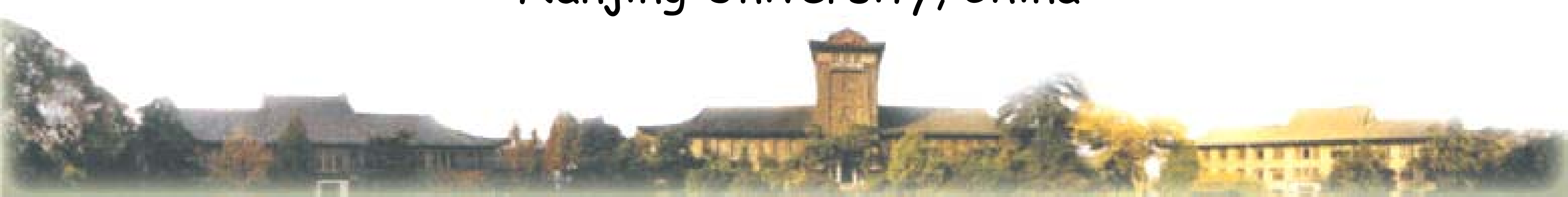
LAMDA
Learning And Mining from DatA
http://lamda.nju.edu.cn

A Framework for Machine Learning with Ambiguous Objects

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

Yin-Xing Li

Yu-Feng Li

Sheng-Jun Huang

... ...

And my collaborators:

Jieping Ye

Shuiwang Ji

Sudir Kumar

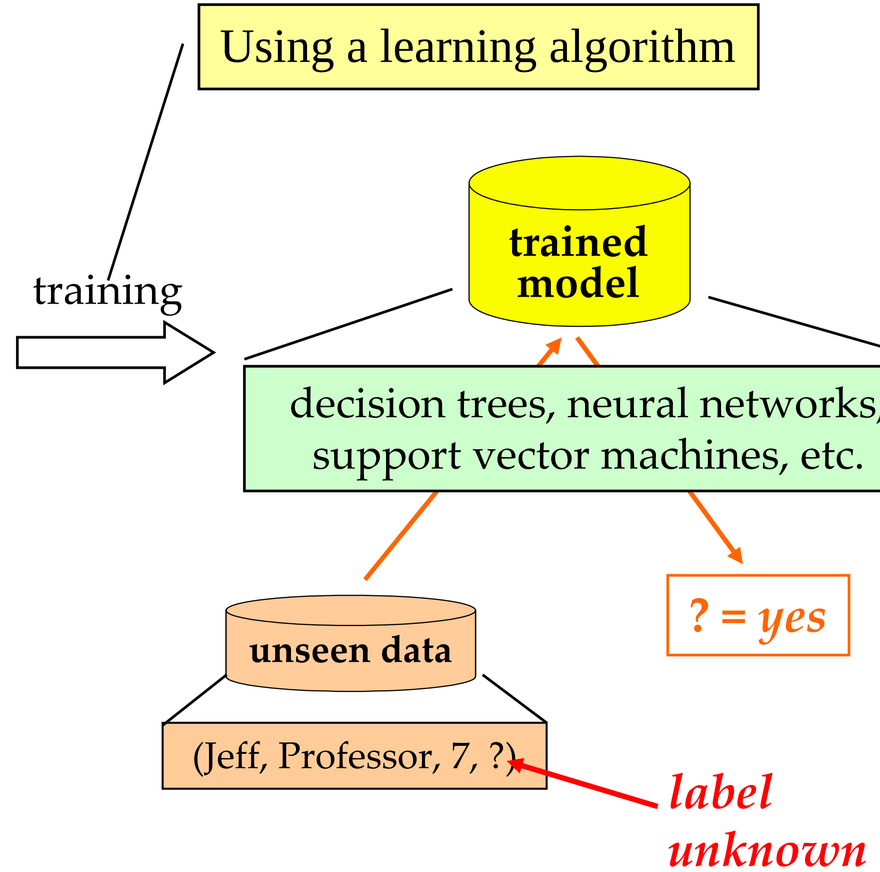
... ...

A typical machine learning process

training data

label

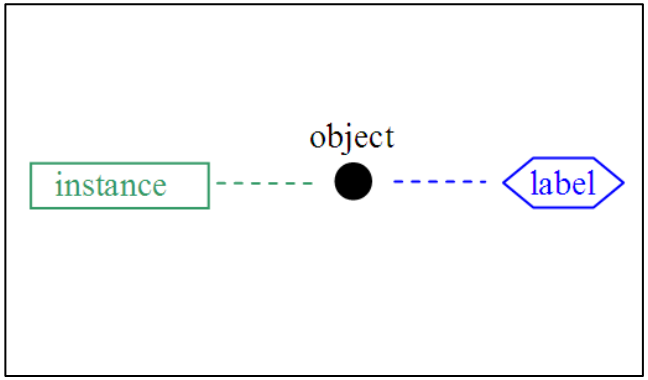
Name	**Rank**	**Years**	**Tenured**
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no



Traditional Machine Learning Setting

In traditional supervised learning:

- A real-world object is represented by an instance (feature vector)
- The instance is associated with a label which indicates the concerned characteristics (such as categorization) of the object



$\mathcal{X}$ - the instance space
$\mathcal{Y}$ - the set of class labels

The task:

To learn a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ from a given data set $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_m, y_m)\}$ where $\boldsymbol{x}_i \in \mathcal{X}$ is an instance and $y_i \in \mathcal{Y}$ is the known label of $\boldsymbol{x}_i$

Ambiguous Data

Elephant ? *Lion ?* *Grassland?*
Tropic ? *Africa ?* … …

Ambiguous Data

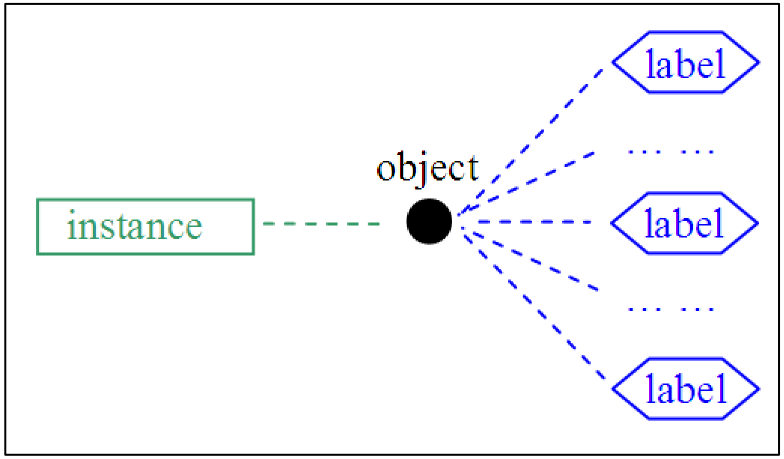
Scientific novel

Jules Verne's writing

Book on traveling

… …

Multi-Label Learning



MLL task:
To learn a function $f_{MLL}: \mathcal{X} \to 2^{\mathcal{Y}}$ from a given data set $\{(\boldsymbol{x}_1, Y_1), (\boldsymbol{x}_2, Y_2), \cdots, (\boldsymbol{x}_m, Y_m)\}$, where $\boldsymbol{x}_i \in \mathcal{X}$ is an instance and $Y_i \subseteq \mathcal{Y}$ is a set of labels $\{y_1^{(i)}, y_2^{(i)}, \cdots, y_{l_i}^{(i)}\}$, $y_k^{(i)} \in \mathcal{Y}$ $(k = 1, 2, \cdots, l_i)$.

$\mathcal{X}$ - the instance space

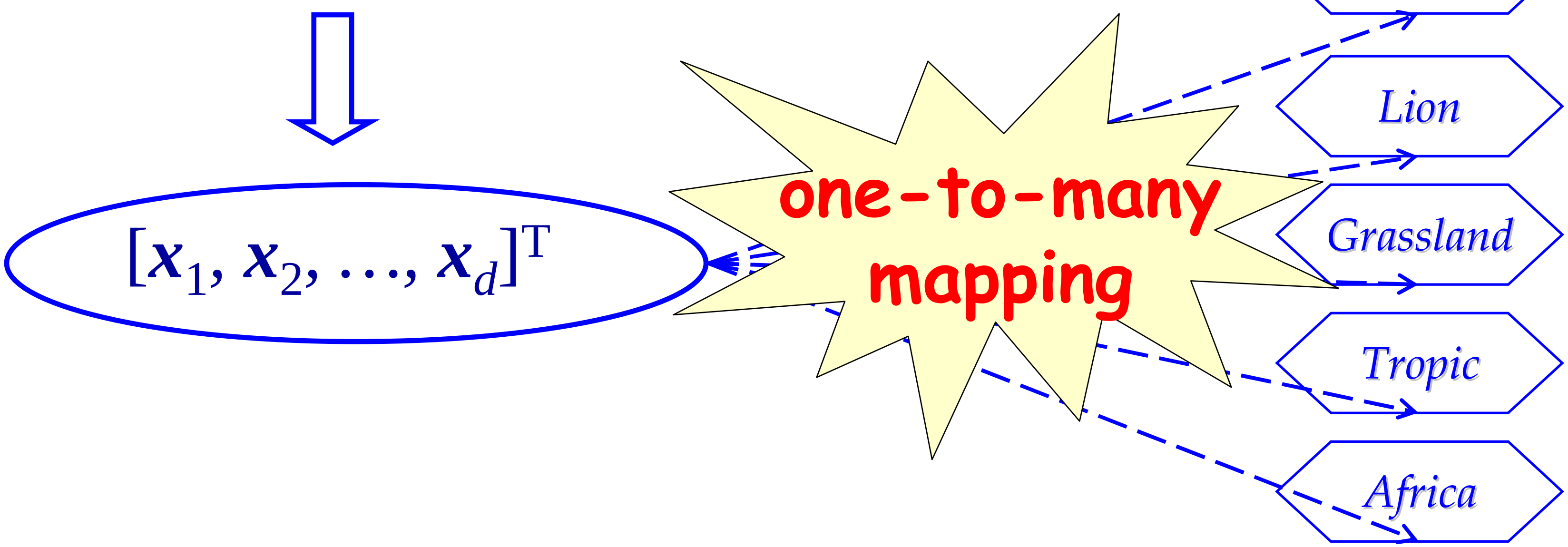
$\mathcal{Y}$ - the set of class labels

l_i - the number of labels in Y_i

Multi-Label Learning Algorithms

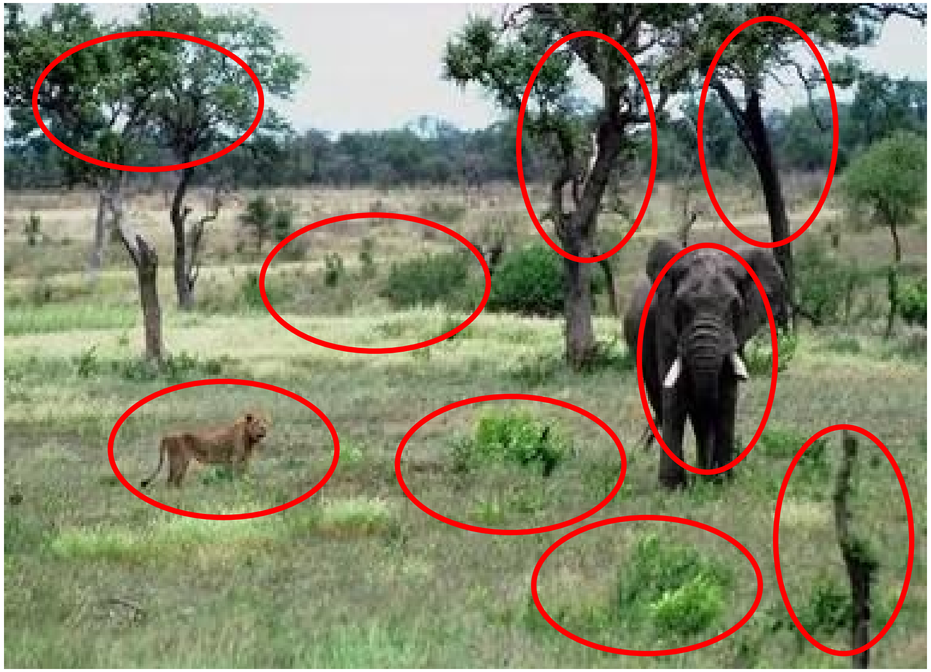
- Decomposing the task into multiple binary classification problems each for a class
  - ✓ MLSVM [Boutell et al., PR04]
  - ✓ . . . . . .
- Considering the ranking among labels
  - ✓ BoosTexter [Schapire & Singer, MLJ00]
  - ✓ BP-MLL [Zhang & Zhou, TKDE06]
  - ✓ RankSVM [Elisseeff & Weston, NIPS'01]
  - ✓ . . . . . .
- Exploring the class correlation
  - ✓ Probabilistic generative models [McCallum, AAAI'99w; Ueda & Saito, NIPS'02]
  - ✓ Maximum entropy methods [Ghamrawi & McCallum, CIKM'05; Zhu et al., SIGIR'05]
  - ✓ . . . . . .

