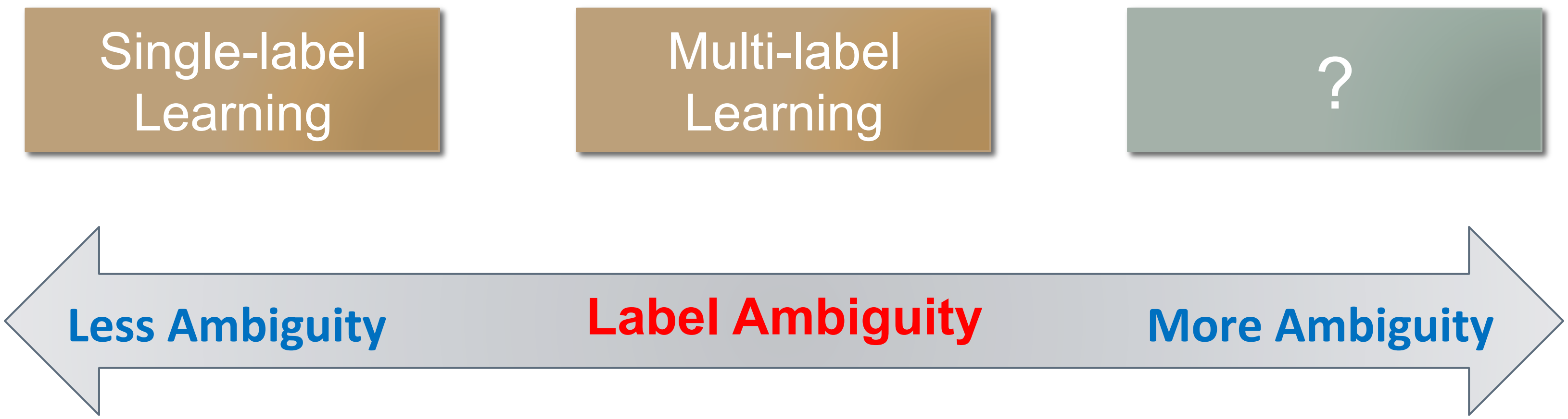
Label Distribution Learning and Its Applications

Xin Geng （耿新）

Pattern Learning and Mining (PALM) Lab
（模式学习与挖掘实验室, http://palm.seu.edu.cn）

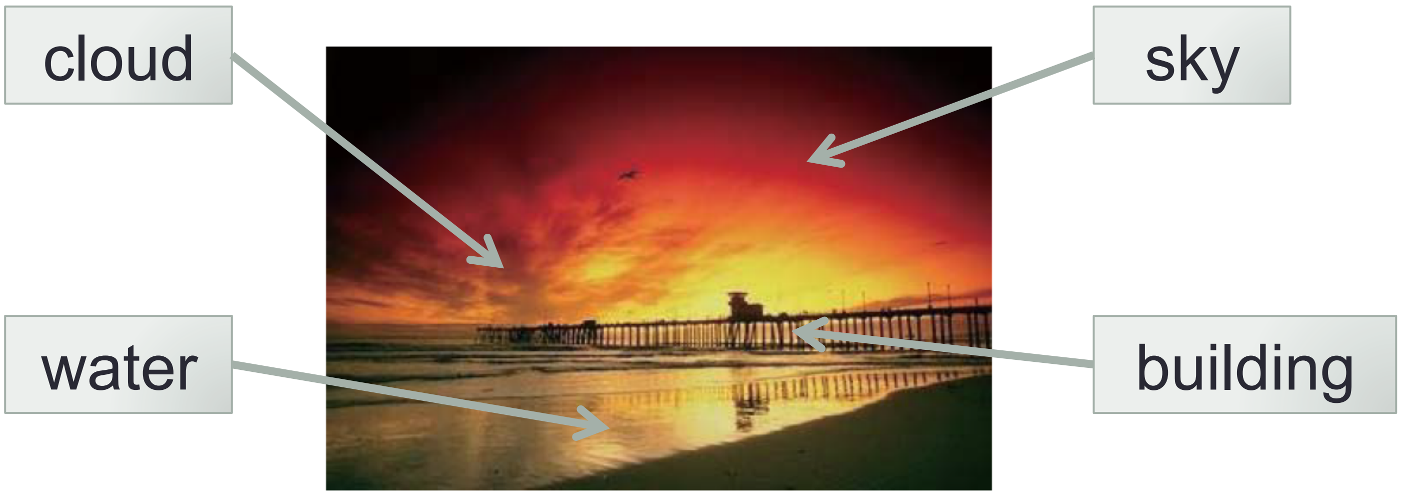
School of Computer Science and Engineering
Southeast University, Nanjing, China
（东南大学）

Learning with Ambiguity



Label Ambiguity

- "**What** describes the instance?"



Multi-label Learning

More Ambiguity?

- “**How** to describe the instance?”

some cloud

mostly sky

much water

a bit of building

How to learn?

- MLL

  Thresholding ➡ Positive labels ➡ MLL

  Not a good choice!

- Label Distribution Learning (LDL)
  - Assign a real number to each label
    - Importance
    - Confidence
    - Probability
    - Level
    - ……

