

基于标记相关性的多标记专属特征学习算法

李华*, 王志杰

(石家庄铁道大学 数理系, 河北 石家庄 050043)

摘要: 基于双标记专属特征的多标记分类算法(BILAS)是一种代表性的多标记学习算法,然而其只考虑了在标记对下取值不同的样本,忽略了取值相同的样本,使得生成的专属特征不能全面准确地刻画标记信息。针对这一不足,基于标记的二阶相关性,对标记对的全部类型样本生成专属特征,提出基于标记相关性的多标记专属特征学习算法。首先,利用基于距离的原型学习方法选择所有标记对的原型,并进一步生成相应的专属特征;然后利用标记幂集的思想构造多标记分类器。在来自MULAN(a Java library for multi-label learning)的5个公开测试数据集上进行实验,与BILAS算法和基于校准标记排序的多标记分类算法(CLR)相比,所提算法在5种多标记评价指标上综合平均排名均是第一,且分别比BILAS和CLR算法提高了20.4%和37.1%,表明了所提算法具有较好的性能。

关键词: 多标记学习;数据降维;相似度;原型学习;标记幂集

中图分类号: TP18

文献标志码: A

文章编号: 0253-2395(2024)01-0059-10

Multi-label Specific Features Learning Algorithm Based on Label Correlation

LI Hua*, WANG Zhijie

(Department of Mathematics and Physics, Shijiazhuang Tiedao University, Shijiazhuang 050043, China)

Abstract: Bilabel-specific features for multi-label classification algorithm (BILAS) is a representative multi-label learning algorithm. However, it only considers samples with different values for the label pair, and ignores samples with the same value, so that the generated label-specific features could not comprehensively and accurately characterize the label information. To weaken this shortcoming, based on the second-order correlation of labels, label-specific features are generated for all types of samples of the label pair, and a multi-label specific features learning algorithm based on label correlation is proposed. Firstly, the distance-based prototype learning method is used to select prototypes of all label pairs, and then the corresponding label-specific features are generated; furthermore, using the idea of label powerset, a multi-label classifier is constructed. Experimental results on five publicly available test datasets from MULAN (a Java library for multi-label learning) show that the proposed algorithm, compared to BILAS and multi-label classification algorithm via calibrated label ranking (CLR), ranks first in terms of the comprehensive average ranking on the five multi-label evaluation metrics. Furthermore, it achieves improvements of 20.4% and 37.1% compared to BILAS and CLR, respectively, demonstrating the effectiveness of the proposed algorithm.

Key words: multi-label learning; dimensionality reduction; similarity; prototype learning; label powerset

0 引言

多标记学习中,一个样本往往与多个标记同时相关联,学习的目的是为新样本预测一组

相关的标记。目前,多标记学习在图像注释^[1],生物信息学^[2],文本分类^[3],临床数据分析^[4]等多个方面有广泛的应用。然而多标记学习会遭受“维数灾难”^[5]问题,解决的方法是将

收稿日期:2023-06-09;接受日期:2023-08-15

基金项目:国家自然科学基金(61806133)

* 通信作者:李华(1978-),女,河北石家庄人,博士,副教授,研究方向为粒计算与数据挖掘。E-mail:lihua@stdu.edu.cn

引文格式:李华,王志杰.基于标记相关性的多标记专属特征学习算法[J].山西大学学报(自然科学版),2024,47(1):59-68. DOI:10.13451/j.sxu.ns.2023141

高维数据映射到低维数据,即维数约简。常见的方法有:特征提取^[6-8]和特征选择^[9-10]等。

基于是否考虑标记专属特征,多标记数据维数约简工作可分为两类。不考虑标记专属特征的多标记维数约简方法主要有基于粒计算理论的属性约简方法^[11-12],如Chen等^[13]基于核模糊粗糙集提出的谱特征选择算法;基于信息论的维数约简方法^[14-15],如Xiong等^[16]基于标记分布和模糊互信息提出的特征选择算法;以及一些利用最优化理论的特征选择方法^[17-18],如Wang等^[19]基于 $l_{2,0}$ 范数稀疏条件和离散条件提出重加权矩阵用于选择特征子集。

上述维数约简算法都是考虑标记与整个特征空间的关系。然而在实际中,标记是可以由专属于自身的某些特征来决定,这种专属于某个标记的特征叫作标记专属特征。一些学者基于标记专属特征进行了多标记数据的维数约简。Zhang等^[20]首次提出了基于标记专属特征的多标记学习算法(LIFT)。Xu等^[21]在LIFT算法的基础上,对标记专属特征空间利用模糊粗糙集理论进行特征选择,以此提高多标记学习算法的分类性能。Huang等^[22]通过学习标记专属特征处理二元分类问题。Guo等^[23]使用距离信息和空间拓扑信息重构标记专属特征空间,从而更好地利用聚类结果进行分类。

然而上述方法大都没有考虑标记之间的相关性,而借助于标记相关性可以极大地提高多标记学习算法的性能^[24]。Weng等^[25]通过局部标记相关性补充正类样本,提出了基于标记专属特征和局部成对标记相关性的多标记学习算法。Huang等^[26]提出局部正、负标记相关分析与标记意识相结合的多标记学习算法。Lin等^[27]通过结合多个标记专属特征,并基于标记相关性的加权和集成技术学习了一个统一的框架实现多标记学习的目标。Zhao等^[28]利用局部与全局标记相关性学习多标记专属特征。

特别地,Zhang等^[29]提出了基于双标记专属特征的多标记分类算法(BILAS)。其首先利用一种原型学习方法选择对于标记对最优的原型,并以此构造双标记专属特征;然后在新的特征空间上构造分类器集成,并且使用经验权

重比生成模型输出。这是首次从标记对的角度研究专属特征,并且提出了一种新的启发式原型学习方法去选择更优的原型,使得生成的专属特征更具可解释性和分离性。

然而BILAS算法仅考虑了在标记对下取值不同的样本,实际上在标记对下取值相同的样本也包含着标记间的相关信息,因此考虑标记对所有类型的样本能够更准确地刻画标记间的相关信息,研究其专属特征能够进一步提高多标记学习算法的性能。基于此,我们提出了基于标记相关性的多标记专属特征学习算法。首先利用基于余弦相似度得到的标记相关性将所有标记对分为相关标记对和不相关标记对,并将样本分成在标记对下取值相同的样本和取值不相同的样本;然后采用基于距离的原型学习方法选择出所有类型样本最优的原型,并利用距离映射生成相应的标记对专属特征;之后利用标记幂集(label powerset, LP)^[30]思想构造了一种多标记分类器。最后,在公开测试数据集上验证所提算法的有效性。

