



标记分布学习范式

耿新

模式学习与挖掘实验室 (PALM)
<http://palm.seu.edu.cn>

东南大学计算机科学与工程学院





报告内容

- **研究背景**
- **概念定义**
- **学习算法**
- **实验**
- **结论**





机器学习中的多义性 (Ambiguity)

非——映射

机器学习

示例



标记端多义性

标记

多示例学习范式
(多对一)

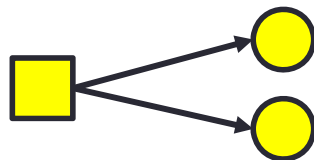
多标记学习范式
(一对多)

多示例多标记学习范式
(多对多)

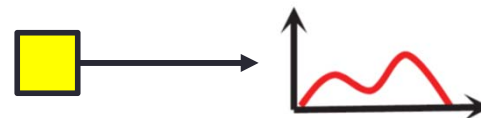
标记端多义性



单标记学习



多标记学习



标记分布学习



举例1：自然场景图像

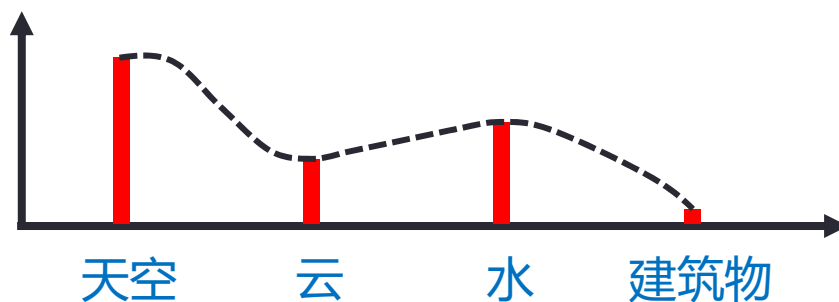
- “**哪些**标记可以用于描述该示例?”



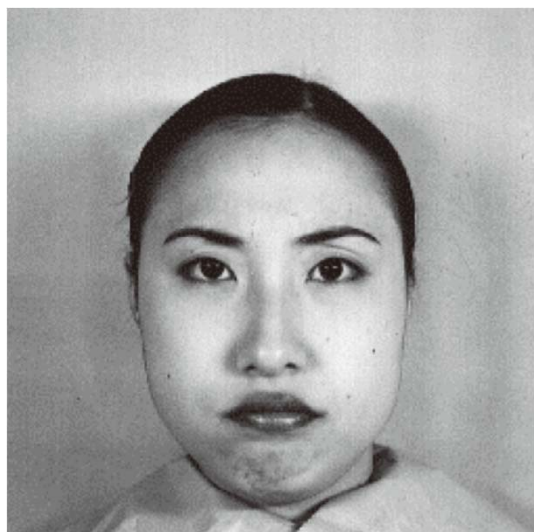
多标记学习

举例1：自然场景图像

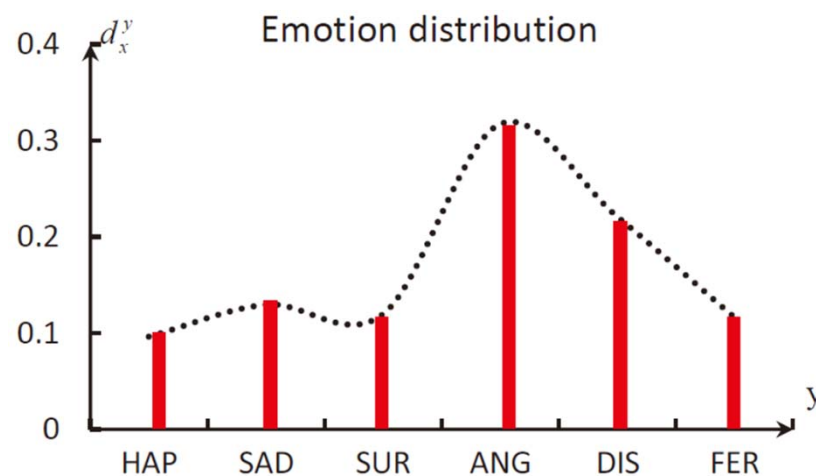
- “每个标记**如何**描述该示例?”



举例2：表情图像



情感	人工打分
高兴 (HAP)	1.48
悲伤 (SAD)	2.48
惊讶 (SUR)	2.19
气愤 (ANG)	4.48
恶心 (DIS)	3.42
害怕 (FER)	2.16





更多数据...

15个标记分布数据集，可从如下链接下载：
<http://cse.seu.edu.cn/PersonalPage/xgeng/LDL>

No.	数据集	样例数	特征数	标记数
1	Yeast-alpha	2,465	24	18
2	Yeast-cdc	2,465	24	15
3	Yeast-elu	2,465	24	14
4	Yeast-diau	2,465	24	7
5	Yeast-heat	2,465	24	6
6	Yeast-spo	2,465	24	6
7	Yeast-cold	2,465	24	4
8	Yeast-dtt	2,465	24	4
9	Yeast-spo5	2,465	24	3
10	Yeast-spoem	2,465	24	2
11	Human Gene	30,542	36	68
12	Natural Scene	2,000	294	9
13	SJAFPE	213	243	6
14	SBU_3DFE	2,500	243	6
15	Movie	7,755	1,869	5



报告内容

- 研究背景
- 概念定义
- 学习算法
- 实验
- 结论





标记分布学习

[Geng, TKDE'16]

对一个示例 x 来说，给每个标记 y 赋予一个实数值 d_x^y ，表示 y 描述 x 的程度；

不失一般性，假设 $d_x^y \in [0, 1]$ ；

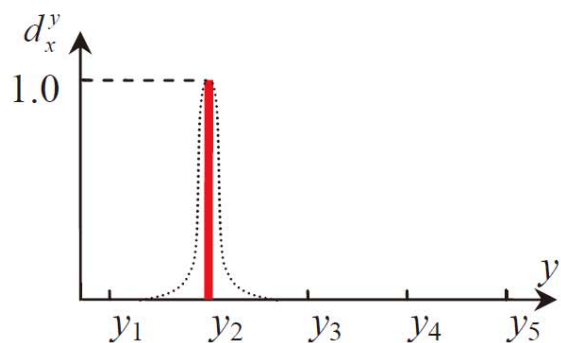
进一步假设标记集合是完备集（用所有该集中的标记一定能够完全描述一个示例），则 $\sum_y d_x^y = 1$ ；

- 满足上述条件的 d_x^y 称为 y 对 x 的**描述度**
- 对一个示例来讲，所有标记的描述度构成的数据结构称为**标记分布** $D_i = \{d_{x_i}^{y_1}, d_{x_i}^{y_2}, \dots, d_{x_i}^{y_c}\}$
- 在以标记分布标注的示例上学习的过程称为**标记分布学习 (Label Distribution Learning, LDL)**