  Keep more, learn more

LDL – Problem Formulation

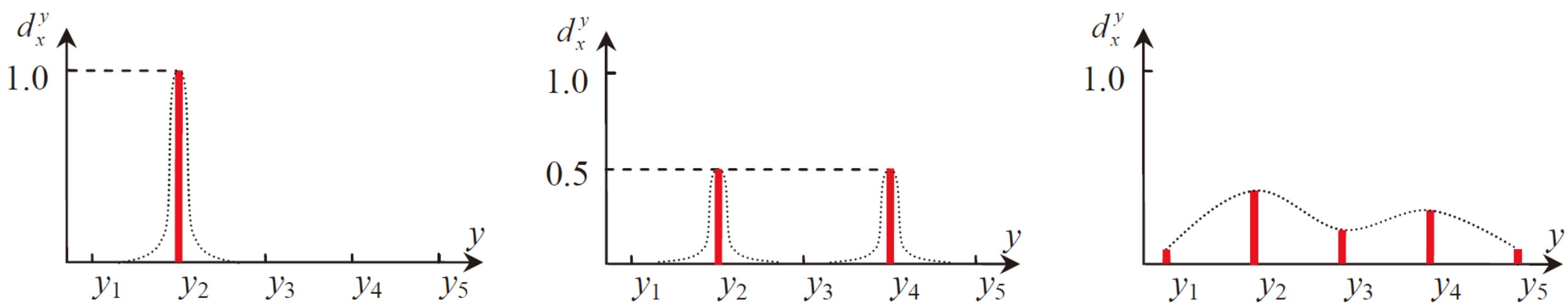
Description Degree

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

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

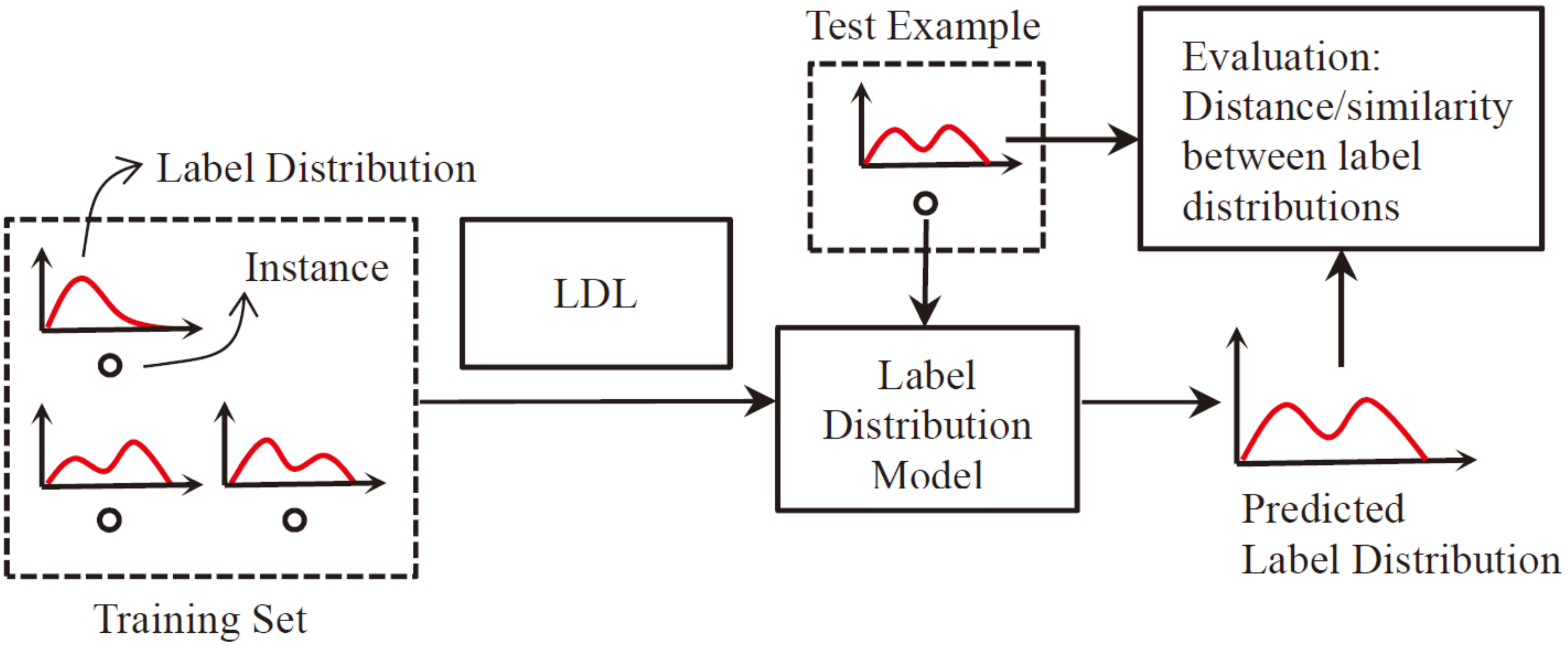
Complete label set ⇨ $\sum_y d_{\boldsymbol{x}}^{y} = 1$

Label Distribution



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

LDL – Problem Formulation



LDL – Algorithms

- Two Categories
  - Conditional Probability Mass Function (Classification)

    Model the mapping from the instance $\boldsymbol{x}$ to the label distribution $\boldsymbol{d}$ via a conditional PMF $p(y|\boldsymbol{x})$

    - IIS-LLD [Geng, Smith-Miles and Zhou, AAAI'10]
    - BFGS-LLD [Geng and Ji, ICDMW'13]
    - CPNN [Geng, Yin and Zhou, TPAMI'13]
  - Multivariate Support Vector Regression (Regression)

    Model the mapping from the instance $\boldsymbol{x}$ to the label distribution $\boldsymbol{d}$ via a multivariate support vector machine

    - LDSVR [Geng and Hou, IJCAI'15]

Reference and Matlab Code
http://cse.seu.edu.cn/PersonalPage/xgeng/LDL

Where are the Label Distributions?

- They come with the original data
  - Emotion Distribution [Zhou, Xue and Geng, ACMMM'15]
  - Movie Rating Distribution [Geng and Hou, IJCAI'15]
- They come from the prior knowledge
  - Facial Age Estimation [Geng, Yin and Zhou, TPAMI'13]
  - Head Pose Estimation [Geng and Xia, CVPR'14]
- They are learnt from the data
  - Multilabel Ranking for Natural Scene Images [Geng and Luo, CVPR'14]
  - Relative Labeling-Importance Aware Multi-label Learning [Li, Zhang and Geng, ICDM'15]

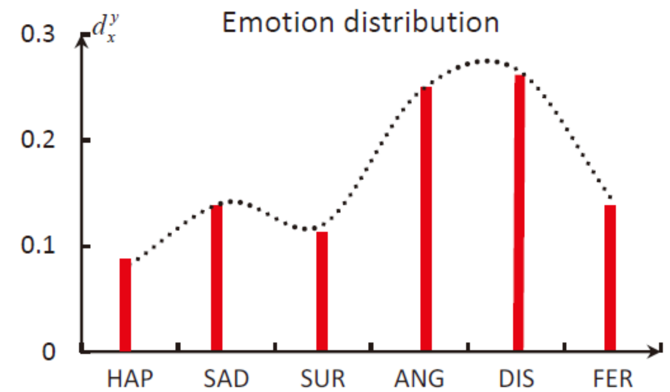
Come with Original Data…

- **Emotion Distribution via Facial Expressions**

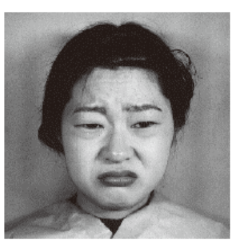
[Zhou, Xue and Geng, ACMMM'15]

