DOI: 10.13718/j.cnki.xdzk.2024.05.004
樊俊，张恒汝，余一帆，等. 面向非均匀分布数据的代价敏感标记分布学习 [J]. 西南大学学报(自然科学版)，2024，46(5)：40-50.

面向非均匀分布数据的代价敏感标记分布学习

樊俊[1,2]，　张恒汝[1,2]，　余一帆[1]，　闵帆[1,2]

1. 西南石油大学 计算机科学学院，成都 610500；2. 西南石油大学 机器学习研究中心，成都 610500

摘要：标记歧义近年来在机器学习和数据挖掘领域备受关注. 标记分布学习(LDL)通过为样本分配概率标记来解决标记歧义问题. 现有的 LDL 方法主要是为处理训练数据均匀分布的情况而设计的. 然而，在实际应用中，训练数据往往呈现非均匀分布. 因此，提出了一种代价敏感的标记分布学习方法(CSLDL)，用以处理这种非均匀分布的数据. 通过充分利用样本的密度信息，设计了一种新的损失函数. 首先，将描述度集平均划分为多个区间，并统计这些区间中的样本个数，从而推导出每个类别标记的经验密度向量. 其次，为了确保不同区间之间的连续性，利用邻居来对目标区间的经验密度进行修正. 将经验密度向量与对称核进行卷积，以使每个区间不仅考虑当前区间，还考虑附近区间. 最后，利用修正后的密度向量构建代价矩阵，并结合 Kullback-Leibler(K-L)散度来处理非均匀分布的训练数据. CSLDL 在 10 个真实世界的数据集上与 6 种最先进的算法进行了对比实验. 实验结果充分验证了提出的方法的有效性和优越性.

关　键　词：标记分布学习；标记歧义；非均匀分布数据；代价敏感；样本密度

中图分类号：TP391　　**文献标志码**：A

开放科学(资源服务)标识码(OSID)：

文 章 编 号：1673-9868(2024)05-0040-11

Cost-sensitive Label Distribution Learning for Non-Uniform Distributed Data

FAN Jun[1,2]，　ZHANG Hengru[1,2]，　YU Yifan[1]，　MIN Fan[1,2]

1. *College of Computer Science*, *Southwest Petroleum University*, *Chengdu 610500*, *China*;
2. *Lab of Machine Learning*, *Southwest Petroleum University*, *Chengdu 610500*, *China*;

Abstract: Learning with label ambiguity has recently been a popular topic in machine learning and data

收稿日期：2023-05-24
基金项目：国家自然科学基金资助项目(61902328)；南充市科技局应用基础研究项目(SXHZ040).
作者简介：樊俊，硕士，主要从事标签分布学习、标签增强等方面的研究.
通信作者：张恒汝，博士，教授.

mining research. Label distribution learning (LDL) deals with label ambiguity by assigning probabilistic labels to each instance. Existing LDL methods are designed for training data that is uniformly distributed. However, in real-world applications, the training data is typically not uniformly distributed. In this paper, we propose a cost-sensitive method for label distribution learning (CSLDL) to deal with the non-uniformly distributed training data. We designed a novel loss function by applying the density information of instances. The descriptive set was firstly averaged over multiple bins. The empirical density vector for each class label was then derived by counting the number of instances in these bins. Secondly, in order to construct the continuity between different bins, we employed neighbor samples to modify the empirical density of the target bins. Specifically, we convolved the empirical density vector with a symmetric kernel so that each bin took into account not just the current bin but also nearby bins. Finally, a cost matrix was constructed using the modified density vectors, combined with Kullback-Leibler (K-L) divergence to deal with non-uniformly distributed training data. Experiments were undertaken on ten real-world datasets compared with six state-of-the-art algorithms. Results demonstrate the effectiveness and superiority of our proposed algorithm.

Key words: label distribution learning; label ambiguity; non-uniformly distributed data; cost-sensitive; density of instances

标记歧义问题[1-2]是当前机器学习和数据挖掘[3-4]领域中备受关注的热门话题. 单标记学习(Single-Label Learning, SLL)[5]和多标记学习(Multi-Label Learning, MLL)[6-8]是当前处理标记歧义问题的 2 种成熟的学习范式. 单标记学习假设每个样本只能与一种逻辑标记相关联，拥有该标记则为 1，否则为 0. 而多标记学习则假设每个样本能与一组逻辑标记相关联. 尽管多标记学习相比于单标记学习能够解决更复杂的标记歧义问题，但是这 2 种学习范式都只能回答标记是否属于一个样本，而无法回答标记对样本的描述程度. 为了解决这一问题，Geng 等人[9]提出了标记分布学习(Label Distribution Learning, LDL). 该学习范式为每个样本分配一组描述度，并假设所有标记的描述度之和为 1. 这种学习范式能够准确区分标记对于样本的描述度差异，可适用于更加广泛的应用场景.

现有的 LDL 算法已经应用了多种方法来提高预测性能. 最初，LDLLC[10]通过应用全局标记相关性来扩展特征空间以提高性能. 随后，LDL-SCL[11]、LDL-FLC[12]建议利用局部标记相关性以提升性能. 最后，LDL-LDM[13]提出了同时利用局部和全局标记相关性的方法. 除了利用标记相关性之外，LDL-HR[14]提议首先学习描述度最高的标记，然后学习其他标记，以避免目标不匹配. 而 LDLSF[15]则提出了一种为每个标记选择特定特征的方法，而不是共享所有特征. 然而，这些算法都是在假设训练集是均匀分布的前提下构建的. 而在实际应用中，情况往往并非如此. 一种标记所对应的训练样本的分布通常是非均匀分布的，如图 1 所示，其中横坐标表示描述度区间，纵坐标表示在该区间内样本的数量，浅绿色为低密度区域，深绿色为高密度区域. 作为一种数据驱动的学习范式，在使用非均匀数据时往往会忽略来自低密度区域的样本，从而导致预测出现错误[16-18].