本文剩余部分组织如下,第1节提出基于标记相关性的多标记专属特征学习算法,包括标记对专属特征生成模型和多标记分类模型;第2节将所提算法与其他多标记学习算法进行对比试验及结果分析;第3节总结全文。

1 基于标记相关性的多标记专属特征学习算法

BILAS^[29]从标记对的角度研究专属特征,使得生成的专属特征更具可解释性和分离性。然而该算法只基于在标记对下取值不同的样本来生成专属特征,没有考虑在标记对下取值相同的样本,而后者也包含着标记间的相关信息。同时利用在标记对下取值相同和取值不相同的样本生成专属特征有利于提高多标记学习算法的性能。因此提出基于标记相关性的多标记专属特征学习算法,其包括两部分:标记对专属特征生成模型和多标记分类模型。

给定一个训练集 $D = \{(x_i, Y_i) | i = 1, 2, \dots, n\}$,其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \in X \subseteq R^d$ 是 d 维特征向量, $Y_i \subseteq L = \{l_1, l_2, \dots, l_q\}$ 是 x_i 的

相关标记集。这里若 x_i 与标记 l_u 相关,则 $l_u(x_i)=1$,反之 $l_u(x_i)=0$ 。多标记学习的目的是学习一个分类函数 $h: X \rightarrow P(L)$,为新的样本预测一组相关的标记,其中 $P(L)$ 为标记集 L 上的幂集。

对于不同标记而言,它们不一定都与所有的特征相关。事实上,标记可以由专属于自身的某些特征来决定^[20],这种专属于某个标记的特征叫作标记专属特征。下面首先基于标记相关性生成标记对专属特征。

1.1 标记对专属特征生成模型

这一节首先利用标记相关性将标记对分为相关标记对和不相关标记对,然后分别基于原型学习方法选择原型,并构造相应的标记对专属特征。

首先利用余弦相似度来计算两个标记间的相关性。

定义 1 给定一个多标记数据集 $D = \{(x_i, Y_i)\}$,其中 $i = 1, 2, \dots, n$, $x_i \in R^d$, $Y_i \subseteq L = \{l_1, l_2, \dots, l_q\}$ 是 x_i 的相关标记集。对于任意的标记 $l_u, l_v \in L$,令 $L_u = (l_u(x_1), l_u(x_2), \dots, l_u(x_n))^T$,则标记 l_u 和标记 l_v 的相关性为

$$r_{uv} = \frac{L_u^T \cdot L_v}{\|L_u\| \cdot \|L_v\|}, \quad (1)$$

由定义可知, $0 \leq r_{uv} \leq 1$ 。

根据标记相关性,本文将标记对分成相关标记对和不相关标记对。

定义 2 给定一个多标记数据集 $D = \{(x_i, Y_i)\}$,其中 $i = 1, 2, \dots, n$, $x_i \in R^d$, $Y_i \subseteq L = \{l_1, l_2, \dots, l_q\}$ 是 x_i 的相关标记集。对于任意的标记 $l_u, l_v \in L$,若 $0.5 \leq r_{uv} \leq 1$,则称 (l_u, l_v) 为相关标记对;若 $0 \leq r_{uv} < 0.5$,则称 (l_u, l_v) 为不相关标记对。

对于标记集 $L = \{l_1, l_2, \dots, l_q\}$ 而言,共有 $q(q-1)$ 个标记对,下面按照相关标记对和不相关标记对分别学习其专属特征。

1.1.1 相关标记对 (l_u, l_v)

对于标记对 (l_u, l_v) 而言,样本可以被分成四类:分别是具有标记 l_u ,不具有标记 l_v 的样本集 U_1^{uv} ;不具有标记 l_u ,具有标记 l_v 的样本集 U_2^{uv} ;不具有标记 l_u ,也不具有标记 l_v 的样本集 U_3^{uv} ;具有标记 l_u ,也具有标记 l_v 的样本集 U_4^{uv} ,即

$$\begin{aligned} U_1^{uv} &= \{x_i | (x_i, Y_i) \in D, l_u \in Y_i, l_v \notin Y_i\}, \\ U_2^{uv} &= \{x_i | (x_i, Y_i) \in D, l_u \notin Y_i, l_v \in Y_i\}, \\ U_3^{uv} &= \{x_i | (x_i, Y_i) \in D, l_u \notin Y_i, l_v \notin Y_i\}, \\ U_4^{uv} &= \{x_i | (x_i, Y_i) \in D, l_u \in Y_i, l_v \in Y_i\}. \end{aligned} \quad (2)$$

这四类样本蕴含了样本集对于这个标记对所有的相关信息,因此可作为生成专属特征的候选原型集合。

首先对于要选择的原型,其应满足在四类样本中具有中心性,同时与不同类型的样本有较大的区分性。也就是说,对于每类样本,原型与其中各个样本的距离都大致相同;而对于不同类型的样本,原型与其中各个样本的距离都较远。

具体来说,对于任意的特征向量 $\mu \in R^d$,不妨设 $\mu \in U_k^{uv}$, $k \in \{1, 2, 3, 4\}$ 。令 $D_1^{uv}, D_2^{uv}, D_3^{uv}, D_4^{uv}$ 分别为 μ 与四类样本的距离构成的集合,即

$$\begin{aligned} D_1^{uv} &= \{d(x, \mu) | x \in U_1^{uv}\}, \\ D_2^{uv} &= \{d(x, \mu) | x \in U_2^{uv}\}, \\ D_3^{uv} &= \{d(x, \mu) | x \in U_3^{uv}\}, \\ D_4^{uv} &= \{d(x, \mu) | x \in U_4^{uv}\}. \end{aligned} \quad (3)$$

需要指出的是,上述四个集合会随着样本 μ 的变化而发生变化。

令

$$d = \sum_{k=1}^4 \frac{1}{|U_k^{uv}|} \sum_{m, n \in D_k^{uv}} |m - n|, \quad (4)$$

其表示对于每个类型样本, μ 与其中各个样本距离的差异性,其值越小,则差异性越小;

$$D = \sum_{j \neq k} \left(\frac{1}{|U_k^{uv}| \cdot |U_j^{uv}|} \sum_{\substack{m \in D_j^{uv} \\ n \in D_k^{uv}}} |m - n| \right), \quad (5)$$

