福州大学
FUZHOU UNIVERSITY

标记增强技术的研究
Research on Label Enhancement

Qinghai Zheng （郑清海）

Jul. 02, Xi’an

Outline

1. Background: Label Distribution Learning and Label Enhancement

2. Overview of Label Enhancement

3. Label Enhancement via Global Sample Correlation

4. Label Enhancement via Label Information Bottleneck

5. Discussions and Further Works

Outline

1. Background: Label Distribution Learning and Label Enhancement

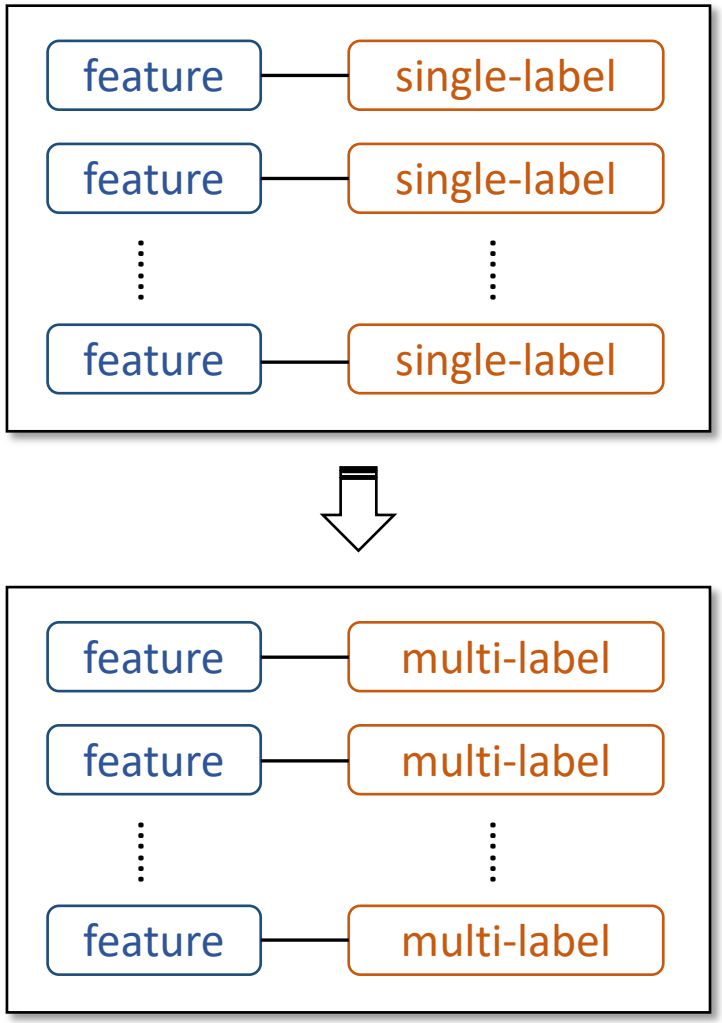
2. Overview of Label Enhancement

3. Label Enhancement via Global Sample Correlation

4. Label Enhancement via Label Information Bottleneck

5. Discussions and Further Works

Traditional Learning Paradigm



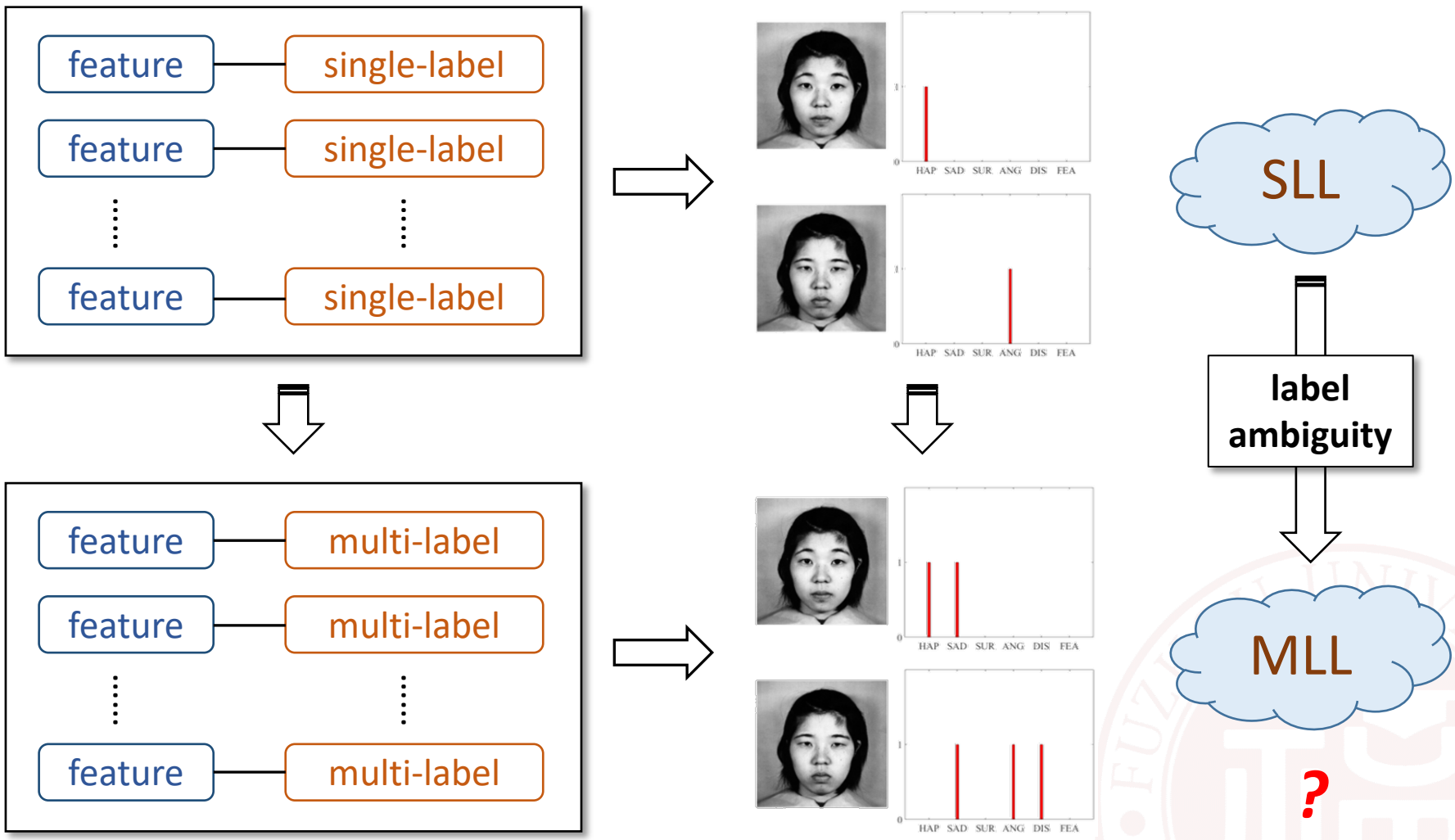
SLL

- single-label learning
- an instance is mapped to one single logical label simply

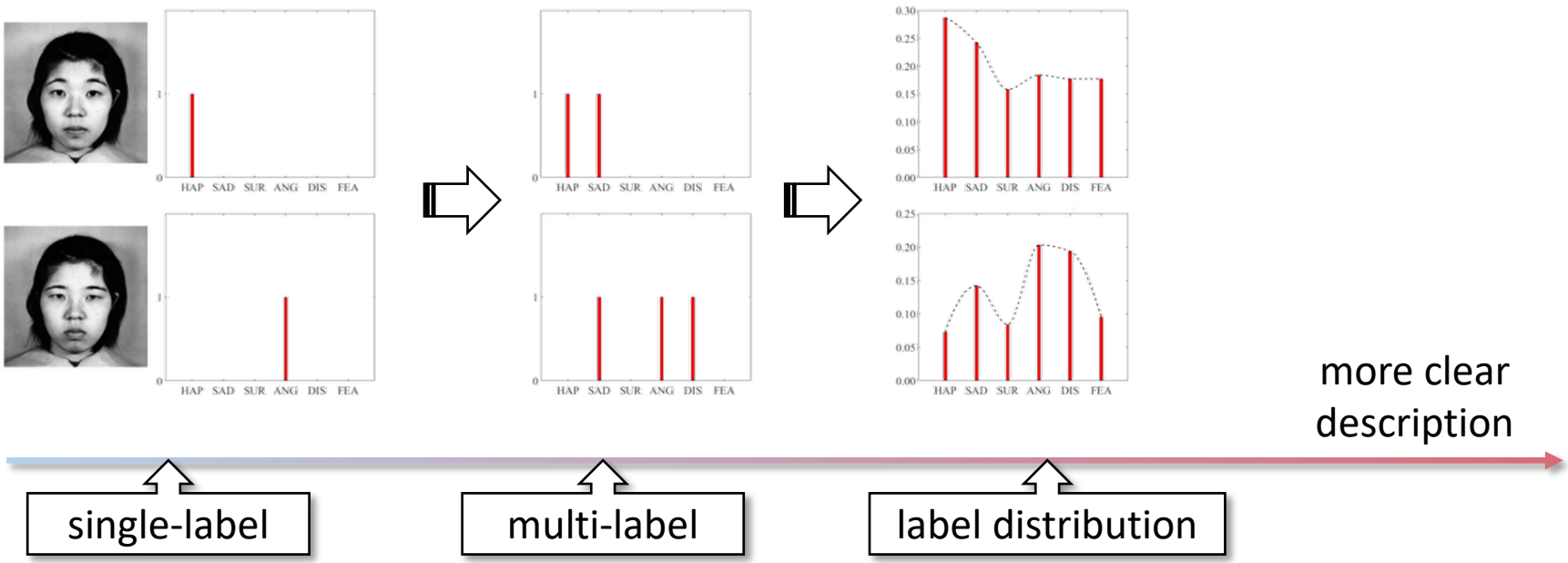
MLL

- multi-label learning
- an instance is mapped to multiple logical labels

Traditional Learning Paradigm



MLL→LDL (Label Distribution Learning)

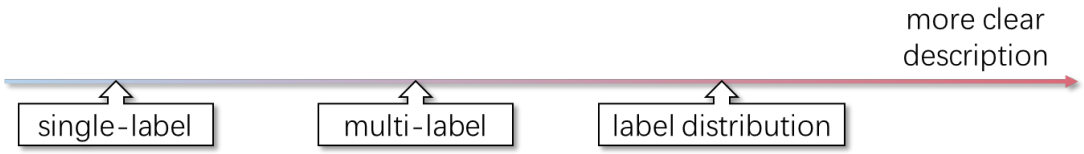


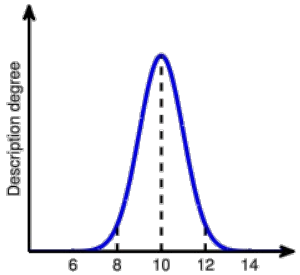
- SLL and MLL:
  - which label can describe the instance?

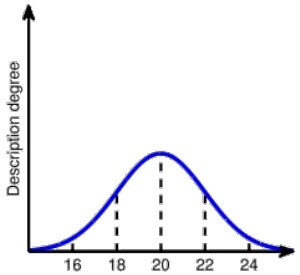
- LDL:
  - how much does each label describe the instance?

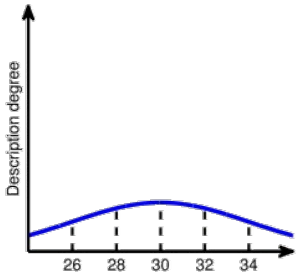
Applications of LDL

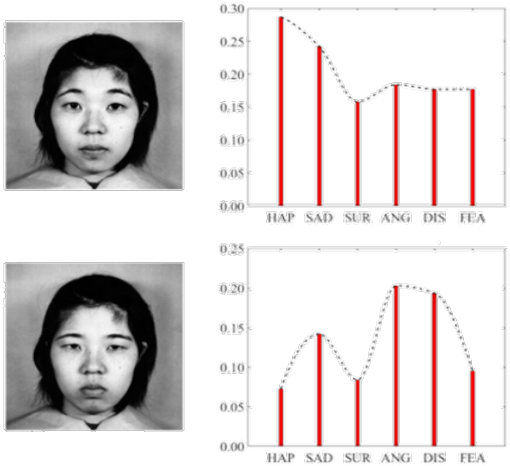
- Emotion analysis from the facial expressions
- Crowd counting
- Image recognition
- Facial age estimation
- …