The Problem


$[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d]^{\mathrm{T}}$
one-to-many mapping
Elephant
Lion
Grassland
Tropic
Africa

Consider …

An image usually contains **multiple** regions each can be represented by an instance



The image can simultaneously belong to **multiple** classes

Elephant

Lion

Grassland

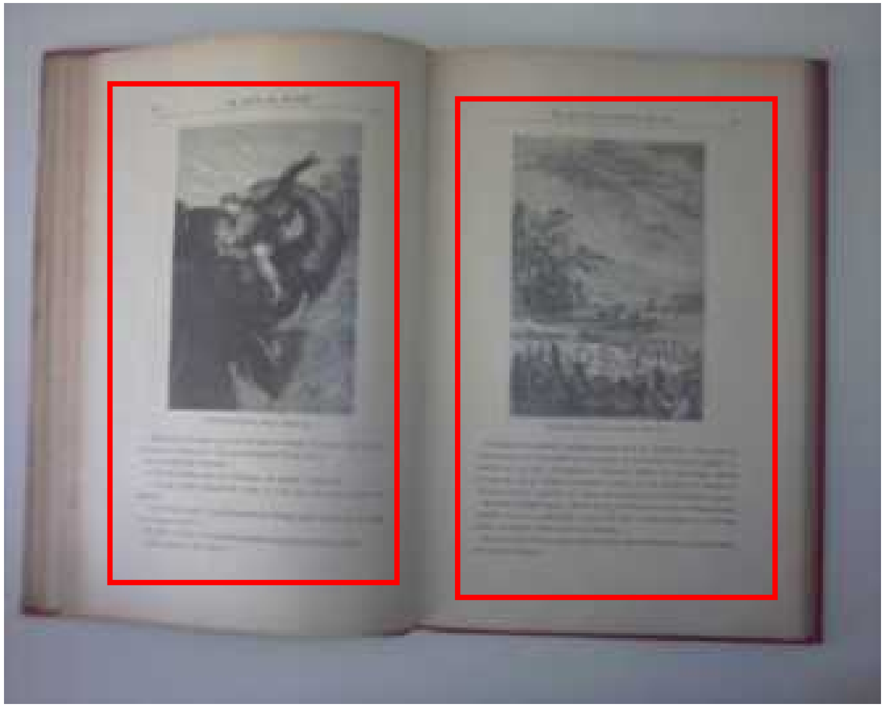
Tropic

Africa

… …

Consider …

A document usually contains **multiple** sections each can be represented by an instance



The document can simultaneously belong to **multiple** categories

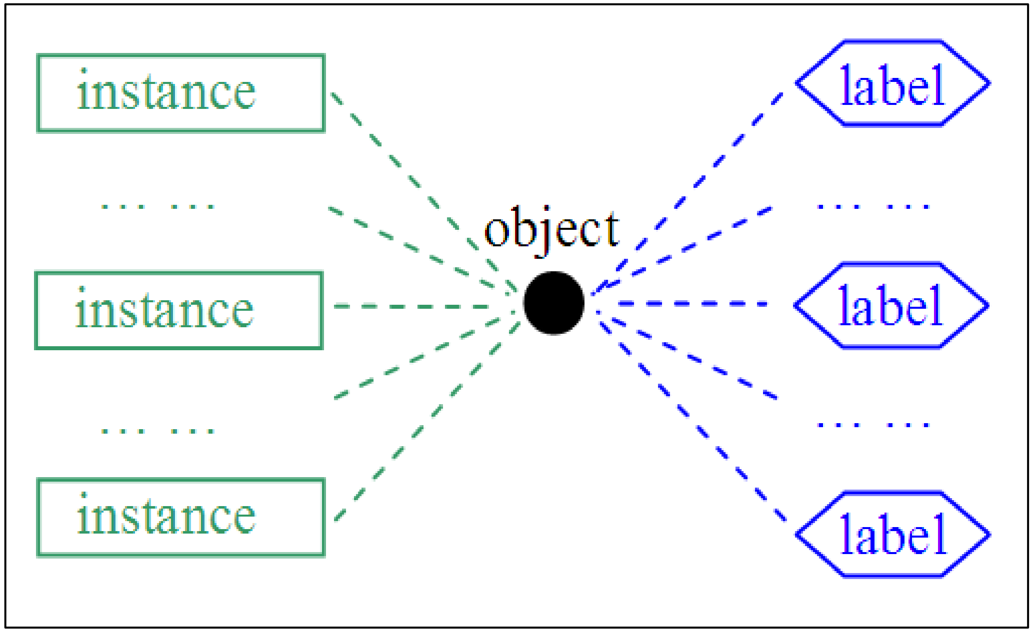
Scientific novel

Jules Verne's writing

Book on traveling

… …

MIML



Multi-Instance Multi-Label (MIML) Learning

多示例多标记学习

[Z.-H. Zhou & M.-L. Zhang, NIPS'06]

Why MIML ?

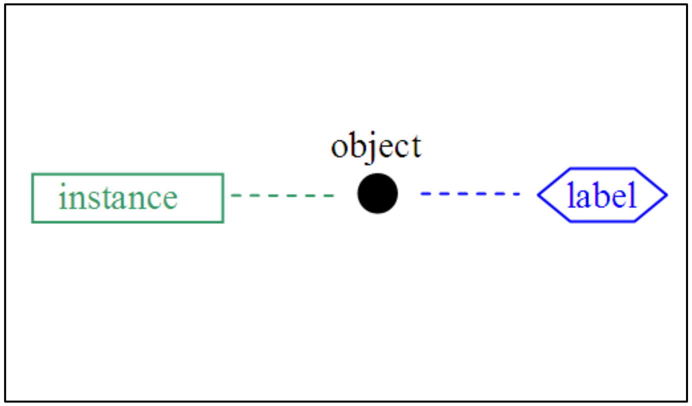
Appropriate representation is important

Having an appropriate representation is as important as having a strong learning algorithm

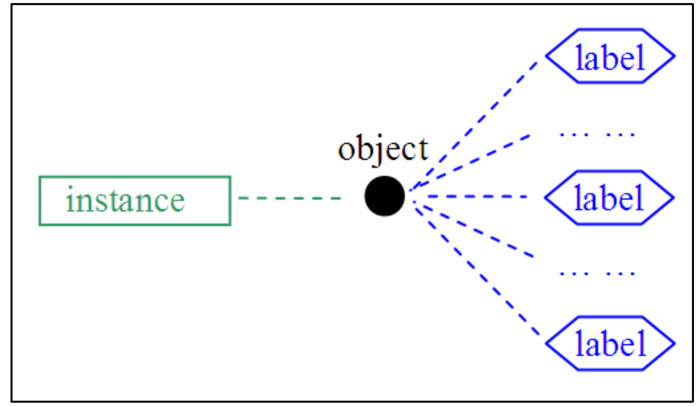
MIML captures more information of ambiguous data

Traditional supervised learning, multi-instance learning and multi-label learning are degenerated versions of MIML

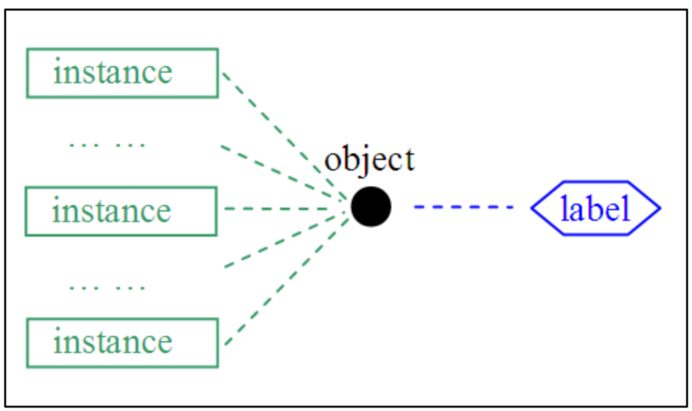
Why MIML ? (con't)



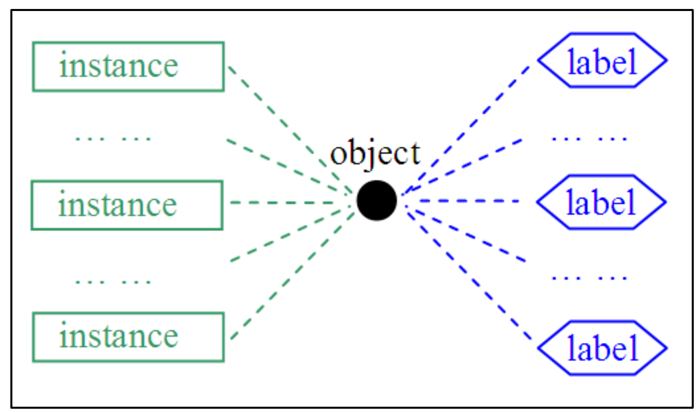
Traditional supervised learning



Multi-label learning



Multi-instance learning



Multi-instance multi-label learning

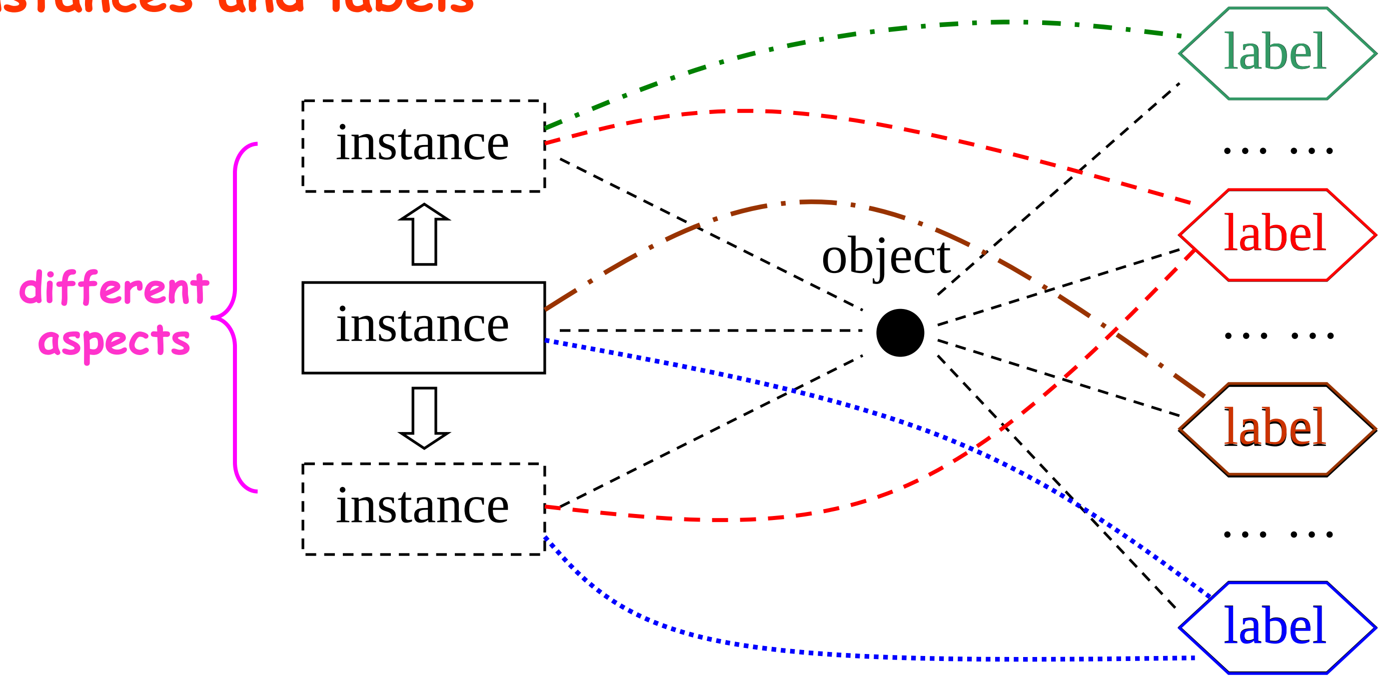
[Z.-H. Zhou & M.-L. Zhang, NIPS'06]

Why MIML ? (con't)

To learn an ***one-to-many*** mapping is an ill-posed problem

Why there are multiple labels?

many-to-many mapping seems better; and moreover, **MIML also offers a possibility for understanding the relationship between instances and labels**