为了处理非均匀分布数据，本文提出了一种新的基于代价敏感的标记分布学习算法(Cost-Sensitive Method for Label Distribution Learning, CSLDL). 我们根据样本的密度构建代价矩阵[19]，并用它来对传统的 K-L 散度目标函数进行优化，使其适用于非均匀分布数据. 首先，将描述度集等分为多个区间，并为每一种标记分别统计其在这些区间中样本的经验密度. 标记分布数据作为一种连续型数据，同一种标记的相邻描述度之间不存在严格的界限[20-21]. 然而，在划分子区间之后，会破坏相邻区间之间的连续性. 为了

重构相邻区间之间的连续性，我们利用邻居对经验密度进行修正，使其不仅考虑当前区间的经验密度，还要考虑邻居区间的经验密度. 例如，区间同时考虑了自身和邻居的经验密度，其中设置贡献权重以考虑与目标区间的距离，距离越远权重越小. 修正后，不仅重建了相邻区间的连续性，还得到了更加平滑的分布，如图1(b)所示. 在现实生活中，连续性数据的变化应该是平滑的，例如人口统计数据，正常情况下，每年的人口只会小幅度增加或减少. 修正后的密度分布也变得更加平滑，可以更好地反映现实世界数据的真实情况，从而增加算法的泛化性. 最后，利用平滑后的密度信息构建代价矩阵，并用它对传统的K-L散度目标函数进行改造，对处于高密度区域的样本设置较小的代价，反之设置较大的代价.

为了验证本文提出的CSLDL方法，我们对多个数据集进行了广泛的实验. 实验结果表明，与6种先进的方法相比，CSLDL预测的结果最贴近真实数据. 本文的主要贡献包括3个方面：① 提出了一种代价敏感的标记分布学习算法(CSLDL)，通过样本密度信息以更均衡的方式处理模型对高密度区域样本的偏重；② 通过利用近邻修正目标区间的经验密度，不仅重建了相邻区间的连续性，而且得到了更符合现实规律的分布数据；③ 应用不同数据集的实验结果证明了提出的算法的有效性和优越性.

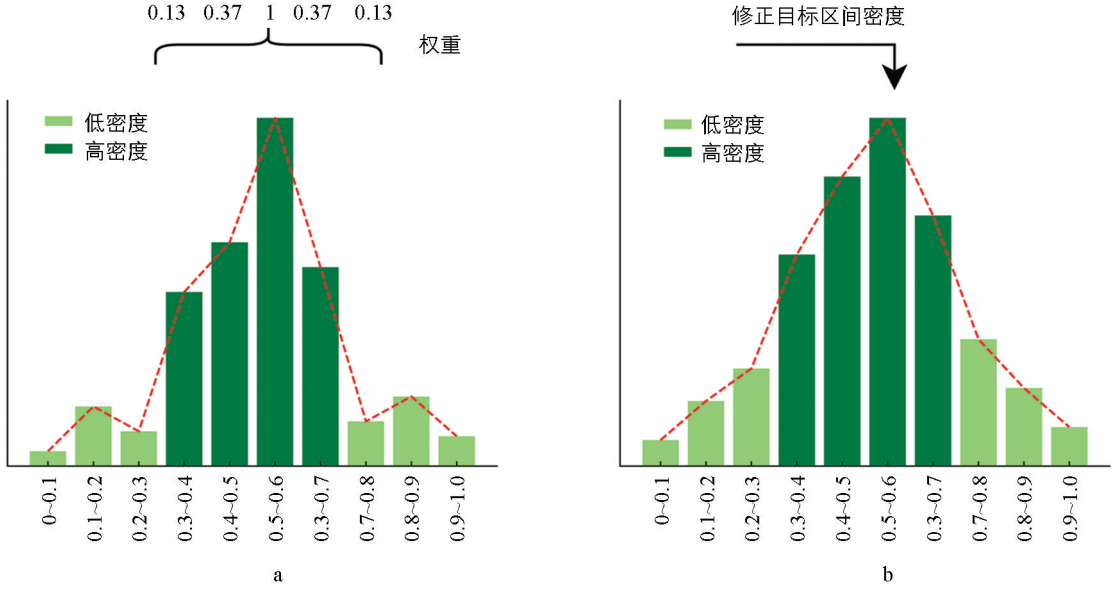

图1 一个标记对应的非均匀数据分布示例

1 相关工作

1.1 符号系统

令 $\boldsymbol{X}=[\boldsymbol{x}_1;\boldsymbol{x}_2;\cdots;\boldsymbol{x}_n]\in R^{n\times q}$ 表示输入特征空间，$\boldsymbol{Y}=\{y_1,y_2,\cdots,y_o\}$ 表示标记集合，$\boldsymbol{D}=[\boldsymbol{d}_1;\boldsymbol{d}_2;\cdots;\boldsymbol{d}_n]\in[0,1]^{n\times o}$ 表示与输入空间相对应的标记空间. 其中，n 为样本数量，q 为特征维度，o 为标记数量. 第 i 个样本表示为 $\boldsymbol{x}_i\in\boldsymbol{X}$，其第 j 个标记的描述度表示为 d_{ij}. 每个样本的描述度应满足

$$\boldsymbol{d}_i=(d_{i1},d_{i2},\cdots,d_{io}),\ \text{s.t.},\ \sum_{j=1}^{o}d_{ij}=1 \tag{1}$$

1.2 标记分布学习算法

当前常见的标记学习算法设计主要基于以下3种策略[22]：

1) 问题转化(Problem Transformation，PT)：这种策略将标记分布学习问题转化为单标记学习问题. 具体而言，训练样本被转化为加权的单标记样本，权重可以由标记的描述度得到，转化后的样本就可以使用单标记学习算法进行处理. 基于此策略的代表算法有PT-Bayes和PT-SVM等.

2）算法改造（Algorithm Adaptation，AA）：这种策略将传统单标记或多标记学习算法改造为能够处理标记分布问题的算法. 该策略的代表算法有基于 k 近邻改造的 AA-kNN 算法以及基于反向传播改造的 AA-BP 算法.

3）专用算法（Specialized Algorithm，SA）：这种策略是根据标记分布学习本身固有特性设计的专用算法. 专用算法通常由 3 个模块组成，包括输出模型、目标函数和优化方法. 现有的专用算法大多采用 K-L 散度作为目标函数，并假设所有标记的贡献相等来进行模型优化. 然而在实际应用中，训练数据通常呈现不均匀分布，使用传统 K-L 散度为目标函数训练的模型会忽略来自低密度区域的样本. 为了解决这个问题，本文提出了一种基于代价敏感的标记分布学习算法，该算法利用样本的密度信息来平衡模型对高密度区域样本的偏重.

