

DOI:10.13232/j.cnki.jnju.2023.01.009

## 基于弱监督对比学习的弱多标记特征选择

王 津<sup>1</sup>, 谭安辉<sup>1,2\*</sup>, 顾沈明<sup>1,2</sup>

(1. 浙江海洋大学信息工程学院, 舟山, 316022;

2. 浙江省海洋大数据挖掘与应用重点实验室, 浙江海洋大学信息工程学院, 舟山, 316022)

**摘 要:** 多标记特征选择已被广泛应用于医疗诊断、模式识别等领域, 然而现实中的数据往往存在维数灾难以及标记大量缺失等问题, 现有的弱多标记特征选择算法又普遍易受缺失标记和噪声的干扰, 使算法模型难以准确地选择重要特征. 针对上述问题, 提出一种用于弱多标记数据集特征选择的弱监督对比学习方法, 旨在缺失和含噪声标记数据集中选择优质特征, 同时挖掘少量人工标注数据中潜在的类间对比模式. 该方法包括三个步骤: 首先, 设计一种弱监督预训练策略, 通过利用实例相似性以及稀疏学习方法获取每个类标记类属属性, 用于恢复缺失标记; 其次, 引入对比学习策略来捕获少量有标记数据的对比模式来削弱噪声数据的影响; 最后, 选取 10 组多标记数据集以及四个评价指标进行实验. 实验结果证明, 与多个先进多标记特征选择算法相比, 提出的方法分类性能更优.

**关键词:** 多标记学习, 特征选择, 缺失标记, 弱监督学习, 对比学习

**中图分类号:** TP181

**文献标志码:** A

## Weak multi-label feature selection based on weakly-supervised contrastive learning

Wang Jin<sup>1</sup>, Tan Anhui<sup>1,2\*</sup>, Gu Shenming<sup>1,2</sup>

(1. School of Information Engineering, Zhejiang Ocean University, Zhoushan, 316022, China;

2. Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, School of Information Engineering, Zhejiang Ocean University, Zhoushan, 316022, China)

**Abstract:** Multi-label feature selection has been widely used in many fields, such as medical diagnosis and pattern recognition. However, most weak multi-label data often suffer from many problems, such as curse of dimensionality and noisy/missing labels in reality. Existing weak label feature selection algorithms are generally vulnerable to missing and noisy labels, which makes it difficult for algorithms to accurately select important features. To solve these problems, we present a weakly-supervised contrastive learning method for feature selection from large weak multi-label data with missing labels that aims at selecting a large number of high-quality features and also exploits inter-class contrastive patterns hidden in a small amount of human-annotated labels of data. The method consists of three steps. Firstly, we design a weakly-supervised pre-training strategy, which select the label-specific feature for each class label via sparse learning method, and constructs the correlation between instance and label-specific feature, which can be used to recover the missing label space. Then, we introduce the contrastive strategy to capture the contrast patterns in a small labeled data, which reduces the influence of noisy labels. In the end, experiments are performed on ten multi-label data sets in terms of four evaluation metrics. Experimental results show that our approach outperforms the other state-of-the-art multi-label feature selection algorithms.

**Key words:** multi-label learning, feature selection, missing labels, weakly-supervised learning, contrastive learning

基金项目: 国家自然科学基金(62076221, 61976194)

收稿日期: 2022-09-30

\* 通讯联系人, E-mail: tananhui86@163.com

真实情景中,单个实例通常会同时呈现多个不相同的类别标记<sup>[1]</sup>,仅靠单个标记无法完整描述其语义,因此,多标记学习框架应运而生.这种学习框架更适用于现实任务的多义性,且已广泛应用于众多领域,如图像自动标注、文本分类和情感分析<sup>[2]</sup>等.和单标记学习不同,多标记学习中经常会遇见具有成百上千特征的高维数据<sup>[3]</sup>,若直接在原始数据上进行学习,容易出现过拟合以及泛化能力差等问题.因此,特征维数灾难是多标记学习的重要挑战之一<sup>[4]</sup>.

为了缓解这一问题,大量关于多标记降维的方法<sup>[5-6]</sup>被提出,这些方法大致分两类,一类是多标记特征抽取方法<sup>[7]</sup>,另一类是多标记特征选择方法<sup>[8-9]</sup>.特征抽取指对原始特征空间进行特定组合,将其映射到一个低维空间,通常可以有效降维并取得不错的分类效果,但不能保留原始特征的物理意义,对应的分类过程不具有解释性.和特征抽取相比,多标记特征选择通过某种度量标准对特征进行排序或者选择特征子集,可以有效去除特征空间中冗余、无关特征,而且保留了原始特征的物理意义,具有更好的解释性,在很大程度上降低了维度灾难对多标记学习的影响.

