



# 智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

## 三角距离相关性的标签分布学习

黄雨婷, 徐媛媛, 张恒汝, 闵帆

引用本文:

黄雨婷, 徐媛媛, 张恒汝, 等. 三角距离相关性的标签分布学习[J]. 智能系统学报, 2021, 16(3): 449–458.

HUANG Yuting, XU Yuanyuan, ZHANG Hengru, et al. Label distribution learning based on triangular distance correlation[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(3): 449–458.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202001027>

## 您可能感兴趣的其他文章

### 一种双优选的半监督回归算法

A dual-optimal semi-supervised regression algorithm

智能系统学报. 2019, 14(4): 689–696 <https://dx.doi.org/10.11992/tis.201805010>

### 鲁棒的半监督多标签特征选择方法

A robust, semi-supervised, and multi-label feature selection method

智能系统学报. 2019, 14(4): 812–819 <https://dx.doi.org/10.11992/tis.201809017>

### 一种具有迁移学习能力的RBF-NN算法及其应用

A RBF-NN algorithm with transfer learning ability and its application

智能系统学报. 2018, 13(6): 959–966 <https://dx.doi.org/10.11992/tis.201705021>

### 基于Spark的多标签超网络集成学习

Multi-label hypernetwork ensemble learning based on Spark

智能系统学报. 2017, 12(5): 624–639 <https://dx.doi.org/10.11992/tis.201706033>

### 基于混合距离学习的鲁棒的模糊C均值聚类算法

Robust FCM clustering algorithm based on hybrid-distance learning

智能系统学报. 2017, 12(4): 450–458 <https://dx.doi.org/10.11992/tis.201607019>

### 基于最小最大概率机的迁移学习分类算法

Transfer learning classification algorithms based on minimax probability machine

智能系统学报. 2016, 11(1): 84–92 <https://dx.doi.org/10.11992/tis.201505024>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202001027

## 三角距离相关性的标签分布学习

黄雨婷, 徐媛媛, 张恒汝, 闵帆

(西南石油大学 计算机科学学院, 四川 成都 610500)

**摘要:** 针对标签相关性的表征问题, 提出一种基于三角距离相关性的标签分布学习算法。首先, 构建距离映射矩阵, 描述标签分布和特征矩阵之间的映射关系。其次, 设计新的三角距离, 以表征标签之间的相关性。最后, 结合标签相关性, 设计基于 Kullback-Leibler 散度的目标函数。在 8 个数据集上的实验结果表明, 与 8 种主流算法相比, 本文提出的算法在 6 个准确性指标上占优势。

**关键词:** 标签分布学习; 标签相关性; 三角距离; 距离映射矩阵; 多标签学习; 最大熵模型; Kullback-Leibler 散度; L-BFGS 方法

中图分类号: TP391 文献标志码: A 文章编号: 1673-4785(2021)03-0449-10

中文引用格式: 黄雨婷, 徐媛媛, 张恒汝, 等. 三角距离相关性的标签分布学习 [J]. 智能系统学报, 2021, 16(3): 449-458.

英文引用格式: HUANG Yuting, XU Yuanyuan, ZHANG Hengru, et al. Label distribution learning based on triangular distance correlation[J]. CAAI transactions on intelligent systems, 2021, 16(3): 449-458.

## Label distribution learning based on triangular distance correlation

HUANG Yuting, XU Yuanyuan, ZHANG Hengru, MIN Fan

(College of Computer Science, Southwest Petroleum University, Chengdu 610500, China)

**Abstract:** Aiming at the representation problem of label correlation, a label distribution learning algorithm based on triangular distance correlation is proposed in this paper. First, a distance-mapping matrix is constructed to describe the mapping relationship between the label distribution and the feature matrix. Then a new triangle distance is designed to characterize the correlation between the labels. Finally, based on the label correlation, the Kullback-Leibler divergence-based objective function is designed. Results on eight datasets show that the proposed algorithm is superior in six evaluation measures in terms of accuracy compared with eight mainstream algorithms.

**Keywords:** label distribution learning; label correlation; triangular distance; distance mapping matrix; multi-label learning; maximum entropy model; Kullback-Leibler divergence; L-BFGS method

标签分布学习 (label distribution learning, LDL) 是多标签学习 (multi-label learning, MLL) 的泛化<sup>[1-8]</sup>。MLL 用标签集的部分标签来描述实例<sup>[9-11]</sup>, LDL 用标签集所有标签的表征程度构成的分布来描述实例<sup>[12-15]</sup>。文献 [12] 将年龄估计问题泛化到 LDL 中, 降低了平均绝对误差 (mean absolute deviation, MAE)。文献 [13] 将人群计数问题泛化到 LDL 中, 提高了人群计数的准确率。

Geng 等<sup>[1]</sup>提出了 SA-IIS(specialized algorithm improiothm lternative scaling) 算法, 将单个标签数据转换为分布数据, 但未考虑标签的相关性。Jia 等<sup>[16]</sup>

提出了 LDLLC(label distribution learning by exploiting label correlation) 算法, 使用皮尔逊相关系数描述了标签之间的相关性。Zheng 等<sup>[17]</sup>提出了 LDL-SCL(label distribution learning by exploiting sample correlation locally) 算法, 考虑实例之间的相关性。后 2 种方法显著提高了模型对标签分布的预测能力。

本文提出了一种三角距离相关性的标签分布学习算法 (label distribution learning based on triangular distance correlation, T-LDL)。首先, 令  $X$  和  $D$  分别表示特征矩阵和标签分布矩阵, 构建距离映射矩阵  $\theta$  描述  $X$  和  $D$  之间的映射关系。其次, 设计新的相似度距离, 以表征标签之间的相关性。最后, 结合标签相关性, 设计基于 KL (kullback-lei-

收稿日期: 2020-01-20.

基金项目: 国家自然科学基金项目 (61902328).