2　面向非均匀分布数据的代价敏感标记分布学习算法

2.1　代价矩阵

为了平衡传统模型对高密度区域样本的偏重，本文采用了对高密度区域样本设置较小代价、对低密度区域样本设置较大代价的方法. 为了满足这一需求，我们使用样本密度的倒数或开方后的倒数作为代价. 经过实验验证，后者的效果更好. 需要注意的是，不同标记所对应的样本分布是不同的，因此每种标记都有一个对应的代价向量.

首先，将$[0, 1]$划分为 m 个子区间 $s_1, s_2, \cdots, s_m$，每个区间拥有相同的间隔 $\frac{1}{m}$，即 $s_1 = \left[0, \frac{1}{m}\right)$，$s_2 = \left[\frac{1}{m}, \frac{2}{m}\right)$，$\cdots$，$s_m = \left[1-\frac{1}{m}, 1\right]$，并统计标记 y_j 在这些子区间的样本密度，从而得到其对应的代价向量：

$$\boldsymbol{c}_j = [c_{j1}, c_{j2}, \cdots, c_{jm}] \tag{2}$$

其中：c_{jb} 表示区间 s_b 样本密度开方后的倒数.

根据标记分布数据的特点，同一标记的相邻描述度之间不存在严格的界限，而是呈现连续性. 然而，对子区间的划分却在相邻区间之间引入了分界线，从而破坏了它们之间的连续性. 为了重新建立相邻区间之间的连续性，本文采用核函数来提取相邻区间的信息，以保持数据的连续性.

$$\hat{c}_{jb} = \sum_{i=b-h}^{b+h} k(i, b) c_{ji} \tag{3}$$

其中：h 表示考虑的左右邻居数量，核函数为对称核函数，满足 $k(i, b) = k(b, i)$ 以及 $\nabla_i k(i, b) + \nabla_b k(b, i) = 0$. 常见的对称核函数有高斯核、拉普拉斯核等，本文采用拉普拉斯核.

考虑了近邻信息后，不仅能够重建不同区域之间的连续性，还能使样本分布更加平滑，更符合真实分布，从而提高算法的泛化能力.

为了方便计算，结合代价向量和训练数据的标记空间构建 $n \times o$ 的代价矩阵：

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1o} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{no} \end{pmatrix} \tag{4}$$

其中：$a_{ij} = \hat{c}_{jb}$，仅当样本 $\boldsymbol{x}_i$ 的第 j 个标记 d_{ij} 位于区域 s_b 时.

2.2　标记分布学习

基于代价敏感的标记分布学习算法（CSLDL）符合专用算法设计策略，由输出模型、目标函数、以及优化方法构成.

标记 y_j 对样本 $\boldsymbol{x}_i$ 的描述度由条件概率的形式表示，其参数模型记作 $p(y_j \mid \boldsymbol{x}_i;\boldsymbol{\theta})$. 本文采用最大熵模型作为输出模型

$$p(y_j \mid \boldsymbol{x}_i;\boldsymbol{\theta})=\frac{1}{Z}\exp\Big(\sum_k \theta_{j,k}x_{ik}\Big) \tag{5}$$

其中：$Z=\sum_j \exp\Big(\sum_k \theta_{j,k}x_{ik}\Big)$ 为归一化项，保证 $\boldsymbol{x}_i$ 所有标记的描述度之和为1；$\boldsymbol{\theta}$ 是描述特征与标记分布之间关系的系数矩阵，$\theta_{j,k}$ 是 $\boldsymbol{\theta}$ 中的一个元素；x_{ik} 是 $\boldsymbol{x}_i$ 的第 k 个特征.

为了获取最优参数 $\boldsymbol{\theta}$，构建如下目标函数：

$$\arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta})+\lambda\Omega(\boldsymbol{\theta}) \tag{6}$$

其中：$L(\boldsymbol{\theta})$ 为定义在训练数据上的损失函数，$\Omega(\boldsymbol{\theta})$ 为约束模型参数复杂度的正则项，λ 控制正则的贡献度.

损失函数用于度量真实标记分布与预测标记分布之间的相似性，常见的损失函数有欧式距离，K-L 散度等. 经过实验验证，采用 K-L 散度效果最佳，

$$L(\boldsymbol{\theta})=\sum_i\sum_j\left(d_{ij}\ln\left(\frac{d_{ij}}{p(y_j \mid \boldsymbol{x}_i;\boldsymbol{\theta})}\right)\right) \tag{7}$$

其中：d_{ij} 为标记 y_j 对样本 $\boldsymbol{x}_i$ 的真实描述度，$p(y_j \mid \boldsymbol{x}_i;\boldsymbol{\theta})$ 为预测描述度.

传统的K-L散度无法平衡模型对高密度区域样本的偏重，因此本文将其与代价矩阵相结合以平衡这种偏重. 结合代价矩阵的损失函数表示为

$$L(\boldsymbol{\theta})=\sum_i\sum_j\left(d_{ij}\ln\left(\frac{d_{ij}}{p(y_j \mid \boldsymbol{x}_i;\boldsymbol{\theta})}\right)A_{ij}\right) \tag{8}$$

本文采用标记分布学习中常用的 $L2$ 范数来约束模型参数 $\boldsymbol{\theta}$，以防止模型过拟合，即：

$$\Omega(\boldsymbol{\theta})=\|\boldsymbol{\theta}\|_F^2 \tag{9}$$

将公式(8)和公式(9)代入公式(6)，最终目标函数表示为

$$\begin{aligned}T(\boldsymbol{\theta})=&\arg\min_{\boldsymbol{\theta}}\sum_i\sum_j\left(d_{ij}\ln\left(\frac{d_{ij}}{p(y_j \mid \boldsymbol{x}_i;\boldsymbol{\theta})}\right)A_{ij}\right)+\lambda\|\boldsymbol{\theta}\|_F^2=\\&\sum_i\sum_j A_{ij}d_{ij}\ln d_{ij}-\sum_i\sum_j A_{ij}d_{ij}\sum_k\theta_{r,k}x+\\&\sum_i\sum_j A_{ij}d_{ij}\ln\sum_r\exp\Big(\sum_k\theta_{r,k}x\Big)+\lambda\|\boldsymbol{\theta}\|_F^2\end{aligned} \tag{10}$$

2.3 优化

本文采用拟牛顿法(limited-memory quasi-Newton method，L-BFGS)[23]对目标函数 $T(\boldsymbol{\theta})$ 进行求解. 对应当前迭代次数的二阶泰勒展开为

$$T(\boldsymbol{\theta}^{(l+1)})\approx T(\boldsymbol{\theta}^{(l)})+\nabla T(\boldsymbol{\theta}^{(l)})^{\mathrm{T}}\Delta+\frac{1}{2}\Delta^{\mathrm{T}}H(\boldsymbol{\theta}^{(l)})\Delta \tag{11}$$

其中：$\Delta=(\boldsymbol{\theta}^{(l+1)}-\boldsymbol{\theta}^{(l)})$ 为参数变化量，$\nabla T(\boldsymbol{\theta}^{(l)})$ 为梯度矩阵，$H(\boldsymbol{\theta}^{(l)})$ 是 Hessian 矩阵.

