Unleash the Power of Label Space: Label Enhancement for Label Distribution Learning

Xin Geng

PAtern Learning and Mining (PALM) Lab
http://palm.seu.edu.cn

School of Computer Science and Engineering
Southeast University

KSEM 2019, Athens
Outline

• Introduction

• Label Enhancement
  • Formulation
  • Algorithms
  • Experiments

• Conclusion
Ambiguity in Machine Learning

Not 1-to-1 mapping

Instance

Multi-instance Learning (Many-to-one)

Multi-label Learning (One-to-many)

Multi-instance Multi-label Learning (Many-to-many)

Label Ambiguity

learning

Label
Label Ambiguity

**Label distribution**: covers all possible labels and explicitly gives the importance of each label to the instance.

- **Single-label learning**
- **Multi-label learning**
- **Label distribution learning (LDL)**

Less Ambiguity    |    Label Ambiguity    |    More Ambiguity
Definition of Label Distribution [Geng, TKDE’16]

A real number $d_{\mathbf{x}}^y$ is assigned to the label $y$ for the instance $\mathbf{x}$.

WLOG $d_{\mathbf{x}}^y \in [0, 1]$,

Complete label set $\sum_y d_{\mathbf{x}}^y = 1$

Description Degree

Label Distribution

(a) Single-label

(b) Multi-label

(c) General case
Example 1: Emotion distribution

<table>
<thead>
<tr>
<th>emotion</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>happy (HAP)</td>
<td>1.48</td>
</tr>
<tr>
<td>sad (SAD)</td>
<td>2.48</td>
</tr>
<tr>
<td>surprise (SUR)</td>
<td>2.19</td>
</tr>
<tr>
<td>angry (ANG)</td>
<td>4.48</td>
</tr>
<tr>
<td>disgust (DIS)</td>
<td>3.42</td>
</tr>
<tr>
<td>fear (FER)</td>
<td>2.16</td>
</tr>
</tbody>
</table>
Example 2: Movie rating distribution

<table>
<thead>
<tr>
<th>Title</th>
<th>Twilight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Rating</td>
<td>5.2/10</td>
</tr>
<tr>
<td>Budget</td>
<td>$37 Million</td>
</tr>
</tbody>
</table>

More real-world label distribution data sets: http://palm.seu.edu.cn/xgeng/LDL/index.htm
Universality of Label Distribution

• The relevance or irrelevance of a label to an instance is essentially relative.
  • The separation between the relevant and irrelevant labels is relative.
Universality of Label Distribution

• The relevance or irrelevance of a label to an instance is essentially relative.
  • When multiple labels are relevant to the same instance, their importance is not likely to be exactly same
Universality of Label Distribution

• The relevance or irrelevance of a label to an instance is essentially relative.
  • The “irrelevance” of each irrelevant label may be very different.
Universality of Label Distribution

• Traditional class label
  0: Irrelevant label
  1: Relevant label

• But

  1. The borderline between 1 and 0 is often vague
  2. 1 and 1 are often different
  3. 0 and 0 are often different

• Logical labels simplify the real world
• Label distribution is closer to the ground-truth!
The Power of Label Space

Instance Space

Logical Label Space

0: Irrelevant label
1: Relevant label

Many Analytic Tools

Limited Expressiveness
Limited Analytic Tools
The Power of Label Space

Logical Label Space

Continuous Label Space
(Label Distribution)
The Power of Label Space

Label distribution space offers more possibilities for ...
Applications with the Power of Label Space

- **Age Estimation**
  - [Gao, et al., IJCAI’18]; [Hou, et al., AAAI’18]; [He, et al., TIP’17]; [Geng, Yin, and Zhou, TPAMI’13]

- **Head Pose Estimation**
  - [Geng and Xia, CVPR’14]

- **Text Emotion Estimation**
  - [Zhou, et al., EMNLP’16]

- **Facial Landmark Detection**
  - [Su and Geng, AAAI’19]

- **Multilabel Ranking**
  - [Geng and Luo, CVPR’14]

- **Video Parsing**
  - [Geng and Ling, AAAI’17]