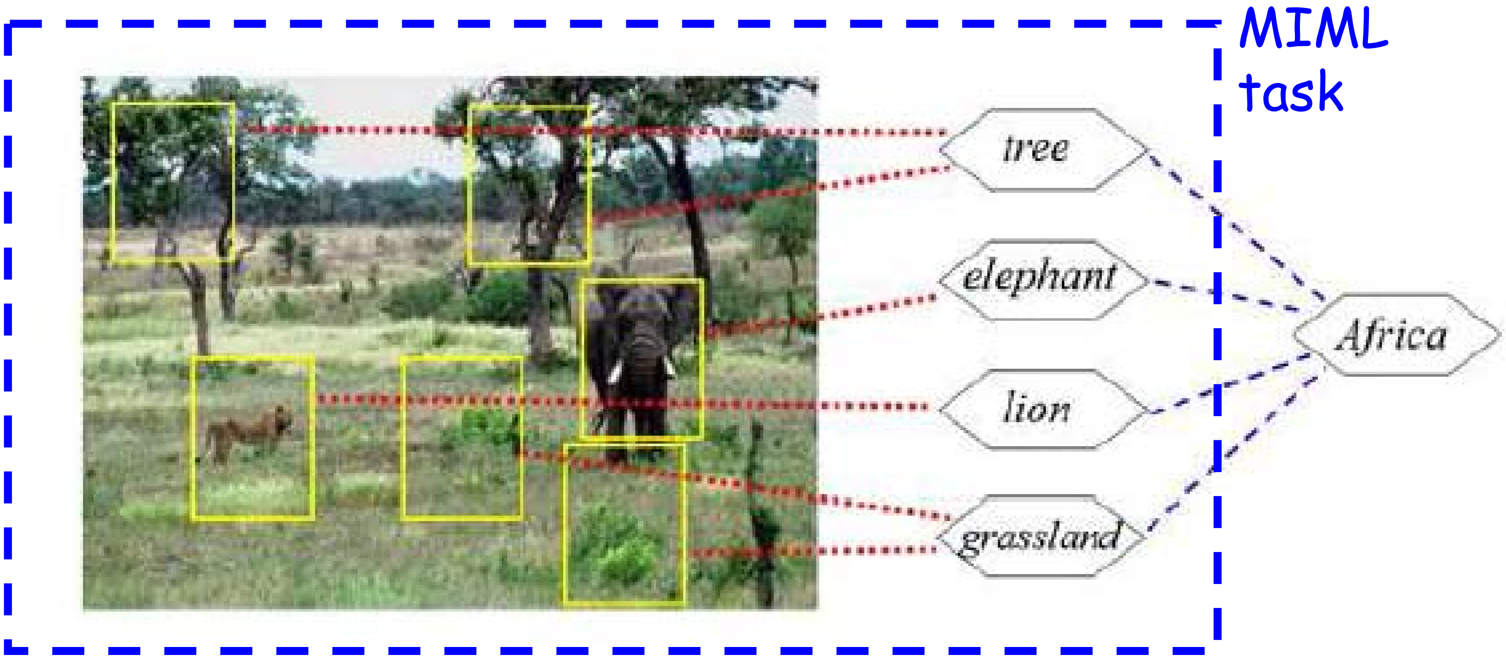
Why MIML ? (con't)

MIML can also be helpful for learning single-label examples involving complicated high-level concepts

(a) *Africa* is a complicated high-level concept

Why MIML ? (con't)

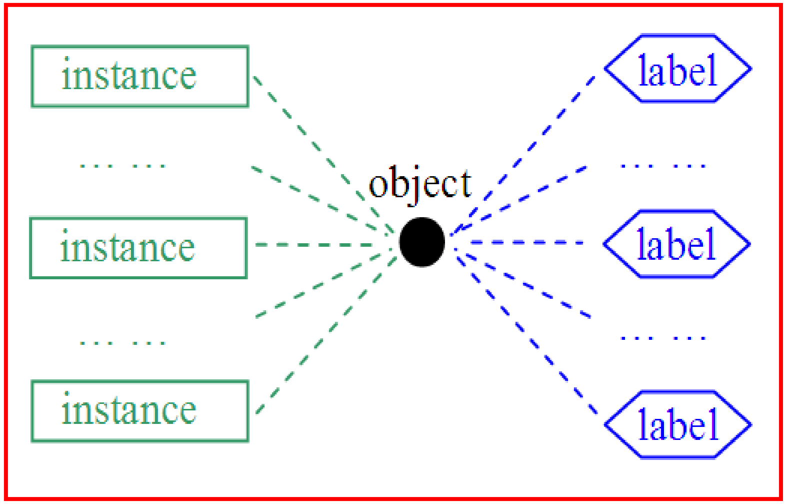
MIML can also be helpful for learning single-label examples involving complicated high-level concepts



(b) The concept *Africa* may become easier to learn through exploiting some sub-concepts

Multi-Instance Multi-Label Learning

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



MIML task:

To learn a function $f_{MIML}: 2^{\mathcal{X}} \rightarrow 2^{\mathcal{Y}}$ from a given data set $\{(X_1, Y_1), (X_2, Y_2), \cdots, (X_m, Y_m)\}$, where $X_i \subseteq \mathcal{X}$ is a set of instances $\{\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(i)}, \cdots, \boldsymbol{x}_{n_i}^{(i)}\}$, $\boldsymbol{x}_j^{(i)} \in \mathcal{X}$ $(j = 1, 2, \cdots, n_i)$, and $Y_i \subseteq \mathcal{Y}$ is a set of labels $\{y_1^{(i)}, y_2^{(i)}, \cdots, y_{l_i}^{(i)}\}$, $y_k^{(i)} \in \mathcal{Y}$ $(k = 1, 2, \cdots, l_i)$.

MIML:

多示例多标记学习

$\mathcal{X}$ - the instance space

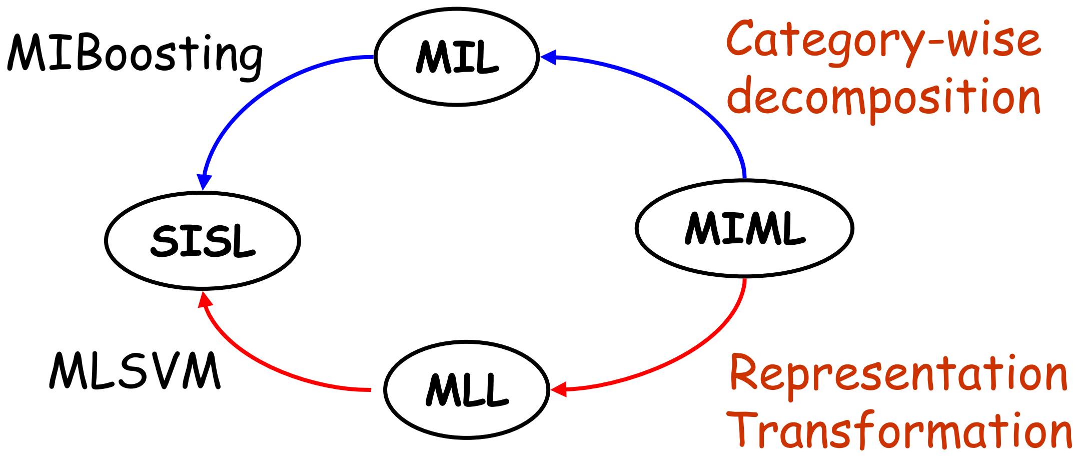
$\mathcal{Y}$ - the set of class labels

n_i - the number of instances in X_i

l_i - the number of labels in Y_i

MIMLBoost & MIMLSVM

MIMLBoost (an illustration of Solution 1)



MIMLSVM (an illustration of Solution 2)

unambiguous → ambiguous

MIMLBoost

Table 1: The MIMLBOOST algorithm

1 Transform each MIML example (X_u, Y_u) $(u = 1, 2, \cdots, m)$ into $|\mathcal{Y}|$ number of multi-instance bags $\{[(X_u, y_1), \Psi(X_u, y_1)], \cdots, [(X_u, y_{|\mathcal{Y}|}), \Psi(X_u, y_{|\mathcal{Y}|})]\}$. Thus, the original data set is transformed into a multi-instance data set containing $m \times |\mathcal{Y}|$ number of multi-instance bags, denoted by $\{[(X^{(i)}, y^{(i)}), \Psi(X^{(i)}, y^{(i)})]\}$ $(i = 1, 2, \cdots, m \times |\mathcal{Y}|)$.

2 Initialize weight of each bag to $W^{(i)} = \frac{1}{m \times |\mathcal{Y}|}$ $(i = 1, 2, \cdots, m \times |\mathcal{Y}|)$.

3 Repeat for $t = 1, 2, \cdots, T$ iterations:

- 3a Set $W_j^{(i)} = W^{(i)}/n_i$ $(i = 1, 2, \cdots, m \times |\mathcal{Y}|)$, assign the bag's label $\Psi(X^{(i)}, y^{(i)})$ to each of its instances $(\boldsymbol{x}_j^{(i)}, y^{(i)})$ $(j = 1, 2, \cdots, n_i)$, and build an instance-level predictor $h_t[(\boldsymbol{x}_j^{(i)}, y^{(i)})] \in \{-1, +1\}$.
- 3b For the ith bag, compute the error rate $e^{(i)} \in [0, 1]$ by counting the number of misclassified instances within the bag, i.e. $e^{(i)} = \frac{\sum_{j=1}^{n_i} [\![h_t[(\boldsymbol{x}_j^{(i)}, y^{(i)})] \neq \Psi(X^{(i)}, y^{(i)})]\!]}{n_i}$.
- 3c If $e^{(i)} < 0.5$ for all $i \in \{1, 2, \cdots, m \times |\mathcal{Y}|\}$, go to Step 4.
- 3d Compute $c_t = \arg\min_{c_t} \sum_{i=1}^{m \times |\mathcal{Y}|} W^{(i)} \exp[(2e^{(i)} - 1)c_t]$.
- 3e If $c_t \leq 0$, go to Step 4.
- 3f Set $W^{(i)} = W^{(i)} \exp[(2e^{(i)} - 1)c_t]$ $(i = 1, 2, \cdots, m \times |\mathcal{Y}|)$ and re-normalize such that $0 \leq W^{(i)} \leq 1$ and $\sum_{i=1}^{m \times |\mathcal{Y}|} W^{(i)} = 1$.

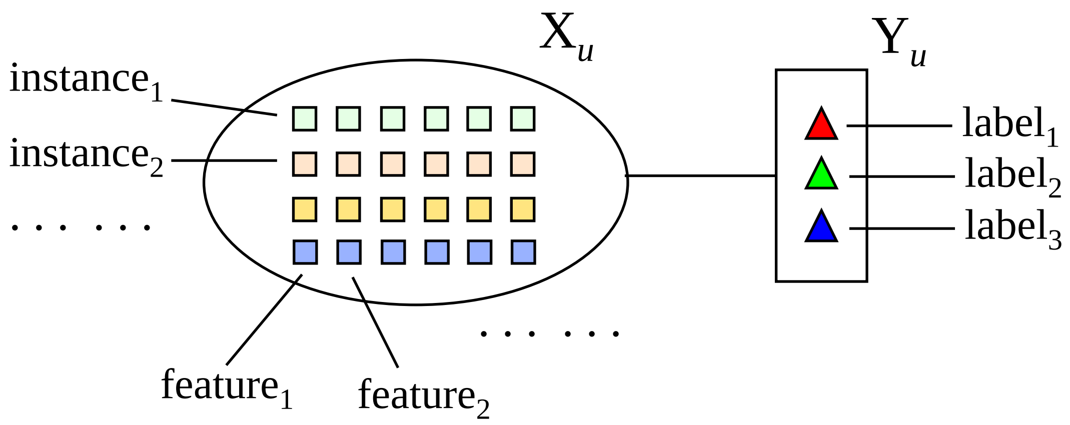
4 Return $Y^* = \{y | \arg_{y \in \mathcal{Y}} sign\left(\sum_j \sum_t c_t h_t[(\boldsymbol{x}_j^*, y)]\right) = +1\}$ ($\boldsymbol{x}_j^*$ is X^*'s jth instance).