通信作者: 张恒汝. E-mail: zhanghrswpu@163.com.

bler divergence) 散度<sup>[18]</sup>的目标函数, 利用从训练数据直接获取的  $X$  和  $D$  拟合  $\theta$  以预测标签分布。

在8个真实数据集上, 将本文提出算法与8种主流算法进行对比实验, 利用 Euclidean 距离<sup>[19]</sup>、Sørensen 距离<sup>[20]</sup>、Squard $\chi^2$  距离<sup>[21]</sup>、KL 散度<sup>[18]</sup>、Intersection 相似度<sup>[22]</sup>和 Fidelity 相似度<sup>[23]</sup>共6种指标进行评价。结果表明, 本文提出的算法在其中3个数据集上所有指标均为最优, 在其余的数据集上部分指标占优。

## 1 相关工作

首先提出 LDL 的问题描述与运行实例, 然后讨论流行的 LDL 算法及其目标函数。表1列出了本文的符号系统。

表1 符号系统  
Table 1 Notations

符号	含义
$R^q$	$q$ 维的输入空间
$Y$	完整标签集
$S$	训练集
$x_i$	第 $i$ 个实例
$d_i$	$x_i$ 对应的实际标签分布
$p_i$	$x_i$ 对应的预测标签分布
$x_{ir}$	第 $i$ 个实例的第 $r$ 个特征
$d_{ij}$	第 $j$ 个标签对 $x_i$ 的实际表征度
$\theta$	距离映射矩阵
$X$	特征矩阵
$D$	实际标签分布矩阵

### 1.1 LDL 问题描述

标签分布学习相对于单标签和多标签学习而言, 以一种更自然的方式去标记实例, 并且为它的每个可能的标签分配一个数值。下面给出它的形式化定义<sup>[1]</sup>。令  $X = R^q$  为  $q$  维输入空间, 表示特征矩阵;  $Y = \{y_1, y_2, \dots, y_c\}$  为完整标签集,  $c$  为标签的数量;  $D$  表示实际标签分布矩阵; 给定一个训练集  $S = \{X, D\} = \{\{x_1, d_1\}, \{x_2, d_2\}, \dots, \{x_n, d_n\}\}$ , 其中  $x_i = [x_{i1} x_{i2} \dots x_{iq}] \in X$  为第  $i$  个实例,  $d_i = [d_{i1} d_{i2} \dots d_{ic}] \in [0, 1]^c$  为  $x_i$  对应的实际标签分布,  $d_{ij}$  是标签  $y_j$  对  $x_i$  的实际表征度, 且  $\sum_{j=1}^c d_{ij} = 1$ 。

### 1.2 运行实例

图1(a)为需要标记的一个示例图片<sup>[24]</sup>, 其完整标签集为{森林, 海洋, 沙漠, 城市}。图1(b)表明 MLL 中仅有{海洋, 城市}2个标签能够描述图1(a)。图1(c)说明 LDL 利用这4个标签构成的分布来描述该图片, 且{海洋, 城市}2个标签对图1(a)的

表征度较高, {森林, 沙漠}2个标签对图1(a)的表征度较低。



(a) 示例图

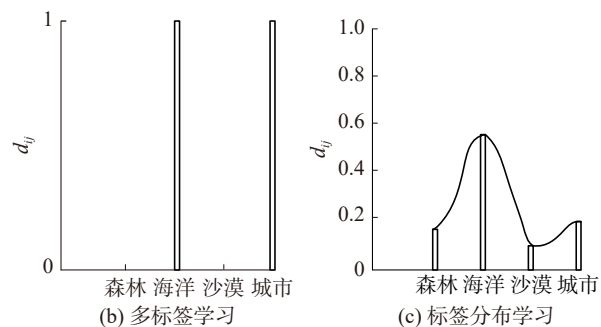


图1 MLL 与 LDL 的比较

Fig. 1 Differences between MLL and LDL

表2和表3为一个标签分布学习的运行实例, 分别为特征矩阵  $X$  和实际标签分布矩阵  $D$ , 这里  $S = \{(x_1, d_1), (x_2, d_2), \dots, (x_4, d_4)\}$ ,  $q = 5$ ,  $c = 4$ 。{天空, 水, 房屋, 沙子, 树木}5个特征表征了图1(a)中包含的信息。{森林, 海洋, 城市, 沙漠}为完整标签集。以加粗行为例,  $x_1 = [0.38, 0.35, 0.00, 0.12, 0.15]$ ,  $d_1 = [0.16, 0.55, 0.10, 0.19]$ , 其中  $x_{11} = 0.38$  表示天空占图片面积的38%,  $d_{11} = 0.16$  表示森林描述该图片的程度为16%。

表2 特征矩阵  $X$   
Table 2 Feature matrix  $X$

特征分布	天空	水	房屋	沙子	树木
$x_1$	<b>0.38</b>	<b>0.35</b>	<b>0.00</b>	<b>0.12</b>	<b>0.15</b>
$x_2$	0.12	0.20	0.50	0.11	0.07
$x_3$	0.11	0.10	0.05	0.19	0.55
$x_4$	0.20	0.01	0.07	0.70	0.02

表3 标签分布矩阵  $D$   
Table 3 Label distribution matrix  $D$

类别	森林	海洋	沙漠	城市
$d_1$	<b>0.16</b>	<b>0.55</b>	<b>0.10</b>	<b>0.19</b>
$d_2$	0.14	0.20	0.10	0.56
$d_3$	0.85	0.05	0.03	0.07
$d_4$	0.10	0.05	0.76	0.09

$X$  和  $D$  之间的映射关系可以通过距离映射矩阵  $\theta$  来描述。给定训练集后, LDL 的目标为学习到该距离映射矩阵  $\theta$ <sup>[16]</sup>, 再通过  $\theta$  计算出预测标

签分布矩阵  $\mathbf{P} = \{p_1, p_2, \dots, p_i\}$ , 其中  $p_i = [p_{i1} p_{i2} \dots p_{ic}]$ ,  $p_{ij}$  为标签  $y_j$  对  $x_i$  的预测表征度, 该表征度用最大熵模型<sup>[25]</sup>表示, 如式 (1) 所示:

$$p(y_j|x_i; \theta) = \frac{\exp\left(\sum_{r=1}^q \theta_{kr} x_{ir}\right)}{\sum_{k=1}^c \exp\left(\sum_{r=1}^q \theta_{kr} x_{ir}\right)} \quad (1)$$

为优化求解  $\theta$ , LDL 算法的目标函数需约束预测分布与真实分布之间的差异。文献 [1] 构建了以 KL 散度为基础的目标函数, 通过求解式 (2), 可得到最优距离映射矩阵  $\theta^*$ , 即

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n \sum_{j=1}^c \left( d_{ij} \ln \frac{d_{ij}}{p(y_j|x_i; \theta)} \right) \quad (2)$$

### 1.3 已有的 LDL 算法

表 4 列出了 4 种流行的 LDL 算法及其目标函数, 表中第一行的 SA-IIS<sup>[1]</sup> 和 SA-BFGS(specialized algorithm effective quasi-newton)<sup>[1]</sup> 两种算法使用相同的目标函数, 它们均采用 KL 散度表征所有实例的真实分布与预测分布之间的差异。前者使用类似于改进迭代缩放的策略作为其优化方法, 后者使用 BFGS 算法作为其优化方法。该目标函数缺少正则项, 易导致欠拟合。

