Research on Partial Label Learning

Min-Ling Zhang
School of Computer Science and Engineering,
MOE Key Lab. of Computer Network & Information Integration
Southeast University, China

February 23, Zhuhai
Traditional Supervised Learning

**Input Space**
represented by a single instance (feature vector) characterizing its properties

**Output Space**
associated with a single label characterizing its semantics
Basic Assumption: Strong Supervision

Key factor for successful learning
(encoding semantics and regularities for the learning problem)

Strong supervision assumption

- Sufficient labeling
  abundant labeled training data are available

- Explicit labeling
  object labeling is unique and unambiguous
But, Supervision Is Usually Weak

Strong supervision
(sufficient & explicit)

Difficult to have!

Constrained by:
- Limited resources
- Physical environment
- Problem properties
- ......

In practice, we usually have to learn with weak supervision

Strong generalization ability
For Example…

**semi-supervised learning**

- instance
- instance

**insufficient labeling**

**multi-instance learning**

- instance
- instance
- instance

**bag-level labeling**

**multi-label learning**

- instance

**non-unique labeling**

**partial-label learning**

- instance

**ambiguous labeling**

For Example…

- insufficient labeling
- non-unique labeling
- bag-level labeling
- ambiguous labeling
Partial Label Learning
- The Framework
Partial Label

Appreciator A -----> Picasso style ✗
Appreciator B -----> Monet style ✗
Appreciator C -----> van Gogh style ✓

Widely exist in real-world applications

- Image classification [Cour et al., JMLR11] [Chen et al., TPAMI18]
- Learning from crowds [Raykar et al., JMLR10] [Yu & Zhang, MLJ17]
- Ecoinformatics [Briggs et al., KDD’12] [Tang & Zhang, AAAI’17]
- POS tagging [Zhou et al., TALLIP18]
- ......
Partial-Label Learning (PLL)

- Each object is associated with multiple candidate labels.
- Only one of the candidate labels is the unknown ground-truth label.
PLL vs. SSL

**semi-supervised learning**

- instance → label
- instance → ???

**insufficient labeling**

**partial-label learning**

- instance → label
- instance → label
- instance → label

**ambiguous labeling**

**Unlabel:**

- ground-truth label assumes the whole label space

**Partial label:**

- ground-truth label is confined within the candidate label set
PLL vs. MLL

Multi-label: all the associated labels are valid ones

Partial label: only one of the associated label is valid
PLL vs. MIL

**Multi-instance learning**

- **Instance**: bag-level labeling

**Partial-label learning**

- **Instance**: ambiguous labeling

**Multi-instance:** one label assigned to a bag of instances, with ambiguity in the input space

**Partial label:** multiple labels assigned to a single instance, with ambiguity in the output space
Partial Label Learning
- Existing Approaches
Formal Definition of PLL

Settings

\( \mathcal{X} : d\)-dimensional feature space \( \mathbb{R}^d \)

\( \mathcal{Y} : \) label space with \( q \) labels \( \{y_1, y_2, \cdots, y_q\} \)

Inputs

\( \mathcal{D} : \) training set with \( m \) examples \( \{(x_i, S_i) \mid 1 \leq i \leq m\} \)

\( x_i \in \mathcal{X} \) is a \( d\)-dimensional feature vector \( (x_{i1}, x_{i2}, \cdots, x_{id})^T \)

\( S_i \subseteq \mathcal{Y} \) is the candidate label set for \( x_i \), with its (unknown) ground-truth label \( y_i \in S_i \)

Outputs

\( h : \) multi-class predictor \( \mathcal{X} \rightarrow \mathcal{Y} \)
Key Challenge

Ambiguous labeling

ground-truth label not accessible by the learning algorithm

Common strategy: Disambiguation

- Disambiguation by ground-truth label identification
- Disambiguation by candidate label averaging
Disambiguation by Identification

Basic strategy

treating the ground-truth label as latent variable

\[ \hat{y}_i = \arg \max_{y \in S_i} F(x_i, y; \theta) \]

identified via iterative refining procedure such as EM

Latent ground-truth label

[Nguyen & Caruana, KDD’08] [Liu & Dietterich, NIPS’12] [Chen et al., CVPR’13] [Zhang et al., KDD’16] [Yu & Zhang, MLJ17] [Chen et al., TPAMI18]
Disambiguation by Identification (Cont.)