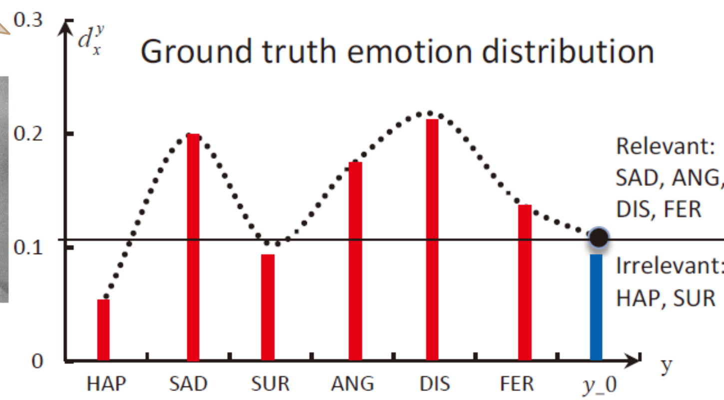
Emotion Distribution Recognition

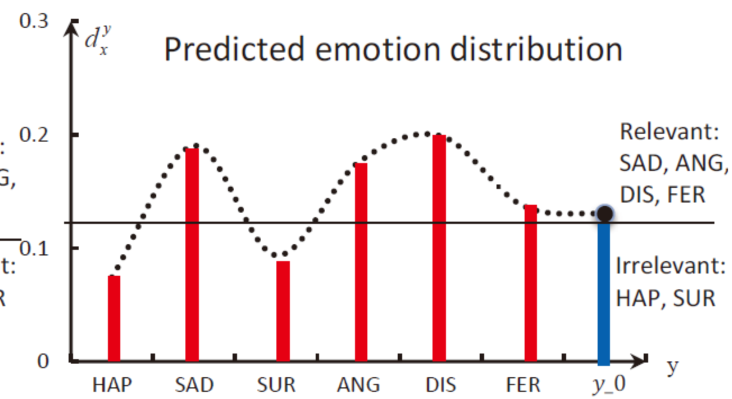
emotion	score	Multi-label
HAP	1.35	-1
SAD	2.32	-1
SUR	1.97	-1
ANG	4.03	1
DIS	4.39	1
FER	2.35	-1



Multi-Emotion Recognition







Come with Original Data…

- **Emotion Distribution via Facial Expressions**

[Zhou, Xue and Geng, ACMMM'15]

Table 1: Experimental results of Label Distribution Learning Methods

database	Algorithm	Evaluation Criterion					
		Euclidean(↓)	Sφrensen(↓)	Squared $\mathcal{X}^2$(↓)	K-L(↓)	Intersection(↑)	Fidelity(↑)
s-JAFFE	EDL	**0.0957±0.0068**	**0.1002±0.0059**	**0.0339±0.0043**	**0.0346±0.0045**	**0.8998±0.0059**	**0.9914±0.0011**
	AA-KNN [4]	0.1306±0.0117•	0.1273±0.0110•	0.0534±0.0086•	0.0556±0.0099•	0.8727±0.0110•	0.9863±0.0023•
	PT-Bayes [4]	0.1682±0.0219•	0.1644±0.0168•	0.0835±0.0215•	0.0916±0.0269•	0.8356±0.0168•	0.9784±0.0059•
	PT-SVM [4]	0.1696±0.0117•	0.1689±0.0099•	0.0812±0.0094•	0.0854±0.0110•	0.8311±0.0099•	0.9792±0.0025•
	AA-BP [4]	0.1908±0.0208•	0.1880±0.0195•	0.1139±0.0210•	0.1100±0.0273•	0.8120±0.0195•	0.9685±0.0058•
s-BU_3DFE	EDL	**0.1055±0.0023**	**0.1061±0.0025**	**0.0402±0.0017**	**0.0420±0.0020**	**0.8939±0.0043**	**0.9898±0.0046**
	AA-KNN [4]	0.1549±0.0036•	0.1464±0.0042•	0.0697±0.0036•	0.0743±0.0031•	0.8536±0.0008•	0.9821±0.0036•
	AA-BP [4]	0.1648±0.0076•	0.1595±0.0065•	0.0760±0.0061•	0.0808±0.0063•	0.8405±0.0017•	0.9804±0.0061•
	PT-Bayes [4]	0.1659±0.0044•	0.1606±0.0049•	0.0766±0.0039•	0.0830±0.0037•	0.8394±0.0010•	0.9803±0.0039•
	PT-SVM [4]	0.1701±0.0032•	0.1638±0.0047•	0.0799±0.0044•	0.0877±0.0029•	0.8362±0.0007•	0.9794±0.0044•

Table 2: Experimental results of Multi-label Learning Methods

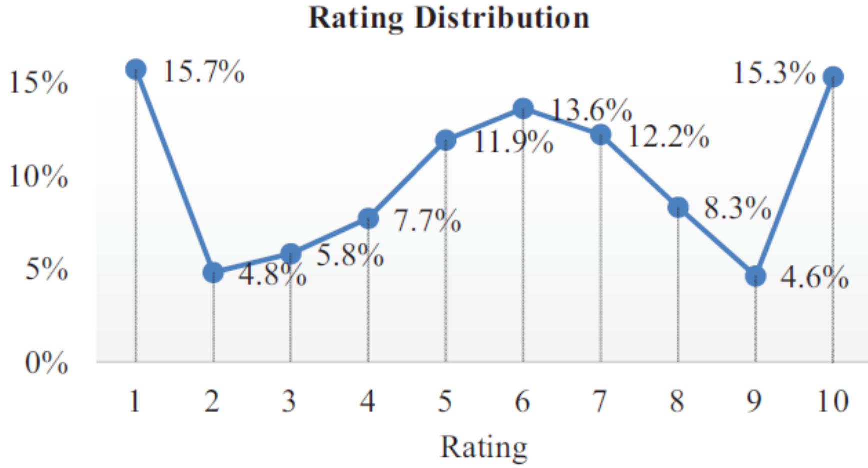
database	Algorithm	Evaluation Criterion				
		Average Precision(↑)	Coverage(↓)	Hamming Loss(↓)	One Error(↓)	Ranking Loss(↓)
s-JAFFE	EDL	**0.9037±0.0300**	**2.6913±0.3680**	**0.2540±0.0352**	**0.1175±0.0687**	**0.1374±0.0316**
	ML-RBF [17]	0.8651±0.0738•	3.2675±0.4157•	0.2484±0.0810•	0.1810±0.1088•	0.2005±0.1085•
	ML-KNN [20]	0.8455±0.0605•	3.4310±0.3822•	0.2790±0.0617•	0.1976±0.0616•	0.2184±0.0878•
	LIFT [18]	0.8050±0.1042•	3.5397±0.3604•	0.3254±0.0777•	0.2690±0.1679•	0.2515±0.1268•
	Rank-SVM [20]	0.7516±0.0982•	3.9198±0.2844•	0.3356±0.1127•	0.3008±0.1390•	0.3276±0.1189•
	MLLOC [6]	0.7502±0.0998•	3.9127±0.2146•	0.5734±0.0639•	0.3214±0.1496•	0.3399±0.1275•
	BP-MLL [19]	0.7435±0.1009•	4.0786±0.2630•	0.3591±0.1133•	0.3444±0.1477•	0.3593±0.1210•
	ECC [12]	0.7380±0.0952•	3.9500±0.2305•	0.3591±0.1133•	0.3484±0.1130•	0.3424±0.1264•
s-BU_3DFE	EDL	**0.7861±0.0216**	**0.5236±0.0758**	**0.1167±0.0069**	**0.3667±0.0349**	**0.1761±0.0178**
	ML-RBF [17]	0.7157±0.0370•	0.6756±0.1139•	0.1271±0.0075•	0.4360±0.0559•	0.2165±0.0300•
	ML-KNN [20]	0.6224±0.0282•	0.9712±0.1103•	0.1387±0.0043•	0.5699±0.0362•	0.2947±0.0243•
	LIFT [18]	0.6062±0.0323•	1.0204±0.1225•	0.1443±0.0045•	0.5913±0.0413•	0.2515±0.0296•
	Rank-SVM [20]	0.6016±0.0321•	0.9876±0.1231•	0.2118±0.0164•	0.6101±0.0380•	0.3069±0.0338•
	MLLOC [6]	0.4970±0.0288•	1.4832±0.2538•	0.1453±0.0042•	0.7140±0.0247•	0.4352±0.0557•
	BP-MLL [19]	0.5367±0.0210•	1.1692±0.1043•	0.1817±0.0098•	0.7024±0.0280•	0.3480±0.0263•
	ECC [12]	0.4419±0.0185•	1.8396±0.0677•	0.1426±0.0046•	0.7484±0.0317•	0.5438±0.0105•