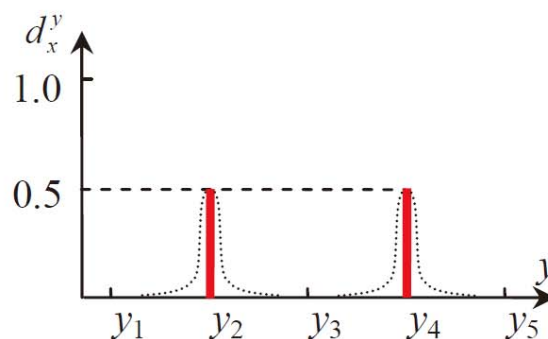


LDL vs. 传统学习范式

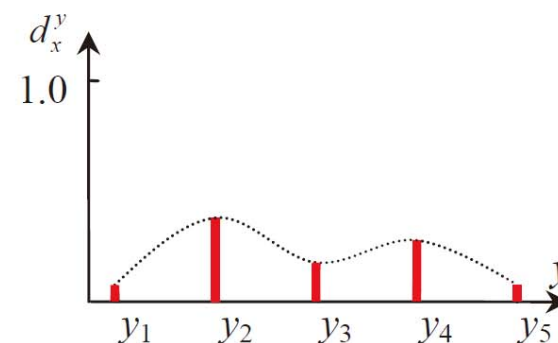
传统单标记学习 (SLL) 和多标记学习 (MLL) 可以看作标记分布学习 (LDL) 的特例



(a) Single-label



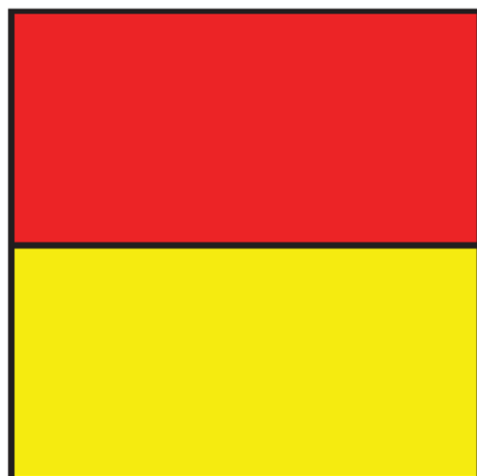
(b) Multi-label



(c) General case

LDL vs. 传统学习范式

两标记问题的决策域



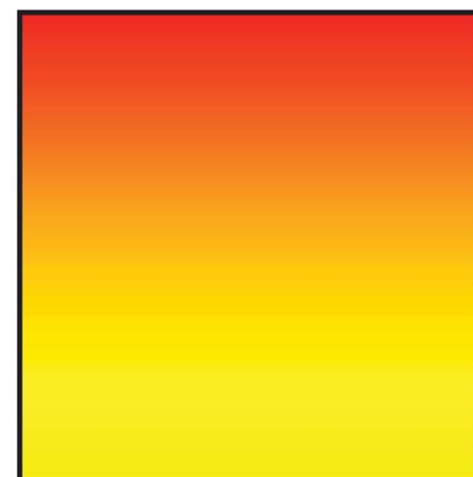
(a) Single-label

c 种可能的输出



(b) Multi-label

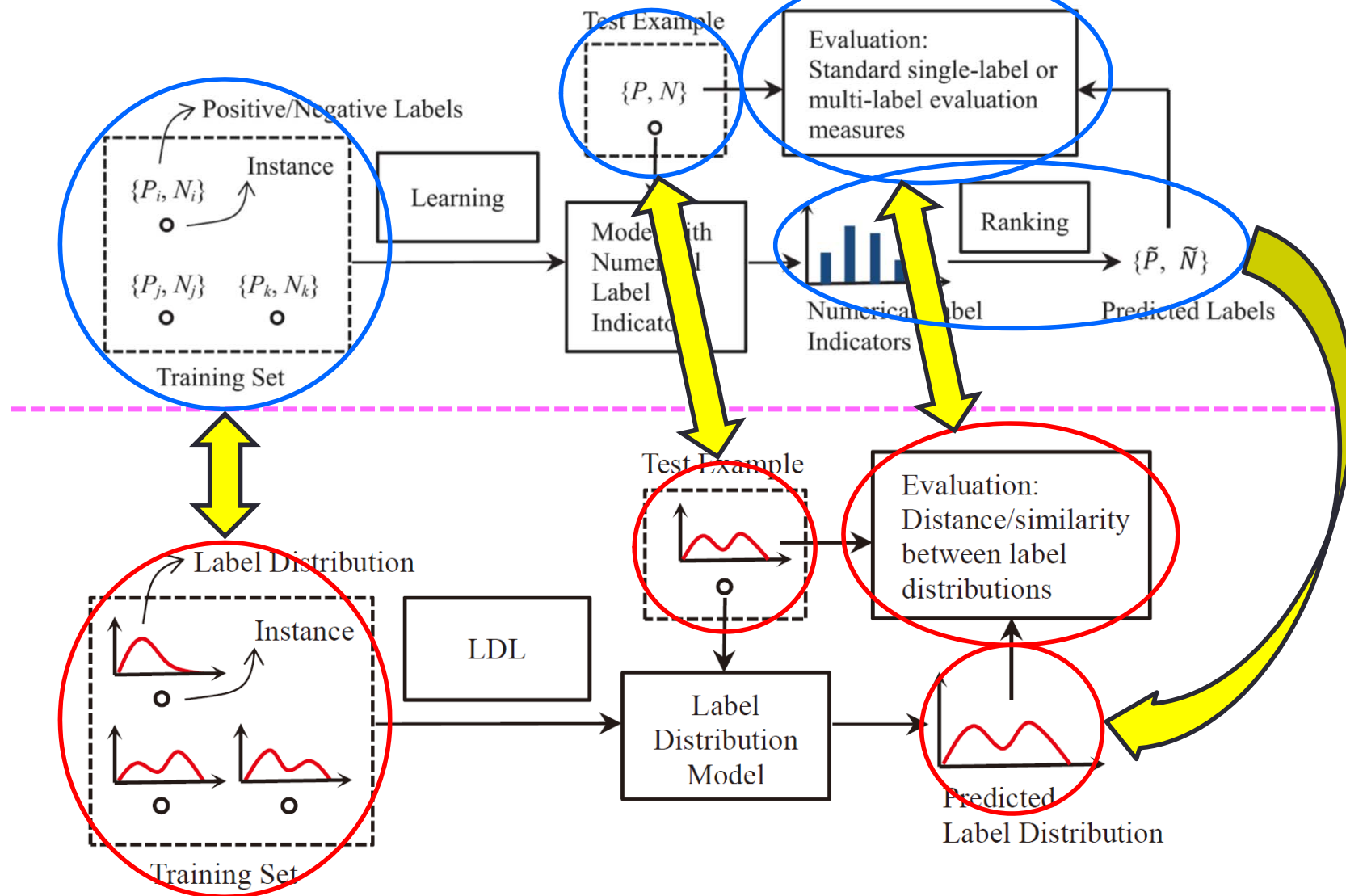
$2^c - 1$ 种可能的输出



(c) Label distribution

无限种可能的输出

LDL vs. 含数值标记指示的学习方法





LDL形式化定义

[Geng, TKDE'16]

$$d_{\mathbf{x}}^y = P(y|\mathbf{x})$$

一种可能的形式化定义形式



Let $\mathcal{X} = \mathbb{R}^q$ denote the input space and $\mathcal{Y} = \{y_1, y_2, \dots, y_c\}$ denote the complete set of labels. Given a training set $S = \{(\mathbf{x}_1, D_1), (\mathbf{x}_2, D_2), \dots, (\mathbf{x}_n, D_n)\}$, the goal of ldl is to learn a conditional probability mass function $p(y|\mathbf{x})$ from S , where $\mathbf{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$.