根据是否依赖特定的学习器,现有的多标记特征选择算法大致分三类:过滤式、包裹式和嵌入式.过滤式方法需要基于数据集完成特征的正确筛选,再选择合适的训练学习器. Lin et al<sup>[10]</sup>通过最大化特征的依赖度的同时最小化特征间的冗余度来进行特征选择.包裹式方法利用最后能用到的学习器的性能作为特征子集的评价准则. Gharroudi et al<sup>[11]</sup>提出基于随机森林和标记依赖性的多标记特征选择算法.嵌入式方法把特征选择过程与学习器训练过程看作一个整体,共同优化,在学习器训练的同时挑选最有意义的特征. Zhang et al<sup>[12]</sup>提出基于流行正则项约束的多标记特征选择算法.尽管上述算法在多标记学习中有不错的表现,但模型性能大多依赖具有大规模高质量人工标注的数据.然而,人工标注多标记数据集不仅耗时费力,还会因为类标记之间的歧义性容易出现漏标的情况,导致标记一致性较低,因此,直接使用弱多标记数据训练多标记学习模型会极大影响模型性能的准确性<sup>[13]</sup>.

针对上述问题,本文提出一种用于弱标记数据集特征选择的弱监督对比学习(Weakly-Supervised Contrastive Feature Selection, WCFS)方法.该方法包括两个核心步骤:(1)使用大规模弱标记数据对模型进行预训练,预训练利用实例相似性以及稀疏学习方法获取的每个类标记类属性恢复缺失标记;(2)对预训练模型再进行调整.受自监督对比学习的启发<sup>[14-15]</sup>,设计了一种弱监督对比学习策略,其目的在于通过构造正负实例,充分利用相同类实例之间的相似性和不同类实例之间的差异性来改进 $k$ 近邻机制,最终达到选择优质特征的目的.本文的主要贡献:

(1)提出一种面向弱多标记特征选择的弱监督对比学习(WCFS)框架,考虑到标记空间的不完备性,设计了一种弱监督预训练策略,通过利用实例相似性以及稀疏学习方法获取的每个类标记类属性来恢复不完备的标记空间.

(2)为了改进 $k$ 近邻机制,引入对比学习,挖掘少量人工标记数据中潜在的类间对比模式,能够充分有效地利用相同类实例之间的相似性以及不同类实例之间的差异性,削弱噪声数据带来的负面影响,减少模型对多标记数据人工标注的依赖度,提升模型的特征选择性能.

(3)在 10 个标准数据集中与目前先进的多标记算法进行对比,实验结果表明本文算法具有较好的特征选择效果.

## 1 相关工作

多标记学习在众多领域都具有广阔的应用背景,但同时也存在巨大的挑战,数据的高维性和标记缺失等问题是目前多标记学习的研究热点.针对类属属性学习和标记缺失,学者们已提出了大量的学习算法.

利用类属属性进行多标记学习可以有效处理数据高维性. Zhang and Wu<sup>[16]</sup>与吴磊和张敏灵<sup>[17]</sup>提出基于类属属性的多标记学习算法(Multi-Label Learning with Label-Specific Features, LIFT),先将 $k$ -means 聚类技术应用于每个标记的正负训练样本,然后利用 SVM (Support Vector Machine)对多标记分类进行建模,在解决多标记分类问题上取得了显著成效.然而这种算法忽略

了标记相关性的影响,因此 Huang et al<sup>[18]</sup>通过特征选择的方式直接学习类属属性,提出类属属性提取学习 (Learning Label-Specific Features and Class-Dependent Labels, LLSF-DL),通过设计优化框架来学习每个标记的低维数据表示,并利用成对相关考虑共享特征. Zhang et al<sup>[19]</sup>提出类属属性和多标记联合学习 (Multi-Label Learning with Label-Specific Features by Resolving Label Correlations, MLFC),该算法的关键是设计一个优化模型来分配特征权重,同时构建附加特征来考虑标记之间的相关性. 标记的类属属性是一组与该标记相关性最强的特征子集,因此,标记的类属属性能为多标记学习提供更有价值的信息.

考虑到实际任务中获得完整的标记空间是非常困难的,因此,在缺失大量标记的环境下,根据多标记信息熵、线性回归和标记一致性等理论的多标记特征选择算法被提出. 例如, Zhu et al<sup>[20]</sup>运用鲁棒的线性回归模型选择最具区分性的特征,并在特征选择的同时恢复缺失标记. Wang et al<sup>[21]</sup>定义了多标记信息熵和多标记互信息两个新概念,并运用特征交互进行特征选择.

针对弱监督分类学习, Vapnik and Sterin<sup>[22]</sup>提出转导支持向量机来估计类标记的线性预测函数. Chapelle et al<sup>[23]</sup>提出直推学习的概念和生成式模型,假设生成数据的概率密度函数为多项式分布,并基于该假设利用有类标记的实例和无类标记的实例进行学习,并估计该模型中的参数.