Come with Original Data…

- Pre-release Prediction of Crowd Opinion on Movies

[Geng and Hou, IJCAI'15]

Title	Twilight
Average Rating	5.2/10
Budget	$ 37 Million
Gross	$ 191 Million



Pre-release Metadata

Attribute	Type	θ	#Values
Genre	C	0	24
Color	C	0	2
Director	C	5	402
1st Actor	C	5	386
2nd Actor	C	5	210
3rd Actor	C	5	103
Country	C	5	33
Language	C	5	23
Writer	C	10	16
Editor	C	10	115
Cinematographer	C	10	173
Art Direction	C	10	39
Costume Designer	C	10	110
Music By	C	10	157
Sound	C	10	26
Production Company	C	20	31
Year	N	–	–
Running Time	N	–	–
Budget	N	–	–

Crowd Rating Distribution

Come with Original Data…

- Pre-release Prediction of Crowd Opinion on Movies

[Geng and Hou, IJCAI'15]

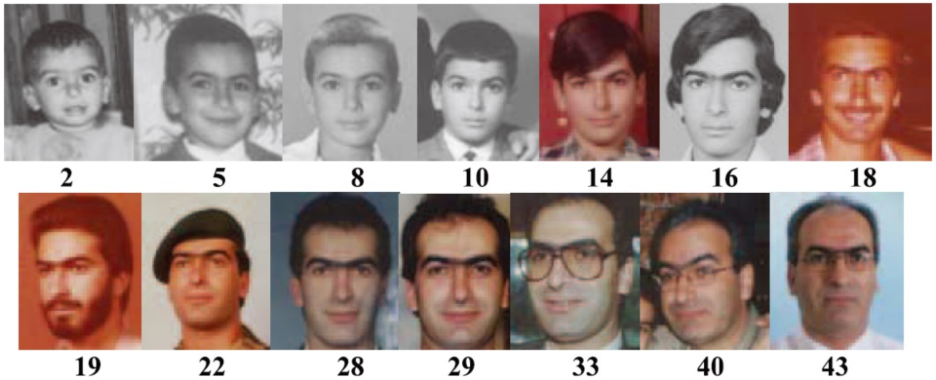
	Euclidean ↓	Sørensen ↓	Squared χ^2 ↓	K-L ↓	Intersection ↑	Fidelity ↑
LDSVR	**.1587±.0026(1)**	**.1564±.0027(1)**	**.0887±.0031(1)**	**.0921±.0035(1)**	**.8436±.0027(1)**	**.9764±.0010(1)**
S-SVR	.1734±.0024(2)	.1723±.0023(2)	.1040±.0030(2)	.1059±.0036(2)	.8277±.0023(2)	.9722±.0009(2)
M-SVR$_p$	.1843±.0031(3)	.1814±.0034(3.5)	.1084±.0033(3)	.1073±.0030(3)	.8186±.0034(3.5)	.9710±.0010(3)
BFGS-LLD	.1853±.0033(4)	.1814±.0033(3.5)	.1176±.0042(4)	.1265±.0050(4)	.8186±.0033(3.5)	.9683±.0012(4)
IIS-LLD	.1866±.0041(5)	.1828±.0044(5)	.1195±.0054(5)	.1288±.0070(6)	.8172±.0044(5)	.9676±.0014(5)
AA-kNN	.1917±.0045(6)	.1899±.0047(6)	.1246±.0062(6)	.1274±.0069(5)	.8101±.0047(6)	.9664±.0018(6)
CPNN	.2209±.0148(7)	.2153±.0150(7)	.1625±.0206(7)	.1826±.0274(7)	.7847±.0150(7)	.9551±.0061(7)

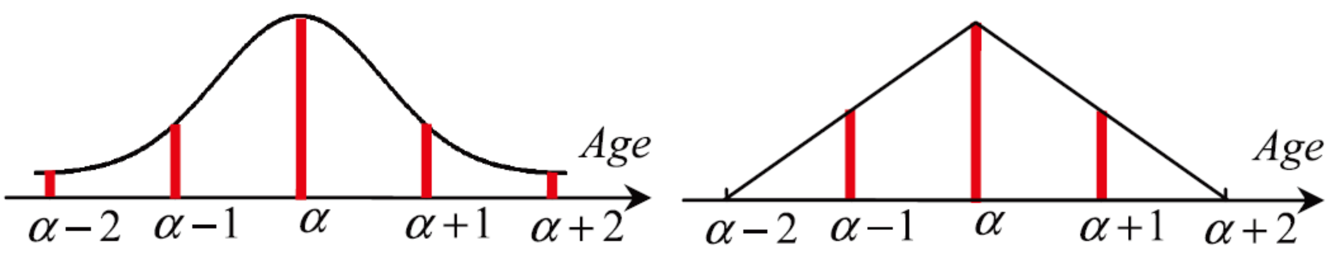
Come from Prior Knowledge…

- Facial Age Estimation [Geng, Yin, and Zhou, TPAMI'13] [Geng, Smith-Miles, and Zhou, AAAI'10]

Prior Knowledge

- Aging is a slow and gradual progress
- The faces at close ages look quite similar





(a) (b)

- **Centered at the chronological age**
- **Highest at the chronological age and gradually decrease on both sides**

Come from Prior Knowledge…

- Facial Age Estimation

[Geng, Yin, and Zhou, TPAMI'13]
[Geng, Smith-Miles, and Zhou, AAAI'10]