式(11)最小化可得

$$\Delta^{(l)}=-H^{-1}(\boldsymbol{\theta}^{(l)})\nabla T(\boldsymbol{\theta}^{(l)}) \tag{12}$$

在L-BFGS优化方法中，$\Delta^{(l)}$ 被视为搜索方向，选择满足 Wolfe 准则的 $\alpha^{(l)}$ 作为步长，则 $\boldsymbol{\theta}$ 的迭代公式为

$$\boldsymbol{\theta}^{(l+1)}=\boldsymbol{\theta}^{(l)}+\alpha^{(l)}\Delta^{(l)} \tag{13}$$

L-BFGS 的基本思想是避免直接计算牛顿法中使用的逆 Hessian 矩阵，而用迭代更新的矩阵 $\boldsymbol{B}$ 来近似逆 Hessian 矩阵，其表示为

$$\boldsymbol{B}^{(l+1)}=(\boldsymbol{I}-\boldsymbol{\rho}^{(l)}\boldsymbol{s}^{(l)}(\boldsymbol{u}^{(l)})^{\mathrm{T}})\boldsymbol{B}^{(l)}(\boldsymbol{I}-\boldsymbol{\rho}^{(l)}\boldsymbol{u}^{(l)}(\boldsymbol{s}^{(l)})^{\mathrm{T}})+\boldsymbol{\rho}^{(l)}\boldsymbol{s}^{(l)}(\boldsymbol{s}^{(l)})^{\mathrm{T}} \tag{14}$$

其中：$\boldsymbol{I}$ 是单位矩阵，其余变量如下表示

$$\boldsymbol{\rho}^{(l)}=\frac{1}{(\boldsymbol{u}^{(l)})^{\mathrm{T}}\boldsymbol{s}^{(l)}} \tag{15}$$

$$\boldsymbol{u}^{(l)}=\nabla T(\boldsymbol{\theta}^{(l+1)})-\nabla T(\boldsymbol{\theta}^{(l)}) \tag{16}$$

$$\boldsymbol{s}^{(l)}=\boldsymbol{\theta}^{(l+1)}-\boldsymbol{\theta}^{(l)} \tag{17}$$

对于目标函数 $T(\boldsymbol{\theta})$ 的优化，L-BFGS 的计算主要与其一阶梯度有关，表示为

$$\frac{\partial T(\boldsymbol{\theta})}{\partial \theta_{r,k}}=\sum_i\sum_j\frac{\exp(\sum_k\theta_{r,k}x_{ir})x_{ir}A_{ij}d_{ij}}{\sum_r\exp(\sum_k\theta_{r,k}x_{ir})}-\sum_i A_{ir}d_{ir}x_{ik}+2\lambda\theta_{r,k} \tag{18}$$

得到最佳参数 $\boldsymbol{\theta}^{*}$ 之后，代入公式(5)求得样本标记分布.

3　实验与结果分析

3.1　数据集

M2B[24]是由 Ren 等人基于面部美感感知的原始数据集[25]整理而来，其标记分布通过 k-wise 模型转化得到. Emotion6[26]包含 1 980 张来自 Flickr 的图像，用于情绪预测研究. 该数据集通过投票标注了 7 种情绪类别(愤怒、厌恶、喜悦、恐惧、悲伤、惊讶和中性)，并通过文献[15]中提出的方法为每张图像提取了一个 168 维的特征向量. Movie[9]是一个计算电影用户评分的数据集，使用 478 656 位用户对 7 755 部电影的 54 242 292 次评分作为数据源. 数据来自 Netflix，评分范围从 1～5(5 种标记). 电影的标记分布为每个评级级别的百分比，电影的特征是从流派、导演、演员、国家、预算等数据中提取的，每部电影的特征向量是 1 869 维. Flickr-LDL 和 Twitter-LDL[27]分别包含 11 150 和 10 045 张图像，其标记为 8 种情感(愤怒、开心、敬畏、满足、厌恶、兴奋、恐惧和悲伤). 按照文献[15]中提出的方法，分别使用 HOG[28]和 Color Moment[29]对 2 个数据集提取特征向量，然后使用 PCA 将特征降维到 168 维. Natural-Scene[9]来自 2 000 个自然场景图像. FBP5500[30]来自 5 500 张彩色人脸照片. 最后 3 个数据集 Yeast-cdc、Yeast-cold 和 Yeast-spoem[31]是从酵母的真实生物学实验中获得的. 表 1 中给出了这些数据集的一些基本统计数据.

表 1　数据集统计信息

数据集	样本数量	特征维度	标记维度
M2B	1 420	250	5
Emotion6	1 980	168	7
Movie	7 755	1 869	5
Twitter-LDL	10 045	168	8
Flickr-LDL	11 150	168	8
Natural-Scene	2 000	294	9
FBP5500	5 500	512	5
Yeast-cdc	2 465	24	15
Yeast-cold	2 465	24	4
Yeast-spoem	2 465	24	2

3.2 评价指标

标记分布学习的评价指标可以划分为2种，分别为距离指标和相似度指标[32]. 常见的距离指标包括Chebyshev距离↓、Squared χ^2 距离↓、Euclidean距离↓以及Sφrensen距离↓，常见的相似度指标有Cosine相关系数↑和Fidelity相似度↑. ↓表示评价指标度量值越低性能越好，↑表示评价指标度量值越高性能越好. 由于同类型指标的实验结果是相似的，为了文章排版的便利性，本文在两类指标中各选取部分指标(Sφrensen、Fidelity)来展示实验结果.