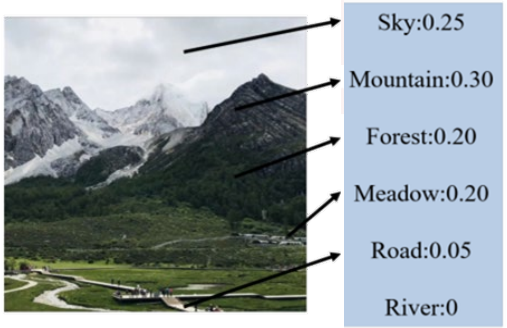




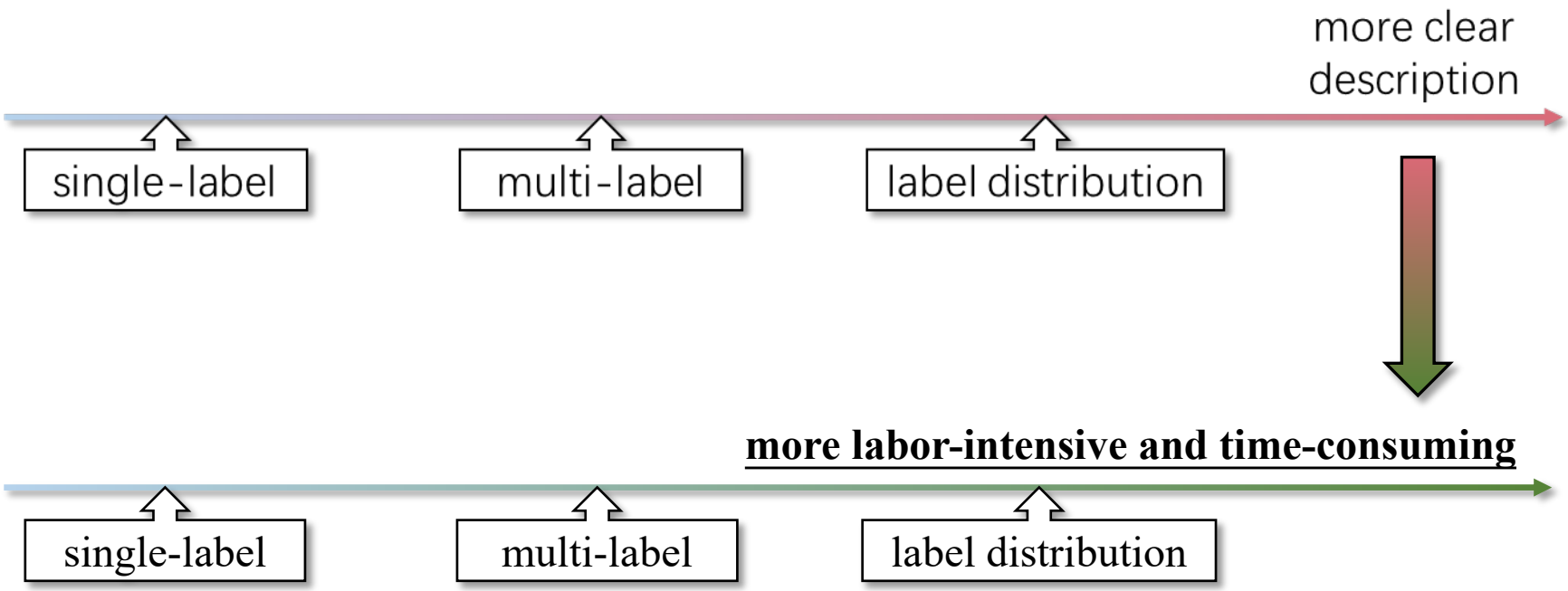








LE (Label Enhancement)



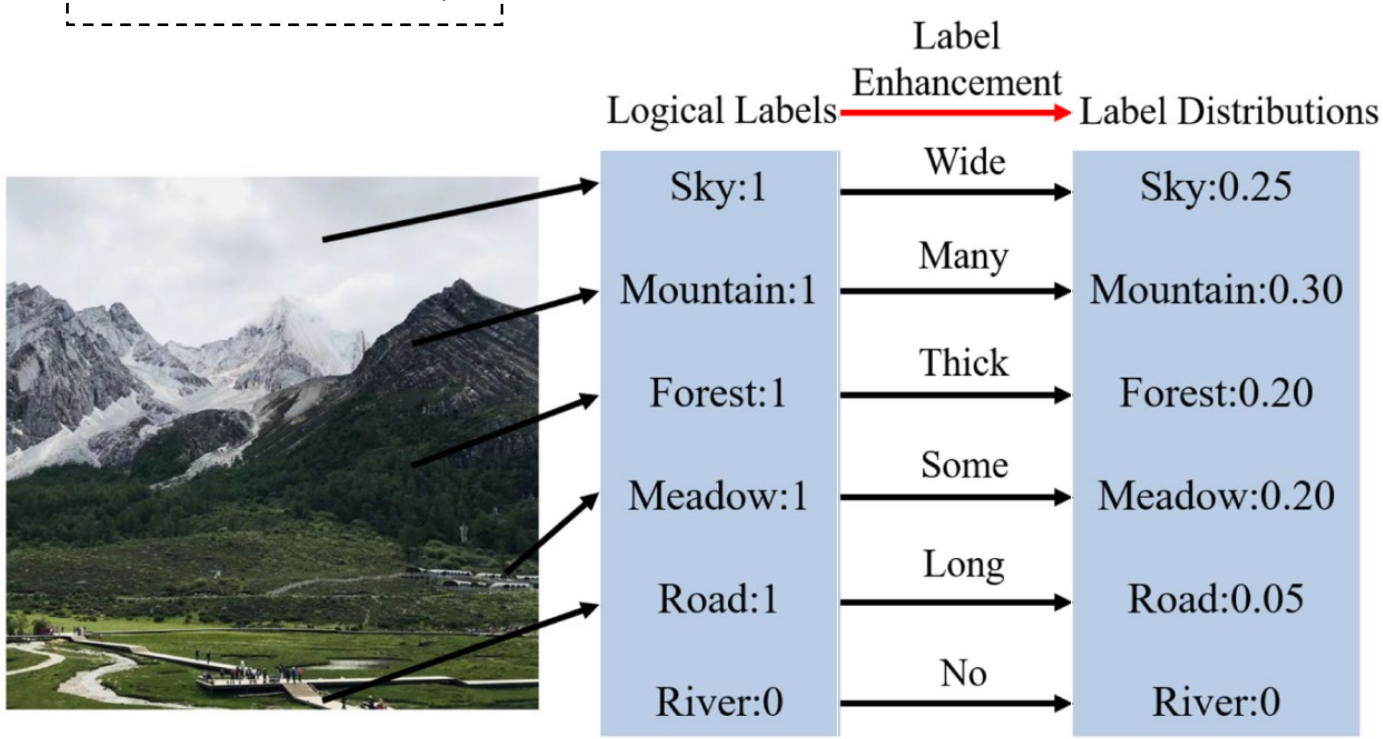
- SLL → MLL → LDL:
  - annotation is more labor-intensive and time-consuming!

LE (Label Enhancement)

- Label Enhancement (LE): LE exactly recovers label distributions from the off-the-shelf logical labels and the implicit information of the given features

  - $X = [x_1; x_2; \dots; x_n]$

    $\Gamma = [L_1; L_2; \dots; L_n]$ ⇒ $\mathfrak{D} = [D_1; D_2; \dots; D_n]$



Outline

1. Background: Label Distribution Learning and Label Enhancement

2. Overview of Label Enhancement

3. Label Enhancement via Global Sample Correlation

4. Label Enhancement via Label Information Bottleneck

5. Discussions and Further Works

Algorithm Adaptation

- Methods that can be extended to tackle the problems of label enhancement
  - FCM (Fuzzy Clustering Method)
    - membership degree: $\omega_{ik} = \dfrac{1}{\sum_{j=1}^{t} \left(\frac{\|x_i - c_k\|_2}{\|x_i - c_j\|_2} \right)^{\frac{1}{\beta-1}}}$
    - prototype label matrix: $Q_j = Q_j + \omega_i, s.t., \ l_{x_i}^{y_j} = 1$
    - label distribution: $D_i = Q \circ \omega_i$
  - KM (Kernel Method)
    - center of one part: $p_+^{y_m} = \dfrac{1}{n_+} \sum_{x \in C_+^{y_m}} f(x_i)$
    - radius and distance: $r_+ = \max \left\| p_+^{y_m} - f(x_i) \right\| \qquad dist_i = \left\| f(x_i) - p_+^{y_m} \right\|$
    - label distribution: $d_{x_i}^{y_m} = \begin{cases} 1 - (\frac{dist_i^2}{r_+^2 + \eta}), \text{if } l_{x_i}^{y_m} = 1 \\ 0, \qquad \text{if } l_{x_i}^{y_m} = 0 \end{cases}$

Specialized Algorithms

- Methods that are specially designed for label enhancement
  - GLLE (Graph Laplacian Label Enhancement)

    Xu N, et al. Label Enhancement for Label Distribution Learning. IJCAI 2018, TKDE 2019

    - basic idea: leverage the topological information of the feature space and the correlation among the labels.
    - similarity matrix: $q_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \text{if } x_j \in K(i) \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$
    - Incorporate the constructed graph similarity matrix into the label space to learn a mapping model:
  - LEVI (LE via Variational Inference)

    Xu N, et al. Variational Label Enhancement. ICML 2020, TPAMI 2022

    - basic idea: consider the label distributions as the latent vectors and infer the label distributions from the logical labels using variational inference
    - overall objective: $$\begin{aligned} T(\boldsymbol{\vartheta}, \boldsymbol{\eta}, \boldsymbol{w}) = & \frac{1}{L}\sum_{m=1}^{L}\frac{1}{2}\|\boldsymbol{z} - \boldsymbol{\rho}^{(m)}\|_2^2 + \lambda\|\boldsymbol{d}^{(m)} - \boldsymbol{l}\|_2^2 \\ & + \sum_{i=1}^{c} l_i \log \tau_i^{(m)} + (1 - l_i)\cdot\log\left(1 - \tau_i^{(m)}\right) \\ & + \frac{1}{2}\{\mathrm{tr}(\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top\boldsymbol{\mu} - k - \log\det(\boldsymbol{\Sigma})\} \end{aligned}$$

Challenges of LE Task

- applications : medical diagnosis、image recognition、…

- goal of LE : exactly recover label distribution

- sample correlations
- label relevant information

- solutions:
  - sample correlations
  - local to global

- solutions:
  - label information bottleneck

applications in practice

abstract goals and key issues

motivations and solutions

Outline

1. Background: Label Distribution Learning and Label Enhancement

2. Overview of Label Enhancement

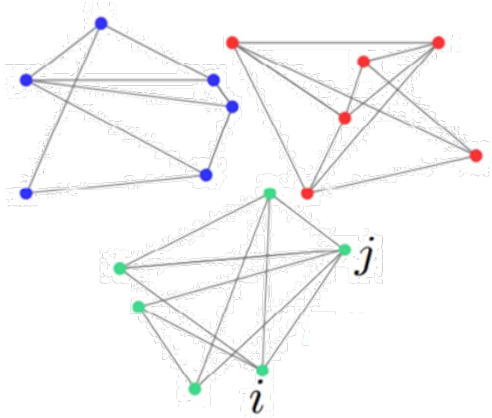
3. Label Enhancement via Global Sample Correlation

4. Label Enhancement via Label Information Bottleneck

5. Discussions and Further Works

LE via Global Sample Correlation

- sample correlations:
  - smoothness assumption
    - the topological structure of the feature space can be transferred to the label space
  - local sample correlations
    - the similarity graph is adopted to achieve local sample correlations

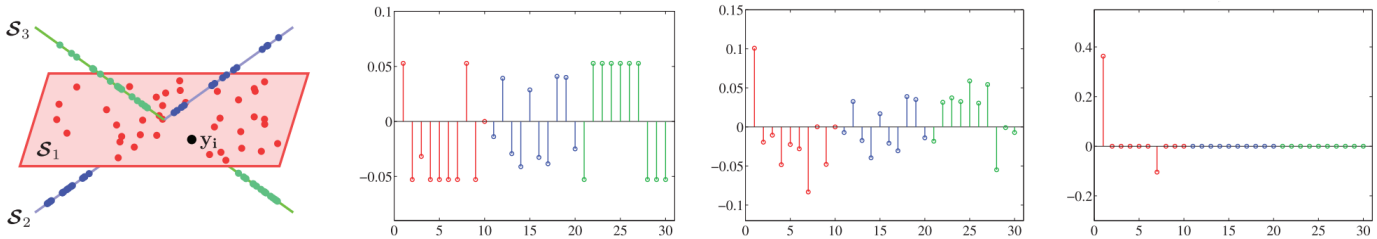


$$q_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), & \text{if } x_j \in K(i) \\ 0, & \text{otherwise} \end{cases}$$

    - limitation: merely consider the local information, the global information is ignored to construct the sample correlations.

LE via Global Sample Correlation

- sample correlations:
  - motivations
    - † from local to global
    - † how to get global sample correlations?
  - global sample correlations: subspace representation
    - † self-expressiveness property: each data point in a union of subspaces can be efficiently reconstructed by a combination of other points

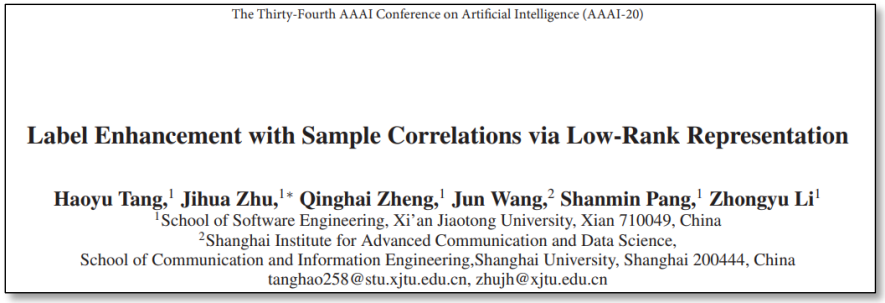


    - † the low-rank subspace representation is adopted:

$$\min_{C,E} \operatorname{rank}(C) + \lambda_2 \|E\|_l, s.t., X = XC + E$$

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - framework
    - global correlations from the features to the recovered the label distributions


The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Label Enhancement with Sample Correlations via Low-Rank Representation

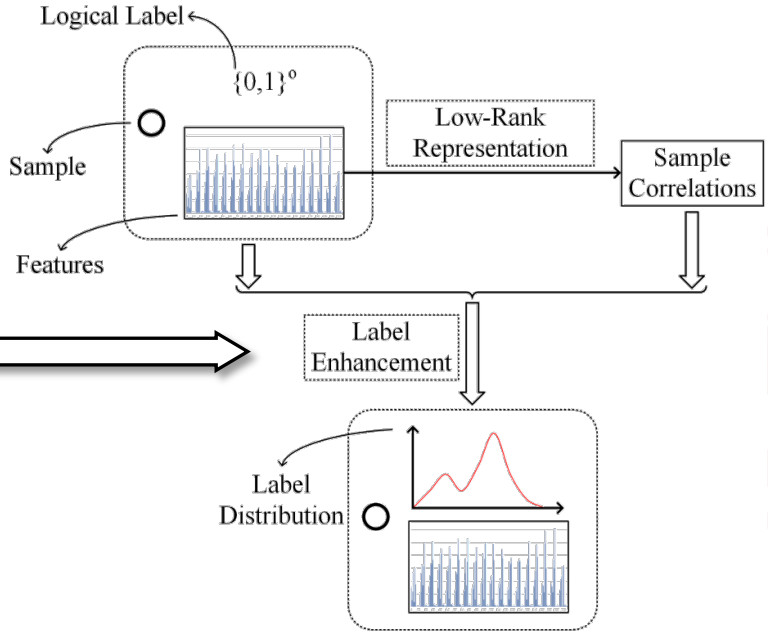
Haoyu Tang,[1] Jihua Zhu,[1*] Qinghai Zheng,[1] Jun Wang,[2] Shanmin Pang,[1] Zhongyu Li[1]
[1]School of Software Engineering, Xi'an Jiaotong University, Xian 710049, China
[2]Shanghai Institute for Advanced Communication and Data Science,
School of Communication and Information Engineering,Shanghai University, Shanghai 200444, China
tanghao258@stu.xjtu.edu.cn, zhujh@xjtu.edu.cn