MAE (in Years) of Different Age Estimators

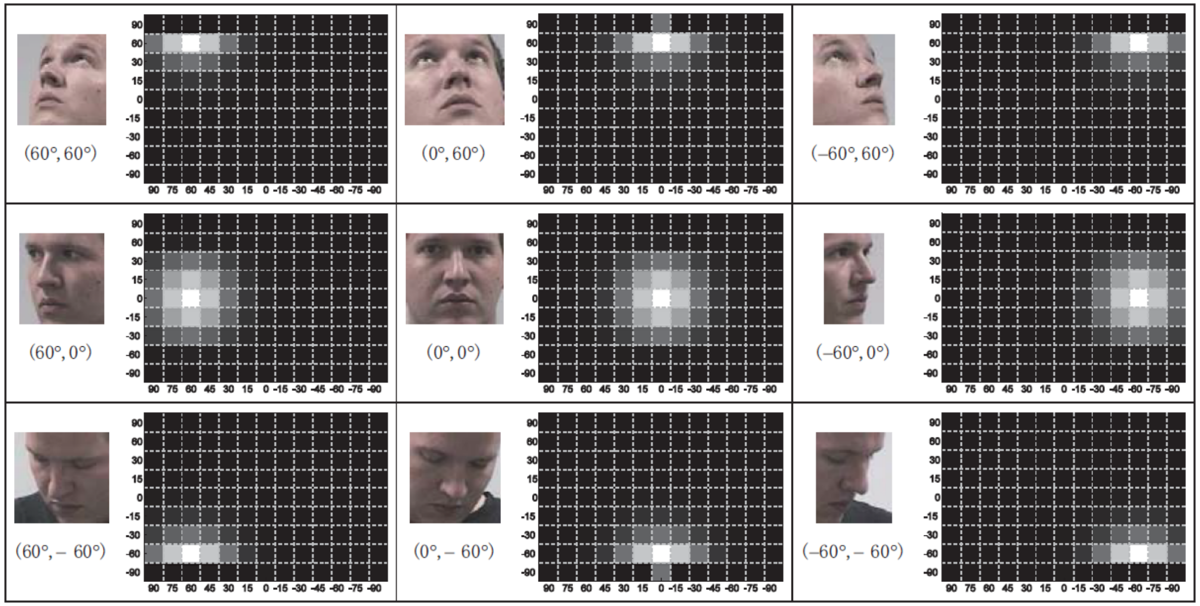
Method		Data Set: FG-NET	Data Set: MORPH
IIS-LLD	Gaussian	**5.77** (1, 1)	**5.67±0.15** (1, 1)
	Triangle	**5.90** (1, 0)	**6.09±0.14** (1, 1)
	Single	**6.27** (1, 0)	**6.35±0.17** (1, 1)
CPNN	Gaussian	**4.76** (1, 1)	**4.87±0.31** (1, 1)
	Triangle	**5.07** (1, 1)	**4.91±0.29** (1, 1)
	Single	**5.31** (1, 1)	**6.59±0.31** (1, 1)
OHRank		**6.27** (1, 0)	**6.28±0.18** (1, 1)
AGES		**6.77** (1, 1)	**6.61±0.11** (1, 1)
WAS		**8.06** (0, 1)	9.21±0.16 (1, 1)
AAS		14.83 (1, 1)	10.10±0.26 (1, 1)
*k*NN		8.24 (0, 1)	9.64±0.24 (1, 1)
BP		11.85 (1, 1)	12.59±1.38 (1, 1)
C4.5		9.34 (1, 1)	**7.48±0.12** (1, 0)
SVM		**7.25** (1, 1)	**7.34±0.17** (1, 0)
ANFIS		8.86 (0, 1)	9.24±0.17 (1, 1)
Human Tests[1]	HumanA	8.13	8.24
	HumanB	6.23	7.23

Come from Prior Knowledge…

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

Prior Knowledge

- Head orientation is intrinsically continuous
- The faces with close head orientations look quite similar



Bivariate Label Distribution

$$d_{\boldsymbol{x}}^{\boldsymbol{y}} = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}|}Z} \exp\left(-\frac{1}{2}(\boldsymbol{y}-\hat{\boldsymbol{y}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\hat{\boldsymbol{y}})\right)$$

- **Centered at the approximate pose**
- **Highest at the approximate pose and gradually decrease around**

Come from Prior Knowledge…

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

Table 1. Head Pose Estimation Results on the Pointing'04 Database.

Method	MAE			Accuracy		
	Yaw	Pitch	Yaw+Pitch	Yaw	Pitch	Yaw+Pitch
MLD-wJ	$\mathbf{4.24^\circ \pm 0.17^\circ}$	$\mathbf{2.69^\circ \pm 0.15^\circ}$	$\mathbf{6.45^\circ \pm 0.29^\circ}$	$\mathbf{73.30\% \pm 1.36\%}$	$\mathbf{86.24\% \pm 0.97\%}$	$\mathbf{64.27\% \pm 1.82\%}$
MLD-J	$5.02^\circ \pm 0.31^\circ$	$3.54^\circ \pm 0.30^\circ$	$7.94^\circ \pm 0.53^\circ$	$67.96\% \pm 2.21\%$	$81.51\% \pm 1.67\%$	$55.66\% \pm 3.28\%$
Kernel PLS	$5.79^\circ \pm 0.32^\circ$	$4.83^\circ \pm 0.29^\circ$	$9.66^\circ \pm 0.33^\circ$	$64.48\% \pm 1.79\%$	$78.35\% \pm 1.11\%$	$51.47\% \pm 1.64\%$
Linear PLS	$9.28^\circ \pm 0.48^\circ$	$8.92^\circ \pm 0.56^\circ$	$15.88^\circ \pm 0.79^\circ$	$46.49\% \pm 2.80\%$	$60.54\% \pm 2.52\%$	$28.10\% \pm 3.28\%$
Kernel SVM	$6.83^\circ \pm 0.36^\circ$	$5.91^\circ \pm 0.31^\circ$	$11.87^\circ \pm 0.39^\circ$	$57.17\% \pm 2.12\%$	$68.24\% \pm 1.71\%$	$34.23\% \pm 2.05\%$
Linear SVM	$8.30^\circ \pm 0.57^\circ$	$8.16^\circ \pm 0.48^\circ$	$14.91^\circ \pm 0.54^\circ$	$50.54\% \pm 2.81\%$	$57.67\% \pm 2.23\%$	$23.80\% \pm 1.75\%$
Kernel SVR	$6.89^\circ \pm 0.47^\circ$	$6.59^\circ \pm 0.62^\circ$	$11.99^\circ \pm 0.76^\circ$	$60.22\% \pm 3.11\%$	$71.72\% \pm 2.22\%$	$44.73\% \pm 3.46\%$
Linear SVR	$8.33^\circ \pm 0.55^\circ$	$8.27^\circ \pm 0.35^\circ$	$14.50^\circ \pm 0.68^\circ$	$52.08\% \pm 3.16\%$	$64.70\% \pm 1.38\%$	$35.16\% \pm 3.08\%$
Stiefelhagen [17][1]	9.5°	9.7°	—	52.0%	66.3%	—
Human Performance [8][2]	11.8°	9.4°	—	40.7%	59.0%	—
Gourier (Associative Memories) [8][3]	10.1°	15.9°	—	50.0%	43.9%	—
Tu (High-order SVD) [18][4]	12.9°	17.97°	—	49.25%	54.84%	—
Tu (PCA) [18][4]	14.11°	14.98°	—	55.20%	57.99%	—
Tu (LEA) [18][4]	15.88°	17.44°	—	45.16%	50.61%	—
Voit [19]	12.3°	12.77°	—	—	—	—
Li (PCA) [12][5]	26.9°	35.1°	—	—	—	—
Li (LDA) [12][5]	25.8°	26.9°	—	—	—	—
Li (LPP) [12][5]	24.7°	22.6°	—	—	—	—
Li (Local-PCA) [12][5]	24.5°	37.6°	—	—	—	—
Li (Local-LDA) [12][5]	19.1°	30.7°	—	—	—	—
Li (Local-LPP) [12][5]	29.2°	40.2°	—	—	—	—
Foytik (Two-layer Phase Cong.) [4][6]	13.0°	—	—	—	—	—