- **Prediction of Crowd Opinion on Movies**
  - [Geng and Hou, IJCAI’15]

- **Indoor Crowd Counting**
  - [Ling and Geng, TIP’19]

- **Expression Recognition**
  - [Zhou, Xue and Geng, ACMMM’15]

- **Beauty Sense**
  - [Ren and Geng, IJCAI’17]
Practical Restrictions

- Directly obtaining description degrees of all labels is difficult:
  - High cost
  - Difficult to quantify
- Most existing data sets simplify the real world: a bipartition of the label set into relevant and irrelevant labels
  - 1: relevant label
  - 0: irrelevant label

We need a way to recover the label distributions from the logical labels in the training set.

Label Enhancement (LE)

{0, 1, 0, 1, 0}
Label Distribution & Label Enhancement

Data with Label Distributions

\{0, 1, 0, 1, 0\}

Data with Logical Labels

Label Distribution Learning

Label Enhancement
Outline

- Introduction
- Label Enhancement
  - Formulation
  - Algorithms
  - Experiments
- Conclusion
Problem Formulation

The logical label vector of $x_i$ is denoted by $L_i = \left(l_{x_i}^{y_1}, l_{x_i}^{y_2}, \ldots, l_{x_i}^{y_c}\right)^T$, where $l_{x_i}^{y_j} \in \{0,1\}$ represents whether $y_j$ describes $x_i$, $c$ is the number of labels. Then, $L_i \in \{0,1\}^c$.

The label distribution of $x_i$ is denoted by $D_i = \left(d_{x_i}^{y_1}, d_{x_i}^{y_2}, \ldots, d_{x_i}^{y_c}\right)^T$, where $d_{x_i}^{y_j} \in [0,1]$ represents the description degree of $y_j$ to $x_i$. Then, $D_i \in [0,1]^c$.

Label Enhancement can be defined as follows.

Given a training set $S = \{(x_i, L_i)|1 \leq i \leq n\}$, label enhancement is to recover the label distribution $D_i$ of $x_i$ from the logical label vector $L_i$, and thus transform $S$ into an LDL training set $E = \{(x_i, D_i)|1 \leq i \leq n\}$.
What Added?

\{0, 1, 0, 1, 0\} + \text{Correlation among labels (y space)} = \text{Topological relationship among feature vectors (x space)}
Outline

• Introduction
• Label Enhancement
  • Formulation
  • Algorithms
  • Experiments
• Conclusion
Label Enhancement Algorithms

• **Fuzzy Label Enhancement**
  - LE based on fuzzy clustering (FCM)
    [Gayar et al., ANNPR'06]
  - LE based on kernel method (KM)
    [Jiang et al., NCA’06]

• **Graph-based Label Enhancement**
  - LE based on label propagation (LP)
    [Li et al., ICDM’15]
  - LE based on manifold learning (ML)
    [Hou et al., AAAI’16]
  - Graph Laplacian Label Enhancement (GLLE)
    [Xu and Geng, IJCAI’18]
LE based on fuzzy clustering (FCM)

Fuzzy C-means clustering (The membership of the instance to the cluster)

The memberships of the instances belonging to the same class are added up to form the cluster-class connection matrix.

By fuzzy composition operation, the memberships of instances to clusters are transformed into the memberships of instances to class labels using the connection matrix.

[Gayar et al., ANNPR'06]
LE based on fuzzy clustering (FCM) [Gayar et al., ANNPR'06]

• Step 1: Fuzzy C-Means clustering (FCM)
  1. Given the cluster number $p$, initialize the $n \times p$ cluster membership matrix $M$ ($m_{ik}$ denotes the membership of $x_i$ to the $k$-th cluster)
  2. Calculate the cluster prototype
     $$\mu_k = \frac{\sum_{i=1}^{n} (m_{ik})^\beta x_i}{\sum_{i=1}^{n} (m_{ik})^\beta}$$
  3. Update the cluster membership matrix $M$
     $$m_{ik} = \frac{1}{\sum_{j=1}^{p} \left( \frac{Dist(x_i, \mu_k)}{Dist(x_i, \mu_j)} \right)^{\beta-1}}$$
  4. Repeat 2 and 3 until convergence