3.3 实验设置

CSLDL与其他6种先进的算法进行比较，包括PT-Bayes[9]、AA-kNN[9]、SA-BFGS[33]、LDL-LDM[13]、LDL-HR[14]以及LDL-SF[15]. 本文对比实验参数与原文中的参数设置一致. 具体来说，对于AA-kNN，邻居数量 k 设置为5. 对于SA-BFGS，$c_1=10^{-4}$，$c_2=0.9$. 对于LDL-HR，$\lambda_1=0.001$，λ_2 和 λ_3 选自$\{10^{-3}, \cdots, 1\}$，$\rho=0.01$. 对于LDL-LDM，λ_1、λ_2 和 λ_3 选自$\{10^{-3}, \cdots, 10^{3}\}$，$g$ 选自$\{1, 2, 3, \cdots, 14\}$. 对于LDL-SF，参数 λ_1、λ_2 和 λ_3 选自$\{10^{-6}, 10^{-5}, \cdots, 10^{-1}\}$，$\rho=10^{-3}$.

3.4 消融实验

为了验证CSLDL的有效性，我们进行了一项涉及以下4个实验的消融研究(ablation study，AS)：

(a) 不考虑代价敏感，且不使用正则项；

(b) 考虑代价敏感，但不使用邻居修正和正则项；

(c) 考虑代价敏感，并使用邻居修正，但不使用正则项；

(d) 考虑代价敏感，并使用邻居修正和正则项.

为了进行公平的比较，将 m 固定为10，h 固定为1，λ 固定为0.1. 实验结果如表2所示，在考虑代价敏感后，如果不进行邻居修正，可能会降低预测性能(a→b)，反之可以得到更好的效果(a→c)，并且在加入正则项后可以得到进一步提升(c→d). 这证明了CSLDL算法的有效性.

表2 消融实验

数据集	AS	Euclidean	Sφrensen	Fidelity
M2B	(a)	0.496 4	0.436 7	0.825 1
	(b)	0.514 7	0.449 0	0.815 5
	(c)	0.494 5	0.435 2	0.827 5
	(d)	**0.479 0**	**0.425 2**	**0.838 7**
Natural-Scene	(a)	0.459 6	0.468 0	0.731 3
	(b)	0.469 9	0.476 0	0.728 0
	(c)	0.458 1	0.467 5	0.733 4
	(d)	**0.418 2**	**0.451 2**	**0.748 1**

3.5 参数优化

本文的主要参数有子区间的数量 m，左右邻居数量 h 以及正则项贡献度 λ. 为了方便排版，仅对部分数据集的表现进行了报告.

首先将 h 固定为1，λ 固定为0.1，m 依次设置为$\{2, 4, 5, 10, 20, 50, 100\}$. 如图2所示，可以观察到当 $m=10$ 时，3个数据集在2种指标之下都能取得较好的结果.

将参数 m 固定为10，λ 固定为0.1，h 依次设置为$\{0, 1, 2, 3, 4, 5\}$. 实验结果如图3所示，当利用邻居进行密度修正后，效果显著提升($h=0$→$h=1$)；当 $h\geqslant 2$ 时，实验结果只在很小范围内变化.

将参数 m 固定为10，h 固定为2，λ 依次设置为$\{0.000\ 1, 0.001, 0.01, 0.1, 1\}$. 实验结果如图4所示，当 λ 等于0.001或者0.01时，效果最好. 综上所述，对所有数据集，参数设置如下，$m=10$，$h=2$，λ 选自$\{0.001, 0.01\}$.