Maximum-likelihood formulation:

\[ \theta^* = \arg \max_\theta \sum_{i=1}^{m} \log \left( \sum_{y \in S_i} F(x_i, y; \theta) \right) \]

Maximum margin formulation:

\[ \theta^* = \arg \max_\theta \sum_{i=1}^{m} \log \left( \max_{y \in S_i} F(x_i, y; \theta) - \max_{y \not\in S_i} F(x_i, y; \theta) \right) \]

......

Potential weakness:

the identified label may turn out to be the false positive label
Disambiguation by Averaging

**Basic strategy**

- treating all the candidate labels in an equal manner
  
  \[ \frac{1}{|S_i|} \sum_{y \in S_i} F(x_i, y; \theta) \]

  Average output over candidate labels

- make final prediction by averaging their modeling outputs

  \[ F(x_i, y; \theta) (y \notin S_i) \]

  Output over non-candidate labels

[Hullermeier & Beringer, IDA06] [Cour et al., CVPR’09] [Cour et al., JMLR11] [Zhang & Yu, IJCAI’15] [Gong et al., TCYB18]
Disambiguation by Averaging (Cont.)

Convex formulation:

\[ \theta^* = \arg \min_\theta \sum_{i=1}^m \Psi \left( \frac{1}{|S_i|} \sum_{y \in S_i} F(x_i, y; \theta) \right) + \sum_{y \notin S_i} \Psi (-F(x_i, y; \theta)) \]

Instance-based formulation:

\[ f(x^*) = \arg \max_{y \in \mathcal{Y}} \sum_{j \in \mathcal{N}(x^*), y \in S_j} \]

......

Potential weakness: ground-truth output overwhelmed by false positive outputs
Partial Label Learning
- Recent Work I
Disambiguation-free PLL

Goal of PLL  Induce a multi-class predictor  $h : \mathcal{X} \rightarrow \mathcal{Y}$

Popular Binary Decomposition

- One-vs-Rest (#classifiers: $q$)
- One-vs-One (#classifiers: $q(q-1)/2$)

Not applicable due to the unknown ground-truth label

**PL-ECOC** (Partial-label Learning with Error-Correcting Output Codes)

- Naturally enable binary decomposition
- Disambiguation-free

Min-Ling Zhang     Feb. 23, Zhuhai
### The PL-ECOC Approach [TKDE17]

#### Illustrative procedure of ECOC

<table>
<thead>
<tr>
<th>$y_1$</th>
<th>+1</th>
<th>-1</th>
<th>+1</th>
<th>+1</th>
<th>+1</th>
<th>-1</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_2$</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>$y_3$</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>$y_4$</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>$y_5$</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>+1</td>
</tr>
</tbody>
</table>

For each multi-class example $(x_i, y_i)$

- $h_1(x_i) = +1$ if $y_i \in \{y_1, y_2, y_4\}$
- $h_1(x_i) = -1$ if $y_i \in \{y_3, y_5\}$

Identify the class with closest codeword to test instance $x^*$
The PL-ECOC Approach (Cont.)

Illustrative procedure of PL-ECOC

<table>
<thead>
<tr>
<th></th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>$y_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_1$</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>$h_2$</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$h_3$</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>$h_4$</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$h_5$</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>$h_6$</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>$h_7$</td>
<td>+1</td>
<td>+1</td>
<td>-1</td>
<td>-1</td>
<td>+1</td>
</tr>
</tbody>
</table>

For each partial-label example $(x_i, S_i)$

- $h_1(x_i) = +1$ if $S_i \subseteq \{y_1, y_2, y_4\}$
- $h_1(x_i) = -1$ if $S_i \subseteq \{y_3, y_5\}$
- ignored w.r.t. $h_1$ otherwise

make prediction in the same way as ECOC

Min-Ling Zhang
Feb. 23, Zhuhai
The PL-ECOC Approach (Cont.)