Each row of $M$, $m_i$, represents the membership of the instance $x_i$ to each cluster
LE based on fuzzy clustering (FCM) [Gayar et al., ANNPR'06]

• **Step 2**: Calculate the cluster-class connection matrix
  1. Initialize $c \times p$ zero matrix $A$
  2. Update each row $A_j$ with
     $$A_j = A_j + m_i, \quad \text{if } l^y_{x_i} = 1$$
  3. Normalized each column of $A$
  4. Normalized each row of $A$

  $a_{jk}$ denotes the connection between class $j$ and cluster $k$

• **Step 3**: Calculate the label distribution of $x_i$
  1. $D_i = A \circ m_i$ (fuzzy composition)
     $$D^j_i = \max_k (a_{jk} \times m_{ik})$$
  2. Normalize $D_i$
LE based on kernel method (KM)

Introduce nonlinearity via kernel method
LE based on kernel method (KM) [Jiang et al., NCA’06]

- Step 1: For each label $y_j$, suppose $C^{y_j}$ contains all the instances labeled by $y_j$, the size of $C^{y_j}$ is $n_j$, then, the center of $C^{y_j}$ is

$$\Psi^{y_j} = \frac{1}{n_j} \sum_{x_i \in C^{y_j}} \phi(x_i)$$

where $\phi(x_i)$ is a nonlinear function determined by the kernel function

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

- Step 2: Calculate the class radius

$$r_j = \max_{x_i \in C^{y_j}} \|\Psi^{y_j} - \phi(x_i)\|,$$

$r_j^2$ can be calculated via inner product of $\phi(x_i)$

- Step 3: Calculate the distance between instance $x_i$ and class center

$$d_{ij}^2 = \|\phi(x_i) - \Psi^{y_j}\|^2$$

$d_{ij}^2$ can be calculated via inner product of $\phi(x_i)$
LE based on kernel method (KM) [Jiang et al., NCA’06]

- **Step 4**: calculate the membership of instance $x_i$ to label $y_j$

$$m_{x_i}^{y_j} = \begin{cases} 
1 - \sqrt{\frac{d_{ij}^2}{(r_j^2 + \delta)}} & \text{if } l_{x_i}^{y_j} = 1 \\
0 & \text{if } l_{x_i}^{y_j} = 0 
\end{cases}$$

Cannot change the membership of irrelevant labels

- **Step 5**: Normalize $m_{x_i} = [m_{x_i}^{y_1}, m_{x_i}^{y_2}, ..., m_{x_i}^{y_c}]$
LE based on label propagation (LP)

[Li, Zhang and Geng, ICDM’15]

Label Propagation in training set
LE based on label propagation (LP) | [Li, Zhang and Geng, ICDM’15]

\[ V = \{ x_i \mid 1 \leq i \leq m \} \]

\[ G = (V, E) \]

\[\forall_{i,j=1}^{m} \quad w_{ij} = \begin{cases} \exp \left( -\frac{\| x_i - x_j \|^2}{2\sigma^2} \right), & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}\]

\[ P = D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad D = \text{diag}[d_1, d_2, \ldots, d_m] \quad d_i = \sum_{j=1}^{m} w_{ij}\]

\[ F^{(0)} = \Phi \]

\[ F^{(t)} = \alpha P F^{(t-1)} + (1 - \alpha) \Phi \]

\[ F^* = (1 - \alpha)(I - \alpha P)^{-1} \Phi \]

\[ \mu_{x_i}^{y_l} = \frac{f_{il}^*}{\sum_{k=0}^{q} f_{ik}^*} \]
LE based on manifold learning (ML)

[Hou, Geng and Zhang, AAAI’16]

- **Feature space**: continuous Euclidean space
- **Label space**: discrete logical space

The manifold structure is transferred from the feature space to the label space.
LE based on manifold learning (ML)

[Hou, Geng and Zhang, AAAI’16]

- **Manifold learning in feature space** [Roweis & Saul, Science, 2000]

  \[
  \arg \min \sum_{i=1}^{n} \| x_i - \sum_{j \neq i} W_{ij} x_j \|^2 \\
  \text{s.t. } 1^T W_i = 1
  \]