  - objective
    - step 1

get global sample correlations

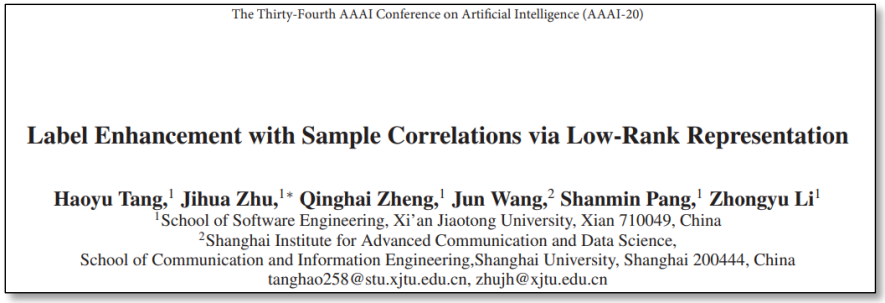
    - step 2

use global sample correlations to guide the recovery process



LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - framework
    - global correlations from the features to the recovered the label distributions


The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Label Enhancement with Sample Correlations via Low-Rank Representation

Haoyu Tang,[1] Jihua Zhu,[1*] Qinghai Zheng,[1] Jun Wang,[2] Shanmin Pang,[1] Zhongyu Li[1]
[1]School of Software Engineering, Xi'an Jiaotong University, Xian 710049, China
[2]Shanghai Institute for Advanced Communication and Data Science,
School of Communication and Information Engineering,Shanghai University, Shanghai 200444, China
tanghao258@stu.xjtu.edu.cn, zhujh@xjtu.edu.cn

  - objective
    - step 1

$$\min_{C,E} \operatorname{rank}(C) + \lambda_2 \|E\|_l, s.t., X = XC + E$$

    - step 2

$$\min_{\hat{\theta}} \mathcal{L}(\hat{\theta}) + \lambda_1 \Psi(\hat{\theta})$$

$$\mathcal{L}(\hat{\theta}) = \sum_{i=1}^{n} \left\| \phi\left(\hat{\theta}, \xi(x_i)\right) - L_i \right\|^2$$

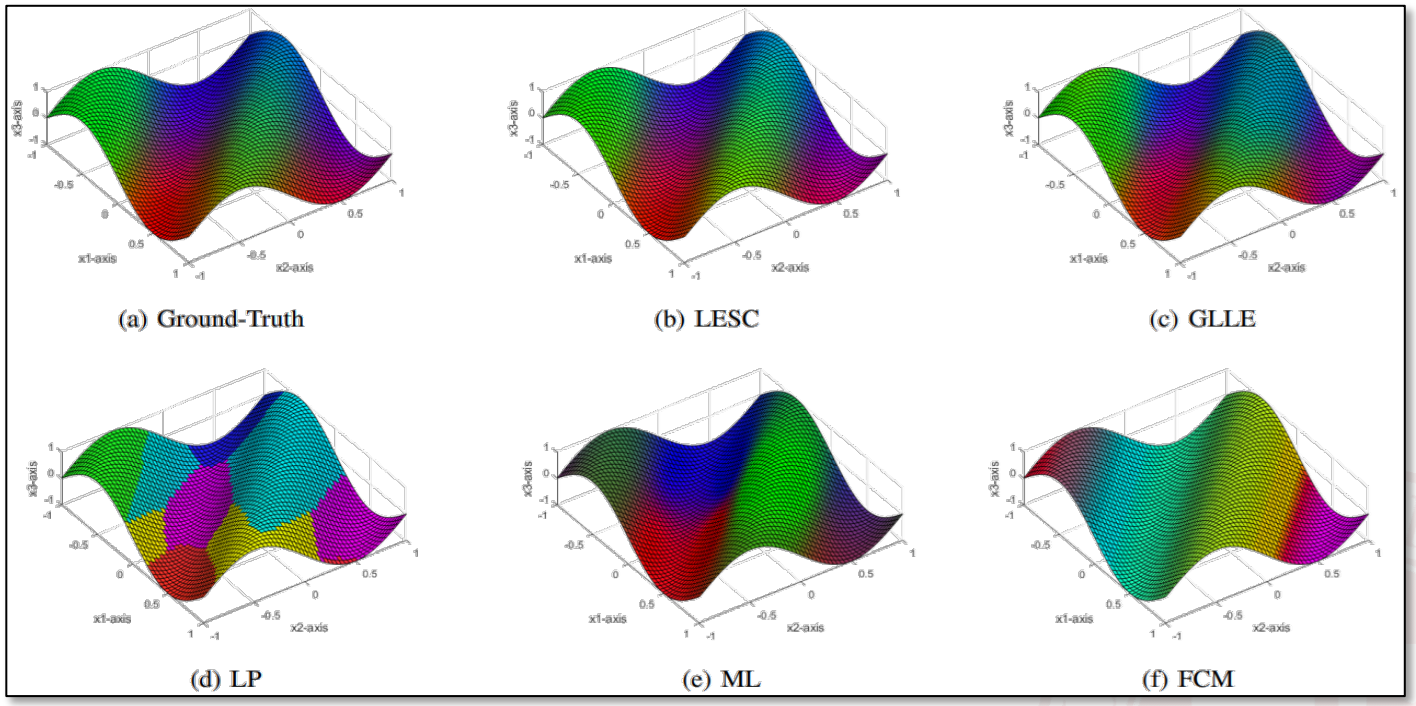
$$\Psi(\hat{\theta}) = \|\mathfrak{D} - \mathfrak{D}C\|_F^2 = \|(I - C^T)\mathfrak{D}^T\|_F^2$$

loss function between d and l

regularizer of smoothness

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - visualization experiments
    - a toy dataset is used



(a) Ground-Truth (b) LESC (c) GLLE

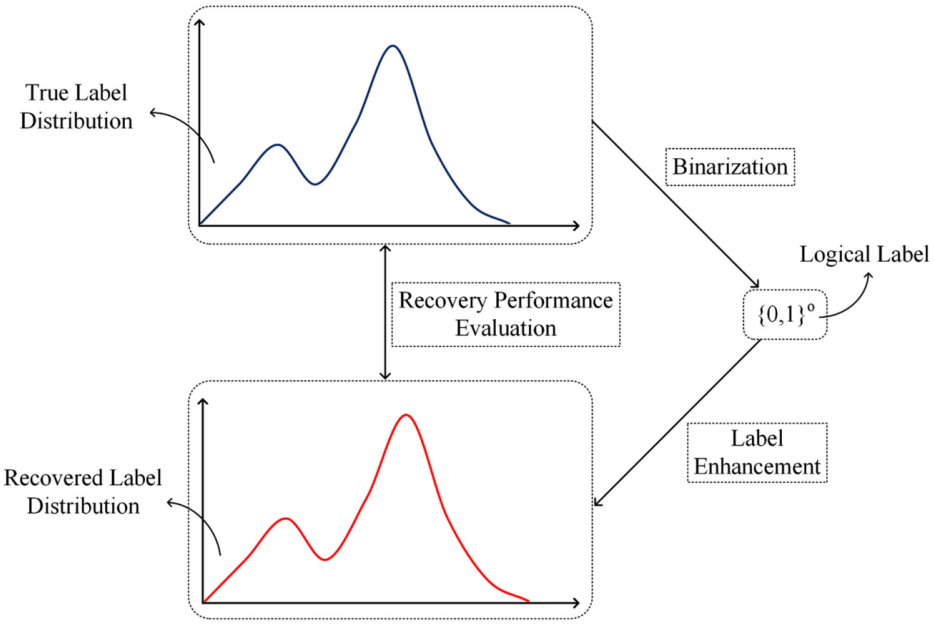
(d) LP (e) ML (f) FCM

    - the color patterns can be directly observed to compare both the ground truth and the recovered label distributions of different methods

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - comparison experiments
    - † real-life datasets are employed

Dataset	# Instances
Artificial	2601
Movie	7755
SBU_3DFE	2500
SJAFFE	213
Yeast-alpha	2465
Yeast-cdc	2465
Yeast-cold	2465
Yeast-diau	2465
Yeast-dtt	2465
Yeast-elu	2465
Yeast-heat	2465
Yeast-spo	2465
Yeast-spo5	2465
Yeast-spoem	2465



    - † 13 real-world datasets and one toy dataset are employed for the quantitative analysis

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - comparison experiments
    - † Cheb↓、Cosine↑ are employed as metrics

Table 3: Recovery Results (value(rank)) Measured by Cheb and Cosine.

Dataset	Measure Results by Cosine ↑					Measure Results by Cheb ↓				
	FCM	LP	ML	GLLE	LESC	FCM	LP	ML	GLLE	LESC
Artificial	0.933(4)	0.974(3)	0.925(5)	0.980(2)	**0.992(1)**	0.230(5)	0.130(3)	0.227(4)	0.108(2)	**0.057(1)**
Movie	0.773(5)	0.929(2)	0.919(3)	0.900(4)	**0.937(1)**	0.188(5)	0.161(3)	0.164(4)	0.160(2)	**0.121(1)**
SBU_3DFE	0.912(3)	0.922(2)	0.815(5)	0.900(4)	**0.932(1)**	0.135(3)	0.123(2)	0.233(5)	0.141(4)	**0.121(1)**
SJAFFE	0.906(4)	0.941(3)	0.857(5)	0.946(2)	**0.973(1)**	0.132(4)	0.107(3)	0.190(5)	0.100(2)	**0.069(1)**
Yeast-alpha	0.922(3)	0.911(4)	0.756(5)	0.973(2)	**0.992(1)**	0.044(4)	0.040(3)	0.057(5)	0.033(2)	**0.015(1)**
Yeast-cdc	0.929(3)	0.916(4)	0.759(5)	0.959(2)	**0.991(1)**	0.051(4)	0.042(3)	0.071(5)	0.038(2)	**0.019(1)**
Yeast-cold	0.922(4)	0.925(3)	0.784(5)	0.969(2)	**0.986(1)**	0.141(4)	0.137(3)	0.242(5)	0.093(2)	**0.056(1)**
Yeast-diau	0.882(4)	0.915(3)	0.803(5)	0.939(2)	**0.985(1)**	0.124(4)	0.099(3)	0.148(5)	0.084(2)	**0.042(1)**
Yeast-dtt	0.959(3)	0.921(4)	0.763(5)	0.983(2)	**0.991(1)**	0.097(3)	0.128(4)	0.244(5)	0.065(2)	**0.043(1)**
Yeast-elu	0.950(3)	0.918(4)	0.763(5)	0.978(2)	**0.991(1)**	0.052(4)	0.044(3)	0.072(5)	0.030(2)	**0.019(1)**
Yeast-heat	0.883(4)	0.932(3)	0.783(5)	0.980(2)	**0.986(1)**	0.169(5)	0.086(3)	0.165(4)	0.056(2)	**0.046(1)**
Yeast-spo	0.909(4)	0.939(3)	0.803(5)	0.968(2)	**0.975(1)**	0.130(4)	0.090(3)	0.171(5)	0.067(2)	**0.060(1)**
Yeast-spo5	0.922(4)	0.969(3)	0.884(5)	**0.974(1)**	**0.974(1)**	0.162(4)	0.114(3)	0.273(5)	**0.092(1)**	**0.092(1)**
Yeast-spoem	0.878(4)	0.950(3)	0.815(5)	0.968(2)	**0.978(1)**	0.233(4)	0.163(3)	0.400(5)	0.108(2)	**0.087(1)**
Avg.Rank	3.71	3.14	4.86	2.21	**1.00**	4.07	3.00	4.79	2.14	**1.00**

    - † LESC can achieve the best recovery results in all cases
    - † are the global correlations work? Yes! LESC can achieve more promising results than GLLE, which use local correlations.