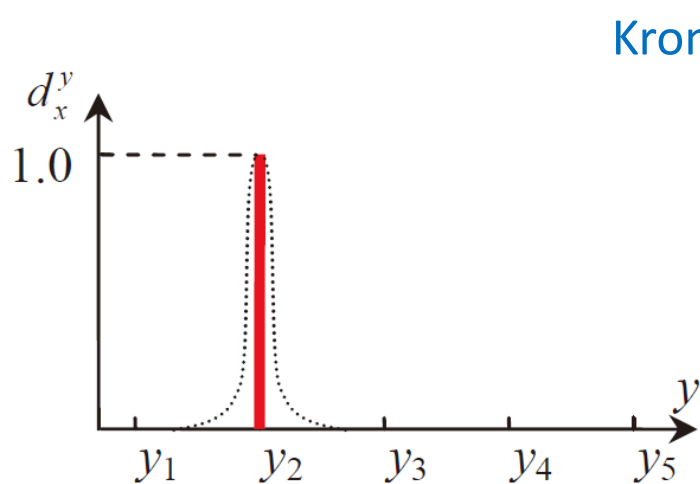


LDL优化目标

- 参数形式 $p(y|x; \theta)$
- 预测分布与真实分布之间的距离度量：K-L散度

$$\theta^* = \operatorname{argmin}_{\theta} \sum_i \sum_j \left(d_{x_i}^{y_j} \ln \frac{d_{x_i}^{y_j}}{p(y_j | x_i; \theta)} \right)$$
$$= \operatorname{argmax}_{\theta} \sum_i \sum_j d_{x_i}^{y_j} \ln p(y_j | x_i; \theta).$$

回望单标记学习



(a) Single-label

Kronecker delta function x_i 的单标记

$$d_{x_i}^{y_j} = Kr(y_j, y(x_i))$$

$$\theta^* = \operatorname{argmax}_{\theta} \sum_i \sum_j d_{x_i}^{y_j} \ln p(y_j | x_i; \theta)$$

$$\theta^* = \operatorname{argmax}_{\theta} \sum_i \ln p(y(x_i) | x_i; \theta)$$

基于 $p(y|x; \theta)$ 的
分类等价于
最大后验决策

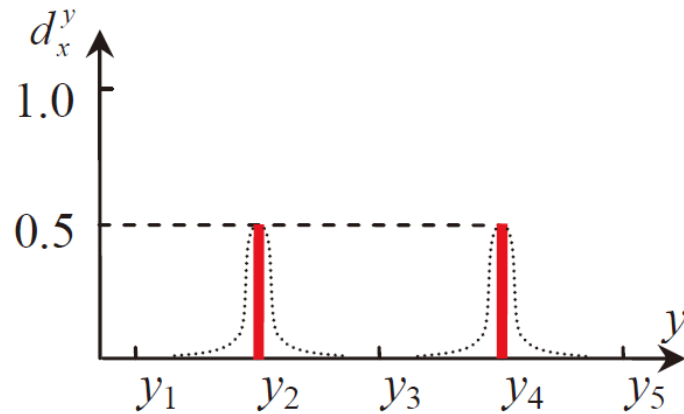
最大对数似然估计 (ML)



回望多标记学习

x_i 的相关
标记集合

$$d_{x_i}^{y_j} = \begin{cases} \frac{1}{|Y_i|}, & y_j \in Y_i \\ 0, & y_j \notin Y_i \end{cases}$$



(b) Multi-label

$$\theta^* = \operatorname{argmax}_{\theta} \sum_i \sum_j d_{x_i}^{y_j} \ln p(y_j | x_i; \theta)$$

$$\theta^* = \operatorname{argmax}_{\theta} \sum_i \frac{1}{|Y_i|} \sum_{y \in Y_i} \ln p(y | x_i; \theta)$$

1. 用Entropy-based Label Assignment (ELA) 将多标记数据转化为加权单标记数据
2. 在加权单标记数据集上应用最大对数似然估计 (ML)

G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," Int. J. Data Warehousing Mining, vol. 3, no. 3, pp. 1–13, 2007.



报告内容

- **研究背景**
- **概念定义**
- **学习算法**
- **实验**
- **结论**





LDL算法

- 三种算法设计策略
 - 问题转化 (Problem Transformation)
将LDL问题转化为传统学习范式
 - 算法改造 (Algorithm Adaption)
将传统学习算法改造为能够处理标记分布
 - 专用算法 (Specialized Algorithms)
为LDL专门设计的算法



问题转化

- 转化为加权单标记样本

$$\begin{array}{ccc} (\mathbf{x}_i, D_i) & \xrightarrow{\text{yellow arrow}} & d_{\mathbf{x}_i}^{y_j} (\mathbf{x}_i, y_j) \\ D_i = \{d_{\mathbf{x}_i}^{y_1}, d_{\mathbf{x}_i}^{y_2}, \dots, d_{\mathbf{x}_i}^{y_c}\} & & \begin{array}{l} i = 1, \dots, n \\ j = 1, \dots, c \end{array} \\ \hline n \text{ examples with LD} & & c \times n \text{ single-label examples} \end{array}$$

- 算法

- Bayes (PT-Bayes)

假设每个类 (标记) 服从Gaussian分布

通过Bayes公式计算每个类的后验概率作为对应标记的描述度 $d_{\mathbf{x}}^{y_j} = P(y_j|\mathbf{x})$

- SVM (PT-SVM)

基于成对耦合的多类分类概率估计 [Wu, Lin, and Weng, JMLR'04]



算法改造

- k -NN (AA- k NN)
 1. 找到测试示例 k 个最近的邻居;
 2. 计算 k 个邻居的标记分布均值, 作为测试示例的标记分布预测.
- BP神经网络 (AA-BP)
 - c 个输出单元, 每个输出一个标记的描述度
 - 用BP算法最小化神经网络的输出与真实标记分布间的误差平方和