  If \( x_j \) is not the neighbor of \( x_i \), then \( W_{ij} = 0 \)

- **Manifold learning in label space**

  \[
  \arg \min \sum_{i=1}^{n} \| \mu_i - \sum_{j \neq i} W_{ij} \mu_j \|^2 \\
  \text{s.t. } \forall 1 \leq i \leq n, 1 \leq l \leq q, \quad y_i^l \mu_i^l \geq \lambda, \lambda > 0
  \]

  \( y_i \) is defined as:

  \[
  y_i = \begin{cases} 
  +1, & \text{if the } l\text{-th label is relevant} \\
  -1, & \text{if the } l\text{-th label is irrelevant}
  \end{cases}
  \]

  Control the sign and scale
Graph Laplacian Label Enhancement (GLLE)

[Xu and Geng, IJCAI’18]

- **Model**

  $D_i = W^T \phi(x_i) + b = \hat{W} \phi_i$

  

- **Goal**

  Determining the best parameter $\hat{W}^*$

- **Target function**

  \[
  \min_{\hat{W}} L(\hat{W}) + \lambda \Omega(\hat{W})
  \]

  Feature space constraint

  Logical label loss
Graph Laplacian Label Enhancement (GLLE)  

[Xu and Geng, IJCAI’18]

- The first part of the target function

\[ L(\mathbf{W}) = \sum_{i=1}^{n} \| \mathbf{W} \phi_i - L_i \|^2 \]  

Least squares (LS)

- The second part of the target function

\[ \Omega(\mathbf{W}) = \sum_{i,j} a_{ij} \| D_i - D_j \|^2 = tr(DGD^T) \]

Smoothness assumption

\[ a_{ij} = \begin{cases} \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) & \text{if } x_j \in N(i) \\ 0 & \text{otherwise} \end{cases} \]

Correlation between \( x_i \) and \( x_j \)

\[ G = \hat{A} - A, \hat{a}_{ij} = \sum_{j=1}^{n} a_{ij} \]

Graph Laplacian
Outline

- Introduction
- Label Enhancement
  - Formulation
  - Algorithms
  - Experiments
- Conclusion
<table>
<thead>
<tr>
<th>No.</th>
<th>Dataset</th>
<th>#Examples</th>
<th>#Features</th>
<th>#Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Artificial</td>
<td>2601</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>SJAFFE</td>
<td>213</td>
<td>243</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Natural Scene</td>
<td>2,000</td>
<td>294</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Yeast-spoem</td>
<td>2,465</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Yeast-spo5</td>
<td>2,465</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Yeast-dtt</td>
<td>2,465</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Yeast-cold</td>
<td>2,465</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Yeast-heat</td>
<td>2,465</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Yeast-spo</td>
<td>2,465</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>Yeast-diau</td>
<td>2,465</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>Yeast-elu</td>
<td>2,465</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>Yeast-cdc</td>
<td>2,465</td>
<td>24</td>
<td>15</td>
</tr>
<tr>
<td>13</td>
<td>Yeast-alpha</td>
<td>2,465</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>14</td>
<td>SBU_3DFE</td>
<td>2,500</td>
<td>243</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>Movie</td>
<td>7,755</td>
<td>1,869</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1: Statistics of the 15 Datasets Used in the Experiments
Artificial dataset

• The label distribution $D = [d_{x,1}^{y_1}, d_{x,2}^{y_2}, d_{x,3}^{y_3}]$ of $x = [x_1, x_2, x_3]$ is generated in the following way

$$t_i = ax_i + bx_i^2 + cx_i^3 + d, i = 1, ..., 3,$$

$$\psi_1 = (\omega_1^T t)^2, \quad \psi_2 = (\omega_2^T t + \lambda_1 \psi_1)^2, \quad \psi_3 = (\omega_3^T t + \lambda_2 \psi_2)^2,$$

$$d_{x,i}^{y_i} = \frac{\psi_i}{\psi_1 + \psi_2 + \psi_3}, i = 1, ..., 3,$$

where $a = 1, \ b = 0.5, \ c = 0.2, \ d = 1, \ \omega_1^T = [4,2,1], \ \omega_2^T = [1,2,4], \ \omega_3^T = [1,4,2], \ \lambda_1 = \lambda_2 = 0.01$.