其表示对于与 μ 不同类型的样本, μ 与其中各个样本距离的差异性,其值越小,则差异性越小。

下面通过如下的公式给每个 μ 打分

$$s = t \cdot P_{\text{sep}} + (1 - t) \cdot P_{\text{dis}}, \quad (6)$$

其中

$$P_{\text{sep}} = \frac{d}{D}, \quad P_{\text{dis}} = \sum_{k=1}^4 \frac{1}{\sqrt{\text{var}(D_k^{uv})}}, \quad (7)$$

这里 $t \in (0, 1)$ 为权重参数; P_{sep} 表示对于同类型样本和不同类型样本, μ 与其中各样本的距离差异性,它的值越小,一方面说明对于每个类

型样本, μ 与其中各个样本距离的差异性越小, 即 μ 与其中各个样本的距离都大致相同; 另一方面说明对于不同类型样本, μ 与其中各个样本距离的差异性越大, 即 μ 与其中各个样本的距离都较远; P_{dis} 表示 μ 与 4 类样本距离标准差的倒数和, 其值越小, 说明 μ 与 4 类样本距离的差异性越大; 综上, μ 的分值越小, μ 越有可能被选作原型。

接下来, 按照 s 的值从低到高排序, 依次从每类样本中选出合适数目的样本。为了解决类不平衡问题, 这里从每类样本中均选择 c 个样本。 c 的选取方法如下

$$c = \left\lceil r \cdot \min_{1 \leq k \leq 4} (|U_k^{wv}|) \right\rceil, \quad (8)$$

其中 $\lceil w \rceil$ 表示对 w 向下取整, $r \in (0, 1]$ 是比例参数。也就是说从 $U_k^{wv} (k=1, 2, 3, 4)$ 中选择分值较小的 c 个样本, 记为 $x_k^1, x_k^2, \dots, x_k^c$, 称之为样本集 U_k^{wv} 的原型。

最后, 对于上述选定的原型, 通过计算样本 x 与所有原型的距离, 可以得到样本 x 的标记对专属特征如下:

$$\begin{aligned} \varphi_{wv}(x) = & (d(x, x_1^1), \dots, d(x, x_1^c), \\ & d(x, x_2^1), \dots, d(x, x_2^c), \\ & d(x, x_3^1), \dots, d(x, x_3^c), \\ & d(x, x_4^1), \dots, d(x, x_4^c)). \end{aligned} \quad (9)$$

注: 对任意的样本 x 和 y , 如果 $\varphi_{wv}(x) = \varphi_{wv}(y)$, 为保持数据集样本容量不变, 这里 $\varphi_{wv}(x)$ 和 $\varphi_{wv}(y)$ 同时保留。

1.1.2 不相关标记对 (l_u, l_v)

对于不相关标记对, 由于具有标记 l_u , 也具有标记 l_v 的样本数目极少或为 0, 因此这里仅考虑 3 种类型的样本集, 即

$$\begin{aligned} U_1^{wv} &= \{x_i | (x_i, Y_i) \in D, l_u \in Y_i, l_v \notin Y_i\}, \\ U_2^{wv} &= \{x_i | (x_i, Y_i) \in D, l_u \notin Y_i, l_v \in Y_i\}, \\ U_3^{wv} &= \{x_i | (x_i, Y_i) \in D, l_u \notin Y_i, l_v \notin Y_i\}. \end{aligned} \quad (10)$$

首先, 对于任意的特征向量 $\mu \in R^d$, 不妨设 $\mu \in U_k^{wv}, k \in \{1, 2, 3\}$ 。可类似相关标记对的原型学习方法, 通过如下的公式给每个 μ 打分

$$s = t \cdot P_{\text{sep}} + (1-t) \cdot P_{\text{dis}}, \quad (11)$$

其中

$$P_{\text{sep}} = \frac{d}{D}, \quad P_{\text{dis}} = \sum_{k=1}^3 \frac{1}{\sqrt{\text{var}(D_k^{wv})}}, \quad (12)$$

这里 d, D, D_k^{wv} 的定义同上节。

然后, 按照 s 的值从低到高排序, 依次从每类样本中选出 c 个样本, 这里

$$c = \left\lceil r \cdot \min_{1 \leq k \leq 3} (|U_k^{wv}|) \right\rceil, \quad (13)$$

也就是说从 $U_k^{wv} (k=1, 2, 3)$ 中选择分值较小的 c 个样本, 记为 $x_k^1, x_k^2, \dots, x_k^c$, 称之为样本集 U_k^{wv} 的原型。

最后, 对于上述选定的原型, 通过计算样本 x 与所有原型的距离, 可以得到样本 x 的标记对专属特征如下:

$$\begin{aligned} \varphi_{wv}(x) = & (d(x, x_1^1), \dots, d(x, x_1^c), \\ & d(x, x_2^1), \dots, d(x, x_2^c), \\ & d(x, x_3^1), \dots, d(x, x_3^c)). \end{aligned} \quad (14)$$

注: 对任意的样本 x 和 y , 如果 $\varphi_{wv}(x) = \varphi_{wv}(y)$, 为保持数据集样本容量不变, 这里 $\varphi_{wv}(x)$ 和 $\varphi_{wv}(y)$ 同时保留。

在生成标记对专属特征之后, 本文利用标记幂集的方法设计了多标记分类模型。

1.2 多标记分类模型

首先对于每个标记对 (l_u, l_v) , 会有四种不同的形式: 分别是具有标记 l_u , 不具有标记 l_v 的标记对, 记为 $(1, 0)$; 不具有标记 l_u , 具有标记 l_v 的标记对, 记为 $(0, 1)$; 不具有标记 l_u , 也不具有标记 l_v 的标记对, 记为 $(0, 0)$; 具有标记 l_u , 也具有标记 l_v 的标记对, 记为 $(1, 1)$ 。根据标记幂集的方法, 令 $\sigma: (l_u, l_v) \mapsto \{1, 2, 3, 4\}$ 是一个单射函数, 即 $\sigma(1, 0) = 1, \sigma(0, 1) = 2, \sigma(0, 0) = 3, \sigma(1, 1) = 4; \sigma^{-1}$ 是相应的逆函数。定义多分类训练集:

$$D_{wv} = \{(\varphi_{wv}(x), \sigma(l_u, l_v)) | x \in U_k^{wv}, k=1, 2, 3, 4\}. \quad (15)$$

由上一小节可知, 当 (l_u, l_v) 为相关标记对时, D_{wv} 的特征维度为 $4c$; 当 (l_u, l_v) 为不相关标记对时, D_{wv} 的特征维度为 $3c$ 。

