Multi-instance multi-label learning in the presence of novel class instances

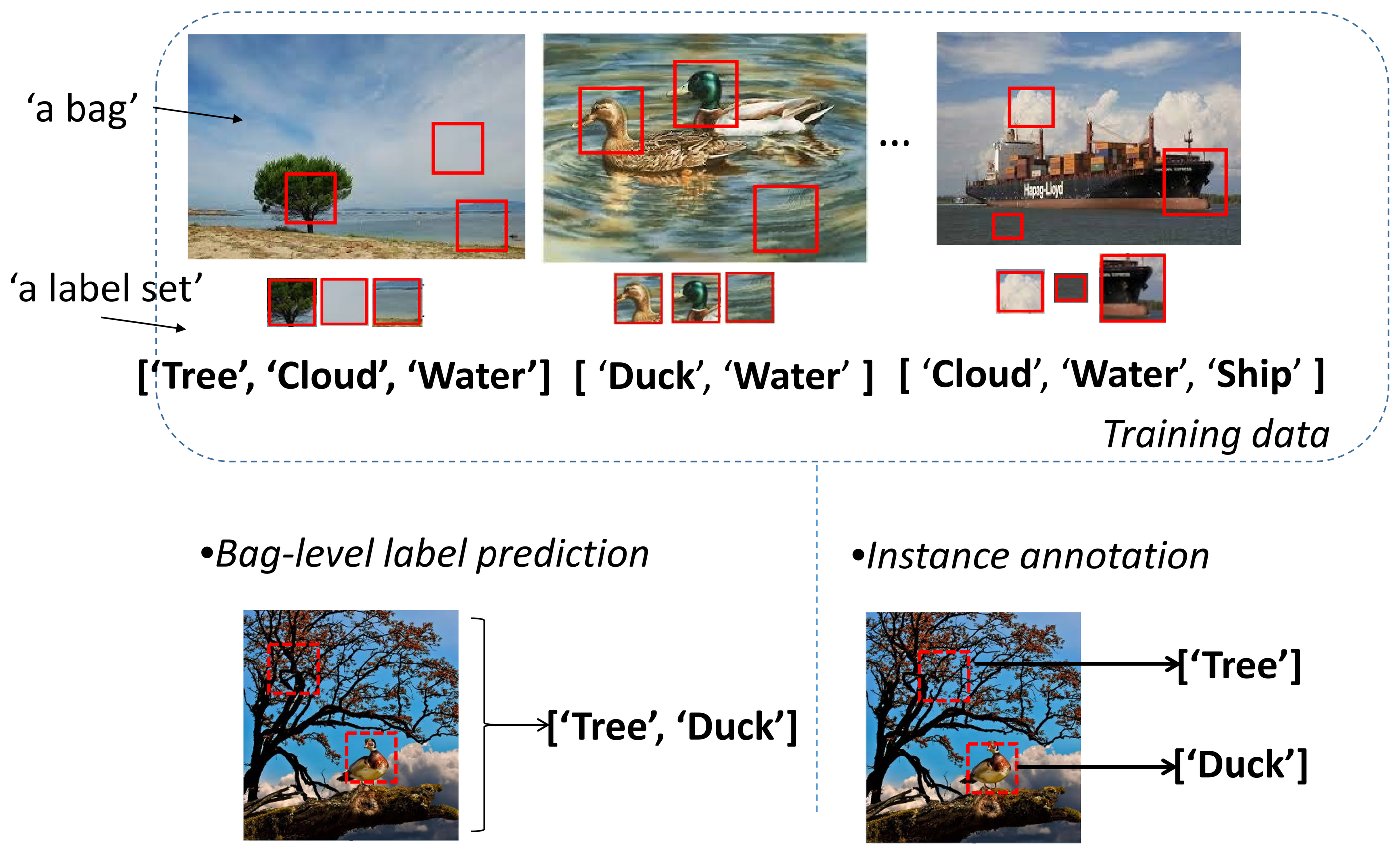
Anh T. Pham[1], Raviv Raich[1], Xiaoli Z. Fern[1], Jesús Pérez Arriaga[2]