• Sampling

The first two components of $x$, $x_1, x_2$ are sampled from a grid of the interval 0.04 within the range $[-1,1]^2$, and there are in total $51 \times 51 = 2601$ instances. The third component $x_3$ is generated by

$$x_3 = \sin((x_1 + x_2) \times \pi)$$
Datasets

- Label distribution binarization

1. Initialize the relevant label set \( y^+ = \emptyset \), the irrelevant label set \( y^- = y \), \( L = 0 \);
2. Select \( y_j \in y^- \) with the highest description degree, then \( y^+ = y^+ + y_j \), \( y^- = y^- - y_j \), \( l_x^{y_j} = 1 \);
3. \( t = \sum_{y_j \in y^+} d_x^{y_j} \), if \( t < T \), go back to step 2, else end.

\[ T = 0.5 \]
Recovery Performance

Ground-truth
Label distributions

Binarization

{0, 1, 0, 1, 0}
Logical labels

Label Enhancement

LDL evaluation measure

Recovered
Label distributions
Recovery Performance

(a) Ground-Truth
(b) GLLE
(c) LP
(d) ML
(e) FCM
(f) KM
Recovery Performance

<table>
<thead>
<tr>
<th>Datasets</th>
<th>FCM</th>
<th>KM</th>
<th>LP</th>
<th>ML</th>
<th>GLLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial</td>
<td>0.188(3)</td>
<td>0.260(5)</td>
<td>0.130(2)</td>
<td>0.227(4)</td>
<td>0.108(1)</td>
</tr>
<tr>
<td>SJAFFE</td>
<td>0.132(3)</td>
<td>0.214(5)</td>
<td>0.107(2)</td>
<td>0.190(4)</td>
<td>0.100(1)</td>
</tr>
<tr>
<td>Natural Scene</td>
<td>0.368(5)</td>
<td>0.306(4)</td>
<td><strong>0.275(1)</strong></td>
<td>0.295(2)</td>
<td>0.296(3)</td>
</tr>
<tr>
<td>Yeast-spoem</td>
<td>0.233(3)</td>
<td>0.408(5)</td>
<td>0.163(2)</td>
<td>0.400(4)</td>
<td><strong>0.108(1)</strong></td>
</tr>
<tr>
<td>Yeast-spo5</td>
<td>0.162(3)</td>
<td>0.277(5)</td>
<td>0.114(2)</td>
<td>0.273(4)</td>
<td><strong>0.092(1)</strong></td>
</tr>
<tr>
<td>Yeast-dtt</td>
<td>0.097(2)</td>
<td>0.257(5)</td>
<td>0.128(3)</td>
<td>0.244(4)</td>
<td><strong>0.065(1)</strong></td>
</tr>
<tr>
<td>Yeast-cold</td>
<td>0.141(3)</td>
<td>0.252(5)</td>
<td>0.137(2)</td>
<td>0.242(4)</td>
<td><strong>0.093(1)</strong></td>
</tr>
<tr>
<td>Yeast-heat</td>
<td>0.169(4)</td>
<td>0.175(5)</td>
<td>0.086(2)</td>
<td>0.165(3)</td>
<td><strong>0.056(1)</strong></td>
</tr>
<tr>
<td>Yeast-spo</td>
<td>0.130(3)</td>
<td>0.175(5)</td>
<td>0.090(2)</td>
<td>0.171(4)</td>
<td><strong>0.067(1)</strong></td>
</tr>
<tr>
<td>Yeast-diau</td>
<td>0.124(3)</td>
<td>0.152(5)</td>
<td>0.099(2)</td>
<td>0.148(4)</td>
<td><strong>0.084(1)</strong></td>
</tr>
<tr>
<td>Yeast-elu</td>
<td>0.052(3)</td>
<td>0.078(5)</td>
<td>0.044(2)</td>
<td>0.072(4)</td>
<td><strong>0.030(1)</strong></td>
</tr>
<tr>
<td>Yeast-cdc</td>
<td>0.051(3)</td>
<td>0.076(5)</td>
<td>0.042(2)</td>
<td>0.071(4)</td>
<td><strong>0.038(1)</strong></td>
</tr>
<tr>
<td>Yeast-alpha</td>
<td>0.044(3)</td>
<td>0.063(5)</td>
<td>0.040(2)</td>
<td>0.057(4)</td>
<td><strong>0.033(1)</strong></td>
</tr>
<tr>
<td>SBU_3DFE</td>
<td>0.135(2)</td>
<td>0.238(5)</td>
<td><strong>0.123(1)</strong></td>
<td>0.233(4)</td>
<td>0.141(3)</td>
</tr>
<tr>
<td>Movie</td>
<td>0.230(4)</td>
<td>0.234(5)</td>
<td>0.161(2)</td>
<td>0.164(3)</td>
<td><strong>0.160(1)</strong></td>
</tr>
</tbody>
</table>