专用算法

- 假设 $p(y|\mathbf{x}; \boldsymbol{\theta})$ 为最大熵模型 (MaxEnt Model)

$$p(y|\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{Z} \exp \left(\sum_k \theta_{y,k} g_k(\mathbf{x}) \right) \quad Z = \sum_y \exp \left(\sum_k \theta_{y,k} g_k(\mathbf{x}) \right)$$



$$\begin{aligned} \boldsymbol{\theta}^* &= \operatorname{argmin}_{\boldsymbol{\theta}} \sum_i \sum_j \left(d_{\mathbf{x}_i}^{y_j} \ln \frac{d_{\mathbf{x}_i}^{y_j}}{p(y_j|\mathbf{x}_i; \boldsymbol{\theta})} \right) \\ &= \operatorname{argmax}_{\boldsymbol{\theta}} \sum_i \sum_j d_{\mathbf{x}_i}^{y_j} \ln p(y_j|\mathbf{x}_i; \boldsymbol{\theta}). \end{aligned}$$



$$T(\boldsymbol{\theta}) = \sum_{i,j} d_{\mathbf{x}_i}^{y_j} \sum_k \theta_{y_j,k} g_k(\mathbf{x}_i) - \sum_i \ln \sum_j \exp \left(\sum_k \theta_{y_j,k} g_k(\mathbf{x}_i) \right)$$



专用算法

- **IIS-LLD** [Geng, Yin, and Zhou, TPAMI'13]
[Geng, Smith-Miles, and Zhou, AAI'10]

Algorithm 1: IIS-LLD

Input: The training set $S = \{(\mathbf{x}_i, D_i)\}_{i=1}^n$ and the convergence criterion ε

Output: $p(y|\mathbf{x}; \boldsymbol{\theta})$

- 1 Initialize the model parameter vector $\boldsymbol{\theta}^{(0)}$;
- 2 $l \leftarrow 0$;
- 3 **repeat**
- 4 $l \leftarrow l + 1$;
- 5 Solve Eq. (4) for $\delta_{y,k}$;
- 6 $\boldsymbol{\theta}^{(l)} \leftarrow \boldsymbol{\theta}^{(l-1)} + \boldsymbol{\Delta}$;
- 7 **until** $T(\boldsymbol{\theta}^{(l)}) - T(\boldsymbol{\theta}^{(l-1)}) < \varepsilon$;
- 8 $p(y|\mathbf{x}; \boldsymbol{\theta}) \leftarrow \frac{1}{Z} \exp\left(\sum_k \theta_{y,k}^{(l)} g_k(\mathbf{x})\right)$;

$$\sum_i p(y_j|\mathbf{x}_i; \boldsymbol{\theta}) g_k(\mathbf{x}_i) \exp(\delta_{y_j,k} s(g_k(\mathbf{x}_i)) g^\#(\mathbf{x}_i)) \quad (4)$$

$$- \sum_i d_{\mathbf{x}_i}^{y_j} g_k(\mathbf{x}_i) = 0,$$



专用算法

- **BFGS-LLD**
[Geng, TKDE'16]

$$\frac{\partial T'(\boldsymbol{\theta})}{\partial \theta_{y_j, k}} = \sum_i \frac{\exp\left(\sum_k \theta_{y_j, k} g_k(\mathbf{x}_i)\right) g_k(\mathbf{x}_i)}{\sum_j \exp\left(\sum_k \theta_{y_j, k} g_k(\mathbf{x}_i)\right)} - \sum_i d_{\mathbf{x}_i}^{y_j} g_k(\mathbf{x}_i). \quad (14)$$

Algorithm 2: BFGS-LLD

Input: The training set $S = \{(\mathbf{x}_i, D_i)\}_{i=1}^n$ and the convergence criterion ε

Output: $p(y|\mathbf{x}; \boldsymbol{\theta})$

- 1 Initialize the model parameter vector $\boldsymbol{\theta}^{(0)}$;
 - 2 Initialize the inverse Hessian approximation $\mathbf{B}^{(0)}$;
 - 3 Compute $\nabla T'(\boldsymbol{\theta}^{(0)})$ by Eq. (14);
 - 4 $l \leftarrow 0$;
 - 5 **repeat**
 - 6 Compute search direction $\mathbf{p}^{(l)} \leftarrow -\mathbf{B}^{(l)} \nabla T'(\boldsymbol{\theta}^{(l)})$;
 - 7 Compute the step length $\alpha^{(l)}$ by a line search procedure to satisfy Eq. (8) and (9);
 - 8 $\boldsymbol{\theta}^{(l+1)} \leftarrow \boldsymbol{\theta}^{(l)} + \alpha^{(l)} \mathbf{p}^{(l)}$;
 - 9 Compute $\nabla T'(\boldsymbol{\theta}^{(l+1)})$ by Eq. (14);
 - 10 $\mathbf{s}^{(l)} \leftarrow \boldsymbol{\theta}^{(l+1)} - \boldsymbol{\theta}^{(l)}$;
 - 11 $\mathbf{u}^{(l)} \leftarrow \nabla T'(\boldsymbol{\theta}^{(l+1)}) - \nabla T'(\boldsymbol{\theta}^{(l)})$;
 - 12 $\boldsymbol{\rho}^{(l)} \leftarrow \frac{1}{\mathbf{s}^{(l)\top} \mathbf{u}^{(l)}}$;
 - 13 $\mathbf{B}^{(l+1)} \leftarrow (\mathbf{I} - \boldsymbol{\rho}^{(l)} \mathbf{s}^{(l)} (\mathbf{u}^{(l)})^\top) \mathbf{B}^{(l)} (\mathbf{I} - \boldsymbol{\rho}^{(l)} \mathbf{u}^{(l)} (\mathbf{s}^{(l)})^\top) + \boldsymbol{\rho}^{(l)} \mathbf{s}^{(l)} (\mathbf{s}^{(l)})^\top$;
 - 14 $l \leftarrow l + 1$;
 - 15 **until** $\|\nabla T'(\boldsymbol{\theta}^{(l)})\| < \varepsilon$;
 - 16 $p(y|\mathbf{x}; \boldsymbol{\theta}) \leftarrow \frac{1}{Z} \exp\left(\sum_k \theta_{y, k} g_k(\mathbf{x})\right)$;
-



专用算法 – 更多...

- 两种算法设计路线

- **分类**

- 用条件质量函数 $p(y|x)$ 为示例 x 到标记分布 D 之间的映射建模

- IIS-LLD [Geng, Smith-Miles and Zhou, AAAI'10]
 - CPNN [Geng, Yin and Zhou, TPAMI'13]
 - BFGS-LLD [Geng, TKDE'16]
 - SCE-LDL [Yang, Geng and Zhou, IJCAI'16]

- **回归**

- 用多元回归为示例 x 到标记分布 D 之间的映射建模

- LDSVR [Geng and Hou, IJCAI'15]
 - LDLogitBoost [Xing, Geng and Xue, CVPR'16]