[1]School of EECS, Oregon State University, Corvallis, OR 97331-5501, USA

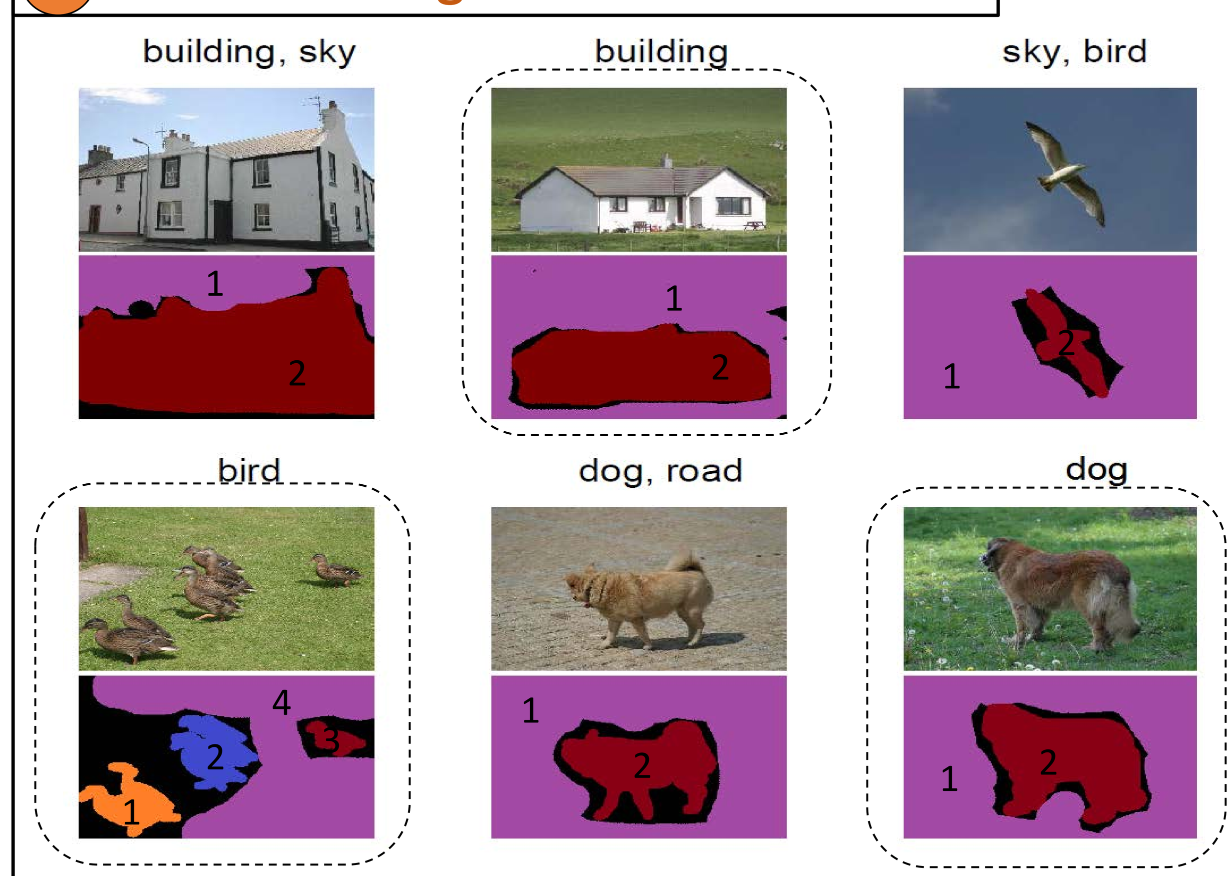
[2]Departamento de Ingenierıa de Comunicaciones, Universidad de Cantabria, 39005 Santander, Spain