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - parameter sensitivity
    - results *w.r.t.* different values of λ_1, λ_2

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.020	0.020	0.020	0.020	0.019	0.020	0.019	0.020
0.0010	0.020	0.020	0.020	0.020	0.020	0.019	0.020	0.020
0.0100	0.019	0.020	0.019	0.020	0.020	0.020	0.019	0.020
0.1000	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
1.0000	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
10.000	0.015	0.016	0.016	0.016	0.016	0.016	0.016	0.016
100.00	0.711	0.639	0.779	0.895	0.920	0.921	0.659	0.578
1000.0	0.929	0.685	0.938	0.928	0.846	0.942	0.635	0.837

Results in the metric of Cheb on Yeast-alpha

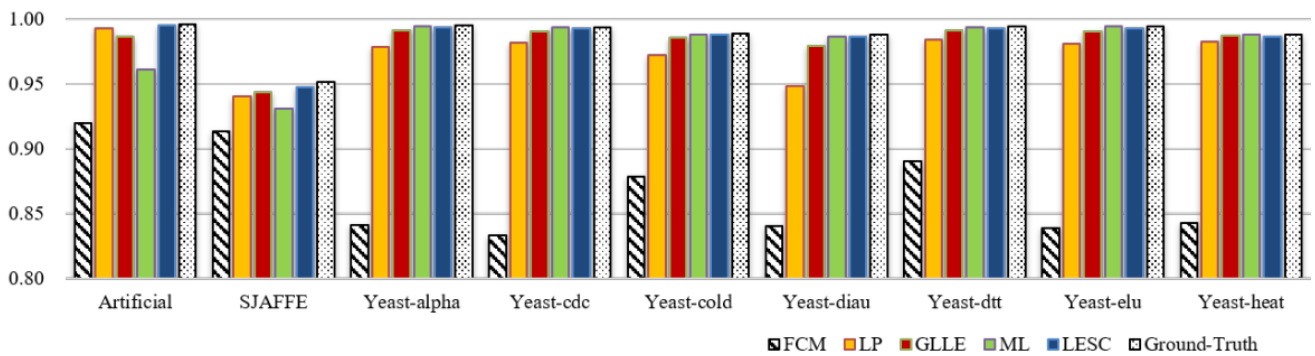
	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
0.0010	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
0.0100	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
0.1000	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988
1.0000	0.991	0.990	0.991	0.991	0.990	0.990	0.991	0.991
10.000	0.992	0.992	0.991	0.992	0.992	0.992	0.992	0.991
100.00	0.302	0.321	0.269	0.246	0.228	0.240	0.312	0.343
1000.0	0.238	0.311	0.237	0.233	0.265	0.234	0.311	0.265

Results in the metric of Cosine on Yeast-alpha

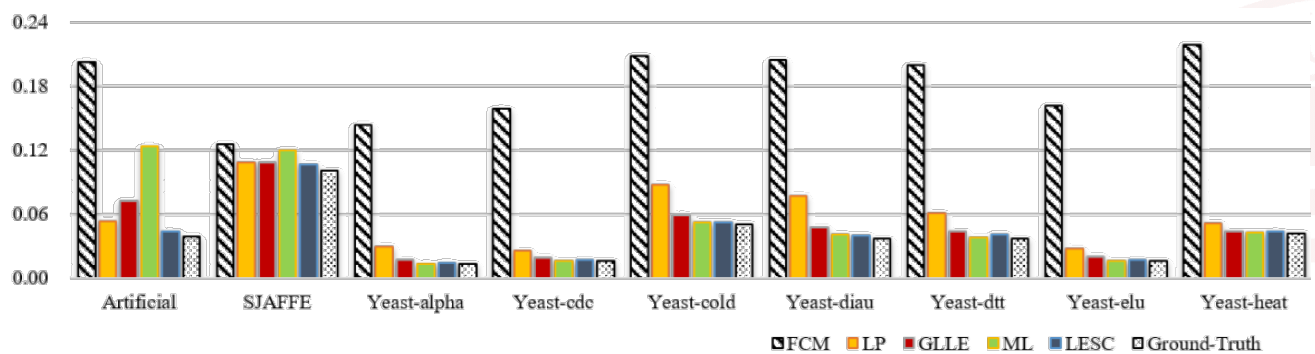
    - LESC is robust to λ_1, λ_2
    - promising recovery results can be achieved with a large range of λ_1, λ_2

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - performance of the recovered label distributions
    - † LDL results based on the recovered label distributions



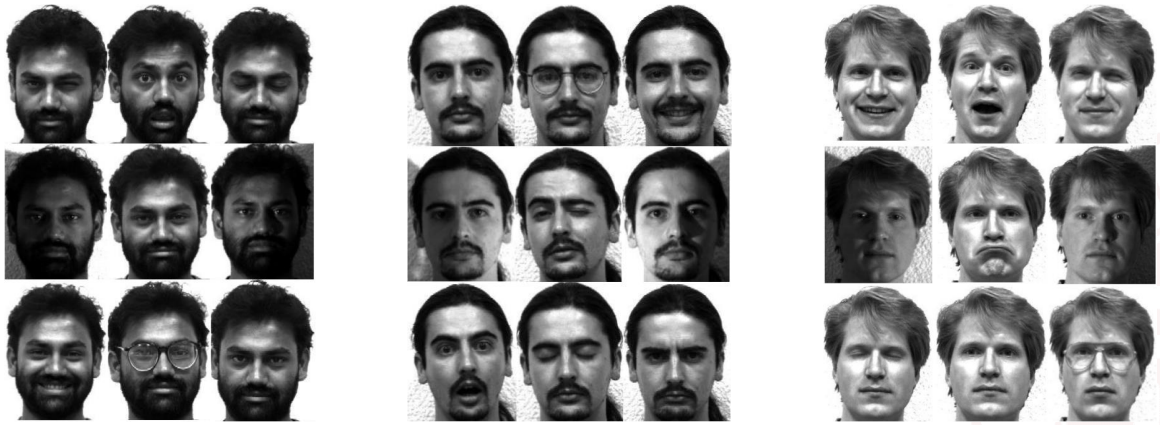
(a) Cosine Measure Result ↑



(b) Cheb Measure Result ↓

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - advantages of LESC
    - global sample correlations are employed during the recovery process
    - better recovery results can be achieved by LESC
  - limitations of LESC
    - **?** (consider the LE task for the following facial emotion images!)



How to get the correct sample correlations?

LE via Global Sample Correlation

- LESC (Label Enhancement with Sample Correlations):
  - advantages of LESC
    - † global sample correlations are employed during the recovery process
    - † better recovery results can be achieved by LESC
  - limitations of LESC
    - † **?** (consider the LE task for the following facial emotion images!)

$$X = [x_1; x_2; \dots; x_n]$$

$$\Downarrow$$

Sample correlations: 1、useful 2、useless	guide the construction of sample correlations $\longrightarrow$	Sample correlations: **1、useful** ~~2、useless~~

$$\Uparrow$$

$$\Gamma = [L_1; L_2; \dots; L_n]$$

How to get the correct sample correlations?

LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - seek more suitable global sample correlations for LE
    - use the ***low-rank tensor constraint***

IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 35, NO. 1

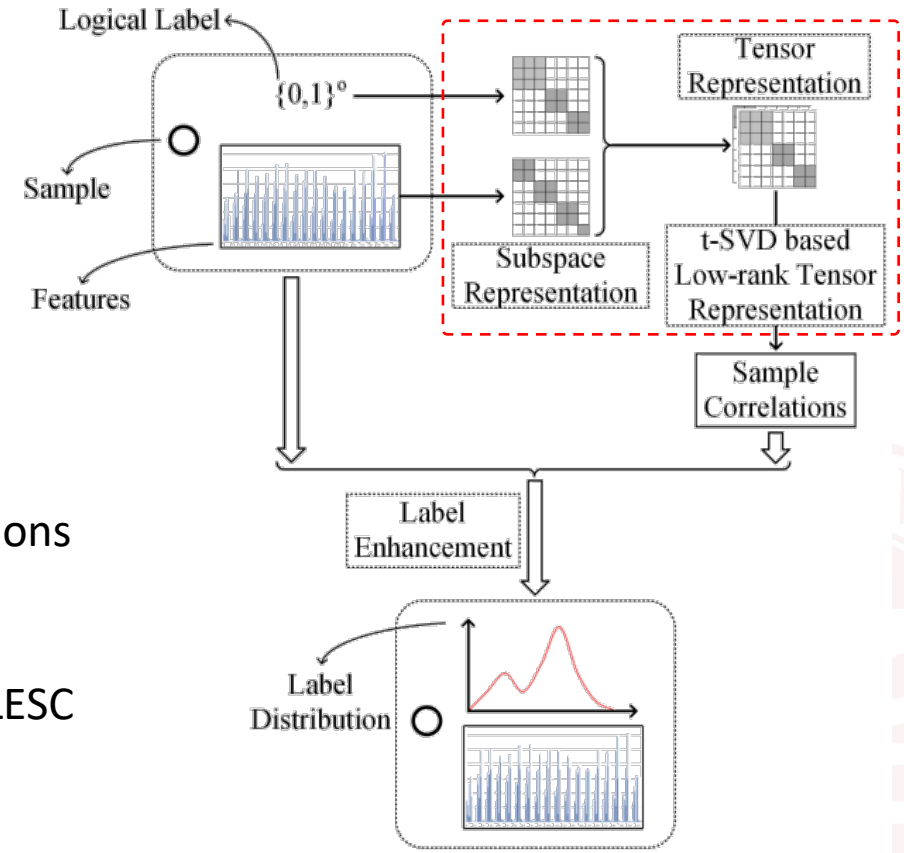
Generalized Label Enhancement With Sample Correlations

Qinghai Zheng, Jihua Zhu, Haoyu Tang, Xinyuan Liu, Zhongyu Li, and Huimin Lu

  - motivation
    - simultaneously utilize the sample correlations in the feature space and the logical label space
    - deal with features and logical labels jointly, the sample correlations in the logical label space act as a kind of tool, extracting the valuable information from the sample correlations in the feature space
    - the *t-SVD-based tensor unclear norm* is adopted for excavating the desired global sample correlations

LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - framework
    - obtain sample correlations from the feature space and the logical label space jointly
    - *low-rank tensor constraint*
  - comparison with LESC
    - mainly different *in the way of sample correlations' construction*
    - more exact sample correlations can be achieved for LE
    - the other components of gLESC are similar to LESC



LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - the t-SVD-based tensor nuclear norm is leveraged to achieve the construction of global sample correlations
    - objective

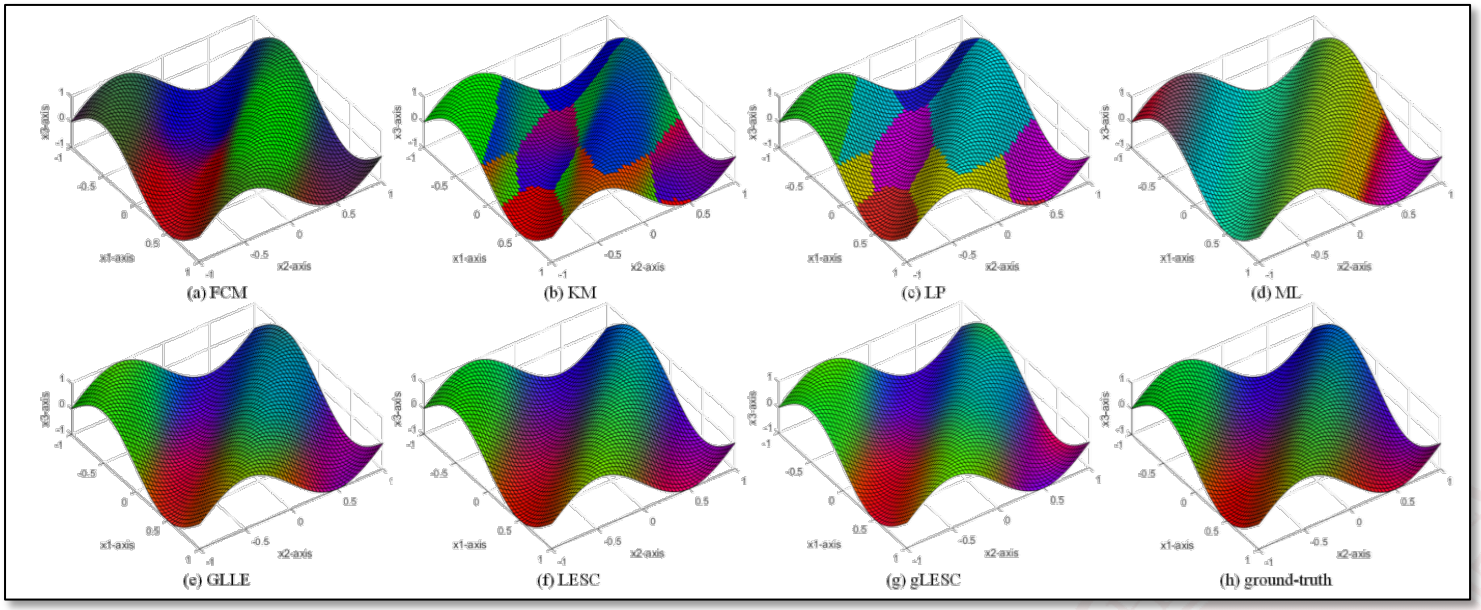
$$\min_{\mathcal{C},\mathcal{E}} \|\mathcal{C}\|_{\circledast} + \lambda_2 \|\mathcal{E}\|_{2,1},$$
$$s.t.\ X = XC^{(1)} + E^{(1)},$$
$$\Gamma = \Gamma C^{(2)} + E^{(2)},$$

    - optimization

$$\begin{aligned}
&\mathrm{ALM}(\mathcal{G},\mathcal{C},\mathcal{E}) = \|\mathcal{G}\|_{\circledast} + \lambda_2 \|\mathcal{E}\|_{2,1} \\
&+\left\langle Y_1, X - XC^{(1)} - E^{(1)} \right\rangle + \frac{\mu}{2}\left(\|X - XC^{(1)} - E^{(1)}\|_F^2\right) \\
&+\left\langle Y_2, \Gamma - \Gamma C^{(2)} - E^{(2)} \right\rangle + \frac{\mu}{2}\left(\|\Gamma - \Gamma C^{(2)} - E^{(2)}\|_F^2\right) \\
&+\langle \mathcal{W}, \mathcal{C} - \mathcal{G} \rangle + \frac{\rho}{2}\|\mathcal{C} - \mathcal{G}\|_F^2,
\end{aligned}$$

LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - visualization experiments
    - a toy dataset is used



    - the more promising recovery performance can be achieved by the proposed LESC and gLESC
    - gLESC outperforms LESC obviously

LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - comparison experiments
    - † more metrics and datasets are employed for comparison

Measure	Formula
Cheb ↓	$Dis_1(D,\hat{D}) = \max_j \left\lvert d^{y_j} - \hat{d}^{y_j} \right\rvert$
Canber ↓	$Dis_2(D,\hat{D}) = \sum_{j=1}^{o} \frac{\lvert d^{y_j} - \hat{d}^{y_j} \rvert}{d^{y_j} + \hat{d}^{y_j}}$
Clark ↓	$Dis_3(D,\hat{D}) = \sqrt{\sum_{j=1}^{o} \frac{(d^{y_j} - \hat{d}^{y_j})^2}{(d^{y_j} + \hat{d}^{y_j})^2}}$
KL ↓	$Dis_4(D,\hat{D}) = \sum_{j=1}^{o} d^{y_j} \ln \frac{d^{y_j}}{\hat{d}^{y_j}}$
Cosine ↑	$Sim_1(D,\hat{D}) = \frac{\sum_{j=1}^{o} d^{y_j} \hat{d}^{y_j}}{\sqrt{\sum_{j=1}^{o} (d^{y_j})^2} \sqrt{\sum_{j=1}^{o} (\hat{d}^{y_j})^2}}$
Intersec ↑	$Sim_2(D,\hat{D}) = \sum_{j=1}^{o} \min\left(d^{y_j}, \hat{d}^{y_j}\right)$

