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

Xin Geng

PAttern 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

Label Ambiguity

Instance → learning → Label

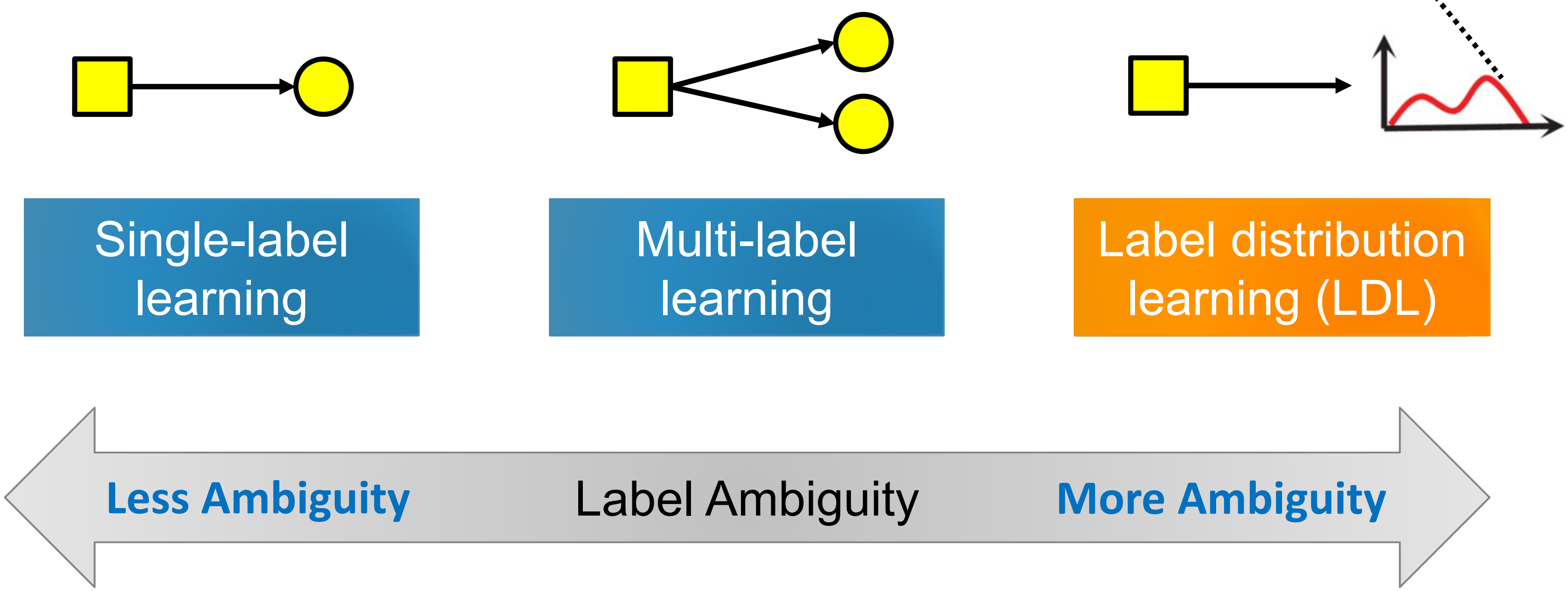
Multi-instance Leaning
(Many-to-one)

Multi-label Leaning
(One-to-many)

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

Label Ambiguity

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



Definition of Label Distribution [Geng, TKDE'16]

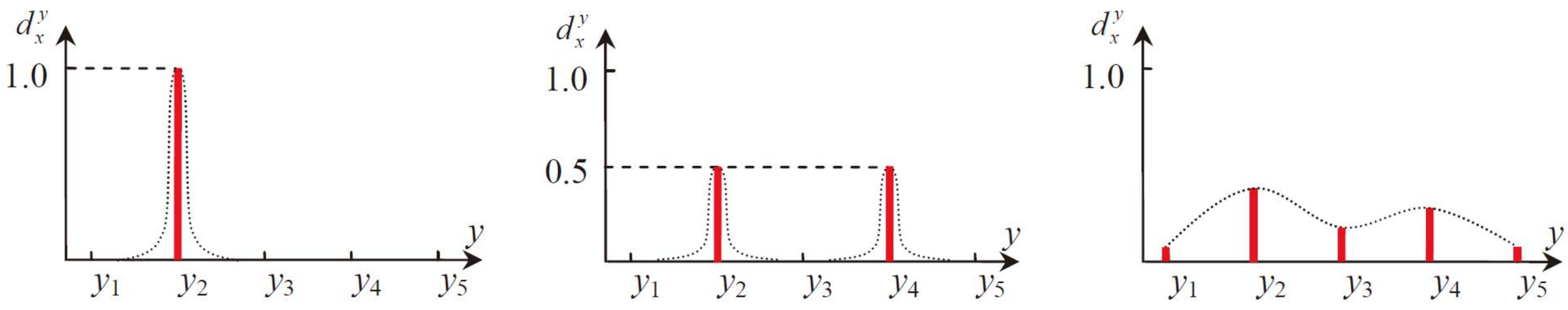
Description Degree

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

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

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

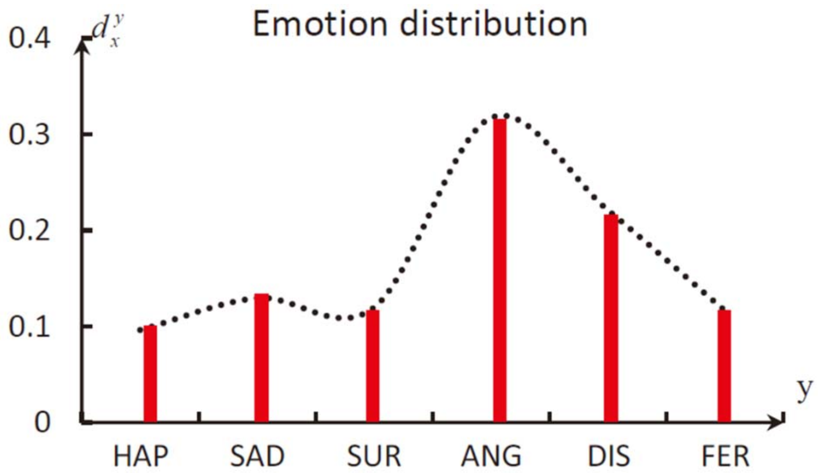
} Label Distribution



(a) Single-label (b) Multi-label (c) General case

Example 1： Emotion distribution

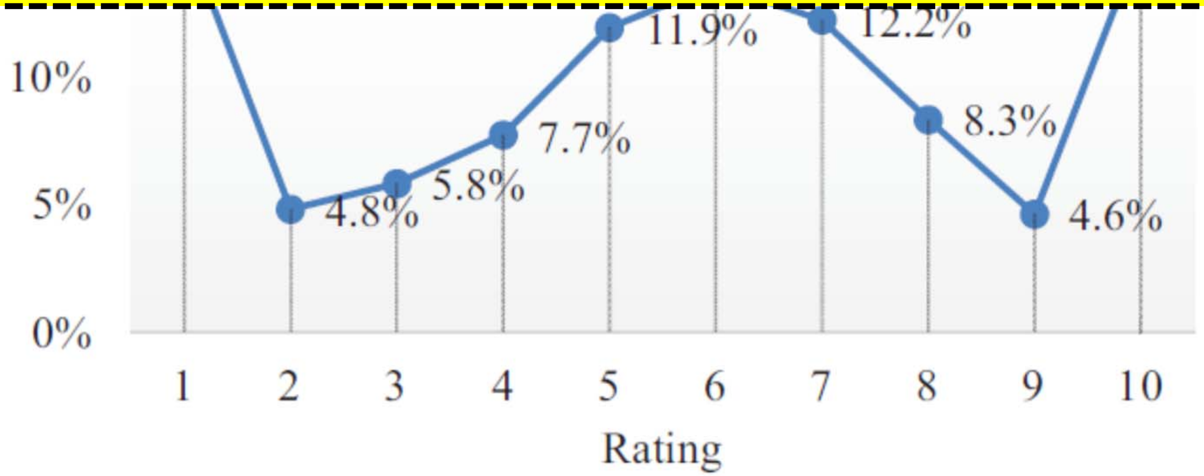
emotion	score
happy （HAP）	1.48
sad （SAD）	2.48
surprise （SUR）	2.19
angry （ANG）	4.48
disgust （DIS）	3.42
fear （FER）	2.16