## 2 基于弱监督对比学习的多标记特征选择

**2.1 算法框架** 在多标记特征选择任务中,给定一个数据集:

$$D = \{(x_i, y_i) | 1 \leq i \leq n\} \quad (1)$$

其中,

$$X = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^{n \times d} \quad (2)$$

为具有  $n$  个实例的特征空间,每个实例

$$x_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T \quad (3)$$

具有  $d$  维特征;每个特征表示为  $f_j, 1 \leq j \leq d$ .

$$Y = [y_1, y_2, \dots, y_n]^T \in \{-1, 0, 1\}^{n \times q} \quad (4)$$

为  $n$  个实例对应的训练标记集,每个实例

$$y_i = [y_{i1}, y_{i2}, \dots, y_{iq}]^T \quad (5)$$

具有  $q$  维标记.  $y_{ij} = 1$  表示第  $i$  个实例具有第  $j$  个标记,反之,  $y_{ij} = -1$ . 如果无法确定第  $i$  个实例是否具有第  $j$  个标记,即该标记缺失,则表示为  $y_{ij} = 0$ . 对于多标记特征选择,给定训练集和标记集,拟合一个系数矩阵  $W \in \mathbb{R}^{d \times q}$ ,其中,  $W_i$  表示  $W$  的第  $i$  行,  $W_j$  表示  $W$  的第  $j$  列,通过优化矩阵  $W$  来筛选标记的重要优质特征.

本节提出一种基于弱监督对比学习的弱多标记特征选择算法 (WCFS) 模型 (图 1),其中,训练数据集中空心圆代表标记有缺失的实例,实心圆代表标记完整的实例. 具体的算法框架如下:

$$\min_{W, F, Q_{ij}, Q_{ij}} L_{\text{RL}}(W, R, L, F) + L_{\text{WSCL}}(X, Q_{ij}) + \Omega(W) \quad (6)$$

其中,第一项  $L_{\text{RL}}(W, R, L, F)$  为损失函数,考虑了特征相似性、标记相似性和标记的一致性;第二项  $L_{\text{WSCL}}(X, Q_{ij})$  为弱监督对比损失函数;第三项  $\Omega(W)$  为控制模型复杂度的正则化项. 每项的具体内容在下面进行介绍.

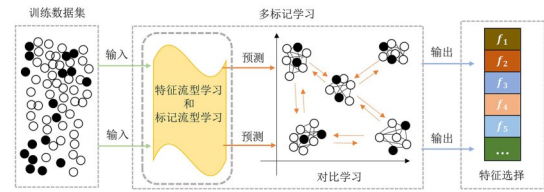


图1 WCFS算法示意图

Fig. 1 Illustration of the proposed method WCFS

**2.2 恢复缺失标记** 首先,在特征选择过程中,为了保证选择的特征数量少且具有代表性,在系数矩阵  $W$  上添加  $l_{2,1}$  范数正则项,其目标函数为:

$$\Omega(W) = \frac{\gamma}{2} \|W\|_{2,1} \quad (7)$$

其中,  $\gamma > 0$  为参数,用于调节经验误差和稀疏正则项之间的平衡.  $l_{2,1}$  范数的计算式为:

$$\|W\|_{2,1} = \sum_{i=1}^d \left( \sum_{j=1}^q |w_{ij}|^2 \right)^{1/2} \quad (8)$$

接着,针对标记空间不完备问题,考虑利用类属属性和标记相似性构建与标记的映射关系,用以恢复缺失标记. 具体地,在多标记学习中一个实例一般具有多个类标记,但每个标记的重要性都不一致<sup>[24]</sup>. 两个强相关的标记比两个弱相关或不相关的标记共享更多的特征,即如果标记  $y_i$  和

标记  $y_j$  是强相关的,那么与之对应的  $W_{\cdot i}$  和  $W_{\cdot j}$  有很大的相似性,否则相似度很小.同时,考虑到实例中两个特征间如果具有强相关性,该实例具备其中一个特征的同时,在很大概率上也会具备另一个特征,否则在很大概率上不具备另一个特征.得到的损失函数为:

$$L_{RL}(W, R, L, F) = \frac{1}{2} \|XW - F\|_F^2 + \frac{\alpha}{2} \text{tr}(RW^T W) + \frac{\beta}{2} \text{tr}(WW^T L) + \eta \|Y - F\|_1 \quad (9)$$