Learnt from the Data…

- Multilabel Ranking for Natural Scene Images
  [Geng and Luo, CVPR'14]
  Learnt from inconsistent rankings from different rankers

- Relative Labeling-Importance Aware Multi-label Learning
  [Li, Zhang and Geng, ICDM'15]
  Learnt from multi-label data

Multilabel Ranking for Natural Scene Images

[Geng and Luo, CVPR’14]

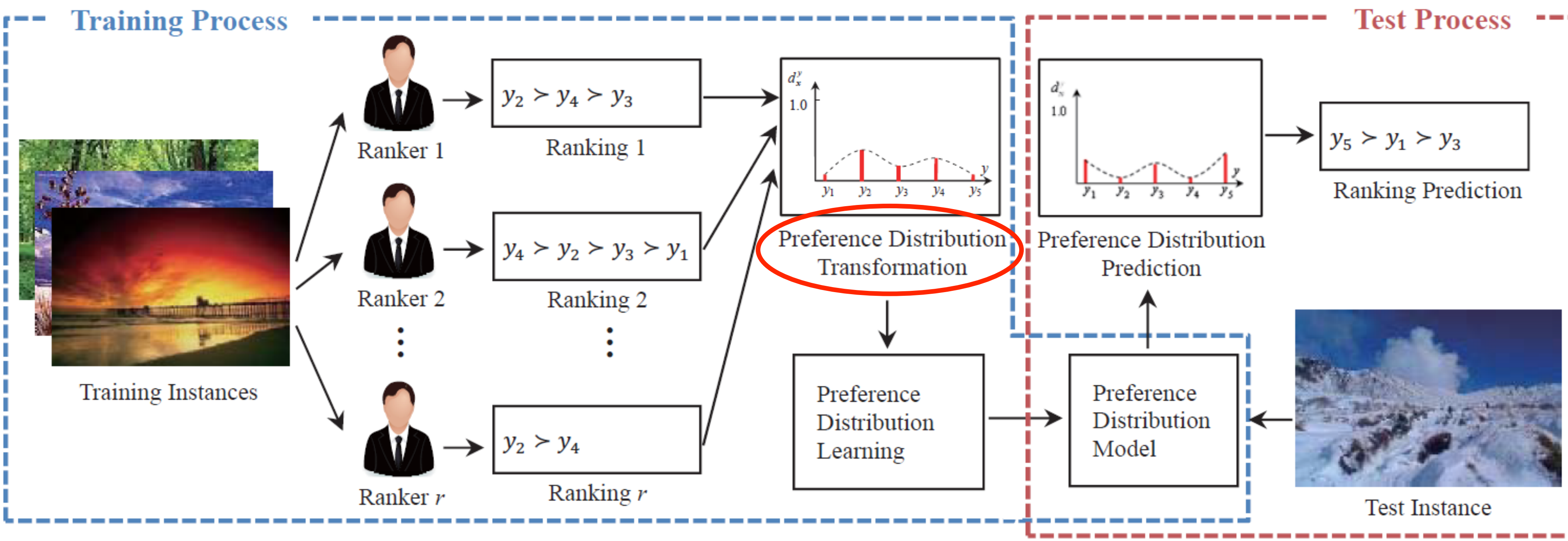
- Multilabel Ranking
  - A bipartition of the relevant (positive) and irrelevant (negative) labels
  - A proper ranking over relevant labels
- Multiple Rankers: Subjective Inconsistent “Ground Truth”

Ranker	Relevant Label Ranking
01	water $\succ$ cloud $\succ$ sky
02	cloud $\succ$ sky $\succ$ water $\succ$ building
03	water $\succ$ cloud $\succ$ sky $\succ$ building
04	water $\succ$ cloud $\succ$ sky
05	water $\succ$ sky $\succ$ cloud
06	water $\succ$ cloud $\succ$ sky
07	water $\succ$ cloud $\succ$ sky
08	sky $\succ$ water
09	water $\succ$ cloud
10	water $\succ$ cloud $\succ$ sky

Multilabel Ranking for Natural Scene Images

[Geng and Luo, CVPR'14]

- Multilabel Ranking by Preference Distribution



Virtual labels as split point between relevant and irrelevant labels

Multilabel Ranking for Natural Scene Images

[Geng and Luo, CVPR'14]

- Preference Distribution Transformation

K-L Divergence

$$\begin{aligned}
min \quad & \sum_{i=1}^{r}\left(\sum_{j=1}^{c} P^j \log\frac{P^j}{P_i^j} + vP^0 \log\frac{P^0}{P_i^0}\right) \\
w.r.t. \quad & P^j, P_i^j, i=1,\cdots,r, j=0,\cdots,c \\
s.t. \quad & P^j \geq 0, P_i^j \geq 0, \\
& \sum_{j=1}^{c} P^j + vP^0 = 1, \sum_{j=1}^{c} P_i^j + vP_i^0 = 1, \\
& P_i^j + \varepsilon \leq P_i^k, \text{if } y_k \succ_{\boldsymbol{x},i} y_j, \\
& P_i^j = 0, \text{if } y_0 \succ_{\boldsymbol{x},i} y_j, \\
& i=1,\cdots,r, j=0,\cdots,c
\end{aligned}$$

Distribution Constraints

Ranking Constraints

Irrelevant Labels

Multilabel Ranking for Natural Scene Images

[Geng and Luo, CVPR'14]