{phaman,raich,xfern}@eecs.oregonstate.edu; jperez@gtas.dicom.unican.es

1 Multi-instance multi-label learning (MIML)



2 MIML learning with novel instances



Known classes:

- Building
- Bird
- Sky
- Dog
- Road

Novel class: (no label for it)

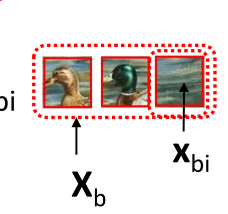
- Grass
- Other

There are grass segments (in boxed images) but there is no label 'grass'

3 Problem formulation

- Training data: B bags, denoted as $\{(\mathbf{X}_b, \mathbf{Y}_b)\}_{b=1}^{B}$

- $\mathbf{X}_b$ is a set of n_b instances for the b^{th} bag $\{\mathbf{x}_{b1}, \mathbf{x}_{b2}, \ldots, \mathbf{x}_{bn_b}\}$, where $\mathbf{x}_{bi} \in \mathcal{X} = \mathbb{R}^d$



- Each instance $\mathbf{x}_{bi}$ is associated with a label $y_{bi} \in \{0, 1, \ldots, C\}$, where C is the number of classes and 0 denotes novel class
- $\mathbf{Y}_b$ is the bag label for the b^{th} bag which is a subset of known labels $\mathcal{Y} = \{1,2,\ldots,C\}$

['Duck', 'Water'] → $\mathbf{Y}_b$

Goals:

1. *Instance annotation:* map an instance in $\mathcal{X}$ to a label in {0,1,2,…,C}
2. *Novelty detection:* map an instance in X to {0, $\mathcal{Y}$}
3. *Bag label prediction:* map a bag in $2^{\mathcal{X}}$ to $2^{\mathcal{Y}}$

4 Graphical model

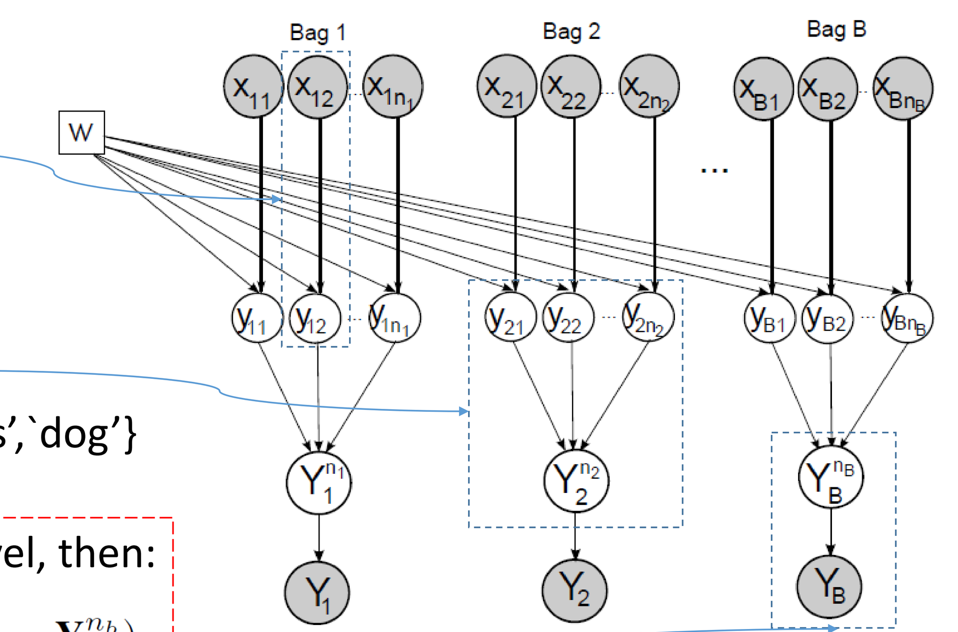
Logistic regression

Union relation

y_{21} ='grass', y_{22} ='dog' → $\mathbf{Y}_2^2$ ={'grass','dog'}

Labels are $\{0,1,2,\ldots,C\}$ where 0 is novel, then:

$$p(\mathbf{Y}_b|\mathbf{Y}_b^{n_b}) = I(\mathbf{Y}_b = \mathbf{Y}_b^{n_b}) + I(\mathbf{Y}_b \bigcup \{0\} = \mathbf{Y}_b^{n_b})$$