然后在 D_{wv} 上采用某种多分类学习算法 B (本文使用 SVM 执行多次二分类过程实现该多分类算法), 训练多分类器 $g_{wv}: \varphi_{wv}(x) \rightarrow \{1, 2, 3, 4\}$, 然后利用 σ^{-1} 将多分类预测结果转化为标记对, 记为 (y_u, y_v) 。

最后对于一个待预测的样本 x' , 计算其关于任意标记 l_u 的隶属度为

$$f(x', l_u) = \frac{\sum_{v=1}^q \mathbb{I}[(y_u, y_v) = (1, y_v)]}{q}, v \neq u, (16)$$

其中 $\mathbb{I}[w] = 1$, 当 w 成立时; 否则 $\mathbb{I}[w] = 0$ 。由公式(16)可知, $f(x', l_u) \in [0, 1]$ 。

相应地, x' 的相关标记集可以被预测为

$$h(x') = \{l_u | f(x', l_u) \geq 0.5, 1 \leq u \leq q\}. (17)$$

综上, 基于标记相关性的多标记专属特征学习算法过程如算法1所示。

算法1 基于标记相关性的多标记专属特征学习算法(LSF_LC)	
输入:	多标记训练集 $\{(x_i, Y_i) 1 \leq i \leq n\}$, 多分类学习算法 B , 平衡参数 t , 比例参数 r , 待预测样本 x'
输出:	x' 的预测标记集 $h(x')$
1.	由公式(1)计算标记相关性 r ;
2.	for $u = 1$ to q do
3.	for $v = 1$ to q do
4.	if $0.5 \leq r_{uv} \leq 1$ then
5.	由公式(2)得到候选原型集合 $U_1^{uv}, U_2^{uv}, U_3^{uv}, U_4^{uv}$;
6.	利用公式(6)计算每个候选原型 $\mu \in R^d$ 的得分 s ;
7.	从每类样本中选择 s 值较低的 c 个原型 $x_k^1, x_k^2, \dots, x_k^c$, $k = 1, 2, 3, 4$;
8.	由公式(9)生成相关标记对专属特征 $\varphi_{uv}(x)$;
9.	由公式(15)定义新的训练集 D_{uv} , 并基于 B 生成分类模型; 然后将分类预测结果转化为标记对 (y_u, y_v) ;
10.	else if $0 \leq r_{uv} < 0.5$ then
11.	由公式(10)得到候选原型集合 $U_1^{uv}, U_2^{uv}, U_3^{uv}$;
12.	由公式(11)计算每个候选原型 $\mu \in R^d$ 的得分 s ;
13.	从每类样本中选择 s 值较低的 c 个原型 $x_k^1, x_k^2, \dots, x_k^c$, $k = 1, 2, 3$;
14.	由公式(14)生成不相关标记对专属特征 $\varphi_{uv}(x)$;
15.	由公式(15)定义新的训练集 D_{uv} , 并基于 B 生成分类模型; 然后将分类预测结果转化为标记对 (y_u, y_v) ;
16.	end if
17.	end for
18.	end for
19.	由公式(16)计算标记隶属度 $f(x', l_u)$;
20.	由公式(17)计算 x' 的预测标记集 $h(x')$ 。

在算法1中, 首先第1步利用余弦相似度计算标记相关性; 然后第2—9步根据原型学习方法选择最优的原型, 进一步生成相关标记对的专属特征, 并在新的训练集上构造分类模型; 第10—18步生成不相关标记对的专属特征, 并在新的训练集上构造分类模型; 最后第19—20步计算标记隶属度并获得待预测样本的预测标记集。

LSF_LC算法的时间复杂度主要由生成标

记对专属特征的过程和构造多标记分类模型的过程决定。其中生成专属特征过程的时间复杂度为 $O(n^2(n+nd)q^2)$, 这里 n 为样本个数, q 为标记个数, d 为特征个数; 构造多标记分类模型过程的时间复杂度为 $O(n^3q^2 + n^2dq^2)$ 。综上整个算法的时间复杂度为 $O(n^2(n+nd)q^2 + n^3q^2 + n^2dq^2)$ 。

2 实验

本节将所提的基于标记相关性的多标记专属特征学习算法(LSF_LC)与两种从标记对的角度处理多标记数据的算法, 即BILAS^[29]和基于校准标记排序的多标记分类算法(CLR)^[31], 分别从5种多标记分类评价指标方面和算法运行时间分析方面进行性能的比较, 并进一步在统计意义下对三种算法的性能展开分析。

下面对三种算法进行说明:

BILAS对于标记对下取值不同的样本研究其专属特征, 在新的特征空间上构造分类器集成, 并且使用经验权重比生成模型输出, 本节依据文献[29]对其进行参数设置;

CLR通过比较标记对的两个标记输出值进行多标记排序, 本节依据文献[31]对其进行参数设置, 值得指出的是该算法不考虑标记专属特征。

LSF_L是本文所提算法, 其中在区间[0.1, 0.9]中以步长0.1进行遍历, 得到实验结果最优的参数组合 $r = 0.1, t = 0.5$ 。

所有实验均在参数为2.40 GHz Intel Core i5-6200U CPU和4 GB RAM的服务器上进行。

首先实验选择了5种常见的多标记数据集, 具体信息如表1所示。同时, 为了保证数据的准确性, 对于每个算法, 在前三个数据集上进行十折交叉验证, 在后两个数据集上进行五折交叉验证。

2.1 评价指标

5个多标记评价指标被用来评估所提的LSF_LC与其他2种多标记学习算法的性能。

对于样本 x_i , 设其真实标记集为 Y_i , 预测标记集为 $h(x_i)$ 。对于某些涉及标记排序的评价指标, 需要将预测函数 $f: X \times L \rightarrow R$ 的结果 $f(x, l_u)$ 转化为排序结果 $r_f(x, l_u)$, 也就是将

表1 多标记数据集

Table 1 Brief description of multi-label datasets

数据集	样本数	特征数	标记数	应用域
flags	194	19	7	图像
music	593	72	6	音乐
scene	2 407	294	6	图像
yeast	2 417	103	14	生物
recreation	5 000	606	22	文本

$f(x, l_u)$ 映射到集合 $\{1, 2, \dots, q\}$, 并且如果 $f(x, l_1) > f(x, l_2)$, 那么 $r_f(x, l_1) < r_f(x, l_2)$ 。

下面是本文用到的5种多标记评价指标。

(1) Hamming Loss^[11]: 该指标表示预测所得标记与样本真实标记之间的不一致程度, 即样本的相关标记被预测成不相关标记, 或样本的不相关标记被预测成相关标记的平均比率。

$$M_{HL}(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|L|} |h(x_i) \Delta Y_i|,$$