[Z.-H. Zhou & M.-L. Zhang, NIPS'06]

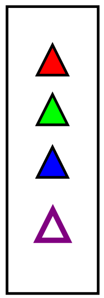
MIMLBoost

Illustration of the category-wise decomposition:

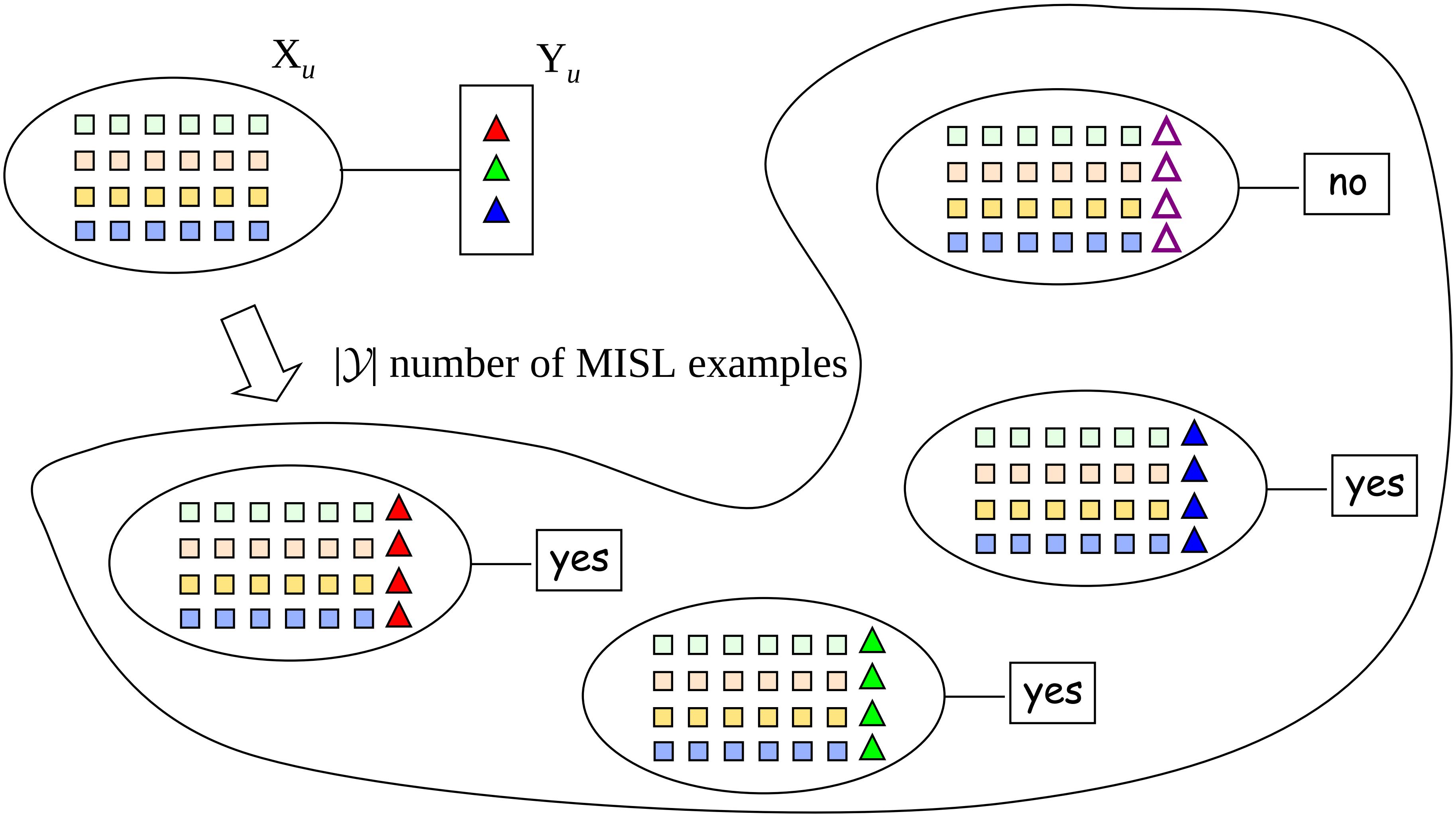
An MIML example (X_u, Y_u)



Label set $\mathcal{Y}$



MIMLBoost (con't)



MIMLSVM

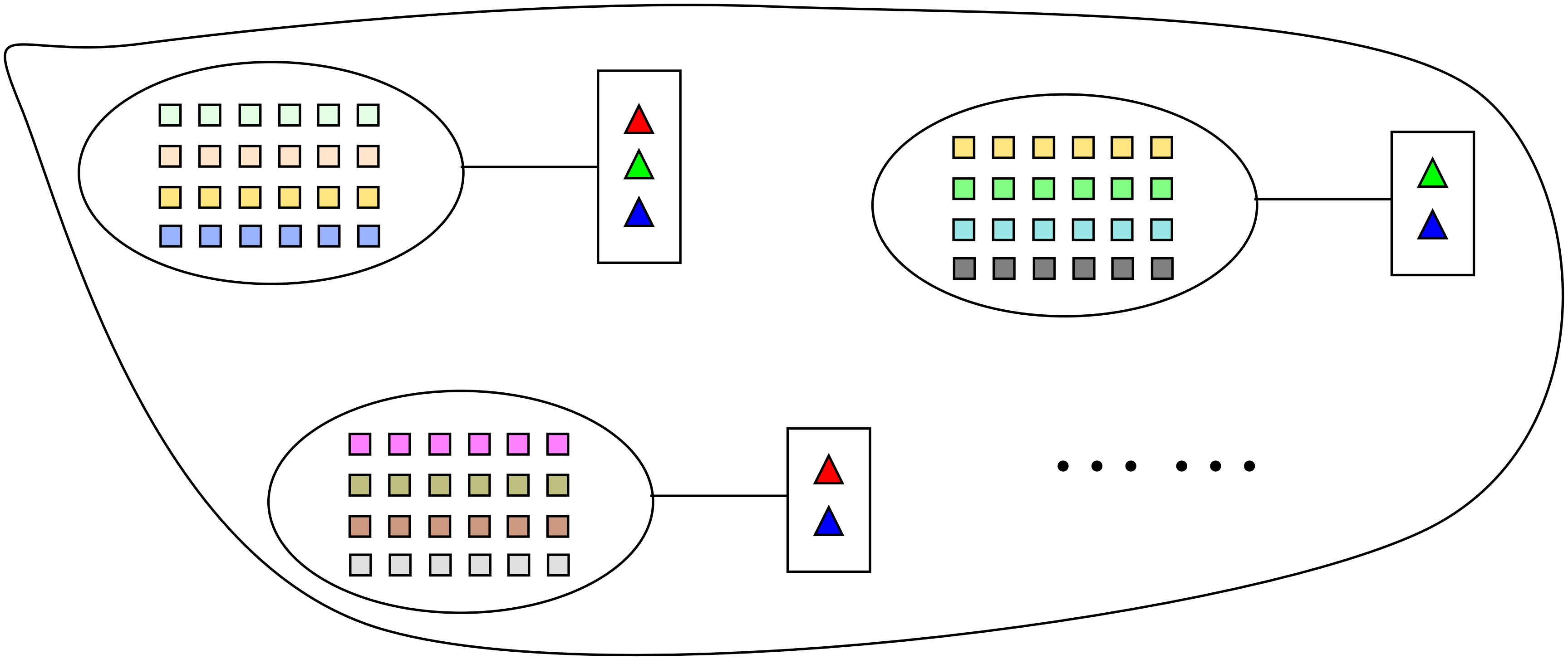
Table 2: The MIMLSVM algorithm

1	For MIML examples (X_u, Y_u) $(u = 1, 2, \cdots, m)$, $\Gamma = \{X_u \mid u = 1, 2, \cdots, m\}$.
2	Randomly select k elements from Γ to initialize the medoids M_t $(t = 1, 2, \cdots, k)$, repeat until all M_t do not change:
	2a $\quad \Gamma_t = \{M_t\}$ $(t = 1, 2, \cdots, k)$.
	2b $\quad$ Repeat for each $X_u \in (\Gamma - \{M_t \mid t = 1, 2, \cdots, k\})$: $index = \arg\min_{t\in\{1,\cdots,k\}} d_H(X_u, M_t)$, $\Gamma_{index} = \Gamma_{index} \cup \{X_u\}$.
	2c $\quad M_t = \arg\min\limits_{A\in\Gamma_t} \sum\limits_{B\in\Gamma_t} d_H(A, B)$ $(t = 1, 2, \cdots, k)$.
3	Transform (X_u, Y_u) into a multi-label example (z_u, Y_u) $(u = 1, 2, \cdots, m)$, where $z_u = (z_{u1}, z_{u2}, \cdots, z_{uk}) = (d_H(X_u, M_1), d_H(X_u, M_2), \cdots, d_H(X_u, M_k))$.
4	For each $y \in \mathcal{Y}$, derive a data set $\mathcal{D}_y = \{(z_u, \Phi(z_u, y)) \mid u = 1, 2, \cdots, m\}$, and then train an SVM $h_y = SVMTrain(\mathcal{D}_y)$.
5	Return $Y^* = \{\arg\max\limits_{y\in\mathcal{Y}} h_y(z^*)\} \cup \{y \mid h_y(z^*) \geq 0, y \in \mathcal{Y}\}$, where $z^* = (d_H(X^*, M_1), d_H(X^*, M_2), \cdots, d_H(X^*, M_k))$.

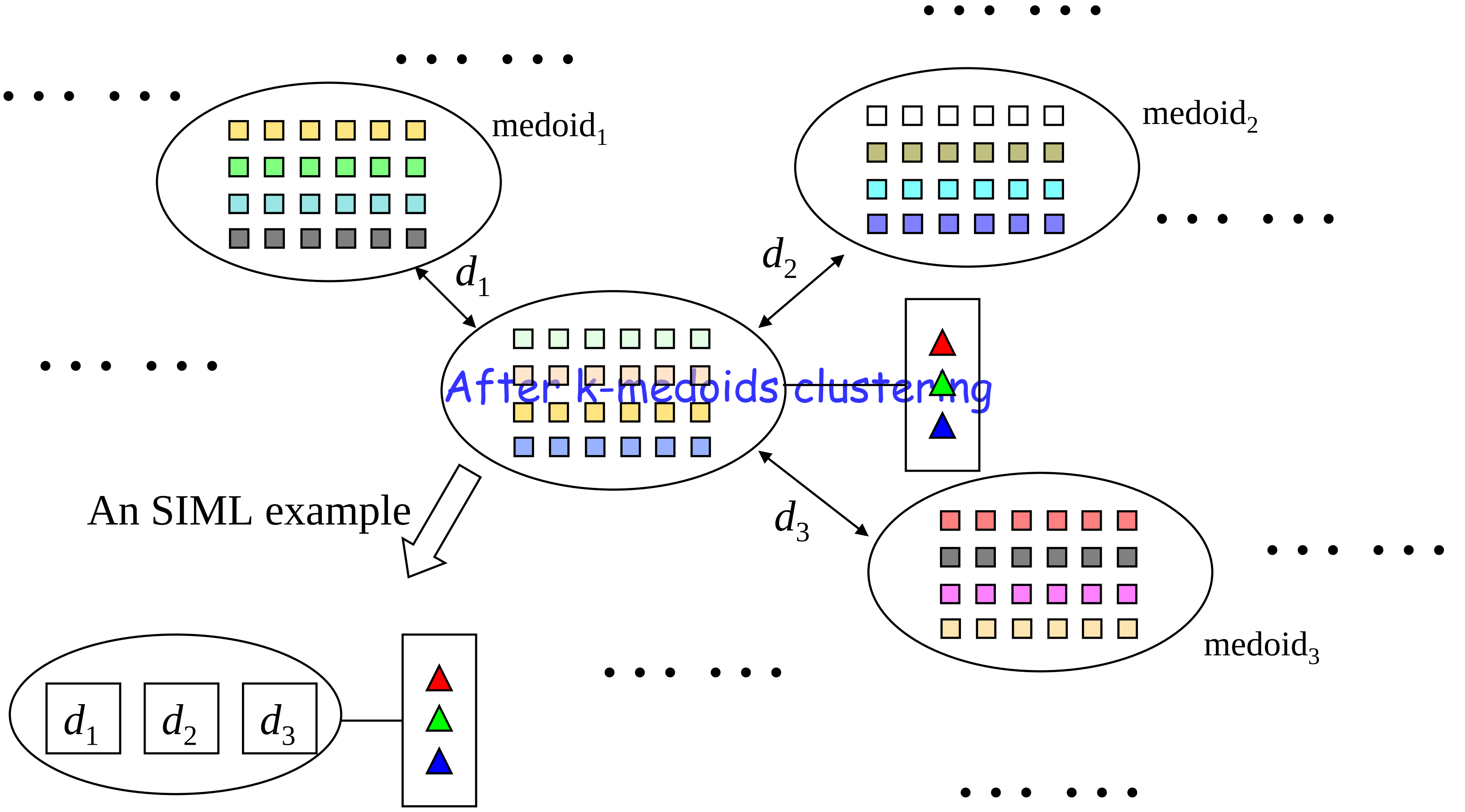
MIMLSVM

Illustration of the representation transformation:

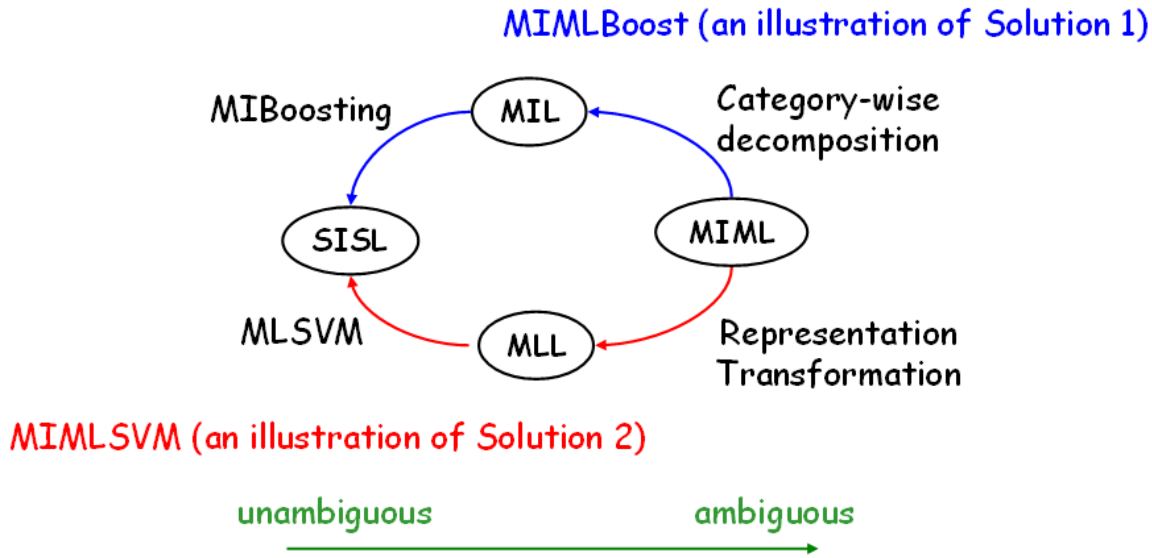
A set of MIML examples



MIMLSVM (con't)



Again, Why MIML?