Example 2: Movie rating distribution

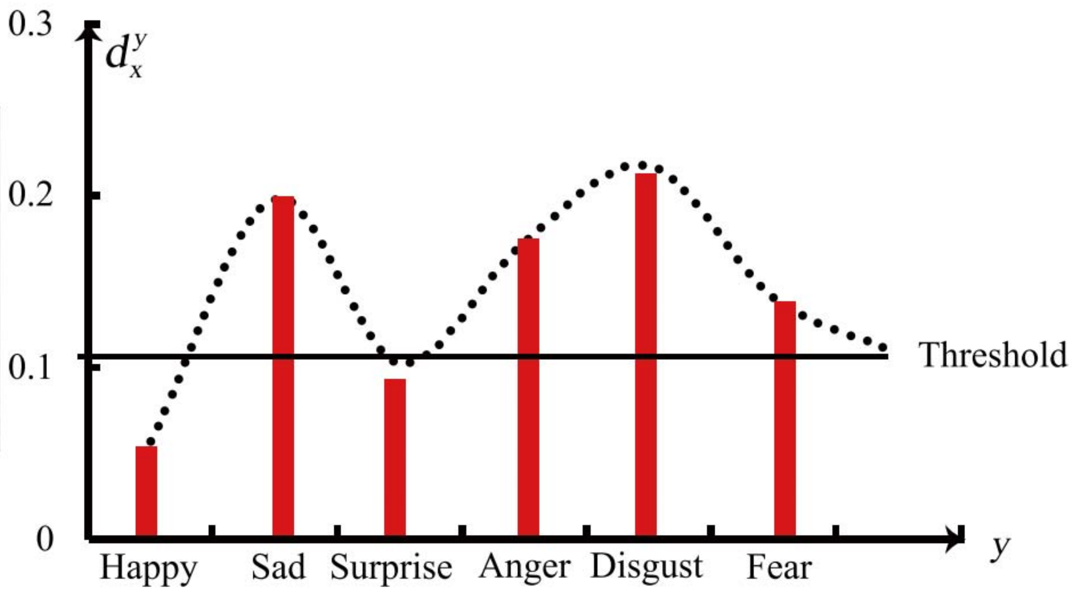
Title	Twilight
Average Rating	5.2/10
Budget	$ 37 Million

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

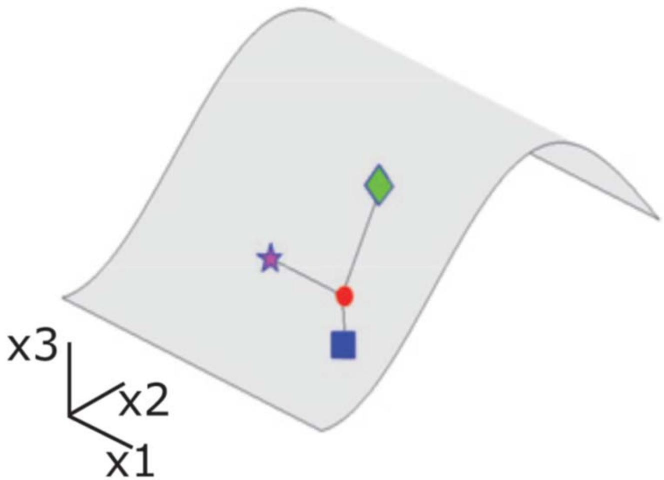
  Logical Labels

- But

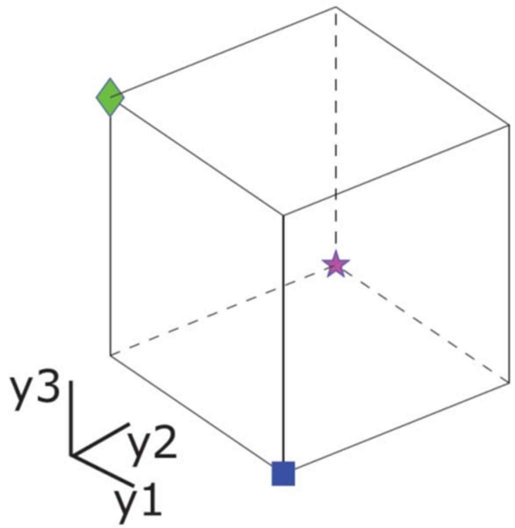
1. The borderline between 1 and 0 is often vague
2. 1 and 1 are often different
3. 0 and 0 are often different

- **Logical labels simplify the real world**
- **Label distribution is closer to the ground-truth!**

The Power of Label Space



Instance Space



Logical Label Space

A@<OPM@ !@SOM<>ODJI

A@<OPM@ N@G@>ODJI

H<IDAJG? !@H=@??DIB

?DH@INDJI<GDOT M@?P>ODJI

Many Analytic Tools

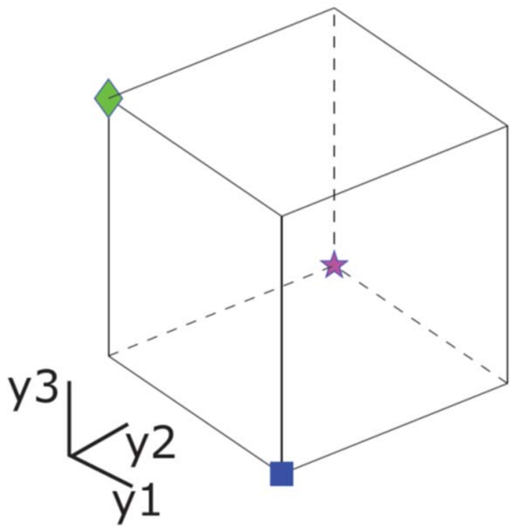
……

0: Irrelevant label

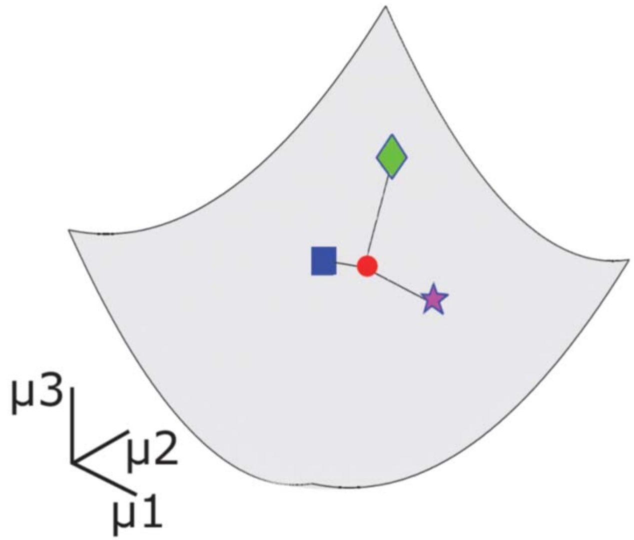
1: Relevant label

Limited Expressiveness
Limited Analytic Tools

The Power of Label Space



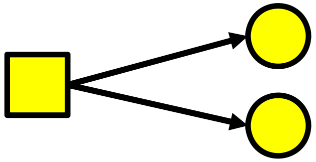
Logical Label Space



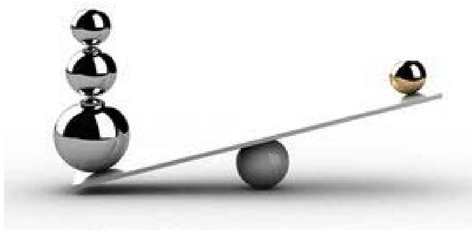
Continuous Label Space
(Label Distribution)