算法代码下载地址：

<http://cse.seu.edu.cn/PersonalPage/xgeng/LDL>

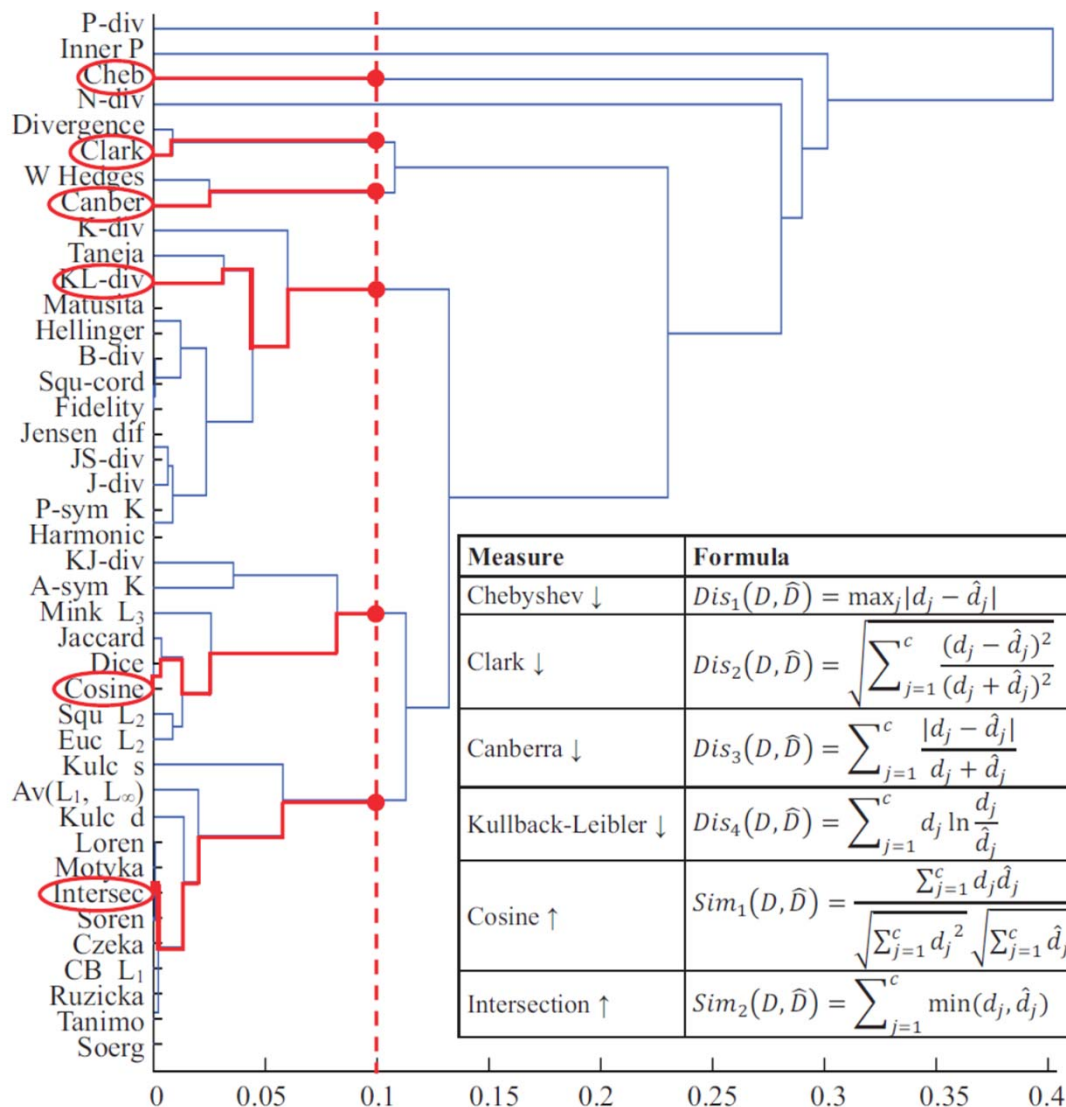
LDL算法评价指标 [Geng, TKDE'16]

- 预测标记分布与真实标记分布之间的平均距离/相似度

从41种度量中，通过一个单连接聚合层次聚类过程选择出6种度量

选择准则：

1. 任意两个度量之间距离大于0.1；
2. 每个度量来自不同语法族；
3. 一般不易受到不稳定情况（如分母为0）影响；
4. 常见度量





报告内容

- **研究背景**
- **概念定义**
- **学习算法**
- **实验**
- **结论**



人造数据（可视化）

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

$$\psi_1 = (w_1^T t)^2,$$

$$\psi_2 = (w_2^T t + \lambda_1 \psi_1)^2,$$

$$\psi_3 = (w_3^T t + \lambda_2 \psi_2)^2,$$

$$d_x^{y_i} = \frac{\psi_i}{\psi_1 + \psi_2 + \psi_3}, i = 1, \dots, 3,$$

$$t = [t_1, t_2, t_3]^T$$

$$a = 1, \quad b = 0.5, \quad c = 0.2, \quad d = 1, \quad w_1 = [4, 2, 1]^T, \quad w_2 = [1, 2, 4]^T$$

$$w_3 = [1, 4, 2]^T, \quad \lambda_1 = \lambda_2 = 0.01.$$

训练集:

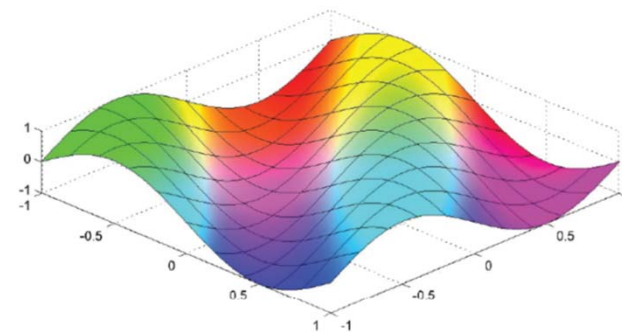
x 的每个分量在 $[-1, 1]$ 内均匀随机采样500次

测试集:

x_1, x_2 分布于 $[-1, 1]$ 内间隔为0.01的网格上
(40401个测试样本)