Complete Pipeline of PL-ECOC

- Coding matrix generation
  - randomly generate a q-bits column vector \( \mathbf{v} \)
  - derive the binary training set \( B_\mathbf{v} \) from the PL training set
  - \( |B_\mathbf{v}| \geq \tau \)
  - set \( \mathbf{v} \) as a column coding of the coding matrix

Repeat until reaching the specified ECOC coding length \( L \)

- Binary classifier induction
  - induce a total of \( L \) binary classifiers, one for each column coding

- Make prediction for unseen instance
  - identify the class whose codeword is closest to the classifiers’ outputs on unseen instance
Experimental Setup

Comparing Algorithms

**PL-ECOC**: Coding length $= \lceil 10 \cdot \log_2(q) \rceil$; Base learner: Libsvm

**averaging-based disambiguation**

**CLPL**: Base learner: SVM with squared hinge loss

**PL-kNN**: # nearest neighbors = 5

**identification-based disambiguation**

**PL-SVM**: Regularization parameter pool $\{10^{-3}, ..., 10^3\}$

**LSB-CMM**: # mixture components = $q$

Experimental Protocol

Ten-fold cross-validation + Pairwise $t$-test
### Controlled UCI Data Sets

<table>
<thead>
<tr>
<th>Data set</th>
<th># Examples</th>
<th># Features</th>
<th># Class Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecoli</td>
<td>336</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Dermatology</td>
<td>364</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>Vehicle</td>
<td>846</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Segment</td>
<td>2,310</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Abalone</td>
<td>4,177</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>Satimage</td>
<td>6,435</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>Usps</td>
<td>9,298</td>
<td>256</td>
<td>10</td>
</tr>
<tr>
<td>Pendigits</td>
<td>10,992</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Letter</td>
<td>20,000</td>
<td>16</td>
<td>26</td>
</tr>
</tbody>
</table>

Generating an **artificial** PL data set from an UCI data set with three controlling parameters $p, r, \epsilon$
Controlled UCI Data Sets

Generating an artificial PL data set from an UCI data set with three controlling parameters $p$, $r$, $\epsilon$

- $p$ : Proportion of examples which are partially labeled ($|S_i| \neq 1$)
- $r$ : # false positive labels in candidate label set ($|S_i| = r + 1$)
- $\epsilon$ : Co-occurring probability for one extra candidate label

Fix $r (=1, 2, 3)$, varying $p \in \{0.1, \ldots, 0.7\}$

Fix $r (=1)$, $p (=1)$, varying $\epsilon \in \{0.1, \ldots, 0.7\}$

28 configurations per UCI data set
Controlled UCI Data Sets (Cont.)

Out of 252 statistical tests (28 configurations x 9 UCI data sets)

- None of the comparing algorithms significantly outperformed PL-ECOC
- PL-ECOC outperforms PL-KNN and CLPL in 61.9% and 71.8% cases respectively
- PL-ECOC outperforms PL-SVM and LSB-CMM in 77.3% and 71.0% cases respectively

TABLE 3

Win/tie/loss counts (pairwise t-test at 0.05 significance level) on the classification performance of PL-ECOC against each comparing algorithm on the controlled UCI data sets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PL-KNN</td>
<td>Figure 1</td>
<td>0/7/0</td>
<td>1/6/0</td>
<td>7/0/0</td>
<td>3/4/0</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>5/2/0</td>
<td>7/0/0</td>
<td>37/26/0</td>
<td>38/25/0</td>
</tr>
<tr>
<td></td>
<td>Figure 2</td>
<td>0/7/0</td>
<td>3/4/0</td>
<td>7/0/0</td>
<td>2/5/0</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>5/2/0</td>
<td>40/23/0</td>
<td>41/22/0</td>
</tr>
<tr>
<td></td>
<td>Figure 3</td>
<td>0/7/0</td>
<td>2/5/0</td>
<td>7/0/0</td>
<td>4/3/0</td>
<td>7/0/0</td>
<td>1/6/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>5/2/0</td>
<td>42/21/0</td>
<td>43/20/0</td>
</tr>
<tr>
<td></td>
<td>Figure 4</td>
<td>2/5/0</td>
<td>3/4/0</td>
<td>7/0/0</td>
<td>2/5/0</td>
<td>7/0/0</td>
<td>3/4/0</td>
<td>7/0/0</td>
<td>6/1/0</td>
<td>4/3/0</td>
<td>45/23/0</td>
<td>46/22/0</td>
</tr>
<tr>
<td>CLPL</td>
<td>Figure 1</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>6/1/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>48/15/0</td>
<td>49/14/0</td>
</tr>
<tr>
<td></td>
<td>Figure 2</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>3/4/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>45/18/0</td>
<td>45/18/0</td>
</tr>
<tr>
<td></td>
<td>Figure 3</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>3/4/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>45/18/0</td>
<td>45/18/0</td>
</tr>
<tr>
<td></td>
<td>Figure 4</td>
<td>0/7/0</td>
<td>1/6/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>43/20/0</td>
<td>43/20/0</td>
</tr>
<tr>
<td>PL-SVM</td>
<td>Figure 1</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>49/14/0</td>
<td>49/14/0</td>
</tr>
<tr>
<td></td>
<td>Figure 2</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>49/14/0</td>
<td>49/14/0</td>
</tr>
<tr>
<td></td>
<td>Figure 3</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>49/14/0</td>
<td>49/14/0</td>
</tr>
<tr>
<td></td>
<td>Figure 4</td>
<td>0/7/0</td>
<td>0/7/0</td>
<td>6/1/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>48/15/0</td>
<td>48/15/0</td>
</tr>
<tr>
<td>LSB-CMM</td>
<td>Figure 1</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>1/6/0</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>43/20/0</td>
<td>43/20/0</td>
</tr>
<tr>
<td></td>
<td>Figure 2</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>1/6/0</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>43/20/0</td>
<td>43/20/0</td>
</tr>
<tr>
<td></td>
<td>Figure 3</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>4/3/0</td>
<td>7/0/0</td>
<td>0/7/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>46/17/0</td>
<td>46/17/0</td>
</tr>
<tr>
<td></td>
<td>Figure 4</td>
<td>7/0/0</td>
<td>2/5/0</td>
<td>1/6/0</td>
<td>7/0/0</td>
<td>2/5/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>7/0/0</td>
<td>47/16/0</td>
<td>47/16/0</td>
</tr>
</tbody>
</table>
## Real-World Data Sets