The Power of Label Space

Label distribution space offers more possibilities for …

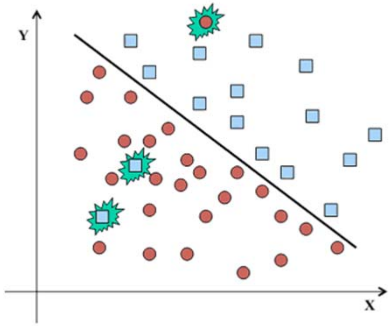


Label Ambiguity

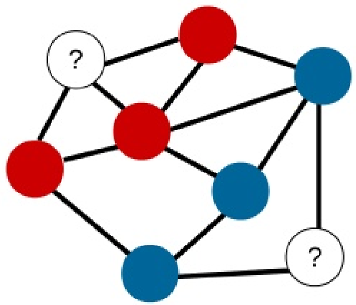


Label Importance

Label Uncertainty



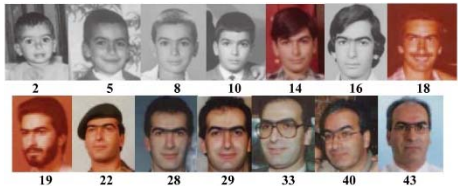
Nosy Label



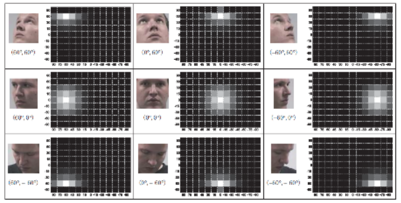
Label Correlation

Prior Knowledge Embedding

Applications with the Power of Label Space

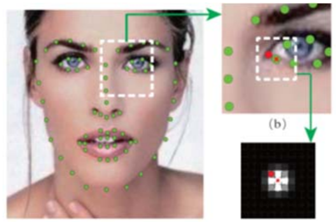


Age Estimation
[Gao, et al., IJCAI'18]; [Hou, et al., AAAI'18]; [He, et al., TIP'17]; [Geng, Yin, and Zhou, TPAMI'13]

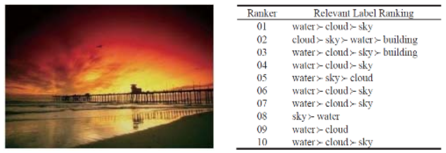


Head Pose Estimation
[Geng and Xia, CVPR'14]

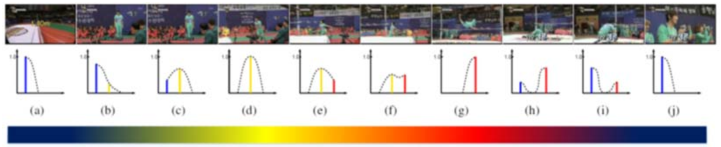
Text Emotion Estimation
[Zhou, et al., EMNLP'16]



Facial Landmark Detection
[Su and Geng, AAAI'19]

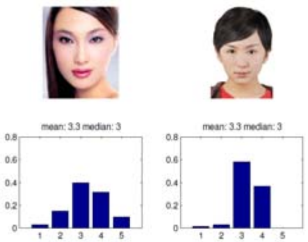


Multilabel Ranking
[Geng and Luo, CVPR'14]



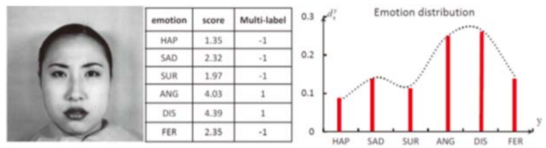
Video Parsing
[Geng and Ling, AAAI'17]

Prediction of Crowd Opinion on Movies
[Geng and Hou, IJCAI'15]



Beauty Sense
[Ren and Geng, IJCAI'17]

Indoor Crowd Counting
[Ling and Geng, TIP'19]



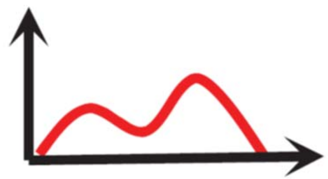
Expression Recognition
[Zhou, Xue and Geng, ACMMM'15]

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

Logical Labels

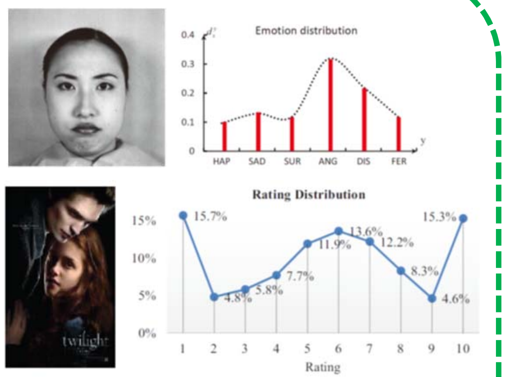
{0, 1, 0, 1, 0}



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

Label Enhancement (LE)

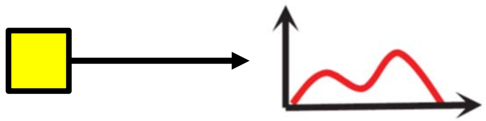
Label Distribution & Label Enhancement



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

[Xu and Geng, IJCAI'18]

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

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

Label Enhancement can be defined as follows.

Given a training set $S = \{(\boldsymbol{x}_i, \boldsymbol{L}_i) | 1 \leq i \leq n\}$, label enhancement is to recover the label distribution $\boldsymbol{D}_i$ of $\boldsymbol{x}_i$ from the logical label vector $\boldsymbol{L}_i$, and thus transform S into an LDL training set $E = \{(\boldsymbol{x}_i, \boldsymbol{D}_i) | 1 \leq i \leq n\}$.

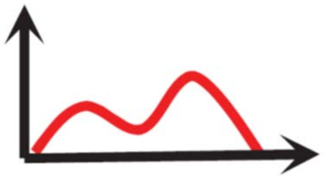
What Added?

{0, 1, 0, 1, 0} +

Topological relationship among feature vectors (x space)

Correlation among labels (y space)

=



Outline

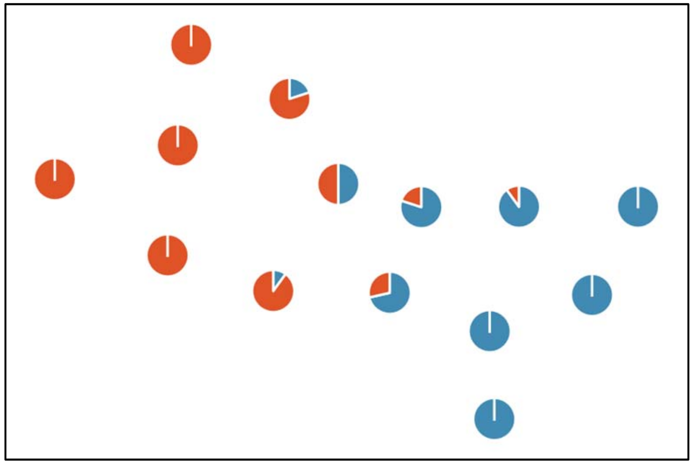
- Introduction
- Label Enhancement
  - Formulation
  - Algorithms
  - Experiments
- Conclusion

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)

[Gayar et al., ANNPR'06]



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.