表 4 已有的 4 种流行的 LDL 算法及其目标函数

Table 4 Objective functions of four popular LDL algorithms

方法	目标函数
SA-IIS、SA-BFGS <sup>[1]</sup>	$\theta^* = \arg \min_{\theta} \sum_{i=1}^n \sum_{j=1}^c \left( d_{ij} \ln \frac{d_{ij}}{p(y_j x_i; \theta)} \right)$
LDLLC <sup>[16]</sup>	$\theta^* = \arg \min_{\theta} \sum_{j=1}^c \left( d_{ij} \ln \frac{d_{ij}}{p(y_j x_i; \theta)} \right) + \lambda_1 \ \theta\ _F^2 + \lambda_2 \sum_{i=1}^n \sum_{j=1}^c \text{sgn}(\rho_{\theta_i, \theta_j}) \text{dis}(\rho_{\theta_i, \theta_j})$
EDL <sup>[26]</sup>	$\theta^* = \arg \min_{\theta} 2 \sum_{i=1}^n \sum_{j=1}^c \frac{(d_{ij} - p(y_j x_i; \theta))^2}{(d_{ij} + p(y_j x_i; \theta))^2} + \frac{\lambda_1}{n} \sum_{k=1}^c \sum_{r=1}^q  \theta_{kr} _1 + \frac{\lambda_2}{n} \sum_{u=1}^c \sum_{j=1}^q \sum_{k=1}^q \omega_{jk} \ \theta_{uj} - \theta_{uk}\ _2^2$

LDLLC<sup>[16]</sup> 在 IIS-LLD 算法的目标函数基础上增加了正则项和标签相关性项。如表 4 中第 2 行所示, 等号右边第 2 项为距离映射矩阵  $\theta$  的 F-范数, 以防止过拟合。第 3 项为符号函数与不同距离共同决定的标签相关性项, 其中符号函数由皮尔逊相关系数决定。但皮尔逊相关系数存在“2 个输入向量间应有线性关系”的约束条件, 而距离映射矩阵  $\theta$  中的任意 2 个向量要满足该条件较为困难。

EDL(emotion distribution learning from texts)<sup>[26]</sup> 通过采用新散度公式表征所有实例的真实分布与预测分布之间的差异, 并增加 2 个约束项。如表 4 中第 3 行所示, 等号右边第 2 项为距离映射矩阵  $\theta$  的 1-范数, 以防止过拟合。第 3 项用不同标签的特征向量之差的 2-范数, 再乘以基于 Plutchik 的情绪轮得到的权重, 表征不同标签之间的关系。该算法在情绪分类场景下表现较好。

## 2 本文工作

常见的 LDL 算法的输入为特征矩阵  $\mathbf{X}$  与实际标签分布矩阵  $\mathbf{D}$ , 输出为预测标签分布矩阵  $\mathbf{P}$ , 构建距离映射矩阵  $\theta$  描述  $\mathbf{X}$  和  $\mathbf{D}$  之间的映射关系。为了得到更精准的预测标签分布矩阵  $\mathbf{P}$ , 设计目标函数是 LDL 算法工作的重点。本节重点介绍如何设计目标函数以及本文提出的 T-LDL 算法。

本文设计的目标函数为

$$T(\theta) = \sum_{i=1}^n \sum_{j=1}^c \left( d_{ij} \ln \frac{d_{ij}}{p(y_j|x_i; \theta)} \right) + \lambda_1 \sum_{i=1}^n \sum_{j=1}^c \eta(\theta_i, \theta_j) \quad (3)$$

式中: 等号右侧第 1 项用 KL 散度表征所有实例的真实分布与预测分布之间的差异; 等号右侧第二项为本文亮点, 设计标签相关性项以获得更好的预测结果。

### 2.1 标签相关性

本文的亮点为结合三元相关性和距离相关性来描述标签之间的相关性, 如式 (4) 所示:

$$\eta(\theta_i, \theta_j) = \text{sgn}(\text{triangle}(\theta_i, \theta_j)) \cdot \text{Dis}(\theta_i, \theta_j) \quad (4)$$

式中:  $\text{sgn}(\text{triangle}(\theta_i, \theta_j))$  表征三元相关性,  $\text{Dis}(\theta_i, \theta_j)$  表征距离相关性。  $\text{sgn}(\text{triangle}(\theta_i, \theta_j))$  用三角距离来表征标签之间存在何种相关性, 即正相关、不相关或负相关;  $\text{Dis}(\theta_i, \theta_j)$  用 Euclidean 距离<sup>[19]</sup> 表征标签之间的相关程度。

由于使用皮尔逊相关系数时需要考虑任意 2 个向量是否存在线性关系, 故提出一种不考虑该约束条件的新三角距离来衡量 2 个向量是否相关。这里, 仅考虑 2 个向量  $\theta_i, \theta_j$  以及 2 个向量之差  $\theta_i - \theta_j$ , 设计该三角距离, 且使得其取值范围为  $[-1, 1]$ , 如式 (5) 所示:

$$\text{triangle}(\theta_i, \theta_j) = 1 - \frac{2 \sqrt{\sum_{k=1}^m (\theta_{ik} - \theta_{jk})^2}}{\sqrt{\sum_{k=1}^m \theta_{ik}^2} + \sqrt{\sum_{k=1}^m \theta_{jk}^2}} \quad (5)$$



将该三角距离代入符号函数,用于判断标签之间存在何种相关性:正相关、不相关或负相关。

$$\text{sgn}(\text{triangle}(\theta_i, \theta_j)) = \begin{cases} 1, & 0 < \text{triangle}(\theta_i, \theta_j) \leq 1 \\ 0, & \text{triangle}(\theta_i, \theta_j) = 0 \\ -1, & -1 \leq \text{triangle}(\theta_i, \theta_j) < 0 \end{cases} \quad (6)$$

式中,  $\text{sgn}(\cdot)$  为 1、0、-1 分别表示标签之间为正相关、不相关或负相关。

由于上述部分只能判断标签之间存在何种相关性,并不能判断标签之间的相关程度,故引入 Euclidean 距离<sup>[19]</sup>表示标签之间的相关程度:

$$\text{Dis}(\theta_i, \theta_j) = \sqrt{\sum_{k=1}^m (\theta_{ik} - \theta_{jk})^2} \quad (7)$$

## 2.2 本文提出的 T-LDL 算法

T-LDL 描述见算法 1。首先将距离映射矩阵  $\theta^{(0)}$  和逆拟 Hessian 矩阵  $B^{(0)}$  初始化为单位矩阵,再通过式 (3) 计算初次目标函数的梯度  $\nabla T(\theta^{(0)})$ 。进入迭代,收敛条件为  $\|\nabla T(\theta^{(l)})\|_2 < \xi$ 。当不满足收敛条件时,采用 L-BFGS 方法<sup>[27]</sup>优化并更新  $\theta$  和  $B$ 。当满足收敛条件时,计算标签  $y_j$  对  $x_i$  的预测表征度  $p(y_j|x_i; \theta)$ 。