<table>
<thead>
<tr>
<th>Data set</th>
<th># Examples</th>
<th># Features</th>
<th># Class Labels</th>
<th>Avg. # CLs</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost</td>
<td>1,122</td>
<td>108</td>
<td>16</td>
<td>2.23</td>
<td>automatic face naming [11]</td>
</tr>
<tr>
<td>MSRCv2</td>
<td>1,758</td>
<td>48</td>
<td>23</td>
<td>3.16</td>
<td>object classification [21]</td>
</tr>
<tr>
<td>BirdSong</td>
<td>4,998</td>
<td>38</td>
<td>13</td>
<td>2.18</td>
<td>bird song classification [4]</td>
</tr>
<tr>
<td>Soccer Player</td>
<td>17,472</td>
<td>279</td>
<td>171</td>
<td>2.09</td>
<td>automatic face naming [27]</td>
</tr>
<tr>
<td>LYN 10</td>
<td>18,313</td>
<td>163</td>
<td>11</td>
<td>2.02</td>
<td>automatic face naming [17]</td>
</tr>
<tr>
<td>LYN 20</td>
<td>19,027</td>
<td>163</td>
<td>21</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td>LYN 50</td>
<td>20,308</td>
<td>163</td>
<td>54</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<td>LYN 100</td>
<td>21,390</td>
<td>163</td>
<td>101</td>
<td>1.94</td>
<td></td>
</tr>
<tr>
<td>LYN 200</td>
<td>22,991</td>
<td>163</td>
<td>219</td>
<td>1.91</td>
<td></td>
</tr>
</tbody>
</table>

**automatic face naming**

**instance:** face cropped from image/video

**candidate labels:** names extracted from associated captions/subtitles

**object classification**

**instance:** image segmentation

**candidate labels:** objects appearing within the same image

**bird song classification**

**instance:** singing syllable of the bird

**candidate labels:** bird species jointly singing within 10-seconds period

**URL:** [http://palm.seu.edu.cn/zhangml/Resources.htm#partial_data](http://palm.seu.edu.cn/zhangml/Resources.htm#partial_data)

Min-Ling Zhang

Feb. 23, Zhuhai
Real-World Data Sets (Cont.)