Dataset	# Instances
Artificial	2601
Movie	7755
SBU_3DFE	2500
SJAFFE	213
Yeast-alpha	2465
Yeast-cdc	2465
Yeast-cold	2465
Yeast-diau	2465
Yeast-dtt	2465
Yeast-elu	2465
Yeast-heat	2465
Yeast-spo	2465
Yeast-spo5	2465
Yeast-spoem	2465

LE via Global Sample Correlation

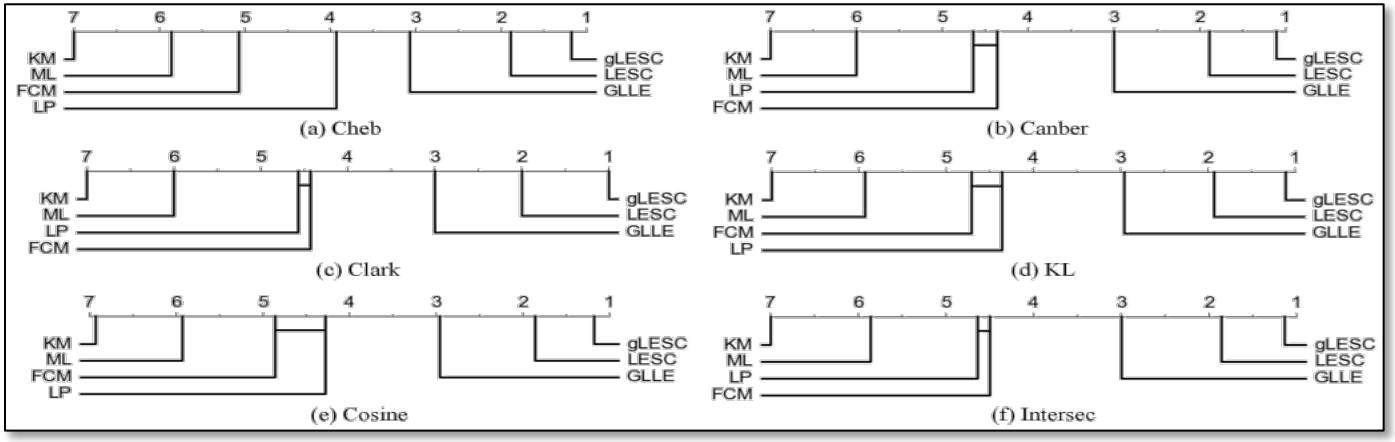
- gLESC (generalized Label Enhancement with Sample Correlations):
  - comparison experiments
    - results

Dataset	Measure Results by Cheb ↓							Measure Results by Canber ↓						
	FCM	KM	LP	ML	GLLE	LESC	gLESC	FCM	KM	LP	ML	GLLE	LESC	gLESC
Artificial	0.188(5)	0.260(7)	0.130(4)	0.227(6)	0.108(3)	0.057(2)	**0.055(1)**	0.797(5)	1.779(7)	0.668(4)	1.413(6)	0.617(3)	0.213(2)	**0.193(1)**
Movie	0.230(6)	0.234(7)	0.161(4)	0.164(5)	0.122(3)	0.121(2)	**0.120(1)**	1.664(4)	3.444(7)	1.720(5)	1.934(6)	1.045(3)	**1.034(1)**	**1.034(1)**
SBU_3DFE	0.135(5)	0.238(7)	0.123(2)	0.233(6)	0.126(4)	**0.122(1)**	0.125(3)	1.020(4)	4.121(7)	1.245(5)	4.001(6)	0.820(3)	**0.799(1)**	0.803(2)
SJAFFE	0.132(5)	0.214(7)	0.107(4)	0.186(6)	0.087(3)	0.069(2)	**0.067(1)**	1.081(5)	4.010(7)	1.064(4)	3.138(6)	0.781(3)	0.561(2)	**0.550(1)**
Yeast-alpha	0.044(5)	0.063(7)	0.040(4)	0.057(6)	0.020(3)	0.015(2)	**0.014(1)**	2.883(4)	11.809(7)	4.544(5)	11.603(6)	1.134(3)	0.846(2)	**0.761(1)**
Yeast-cdc	0.051(5)	0.076(7)	0.042(4)	0.071(6)	0.022(3)	0.019(2)	**0.017(1)**	2.415(4)	9.875(7)	3.644(5)	9.695(6)	0.959(3)	0.765(2)	**0.695(1)**
Yeast-cold	0.141(5)	0.252(7)	0.137(4)	0.242(6)	0.066(3)	0.056(2)	**0.052(1)**	0.734(4)	2.566(7)	0.924(5)	2.519(6)	0.305(3)	0.263(2)	**0.242(1)**
Yeast-diau	0.124(5)	0.152(7)	0.099(4)	0.148(6)	0.053(3)	0.042(2)	**0.039(1)**	1.895(4)	4.261(7)	1.748(5)	4.180(6)	0.671(3)	0.480(2)	**0.452(1)**
Yeast-dtt	0.097(4)	0.257(7)	0.128(5)	0.244(6)	0.052(3)	0.043(2)	**0.037(1)**	0.501(4)	2.594(7)	0.941(5)	2.549(6)	0.248(3)	0.206(2)	**0.175(1)**
Yeast-elu	0.052(5)	0.078(7)	0.044(4)	0.072(6)	0.023(3)	0.019(2)	**0.017(1)**	1.689(4)	9.110(7)	3.381(5)	8.949(6)	0.902(3)	0.727(2)	**0.628(1)**
Yeast-heat	0.169(6)	0.175(7)	0.086(4)	0.165(5)	0.049(3)	0.046(2)	**0.043(1)**	1.157(4)	3.849(7)	1.293(5)	3.779(6)	0.430(3)	0.401(2)	**0.372(1)**
Yeast-spo	0.130(5)	0.175(7)	0.090(4)	0.171(6)	0.062(3)	0.060(2)	**0.059(1)**	0.998(4)	3.854(7)	1.231(5)	3.772(6)	0.548(3)	0.533(2)	**0.521(1)**
Yeast-spo5	0.162(5)	0.277(7)	0.114(4)	0.273(6)	0.099(3)	**0.092(1)**	**0.092(1)**	0.563(5)	1.382(7)	0.401(4)	1.355(6)	0.305(3)	0.284(2)	**0.283(1)**
Yeast-spoem	0.233(5)	0.408(7)	0.163(4)	0.403(6)	0.088(3)	0.087(2)	**0.084(1)**	0.534(5)	1.253(7)	0.365(4)	1.226(6)	0.183(3)	0.180(2)	**0.175(1)**
Avg.Rank	5.07	7.00	3.93	5.86	3.07	1.86	**1.14**	4.27	7.00	4.71	6.00	3.00	1.86	**1.07**

Dataset	Measure Results by Cosine ↑							Measure Results by Intersec ↑						
	FCM	KM	LP	ML	GLLE	LESC	gLESC	FCM	KM	LP	ML	GLLE	LESC	gLESC
Artificial	0.933(5)	0.918(7)	0.974(4)	0.925(6)	0.980(3)	**0.992(1)**	0.991(2)	0.812(5)	0.740(7)	0.870(4)	0.773(6)	0.892(3)	0.943(2)	**0.945(1)**
Movie	0.773(7)	0.880(6)	0.929(4)	0.919(5)	0.936(3)	0.937(2)	**0.938(1)**	0.677(6)	0.649(7)	0.778(5)	0.779(4)	0.831(3)	**0.833(1)**	**0.833(1)**
SBU_3DFE	0.912(5)	0.812(7)	0.922(4)	0.815(6)	0.927(3)	**0.932(1)**	**0.931(2)**	0.827(4)	0.579(7)	0.810(5)	0.587(6)	0.850(3)	**0.855(1)**	0.854(2)
SJAFFE	0.906(5)	0.827(7)	0.941(4)	0.857(6)	0.958(3)	0.973(2)	**0.975(1)**	0.821(5)	0.593(7)	0.837(4)	0.661(6)	0.872(3)	0.905(2)	**0.908(1)**
Yeast-alpha	0.922(4)	0.751(7)	0.911(5)	0.756(6)	0.987(3)	0.992(2)	**0.994(1)**	0.844(4)	0.532(7)	0.774(5)	0.537(6)	0.938(3)	0.953(2)	**0.958(1)**
Yeast-cdc	0.929(4)	0.754(7)	0.916(5)	0.759(6)	0.987(3)	0.991(2)	**0.992(1)**	0.847(4)	0.533(7)	0.779(5)	0.538(6)	0.937(3)	0.950(2)	**0.954(1)**
Yeast-cold	0.922(5)	0.779(7)	0.925(4)	0.784(6)	0.982(3)	0.986(2)	**0.988(1)**	0.833(4)	0.559(7)	0.794(5)	0.565(6)	0.924(3)	0.935(2)	**0.940(1)**
Yeast-diau	0.882(5)	0.799(7)	0.915(4)	0.803(6)	0.975(3)	0.985(2)	**0.987(1)**	0.760(5)	0.588(7)	0.788(4)	0.593(6)	0.906(3)	0.933(2)	**0.937(1)**
Yeast-dtt	0.959(4)	0.759(7)	0.921(5)	0.763(6)	0.988(3)	0.991(2)	**0.994(1)**	0.894(4)	0.541(7)	0.786(5)	0.546(6)	0.939(3)	0.949(2)	**0.957(1)**
Yeast-elu	0.950(4)	0.758(7)	0.918(5)	0.763(6)	0.987(3)	0.991(2)	**0.993(1)**	0.883(4)	0.539(7)	0.782(5)	0.544(6)	0.936(3)	0.949(2)	**0.956(1)**
Yeast-heat	0.883(5)	0.779(7)	0.932(4)	0.783(6)	0.984(3)	0.986(2)	**0.987(1)**	0.807(4)	0.559(7)	0.805(5)	0.564(6)	0.929(3)	0.934(2)	**0.939(1)**
Yeast-spo	0.909(5)	0.800(7)	0.939(4)	0.803(6)	0.974(3)	0.975(2)	**0.976(1)**	0.836(4)	0.575(7)	0.819(5)	0.580(6)	0.909(3)	0.912(2)	**0.914(1)**
Yeast-spo5	0.922(5)	0.882(7)	0.969(4)	0.884(6)	0.971(3)	**0.974(1)**	**0.974(1)**	0.838(5)	0.724(7)	0.886(4)	0.727(6)	0.901(3)	**0.908(1)**	**0.908(1)**
Yeast-spoem	0.878(5)	0.812(7)	0.950(4)	0.815(6)	0.978(2)	0.978(2)	**0.979(1)**	0.767(5)	0.592(7)	0.837(4)	0.597(6)	0.912(3)	0.913(2)	**0.916(1)**
Avg.Rank	4.86	6.93	4.29	5.93	2.93	1.79	**1.14**	4.50	7.00	4.64	5.86	3.00	2.13	**1.07**

LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - model discussions
    - statistic analysis



Dataset	Measure Results by Cheb						Measure Results by Canber					
	GLLE and LESC		GLLE and gLESC		LESC and gLESC		GLLE and LESC		GLLE and gLESC		LESC and gLESC	
	t-value	*p*	*t*-value	*p*	*t*-value	*p*	*t*-value	*p*	*t*-value	*p*	*t*-value	*p*
Movie	2.8896	0.0179	3.7356	0.0047	10.1806	3.08E-06	4.1985	0.0023	4.9693	7.71E-04	14.1472	1.87E-07
SBU_3DFE	6.0865	1.82E-04	7.5646	3.45E-05	-0.906	0.3882	6.0392	1.93E-04	4.8465	9.13E-04	-2.1394	0.061
SJAFFE	12.9504	4.01E-07	8.9985	8.55E-06	70.0691	0.9464	16.8084	4.18E-08	10.0808	3.35E-06	1.6848	10.1263
Yeast-alpha	3.3151	0.009	4.5912	0.0013	5.3743	4.48E-04	4.9394	8.03E-04	6.1692	1.65E-04	12.2863	6.30E-07
Yeast-cdc	12.803	4.43E-07	46.4826	4.94E-12	1.63	0.1375	21.9595	3.98E-09	91.6064	1.12E-14	4.2433	0.0022
Yeast-cold	18.0582	2.23E-08	11.3127	1.27E-06	1.7457	0.1148	27.6061	5.21E-10	9.0292	8.31E-06	1.6152	0.1407
Yeast-diau	18.6861	1.65E-08	25.0726	1.23E-09	4.9986	7.40E-04	30.2152	2.33E-10	35.6724	5.29E-11	8.5495	1.30E-05
Yeast-dtt	9.6781	4.70E-06	24.3695	1.58E-09	16.6552	4.53E-08	7.6114	3.29E-05	24.5951	1.46E-09	15.0992	1.07E-07
Yeast-elu	23.7774	1.96E-09	28.7222	3.66E-10	8.4796	1.39E-05	32.2289	1.31E-10	54.6123	1.16E-12	16.6537	4.53E-08
Yeast-heat	5.1934	5.69E-04	9.7099	4.57E-06	5.5428	3.60E-04	6.5476	1.05E-04	10.8416	1.82E-06	10.7182	2.00E-06
Yeast-spo	1.5796	0.1487	2.0551	0.07	0.132	0.8979	2.458	0.0363	3.6905	0.005	0.6831	0.5117
Yeast-spo5	4.4747	0.0015	5.1503	6.03E-04	2.0249	0.0735	3.9804	0.0032	4.8572	8.99E-04	2.1382	0.0612
Yeast-spoem	1.5213	0.1625	2.2097	0.0545	0.9828	0.3514	1.2646	0.2378	1.9638	0.0811	1.0861	0.3057