### 算法 1 T-LDL 算法

输入  $X, D, \xi$ ;

输出  $p(y|x; \theta)$ 。

1) 初始化距离映射矩阵  $\theta^{(0)}$  和逆拟 Hessian 矩阵  $B^{(0)}$ ;

2) 通过式 (3) 计算梯度  $\nabla T(\theta^{(0)})$ ;

3) 如果  $\|\nabla T(\theta^{(l)})\|_2 > \xi$ , 使用 L-BFGS 方法<sup>[27]</sup>优化更新  $\theta$  和  $B$ ;

4) end if;

5)  $l \leftarrow l + 1$ ;

6) 通过式 (1) 计算  $p(y_j|x_i; \theta)$ 。

## 3 实验及结果分析

本节首先介绍实验使用的 8 个数据集和 6 个评价指标,再将本文提出的 T-LDL 算法与 LDLLC<sup>[16]</sup>、PT-Bayes<sup>[1]</sup>、PT-SVM<sup>[1, 17]</sup>、AA-kNN<sup>[1, 4]</sup>、AA-BP<sup>[1]</sup>、SA-IIS<sup>[1, 16]</sup>、SA-BFGS(specialized algorithm effective quasi-newton)<sup>[1, 2]</sup>和 EDL<sup>[26]</sup>8 种主流的 LDL 算法进行比较,最后对实验结果进行讨论。

### 3.1 数据集

表 5 列出了从芽殖酵母的 8 个生物学实验中收集得到的 8 个真实数据集<sup>[28]</sup>。实例为 2 465 个酵母基因,特征是长度为 24 的系统发育谱,标签为不同生物实验中的离散时间点,数量范围为 4~18。

表 5 数据集  
Table 5 Datasets

数据集名称	实例个数	特征个数	标签个数
Alpha	2 465	24	18
Cdc	2 465	24	15
Elu	2 465	24	14
Diau	2 465	24	7
Heat	2 465	24	6
Spo	2 465	24	6
Cold	2 465	24	4
Dtt	2 465	24	4

Alpha 数据集记录在  $\alpha$  因子的影响下酵母在有丝分裂期间的基因表达情况; Cdc 数据集记录酵母在细胞分裂期间停滞的 cdc-15 基因表达情况; Elu 数据集记录酵母经离心淘洗后的基因表达情况; Diau 数据集记录酵母在双峰转换过程中的基因表达情况; Heat 数据集记录酵母在经过高温冲击后的基因表达情况; Spo 数据集记录酵母在孢子形成过程中的基因表达情况; Cold 数据集记录酵母经低温处理后的基因表达情况; Dtt 数据集记录酵母经还原剂处理后的基因表达情况<sup>[28]</sup>。

### 3.2 评价指标

表 6 列出了评估 LDL 算法的 6 个评价指标的名称和公式。其中,  $p_{ij}$  是标签  $y_j$  对  $x_i$  的预测表征度;  $d_{ij}$  是标签  $y_j$  对  $x_i$  的实际表征度; “↓”表示“越小越好”; “↑”表示“越大越好”。

表 6 LDL 算法的评价指标  
Table 6 Evaluation measures for the LDL algorithms

指标名称	公式
Euclidean <sup>[19]</sup> ↓	$\text{dis} = \sqrt{\sum_{j=1}^c (p_{ij} - d_{ij})^2}$
Sørensen <sup>[20]</sup> ↓	$\text{dis} = \frac{\sum_{j=1}^c  p_{ij} - d_{ij} }{\sum_{j=1}^c  p_{ij} + d_{ij} }$
Squard $\chi^2$ <sup>[21]</sup> ↓	$\text{dis} = \sum_{j=1}^c \frac{(p_{ij} - d_{ij})^2}{p_{ij} + d_{ij}}$
Kullback-Leibler (KL) <sup>[18]</sup> ↓	$\text{dis} = \sum_{j=1}^c p_{ij} \ln \frac{p_{ij}}{d_{ij}}$
Intersection <sup>[22]</sup> ↑	$\text{sim} = \sum_{j=1}^c \min(p_{ij}, d_{ij})$
Fidelity <sup>[29]</sup> ↑	$\text{sim} = \sum_{j=1}^c \sqrt{p_{ij} d_{ij}}$

### 3.3 实验结果

表7~14的第1~6列列出了10次实验的平均结果±标准差(当前方法性能的排名),末列为前6列平均性能排名。首先比较表7~14中的平均值,如果平均值相同,再比较标准差。

对于数据集 Elu 和 Cold, 本文提出的方法在所有评价指标上都比其他8种方法表现更好。对于数据集 Alpha、Cdc 和 Heat, 本文提出的方法在大多数评价指标上排名第一。对于其余3个数据集, 本文提出的方法排在第二或者第三。

表7 Alpha数据集上的实验结果  
Table 7 Experimental results on the Alpha dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	<b>0.023 1±0.000 2</b> (1)	<b>0.037 8±0.000 5</b> (1)	<b>0.005 5±0.000 3</b> (1)	<b>0.005 5±0.000 2</b> (1)	<b>0.962 2±0.000 5</b> (1)	0.998 6±.000 2 (2)	1.2
LDLLC	0.023 2±0.000 4 (2)	0.037 9±0.000 6 (2)	0.005 6±0.000 3 (2)	0.005 5±0.000 3 (2)	0.962 1±0.000 6 (2)	<b>0.998 6±.000 1</b> (1)	1.8
PT-Bayes	0.229 8±0.012 4 (9)	0.348 5±0.015 4 (9)	0.387 9±0.027 7 (9)	0.560 7±0.071 0 (9)	0.651 5±0.015 4 (9)	0.877 7±0.010 0 (9)	9.0
PT-SVM	0.027 6±0.000 6 (6)	0.044 5±0.000 9 (6)	0.007 1±0.000 3 (6)	0.007 1±0.000 3 (6)	0.956 5±0.000 9 (6)	0.998 1±0.000 1 (6)	6.0
AA-kNN	0.027 9±0.000 6 (7)	0.044 9±0.001 2 (7)	0.007 3±0.000 3 (7)	0.007 4±0.000 4 (8)	0.956 1±0.001 2 (7)	0.998 0±0.000 1 (7)	7.2
AA-BP	0.087 1±0.007 0 (8)	0.147 5±0.013 1 (8)	0.139 9±0.050 1 (8)	0.007 3±0.005 8 (7)	0.853 8±0.011 7 (8)	0.983 9±0.001 7 (8)	7.8
SA-IIS	0.026 9±0.000 4 (5)	0.042 9±0.001 2 (4)	0.006 9±0.000 4 (5)	0.006 9±0.000 4 (5)	0.957 1±0.001 2 (4)	0.998 3±0.001 1 (5)	4.7
SA-BFGS	0.025 1±0.000 4 (3)	0.040 8±0.001 1 (3)	0.006 3±0.000 8 (3)	0.006 3±0.000 4 (3)	0.957 4±0.000 9 (3)	0.998 5±0.001 1 (3)	3.0
EDL	0.026 0±0.001 1 (4)	0.042 9±0.002 2 (5)	0.006 7±0.000 6 (4)	0.006 8±0.000 6 (4)	0.957 0±0.002 2 (5)	0.998 3±0.000 2 (4)	4.3