其中 Δ 是两个集合之间的对称差。 $M_{HL}(f)$ 的值越小, 性能越好。

(2) macro-averaging AUC^[29]: 该指标度量了所有类标记上的平均 AUC 值, 其中 AUC 值表示正样本预测值大于负样本预测值的比率。

$$M_{AUC}(f) = \frac{1}{q} \sum_{u=1}^q \frac{|\{(x', x'') | f(x', l_u) \geq f(x'', l_u)\}|}{|P_u| |N_u|},$$

其中 $(x', x'') \in P_u \times N_u$, $P_u = \{x_i | l_u \in Y_i, 1 \leq i \leq n\}$, $N_u = \{x_i | l_u \notin Y_i, 1 \leq i \leq n\}$ 。这里 $M_{AUC}(f)$ 值越大, 性能越好。

(3) Ranking Loss^[11]: 该指标表示标记对被错误排序的平均比率。

$$M_{RL}(f) = \frac{1}{n} \sum_{i=1}^n \frac{|\{(l, l') | f(x_i, l) \leq f(x_i, l')\}|}{|Y_i| |\overline{Y}_i|},$$

其中 $(l, l') \in Y_i \times \overline{Y}_i$, $\overline{Y}_i = L - Y_i$ 是不相关的标记集。 $M_{RL}(f)$ 值越小, 性能越好。

(4) Average Precision^[11]: 该指标表示在样本的标记排序队列中, 排在相关标记前面的标记仍是相关标记的平均比率。

$$M_{AP}(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|Y_i|} \sum_{l \in Y_i} \frac{|\{l' \in Y_i | r_f(x_i, l') \leq r_f(x_i, l)\}|}{r_f(x_i, l)},$$

其中 n 为对象数。 $M_{AP}(f)$ 值越大, 性能越好。

(5) Coverage^[11]: 该指标表示在样本的标记排序队列中, 从排序最高的相关标记开始,

平均需要搜索多少次才能覆盖样本的所有相关标记。

$$M_{COV}(f) = \frac{1}{n} \sum_{i=1}^n \max_{l \in Y_i} r_f(x_i, l) - 1,$$

这里 $M_{COV}(f)$ 值越小, 性能越好。

2.2 实验结果与分析

这一小节从5种评价指标和运行时间分析两个方面对比3种算法的性能, 并在统计意义下对三种算法的性能展开分析。

(1) 分类性能方面。3种算法在5个数据集上对应于5种指标的结果如表2至表6所示, 其中“ \uparrow ”表示指标值越大性能越好, “ \downarrow ”表示指标值越小性能越好。每个数据集上对应指标的最好结果都由粗体表示。

从表3和表5中可以看出在五个数据集上 LSF_LC 算法关于 Coverage 和 macro-averaging AUC 评价指标的结果都是最优的, 其次是 BILAS 算法, 再次是 CLR 算法。也就是说, LSF_LC 算法能够更好地预测样本的全部相关标记。

从表2可以看出 LSF_LC 算法对于 Average Precision 指标而言, 在4个数据集上的结果是最优的, 而在剩下 yeast 数据集上的结果仅次于 BILAS 算法。这表明 LSF_LC 算法可以更有效地预测相关标记的排序。

表4和表6中展示的结果表明 LSF_LC 算法的性能结果和其他算法相比大致相同。例如 LSF_LC 算法在 music 数据集上关于 Hamming Loss 和 Ranking Loss 的结果要优于其他对比算法, 而在 yeast 数据集上的结果则相反。

从表2至表6可以得到, LSF_LC 算法、BILAS 算法和 CLR 算法在五种评价度量上的平均排名分别为 1.56、1.96 和 2.48。因此, LSF_LC 算法相较于 BILAS 算法和 CLR 算法的平均排名分别高了 20.4% 和 37.1%。这表明所提 LSF_LC 算法具有较好的分类性能。

另一方面, 图1至图5展示了3种算法在各个数据集上关于5个评价指标值的性能对比, 其中最优结果由红色字体表示。

比较两种基于标记专属特征的多标记学习算法, 即 LSF_LC 算法和 BILAS 算法, 从图2能够看到 LSF_LC 算法在 music 数据集上的实验

表2 各算法在5个数据集上的 Average Precision 结果(平均值±方差)

Table 2 Average Precision results of different algorithms on 5 datasets (mean ± std)

对比算法	Average Precision ↑					平均排名
	scene	music	flags	yeast	recreation	
LSF_LC	0.819 2±0.000 1	0.742 3±0.000 7	0.566 7±0.018 7	0.464 5±0.000 1	0.273 3±0.000 0	1.2
BILAS	0.693 9±0.001 8	0.538 3±0.002 4	0.550 7±0.001 4	0.644 3±0.000 9	0.224 2±0.000 1	1.8
CLR	0.430 8±0.005 2	0.535 5±0.009 2	0.479 8±0.009 9	0.413 1±0.002 3	0.180 0±0.002 3	3.0

表3 各算法在5个数据集上的 Coverage 结果(平均值±方差)

Table 3 Coverage results of different algorithms on 5 datasets (mean ± std)

对比算法	Coverage ↓					平均排名
	scene	music	flags	yeast	recreation	
LSF_LC	0.105 4±0.000 0	0.346 5±0.001 1	0.779 1±0.001 3	0.724 7±0.000 1	0.472 1±0.000 0	1.0
BILAS	0.233 3±0.001 1	0.610 9±0.002 5	0.794 4±0.000 6	0.730 4±0.002 8	0.650 8±0.000 6	2.4
CLR	0.414 0±0.006 6	0.540 7±0.010 9	0.808 7±0.007 8	0.774 1±0.002 2	0.643 1±0.003 8	2.6

表4 各算法在5个数据集上的 Hamming Loss 结果(平均值±方差)

Table 4 Hamming Loss results of different algorithms on 5 datasets (mean ± std)

对比算法	Hamming Loss ↓					平均排名
	scene	music	flags	yeast	recreation	
LSF_LC	0.222 8±0.000 0	0.243 6±0.000 4	0.562 7±0.001 5	0.464 9±0.000 2	0.399 7±0.000 2	2.6
BILAS	0.163 2±0.000 2	0.328 6±0.001 3	0.532 1±0.002 1	0.250 5±0.000 3	0.065 4±0.000 0	1.4
CLR	0.221 3±0.002 6	0.699 1±0.015 0	0.485 2±0.011 2	0.303 5±0.002 9	0.318 5±0.035 6	2.0