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 $\boldsymbol{M}$ (m_{ik} denotes the membership of $\boldsymbol{x}_i$ to the k-th cluster)
  2. Calculate the cluster prototype
  $$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{n}(m_{ik})^{\beta}\boldsymbol{x}_i}{\sum_{i=1}^{n}(m_{ik})^{\beta}}$$
  3. Update the cluster membership matrix $\boldsymbol{M}$
  $$m_{ik} = \frac{1}{\sum_{j=1}^{p}\left(\frac{Dist(\boldsymbol{x}_i, \boldsymbol{\mu}_k)}{Dist(\boldsymbol{x}_i, \boldsymbol{\mu}_j)}\right)^{\frac{2}{\beta-1}}}$$
  4. Repeat 2 and 3 until convergence

Each row of $\boldsymbol{M}$, $\boldsymbol{m}_i$, represents the membership of the instance $\boldsymbol{x}_i$ 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 $\boldsymbol{A}$
  2. Update each row $\boldsymbol{A}_j$ with

$$\boldsymbol{A}_j = \boldsymbol{A}_j + \boldsymbol{m}_i, \qquad if\ l_{\boldsymbol{x}_i}^{y_j} = 1$$

  3. Normalized each column of $\boldsymbol{A}$
  4. Normalized each row of $\boldsymbol{A}$

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

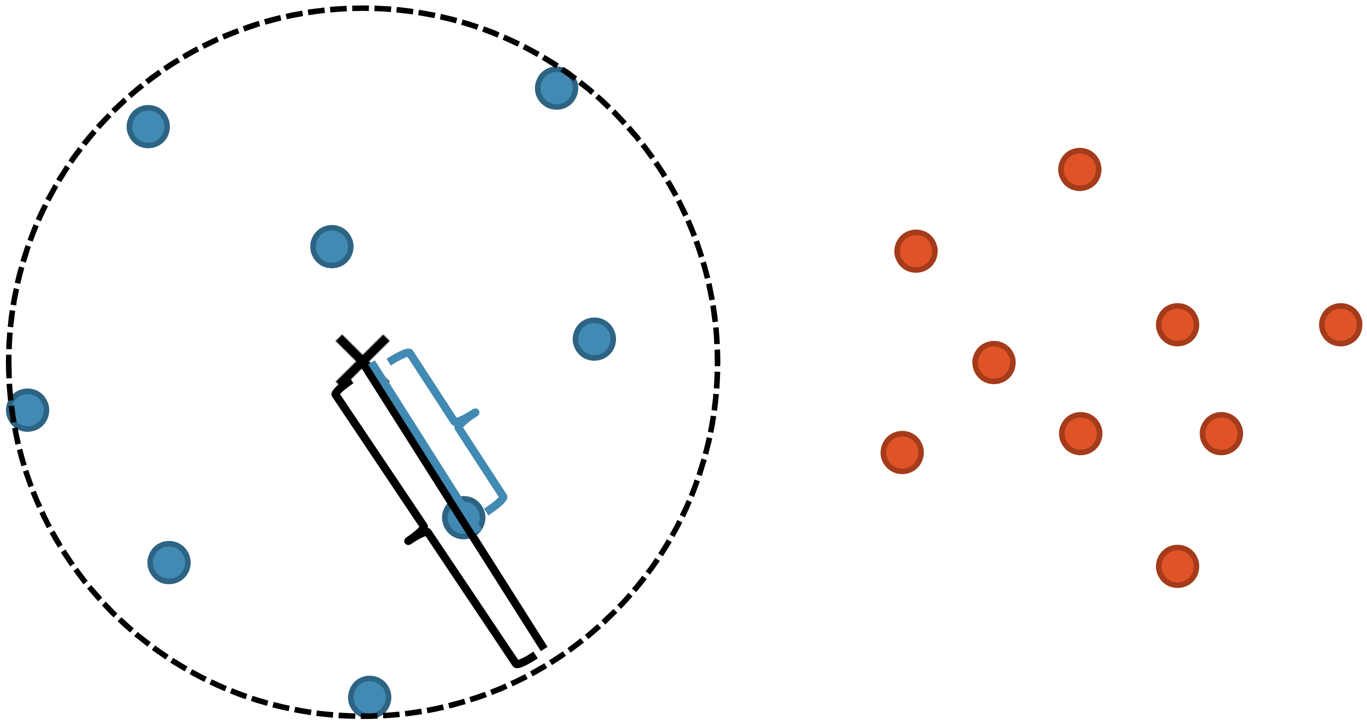
- Step 3： Calculate the label distribution of $\boldsymbol{x}_i$
  1. $\boldsymbol{D}_i = \boldsymbol{A} \circ \boldsymbol{m}_i$ (fuzzy composition)

$$D_i^j = \max_k (a_{jk} \times m_{ik})$$

  2. Normalize $\boldsymbol{D}_i$

LE based on kernel method (KM)

[Jiang et al., NCA'06]



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

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

$(\boldsymbol{\Psi}^{y_j})^2$ can be calculated via inner product of $\phi(\boldsymbol{x}_i)$

where $\phi(\boldsymbol{x}_i)$ is a nonlinear function determined by the kernel function

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i) \cdot \phi(\boldsymbol{x}_j)$$

- Step 2: Calculate the class radius

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

r_j^2 can be calculated via inner product of $\phi(\boldsymbol{x}_i)$

- Step 3: Calculate the distance between instance $\boldsymbol{x}_i$ and class center

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

d_{ij}^2 can be calculated via inner product of $\phi(\boldsymbol{x}_i)$

LE based on kernel method (KM)

[Jiang et al., NCA'06]

- Step 4: calculate the membership of instance $\boldsymbol{x}_i$ to label y_j

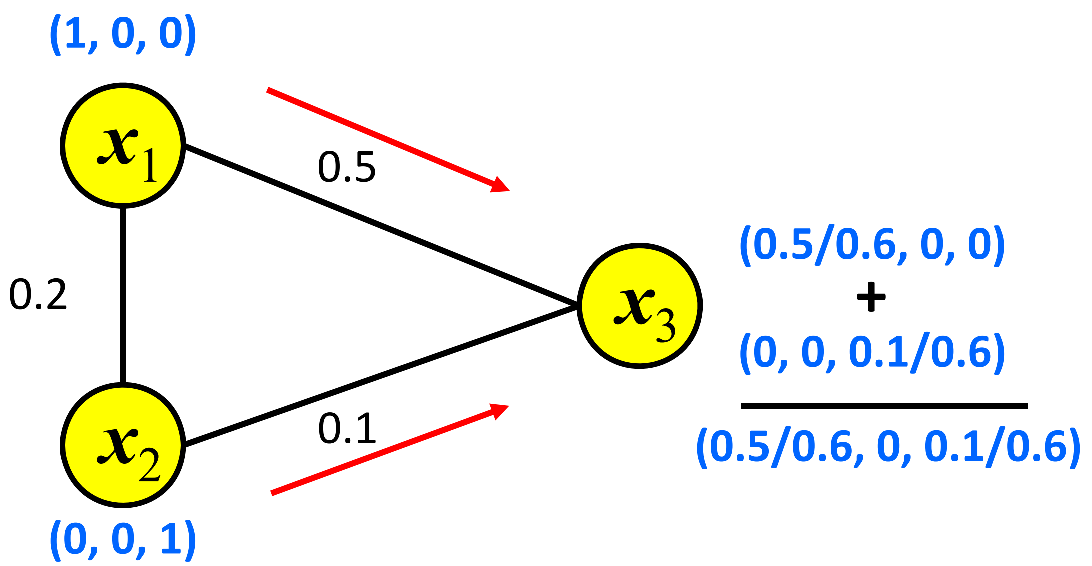
$$m_{\boldsymbol{x}_i}^{y_j} = \begin{cases} 1 - \sqrt{\dfrac{\left\| d_{ij}^2 \right\|}{\left(r_j^2 + \delta\right)}} & if\ l_{\boldsymbol{x}_i}^{y_j} = 1 \\ 0 & if\ l_{\boldsymbol{x}_i}^{y_j} = 0 \end{cases}$$

Cannot change the membership of irrelevant labels

- Step 5: Normalize $\boldsymbol{m}_{\boldsymbol{x}_i} = \left[m_{\boldsymbol{x}_i}^{y_1}, m_{\boldsymbol{x}_i}^{y_2}, \dots, m_{\boldsymbol{x}_i}^{y_c} \right]$

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]

$$G = (V, E)$$

$$V = \{\boldsymbol{x}_i \mid 1 \leq i \leq m\}$$

$$\forall_{i,j=1}^{m}: \quad w_{ij} = \begin{cases} \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{2\sigma^2}\right), & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}$$

Propagation Matrix

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

$$\mathbf{F}^{(0)} = \mathbf{\Phi}$$

$$\forall_{i=1}^{m} \forall_{l=0}^{q}: \quad \phi_{il} = \begin{cases} \tau, & \text{if } y_l = y_0 \\ 1, & \text{if } y_l \in Y_i \\ 0, & \text{otherwise} \end{cases}$$