- The MIML framework incorporates more information (+)
- These solutions degenerate MIML to solve, while the degeneration loses information (-)

If (+) > (-), then it is worth doing

Scene Classification: Result

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

Table 3
Results (mean±std.) on scene classification ('↓' indicates 'the smaller the better'; '↑' indicates 'the larger the better')

Compared Algorithms	Evaluation Criteria				
	hloss ↓	*one-error* ↓	*coverage* ↓	*rloss* ↓	*aveprec* ↑
MimlBoost	.192±.004	**.349±.016**	**.986±.041**	**.179±.008**	**.778±.009**
MimlSvm	**.190±.009**	.350±.020	1.083±.050	.201±.001	.766±.013
AdtBoost.MH	.210±.006	.436±.019	1.223±.049	N/A	.718±.012
RankSvm	.219±.020	.400±.062	1.177±.160	.225±.041	.739±.040
Ml-*k*NN	.191±.006	.370±.017	1.085±.047	.203±.010	.759±.010

The MIML algorithms are apparently superior to non-MIML algorithms

[Z.-H. Zhou et al, CORR abs/0808.3231]

Text Categorization: Result

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

Table 4
Results (mean±std.) on text categorization ('↓' indicates 'the smaller the better'; '↑' indicates 'the larger the better')

Compared Algorithms	Evaluation Criteria				
	hloss ↓	*one-error* ↓	*coverage* ↓	*rloss* ↓	*aveprec* ↑
MimlBoost	.054±.004	.092±.013	.401±.035	.037±.004	.937±.007
MimlSvm	**.034±.003**	**.071±.009**	**.315±.029**	**.024±.003**	**.955±.006**
AdtBoost.MH	.055±.004	.120±.016	.409±.046	N/A	.925±.010
RankSvm	.093±.007	.205±.055	.639±.161	.078±.027	.867±.037
Ml-*k*nn	.067±.005	.191±.017	.683±.052	.085±.008	.871±.010

The MIML algorithms are apparently superior to non-MIML algorithms

[Z.-H. Zhou et al, CORR abs/0808.3231]

MIML Results

Solving MIML problems by degeneration:

- MIMLBoost [Z.-H. Zhou & M.-L. Zhang, NIPS'06]
- MIMLSVM [Z.-H. Zhou & M.-L. Zhang, NIPS'06]

Solving MIML problems by regularization:

- D-MIMLSVM [Z.-H. Zhou et al., CORR abs/0808.3231]

Large margin MIML algorithm:

- M3MIML [M.-L. Zhang & Z.-H. Zhou, ICDM'08]

MIML Results (con't)

The usefulness of MIML when there is no access to raw objects:

- INSDIF [M.-L. Zhang & Z.-H. Zhou, AAAI'07]

MIML to help the learning of complicated high-level concepts:

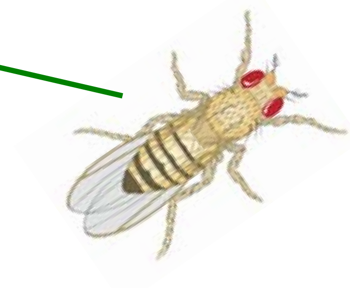
- SUBCOD [Z.-H. Zhou et al., CORR abs/0808.3231]

MIML for image annotation [Z.-J. Zha et al., CVPR'08]

MIML metric learning [R. Jin et al., CVPR'09]

Drosophila Gene Expression Pattern

Drosophila, or *fruit fly*, is a model organism widely studied in developmental biology



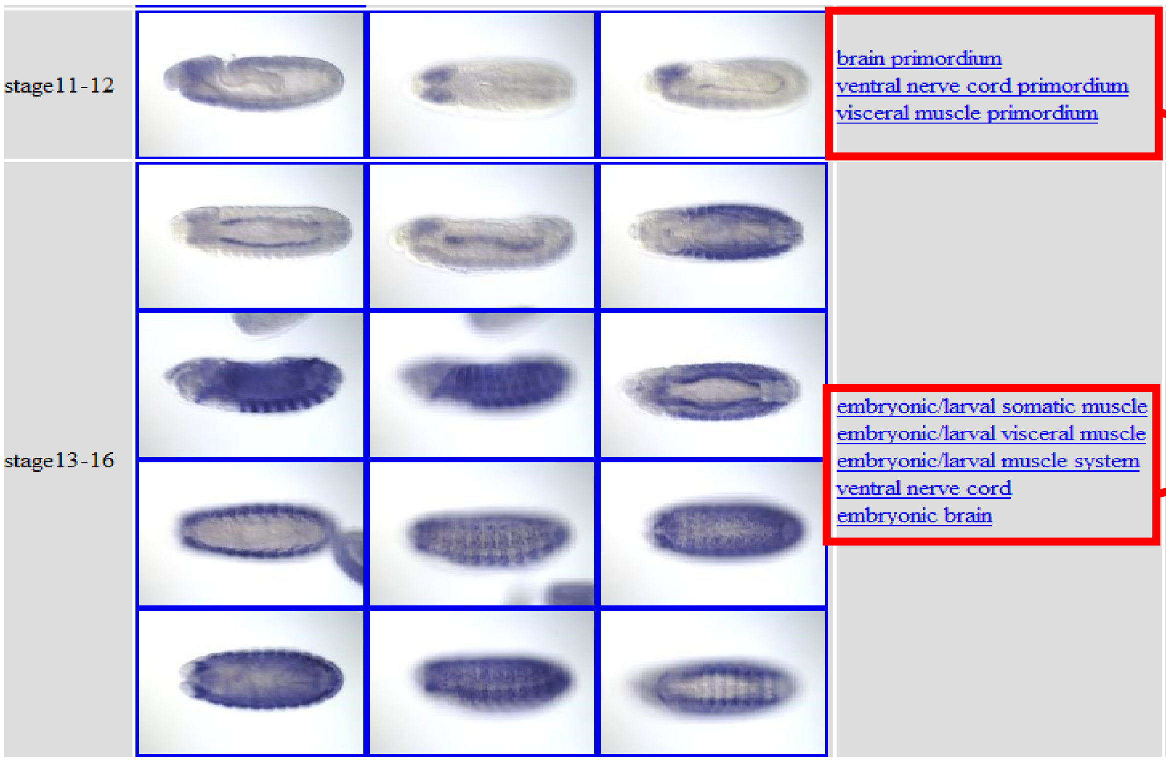
Gene RhoGAP71E expressed
stage: 7-8

Gene expression pattern by RNA *in situ* hybridization during Drosophila embryogenesis

stage（1-3） stage（4-6） stage（7-8） stage（9-10） stage（11-12） stage（13-16）

The BDGP Project

The Berkeley *Drosophila* Genome Project (BDGP) produced a large amount of spatial-temporal gene expression images

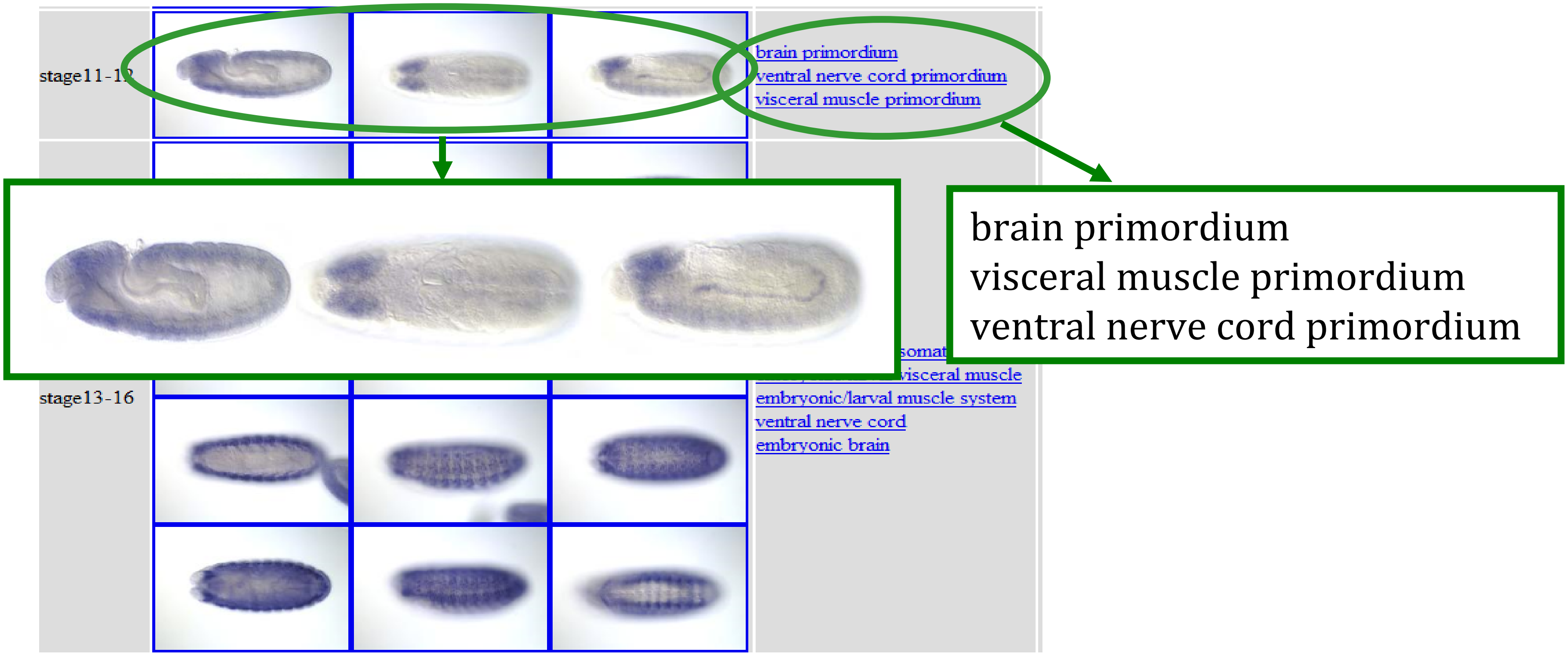


anatomical and developmental ontology terms manually labeled by human curators

Gene: Actn

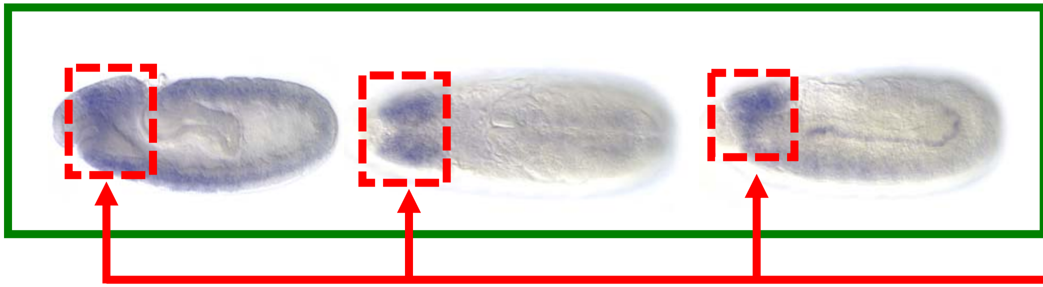
Learning And Mining from DatA
http://lamda.nju.edu.cn

Difficulty for Automatic Annotation



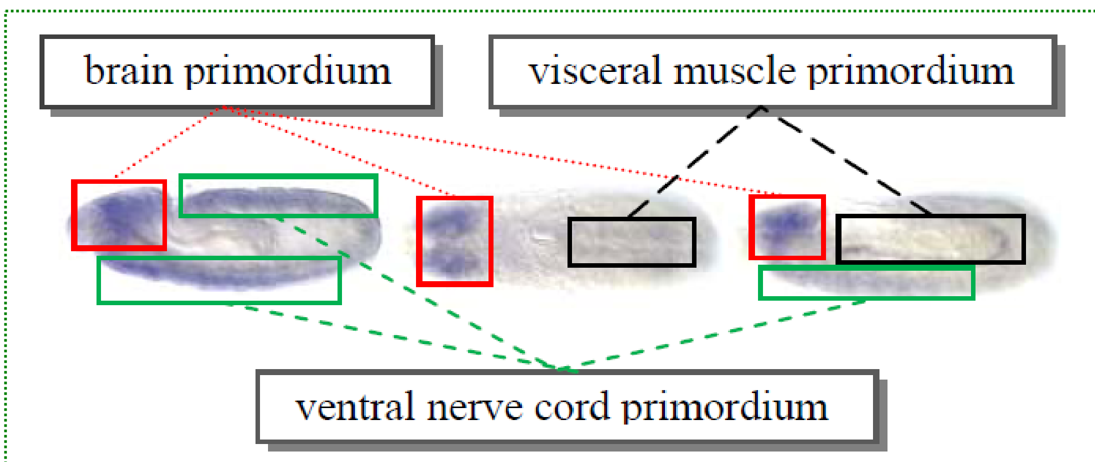
Difficulty for Automatic Annotation

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



brain primordium
visceral muscle primordium
ventral nerve cord primordium

The terms are body-part related



We do not know which term is associated with which region in the images !!