<table>
<thead>
<tr>
<th></th>
<th>PL-ECOC</th>
<th>PL-KNN</th>
<th>CLPL</th>
<th>PL-SVM</th>
<th>LSB-CMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost</td>
<td>0.703±0.052</td>
<td>0.424±0.041</td>
<td>0.742±0.038</td>
<td>0.729±0.040</td>
<td>0.707±0.055</td>
</tr>
<tr>
<td>MSRCv2</td>
<td>0.505±0.027</td>
<td>0.448±0.037</td>
<td>0.413±0.039</td>
<td>0.482±0.043</td>
<td>0.456±0.031</td>
</tr>
<tr>
<td>BirdSong</td>
<td>0.740±0.016</td>
<td>0.614±0.024</td>
<td>0.632±0.017</td>
<td>0.663±0.032</td>
<td>0.717±0.024</td>
</tr>
<tr>
<td>Soccer Player</td>
<td>0.537±0.020</td>
<td>0.497±0.014</td>
<td>0.368±0.010</td>
<td>0.443±0.014</td>
<td>0.525±0.015</td>
</tr>
<tr>
<td>LYN 10</td>
<td>0.694±0.010</td>
<td>0.460±0.012</td>
<td>0.605±0.013</td>
<td>0.692±0.009</td>
<td>0.703±0.010</td>
</tr>
<tr>
<td>LYN 20</td>
<td>0.697±0.012</td>
<td>0.469±0.015</td>
<td>0.585±0.010</td>
<td>0.686±0.011</td>
<td>0.702±0.011</td>
</tr>
<tr>
<td>LYN 50</td>
<td>0.694±0.008</td>
<td>0.472±0.014</td>
<td>0.540±0.012</td>
<td>0.666±0.002</td>
<td>0.679±0.007</td>
</tr>
<tr>
<td>LYN 100</td>
<td>0.680±0.012</td>
<td>0.459±0.010</td>
<td>0.507±0.011</td>
<td>0.655±0.010</td>
<td>0.673±0.010</td>
</tr>
<tr>
<td>LYN 200</td>
<td>0.662±0.010</td>
<td>0.457±0.014</td>
<td>0.462±0.009</td>
<td>0.636±0.010</td>
<td>0.648±0.007</td>
</tr>
</tbody>
</table>

- On BirdSong, LYN 50 and LYN 200, PL-ECOC is superior to all the comparing algorithms
- On Soccer Player, LYN 20, LYN 100 and MSRCv2, PL-ECOC is superior or at least comparable to all the comparing algorithms
- On Lost and LYN 10, PL-ECOC is inferior to the comparing algorithms in only two cases (CLPL on Lost; LSB-CMM on LYN 10)
Sensitivity Analysis for Coding Length

- Accuracy improves as the coding length increases
- Becomes stable as coding length approaches $[10 \cdot \log_2(q)]$

Fig. 5. Classification accuracy of PL-ECOC changes as the codeword length $L$ increases from $[\log_2(q)]$ to $[15 \cdot \log_2(q)]$ with step-size $[\log_2(q)]$. 
Partial Label Learning

- Recent Work II
Class-Imbalance Aware PLL

**Goal of PLL**  Induce a multi-class predictor $h : \mathcal{X} \rightarrow \mathcal{Y}$

**Inherent Issue:** Class-imbalance

- **Data-level:** pre-process original data
- **Algorithm-level:** design specific algorithm

**CIMAP** (Class-IMbalance Aware Partial label learning)

- **Two major advantages**
  - Class-imbalance sensitive
  - Independent of PLL algorithms
The CIMAP Approach \[\text{[KDD'18]}\]

Data-level manipulation

**Candidate Label Set Disambiguation** + **Training Set Replenishment**

\[b^S = [b_1^S, \ldots, b_q^S]^\top: \text{binary vector for candidate label set}\]

\[\forall \ 1 \leq j \leq q: \quad b_j^S = \begin{cases} 1, & \text{if } y_j \in S \\ 0, & \text{otherwise} \end{cases}\]

**Notations**

\(N(x_i): \text{index set of } k\text{-nearest neighbors of } x_i\)

\(\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_m]^\top: m \times q \text{ labeling confidence matrix}\)

\(G_{ij}: \text{disambiguated data set with label } y_j\)
The CIMAP Approach (Cont.)

Candidate label set disambiguation

1. Estimate confidence vector $\gamma_i$ of $x_i$ via weighted $k$NN aggregation

$$\gamma_i = \sum_{j \in \mathcal{N}(x_i)} \left( 1 - \frac{d(x_i, x_j)}{\sum_{k \in \mathcal{N}(x_i)} d(x_i, x_k)} \right) \cdot b^{S_j}$$

2. Derive a multi-class data set $\mathcal{M}$ by disambiguating $\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_m]^T$

$$\mathcal{M} = \bigcup_{j=1}^{q} G_j$$

where $G_j = \{(x_i, y_j) \mid 1 \leq i \leq m, j = \arg \max_{1 \leq k \leq q} \gamma_{ik}\}$

3. Adjust $\mathcal{M}$ to ensure a threshold constraint on the size of $G_j \ (1 \leq j \leq q)$
The CIMAP Approach (Cont.)