Graphical model for the problem

$\mathbf{Y}_b^{n_b}$ may include novel label 0 $\longrightarrow \mathbf{Y}_b^{n_b} \subseteq \{0,1,2,\ldots,C\}$

$\mathbf{Y}_b$ removes the novel label 0 from $\mathbf{Y}_b^{n_b} \longrightarrow \mathbf{Y}_b \subseteq \{1,\ldots,C\}$

5 Inference

- Maximum likelihood inference

$$p(\mathbf{Y}_D, \mathbf{X}_D|\mathbf{w}) = p(\mathbf{X}_D)\prod_{b=1}^{B} p(\mathbf{Y}_b|\mathbf{X}_b, \mathbf{w}) \text{ where}$$

$$p(\mathbf{Y}_b|\mathbf{X}_b, \mathbf{w}) = \sum_{y_{b1}=0}^{C} \cdots \sum_{y_{bn_b}=0}^{C} [\{I(\mathbf{Y}_b = \bigcup_{j=1}^{n_b} y_{bj}) + I(\mathbf{Y}_b \bigcup\{0\} = \bigcup_{j=1}^{n_b} y_{bj})\} \times \prod_{i=1}^{n_b} p(y_{bi}|\mathbf{x}_{bi}, \mathbf{w})]$$

No close form → EM

- Surrogate function:

$$g(\mathbf{w}, \mathbf{w}') = \sum_{b=1}^{B}\sum_{i=1}^{n_b}[\sum_{c=0}^{C} p(y_{bi} = c|\mathbf{Y}_b, \mathbf{X}_b, \mathbf{w}')\mathbf{w}_c^T\mathbf{x}_{bi} - \log(\sum_{c=0}^{C} e^{\mathbf{w}_c^T\mathbf{x}_{bi}})] + \zeta$$

- E-step:

Compute $p(y_{bi} = c|\mathbf{Y}_b, \mathbf{X}_b, \mathbf{w}^{(k)})$ Brute-force computation: $O(C^{n_b})$

- M-step:

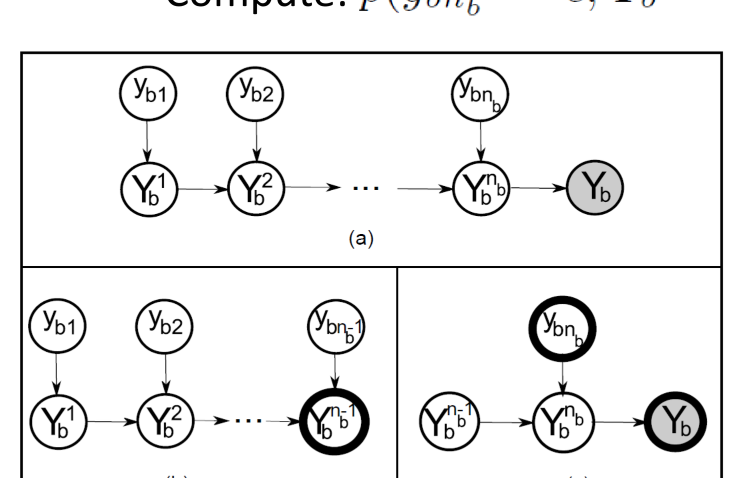
Find $\mathbf{w}^{(k+1)}$ such that $g(\mathbf{w}^{(k+1)}, \mathbf{w}^{(k)}) \geq g(\mathbf{w}^{(k)}, \mathbf{w}^{(k)})$