Generality of the Problem

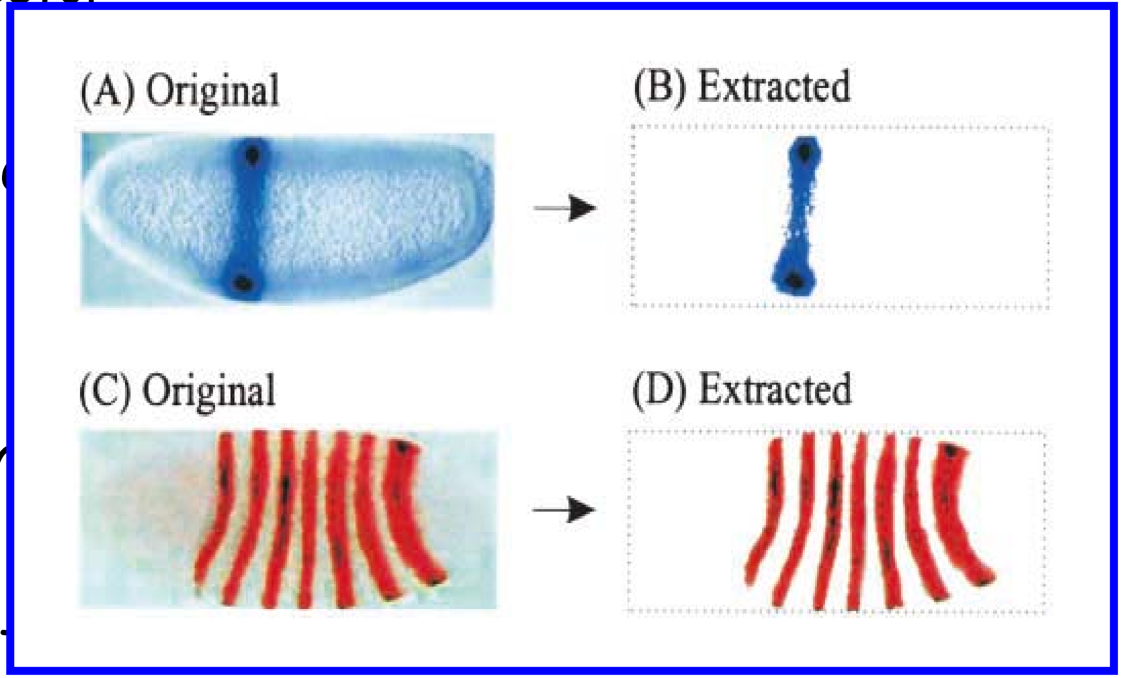
A good solution to the *Drosophila* gene expression pattern annotation task will also benefit other bio-problems

e.g., Protein functional prediction

- ✓ many conformations, varying functions
- ✓ Lack knowledge of which conformation is responsible for a specific function

Previous Solutions

- ✓ BESTi Algorithm [Kumar et al., Genetics02]
  - use images from literatures
  - use binary feature vector

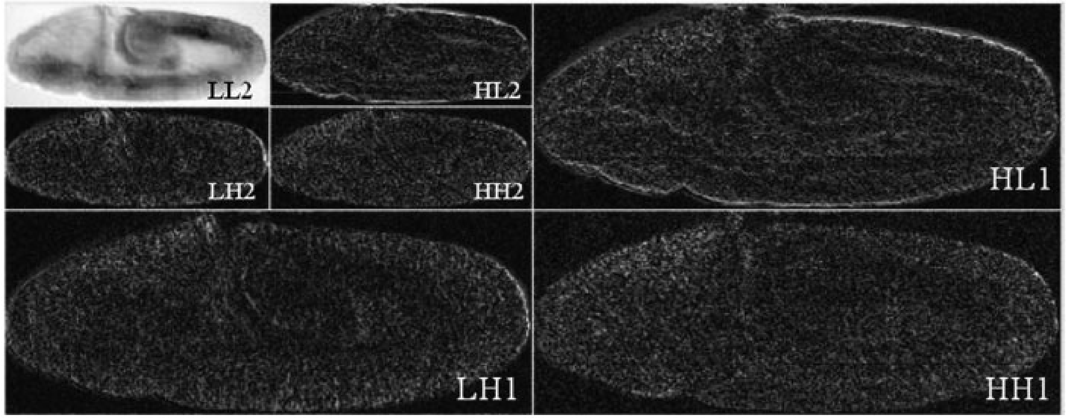


- ✓ 2D Wavelet featur
  - use BDGP images
- ✓ Multi-kernel learnin
  - use BDGP images
  - use multi-pyramid ma

Previous Solutions

- ✓ BESTi Algorithm [Kumar et al., Genetics02]
  - use images from literatures
  - use binary feature vector

- ✓ 2D Wavelet features, LDA classifier
  - use BDGP images



- ✓ Multi-kernel lear
  - use BDGP images
  - use multi-pyramid

Previous Solutions

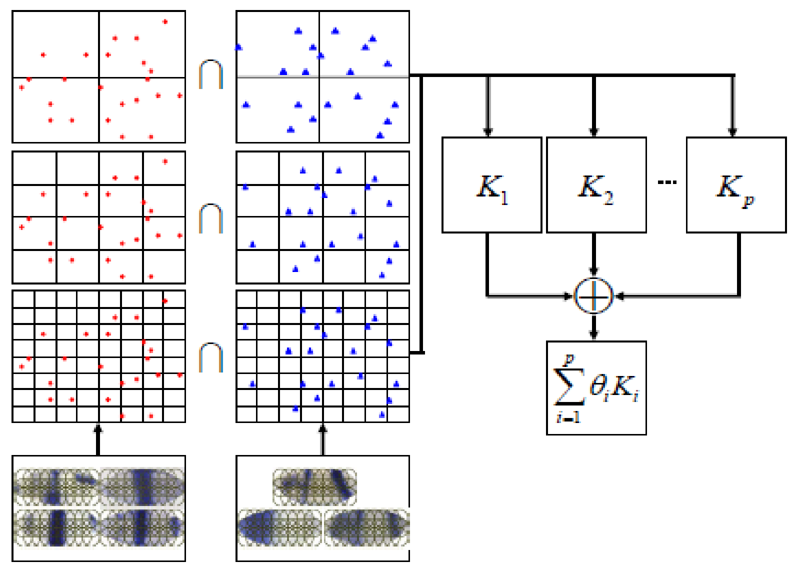
✓ BESTi ... al., Genetics02]

- use im

- use bi

✓ 2D Wa

- use BD ... ioinformatics07]

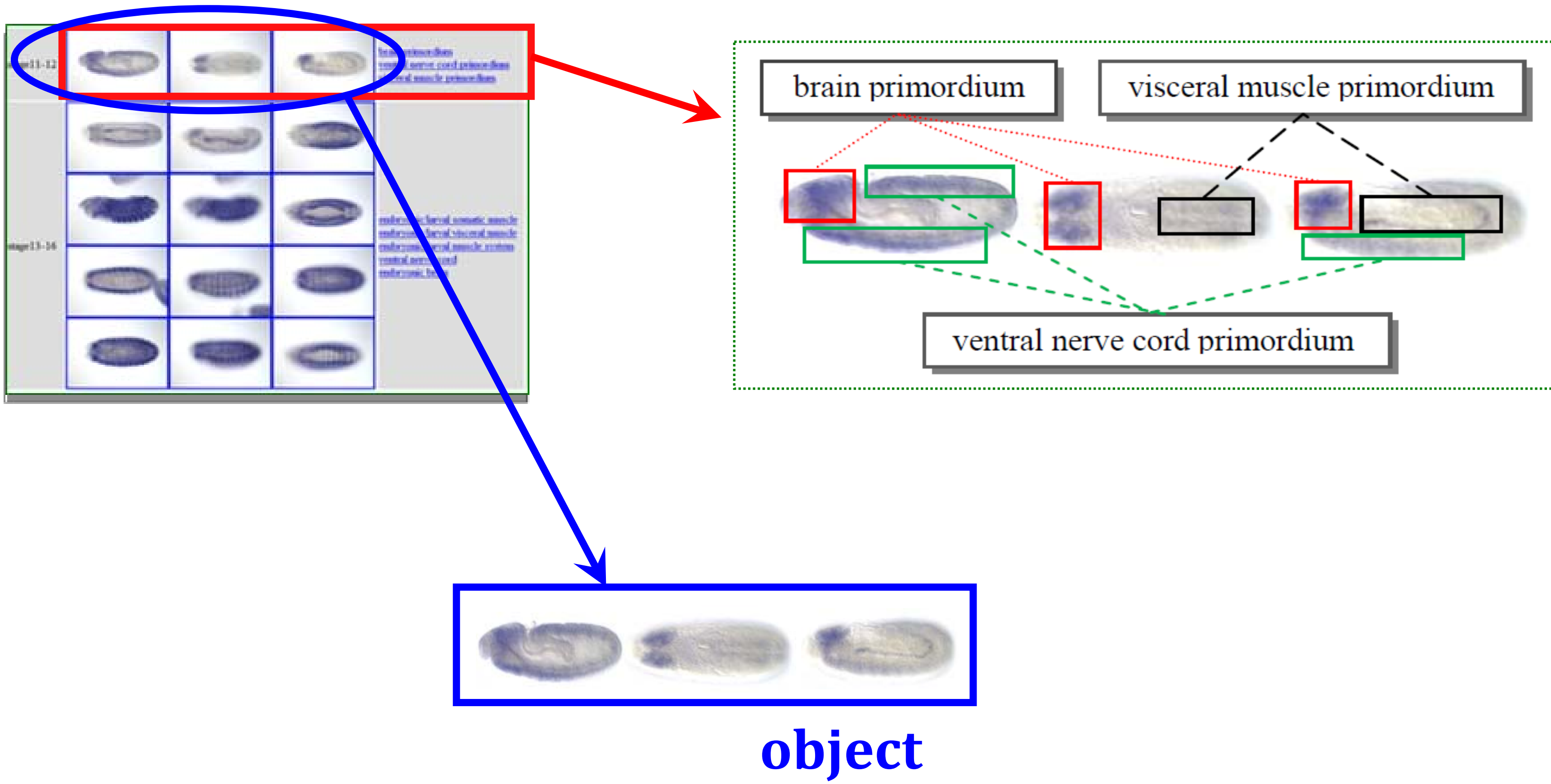


✓ Multi-kernel learning with hypergraph

- use BDGP images [S. Ji et al., Bioinformatics08]

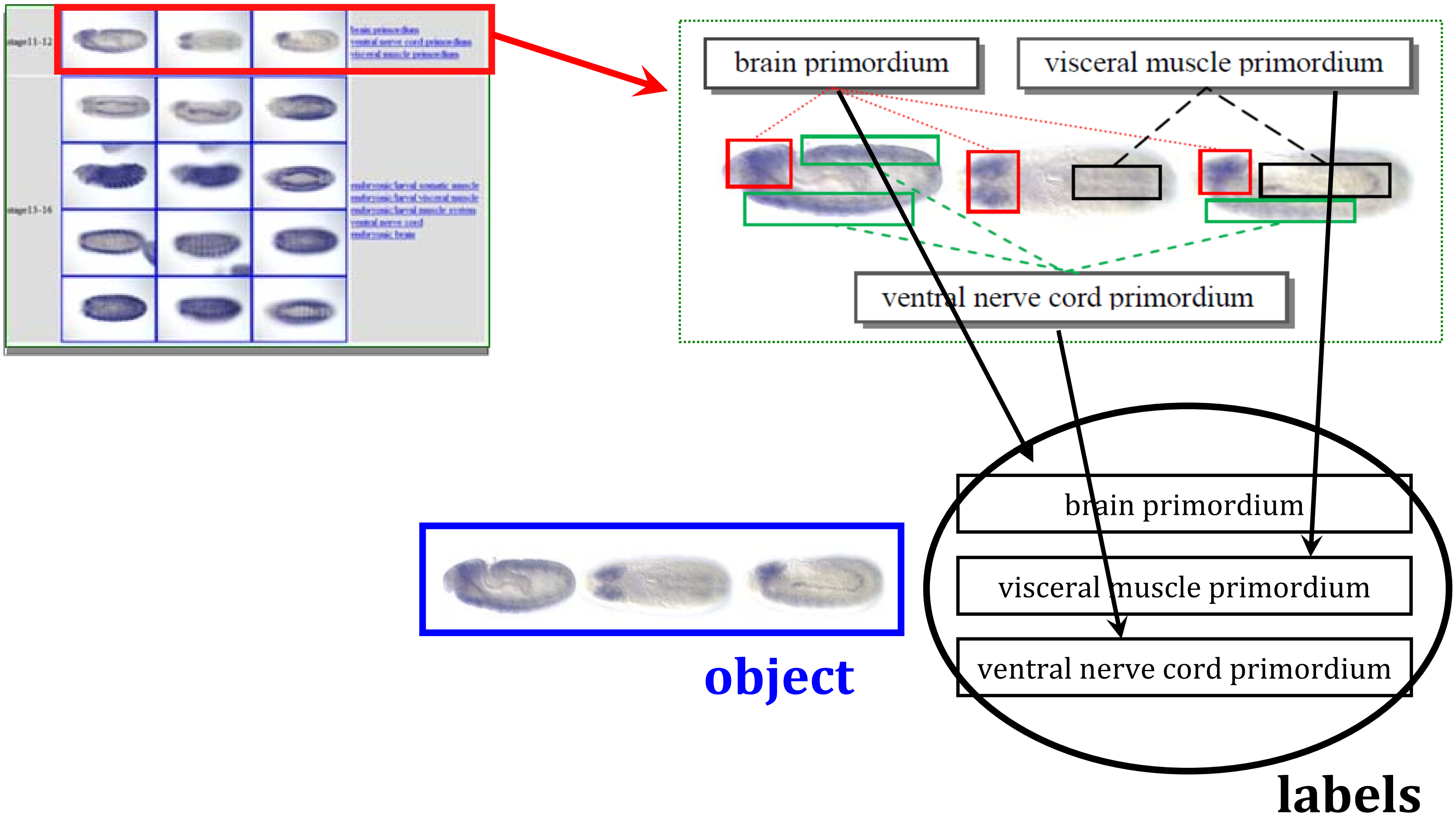
- use multi-pyramid match kernel and hypergraph learning