Training set replenishment

Let \( j^* = \arg \max_{1 \leq j \leq q} |G_j| \), a total of \( |G_{j^*}| - |G_j| \) PL examples will be generated for \( y_j \) to replenish the original PL training set \( \mathcal{D} \)

- **CIMAP-Ros** (random over-sampling)
  \[ \mathcal{D} = \mathcal{D} \cup \{(\hat{x}_i, S_i)\} \]

- **CIMAP-SMOTE** (synthetic over-sampling)
  \[ \mathcal{D} = \mathcal{D} \cup \{(\hat{x}_i, S_i)\} \text{ with } \hat{x}_{ia} = x_{ia} + (x_{ra} - x_{ia}) \cdot \omega_a \]

- **CIMAP-Pos** (perturbation over-sampling)
  \[ \mathcal{D} = \mathcal{D} \cup \{(\hat{x}_i, \hat{S}_i)\} \]
  \[ \text{with } \hat{S}_i = \text{sign} \left[ b^{S_i} + ((x_r - x_i)^\top \omega) \cdot (b^{S_r} - b^{S_i}) \right] \]
Experimental Setup

Performance Metrics

- Average Precision: $\text{AvgP} = \frac{1}{q} \sum_{j=1}^{q} P_j$
- Average F-measure: $\text{AvgF} = \frac{1}{q} \sum_{j=1}^{q} F_j$
- Average Recall: $\text{AvgR} = \frac{1}{q} \sum_{j=1}^{q} R_j$
- MAUC: $\text{MAUC} = \frac{2}{q(q-1)} \sum_{1 \leq j < k \leq q} \frac{A_{jk} + A_{kj}}{2}$

Comparing Algorithms

PLL algorithm vs. its class-imbalance aware version
(coupled with three variants of CIMAP)

Parameter values: $k = 5, \tau = 5$

Experimental Protocol

Ten-fold cross-validation $+$ Pairwise $t$-test
## Controlled UCI Data Sets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Examples</th>
<th>#Features</th>
<th>#Class Labels</th>
<th>IR</th>
<th>Class Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass</td>
<td>214</td>
<td>9</td>
<td>6</td>
<td>8.44</td>
<td>76/70/29/17/13/9</td>
</tr>
<tr>
<td>Ecoli</td>
<td>332</td>
<td>7</td>
<td>6</td>
<td>28.6</td>
<td>143/77/52/35/20/5</td>
</tr>
<tr>
<td>Abalone</td>
<td>4,153</td>
<td>7</td>
<td>19</td>
<td>49.21</td>
<td>689/634/568/487/391/267/259/203/126/115/1</td>
</tr>
</tbody>
</table>

- $\hat{p}$: Proportion of examples which are partially labeled ($|S_i| \neq 1$)
- $r$: # false positive labels in candidate label set ($|S_i| = r + 1$)
- $\epsilon$: Co-occurring probability for one extra candidate label

- Fix $r$ (=1, 2, 3), varying $p \in \{0.1, \ldots, 0.7\}$
- Fix $r$ (=1), $\hat{p}$ (=1), varying $\epsilon \in \{0.1, \ldots, 0.7\}$

28 configurations per UCI data set
Controlled UCI Data Sets (Cont.)

For **PL-SVM**, all three CIMAP variants achieve superior or at least comparable performance than the coupling algorithm.

For **IPAL**, CIMAP-POS achieves superior or at least comparable performance while CIMAP-ROS and CIMAP-SMOTE outperforms the coupling algorithm in 61.9% and 88.0% cases respectively.

For **PL-KNN** and **CLPL**, the three CIMAP variants are outperformed by the two coupling algorithms in only 2.2% and 1.6% cases respectively.

