIL-learners to MIL-representation; MI-Ensemble:

- ✓ Z.-H. Zhou and M.-L. Zhang. Ensembles of multi-instance learners. In: Proceedings of the 14th European Conference on Machine Learning (ECML'03), Cavtat-Dubrovnik, Croatia, LNAI 2837, 2003, pp.492-502

  Code: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/MIL-Ensemble.htm

- ✓ Z.-H. Zhou. Multi-instance learning from supervised view. Journal of Computer Science and Technology, 2006, 21(5): 800-809

Adapting MIL-representation to SIL-learners:

- ✓ Z.-H. Zhou and M.-L. Zhang. Solving multi-instance problems with classifier ensemble based on constructive clustering. Knowledge and Information Systems, 2007, 11(2): 155-170

  Code: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/CCE.htm

Multi-Instance Learning … more details

MIL & SSL:

✓ Z.-H. Zhou and J.-M. Xu. On the relation between multi-instance learning and semi-supervised learning. In: Proceedings of the 24th International Conference on Machine Learning (ICML'07), Corvallis, OR, 2007, pp.1167-1174.

Code: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/MissSVM.htm

MIL without i.i.d. assumption:

✓ Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li. Multi-instance learning by treating instances as non-i.i.d. samples. In: Proceedings of the 26th International Conference on Machine Learning (ICML'09), Montreal, Canada, 2009, pp.1249-1256. (CORR abs/0807.1997)

Code: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/miGraph.htm

Data: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/mil-text-data.htm

Multi-Instance Learning … more details

Identifying key instances:

✓ Z.-H. Zhou, X.-B. Xue, and Y. Jiang. Locating regions of interest in CBIR with multi-instance learning techniques. In: Proceedings of the 18th Australian Joint Conference on Artificial Intelligence (AJCAI'05), Sydney, Australia, LNAI 3809, 2005, pp.92-101.

✓ Y.-F. Li, J. T. Kwok, I. W. Tsang, and Z.-H. Zhou. A convex method for locating regions of interest with multi-instance learning. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'09), Bled, Slovenia, Part II, LNAI 5782, 2009, pp.15-30.

Code: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/KISVM.htm

Multi-Instance Learning … more details

MIL-based Webpage recommendation:

- ✓ Z.-H. Zhou, K. Jiang, and M. Li. Multi-instance learning based web mining. Applied Intelligence, 2005, 22(2): 135-147

  Data: http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/milweb-datafile.htm

Multi-instance bag generator:

- ✓ Z.-H. Zhou, M.-L. Zhang, and K.-J. Chen. A novel bag generator for image database retrieval with multi-instance learning techniques. In: Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03), Sacramento, CA, 2003, pp.565-569.

Thanks !