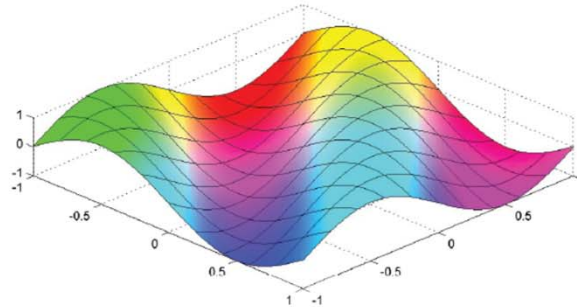
$$x_3 = \sin((x_1 + x_2) \times \pi)$$

三元标记分布转化为RGB颜色

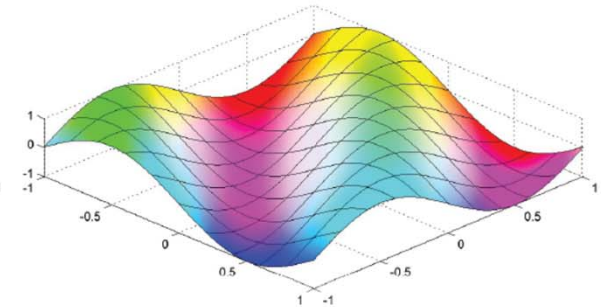


(a) Ground-Truth

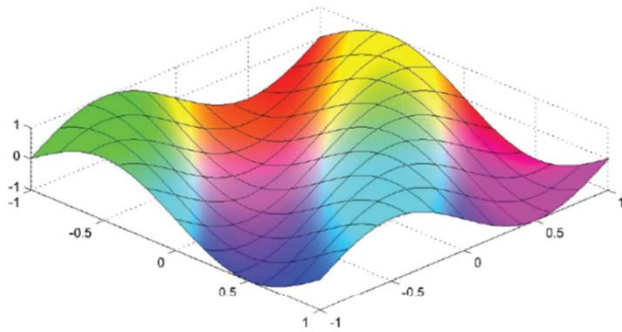
人造数据（定性分析）



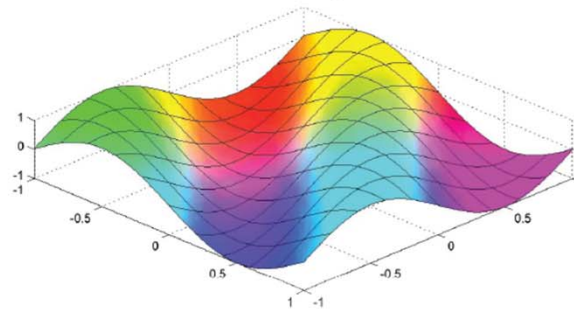
(b) PT-Bayes



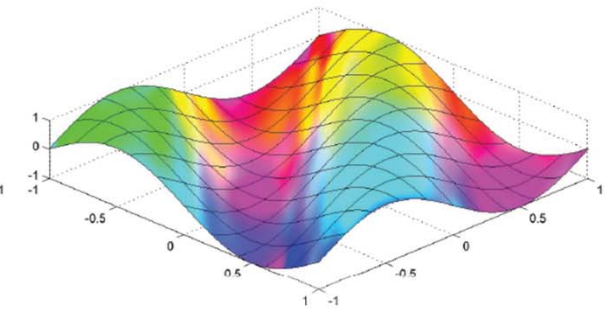
(c) PT-SVM



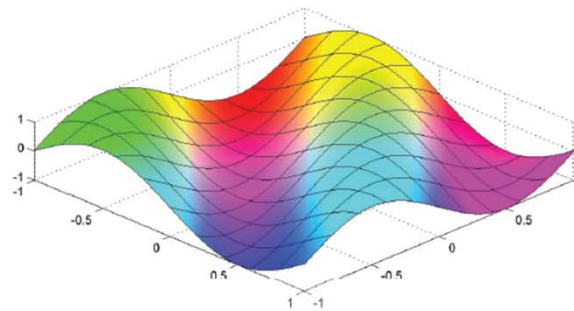
(a) Ground-Truth



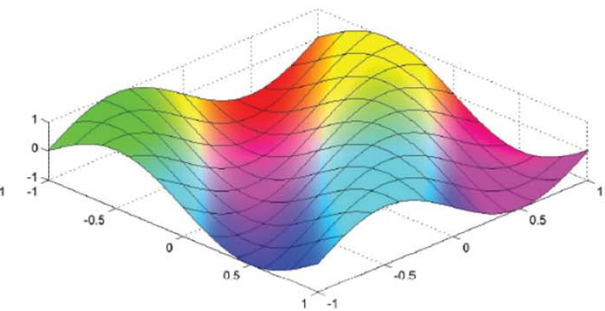
(d) AA- k NN



(e) AA-BP



(f) SA-IIS



(g) SA-BFGS



人造数据（定量分析）

Experimental Results on the Artificial Dataset

Criterion	PT-Bayes	PT-SVM	AA-kNN	AA-BP	SA-IIS	SA-BFGS
Chebyshev ↓	0.080(3)	0.653(6)	0.086(4)	0.101(5)	0.0767(2)	0.0766(1)
Clark ↓	0.341(1)	1.135(6)	0.382(4)	0.520(5)	0.349(2)	0.352(3)
Canberra ↓	0.488(1)	1.823(6)	0.564(4)	0.699(5)	0.489(2)	0.495(3)
Kullback-Leibler ↓	0.030(3)	1.482(6)	0.035(4)	0.066(5)	0.029(1)	0.030(2)
Cosine ↑	0.990(3)	0.377(6)	0.989(4)	0.983(5)	0.99116(2)	0.99120(1)
Intersection ↑	0.920(3)	0.347(6)	0.914(4)	0.899(5)	0.9233(2)	0.9234(1)
Avg. Rank	2.33	6.00	4.00	5.00	1.83	1.83
Running Time (ms)	22/45	391/1,153	0/79,961	101,568/149	1,168/33	187/33

观察：

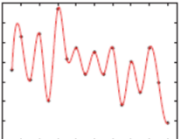
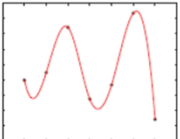
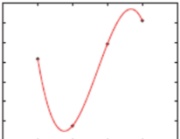
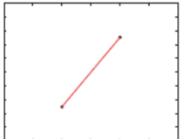
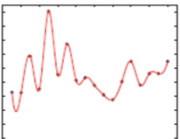
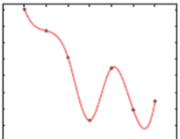
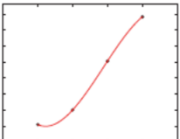
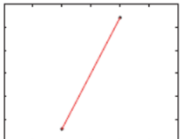
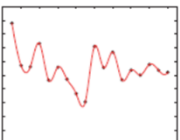
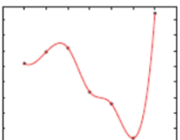
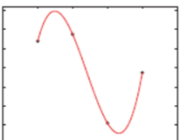
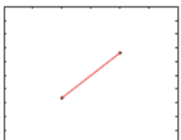
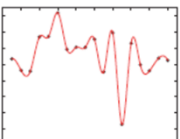
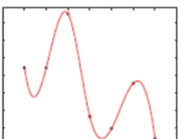
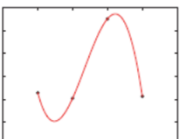
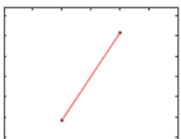
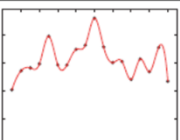
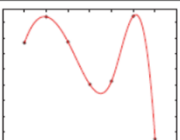
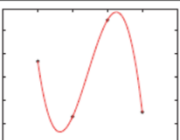
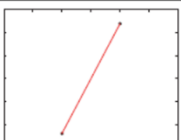
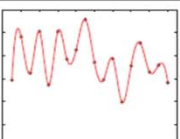
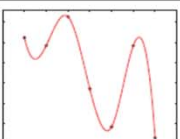
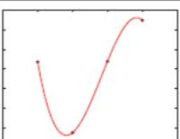

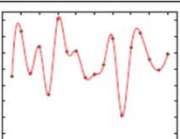
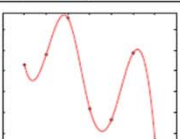
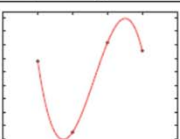
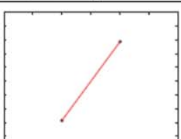
1. 同一算法在不同评价指标上的表现不同；
2. 平均排序：SA-BFGS = SA-IIS > PT-Bayes > AA-kNN > AA-BP > PT-SVM;
3. SA-BFGS和PT-Bayes较为高效