Formulated as an MIML Problem



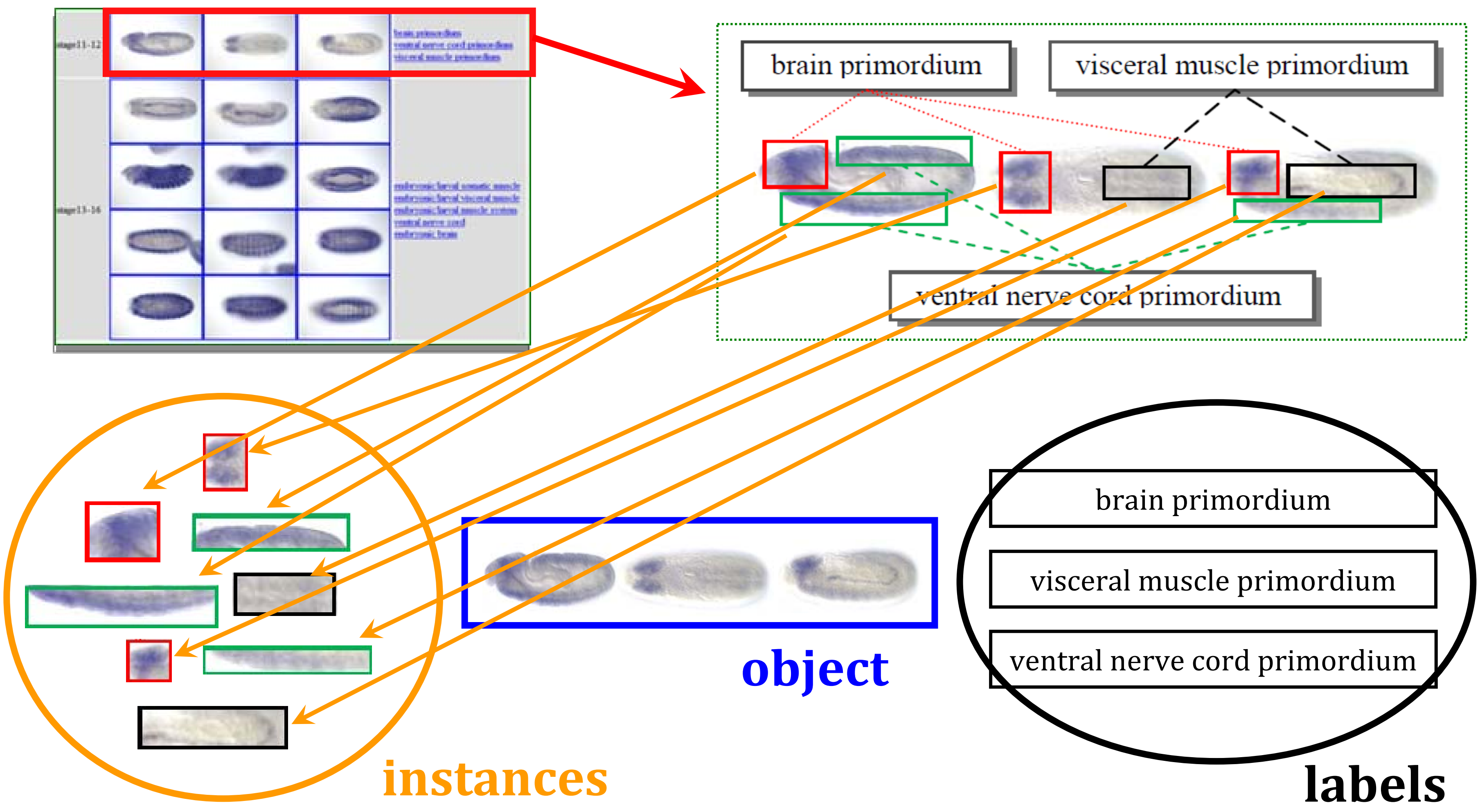
Formulated as an MIML Problem

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

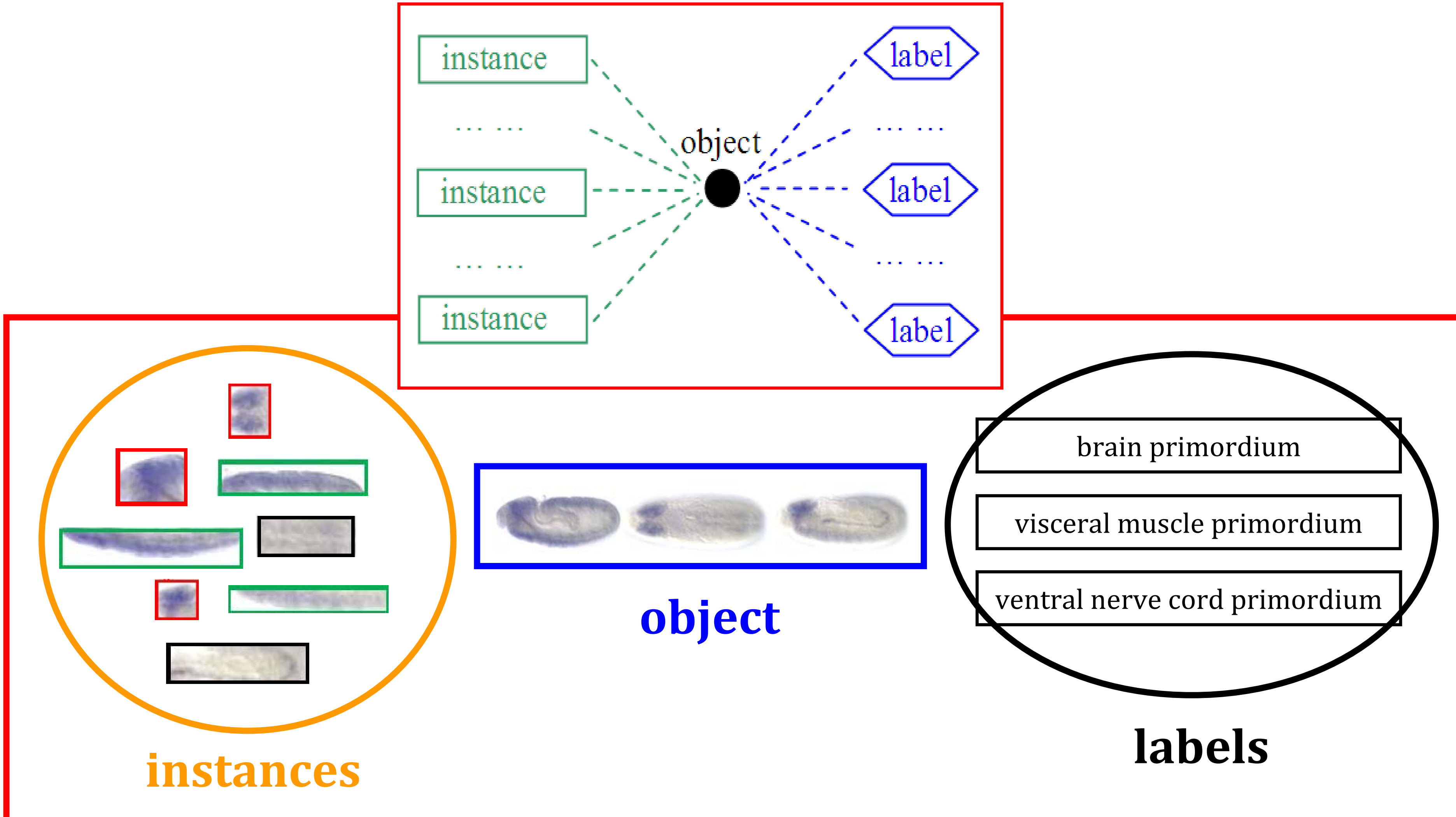


Learning And Mining from DatA
http://lamda.nju.edu.cn

Formulated as an MIML Problem



Formulated as an MIML Problem (con't)



The MIMLSVM+ Algorithm

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

For each label $y \in \mathcal{Y}$, let $\varphi(X_i, y) = +1$ if $y \in Y_i$ and -1 otherwise

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + \boxed{C^+} \sum_{\varphi(X_i,y)=1} \xi_i + \boxed{C^-} \sum_{\varphi(X_i,y)=-1} \xi_i$$

$$\text{subject to:} \quad \varphi(X_i, y)(w'\phi(X_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0 \qquad\qquad (i = 1, 2, \cdots, n)$$

We set $C^+ > C^-$ to make the classifier biased toward positive class

The MIMLSVM+ Algorithm

For each label $y \in \mathcal{Y}$, let $\varphi(X_i, y) = +1$ if $y \in Y_i$ and -1 otherwise

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C^+ \sum_{\varphi(X_i,y)=1} \xi_i + C^- \sum_{\varphi(X_i,y)=-1} \xi_i$$

$$\text{subject to:} \quad \varphi(X_i, y)(w' \boxed{\phi(X_i)} + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0 \qquad\qquad (i = 1, 2, \cdots, n)$$

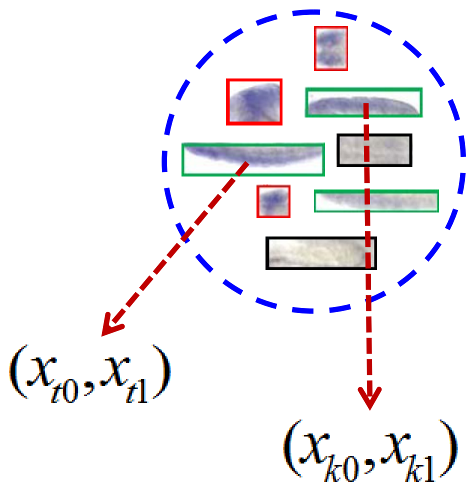
This involves a kernel function mapping a <u>bag of instances</u> into kernel space. We simply use the *set kernel*:

$$K_{SET}(X, X') = \sum_{i=1}^{n} \sum_{j=1}^{m} K(x_i, x_j')$$

[Y.-X. Li et al, IJCAI'09]

Features Used to Describe Instances

- ✓ visual features of gene expression of patches
- ✓ spatial information of patches



$$X_i = \{x_t\} = \{x_{t0}, x_{t1}\}$$

visual features	spatial information

$$K(x_t, x_k) = e^{-\gamma_1 \|x_{t0} - x_{k0}\|^2 - \gamma_2 \|x_{t1} - x_{k1}\|^2}$$

Experimental Configuration

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

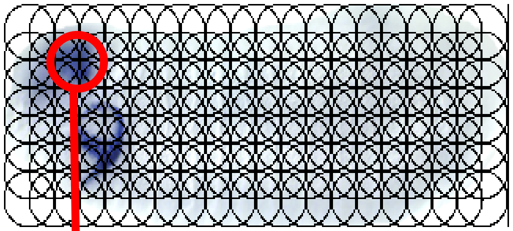
Dataset

2,816 bags, **2,052,722 instances** (15,434 x 133), **119 labels**

(2,816 image groups, 15,434 images, 133 instances per image, 119 terms)

Feature

SIFT on dense regular patches
Center coordinates of patches



sift & coordinates

Evaluation Measures

Extended from traditional measures

- ✓ Macro-F1 the larger, the better
- ✓ Micro-F1 the larger, the better
- ✓ AUC (Area under ROC curve) the larger, the better

Multi-Label measures

- ✓ Average precision the larger, the better
- ✓ One-error the smaller, the better
- ✓ Coverage the smaller, the better
- ✓ Ranking loss the smaller, the better
- ✓ Hamming loss the smaller, the better