- Experiment

	Measure	PDL	CRPC	CLRT	kNN-MLR
Ranking	Kendall's Tau-b ↑	**0.3869±0.0105**	0.3414±0.0117 (1)	0.3120±0.0109 (1)	0.3686±0.0079 (1)
	Kendall's Tau-c ↑	**0.3405±0.0091**	0.3003±0.0104 (1)	0.2750±0.0095 (1)	0.3243±0.0074 (1)
	Spearman's Rho ↑	**0.5660±0.0097**	0.5257±0.0090 (1)	0.5025±0.0141 (1)	0.5501±0.0074 (1)
	Gamma ↑	**0.6241±0.0215**	0.5533±0.0228 (1)	0.5002±0.0212 (1)	0.5977±0.0137 (1)
	SAG ↑	**0.2534±0.0073**	0.2227±0.0088 (1)	0.2057±0.0068 (1)	0.2404±0.0077 (1)
Classification	Hamming loss ↓	**0.2090±0.0106**	0.2282±0.0100 (1)	0.2336±0.0147 (1)	0.2204±0.0107 (1)
	One-error ↓	**0.3382±0.0262**	0.3922±0.0248 (1)	0.4375±0.0335 (1)	0.3700±0.0252 (1)
	Coverage ↓	**2.9227±0.1573**	3.2790±0.1475 (1)	3.3713±0.1615 (1)	3.0663±0.1113 (1)
	Ranking loss ↓	**0.1688±0.0110**	0.2086±0.0110 (1)	0.2327±0.0107 (1)	0.1837±0.0069 (1)
	Average precision ↑	**0.7426±0.0162**	0.6974±0.0142 (1)	0.6714±0.0145 (1)	0.7215±0.0151 (1)

Relative Labeling-Importance Aware Multi-label Learning

[Li, Zhang and Geng, ICDM'15]

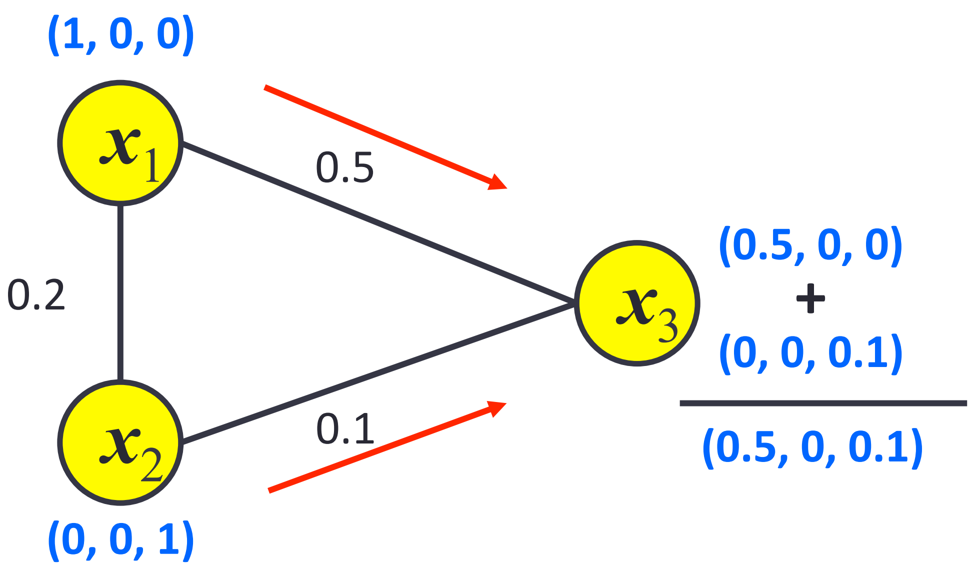
- **State-of-the-art:**
  Existing multi-label Learning methods usually assume equal label importance.
- **Fact:**
  When multiple labels are associated to the same instance, their importance to the instance can hardly be exactly same.

Relative Labeling-Importance Aware Multi-label Learning

[Li, Zhang and Geng, ICDM'15]

- Implicit Relative Labeling-Importance

Label Propagation on the Training Set



Relative Labeling-Importance Aware Multi-label Learning

[Li, Zhang and Geng, ICDM'15]

- Implicit Relative Labeling-Importance

Label Propagation on the Training Set

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

$$G = (V, E)$$

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

Propagation Matrix

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

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

Label Propagation

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

Converge to

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

Label Distribution

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

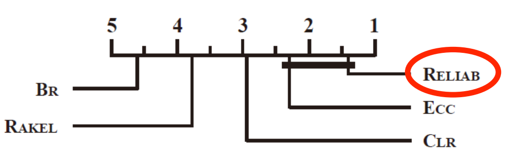
Relative Labeling-Importance Aware Multi-label Learning

[Li, Zhang and Geng, ICDM'15]

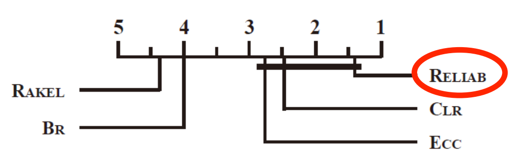
- Experiments

Comparing algorithm	One-error ↓								
	cal500	emotions	medical	llog	msra	image	scene	yeast	slashdot
RELIAB	**0.129±0.019**	**0.273±0.019**	**0.160±0.012**	**0.745±0.007**	**0.066±0.014**	**0.348±0.016**	**0.248±0.007**	**0.223±0.011**	0.509±0.014
BR	0.906±0.025	0.375±0.027	0.306±0.031	0.885±0.013	0.362±0.013	0.527±0.011	0.472±0.016	0.284±0.010	0.731±0.014
CLR	0.375±0.118	0.356±0.030	0.706±0.149	0.883±0.023	0.152±0.009	0.502±0.016	0.367±0.017	0.272±0.012	0.978±0.003
ECC	0.255±0.028	0.353±0.040	0.187±0.016	0.794±0.011	0.211±0.011	0.475±0.011	0.378±0.015	0.261±0.010	**0.476±0.015**
RAKEL	0.672±0.029	0.394±0.027	0.252±0.025	0.876±0.015	0.288±0.014	0.498±0.013	0.440±0.016	0.297±0.012	0.596±0.011