| Avg. Rank     | 3.13 | 4.93 | 1.93 | 3.73 | 1.27 |

Table 2: Recovery Results (value(rank)) Measured by Cheb
## Recovery Performance

<table>
<thead>
<tr>
<th>Datasets</th>
<th>FCM</th>
<th>KM</th>
<th>LP</th>
<th>ML</th>
<th>GLLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial</td>
<td>0.933(3)</td>
<td>0.918(5)</td>
<td>0.974(2)</td>
<td>0.925(4)</td>
<td>0.980(1)</td>
</tr>
<tr>
<td>SJAFFE</td>
<td>0.906(3)</td>
<td>0.827(5)</td>
<td>0.941(2)</td>
<td>0.857(4)</td>
<td>0.946(1)</td>
</tr>
<tr>
<td>Natural Scene</td>
<td>0.593(5)</td>
<td>0.748(4)</td>
<td><strong>0.860(1)</strong></td>
<td>0.818(2)</td>
<td>0.769(3)</td>
</tr>
<tr>
<td>Yeast-spoem</td>
<td>0.878(3)</td>
<td>0.812(5)</td>
<td>0.950(2)</td>
<td>0.815(4)</td>
<td><strong>0.968(1)</strong></td>
</tr>
<tr>
<td>Yeast-spo5</td>
<td>0.922(3)</td>
<td>0.882(5)</td>
<td>0.969(2)</td>
<td>0.884(4)</td>
<td><strong>0.974(1)</strong></td>
</tr>
<tr>
<td>Yeast-dtt</td>
<td>0.959(2)</td>
<td>0.759(5)</td>
<td>0.921(3)</td>
<td>0.763(4)</td>
<td><strong>0.983(1)</strong></td>
</tr>
<tr>
<td>Yeast-cold</td>
<td>0.922(3)</td>
<td>0.779(5)</td>
<td>0.925(2)</td>
<td>0.784(4)</td>
<td><strong>0.969(1)</strong></td>
</tr>
<tr>
<td>Yeast-heat</td>
<td>0.883(3)</td>
<td>0.779(5)</td>
<td>0.932(2)</td>
<td>0.783(4)</td>
<td><strong>0.980(1)</strong></td>
</tr>
<tr>
<td>Yeast-spo</td>
<td>0.909(3)</td>
<td>0.800(5)</td>
<td>0.939(2)</td>
<td>0.803(4)</td>
<td><strong>0.968(1)</strong></td>
</tr>
<tr>
<td>Yeast-diau</td>
<td>0.882(3)</td>
<td>0.799(5)</td>
<td>0.915(2)</td>
<td>0.803(4)</td>
<td><strong>0.939(1)</strong></td>
</tr>
<tr>
<td>Yeast-elu</td>
<td>0.950(2)</td>
<td>0.758(5)</td>
<td>0.918(3)</td>
<td>0.763(4)</td>
<td><strong>0.978(1)</strong></td>
</tr>
<tr>
<td>Yeast-cdc</td>
<td>0.929(2)</td>
<td>0.754(5)</td>
<td>0.916(3)</td>
<td>0.759(4)</td>
<td><strong>0.959(1)</strong></td>
</tr>
<tr>
<td>Yeast-alpha</td>
<td>0.922(2)</td>
<td>0.751(5)</td>
<td>0.911(3)</td>
<td>0.756(4)</td>
<td><strong>0.973(1)</strong></td>
</tr>
<tr>
<td>SBU_3DFE</td>
<td>0.912(2)</td>
<td>0.812(5)</td>
<td><strong>0.922(1)</strong></td>
<td>0.815(4)</td>
<td>0.900(3)</td>
</tr>
<tr>
<td>Movie</td>
<td>0.773(5)</td>
<td>0.880(4)</td>
<td><strong>0.929(1)</strong></td>
<td>0.919(2)</td>
<td>0.900(3)</td>
</tr>
<tr>
<td><strong>Avg. Rank</strong></td>
<td>2.93</td>
<td>4.87</td>
<td>2.07</td>
<td>3.73</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 3: Recovery Results (value(rank)) Measured by Cosine ↑
LDL Predictive Performance