LE via Global Sample Correlation

- gLESC (generalized Label Enhancement with Sample Correlations):
  - model discussions
    - parameter sensitivity

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.123	0.123	0.123	0.123	0.123	0.123	0.124	0.124
0.0010	0.124	0.123	0.122	0.124	0.123	0.123	0.122	0.124
0.0100	0.123	0.124	0.123	0.123	0.123	0.123	0.123	0.123
0.1000	0.123	0.123	0.124	0.123	0.123	0.123	0.123	0.123
1.0000	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
10.000	0.133	0.134	0.134	0.134	0.133	0.133	0.134	0.133
100.00	0.139	0.139	0.138	0.138	0.138	0.138	0.138	0.139
1000.0	0.652	0.844	0.812	0.716	0.748	0.593	0.696	0.778

(a) Results in the metric of Cheb on SBU_3DFE

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.020	0.020	0.020	0.019	0.020	0.020	0.020	0.020
0.0010	0.020	0.020	0.020	0.020	0.019	0.019	0.020	0.019
0.0100	0.020	0.020	0.019	0.020	0.019	0.019	0.020	0.020
0.1000	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019
1.0000	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
10.000	0.015	0.015	0.014	0.014	0.014	0.014	0.015	0.015
100.00	0.893	0.015	0.014	0.015	0.014	0.014	0.015	0.015
1000.0	0.561	0.872	0.892	0.906	0.629	0.920	0.816	0.643

(c) Results in the metric of Cheb on Yeast-alpha

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.066	0.066	0.066	0.067	0.066	0.066	0.066	0.066
0.0010	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.065
0.0100	0.066	0.066	0.066	0.066	0.065	0.066	0.066	0.065
0.1000	0.064	0.064	0.064	0.065	0.065	0.065	0.065	0.065
1.0000	0.058	0.058	0.058	0.058	0.058	0.058	0.058	0.058
10.000	0.052	0.052	0.052	0.052	0.053	0.052	0.052	0.052
100.00	0.055	0.055	0.055	0.055	0.054	0.055	0.055	0.055
1000.0	0.619	0.507	0.581	0.741	0.755	0.589	0.750	0.759

(e) Results in the metric of Cheb on Yeast-cold

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.930	0.930	0.930	0.930	0.930	0.930	0.929	0.929
0.0010	0.929	0.931	0.931	0.929	0.930	0.930	0.931	0.930
0.0100	0.930	0.929	0.930	0.930	0.929	0.931	0.931	0.930
0.1000	0.931	0.931	0.929	0.930	0.931	0.930	0.930	0.931
1.0000	0.930	0.931	0.931	0.930	0.931	0.930	0.930	0.931
10.000	0.924	0.924	0.923	0.923	0.924	0.924	0.924	0.925
100.00	0.917	0.917	0.918	0.918	0.918	0.918	0.919	0.918
1000.0	0.450	0.355	0.370	0.433	0.400	0.484	0.424	0.385

(b) Results in the metric of Cosine on SBU_3DFE

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
0.0010	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
0.0100	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
0.1000	0.987	0.988	0.987	0.987	0.988	0.987	0.988	0.987
1.0000	0.991	0.990	0.990	0.990	0.990	0.990	0.990	0.990
10.000	0.993	0.993	0.994	0.994	0.993	0.994	0.993	0.993
100.00	0.248	0.993	0.993	0.993	0.993	0.993	0.993	0.993
1000.0	0.343	0.259	0.241	0.251	0.336	0.240	0.266	0.322

(d) Results in the metric of Cosine on Yeast-alpha

	0.0001	0.0010	0.0100	0.1000	1.0000	10.000	100.00	1000.0
0.0001	0.982	0.982	0.982	0.982	0.982	0.982	0.982	0.982
0.0010	0.982	0.982	0.982	0.982	0.982	0.982	0.982	0.982
0.0100	0.982	0.982	0.982	0.982	0.982	0.982	0.982	0.982
0.1000	0.983	0.983	0.983	0.983	0.983	0.983	0.983	0.983
1.0000	0.986	0.986	0.986	0.986	0.986	0.986	0.986	0.986
10.000	0.988	0.988	0.988	0.988	0.988	0.988	0.988	0.988
100.00	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987
1000.0	0.563	0.630	0.567	0.492	0.482	0.568	0.484	0.476

(f) Results in the metric of Cosine on Yeast-cold

Outline

1. Background: Label Distribution Learning and Label Enhancement

2. Overview of Label Enhancement

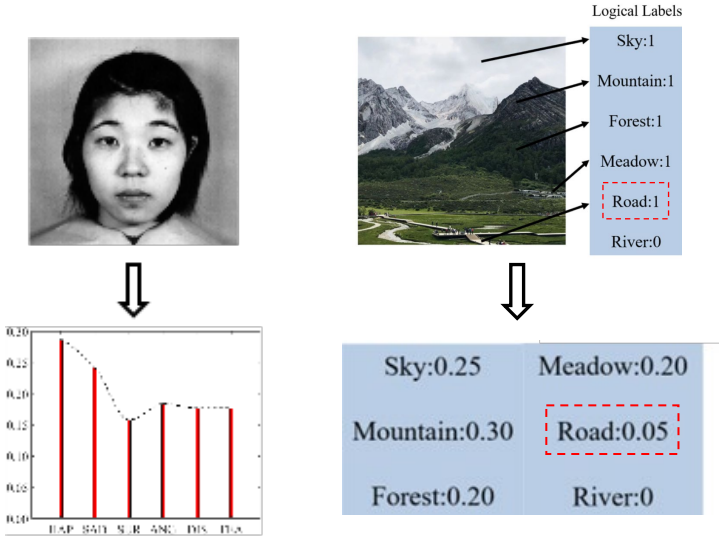
3. Label Enhancement via Global Sample Correlation

4. Label Enhancement via Label Information Bottleneck

5. Discussions and Further Works

LE via Label Information Bottleneck

- sample correlations only work on certain scenarios
  - examples
    - † facial expression analysis (?)
    - † landscape Image recognition (✘)
  - how to address the samples, the sample correlations of which are inconspicuous?
    - † a new framework is required



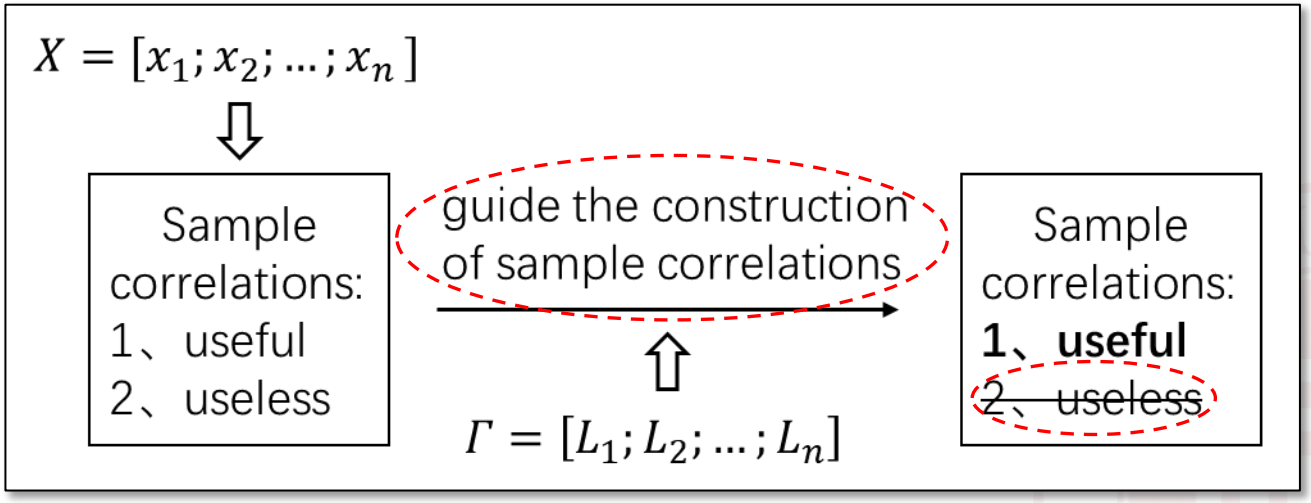
- the common fault (limitation) existing in all existing LE methods
  - the following extra constraint is required:
    - † the recovered label distributions should be close to existing logical labels, *i.e.*, $\|\boldsymbol{d}-\boldsymbol{l}\|_2^2$.

$$\min_{\theta} \|f_\theta(\boldsymbol{X})-\boldsymbol{L}\|_F^2 + \gamma reg(f_\theta(\boldsymbol{X}))$$

    - † it is unreasonable in many cases
    - † a new framework is required

LE via Label Information Bottleneck

- motivations
  - avoid deal with sample correlations and the extra constraint
    - † new idea and framework for LE
    - † think in different way while keeping the advantages of the previous strategies for LE
    - † let us rethink the learning strategy introduced in gLESC



sample correlations → label relevant information

LE via Label Information Bottleneck

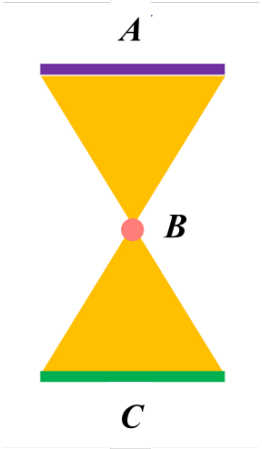
- motivations
  - avoid deal with sample correlations and the extra constraint
    - † new idea and framework for LE
    - † think in different way while keeping the advantages of the previous strategies for LE
    - † let us rethink the learning strategy introduced in gLESC

Label Information Bottleneck for Label Enhancement

Qinghai Zheng[1], Jihua Zhu[2]*, Haoyu Tang[3]
[1]College of Computer and Data Science, Fuzhou University, China
[2]School of Software Engineering, Xi'an Jiaotong University, Xi'an, China
[3]School of Software, Shandong University, Jinan, China

LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - deal with the LE from the perspective of information theory
  - decompose the label relevant information into:
    - † the information about the assignments of labels to instance
    - † the information about the label gaps between logical labels and distribution labels
  - briefly introduction to information bottleneck



✓ objective of information bottleneck

$$\min_{\boldsymbol{B}} \; -I(\boldsymbol{B},\boldsymbol{C}), \text{ s.t.}, I(\boldsymbol{A},\boldsymbol{B}) \leqslant I_c$$

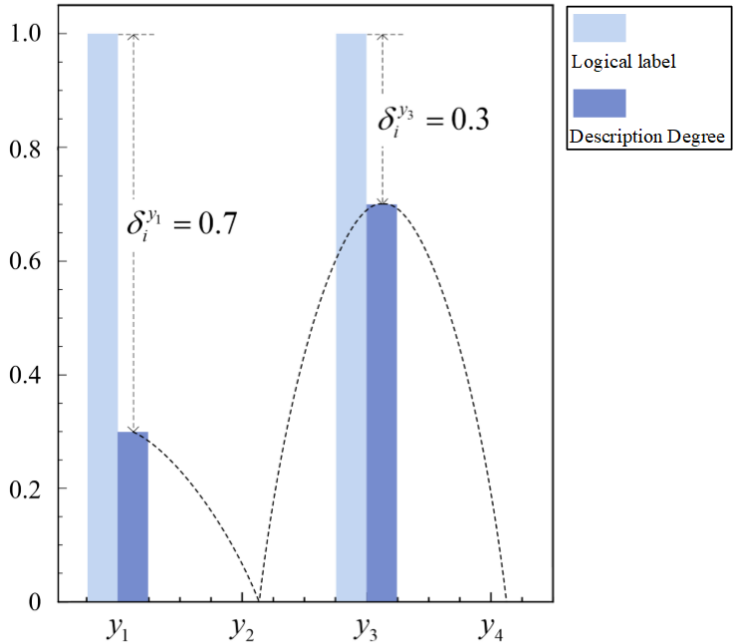
✓ classical works of information bottleneck

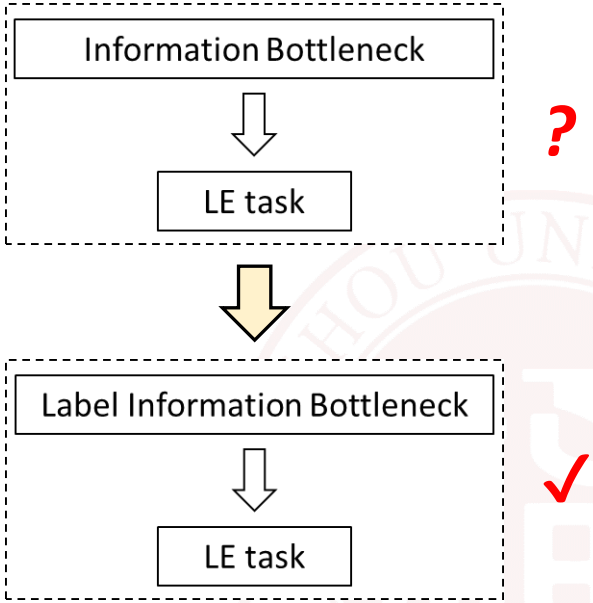
‡ Tishby N, et. al. The information bottleneck method. ArXiv preprint physics/0004057, 2000.