6 E-step: Compute $p(y_{bi} = c|\mathbf{Y}_b, \mathbf{X}_b, \mathbf{w}^{(k)})$

- Conditional rule:

$$p(y_{bi} = c|\mathbf{Y}_b = \mathbf{L}, \mathbf{X}_b, \mathbf{w}) = \frac{p(y_{bi} = c, \mathbf{Y}_b = \mathbf{L}|\mathbf{X}_b, \mathbf{w})}{\sum_{c \in \mathbf{L}\cup\{0\}} p(y_{bi} = c, \mathbf{Y}_b = \mathbf{L}|\mathbf{X}_b, \mathbf{w})}$$

- Compute: $p(y_{bn_b} = c, \mathbf{Y}_b = \mathbf{L}|\mathbf{X}_b, \mathbf{w})$ linear computational complexity w.r.t. n_b



- Introduce a partial bag label $\mathbf{Y}_b^i = \bigcup_{j=1}^{k} y_{bk}$

Recall: $p(\mathbf{Y}_b|\mathbf{Y}_b^{n_b}) = I(\mathbf{Y}_b = \mathbf{Y}_b^{n_b}) + I(\mathbf{Y}_b \bigcup\{0\} = \mathbf{Y}_b^{n_b})$

- Allows for a recursive computation as follows
  - Compute $p(\mathbf{Y}_b^1)$
  - Compute $p(\mathbf{Y}_b^2)$
  - ….
  - Finally, from $p(\mathbf{Y}_b^{n_b-1})$ and $p(y_{bn_b})$ → $p(y_{bn_b}, \mathbf{Y}_b)$

7 Dataset description

Synthetic dataset

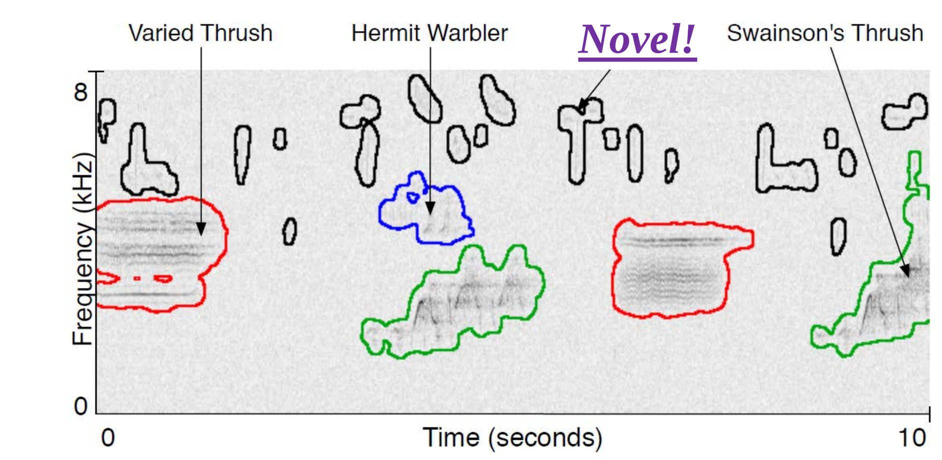
MNIST dataset

(1 is novel class)

Real dataset

Corel5k-10 dataset

novel instances are outside top 10 classes



HJA bird song dataset

novel instances: noise segments, rain…(Briggs et al.)

AUC metric for novelty detection

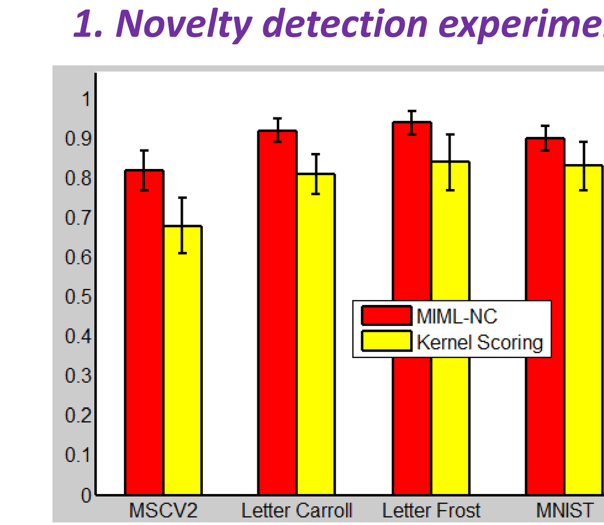
Hamming loss metric for bag label prediction