Ground-truth Label distributions

SA-BFGS

Predicted Label distributions

Binarization

LDL evaluation measure

{0, 1, 0, 1, 0} Logical labels

LE

Recovered Label distributions

SA-BFGS

Predicted Label distributions
**LDL Predictive Performance**

![Graph 1: LDL Predictive Performance](image1)

**Figure 3:** Comparison of the LDL after the LE pre-process against the direct LDL measured by Cheb ↓.

![Graph 2: LDL Predictive Performance](image2)

**Figure 4:** Comparison of the LDL after the LE pre-process against the direct LDL measured by Cosine ↑.
### LDL Predictive Performance

#### Table 4: The Average Ranks of Five Algorithms on Six Measures

<table>
<thead>
<tr>
<th>Criterion</th>
<th>FCM</th>
<th>KM</th>
<th>LP</th>
<th>ML</th>
<th>GLLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheb</td>
<td>4.40</td>
<td>4.20</td>
<td>2.00</td>
<td>3.13</td>
<td>1.27</td>
</tr>
<tr>
<td>Clark</td>
<td>4.33</td>
<td>4.07</td>
<td>2.27</td>
<td>3.07</td>
<td>1.27</td>
</tr>
<tr>
<td>Canber</td>
<td>4.20</td>
<td>4.13</td>
<td>2.27</td>
<td>3.13</td>
<td>1.27</td>
</tr>
<tr>
<td>KL</td>
<td>4.37</td>
<td>4.30</td>
<td>2.00</td>
<td>3.13</td>
<td>1.20</td>
</tr>
<tr>
<td>Cosine</td>
<td>4.53</td>
<td>4.27</td>
<td>1.93</td>
<td>3.07</td>
<td>1.20</td>
</tr>
<tr>
<td>Intersec</td>
<td>4.40</td>
<td>4.20</td>
<td>1.93</td>
<td>3.13</td>
<td>1.33</td>
</tr>
</tbody>
</table>
Outline

• Introduction

• Label Enhancement
  • Formulation
  • Algorithms
  • Experiments

• Conclusion
Conclusion

• **Label distribution learning**
  - is more general a framework than single-label and multi-label learning
  - deals with different importance of labels
  - is generally suitable for many practical problems
  - lack of label distribution annotation limits the application of LDL

• **Label enhancement**
  - recovers label distributions from logical labels
  - leverages the topological information of the feature space and/or the correlation among the labels
  - is the precondition for the universality of LDL
Interested in LDL & LE?

All the papers, codes and datasets are available at:
http://palm.seu.edu.cn/xgeng/SDL/index.htm
Acknowledgements

• Collaborators

Prof. Zhi-Hua Zhou  Prof. Kate Smith-Miles  Prof. Min-Ling Zhang  Prof. Deyu Zhou  Dr. Hui Xue

• Students

Chao Yin  Yu-Kun Li  Ying Zhou
Peng Hou  Yu Xia  Longrun Luo
Rongzi Ji  Chao Xing  Xu Yang
Miaogen Ling  Quan Zhao  Xuan Zhang
Ning Xu  Ruifeng Shao  Jing Wang
THANK YOU!

http://palm.seu.edu.cn