表5 各算法在5个数据集上的 macro-averaging AUC 结果(平均值±方差)

Table 5 Macro-averaging AUC results of different algorithms on 5 datasets (mean ± std)

对比算法	macro-averaging AUC ↑					平均排名
	scene	music	flags	yeast	recreation	
LSF_LC	0.888 4±0.000 0	0.780 9±0.000 7	0.545 6±0.004 2	0.650 7±0.000 3	0.706 2±0.000 1	1.0
BILAS	0.779 4±0.001 0	0.523 7±0.001 2	0.495 0±0.003 0	0.511 3±0.001 2	0.453 4±0.001 0	2.0
CLR	0.520 6±0.007 7	0.506 9±0.010 9	0.299 7±0.010 9	0.472 9±0.003 7	0.403 6±0.002 7	3.0

表6 各算法在5个数据集上的 Ranking Loss 结果(平均值±方差)

Table 6 Ranking Loss results of different algorithms on 5 datasets (mean ± std)

对比算法	Ranking Loss ↓					平均排名
	scene	music	flags	yeast	recreation	
LSF_LC	0.147 9±0.000 0	0.271 6±0.000 9	0.692 8±0.004 2	0.551 3±0.000 1	0.455 3±0.000 0	2.0
BILAS	0.258 1±0.001 6	0.527 8±0.004 8	0.621 8±0.003 2	0.361 6±0.001 8	0.604 5±0.000 6	2.2
CLR	0.520 6±0.007 7	0.506 9±0.014 1	0.304 6±0.012 4	0.472 9±0.003 9	0.403 6±0.002 7	1.8

结果关于五个指标都是最优的;从图1可以看到LSF_LC算法在scene数据集有4个指标的结果是最优的;由图3—图5可知,总体而言,在其余的数据集上,LSF_LC算法优于BILAS算法。这表明学习标记对所有类型样本的专属特征,是能够更有效地提高多标记学习算法性能。

由图1至图5可知,总体而言,两种基于标记专属特征的多标记学习算法,即LSF_LC算

法和BILAS算法在5种评价指标上优于不考虑标记的专属特征的多标记学习算法CLR。这表明标记专属特征可以提高多标记学习算法的性能。

(2)算法运行时间方面。LSF_LC、BILAS^[29]、CLR^[31]三个算法的时间复杂度分别为 $O(n^2(n+nd)q^2+n^3q^2+n^2dq^2)$ 、 $O(ndq^2+n)$ 、 $O(nq)$ 。图6给出了三个算法运行时间上的比较。从中可以看出,LSF_LC算法运行时间最长,这是由于它

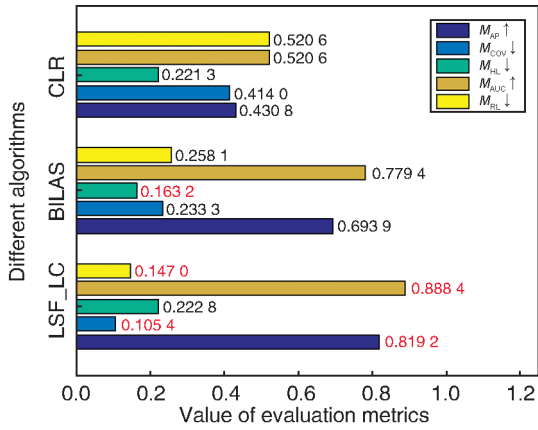


图 1 各算法在 scene 数据集上五种评价度量的结果
Fig. 1 Results of five evaluation measures of different algorithms on scene data set

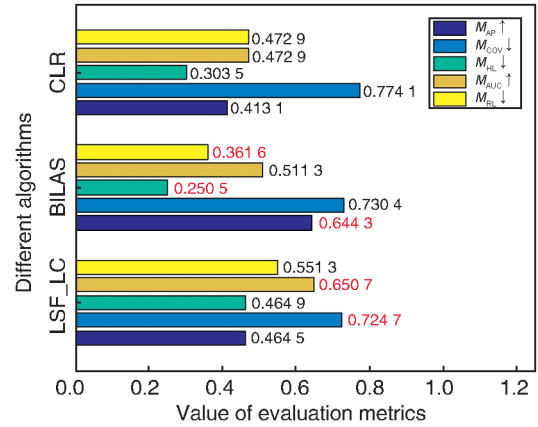


图 4 各算法在 yeast 数据集上五种评价度量的结果
Fig. 4 Results of five evaluation measures of different algorithms on scene data set

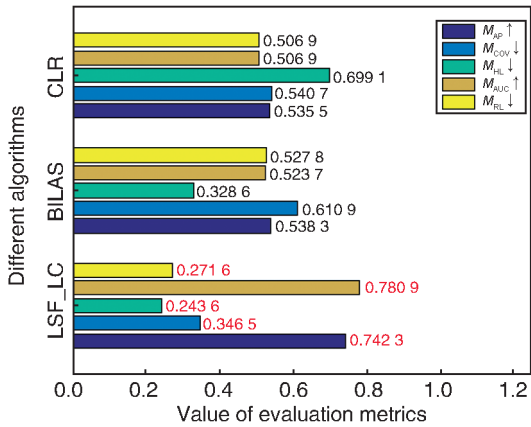


图 2 各算法在 music 数据集上五种评价度量的结果
Fig. 2 Results of five evaluation measures of different algorithms on scene data set

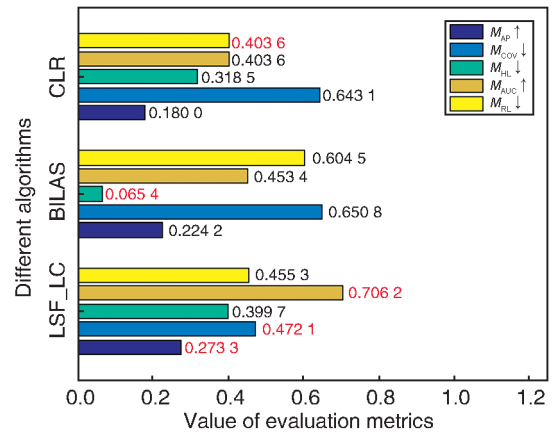


图 5 各算法在 recreation 数据集上五种评价度量的结果
Fig. 5 Results of five evaluation measures of different algorithms on scene data set