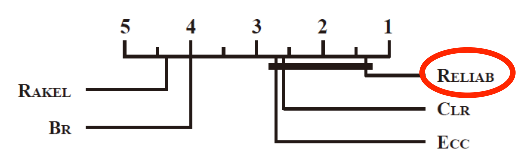
Comparing algorithm	One-error ↓							
	corel5k	rcv1-s1	rcv1-s2	rcv1-s3	rcv1-s4	rcv1-s5	bibtex	mediamill
RELIAB	0.795±0.009	0.510±0.005	**0.479±0.006**	**0.487±0.007**	**0.466±0.008**	**0.467±0.012**	**0.418±0.007**	0.192±0.007
BR	0.921±0.004	0.736±0.006	0.758±0.008	0.755±0.003	0.737±0.010	0.763±0.008	0.880±0.004	0.185±0.004
CLR	**0.748±0.011**	0.503±0.006	0.549±0.006	0.549±0.025	0.584±0.076	0.678±0.092	0.514±0.003	**0.147±0.002**
ECC	0.911±0.004	**0.490±0.005**	0.515±0.007	0.512±0.006	0.485±0.004	0.495±0.005	0.907±0.003	0.158±0.002
RAKEL	0.867±0.004	0.626±0.008	0.622±0.008	0.637±0.008	0.618±0.010	0.614±0.013	0.779±0.015	0.200±0.003



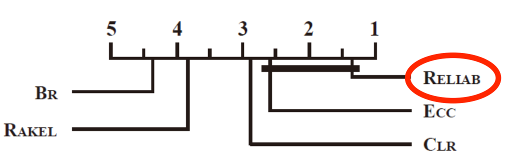
(a) *One-error*



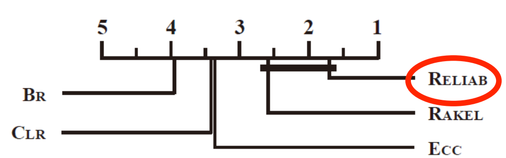
(b) *Coverage*



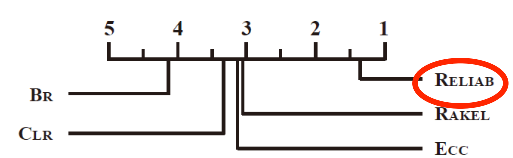
(c) *Ranking loss*



(d) *Average precision*



(e) *Macro-averaging F1*



(f) *Micro-averaging F1*

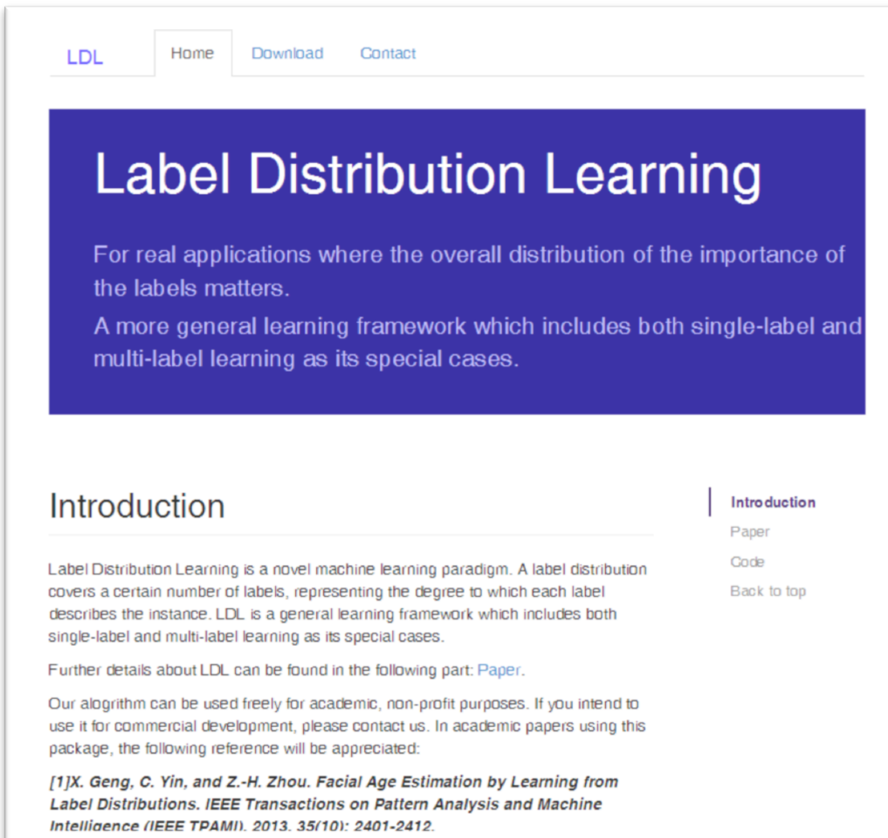
Conclusion

- Label distribution learning
  - More general framework than single-label and multi-label learning
  - Deals with different importance of labels
  - Matches certain problems better
  - Needs special design
- Label distribution comes into play when
  - **There is a natural measure of description degree**
  - **The labels are correlated to each other**
  - **There are multiple labeling sources for one instance**
  - **Multiple labels are associated to the same instance with different importance**
  - **……**

15 real-world datasets collected

Interested?

Download the LDL Matlab package from
http://cse.seu.edu.cn/PersonalPage/xgeng/



References

1. Xin Geng, Chao Yin, and Zhi-Hua Zhou. Facial Age Estimation by Learning from Label Distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence (**IEEE TPAMI**), 2013, 35(10): 2401-2412.
2. Yu-Kun Li, Min-Ling Zhang, Xin Geng. Leveraging implicit relative labeling-importance information for effective multi-label learning. In: *Proceedings of the 15th IEEE International Conference on Data Mining* ***(ICDM'15)***, Atlantic City, NJ, in press.
3. Ying Zhou, Hui Xue, Xin Geng, Emotion Distribution Recognition from Facial Expressions. In: *Proceedings of the 23rd ACM International Conference on Multimedia* ***(ACM MM'15)***, Brisbane, Australia, 2015, 1247-1250.
4. Xin Geng and Peng Hou. Pre-release Prediction of Crowd Opinion on Movies by Label Distribution Learning. In: *Proceedings of the International Joint Conference on Artificial Intelligence* ***(IJCAI'15)***, Buenos Aires, Argentina, 2015, 3511-3517.
5. Xin Geng and Yu Xia. Head Pose Estimation Based on Multivariate Label Distribution. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* (***CVPR'14***), Columbus, OH, 2014, pp. 1837-1842.
6. Xin Geng and Longrun Luo. Multilabel Ranking with Inconsistent Rankers. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* (***CVPR'14***), Columbus, OH, 2014, pp. 3742-3747.
7. Xin Geng and Rongzi Ji. Label Distribution Learning. In Proceedings of the 2013 International Conference on Data Mining Workshops (***ICDMW'13***), Dallas, TA, 2013, pp. 377-383.
8. Xin Geng, Kate Smith-Miles, Zhi-Hua 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.

Acknowledgement

- Colleagues

Prof. Zhi-Hua Zhou

Prof. Kate Smith-Miles

Prof. Min-Ling Zhang

Dr. Hui Xue

- Students

Chao Yin	Yu-Kun Li	Ying Zhou
Peng Hou	Yu Xia	Longrun Luo
Rongzi Ji		

THANK YOU

http:// palm.seu.edu.cn