表8 Cdc数据集上的实验结果  
Table 8 Experimental results on the Cdc dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	<b>0.028 0±0.000 3</b> (1)	0.042 8±0.000 7 (2)	0.007 1±0.000 5 (2)	0.006 9±0.000 1 (2)	<b>0.958 7±0.000 4</b> (1)	<b>0.998 4±0.000 2</b> (1)	1.5
LDLLC	0.028 0±0.000 5 (2)	<b>0.042 7±0.000 9</b> (1)	0.007 1±0.000 7 (3)	<b>0.006 7±0.000 5</b> (1)	0.957 3±0.000 9 (2)	0.998 2±0.000 3 (4)	2.2
PT-Bayes	0.239 9±0.010 3 (9)	0.345 5±0.011 1 (9)	3.853±0.021 0 (9)	0.537 4±0.050 3 (9)	0.654 5±0.011 1 (9)	0.877 8±0.007 5 (9)	9.0
PT-SVM	0.029 8±0.000 7 (5)	0.045 8±0.001 2 (6)	0.007 7±0.000 4 (6)	0.007 6±0.000 4 (6)	0.955 4±0.001 2 (6)	0.998 0±0.000 1 (6)	5.8
AA-kNN	0.030 1±0.000 9 (7)	0.046 2±0.001 3 (7)	0.008 0±0.000 4 (7)	0.007 9±0.000 4 (7)	0.953 8±0.001 3 (7)	0.998 0±0.000 1 (6)	6.8
AA-BP	0.076 9±0.008 1 (8)	0.119 2±0.010 9 (8)	0.084 2±0.028 1 (8)	0.051 1±0.012 1 (8)	0.882 9±0.013 4 (8)	0.987 9±0.005 1 (8)	8.0
SA-IIS	0.029 0±0.001 0 (6)	0.044 5±0.001 5 (4)	0.007 3±0.000 5 (5)	0.007 2±0.000 5 (5)	0.955 6±0.001 5 (5)	0.998 2±0.001 2 (5)	5.0
SA-BFGS	0.028 4±0.001 1 (4)	0.044 9±0.001 6 (5)	<b>0.007 0±0.000 4</b> (1)	0.007 0±0.000 5 (3)	0.955 8±0.001 6 (4)	0.998 3±0.001 1 (2)	3.2
EDL	0.028 3±0.000 6 (3)	0.042 9±0.000 8 (3)	0.007 2±0.000 4 (4)	0.007 2±0.000 4 (4)	0.957 1±0.000 8 (3)	0.998 2±0.000 1 (3)	3.3

表9 Elu 数据集上的实验结果

Table 9 Experimental results on the Elu dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	<b>0.0279±0.0003</b> (1)	<b>0.0415±0.0005</b> (1)	<b>0.0063±0.0005</b> (1)	<b>0.0062±0.0004</b> (1)	<b>0.9585±0.0005</b> (1)	<b>0.9984±0.0003</b> (1)	1.0
LDLLC	0.0279±0.0005 (2)	0.0415±0.0007 (2)	0.0063±0.0008 (2)	0.0062±0.0006 (2)	0.9585±0.0007 (2)	0.9984±0.0004 (2)	2.0
PT-Bayes	0.2588±0.0203 (9)	0.3558±0.0198 (9)	0.4081±0.0408 (9)	0.6062±0.1030 (9)	0.6442±0.0198 (9)	0.8689±0.0156 (9)	9.0
PT-SVM	0.0293±0.0008 (4)	0.0438±0.0012 (4)	0.0068±0.0005 (4)	0.0068±0.0005 (4)	0.9562±0.0012 (4)	0.9983±0.0002 (4)	4.0
AA-kNN	0.0297±0.0010 (5)	0.0443±0.0014 (5)	0.0071±0.0006 (6)	0.0071±0.0006 (6)	0.9557±0.0014 (5)	0.9982±0.0002 (5)	5.3
AA-BP	0.0733±0.0037 (8)	0.1100±0.0048 (8)	0.0731±0.0026 (8)	0.0481±0.0061 (8)	0.8891±0.0064 (8)	0.9890±0.0025 (8)	8.0
SA-IIS	0.0307±0.0009 (6)	0.0472±0.0014 (6)	0.0071±0.0004 (5)	0.0071±0.0004 (5)	0.9528±0.0015 (7)	0.9982±0.0035 (6)	5.8
SA-BFGS	0.0308±0.0009 (7)	0.0475±0.0012 (7)	0.0075±0.0004 (7)	0.0073±0.0003 (7)	0.9552±0.0017 (6)	0.9979±0.0009 (7)	6.8
EDL	0.0289±0.0005 (3)	0.0431±0.0008 (3)	0.0067±0.0003 (3)	0.0067±0.0003 (3)	0.9569±0.0007 (3)	0.9983±0.0001 (3)	3.0