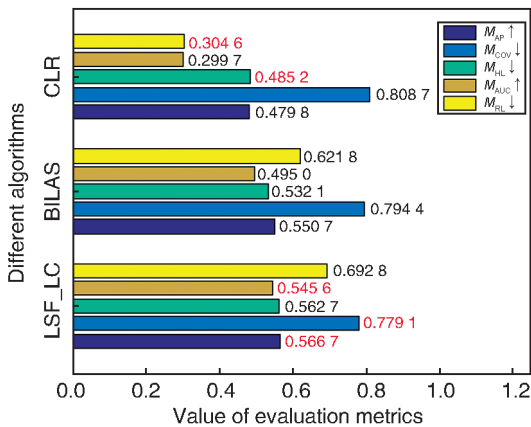


图 3 各算法在 flags 数据集上五种评价度量的结果
Fig. 3 Results of five evaluation measures of different algorithms on scene data set

BILAS算法和CLR算法总共有 $q(q-1)/2$ 个标记对,而本文所提的LSF_LC算法有 $q(q-1)$ 个标记对。此外由于本文所提的LSF_LC算法考虑了标记对下取值不同和取值相同的样本,导致计算量翻倍增长。

综上,结合分类性能的表现,LSF_LC算法与另外两种算法相比,虽然分类性能有所提高,但运行时间也相应骤增。

下面从统计学角度分析各算法之间的性能,这里采用常见的Friedman测试^[29]来比较在 N 个数据集上的 K 个比较算法。首先将每个算法在各个数据集上的结果进行排序。Friedman测试比较的是每个算法的平均排序,即

$$R_j = \frac{1}{N} \sum_{i=1}^K r_i^j,$$

的时间复杂度最高。另一个影响算法运行时间的因素是这三种算法计算标记对的个数,其中

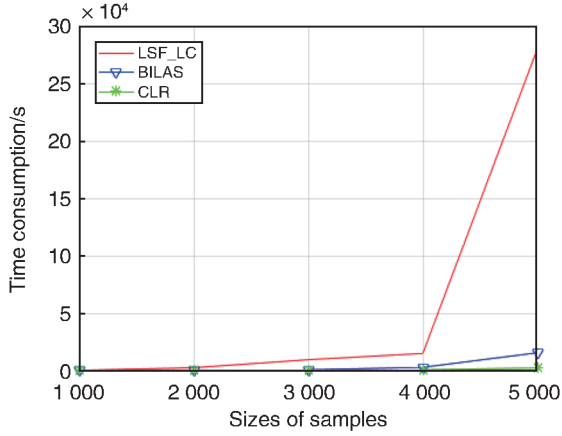


图6 各算法运行时间的对比

Fig. 6 Comparison of time consumption among different algorithms

其中 r_i^j 表示的是第 j 个算法在第 i 个数据集上的性能结果排序。在原假设下,所有的算法性能是相同的,即 R_j 均相同。Friedman 统计 F_F 是服从自由度为 $(K-1)$ 和 $(K-1)(N-1)$ 的 F 分布,其定义为

$$F_F = \frac{(N-1)\chi_F^2}{N(K-1) - \chi_F^2},$$

其中 $\chi_F^2 = \frac{12N}{K(K+1)} \left(\sum_{i=1}^K R_i^2 - \frac{K(K+1)^2}{4} \right)$ 。

当 F_F 值大于在显著性水平 α 下的临界值,则拒绝所有算法性能相同的原假设。本文中 $N=5, K=3$, 显著性水平 $\alpha=0.05$ 。表7给出了在5个评价指标上的 F_F 值和显著性水平 $\alpha=0.05$ 下的临界值。

表7 在5个评价度量上的Friedman统计 F_F Table 7 Friedman statistics F_F on 5 evaluation measures

评价度量	F_F	临界值($\alpha=0.05$)
Average Precision	7.111 1	
Coverage	12.666 7	
Hamming Loss	-0.491 2	2.239
macro-averaging AUC	Inf(∞)	
Ranking Loss	0.166 7	

从表7可以看出,在 Average Precision, Coverage, macro-averaging AUC 三个指标上,所有算法性能相同的原假设被拒绝,即 LSF_LC 算法具有较好性能。而在 Hamming Loss 和 Ranking Loss 两个指标上不能拒绝原假设,因此从统计角度来看,LSF_LC 算法的性能结果和其他算法相比大致相同。

3 结论

基于标记的二阶相关性,本文所提的多标记专属特征学习算法是同时考虑标记对下取值相同和不同的样本,采用基于距离的原型学习方法选择出所有类型样本的原型,并利用距离映射生成相应的标记对专属特征,之后利用标记聚集思想在新的特征空间上构造多标记分类模型。多标记公开测试数据集上的实验表明所提算法能够有效地提高多标记学习算法性能。

然而本文在构造多标记分类模型时,没有充分考虑标记的相关性,下一步将在考虑标记相关性的基础上,提出具有更好性能的多标记分类模型。同时,本文提出的算法时间复杂度过高,下一步将研究优化算法,提高算法效率。

参考文献:

- [1] LI P, CHEN P, ZHANG D Z. Cross-modal Feature Representation Learning and Label Graph Mining in a Residual Multi-attentional CNN-LSTM Network for Multi-label Aerial Scene Classification[J]. *Remote Sens*, 2022, **14**(10): 2424. DOI: 10.3390/rs14102424.
- [2] LI Z J, LIAO B, LI Y, et al. Gene Function Prediction Based on Combining Gene Ontology Hierarchy with Multi-instance Multi-label Learning[J]. *RSC Adv*, 2018, **8**(50): 28503-28509. DOI: 10.1039/C8RA05122D.
- [3] YAN K B, LI Z X, ZHANG C L. A New Multi-instance Multi-label Learning Approach for Image and Text Classification[J]. *Multimed Tools Appl*, 2016, **75**(13): 7875-7890. DOI: 10.1007/s11042-015-2702-6.
- [4] HUANG M X, HAN H R, WANG H, et al. A Clinical Decision Support Framework for Heterogeneous Data Sources[J]. *IEEE J Biomed Health Inform*, 2018, **22**(6): 1824-1833. DOI: 10.1109/JBHI.2018.2846626.
- [5] ZHANG Y, ZHOU Z H. Multilabel Dimensionality Reduction via Dependence Maximization[J]. *ACM Trans Knowl Discov From Data*, **4**(3): 1-21. DOI: 10.1145/1839490.1839495.
- [6] JOLLIFFE I T. Principal Component Analysis[M]. 2nd ed. New York: Springer, 2002.
- [7] ZHANG H J, QIANG W W, ZHANG J X, et al. Unified Feature Extraction Framework Based on Contrastive Learning[J]. *Knowl Based Syst*, 2022, **258**: 110028. DOI: 10.1016/j.knsys.2022.110028.
- [8] ZHANG H J, ZHAO S Y, QIANG W W, et al. Feature Extraction Framework Based on Contrastive Learning with