真实世界数据

No.	数据集	样例数	特征数	标记数
1	Yeast-alpha	2,465	24	18
2	Yeast-cdc	2,465	24	15
3	Yeast-elu	2,465	24	14
4	Yeast-diau	2,465	24	7
5	Yeast-heat	2,465	24	6
6	Yeast-spo	2,465	24	6
7	Yeast-cold	2,465	24	4
8	Yeast-dtt	2,465	24	4
9	Yeast-spo5	2,465	24	3
10	Yeast-spoem	2,465	24	2
11	Human Gene	30,542	36	68
12	Natural Scene	2,000	294	9
13	SJAFPE	213	243	6
14	SBU_3DFE	2,500	243	6
15	Movie	7,755	1,869	5

典型 预测 结果 举例

	18 Labels	7 Labels	4 Labels	2 Labels
Real				
PT-Bayes	 {0.048, 0.552, 1.885, 0.036} {0.960, 0.891}	 {0.042, 0.230, 0.492, 0.016} {0.984, 0.929}	 {0.170, 0.516, 0.911, 0.128} {0.907, 0.786}	 {0.192, 0.310, 0.421, 0.086} {0.939, 0.808}
PT-SVM	 {0.006, 0.105, 0.368, 0.0012} {0.9988, 0.980}	 {0.040, 0.190, 0.359, 0.011} {0.990, 0.949}	 {0.041, 0.108, 0.188, 0.006} {0.995, 0.954}	 {0.018, 0.025, 0.036, 0.001} {0.999, 0.982}
AA-kNN	 {0.005, 0.091, 0.282, 0.0009} {0.9991, 0.9844}	 {0.028, 0.151, 0.381, 0.007} {0.993, 0.945}	 {0.030, 0.080, 0.124, 0.003} {0.997, 0.970}	 {0.058, 0.085, 0.118, 0.007} {0.994, 0.942}
AA-BP	 {0.008, 0.156, 0.583, 0.003} {0.997, 0.968}	 {0.014, 0.076, 0.172, 0.002} {0.998, 0.975}	 {0.036, 0.085, 0.145, 0.004} {0.996, 0.964}	 {0.069, 0.102, 0.143, 0.010} {0.991, 0.931}
SA-IIS	 { 0.004 , 0.083 , 0.283, 0.00077 } { 0.99923 , 0.9843}	 {0.013, 0.072, 0.160, 0.002} {0.998, 0.977}	 {0.016, 0.042, 0.067, 0.0008} {0.9992, 0.984}	 {0.012, 0.018, 0.025, 0.0003} {0.9997, 0.988}
SA-BFGS	 {0.006, 0.086, 0.260 , 0.00081} {0.99919, 0.986 }	 { 0.012 , 0.056 , 0.120 , 0.001 } { 0.999 , 0.983 }	 { 0.014 , 0.034 , 0.055 , 0.0006 } { 0.9994 , 0.987 }	 { 0.007 , 0.010 , 0.014 , 0.0001 } { 0.9999 , 0.993 }





真实世界数据（定量分析）

Experimental Results (mean \pm std(rank)) on the Real-World Datasets Measured by Kullback-Leibler Divergence \downarrow

Dataset	PT-Bayes	PT-SVM	AA- k NN	AA-BP	SA-IIS	SA-BFGS
Yeast-alpha	0.719 \pm 0.080(6)	0.009 \pm 0.002(4)	0.0066 \pm 0.001(2)	0.081 \pm 0.011(5)	0.0067 \pm 0.001(3)	0.006 \pm 0.001(1)
Yeast-cdc	0.603 \pm 0.073(6)	0.010 \pm 0.002(4)	0.0083 \pm 0.001(3)	0.060 \pm 0.007(5)	0.0082 \pm 0.001(2)	0.007 \pm 0.001(1)
Yeast-elu	0.556 \pm 0.071(6)	0.008 \pm 0.001(4)	0.0074 \pm 0.0004(3)	0.051 \pm 0.009(5)	0.0073 \pm 0.0005(2)	0.006 \pm 0.0004(1)
Yeast-diau	0.306 \pm 0.036(6)	0.019 \pm 0.002(4)	0.015 \pm 0.001(3)	0.024 \pm 0.004(5)	0.014 \pm 0.001(2)	0.013 \pm 0.001(1)
Yeast-heat	0.255 \pm 0.040(6)	0.0148 \pm 0.001(4)	0.0145 \pm 0.001(3)	0.021 \pm 0.004(5)	0.0133 \pm 0.0004(2)	0.0126 \pm 0.0005(1)
Yeast-spo	0.281 \pm 0.031(6)	0.0304 \pm 0.005(4)	0.0302 \pm 0.002(3)	0.034 \pm 0.006(5)	0.0254 \pm 0.003(2)	0.0246 \pm 0.003(1)
Yeast-cold	0.208 \pm 0.031(6)	0.0147 \pm 0.001(4)	0.014 \pm 0.001(3)	0.0149 \pm 0.002(5)	0.013 \pm 0.001(2)	0.012 \pm 0.001(1)
Yeast-dtt	0.206 \pm 0.029(6)	0.0073 \pm 0.001(4)	0.0072 \pm 0.001(3)	0.009 \pm 0.001(5)	0.0070 \pm 0.001(2)	0.006 \pm 0.001(1)
Yeast-spo5	0.214 \pm 0.025(6)	0.03010 \pm 0.003(3)	0.033 \pm 0.003(5)	0.031 \pm 0.003(4)	0.03007 \pm 0.003(2)	0.029 \pm 0.003(1)
Yeast-spoem	0.190 \pm 0.038(6)	0.0280 \pm 0.004(4)	0.0285 \pm 0.003(5)	0.026 \pm 0.003(3)	0.025 \pm 0.003(2)	0.024 \pm 0.003(1)
Human Gene	1.887 \pm 0.766(6)	0.240 \pm 0.019(3)	0.301 \pm 0.026(4)	0.500 \pm 0.068(5)	0.238 \pm 0.019(2)	0.236 \pm 0.019(1)
Natural Scene	3.065 \pm 0.487(6)	1.447 \pm 0.243(4)	2.767 \pm 0.137(5)	0.875 \pm 0.029(3)	0.870 \pm 0.026(2)	0.854 \pm 0.062(1)
s-JAFFE	0.074 \pm 0.014(4)	0.086 \pm 0.016(5)	0.071 \pm 0.023(3)	0.113 \pm 0.030(6)	0.070 \pm 0.012(2)	0.064 \pm 0.016(1)
s-BU_3DFE	0.079 \pm 0.004(4)	0.089 \pm 0.007(6)	0.065 \pm 0.002(2)	0.085 \pm 0.009(5)	0.068 \pm 0.004(3)	0.049 \pm 0.002(1)
Movie	0.953 \pm 0.352(6)	0.268 \pm 0.079(5)	0.201 \pm 0.011(4)	0.179 \pm 0.03(3)	0.137 \pm 0.013(1)	0.140 \pm 0.020(2)
Avg. Rank	5.73	4.13	3.40	4.60	2.07	1.07