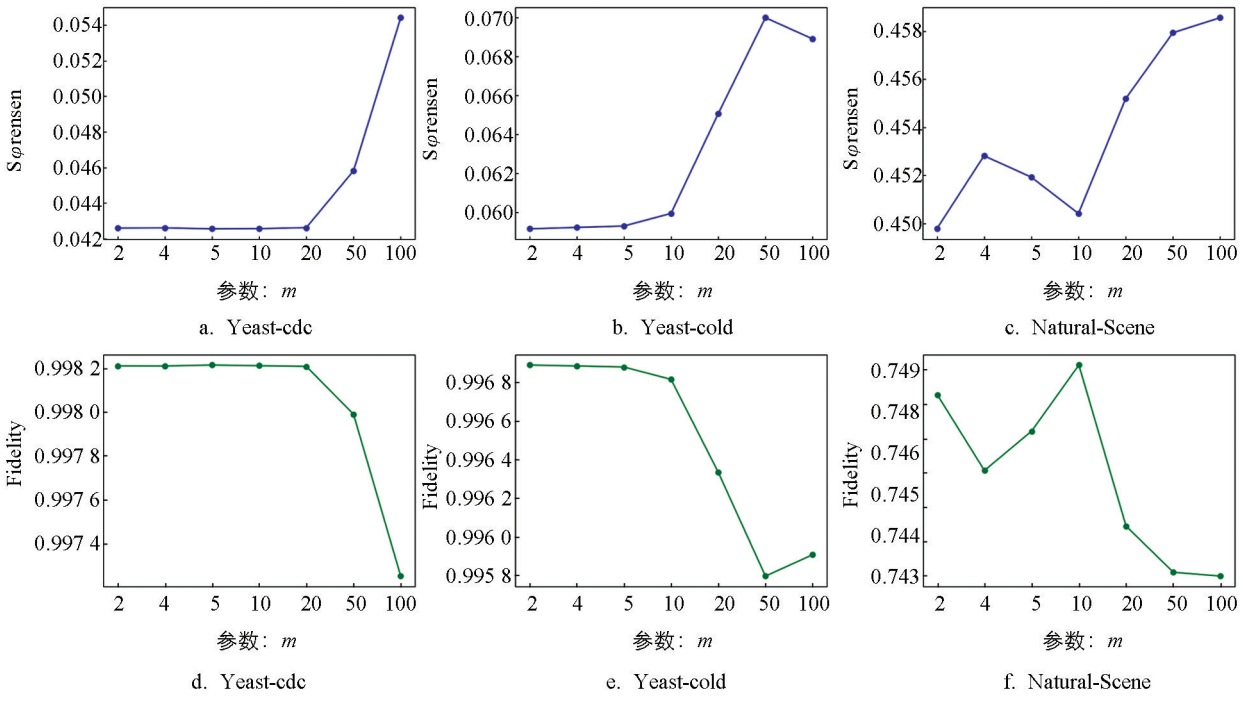

图 2　参数 *m* 对数据集 Yeast-cold、Yeast-cdc、Natural-Scene 的影响

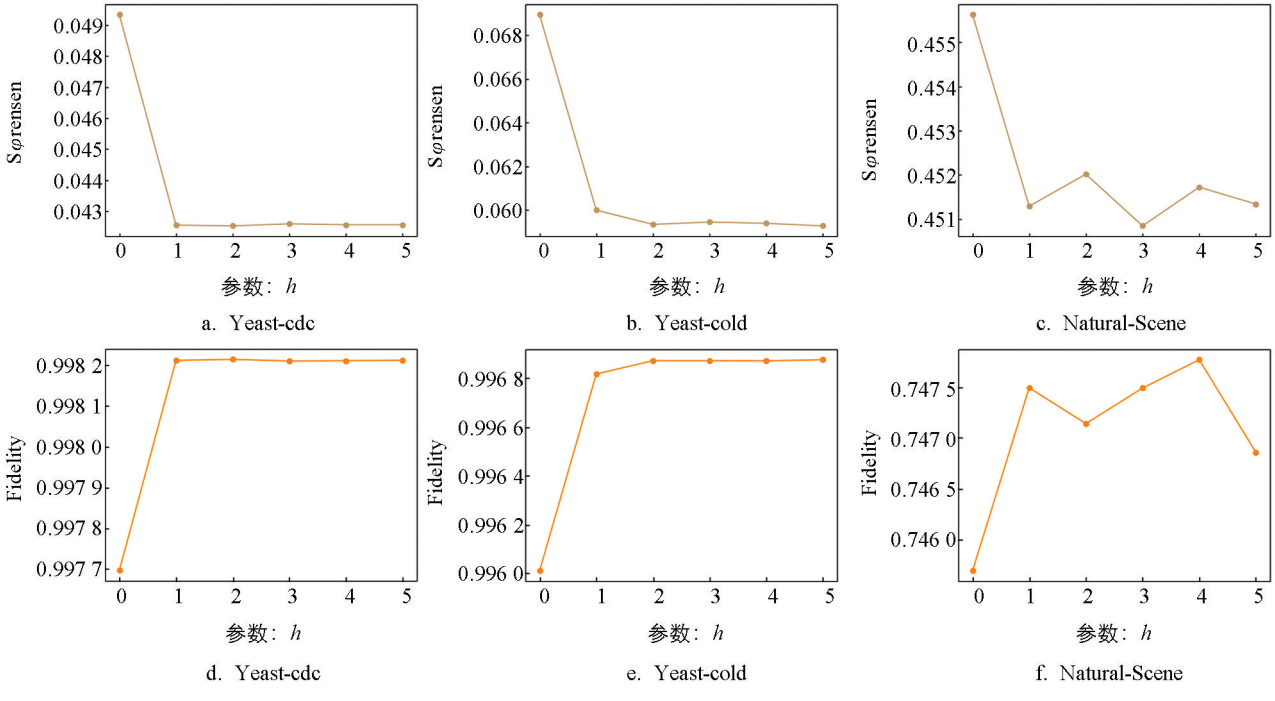

图 3　参数 *h* 对数据集 Yeast-cold、Yeast-cdc、Natural-Scene 的影响

3.6　实验结果

本文采用十折交叉验证的方法，将训练数据随机划分为 10 份，每次选取其中 9 份用作训练集，1 份作为测试集. 该过程被重复执行 10 次，并记录下数据集中每种算法排名的平均值和标准差，以确保评估结果的可靠性. 实验结果如表 3～4 所示，其中黑体数字表示在当前数据集下，该算法表现最佳. 通过对 7 种算法在所有数据集下的表现进行平均排名，可以清晰地看出 CSLDL 算法的有效性和优越性.

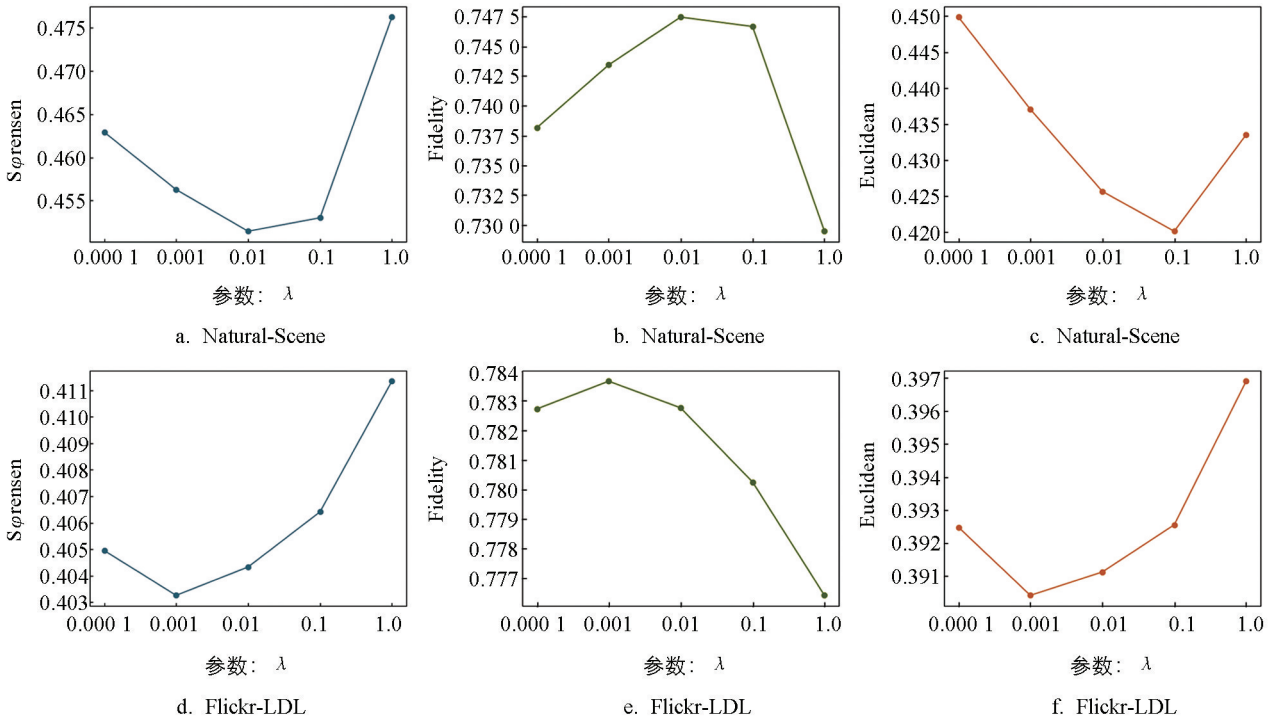

图4 参数λ对数据集Flickr-LDL、Natural-Scene的影响

表3 不同LDL算法的Sørensen比较结果(平均值±标准差)