表10 Diau 数据集上的实验结果

Table 10 Experimental results on the Diau dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	0.0543±0.0008 (3)	0.0597±0.0009 (3)	0.0132±0.0011 (2)	0.0130±0.0008 (2)	0.9403±0.0009 (3)	0.9967±0.0005 (2)	2.5
LDLLC	0.0545±0.0009 (4)	0.0599±0.0010 (4)	0.0133±0.0011 (3)	0.0132±0.0008 (3)	0.9401±0.0010 (4)	0.9966±0.0005 (3)	3.5
PT-Bayes	0.4027±0.0183 (9)	0.4177±0.0170 (9)	0.5280±0.0281 (9)	0.8512±0.0772 (9)	0.5823±0.0170 (9)	0.8230±0.0107 (9)	9.0
PT-SVM	0.0628±0.0037 (8)	0.0686±0.0041 (7)	0.0169±0.0018 (7)	0.0167±0.0017 (7)	0.9314±0.0041 (7)	0.9957±0.0004 (7)	7.2
AA-kNN	0.0567±0.0019 (5)	0.0622±0.0022 (5)	0.0145±0.0011 (5)	0.0145±0.0010 (5)	0.9378±0.0022 (5)	0.9963±0.0003 (5)	5.0
AA-BP	0.0802±0.0051 (7)	0.0863±0.0059 (8)	0.0276±0.0013 (8)	0.0291±0.0069 (8)	0.9142±0.0067 (8)	0.9929±0.0031 (8)	8.0
SA-IIS	0.0539±0.0031 (2)	0.0593±0.0032 (2)	0.0144±0.0014 (4)	0.0141±0.0013 (4)	0.9407±0.0003 (2)	0.9964±0.0036 (4)	3.0
SA-BFGS	<b>0.0444±0.0022</b> (1)	<b>0.0476±0.0023</b> (1)	<b>0.0089±0.0008</b> (1)	<b>0.0083±0.0009</b> (1)	<b>0.9513±0.0027</b> (1)	<b>0.9978±0.0031</b> (1)	1.0
EDL	0.0597±0.0010 (6)	0.0653±0.0010 (6)	0.0158±0.0005 (6)	0.0155±0.0005 (6)	0.9347±0.0010 (6)	0.9960±0.0002 (6)	6.0

表11 Heat 数据集上的实验结果

Table 11 Experimental results on the Heat dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	0.0591±0.0009 (2)	<b>0.0597±0.0008</b> (1)	<b>0.0127±0.0011</b> (1)	<b>0.0125±0.0007</b> (1)	<b>0.9403±0.0008</b> (1)	<b>0.9968±0.0007</b> (1)	1.2

续表 11

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
LDLLC	<b>0.059 1±0.000 8</b> (1)	<b>0.059 7±0.000 8</b> (1)	<b>0.012 7±0.001 1</b> (1)	<b>0.012 5±0.000 7</b> (1)	<b>0.940 3±0.000 8</b> (1)	0.996 8±0.000 8 (2)	1.2
PT-Bayes	0.450 0±0.023 1 (9)	0.435 4±0.019 3 (9)	0.545 0±0.036 1 (9)	0.867 8±0.119 8 (9)	0.564 6±0.019 3 (9)	0.818 0±0.013 1 (9)	9.0
PT-SVM	0.062 5±0.002 3 (4)	0.062 7±0.002 2 (3)	0.014 1±0.001 0 (3)	0.014 1±0.001 0 (3)	0.937 3±0.002 2 (4)	0.996 4±0.000 3 (3)	3.3
AA-kNN	0.062 4±0.002 0 (3)	0.063 2±0.001 8 (4)	0.014 1±0.001 0 (3)	0.014 1±0.001 0 (3)	0.936 8±0.001 8 (3)	0.996 4±0.000 3 (3)	3.2
AA-BP	0.079 3±0.006 8 (8)	0.082 2±0.007 1 (8)	0.023 5±0.004 7 (8)	0.024 6±0.005 3 (8)	0.919 8±0.006 1 (8)	0.993 7±0.002 8 (8)	8.0
SA-IIS	0.070 3±0.003 6 (6)	0.069 2±0.003 3 (6)	0.018 2±0.001 6 (6)	0.018 2±0.001 6 (6)	0.930 9±0.003 3 (6)	0.995 4±0.004 2 (7)	6.2
SA-BFGS	0.072 8±0.003 1 (7)	0.079 1±0.002 9 (7)	0.018 8±0.001 6 (7)	0.018 6±0.001 5 (7)	0.930 4±0.003 4 (7)	0.996 1±0.004 8 (6)	6.8
EDL	0.062 9±0.001 6 (5)	0.063 3±0.001 7 (5)	0.014 3±0.000 8 (5)	0.014 3±0.000 8 (5)	0.936 6±0.001 7 (5)	0.996 3±0.000 3 (5)	5.0

表 12 Spo 数据集上的实验结果

Table 12 Experimental results on the Spo dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	<b>0.081 7±0.001 4</b> (1)	0.084 2±0.001 4 (2)	0.024 7±0.001 6 (2)	0.024 3±0.001 6 (2)	0.915 8±0.001 4 (2)	0.993 7±0.000 6 (3)	2.0
LDLLC	0.081 9±0.001 3 (2)	0.084 4±0.001 3 (3)	0.024 8±0.001 4 (3)	0.024 5±0.001 3 (3)	0.915 6±0.001 3 (3)	0.993 7±0.000 5 (2)	2.7
PT-Bayes	0.403 8±0.016 2 (9)	0.403 0±0.013 4 (9)	0.497 2±0.024 6 (9)	0.717 2±0.084 0 (9)	0.597 1±0.013 4 (9)	0.834 2±0.009 5 (9)	9.0
PT-SVM	0.087 8±0.001 9 (7)	0.089 3±0.002 2 (6)	0.028 0±0.001 5 (6)	0.028 4±0.001 5 (6)	0.910 7±0.002 2 (6)	0.992 9±0.000 4 (6)	6.2
AA-kNN	0.087 9±0.003 0 (6)	0.089 9±0.002 4 (7)	0.028 6±0.002 0 (7)	0.028 6±0.000 2 (7)	0.909 6±0.003 4 (7)	0.992 7±0.000 5 (7)	6.8
AA-BP	0.097 9±0.004 1 (8)	0.101 2±0.003 8 (8)	0.034 4±0.003 8 (8)	0.035 9±0.003 9 (8)	0.898 2±0.003 7 (8)	0.990 6±0.001 0 (8)	8.0
SA-IIS	0.086 3±0.004 1 (5)	0.086 1±0.003 6 (4)	0.025 1±0.003 6 (4)	0.025 2±0.002 2 (4)	0.913 9±0.003 6 (4)	0.993 7±0.000 5 (4)	4.2
SA-BFGS	0.081 9±0.004 5 (3)	<b>0.083 3±0.003 8</b> (1)	<b>0.022 9±0.001 9</b> (1)	<b>0.022 6±0.002 1</b> (1)	<b>0.916 8±0.003 9</b> (1)	<b>0.995 1±0.000 7</b> (1)	1.3
EDL	0.084 3±0.002 9 (4)	0.087 2±0.002 9 (5)	0.026 8±0.001 5 (5)	0.026 9±0.001 6 (5)	0.912 8±0.002 8 (5)	0.993 2±0.000 4 (5)	4.8