‡ Alemi A A, et. al. Deep Variational Information Bottleneck. ICLR 2017

LE via Label Information Bottleneck

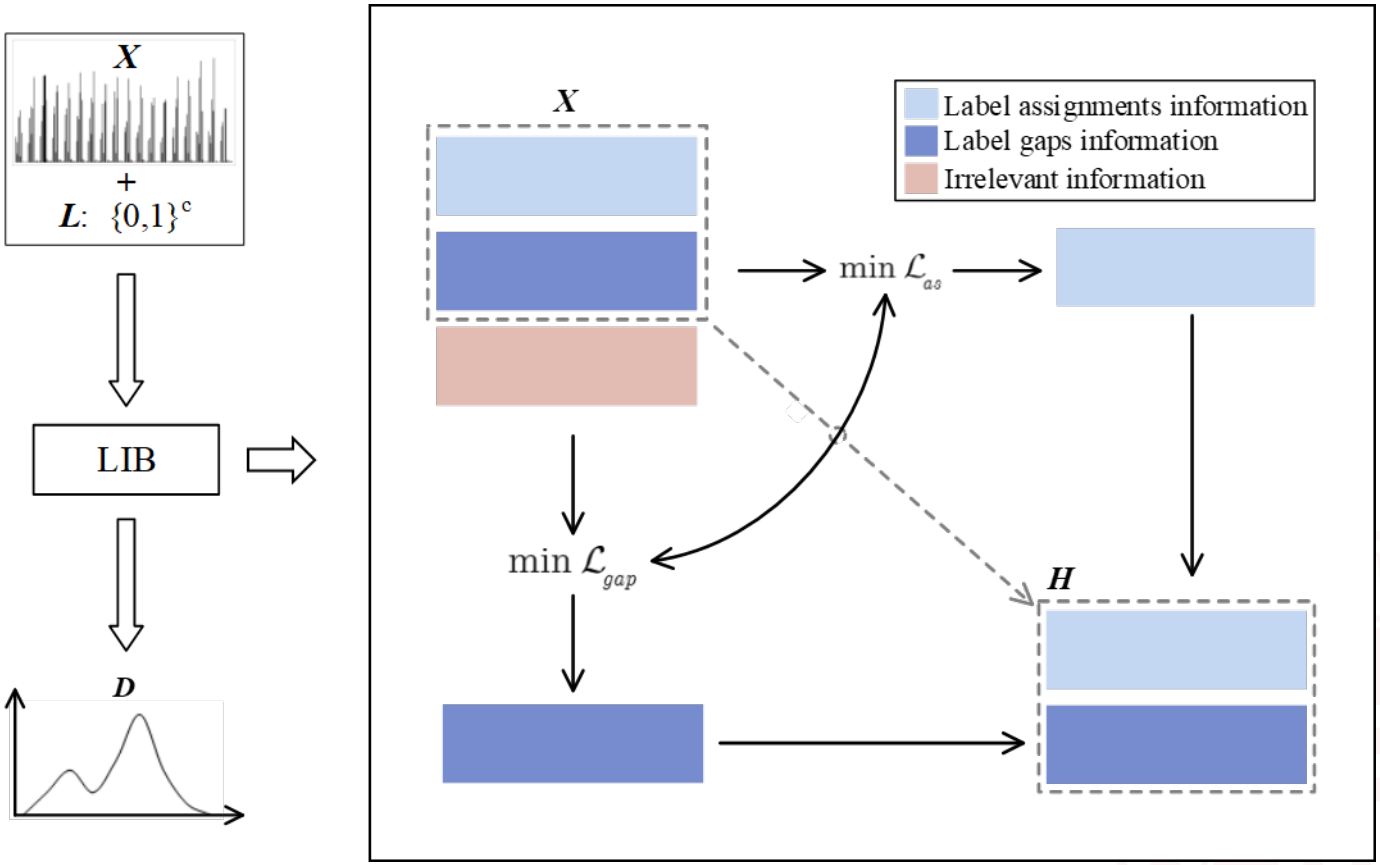
- LIB (Label Information Bottleneck)
  - deal with the LE from the perspective of information theory
  - decompose the label relevant information into:
    - the information about the assignments of labels to instance
    - the information about the label gaps between logical labels and distribution labels





LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - framework



LE via Label Information Bottleneck

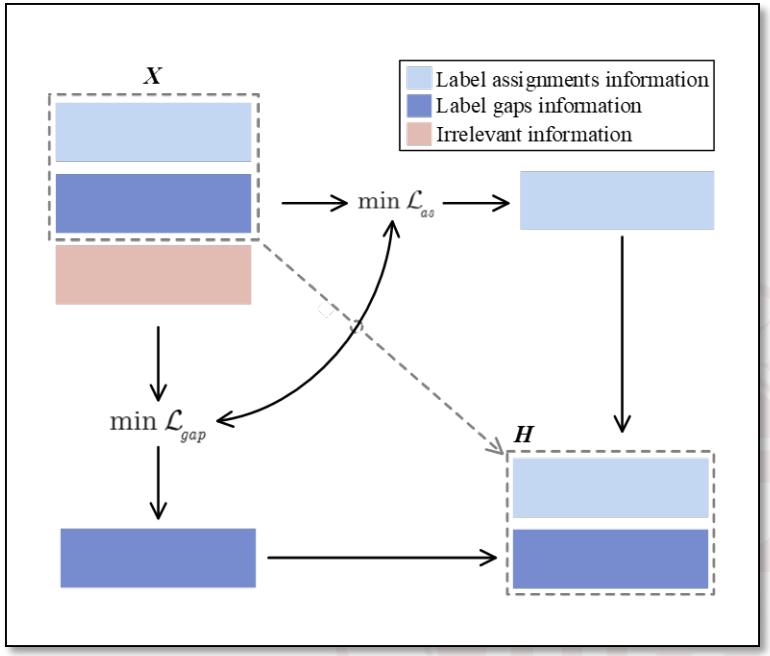
- LIB (Label Information Bottleneck)
  - we formulates the LE problem as the following two joint processes
    - † Learn the representation with the label relevant information
    - † Recover label distributions based on the learned representation
  - objective

$$\min_{\boldsymbol{H}} \ \mathcal{L}_{as} + \alpha \mathcal{L}_{gap}, \ \text{s.t.}, I(\boldsymbol{X}, \boldsymbol{H}) \leqslant I_c$$

† $\mathcal{L}_{as}$: the label assignments information modeling for LE

† $\mathcal{L}_{gap}$: the label gaps information modeling for LE

† $I(\boldsymbol{X}, \boldsymbol{H}) \leqslant I_c$: the label irrelevant information modeling



LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - the label assignments information modeling
    - † $\mathcal{L}_{as}$: $\mathcal{L}_{as} = -I(\boldsymbol{H}, \boldsymbol{L}) \Rightarrow \mathcal{L}_{as} = -\sum_{\boldsymbol{h}} \sum_{\boldsymbol{l}} p(\boldsymbol{h}, \boldsymbol{l}) \log \frac{p(\boldsymbol{l}|\boldsymbol{h})}{p(\boldsymbol{l})}$

$$\mathrm{KL}(p(\boldsymbol{l}|\boldsymbol{h})||q(\boldsymbol{l}|\boldsymbol{h})) = \sum_{\boldsymbol{l}} p(\boldsymbol{l}|\boldsymbol{h}) \log \frac{p(\boldsymbol{l}|\boldsymbol{h})}{q(\boldsymbol{l}|\boldsymbol{h})} \geqslant 0$$

$$\Rightarrow \sum_{\boldsymbol{l}} p(\boldsymbol{l}|\boldsymbol{h}) \log p(\boldsymbol{l}|\boldsymbol{h}) \geqslant \sum_{\boldsymbol{l}} p(\boldsymbol{l}|\boldsymbol{h}) \log q(\boldsymbol{l}|\boldsymbol{h}),$$

$$\mathbb{E}_{p(\boldsymbol{l})}[-\log p(\boldsymbol{l})] = -\sum_{\boldsymbol{l}} p(\boldsymbol{l}) \log p(\boldsymbol{l}) \geqslant 0,$$

- ✓ KL divergence and the entropy are positive
- ✓ Markov chain: L ← X → H

$$\mathcal{L}_{as} \leqslant -\sum_{\boldsymbol{x}} \sum_{\boldsymbol{l}} \sum_{\boldsymbol{h}} p(\boldsymbol{x}, \boldsymbol{l}) p(\boldsymbol{h}|\boldsymbol{x}) \log q(\boldsymbol{l}|\boldsymbol{h})$$

- † it can be seen that $\mathcal{L}_{as}$ excavates the information about the assignments of labels to the instance

LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - the label gaps information modeling
    - $\mathcal{L}_{gap}$: $\mathcal{L}_{gap} = I(\boldsymbol{\Delta}|\boldsymbol{H}) = -\log p(\boldsymbol{\Delta}|\boldsymbol{H})$

$$= -\sum_{\boldsymbol{\delta}} \sum_{\boldsymbol{h}} \log p(\boldsymbol{\delta}|\boldsymbol{h})$$

$$= -\sum_{\boldsymbol{l}} \sum_{\boldsymbol{h}} \log p(\boldsymbol{l} - \hat{\boldsymbol{d}}|\boldsymbol{h})$$

    - We consider the conditional self-information here
    - It can be also interpreted and derived from the view of the probability distribution.
  - the label irrelevant information modeling
    - $I(\boldsymbol{X}, \boldsymbol{H}) = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{h}} p(\boldsymbol{x}, \boldsymbol{h}) \log \frac{p(\boldsymbol{h}|\boldsymbol{x})}{p(\boldsymbol{h})}$

$$\Downarrow$$

$$I(\boldsymbol{X}, \boldsymbol{H}) \leqslant \sum_{\boldsymbol{x}} \sum_{\boldsymbol{h}} p(\boldsymbol{x}, \boldsymbol{h}) \log \frac{p(\boldsymbol{h}|\boldsymbol{x})}{q(\boldsymbol{h})} = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{l}} p(\boldsymbol{x}, \boldsymbol{l}) \mathrm{KL}(p(\boldsymbol{h}|\boldsymbol{x}) || q(\boldsymbol{h}))$$

LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - the overall objective and optimization

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{as} + \alpha\mathcal{L}_{gap} + \beta I(\boldsymbol{X}, \boldsymbol{H}) \\ &\leqslant -\sum_{\boldsymbol{x}}\sum_{\boldsymbol{l}}\sum_{\boldsymbol{h}} p(\boldsymbol{x},\boldsymbol{l})p(\boldsymbol{h}|\boldsymbol{x},\boldsymbol{l})\log q(\boldsymbol{l}|\boldsymbol{h}) \\ &\quad -\alpha\sum_{\boldsymbol{l}}\sum_{\boldsymbol{h}}\log p(\boldsymbol{l}-\hat{\boldsymbol{d}}|\boldsymbol{h}) \\ &\quad +\beta\sum_{\boldsymbol{x}}\sum_{\boldsymbol{l}} p(\boldsymbol{x},\boldsymbol{l})\mathrm{KL}(p(\boldsymbol{h}|\boldsymbol{x})||q(\boldsymbol{h})).\end{aligned}$$

✓ the empirical Monte Carlo approximation of sampling

$$\begin{aligned}\mathcal{L}_{LIB} &= \frac{1}{n}\sum_{i=1}^{n}[-\sum_{\boldsymbol{h}} p(\boldsymbol{h}|\boldsymbol{x}_i)\log q(\boldsymbol{l}_i|\boldsymbol{h}) \\ &+\beta\mathrm{KL}(p(\boldsymbol{h}|\boldsymbol{x}_i)||q(\boldsymbol{h}))] - \alpha\sum_{\boldsymbol{l}}\sum_{\boldsymbol{h}}\log p(\boldsymbol{l}-\hat{\boldsymbol{d}}|\boldsymbol{h})\end{aligned}$$

$$\begin{aligned}&\min_{\theta_{en},\theta_{de},\theta_{gd},\theta_{ld}} \mathcal{L}_{LIB} \\ \Rightarrow &\min_{\theta_{en},\theta_{de},\theta_{gd},\theta_{ld}} \frac{1}{n}\sum_{\boldsymbol{l}}[\frac{1}{2}\|\boldsymbol{\mu}_{\boldsymbol{l}|\boldsymbol{h}}-\boldsymbol{l}\|_2^2 \\ &+\alpha(\frac{1}{2}(\boldsymbol{l}-\hat{\boldsymbol{d}})^T(\boldsymbol{\sigma}_{\boldsymbol{\delta}|\boldsymbol{h}}^{-2}\boldsymbol{I})(\boldsymbol{l}-\hat{\boldsymbol{d}}) + \log\det(\boldsymbol{\sigma}_{\boldsymbol{\delta}|\boldsymbol{h}}^{2}\boldsymbol{I}))] \\ &+\frac{\beta}{2}\sum_{\boldsymbol{x}}[\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{x}}^T\boldsymbol{\mu}_{\boldsymbol{h}|\boldsymbol{x}} + \mathrm{tr}(\boldsymbol{\sigma}_{\boldsymbol{h}|\boldsymbol{x}}^2\boldsymbol{I}) - \log\det(\boldsymbol{\sigma}_{\boldsymbol{h}|\boldsymbol{x}}^2\boldsymbol{I})]\end{aligned}$$