Label Propagation

$$\mathbf{F}^{(t)} = \alpha \mathbf{P}\mathbf{F}^{(t-1)} + (1-\alpha)\mathbf{\Phi}$$

Converge to

$$\mathbf{F}^* = (1-\alpha)(\mathbf{I} - \alpha\mathbf{P})^{-1}\mathbf{\Phi}$$

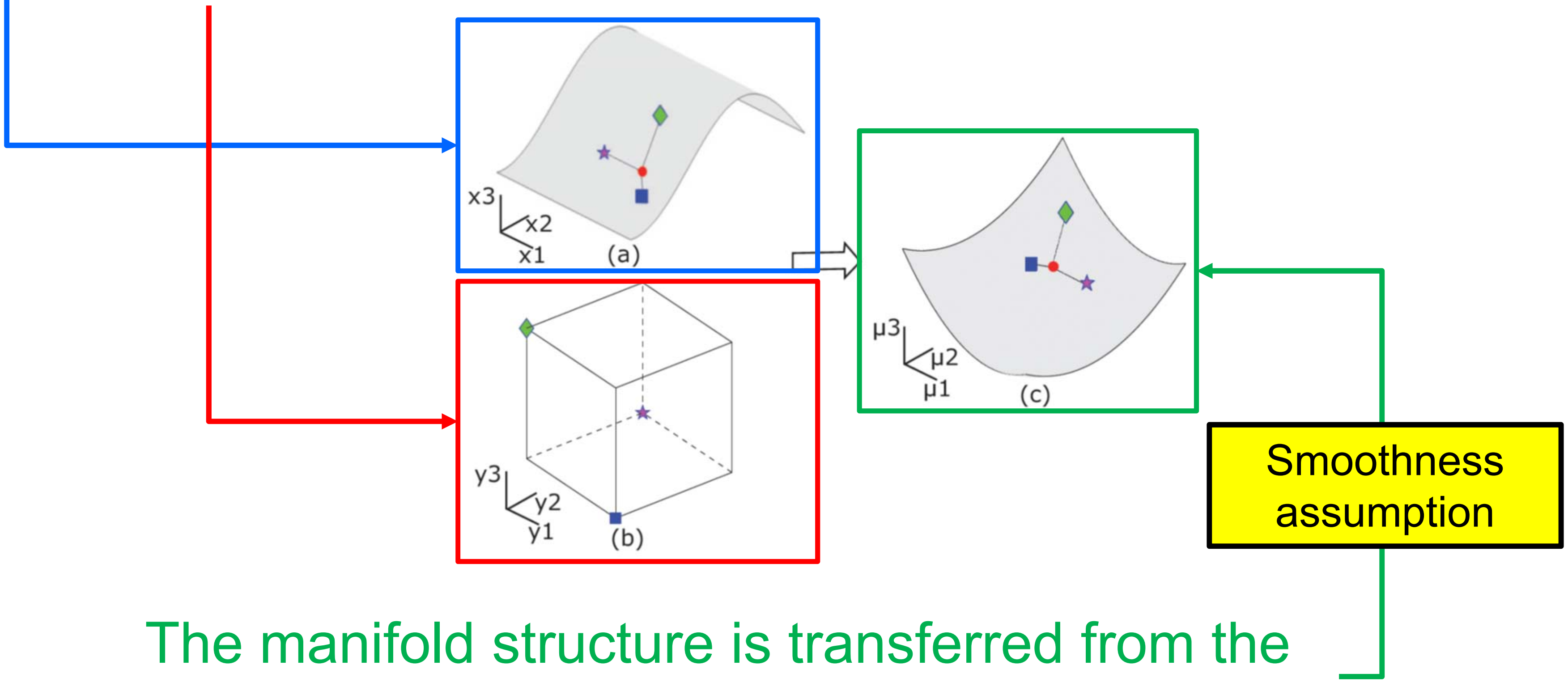
$$\mu_{\boldsymbol{x}_i}^{y_l} = \frac{f_{il}^*}{\sum_{k=0}^{q} f_{ik}^*}$$

Label Distribution

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]

$$\arg\min_{\boldsymbol{W}} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \sum_{j \neq i} W_i^j \boldsymbol{x}_j\|^2$$

$$\text{s.t.} \quad \mathbf{1}^{\mathrm{T}} \boldsymbol{W}_i = 1$$

If $\boldsymbol{x}_j$ is not the neighbor of $\boldsymbol{x}_i$, then $\boldsymbol{W}_i^j = 0$

Local topological structure

- Manifold learning in label space

$$\arg\min_{\boldsymbol{\mu}} \sum_{i=1}^{n} \|\boldsymbol{\mu}_i - \sum_{j \neq i} W_i^j \boldsymbol{\mu}_j\|^2$$

$$\text{s.t.} \quad \forall 1 \leq i \leq n, 1 \leq l \leq q \quad y_i^l \mu_i^l \geq \lambda, \lambda > 0$$

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

Control the sign and scale

Graph Laplacian Label Enhancement（GLLE）

[Xu and Geng, IJCAI'18]

- Model

Nonlinear transformation

$$\boldsymbol{D}_i = \boldsymbol{W}^{\top}\varphi(\boldsymbol{x}_i) + \boldsymbol{b} = \widehat{\boldsymbol{W}}\boldsymbol{\phi}_i$$

Goal Determining the best parameter $\widehat{\boldsymbol{W}}^*$

- Target function

Logical label loss

$$\min_{\widehat{\boldsymbol{W}}} L(\widehat{\boldsymbol{W}}) + \lambda\Omega(\widehat{\boldsymbol{W}})$$

Feature space constraint

Graph Laplacian Label Enhancement（GLLE）

[Xu and Geng, IJCAI'18]

- The first part of the target function

$$L(\widehat{\boldsymbol{W}}) = \sum_{i=1}^{n} \left\| \widehat{\boldsymbol{W}} \boldsymbol{\phi}_i - \boldsymbol{L}_i \right\|^2$$

Least squares（LS）

- The second part of the target function

Smoothness assumption

$$\Omega(\widehat{\boldsymbol{W}}) = \sum_{i,j} a_{ij} \left\| \boldsymbol{D}_i - \boldsymbol{D}_j \right\|^2 = \boldsymbol{tr}(\boldsymbol{D}\boldsymbol{G}\boldsymbol{D}^{\top})$$

$$a_{ij} = \begin{cases} \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right) & \text{if } \boldsymbol{x}_j \in N(i) \\ 0 & \text{otherwise} \end{cases}$$

Correlation between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$

Graph Laplacian

$$\boldsymbol{G} = \widehat{\boldsymbol{A}} - \boldsymbol{A},\ \hat{a}_{ij} = \sum_{j=1}^{n} a_{ij}$$

Outline

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

Datasets

No.	Dataset	#Examples	#Features	#Labels
1	Artificial	2601	3	3
2	SJAFFE	213	243	6
3	Natural Scene	2,000	294	9
4	Yeast-spoem	2,465	24	2
5	Yeast-spo5	2,465	24	3
6	Yeast-dtt	2,465	24	4
7	Yeast-cold	2,465	24	4
8	Yeast-heat	2,465	24	6
9	Yeast-spo	2,465	24	6
10	Yeast-diau	2,465	24	7
11	Yeast-elu	2,465	24	14
12	Yeast-cdc	2,465	24	15
13	Yeast-alpha	2,465	24	18
14	SBU_3DFE	2,500	243	6
15	Movie	7,755	1,869	5

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

Artificial dataset

- The label distribution $\boldsymbol{D} = \left[d_{\boldsymbol{x}}^{y_1}, d_{\boldsymbol{x}}^{y_2}, d_{\boldsymbol{x}}^{y_3}\right]$ of $\boldsymbol{x} = [x_1, x_2, x_3]$ is generated in the following way

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

$$\psi_1 = (\boldsymbol{\omega}_1^T \boldsymbol{t})^2, \quad \psi_2 = (\boldsymbol{\omega}_2^T \boldsymbol{t} + \lambda_1 \psi_1)^2, \quad \psi_3 = (\boldsymbol{\omega}_3^T \boldsymbol{t} + \lambda_2 \psi_2)^2,$$