<table>
<thead>
<tr>
<th>Evaluation metric</th>
<th>CIMAP-ROS vs. coupling algorithm</th>
<th>CIMAP-SMOTE vs. coupling algorithm</th>
<th>CIMAP-POS vs. coupling algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PL-KNN</td>
<td>PL-SVM</td>
<td>CLPL</td>
</tr>
<tr>
<td>Avg. Precision</td>
<td>60/23/1</td>
<td>79/5/0</td>
<td>65/19/0</td>
</tr>
<tr>
<td>Avg. Recall</td>
<td>53/29/2</td>
<td>78/6/0</td>
<td>56/28/10</td>
</tr>
<tr>
<td>Avg. F-measure</td>
<td>54/25/5</td>
<td>79/5/0</td>
<td>65/19/0</td>
</tr>
<tr>
<td>MAUC</td>
<td>40/43/1</td>
<td>60/24/0</td>
<td>52/31/1</td>
</tr>
</tbody>
</table>
## Real-World Data Sets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Examples</th>
<th>#Features</th>
<th>#Class Labels</th>
<th>avg.#CLS</th>
<th>IR</th>
<th>Class Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>FG-NET</td>
<td>1,002</td>
<td>262</td>
<td>63</td>
<td>7.34</td>
<td>47.00</td>
<td>(a)</td>
</tr>
<tr>
<td>Lost</td>
<td>1,122</td>
<td>106</td>
<td>14</td>
<td>3.22</td>
<td>11.33</td>
<td>(b)</td>
</tr>
<tr>
<td>MSRCv2</td>
<td>1,755</td>
<td>48</td>
<td>22</td>
<td>3.15</td>
<td>10.63</td>
<td>(c)</td>
</tr>
<tr>
<td>BirdSong</td>
<td>4,908</td>
<td>38</td>
<td>13</td>
<td>2.18</td>
<td>40.00</td>
<td>(d)</td>
</tr>
<tr>
<td>Soccer Player</td>
<td>8,883</td>
<td>279</td>
<td>170</td>
<td>1.77</td>
<td>25.44</td>
<td>(e)</td>
</tr>
<tr>
<td>Yahoo!News</td>
<td>17,262</td>
<td>163</td>
<td>17</td>
<td>1.85</td>
<td>48.03</td>
<td>(f)</td>
</tr>
</tbody>
</table>

(b) 204/198/142/103/103/88/76/61/33/26/25/25/20/18;  
(d) 1280/810/602/501/494/345/277/190/139/126/120/82/32;  
**Real-World Data Sets (Cont.)**

Detailed Performance on the real-world PL data sets in terms of the average **MAUC**.