真实世界数据（定量分析）

- 算法比较（6个指标上平均排序一致）

SA-BFGS > SA-IIS > AA-kNN > PT-SVM > AA-BP > PT-Bayes

- 专用算法（SA-BFGS和SA-IIS）比传统算法转化（PT）或改造（AA）来的算法更好；
- SA-BFGS比SA-IIS更好一些；
- PT-Bayes的Gaussian假设可能不适用真实世界数据；
- AA-BP较容易过配；
- AA-kNN保留了标记分布的整体性，而PT-SVM破坏了这种整体性

实际应用

	<p>Facial Age Estimation</p> <ul style="list-style-type: none"> X. Geng, Q. Wang, and Y. Xia. Facial Age Estimation by Adaptive Learning. <i>Proceedings of the 2014 IEEE International Conference on Computer Vision (ICCV'14)</i>, Stockholm, Sweden, 2014, pp. 4465 - 4470. X. Geng, C. Yin, and Z.-H. Zhou. Facial Age Estimation by Adaptive Learning. <i>Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV'13)</i>, 2013, pp. 2401-2412. X. Geng, K. Smith-Miles, Z.-H. Zhou. Facial Age Estimation by Adaptive Learning. <i>Proceedings of the 2010 IEEE International Conference on Systems, Man, and Cybernetics (SMC'10)</i>, Atlanta, GA, 2010, pp. 451-456. 	<h2>人脸年龄估计</h2>																						
	<p>Head Pose Estimation</p> <ul style="list-style-type: none"> X. Geng and Y. Xia. Head Pose Estimation Based on Multi-View Learning. <i>Proceedings of the 2014 IEEE International Conference on Computer Vision (ICCV'14)</i>, Columbus, OH, 2014, pp. 1837-1842. 	<h2>头部姿态估计</h2>																						
	<p>Pre-release Prediction of Movies</p> <ul style="list-style-type: none"> X.Geng and P.Hou. Pre-release Prediction of Crowd Opinion on Movies by Label Distribution Learning. In: <i>Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'15)</i>, Buenos Aires, Argentina, 2015, 3511-3517. 	<h2>电影评价预测</h2>																						
 <table border="1"> <thead> <tr> <th>Ranker</th> <th>Relevant Label Ranking</th> </tr> </thead> <tbody> <tr><td>01</td><td>water > cloud > sky</td></tr> <tr><td>02</td><td>cloud > sky > water > building</td></tr> <tr><td>03</td><td>water > cloud > sky > building</td></tr> <tr><td>04</td><td>water > cloud > sky</td></tr> <tr><td>05</td><td>water > sky > cloud</td></tr> <tr><td>06</td><td>water > cloud > sky</td></tr> <tr><td>07</td><td>water > cloud > sky</td></tr> <tr><td>08</td><td>sky > water</td></tr> <tr><td>09</td><td>water > cloud</td></tr> <tr><td>10</td><td>water > cloud > sky</td></tr> </tbody> </table>	Ranker	Relevant Label Ranking	01	water > cloud > sky	02	cloud > sky > water > building	03	water > cloud > sky > building	04	water > cloud > sky	05	water > sky > cloud	06	water > cloud > sky	07	water > cloud > sky	08	sky > water	09	water > cloud	10	water > cloud > sky	<p>Multi-label Ranking</p> <ul style="list-style-type: none"> X. Geng and L.-L. Luo. Multilabel Ranking with Inconsistent Rankers. In: <i>Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'14)</i>, Columbus, OH, 2014, pp. 3742-3747. 	<h2>自然场景图像多标记排序</h2>
Ranker	Relevant Label Ranking																							
01	water > cloud > sky																							
02	cloud > sky > water > building																							
03	water > cloud > sky > building																							
04	water > cloud > sky																							
05	water > sky > cloud																							
06	water > cloud > sky																							
07	water > cloud > sky																							
08	sky > water																							
09	water > cloud																							
10	water > cloud > sky																							
	<p>Emotion Distribution Recognition</p> <ul style="list-style-type: none"> X.Geng and P.Hou, Y. Zhou, H. Xue and X. Geng, Emotion Distribution Recognition from Facial Expressions. In: <i>Proceedings of the 23rd ACM International Conference on Multimedia (ACM MM'15)</i>, Brisbane, Australia, 2015, 1247-1250. 	<h2>情感分布识别</h2>																						
	<p>Multi-label Learning</p> <ul style="list-style-type: none"> P. Hou, X. Geng and M.-L. Zhang. Multi-Label Manifold Learning. In: <i>Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)</i>, Phoenix, AZ, 2016, in press. Y.-K. Li, M.-L. Zhang and X. Geng. Leveraging implicit relative labeling-importance information for effective multi-label learning. In: <i>Proceedings of the 15th IEEE International Conference on Data Mining (ICDM'15)</i>, Atlantic City, NJ, 2015, 251-260. 																							
	<p>Crowd Counting</p> <ul style="list-style-type: none"> Z. Zhang, M. Wang, X. Geng. Crowd Counting in Public Video Surveillance by Label Distribution Learning. <i>Neurocomputing</i>, 2015, vol. 166: 151-163. 	<h2>人群计数</h2>																						



报告内容

- **研究背景**
- **概念定义**
- **学习算法**
- **实验**
- **结论**





总结

- **标记分布学习**

- 是一种比传统单标记和多标记学习更为泛化的学习范式
- 能够处理标记的不同重要程度（描述度）
- 对某些实际问题匹配的更好
- 需要专门的算法设计

- **标记分布学习适用于**

- 数据本身具有某种天然的描述度度量
- 标记之间有较强相关性
- 同一示例由多个标注源标注并产生不一致性
- 同一示例与多个标记相关，且标记重要程度不同
-



对LDL感兴趣？

所有**LDL论文**、**算法代码**和**数据集**可从如下地址下载：
<http://cse.seu.edu.cn/PersonalPage/xgeng/LDL>

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

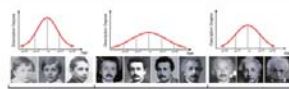
In academic papers using our codes and data, the following references will be appreciated:

[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.



参考文献

1. D. Zhou, Y. Zhou, X. Zhang, Q. Zhao and X. Geng. Emotion Distribution Learning from Texts. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'16)*, Austin, TX, 2016, in press.
2. X. Geng. Label Distribution Learning. *IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)*, 2016, 28(7): 1734-1748.
3. X. Yang, X. Geng and D.-Y. Zhou. Sparsity Conditional Energy Label Distribution Learning for Age Estimation. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'16)*, New York City, NY, 2016, 2259-2265.
4. C. Xing and X. Geng. Logistic Boosting Regression for Label Distribution Learning. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'16)*, Las Vegas, NV, 2016, 4489-4497.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.



致谢

- 合作者



Prof. Zhi-Hua Zhou



Prof. Kate Smith-Miles



Prof. Min-Ling Zhang



Dr. Deyu Zhou



Dr. Hui Xue

- 学生

Chao Yin
Peng Hou
Rongzi Ji

Yu-Kun Li
Yu Xia
Chao Xing

Ying Zhou
Longrun Luo
Xu Yang



谢谢！



[http:// palm.seu.edu.cn](http://palm.seu.edu.cn)