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

where $a = 1$, $b = 0.5$, $c = 0.2$, $d = 1$, $\boldsymbol{\omega}_1^T = [4,2,1]$, $\boldsymbol{\omega}_2^T = [1,2,4]$, $\boldsymbol{\omega}_3^T = [1,4,2]$, $\lambda_1 = \lambda_2 = 0.01$。

- Sampling

The first two components of $\boldsymbol{x}$, 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 51×51 = 2601 instances. The third component x_3 is generated by

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

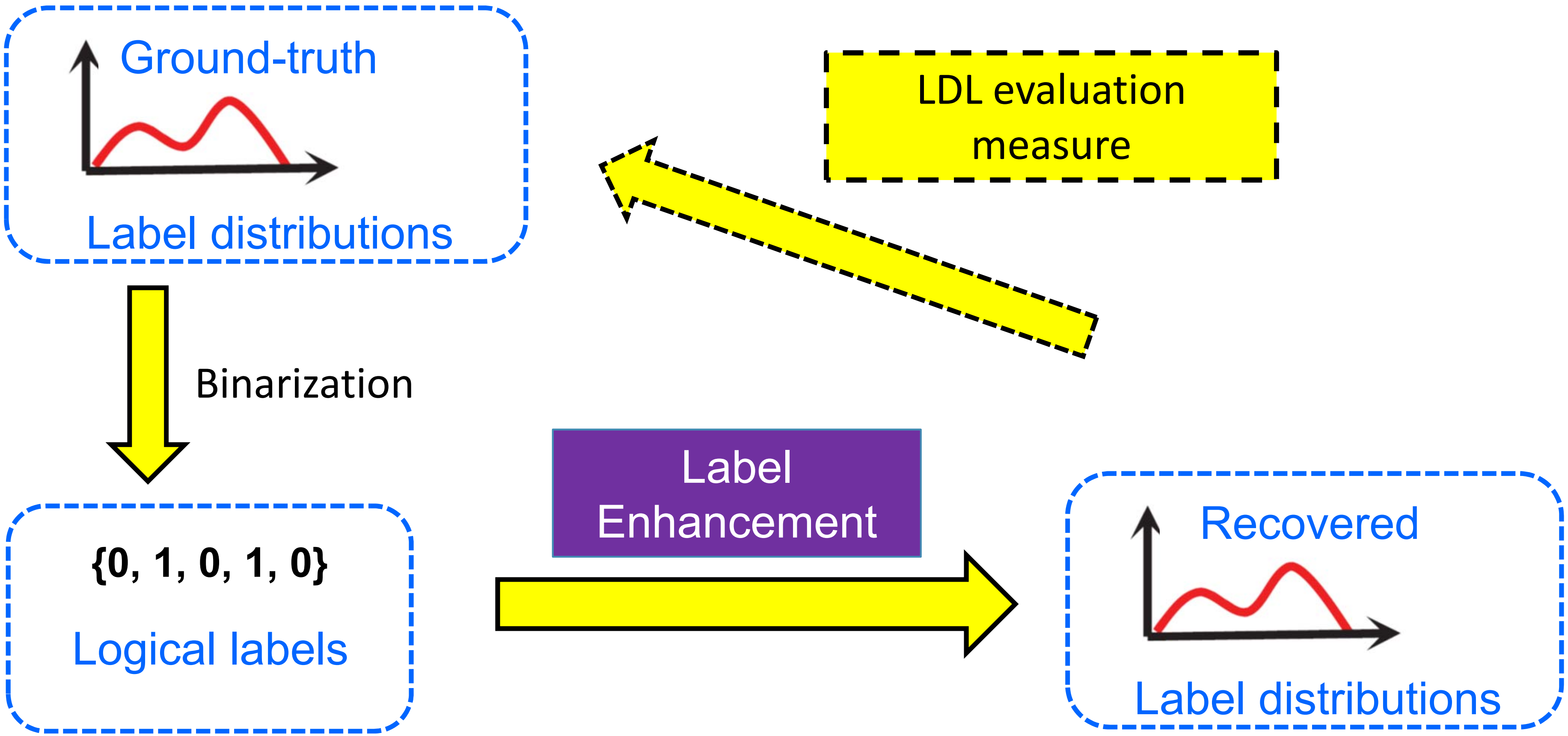
Datasets

- Label distribution binarization

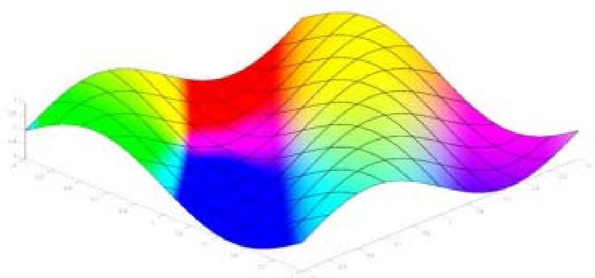
1. Initialize the relevant label set $\mathcal{Y}^+ = \emptyset$, the irrelevant label set $\mathcal{Y}^- = \mathcal{Y}$, $\boldsymbol{L} = \mathbf{0}$;
2. Select $y_j \in \mathcal{Y}^-$ with the highest description degree, then $\mathcal{Y}^+ = \mathcal{Y}^+ + y_j$, $\mathcal{Y}^- = \mathcal{Y}^- - y_j$, $l_{\boldsymbol{x}}^{y_j} = 1$;
3. $t = \sum_{y_j \in \mathcal{Y}^+} d_{\boldsymbol{x}}^{y_j}$, if $t < T$, go back to step 2, else end.