表 13 Cold 数据集上的实验结果

Table 13 Experimental results on the Cold dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	<b>0.068 1±0.001 5</b> (1)	<b>0.059 1±0.001 4</b> (1)	<b>0.012 2±0.002 3</b> (1)	<b>0.012 0±0.001 3</b> (1)	<b>0.940 9±0.001 4</b> (1)	<b>0.996 9±0.001 3</b> (1)	1.0
LDLLC	0.068 3±0.001 9 (2)	0.059 2±0.001 7 (2)	0.012 2±0.002 5 (2)	0.012 1±0.001 7 (2)	0.940 8±0.001 7 (2)	0.996 9±0.001 2 (3)	2.2
PT-Bayes	0.525 2±0.022 4 (9)	0.447 9±0.018 9 (9)	0.587 3±0.035 2 (9)	0.908 9±0.104 2 (9)	0.552 1±0.018 9 (9)	0.799 1±0.013 4 (9)	9.0



续表 13

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
PT-SVM	0.075 3±0.008 0	0.065 4±0.006 9	0.014 7±0.003 3	0.014 6±0.003 3	0.934 6±0.006 9	0.996 3±0.000 8	5.3
	(5)	(6)	(5)	(5)	(6)	(5)	
AA-kNN	0.072 4±0.002 7	0.063 0±0.002 4	0.013 6±0.001 1	0.013 6±0.001 1	0.937 0±0.002 4	0.996 6±0.000 3	3.2
	(3)	(3)	(3)	(3)	(3)	(4)	
AA-BP	0.083 8±0.004 5	0.071 0±0.002 7	0.017 8±0.001 1	0.016 3±0.003 0	0.932 8±0.002 9	0.995 2±0.001 7	8.0
	(8)	(8)	(8)	(8)	(8)	(8)	
SA-IIS	0.076 7±0.000 4	0.065 3±0.003 4	0.015 7±0.001 5	0.015 5±0.001 5	0.934 7±0.003 4	0.996 0±0.003 9	6.2
	(6)	(5)	(7)	(7)	(5)	(7)	
SA-BFGS	0.074 5±0.000 4	0.064 1±0.003 5	0.013 9±0.001 3	0.014 3±0.001 5	0.934 8±0.003 5	0.996 8±0.003 6	3.0
	(4)	(3)	(3)	(3)	(3)	(2)	
EDL	0.077 1±0.001 8	0.066 8±0.001 6	0.015 4±0.000 9	0.015 3±0.000 9	0.933 2±0.001 6	0.996 1±0.000 3	6.5
	(7)	(7)	(6)	(6)	(7)	(6)	

表 14 Dtt 数据集上的实验结果

Table 14 Experimental results on the Dtt dataset

算法	Euclidean↓	Sørensen↓	Squard $\chi^2$ ↓	KL↓	Intersection↑	Fidelity↑	平均值
T-LDL	<b>0.047 7±0.001 5</b>	0.041 5±0.001 3	0.006 2±0.002 7	0.006 0±0.001 8	<b>0.958 5±0.001 3</b>	0.998 4±0.001 2	1.8
	(1)	(2)	(2)	(2)	(1)	(3)	
LDLLC	0.048 0±0.002 0	0.041 7±0.001 7	0.006 2±0.002 8	0.006 1±0.002 1	0.958 3±0.001 7	0.998 4±0.001 1	2.7
	(2)	(3)	(3)	(3)	(3)	(2)	
PT-Bayes	0.487 9±0.024 2	0.415 6±0.019 2	0.541 6±0.043 8	0.906 9±0.158 0	0.584 4±0.019 2	0.811 3±0.018 6	9.0
	(9)	(9)	(9)	(9)	(9)	(9)	
PT-SVM	0.051 6±0.002 9	0.044 7±0.002 4	0.007 1±0.000 9	0.007 1±0.000 9	0.955 3±0.002 4	0.998 2±0.000 3	6.3
	(6)	(6)	(7)	(7)	(6)	(6)	
AA-kNN	0.051 2±0.001 9	0.044 3±0.001 7	0.007 1±0.000 7	0.007 0±0.000 7	0.955 7±0.001 7	0.998 2±0.000 2	5.3
	(5)	(5)	(6)	(6)	(5)	(5)	
AA-BP	0.062 2±0.003 2	0.053 1±0.002 9	0.009 7±0.001 2	0.012 2±0.003 7	0.946 5±0.002 4	0.996 9±0.001 1	8.0
	(8)	(8)	(8)	(8)	(8)	(8)	
SA-IIS	0.053 5±0.002 3	0.048 0±0.002 3	0.006 8±0.000 5	0.006 8±0.000 5	0.952 0±0.002 3	0.998 3±0.001 3	5.5
	(7)	(7)	(4)	(4)	(7)	(4)	
SA-BFGS	0.049 5±0.001 9	<b>0.040 9±0.001 7</b>	<b>0.005 8±0.000 5</b>	<b>0.005 4±0.000 4</b>	0.958 4±0.002 3	<b>0.998 9±0.001 0</b>	1.5
	(3)	(1)	(1)	(1)	(2)	(1)	
EDL	0.050 8±0.002 2	0.044 0±0.001 8	0.006 9±0.000 7	0.006 8±0.000 8	0.956 0±0.001 8	0.998 2±0.000 3	4.7
	(4)	(4)	(5)	(5)	(4)	(6)	

### 3.4 讨论

各种算法通常在不同的数据集上具有不同的排名,表明每种算法都有其合适的应用场景,如EDL算法更适用于文本情绪分类场景。不同评价指标下同一算法的不同排名,反映了6项评价指标的多样性。在比较不同方法对新数据集的预测效果时,应综合考虑多个评价指标。

与同样考虑标签相关性的LDLLC算法相比,T-LDL算法在绝大多数数据集上的表现均优于LDLLC算法。LDLLC算法基于皮尔逊相关系数表征标签相关性,而T-LDL算法基于本文设计的三角距离。皮尔逊相关系数要求输入的2个向量满足线性相关,而本文设计的三角距离无该约束

条件。实验证明在本文场景中,三角距离更加合适。

## 4 结束语

为了进一步提高标签分布学习算法的预测性能,本文提出了三角距离相关性的标签分布学习算法。新的三角距离可以充分考虑向量本身和向量之差,能更好地描述标签之间的相关性。实验结果表明,本文的方法比大多数现有的方法表现更好。

未来的工作将尝试从以下几个方面提高标签分布学习方法的性能:1)采用属性约简以降低算法的时间复杂度;2)使用其他度量取代作为目标

函数基础的 KL 散度; 3) 利用新的距离映射函数表示标签的相关性。

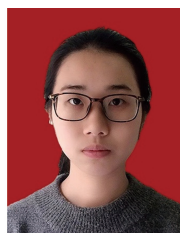
## 参考文献:

- [1] GENG Xin. Label distribution learning[J]. *IEEE transactions on knowledge and data engineering*, 2016, 28(7): 1734–1748.
- [2] JIA Xiuyi, ZHENG Xiang, LI Weiwei, et al. Facial emotion distribution learning by exploiting low-rank label correlations locally[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019: 9841–9850.
- [3] YANG Xu, GAO Binbin, XING Chao, et al. Deep label distribution learning for apparent age estimation[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision Workshops*. Santiago, Chile, 2015: 102–108.
- [4] ZHANG Hengru, HUANG Yuting, XU Yuanyuan, et al. COS-LDL: label distribution learning by cosine-based distance-mapping correlation[J]. *IEEE access*, 2020, 8: 63961–63970.
- [5] 邵东恒, 杨文元, 赵红. 应用  $k$ -means 算法实现标记分布学习 [J]. *智能系统学报*, 2017, 12(3): 325–332.  
SHAO Dongheng, YANG Wenyuan, ZHAO Hong. Label distribution learning based on  $k$ -means algorithm[J]. *CAAI transactions on intelligent systems*, 2017, 12(3): 325–332.
- [6] 刘玉杰, 唐顺静, 高永标, 等. 基于标签分布学习的视频摘要算法 [J]. *计算机辅助设计与图形学学报*, 2019, 31(1): 104–110.  
LIU Yujie, TANG Shunjing, GAO Yongbiao, et al. Label distribution learning for video summarization[J]. *Journal of computer-aided design & computer graphics*, 2019, 31(1): 104–110.
- [7] 王一宾, 田文泉, 程玉胜. 基于标记分布学习的异态集成学习算法 [J]. *模式识别与人工智能*, 2019, 32(10): 945–954.  
WANG Yibin, TIAN Wenquan, CHENG Yusheng. Heterogeneous ensemble learning algorithm based on label distribution learning[J]. *Pattern recognition and artificial intelligence*, 2019, 32(10): 945–954.
- [8] 耿新, 徐宁. 标记分布学习与标记增强 [J]. *中国科学: 信息科学*, 2018, 48(5): 521–530.  
GENG Xin, XU Ning. Label distribution learning and label enhancement[J]. *Scientia sinica informationis*, 2018, 48(5): 521–530.
- [9] ZHANG Mingling, ZHANG Kun. Multi-label learning by exploiting label dependency[C]//*Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington, USA, 2010: 999–1007.
- [10] BI Wei, KWOK J T. Multilabel classification with label correlations and missing labels[C]//*Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Québec City, Canada, 2014: 1680–1686.
- [11] HUANG Shengjun, ZHOU Zhihua. Multi-label learning by exploiting label correlations locally[C]//*Proceedings of the 26th AAAI Conference on Artificial Intelligence*. Toronto, Canada, 2012: 949–955.
- [12] GENG Xin, WANG Qin, XIA Yu. Facial age estimation by adaptive label distribution learning[C]//*Proceedings of the 22nd International Conference on Pattern Recognition*. Stockholm, Sweden, 2014: 4465–4470.
- [13] ZHANG Zhaoxiang, WANG Mo, GENG Xin. Crowd counting in public video surveillance by label distribution learning[J]. *Neurocomputing*, 2015, 166: 151–163.
- [14] GENG Xin, YIN Chao, ZHOU Zhihua. Facial age estimation by learning from label distributions[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(10): 2401–2412.
- [15] GENG Xin, LING Miaogen. Soft video parsing by label distribution learning[C]. *Proceedings of the 31th AAAI Conference on Artificial Intelligence*. San Francisco, USA, 2017: 1331–1337.
- [16] JIA Xiuyi, LI Weiwei, LIU Junyu, et al. Label distribution learning by exploiting label correlations[C]//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans, USA, 2018: 3310–3317.
- [17] ZHENG Xiang, JIA Xiuyi, LI Weiwei. Label distribution learning by exploiting sample correlations locally[C]//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans, USA, 2018: 4556–4563.
- [18] KULLBACK S, LEIBLER R A. On information and sufficiency[J]. *The annals of mathematical statistics*, 1951, 22(1): 79–86.
- [19] DANIELSSON P E. Euclidean distance mapping[J]. *Computer graphics and image processing*, 1980, 14(3): 227–248.
- [20] SØRENSEN T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons[J]. *Kongelige danske videnskabernes selskab biologiske skrifter*, 1948, 5(4): 1–34.
- [21] GAVIN D G, OSWALD W W, WAHL E R, et al. A statistical approach to evaluating distance metrics and analog assignments for pollen records[J]. *Quaternary research*, 2003, 60(3): 356–367.
- [22] DUDA R O, HART P E, STORK D G. *Pattern classification*[M]. 2nd ed. New York: Wiley, 2000.
- [23] DEZA E, DEZA M M. *Dictionary of distances*[M]. Amsterdam: Elsevier, 2006.
- [24] JEGOU H, DOUZE M, SCHMID C. Hamming embed-

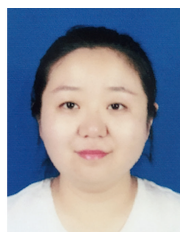
ding and weak geometric consistency for large scale image search[C]//Proceedings of the 10th European Conference on Computer Vision. Marseille, France, 2008: 304–317.

- [25] BERGER A L, PIETRA V J D, PIETRA S A D. A maximum entropy approach to natural language processing[J]. Computational linguistics, 1996, 22(1): 39–71.
- [26] ZHOU Deyu, ZHANG Xuan, ZHOU Yin, et al. Emotion distribution learning from texts[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas, 2016: 638–647.
- [27] YUAN Yaxiang. A modified BFGS algorithm for unconstrained optimization[J]. *IMA journal of numerical analysis*, 1991, 11(3): 325–332.
- [28] EISEN M B, SPELLMAN P T, BROWN P O, et al. Cluster analysis and display of genome-wide expression patterns[J]. *Proceedings of the national academy of sciences of the united states of America*, 1998, 95(25): 14863–14868.
- [29] CHA Su H. Comprehensive survey on distance/similarity measures between probability density functions[J]. *International journal of mathematical models and methods in applied sciences*, 2007, 1(4): 300–307.

## 作者简介:



黄雨婷, 硕士研究生, 主要研究方向为标签分布学习和推荐系统。



徐媛媛, 讲师, 主要研究方向为信号处理和推荐系统。主持教育部产学合作协同育人项目 2 项。发表学术论文 2 篇。



张恒汝, 教授, 主要研究方向为标签分布学习、粒计算、推荐系统和数据挖掘。主持四川省科技厅项目 1 项, 参与省部级科研及教研项目 3 项。发表学术论文 30 余篇。