数据集	PT-Bayes	AA-kNN	SA-BFGS	LDL-HR	LDL-LDM	LDL-SF	CSLDL
M2B	0.709 9±0.000 5	0.446 2±0.000 6	0.444 2±0.002 6	0.465 6±0.008 5	0.432 6±0.004 2	0.432 7±0.002 2	**0.427 3±0.002 6**
Emotion6	0.603 1±0.004 8	0.450 1±0.004 5	0.427 9±0.000 8	0.444 8±0.007 0	0.427 3±0.002 9	0.415 2±0.003 4	**0.412 6±0.001 1**
Movie	0.280 0±0.000 8	0.176 0±0.006 2	0.180 8±0.001 3	0.229 4±0.002 4	**0.172 5±0.000 4**	0.188 2±0.000 4	0.173 9±0.000 3
Twitter-LDL	0.543 6±0.006 3	0.399 9±0.001 9	0.381 2±0.002 0	0.642 5±0.002 7	0.484 3±0.001 8	0.377 2±0.000 4	**0.376 8±0.000 4**
Flickr-LDL	0.605 4±0.003 1	0.445 8±0.687 9	0.408 3±0.000 1	0.573 2±0.002 1	0.459 8±0.000 4	0.407 3±0.000 1	**0.405 4±0.000 2**
Natural-Scene	0.651 7±0.000 8	0.440 0±0.000 6	0.463 1±0.003 0	0.505 9±0.017 1	0.502 1±0.007 7	0.460 3±0.000 6	**0.450 8±0.000 8**
FBP5500	0.498 5±0.000 3	0.171 5±0.000 1	0.201 0±0.001 8	0.389 3±0.016 8	0.178 3±0.002 1	0.163 2±0.000 5	**0.154 5±0.000 9**
Yeast-cdc	0.235 7±0.003 5	0.047 2±0.000 1	0.042 9±0.000 4	0.047 6±0.000 8	0.043 2±0.000 3	0.042 7±0.000 2	**0.042 5±0.000 1**
Yeast-cold	0.211 3±0.003 5	0.064 3±0.000 1	0.059 5±0.000 9	0.059 8±0.000 9	0.059 6±0.000 3	0.059 3±0.000 2	**0.059 3±0.000 1**
Yeast-spoem	0.179 5±0.010 9	0.091 8±0.000 3	0.087 2±0.000 4	0.087 3±0.003 7	0.087 1±0.001 4	**0.086 9±0.000 2**	0.087 4±0.000 1
Avg. Rank	6.9	4.3	3.6	5.7	4.0	2.4	1.5

表4 不同LDL算法的Fidelity比较结果(平均值±标准差)

数据集	PT-Bayes	AA-kNN	SA-BFGS	LDL-HR	LDL-LDM	LDL-SF	CSLDL
M2B	0.489 6±0.001 5	0.819 7±0.001 3	0.818 5±0.001 9	0.832 1±0.005 5	**0.849 7±0.002 3**	0.822 7±0.002 0	0.836 7±0.001 9
Emotion6	0.614 7±0.003 9	0.772 3±0.003 6	0.795 9±0.000 2	0.783 6±0.006 3	0.795 0±0.002 5	0.801 5±0.002 3	**0.806 9±0.000 7**
Movie	0.928 5±0.000 4	0.970 2±0.000 3	0.968 7±0.000 3	0.956 6±0.000 9	**0.972 6±0.000 1**	0.961 9±0.000 2	0.971 8±0.000 1
Twitter-LDL	0.620 7±0.008 0	0.756 7±0.001 3	0.790 3±0.001 3	0.568 1±0.002 2	0.714 9±0.001 5	0.790 6±0.000 2	**0.793 1±0.000 4**
Flickr-LDL	0.586 9±0.001 4	0.737 2±0.462 9	0.777 9±0.000 4	0.645 6±0.001 7	0.736 8±0.000 6	0.777 2±0.000 1	**0.780 8±0.000 1**
Natural-Scene	0.588 7±0.000 8	0.728 0±0.000 9	0.737 7±0.001 7	0.673 0±0.016 4	0.708 6±0.005 3	0.730 0±0.000 5	**0.747 7±0.000 5**
FBP5500	0.751 2±0.000 2	0.960 2±0.000 1	0.944 6±0.000 3	0.807 1±0.014 4	0.956 2±0.001 4	0.959 9±0.000 2	**0.968 5±0.000 2**
Yeast-cdc	0.938 9±0.001 6	0.997 9±0.000 0	0.998 2±0.000 1	0.997 9±0.000 1	**0.998 2±0.000 0**	0.998 2±0.000 1	0.998 2±0.000 1
Yeast-cold	0.953 2±0.001 1	0.996 4±0.000 1	0.996 9±0.000 1	0.996 8±0.000 2	**0.996 9±0.000 0**	0.996 9±0.000 1	0.996 9±0.000 1
Yeast-spoem	0.966 5±0.003 6	0.992 8±0.000 1	0.993 5±0.000 1	0.993 2±0.000 9	0.993 2±0.000 0	**0.993 6±0.000 1**	**0.993 6±0.000 1**
Avg. Rank	6.9	4.5	3.2	5.5	3.1	2.6	1.4

4　总结与未来工作

标记分布学习展现出了强大的泛化能力，相较于单标记学习和多标记学习，它可以更有效地应对标记歧义问题. 然而，现有的标记分布学习算法主要是针对均匀分布数据而设计的，却常忽视了实际情况下训练数据中普遍存在的不均匀分布现象. 本文提出了一种基于代价敏感的标记分布学习算法，旨在处理非均匀分布数据. 实验结果表明，CSLDL 算法在处理这类数据时表现优于大多数现有的标记分布学习算法.