$T = 0.5$

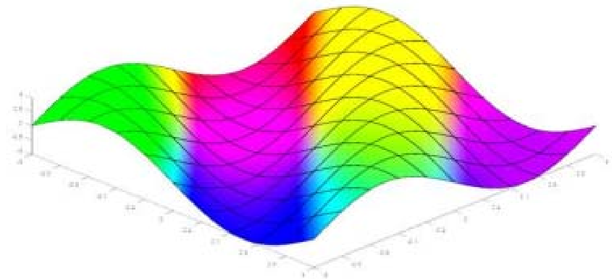
Recovery Performance



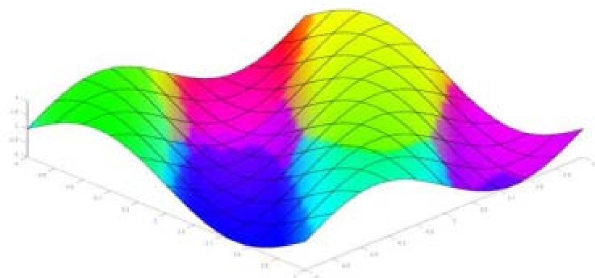
Recovery Performance



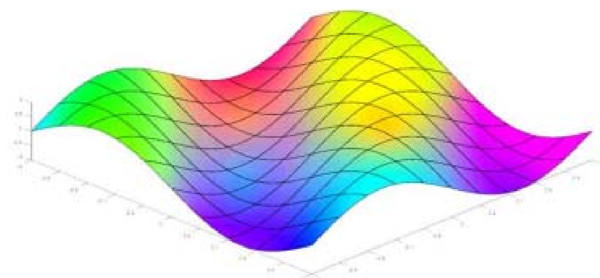
(a) Ground-Truth



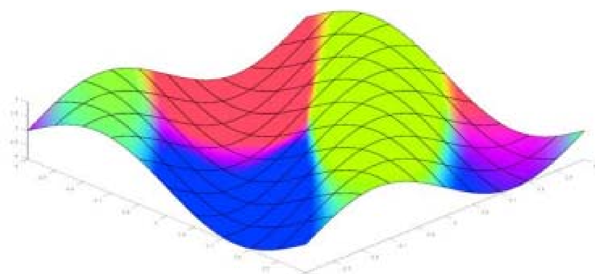
(b) GLLE



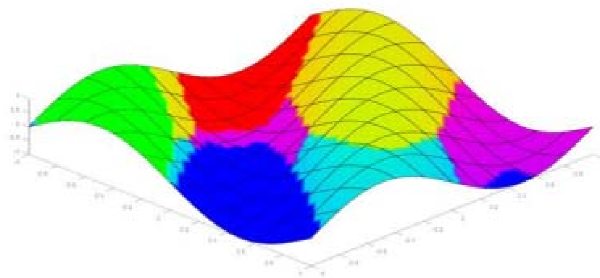
(c) LP



(d) ML



(e) FCM



(f) KM

Recovery Performance

Datasets	FCM	KM	LP	ML	GLLE
Artificial	0.188(3)	0.260(5)	0.130(2)	0.227(4)	**0.108(1)**
SJAFFE	0.132(3)	0.214(5)	0.107(2)	0.190(4)	**0.100(1)**
Natural Scene	0.368(5)	0.306(4)	**0.275(1)**	0.295(2)	0.296(3)
Yeast-spoem	0.233(3)	0.408(5)	0.163(2)	0.400(4)	**0.108(1)**
Yeast-spo5	0.162(3)	0.277(5)	0.114(2)	0.273(4)	**0.092(1)**
Yeast-dtt	0.097(2)	0.257(5)	0.128(3)	0.244(4)	**0.065(1)**
Yeast-cold	0.141(3)	0.252(5)	0.137(2)	0.242(4)	**0.093(1)**
Yeast-heat	0.169(4)	0.175(5)	0.086(2)	0.165(3)	**0.056(1)**
Yeast-spo	0.130(3)	0.175(5)	0.090(2)	0.171(4)	**0.067(1)**
Yeast-diau	0.124(3)	0.152(5)	0.099(2)	0.148(4)	**0.084(1)**
Yeast-elu	0.052(3)	0.078(5)	0.044(2)	0.072(4)	**0.030(1)**
Yeast-cdc	0.051(3)	0.076(5)	0.042(2)	0.071(4)	**0.038(1)**
Yeast-alpha	0.044(3)	0.063(5)	0.040(2)	0.057(4)	**0.033(1)**
SBU_3DFE	0.135(2)	0.238(5)	**0.123(1)**	0.233(4)	0.141(3)
Movie	0.230(4)	0.234(5)	0.161(2)	0.164(3)	**0.160(1)**
Avg. Rank	3.13	4.93	1.93	3.73	1.27

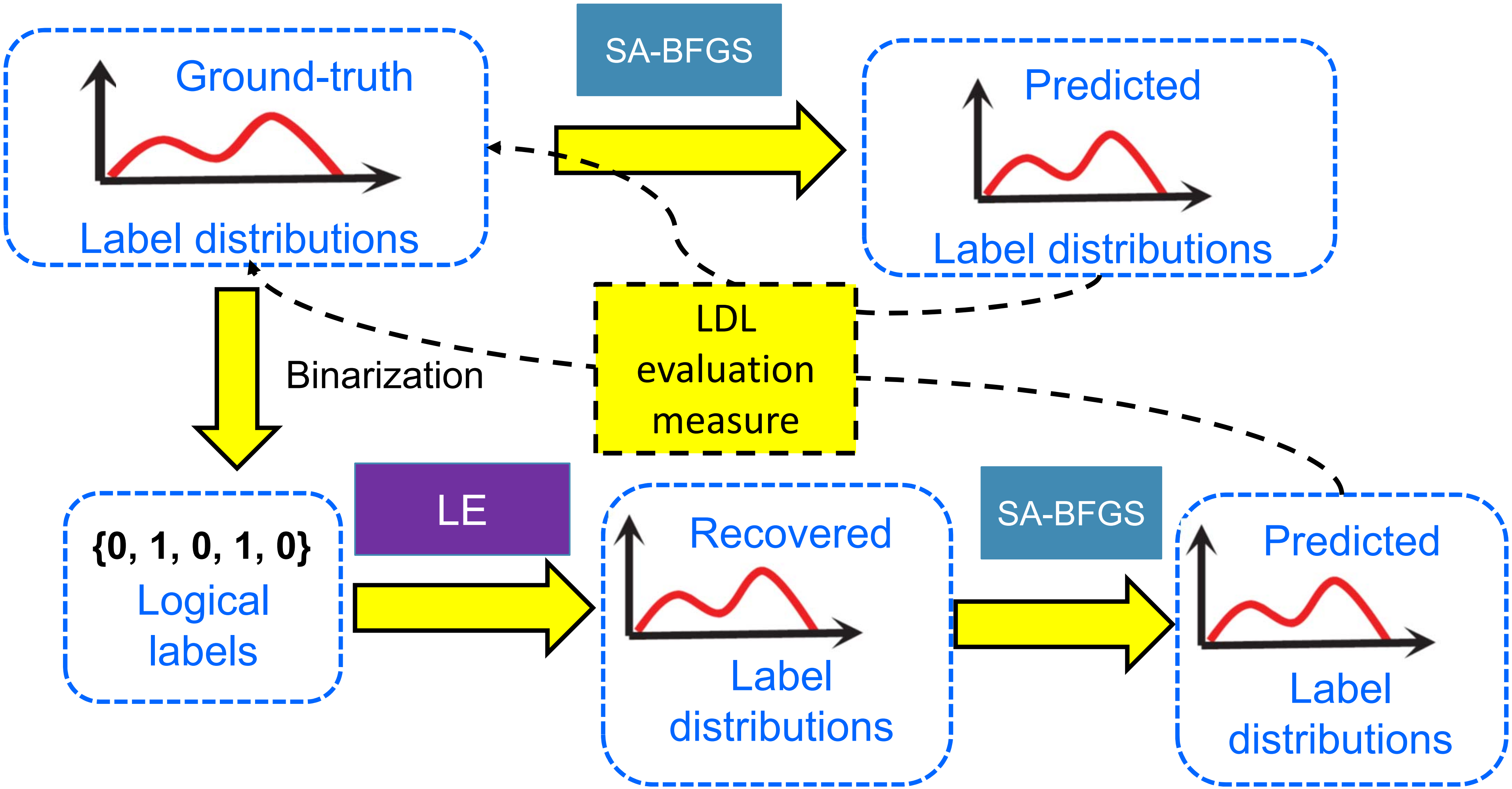
Table 2: Recovery Results (value(rank)) Measured by Cheb ↓

Recovery Performance

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

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

LDL Predictive Performance



LDL Predictive Performance

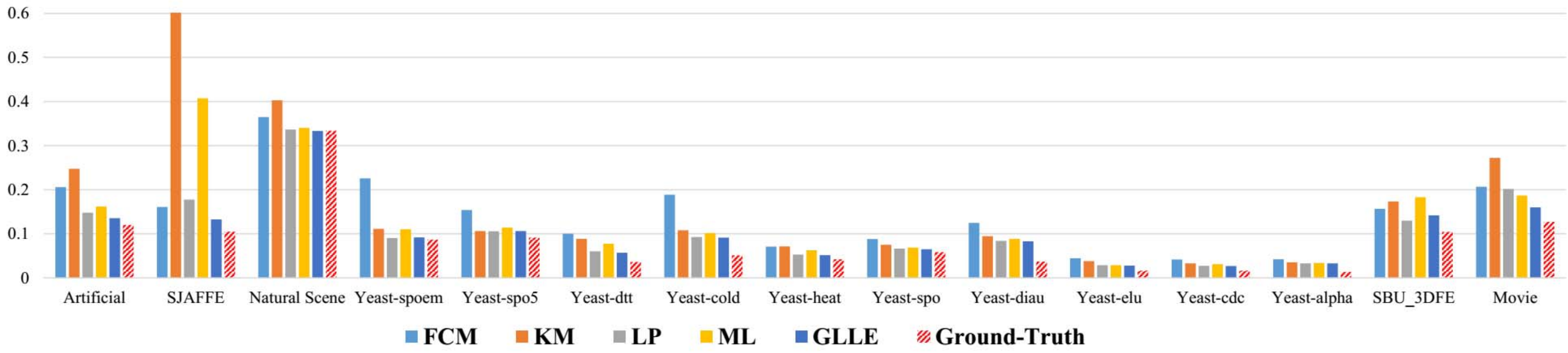


Figure 3: Comparison of the LDL after the LE pre-process against the direct LDL measured by Cheb ↓.