<table>
<thead>
<tr>
<th>PL algorithm and its CIMAP variants</th>
<th>FG-NET</th>
<th>Lost</th>
<th>MSRCv2</th>
<th>BirdSong</th>
<th>Soccer Player</th>
<th>Yahoo! News</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL-KNN</td>
<td>0.100±0.021</td>
<td>0.474±0.100</td>
<td>0.416±0.068</td>
<td>0.703±0.048</td>
<td>0.534±0.024</td>
<td>0.993±0.001</td>
</tr>
<tr>
<td>CIMAP-Ros</td>
<td>0.400±0.041</td>
<td>0.652±0.062</td>
<td>0.559±0.033</td>
<td>0.692±0.062</td>
<td>0.999±0.001</td>
<td>0.994±0.001</td>
</tr>
<tr>
<td>CIMAP-SMOTE</td>
<td>0.578±0.023</td>
<td>0.906±0.073</td>
<td>0.899±0.069</td>
<td>0.758±0.073</td>
<td>0.999±0.001</td>
<td>0.993±0.001</td>
</tr>
<tr>
<td>CIMAP-Pos</td>
<td>0.540±0.022</td>
<td>0.906±0.072</td>
<td>0.933±0.047</td>
<td>0.970±0.002</td>
<td>0.998±0.001</td>
<td>0.988±0.001</td>
</tr>
<tr>
<td>PL-SVM</td>
<td>0.197±0.042</td>
<td>0.804±0.086</td>
<td>0.422±0.085</td>
<td>0.718±0.099</td>
<td>0.816±0.031</td>
<td>0.990±0.004</td>
</tr>
<tr>
<td>CIMAP-Ros</td>
<td>0.456±0.040</td>
<td>0.981±0.020</td>
<td>0.751±0.146</td>
<td>0.834±0.070</td>
<td>0.994±0.004</td>
<td>0.992±0.002</td>
</tr>
<tr>
<td>CIMAP-SMOTE</td>
<td>0.338±0.026</td>
<td>0.983±0.010</td>
<td>0.605±0.900</td>
<td>0.867±0.065</td>
<td>0.994±0.010</td>
<td>0.987±0.001</td>
</tr>
<tr>
<td>CIMAP-Pos</td>
<td>0.378±0.045</td>
<td>0.982±0.007</td>
<td>0.556±0.066</td>
<td>0.990±0.031</td>
<td>0.987±0.006</td>
<td>0.989±0.002</td>
</tr>
<tr>
<td>CPLPL</td>
<td>0.197±0.042</td>
<td>0.804±0.086</td>
<td>0.422±0.085</td>
<td>0.718±0.099</td>
<td>0.816±0.031</td>
<td>0.990±0.004</td>
</tr>
<tr>
<td>CIMAP-Ros</td>
<td>0.456±0.040</td>
<td>0.981±0.020</td>
<td>0.751±0.146</td>
<td>0.834±0.070</td>
<td>0.994±0.004</td>
<td>0.992±0.002</td>
</tr>
<tr>
<td>CIMAP-SMOTE</td>
<td>0.338±0.026</td>
<td>0.983±0.010</td>
<td>0.605±0.900</td>
<td>0.867±0.065</td>
<td>0.994±0.010</td>
<td>0.987±0.001</td>
</tr>
<tr>
<td>CIMAP-Pos</td>
<td>0.378±0.045</td>
<td>0.982±0.007</td>
<td>0.556±0.066</td>
<td>0.990±0.031</td>
<td>0.987±0.006</td>
<td>0.989±0.002</td>
</tr>
<tr>
<td>IPAL</td>
<td>0.297±0.053</td>
<td>0.717±0.152</td>
<td>0.774±0.102</td>
<td>0.922±0.081</td>
<td>0.875±0.030</td>
<td>0.999±0.001</td>
</tr>
<tr>
<td>CIMAP-Ros</td>
<td>0.392±0.037</td>
<td>0.766±0.057</td>
<td>0.714±0.046</td>
<td>0.826±0.014</td>
<td>0.880±0.002</td>
<td>0.883±0.003</td>
</tr>
<tr>
<td>CIMAP-SMOTE</td>
<td>0.804±0.027</td>
<td>0.998±0.001</td>
<td>0.999±0.001</td>
<td>0.999±0.001</td>
<td>0.994±0.001</td>
<td>0.999±0.001</td>
</tr>
<tr>
<td>CIMAP-Pos</td>
<td>0.758±0.045</td>
<td>0.997±0.001</td>
<td>0.999±0.001</td>
<td>0.999±0.001</td>
<td>0.994±0.001</td>
<td>0.999±0.001</td>
</tr>
</tbody>
</table>
Partial Label Learning

- Related Resources
Introductory Papers


...
Data Sets

- Partial label learning (PLL)
  - [http://palm.seu.edu.cn/zhangml/Resources.htm#partial_data](http://palm.seu.edu.cn/zhangml/Resources.htm#partial_data)
  - [http://www.timotheecour.com/tv_data/tv_data.html](http://www.timotheecour.com/tv_data/tv_data.html)
  - [http://web.engr.oregonstate.edu/~briggsf/](http://web.engr.oregonstate.edu/~briggsf/)

- Multi-label learning (MLL)
  - [http://mulan.sourceforge.net/datasets.html](http://mulan.sourceforge.net/datasets.html)
  - [http://meka.sourceforge.net/#datasets](http://meka.sourceforge.net/#datasets)

- Multi-instance multi-label learning (MIML)
  - [http://lamda.nju.edu.cn/data_MIMLimage.ashx](http://lamda.nju.edu.cn/data_MIMLimage.ashx)
  - [http://lamda.nju.edu.cn/data_MIText.ashx](http://lamda.nju.edu.cn/data_MIText.ashx)
  - [http://lamda.nju.edu.cn/data_MIMLprotein.ashx](http://lamda.nju.edu.cn/data_MIMLprotein.ashx)
Codes

- Partial label learning (PLL)
  - http://web.engr.oregonstate.edu/~liuli/files/LSB-CMM_1.0.tar.gz
  - http://cse.seu.edu.cn/PersonalPage/zhangml/Resources.htm#codes

- Multi-label learning (MLL)
  - http://meka.sourceforge.net/
  - http://cse.seu.edu.cn/people/zhangml/Resources.htm#codes_mll

- Multi-instance multi-label learning (MIML)
  - http://lamda.nju.edu.cn/code_MIML.ashx

- Label distribution learning (LDL)
  - http://cse.seu.edu.cn/PersonalPage/xgeng/LDL
Thanks!