在未来的工作中，我们将尝试探索其他方法来处理非均匀分布数据，以取得更加高效的结果. 例如，可以考虑将数据合成、进行重采样等方法迁移到标记分布学习领域；也可以对原始数据进行增强处理，使其更容易被标记所描述等.

参考文献：

[1] CHEN C H，PATEL V M，CHELLAPPA R. Matrix Completion for Resolving Label Ambiguity [C] //2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston，MA，USA：IEEE，2015：4110-4118.

[2] GAO B B，XING C，XIE C W，et al. Deep Label Distribution Learning with Label Ambiguity [J]. IEEE Transactions on Image Processing，2017，26(6)：2825-2838.

[3] 周亮，陈辰，李宁. 基于机器学习和经验模态分解的跨期套利研究 [J]. 西南大学学报(自然科学版)，2022，44(1)：148-159.

[4] 杭立，车进，宋培源，等. 基于机器学习和图像处理技术的病虫害预测 [J]. 西南大学学报(自然科学版)，2020，42(1)：134-141.

[5] COUR T，SAPP B，TASKAR B. Learning from Partial Labels [J]. Journal of Machine Learning Research，2011，12：1501-1536.

[6] TSOUMAKAS G，KATAKIS I. Multi-Label Classification [J]. International Journal of Data Warehousing and Mining，2007，3(3)：1-13.

[7] ZHU Y，KWOK J T，ZHOU Z H. Multi-Label Learning with Global and Local Label Correlation [J]. IEEE Transactions on Knowledge and Data Engineering，2018，30(6)：1081-1094.

[8] ZHANG M L，WU L. Lift：Multi-Label Learning with Label-Specific Features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，37(1)：107-120.

[9] GENG X. Label Distribution Learning [J]. IEEE Transactions on Knowledge and Data Engineering，2016，28(7)：1734-1748.

[10] JIA X Y，LI W W，LIU J Y，et al. Label Distribution Learning by Exploiting Label Correlations [C] // Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans，Lousiana，USA：AAAI Press，2018：3310-3317.

[11] ZHENG X，JIA X Y，LI W W. Label Distribution Learning by Exploiting Sample Correlations Locally [C] // Proceedings of the AAAI Conference on Artificial Intelligence，New Orleans，Lousiana，USA：AAAI Press，2018：4556-4563.

[12] 容斌元，徐媛媛，吕亚兰，等. 融合标签局部相关性的标签分布学习 [J]. 山东大学学报(理学版)，2022，57(7)：53-64.

[13] WANG J，GENG X. Label Distribution Learning by Exploiting Label Distribution Manifold [J]. IEEE Transactions on Neural Networks and Learning Systems，2023，34(2)：839-852.

[14] KRAWCZYK B. Learning from Imbalanced Data：Open Challenges and Future Directions [J]. Progress in Artificial Intelligence，2016，5(4)：221-232.

[15] REN T T，JIA X Y，LI W W，et al. Label Distribution Learning with Label-Specific Features [C] //Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California：AAAI Press，2019：3318-3324.

[16] CHARTE F，RIVERA A J，DEL JESUS M J，et al. MLSMOTE：Approaching Imbalanced Multilabel Learning through Synthetic Instance Generation [J]. Knowledge-Based Systems，2015，89：385-397.

[17] TORGO L, RIBEIRO R P, PFAHRINGER B, et al. SMOTE for Regression [C] //Portuguese Conference on Artificial Intelligence. Berlin, Heidelberg: Springer, 2013: 378-389.

[18] Branco P, Torgo L, Ribeiro R P. Smogn: A Pre-processing Approach for Imbalanced Regression [C] // Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications. Skopje, Macedonia: Microtome Publishing, 2017: 36-50.

[19] WU G Q, TIAN Y J, LIU D L. Cost-Sensitive Multi-Label Learning with Positive and Negative Label Pairwise Correlations [J]. Neural Networks, 2018, 108: 411-423.

[20] ZHAO X Y, AN Y X, XU N, et al. Continuous Label Distribution Learning [J]. Pattern Recognition, 2023, 133: 109056.

[21] YANG Y Z, ZHA K W, CHEN Y C, et al. Delving into Deep Imbalanced Regression [C] //International Conference on Machine Learning. Virtual Event: Microtome Publishing, 2021: 11842-11851.

[22] 黄雨婷, 徐媛媛, 张恒汝, 等. 融合标签结构依赖性的标签分布学习 [J]. 南京大学学报(自然科学), 2020, 56(4): 524-532.

[23] NOCEDAL J, WRIGHT S J. Numerical optimization [M]. 2nd ed. New York: Springer, 2006.

[24] R Y, G X . Sense Beauty by Label Distribution Learning [C] // Proceedings of the 32th International Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017: 2648-2654.

[25] NGUYEN T V, LIU S, NI B B, et al. Sense Beauty via Face, Dressing, and/or Voice [C] //Proceedings of the 20th ACM International Conference on Multimedia. Nara, Japan: ACM, 2012: 239-248.

[26] PENG K C, CHEN T, SADOVNIK A, et al. A Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions [C] //2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 860-868.

[27] YANG J F, SUN M, SUN X X. Learning Visual Sentiment Distributions via Augmented Conditional Probability Neural Network [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 224-230.

[28] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection [C] //2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005: 886-893.

[29] STRICKER M A, ORENGO M. Similarity of Color Images [C] //Proc SPIE 2420, Storage and Retrieval for Image and Video Databases III. San Jose, CA, United States: SPIE Press, 1995: 381-392.

[30] LIANG L Y, LIN L J, JIN L W, et al. SCUT-FBP5500: a Diverse Benchmark Dataset for Multi-Paradigm Facial Beauty Prediction [C] //2018 24th International Conference on Pattern Recognition (ICPR). Beijing, China: IEEE, 2018: 1598-1603.

[31] LYONS M, AKAMATSU S, KAMACHI M, et al. Coding Facial Expressions with Gabor Wavelets [C] //Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. Nara, Japan: IEEE, 1998: 200-205.

[32] JIA X Y, LU Y N, ZHANG F W. Label Enhancement by Maintaining Positive and Negative Label Relation [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(2): 1708-1720.

[33] GENG X, YIN C, ZHOU Z H. Facial Age Estimation by Learning from Label Distributions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(10): 2401-2412.

责任编辑 崔玉洁