✓ Reparameterization trick
✓ Variational Inference

LE via Label Information Bottleneck

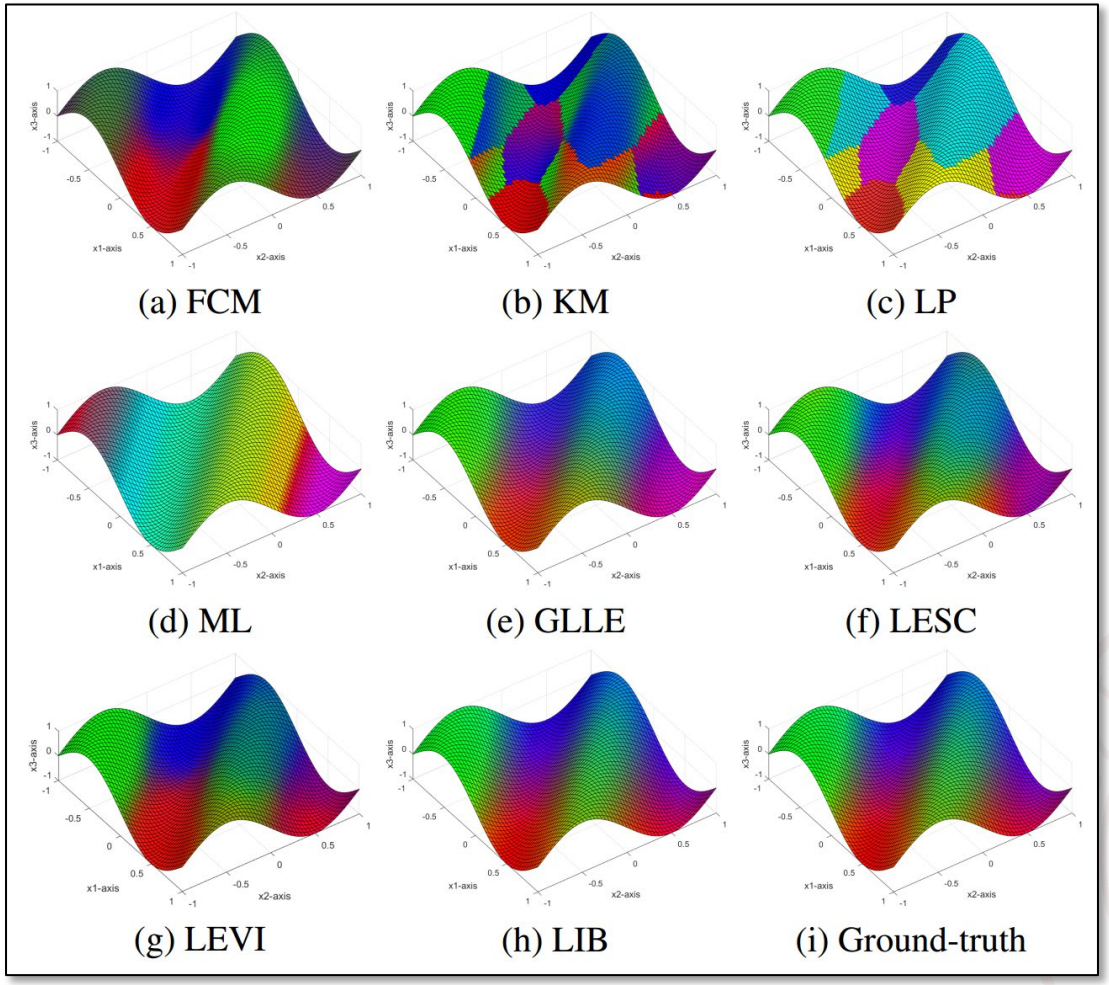
- LIB (Label Information Bottleneck)
  - model discussions
    - LIB address the problem from the perspective of information bottleneck
    - it achieves the following term

$$\frac{1}{2}(\boldsymbol{l}-\hat{\boldsymbol{d}})^T(\boldsymbol{\sigma}_{\boldsymbol{\delta}|\boldsymbol{h}}^{-2}\boldsymbol{I})(\boldsymbol{l}-\hat{\boldsymbol{d}})+\log\det(\boldsymbol{\sigma}_{\boldsymbol{\delta}|\boldsymbol{h}}^{2}\boldsymbol{I}) \qquad \|\boldsymbol{d}-\boldsymbol{l}\|_2^2$$

    - It can be deduced by excavating the label relevant information about the label gaps between logical labels and label distributions reasonably.
  - compare with LEVI further
    - LIB makes attempts from the perspective of information bottleneck, while LEVI from the view of variational inference
    - The formulas of LEVI and LIB are just partially similar in form, since the variational inference is employed as the optimization tool in LIB

LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - visualization experiments



LE via Label Information Bottleneck

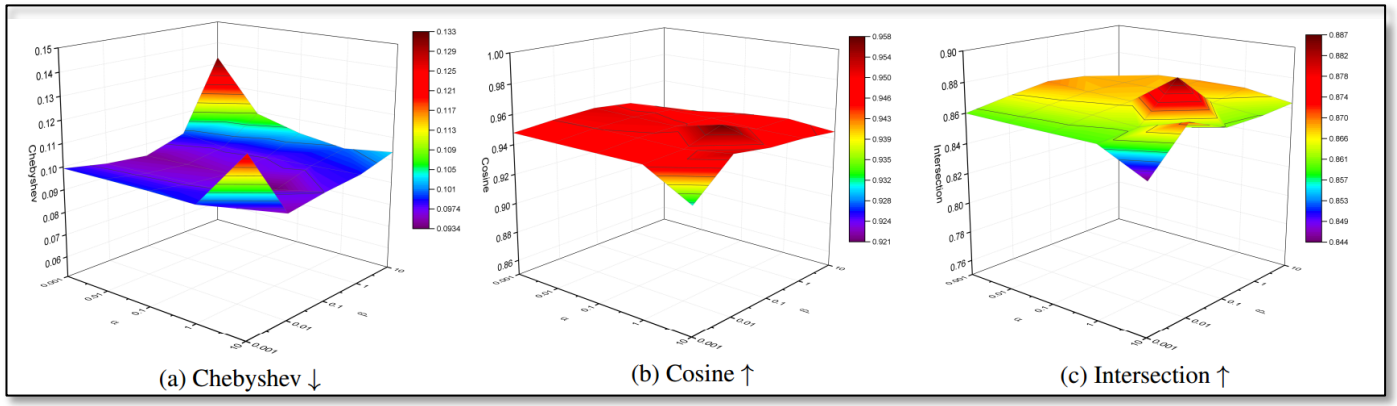
- LIB (Label Information Bottleneck)
  - comparison experiments

Metric	Chebyshev ↓								Clark ↓							
Method	FCM	KM	LP	ML	GLLE	LESC	LEVI	LIB	FCM	KM	LP	ML	GLLE	LESC	LEVI	LIB
Movie	0.230	0.234	0.161	0.164	0.122	0.121	0.110	**0.107**	0.859	1.766	0.913	1.140	0.569	0.564	0.551	**0.517**
SUB-3DFE	0.135	0.238	0.123	0.233	0.126	0.122	0.095	**0.094**	0.482	1.907	0.580	1.848	0.391	0.378	0.303	**0.297**
SJAFFE	0.132	0.214	0.107	0.186	0.087	**0.069**	0.075	0.071	0.522	1.874	0.502	1.519	0.377	0.276	0.290	**0.262**
Yeast-alpha	0.044	0.063	0.040	0.057	0.020	0.015	**0.012**	0.017	0.821	3.153	1.185	3.088	0.337	**0.253**	0.319	0.275
Yeast-cdc	0.051	0.076	0.042	0.071	0.022	0.019	**0.016**	0.017	0.739	2.885	1.014	2.825	0.306	0.251	0.323	**0.242**
Yeast-cold	0.141	0.252	0.137	0.242	0.066	0.056	0.082	**0.054**	0.433	1.472	0.503	1.440	0.176	0.152	0.269	**0.146**
Yeast-diau	0.124	0.152	0.099	0.148	0.053	**0.042**	0.044	0.049	0.838	1.886	0.788	1.844	0.296	**0.224**	0.295	0.273
Yeast-dtt	0.097	0.257	0.128	0.244	0.052	0.043	0.084	**0.034**	0.329	1.477	0.499	1.446	0.143	0.119	0.294	**0.092**
Yeast-elu	0.052	0.078	0.044	0.072	0.023	0.019	**0.017**	0.018	0.579	2.768	0.973	2.711	0.295	0.241	0.317	**0.224**
Yeast-heat	0.169	0.175	0.086	0.165	0.049	0.046	0.052	**0.039**	0.580	1.802	0.568	1.764	0.213	0.199	0.288	**0.165**
Yeast-spo	0.130	0.175	0.090	0.171	0.062	0.060	0.055	**0.053**	0.520	1.811	0.558	1.768	0.266	0.258	0.277	**0.224**
Yeast-spo5	0.162	0.277	0.114	0.273	0.099	0.092	0.091	**0.076**	0.395	1.059	0.274	1.036	0.197	0.185	0.209	**0.158**
Yeast-sopem	0.233	0.408	0.163	0.403	0.088	0.087	0.115	**0.069**	0.401	1.028	0.272	1.004	0.132	0.129	0.182	**0.104**
Avg.Rank	6.077	8.000	5.000	6.846	3.769	2.308	2.463	**1.538**	5.385	8.000	5.615	7.000	3.385	1.923	3.462	**1.231**

Metric	Cosine ↑								Intersection ↑							
Method	FCM	KM	LP	ML	GLLE	LESC	LEVI	LIB	FCM	KM	LP	ML	GLLE	LESC	LEVI	LIB
Movie	0.773	0.880	0.929	0.919	0.936	0.937	0.954	**0.955**	0.677	0.649	0.778	0.779	0.831	0.833	0.849	**0.859**
SUB-3DFE	0.912	0.812	0.922	0.815	0.927	0.932	0.956	**0.958**	0.827	0.579	0.810	0.587	0.850	0.855	0.882	**0.887**
SJAFFE	0.906	0.827	0.941	0.857	0.958	0.973	0.969	**0.974**	0.821	0.593	0.837	0.661	0.872	0.905	0.897	**0.909**
Yeast-alpha	0.922	0.751	0.911	0.756	0.987	**0.992**	0.989	**0.992**	0.844	0.532	0.774	0.537	0.938	**0.953**	0.932	0.951
Yeast-cdc	0.929	0.754	0.916	0.759	0.987	0.991	0.987	**0.992**	0.847	0.533	0.779	0.538	0.937	0.950	0.925	**0.951**
Yeast-cold	0.922	0.779	0.925	0.784	0.982	0.986	0.970	**0.988**	0.833	0.559	0.794	0.565	0.924	0.935	0.881	**0.938**
Yeast-diau	0.882	0.799	0.915	0.803	0.975	**0.985**	0.980	0.979	0.760	0.588	0.788	0.593	0.906	**0.933**	0.908	0.913
Yeast-dtt	0.959	0.759	0.921	0.763	0.988	0.991	0.965	**0.995**	0.894	0.541	0.786	0.546	0.939	0.949	0.866	**0.961**
Yeast-elu	0.950	0.758	0.918	0.763	0.987	0.991	0.987	**0.992**	0.883	0.539	0.782	0.544	0.936	0.949	0.924	**0.952**
Yeast-heat	0.883	0.779	0.932	0.783	0.984	0.986	0.977	**0.990**	0.807	0.559	0.805	0.564	0.929	0.934	0.897	**0.946**
Yeast-spo	0.909	0.800	0.939	0.803	0.974	0.975	0.978	**0.982**	0.836	0.575	0.819	0.580	0.909	0.912	0.903	**0.925**
Yeast-spo5	0.922	0.882	0.969	0.884	0.971	0.974	0.979	**0.983**	0.838	0.724	0.886	0.727	0.901	0.908	0.909	**0.924**
Yeast-sopem	0.878	0.812	0.950	0.815	0.978	0.978	0.972	**0.985**	0.767	0.592	0.837	0.597	0.912	0.913	0.885	**0.931**
Avg.Rank	5.846	7.923	5.308	6.923	3.462	2.154	2.923	**1.231**	5.385	8.000	5.692	6.846	3.385	2.007	3.462	**1.154**

LE via Label Information Bottleneck

- LIB (Label Information Bottleneck)
  - sensitivity analysis



(a) Chebyshev ↓ (b) Cosine ↑ (c) Intersection ↑

  - ablation study

Metric	Chebyshev ↓		Clark ↓		Canberra ↓		Kullback-Leibler ↓		Cosine ↑		Intersection ↑	
Method	LIB_{gap}	LIB	LIB_{gap}	LIB	LIB_{gap}	LIB	LIB_{gap}	LIB	LIB_{gap}	LIB	LIB_{gap}	LIB
Movie	0.120	**0.107**	0.563	**0.517**	1.029	**0.920**	0.099	**0.077**	0.938	**0.955**	0.834	**0.859**
SUB-3DFE	0.130	**0.094**	0.395	**0.297**	0.849	**0.611**	0.079	**0.041**	0.923	**0.958**	0.846	**0.887**
SJAFFE	0.113	**0.071**	0.391	**0.262**	0.816	**0.531**	0.066	**0.027**	0.938	**0.973**	0.860	**0.909**
Yeast-alpha	0.018	**0.017**	0.281	**0.275**	0.920	**0.893**	0.010	**0.009**	0.991	**0.992**	0.950	**0.951**
Yeast-cdc	0.019	**0.017**	0.254	**0.242**	0.782	**0.747**	0.009	**0.008**	0.991	**0.992**	0.948	**0.951**
Yeast-cold	0.061	**0.017**	0.162	**0.146**	0.280	**0.250**	0.016	**0.012**	0.985	**0.988**	0.930	**0.938**
Yeast-diau	0.050	**0.049**	0.288	**0.273**	0.659	**0.621**	0.025	**0.022**	0.977	**0.979**	0.908	**0.913**
Yeast-dtt	0.045	**0.034**	0.124	**0.092**	0.217	**0.158**	0.010	**0.005**	0.991	**0.995**	0.946	**0.961**
Yeast-elu	0.019	**0.018**	0.237	**0.224**	0.714	**0.670**	0.009	**0.008**	0.992	**0.992**	0.949	**0.952**
Yeast-heat	0.045	**0.039**	0.193	**0.165**	0.388	**0.327**	0.014	**0.011**	0.986	**0.990**	0.936	**0.946**
Yeast-spo	0.059	**0.053**	0.253	**0.224**	0.523	**0.454**	0.025	**0.019**	0.976	**0.982**	0.914	**0.925**
Yeast-spo5	0.097	**0.076**	0.193	**0.158**	0.300	**0.241**	0.032	**0.021**	0.971	**0.983**	0.903	**0.924**
Yeast-sopem	0.088	**0.069**	0.130	**0.104**	0.181	**0.144**	0.027	**0.018**	0.977	**0.985**	0.912	**0.931**

Outline

1. Background: Label Distribution Learning and Label Enhancement

2. Overview of Label Enhancement

3. Label Enhancement via Global Sample Correlation

4. Label Enhancement via Label Information Bottleneck

5. Discussions and Further Works

Discussions and Further Works

- key issues of LE
  - how to extract the underlying information effectively
    - sample correlations (AAAI 2020, TKDE 2021)
    - label information bottleneck (CVPR 2023)

- further works
  - LE based on multi-view data
    - Tensor based Multi-View Label Enhancement for Multi-Label Learning (Zhang F, Jia X, and Li W. IJCAI 2020)
  - LE in different real-life scenarios
    - medically assisted diagnosis
    - LE and generative model
    - …

Thanks!