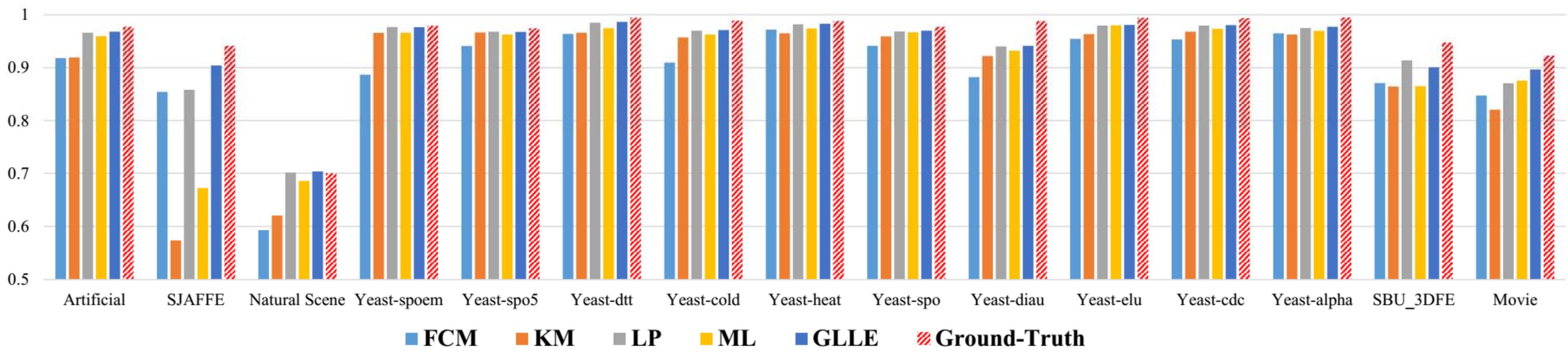


Figure 4: Comparison of the LDL after the LE pre-process against the direct LDL measured by Cosine ↑.

LDL Predictive Performance

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

Criterion	FCM	KM	LP	ML	GLLE
Cheb	4.40	4.20	2.00	3.13	1.27
Clark	4.33	4.07	2.27	3.07	1.27
Canber	4.20	4.13	2.27	3.13	1.27
KL	4.37	4.30	2.00	3.13	1.20
Cosine	4.53	4.27	1.93	3.07	1.20
Intersec	4.40	4.20	1.93	3.13	1.33

Outline

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

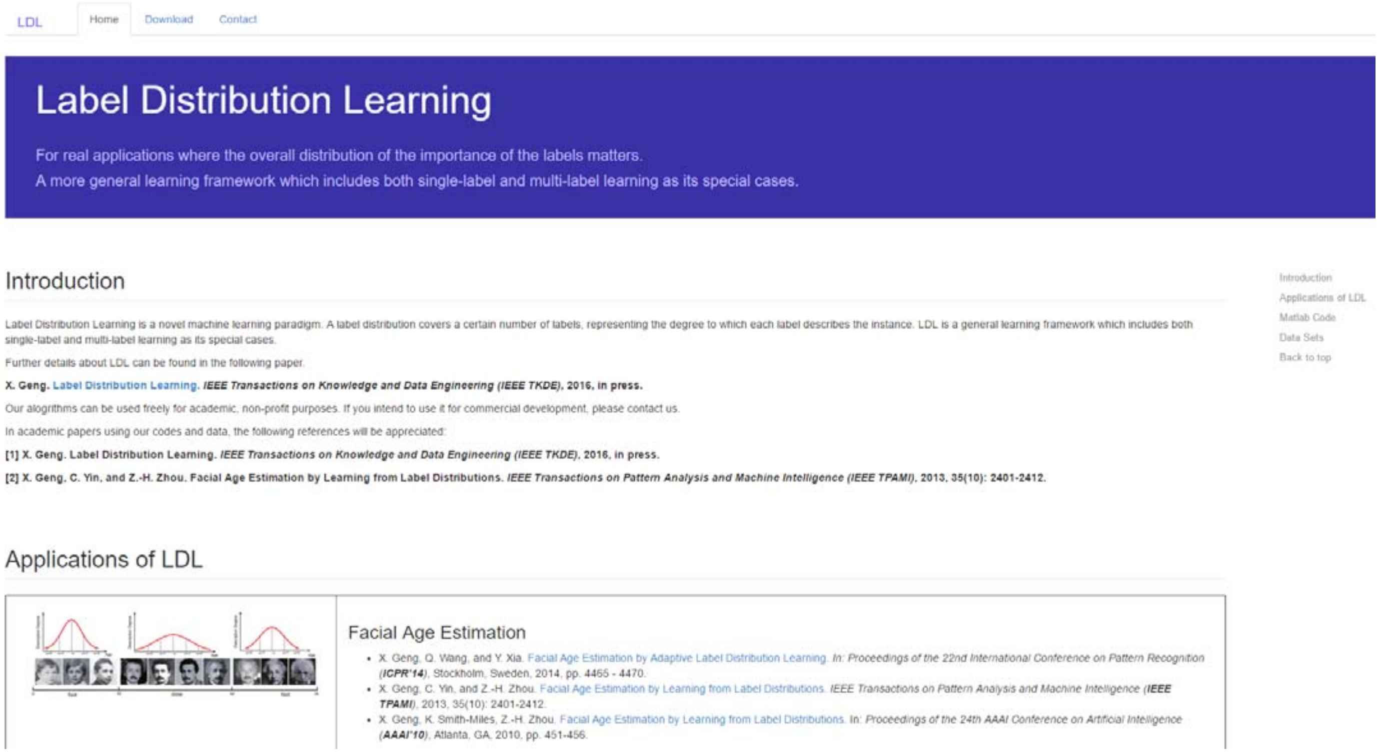
Conclusion

- **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/LDL/index.htm


LDL | Home | Download | Contact

Label Distribution Learning

For real applications where the overall distribution of the importance of the labels matters.
A more general learning framework which includes both single-label and multi-label learning as its special cases.

Introduction

Label Distribution Learning is a novel machine learning paradigm. A label distribution covers a certain number of labels, representing the degree to which each label describes the instance. LDL is a general learning framework which includes both single-label and multi-label learning as its special cases.

Further details about LDL can be found in the following paper.

X. Geng. Label Distribution Learning. *IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)*, 2016, in press.

Our alogrithms can be used freely for academic, non-profit purposes. If you intend to use it for commercial development, please contact us.

In academic papers using our codes and data, the following references will be appreciated:

[1] X. Geng. Label Distribution Learning. *IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)*, 2016, in press.

[2] X. Geng, C. Yin, and Z.-H. Zhou. Facial Age Estimation by Learning from Label Distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 2013, 35(10): 2401-2412.

Introduction
Applications of LDL
Matlab Code
Data Sets
Back to top

Applications of LDL

Facial Age Estimation

- X. Geng, Q. Wang, and Y. Xia. Facial Age Estimation by Adaptive Label Distribution Learning. *In: Proceedings of the 22nd International Conference on Pattern Recognition (**ICPR'14**)*, Stockholm, Sweden, 2014, pp. 4465 - 4470.
- X. Geng, C. Yin, and Z.-H. Zhou. Facial Age Estimation by Learning from Label Distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (**IEEE TPAMI**)*, 2013, 35(10): 2401-2412.
- X. Geng, K. Smith-Miles, Z.-H. Zhou. Facial Age Estimation by Learning from Label Distributions. In: *Proceedings of the 24th AAAI Conference on Artificial Intelligence (**AAAI'10**)*, Atlanta, GA, 2010, pp. 451-456.

Acknowledgements

- Collaborators

Prof. Zhi-Hua Zhou

Prof. Kate Smith-Miles

Prof. Min-Ling Zhang

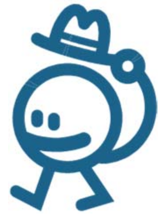
Prof. Deyu Zhou

Dr. Hui Xue

- Students

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

THANK YOU！



http:// palm.seu.edu.cn