- Adaptive Positive and Negative Samples[J]. *Neural Netw*, 2022, **156**: 244–257. DOI: 10.1016/j.neunet.2022.09.029.
- [9] ZHANG P, LIU G X, GAO W F, *et al.* Multi-label Feature Selection Considering Label Supplementation[J]. *Pattern Recognit*, 2021, **120**: 108137. DOI: 10.1016/j.patcog.2021.108137.
- [10] YANG X L, CHEN H M, LI T R, *et al.* Neighborhood Rough Sets with Distance Metric Learning for Feature Selection[J]. *Knowl Based Syst*, 2021, **224**: 107076. DOI: 10.1016/j.knosys.2021.107076.
- [11] LI H, LI D Y, ZHAI Y H, *et al.* A Novel Attribute Reduction Approach for Multi-label Data Based on Rough Set Theory[J]. *Inf Sci*, 2016, **367/368**: 827–847. DOI: 10.1016/j.ins.2016.07.008.
- [12] TAN A H, LIANG J Y, WU W Z, *et al.* Fuzzy Rough Discrimination and Label Weighting for Multi-label Feature Selection[J]. *Neurocomputing*, 2021, **465**: 128–140. DOI: 10.1016/j.neucom.2021.09.007.
- [13] CHEN J K, LIN Y J, MI J S, *et al.* A Spectral Feature Selection Approach with Kernelized Fuzzy Rough Sets[J]. *IEEE Trans Fuzzy Syst*, 2022, **30**(8): 2886–2901. DOI: 10.1109/TFUZZ.2021.3096212.
- [14] WANG Z H, CHEN H M, YUAN Z, *et al.* Exploiting Fuzzy Rough Mutual Information for Feature Selection[J]. *Appl Soft Comput*, 2022, **131**: 109769. DOI: 10.1016/j.asoc.2022.109769.
- [15] HU L, GAOLB, LI Y H, *et al.* Feature-specific Mutual Information Variation for Multi-label Feature Selection[J]. *Inf Sci*, 2022, **593**: 449–471. DOI: 10.1016/j.ins.2022.02.024.
- [16] XIONG C Z, QIAN W B, WANG Y L, *et al.* Feature Selection Based on Label Distribution and Fuzzy Mutual Information[J]. *Inf Sci*, 2021, **574**: 297–319. DOI: 10.1016/j.ins.2021.06.005.
- [17] RANJAN R, CHHABRA J K. Automatic Clustering and Feature Selection Using Multi-objective Crow Search Algorithm[J]. *Appl Soft Comput*, 2023, **142**: 110305. DOI: 10.1016/j.asoc.2023.110305.
- [18] LIN Q, YANG L R, ZHONG P, *et al.* Robust Supervised Multi-view Feature Selection with Weighted Shared Loss and Maximum Margin Criterion[J]. *Knowl Based Syst*, 2021, **229**: 107331. DOI: 10.1016/j.knosys.2021.107331.
- [19] WANG J Y, WANG H M, NIE F, *et al.* Feature Selection with Multi-class Logistic Regression[J]. *Neurocomputing*, 2023, **543**: 126268. DOI: 10.1016/j.neucom.2023.126268.
- [20] ZHANG M L, WU L. Lift: Multi-label Learning with Label-specific Features[J]. *IEEE Trans Pattern Anal Mach Intell*, 2015, **37**(1): 107–120. DOI: 10.1109/TPAMI.2014.2339815.
- [21] XU S P, YANG X B, YU H L, *et al.* Multi-label Learning with Label-specific Feature Reduction[J]. *Knowl Based Syst*, 2016, **104**: 52–61. DOI: 10.1016/j.knosys.2016.04.012.
- [22] HUANG J, LI G R, HUANG Q M, *et al.* Learning Label-specific Features and Class-dependent Labels for Multi-label Classification[J]. *IEEE Trans Knowl Data Eng*, 2016, **28**(12): 3309–3323. DOI: 10.1109/TKDE.2016.2608339.
- [23] GUO Y M, CHUNG F, LI G Z, *et al.* Leveraging Label-specific Discriminant Mapping Features for Multi-label Learning[J]. *ACM Trans Knowl Discov From Data*, **13**(2): 1–23. DOI: 10.1145/3319911.
- [24] CHE X Y, CHEN D G, MI J S. Label Correlation in Multi-label Classification Using Local Attribute Reductions with Fuzzy Rough Sets[J]. *Fuzzy Sets Syst*, 2022, **426**: 121–144. DOI: 10.1016/j.fss.2021.03.016.
- [25] WENG W, LIN Y J, WU S X, *et al.* Multi-label Learning Based on Label-specific Features and Local Pairwise Label Correlation[J]. *Neurocomputing*, 2018, **273**: 385–394. DOI: 10.1016/j.neucom.2017.07.044.
- [26] HUANG R, KANG L Y. Local Positive and Negative Label Correlation Analysis with Label Awareness for Multi-label Classification[J]. *Int J Mach Learn Cybern*, 2021, **12**(9): 2659–2672. DOI: 10.1007/s13042-021-01352-2.
- [27] LIN Y J, HU Q H, LIU J H, *et al.* MULFE: Multi-label Learning via Label-specific Feature Space Ensemble[J]. *ACM Trans Knowl Discov From Data*, **16**(1): 1–24. DOI: 10.1145/3451392.
- [28] ZHAO D W, GAO Q W, LU Y X, *et al.* Learning Multi-label Label-specific Features via Global and Local Label Correlations[J]. *Soft Comput*, 2022, **26**(5): 2225–2239. DOI: 10.1007/s00500-021-06645-w.
- [29] ZHANG M L, FANG J P, WANG Y B. BiLabel-specific Features for Multi-label Classification[J]. *ACM Trans Knowl Discov From Data*, **16**(1): 1–23. DOI: 10.1145/3458283.
- [30] SHEN X P, BOUTELL M, LUO J B, *et al.* Multilabel Machine Learning and Its Application to Semantic Scene Classification[C]//Proc SPIE 5307, Storage and Retrieval Methods and Applications for Multimedia 2004, 2003, **5307**: 188–199. DOI: 10.1117/12.523428.
- [31] FÜRNRKRAZ J, HÜLLERMEIER E, MENCÍA E L, *et al.* Multilabel Classification via Calibrated Label Ranking [J]. *Mach Learn*, 2008, **73**(2): 133–153. DOI: 10.1007/s10994-008-5064-8.