Compared Methods

Existing methods

- ✓ MKL-PMK [S. Ji et al., Bioinformatics08]
- ✓ MIML-SVM [Z.-H. Zhou and M.-L. Zhang, NIPS'06]

Degenerated variants of MIMLSVM+

- ✓ MIML-SVM_{sv}^{+} : Concatenate visual and spatial information
- ✓ MIML-SVM_{v}^{+} : Use only visual features

Experimental Results

Learning And Mining from DatA

http://lamda.nju.edu.cn

50% train 50% test, 30 runs with random partitions

# terms	# groups	Algorithms	macro-F1 ↑	micro-F1 ↑	AUC ↑	Ave. Precision ↑	one-error ↓	coverage ↓	Rankloss ↓	Hammloss ↓
10	2228	MIMLSVM$^+$	**0.643±0.011**	**0.689±0.007**	**0.883±0.004**	**0.779±0.005**	**0.272±0.008**	**2.994±0.056**	**0.150±0.006**	**0.150±0.004**
		MIMLSVM$^+_{SV}$	0.627±0.010	0.676±0.006	0.869±0.004	0.773±0.005	0.277±0.011	3.073±0.048	0.157±0.004	0.156±0.003
		MIMLSVM^+_V	0.619±0.011	0.667±0.007	0.863±0.004	0.764±0.005	0.291±0.009	3.139±0.044	0.164±0.004	0.160±0.003
		MKL-PMK	0.584±0.009	0.621±0.009	0.825±0.006	0.722±0.007	0.343±0.011	3.483±0.072	0.198±0.006	0.196±0.006

# terms	# groups	Algorithms	macro-F1 ↑	micro-F1 ↑	AUC ↑	Ave. Precision ↑	one-error ↓	coverage ↓	Rankloss ↓	Hammloss ↓
20	2476	MIMLSVM$^+$	**0.468±0.015**	**0.587±0.007**	**0.862±0.003**	**0.673±0.008**	**0.357±0.011**	**6.189±0.117**	**0.152±0.005**	**0.114±0.002**
		MIMLSVM$^+_{SV}$	0.454±0.012	0.574±0.008	0.845±0.003	0.660±0.009	0.364±0.013	6.481±0.119	0.163±0.005	0.118±0.003
		MIMLSVM^+_V	0.445±0.012	0.566±0.006	0.840±0.004	0.651±0.008	0.377±0.011	6.609±0.114	0.169±0.004	0.119±0.002
		MKL-PMK	0.410±0.007	0.506±0.006	0.771±0.006	0.580±0.007	0.445±0.009	8.082±0.122	0.230±0.005	0.144±0.003

# terms	# groups	Algorithms	macro-F1 ↑	micro-F1 ↑	AUC ↑	Ave. Precision ↑	one-error ↓	coverage ↓	Rankloss ↓	Hammloss ↓
30	2646	MIMLSVM$^+$	**0.368±0.012**	**0.541±0.007**	**0.850±0.003**	**0.623±0.007**	**0.377±0.010**	**9.406±0.173**	**0.153±0.003**	**0.087±0.002**
		MIMLSVM$^+_{SV}$	0.354±0.001	0.527±0.006	0.829±0.004	0.605±0.007	0.388±0.010	9.964±0.195	0.166±0.004	0.090±0.002
		MIMLSVM^+_V	0.340±0.012	0.517±0.007	0.822±0.004	0.596±0.007	0.399±0.010	10.183±0.189	0.171±0.004	0.091±0.002
		MKL-PMK	0.310±0.008	0.455±0.008	0.741±0.007	0.511±0.008	0.488±0.011	13.010±0.2413	0.243±0.006	0.142±0.003

MIMLSVM+ achieves the best performance on ALL cases and ALL evaluation measures

Experimental Results (con't)

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

Since MIMLSVM could not work on the previous large data sets, we extract a smaller data set via random sampling

167 bags, **57,323 instances** (431 x 133), **10 labels**

(167 image groups, 431 images, 133 inst per image, 10 terms)

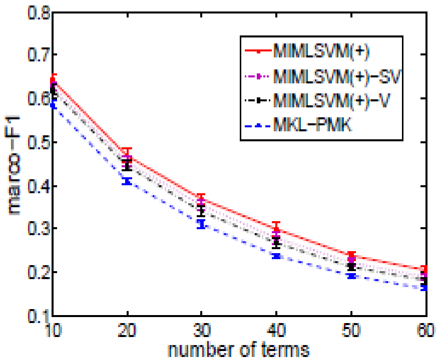
20 runs with random splits of training/test sets

# terms	# groups	Algorithms	macro-F1 ↑	micro-F1 ↑	AUC ↑	Ave. Precision ↑	one-error ↓	coverage ↓	Rankloss ↓	Hammloss ↓
10	167	MIMLSVM$^+$	**0.460±0.041**	**0.606±0.026**	**0.807±0.191**	**0.733±0.019**	**0.311±0.034**	**3.508±0.262**	**0.186±0.015**	**0.171±0.019**
		MIMLSVM$^+_{SV}$	0.424±0.049	0.569±0.033	0.774±0.017	0.710±0.027	0.354±0.047	3.667±0.199	0.204±0.016	0.191±0.015
		MIMLSVM	0.176±0.047	0.367±0.054	0.629±0.041	0.592±0.028	0.468±0.060	4.792±0.300	0.318±0.029	0.241±0.097

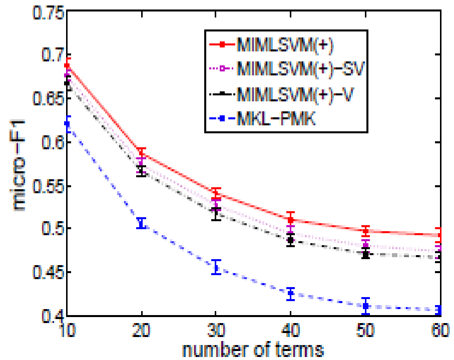
MIMLSVM+ achieves the best performance on ALL evaluation measures

[Y.-X. Li et al, IJCAI'09]

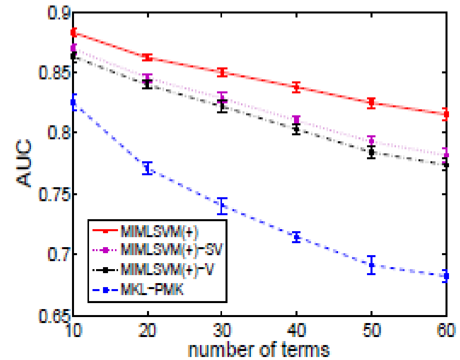
Experimental Results (con't)



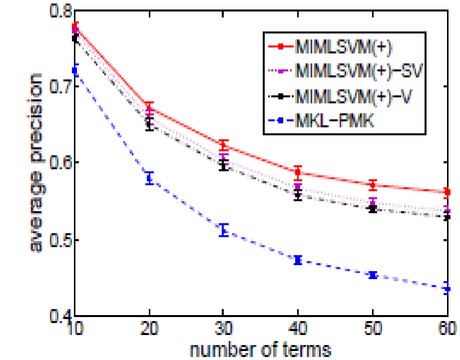
(a) macro-F1 $\uparrow$



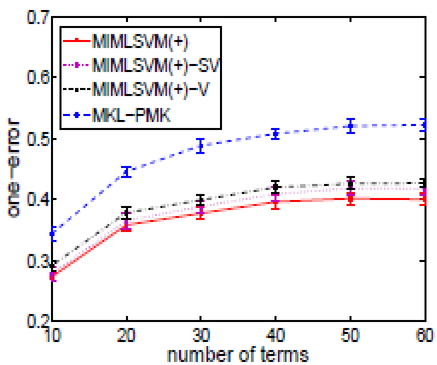
(b) micro-F1 $\uparrow$



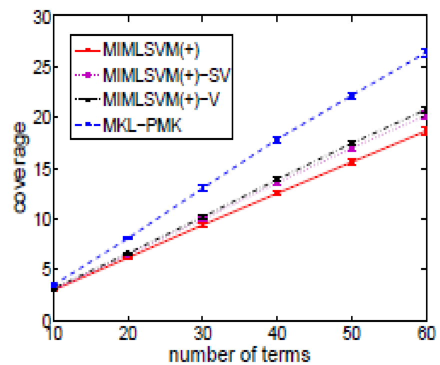
(c) AUC $\uparrow$



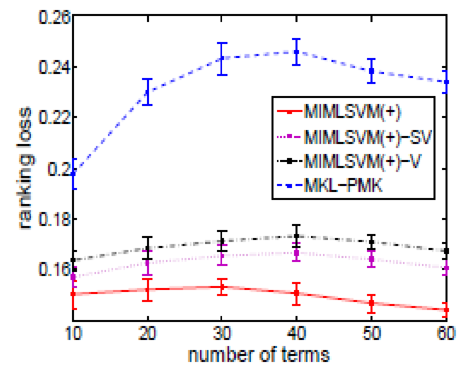
(d) average precision $\uparrow$



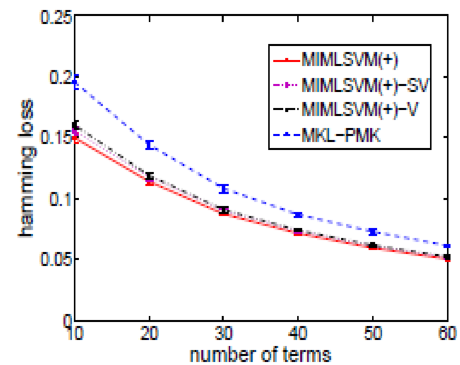
(e) one-error $\downarrow$



(f) coverage $\downarrow$



(g) ranking loss $\downarrow$



(h) hamming loss $\downarrow$

The comparison under different number of labels (annotation terms)

[Y.-X. Li et al, IJCAI'09]

MIML Papers

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

- ✓ Y.-X. Li, S. Ji, J. Ye, S. Kumar, and Z.-H. Zhou. Drosophila gene expression pattern annotation through multi-instance multi-label learning. IJCAI'09
- ✓ S. Wang, R. Jin, and Z.-H. Zhou. Learn a distance metric from multi-instance multi-label data. CVPR'09
- ✓ M.-L. Zhang and Z.-H. Zhou. M^{3}MIML: A maximum margin method for multi-instance multi-label learning. ICDM'08
- ✓ Z.-H. Zhou, M.-L. Zhang, S.-J. Huang, Y.-F. Li. MIML: A Framework for Learning with Ambiguous Objects. CORR abs/0808.3231
- ✓ M.-L. Zhang, Z.-H. Zhou. Multi-label learning by instance differentiation. AAAI'07, pp.669-674
- ✓ Z.-H. Zhou, M.-L. Zhang. Multi-instance multi-label learning with application to scene classification. NIPS'06, pp.1609-1616.

MIML Resources

Codes:

- MIMLBoost & MIMLSVM:

  http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/MIMLBoost&MIMLSVM.htm

- InsDif:

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

- M^3MIML:

  http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/M3MIML.htm

Data:

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

Thanks!

机器学习周(南京大学)

Machine Learning Week (Nanjing University)

南京大学
Nanjing University

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

ACML2009 Nov.2-4

The 1st Asian Conference on Machine Learning

The 1st Asian Conference on Machine Learning (ACML'09) aims at providing a leading international forum for researchers in machine learning and related fields to share their new ideas and achievements.

Keynote speaker:
Thomas Dietterich
(Oregon State University; Founding Chair of IMLS)

Invited speaker:
Masashi Sugiyama
(Tokyo Institute of Technology)
Qiang Yang
(Hong Kong University of Science and Technology)

Registration and other details please visit:
http://lamda.nju.edu.cn/conf/acml09

(ACML2009将为中国大陆学生提供低价旁听注册，详情请见会议网站)

MLA2009 11月6-8日

第七届机器学习及其应用研讨会

第七届“机器学习及其应用”研讨会由南京大学计算机软件新技术国家重点实验室主办，将于2009年11月6-8日在南京举行。该研讨会拟邀请海内外活跃在机器学习及相关领域研究第一线的10余位华人学者前来做特邀报告，研讨会仍将举行Top Conference Review（“过去一年顶级会议回顾”）以及墙展交流等活动。

按惯例，该研讨会不征文、不收费，欢迎机器学习及相关领域的学者、研究生前来旁听特邀报告并参加讨论。

具体情况请见：http://lamda.nju.edu.cn/conf/mla09

SSMLA2009 11月6-8日

第四届机器学习及其应用学生研讨会

为了促进机器学习及相关领域的研究生之间以及研究生与资深学者之间的交流，在第7届机器学习及其应用研讨会期间，将举行第4届机器学习及其应用学生研讨会（SSMLA'09）。SSMLA'09以Poster（墙展张贴）的方式进行。除部分特邀Poster外，SSMLA'09将公开征集学生Poster，通过评审的学生Poster将获得到场展示的机会。会议还将评出最佳学生Poster奖。

具体情况请见：http://lamda.nju.edu.cn/conf/ssmla09