8 Experimental results

Baseline methods:

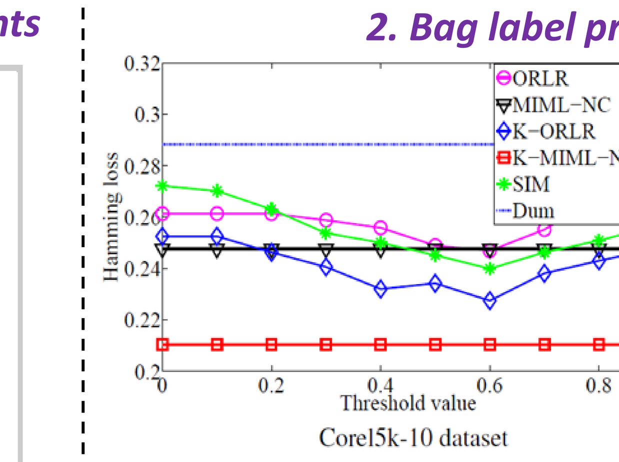
1. OR-Logistic Regression (**ORLR**) (*unaware of novel class*)
2. Kernel OR-Logistic Regression (**K-ORLR**)
3. Kernel Scoring
4. Support Instance Machine (**SIM**) (*uses max principle*)
5. Dummy classifier (**Dum**) (*trained without instance features*)

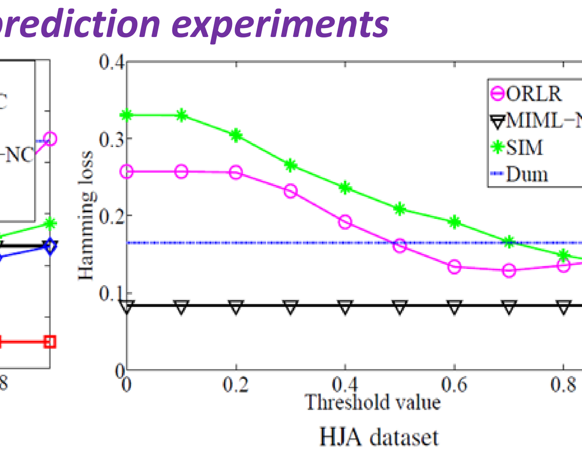
1. Novelty detection experiments



AUC (Area Under a Curve)

2. Bag label prediction experiments





- ORLR and SIM are not designed to work under the presence of novel class instances → need a threshold to filer out novel class instances
- The proposed method is parameter free

9 References

1. Briggs, F., Fern, X. Z., and Raich, R. *Rank-loss support instance machines for MIML instance annotation*. In KDD, 2012.
2. Liu, L. and Dietterich, T. G. *A conditional multinomial mixture model for superset label learning*. In NIPS, 2012.
3. Lou, Q., Raich, R., Briggs, F., and Fern, X. Z. *Novelty detection under multi-label multi-instance framework*. In MLSP, 2013.
4. Pham, A. T., Raich, R., and Fern, X. Z. *Efficient instance annotation in multi-instance learning*. In SSP, 2014.
5. Saligrama, V. and Zhao, M. *Local anomaly detection*. In AISTATS, 2012.
6. Huang, S.-J., Gao,W., and Zhou, Z.-H. *Fast multi-instance multi-label learning*. In AAAI, 2014.

Acknowledgements: This work is partially supported by the National Science Foundation grants CCF-1254218, DBI-1356792, and IIS-1055113.