s.t.  $F = QYC$

其中,  $\alpha, \beta, \eta > 0$ , 都是参数,  $\alpha$  控制标记相关性对  $W$  的影响,  $\beta$  控制特征相似性对  $W$  的影响,  $\eta$  控制标记一致性.  $R_{ij} = 1 - C_{ij}$ ,  $C_{ij}$  为标记  $y_i$  和  $y_j$  之间的余弦相似度,  $C \in R^{q \times q}$  为标记相似矩阵, 取值范围为  $[-1, 1]$ .  $L = A - S$  表示拉普拉斯矩阵, 其中,  $A$  是对角矩阵,  $A_{ii} = \sum_{j=1}^d S_{ij}$ ,  $S \in R^{d \times d}$  表示特征相似度矩阵, 具体的计算方法如下:

$$S_{ij} = \begin{cases} \exp\left(-\frac{\|f_i - f_j\|_2^2}{2\sigma^2}\right), & f_i \in N_{f_j} \text{ 或 } f_j \in N_{f_i} \\ 0, & \text{其他} \end{cases} \quad (10)$$

其中,  $S_{ij}$  表示特征  $f_i$  和  $f_j$  之间的相似度, 取值范围  $[0, 1]$ ,  $N_{f_i}$  表示距离特征  $f_i$  最接近的  $k$  个近邻;  $F \in R^{n \times q}$  表示预测的标记矩阵; 最后一项为  $l_1$  范数, 计算式为  $\|X\|_1 = \sum_{i=1}^n |x_i|$ . 约束条件中  $Q \in R^{n \times n}$  是实例相似矩阵, 为了使实例之间相似性更符合真实分布, 先对每个实例  $x_i$  根据欧式距离检索出  $k$  个最近邻  $N_{x_i}$ , 再用余弦值计算两个实例的相似性. 具体如下:

$$Q_{ij} = \begin{cases} \frac{x_i x_j^T}{\|x_i\| \|x_j\|}, & x_i \in N_{x_j} \text{ 或 } x_j \in N_{x_i} \\ 0, & \text{其他} \end{cases} \quad (11)$$

最后, 为了削弱噪声数据的负面影响, 充分挖掘实例自身信息, 引入弱监督对比学习, 具体方法在下一节介绍.

**2.3 弱监督对比学习** 考虑到  $k$  近邻中可能会存在噪声数据, 即两个实例之间的相似性很高, 但两个实例对应的标记相似性不高. 针对这一情况, 受自监督对比学习模型<sup>[14-15]</sup>启发, 本文设计了面向弱多标记数据的弱监督对比学习损失函数,

充分挖掘数据中的同类和非同类实例之间的对比模式, 改进实例相似性中  $k$  近邻机制. 具体如下:

$$L_{WSCL}(X, Q_{ij}) = \sum_i \sum_{x_j \in N_{x_i}} I_{\lambda_{ij} \leq 0} L_{WSCL}^{ij} \quad (12)$$

$$L_{WSCL}^{ij} = -\lambda_{ij} \lg \frac{\exp(Q_{ij}/\tau)}{\sum_{x_k \in N_{x_i}, k \neq j} \exp(Q_{ik}/\tau) + \exp(Q_{ij}/\tau)} \quad (13)$$

$L_{WSCL}(X, Q_{ij})$  表示弱监督对比学习目标函数 (Weakly-Supervised Contrastive Learning, WSCL). 其中,  $I_{\lambda_{ij} \leq 0} \in [-1, 1]$  是指示函数, 当且仅当  $\lambda_{ij} \leq 0$  时返回  $-1$ , 否则返回  $1$ ;  $\tau > 0$  表示温度系数, 是一个超参数, 调节对噪声数据的关注度;  $z_{ij} = y_i^T y_j$  表示两个实例共有的标记数; 动态系数  $\lambda_{ij}$  是  $z_{ij}$  归一化后的结果, 计算式如下:

$$\lambda_{ij} = \frac{z_{ij}}{\sum_{x_k \in N_{x_i}} |z_{ik}|} \quad (14)$$

$\lambda_{ik} \leq 0$  表示实例  $x_k$  为实例  $x_i$  的负实例,  $\lambda_{ij} > 0$  表示实例  $x_k$  为实例  $x_i$  的正实例.

$k$  近邻中, 如果两个实例拥有相同的标记越多, 它们的距离越近, 反之则距离越远. 因此本文考虑在弱监督对比学习阶段对实例相似矩阵  $Q$  进行迭代优化. 具体地,  $k$  近邻中的实例经过  $\lambda_{ik}$  的“筛选”被归纳到对应的计算结构中, 充分利用数据已有的自身信息, 通过学习同类实例间的相似性与非同类实例间的差异性进一步增加同类实例间的相似性, 减小非同类之间相似性, 削弱噪声数据对训练过程产生的负面影响, 让实例相似矩阵  $Q$  更加准确地描述真实的实例分布情况, 进而提升预测标记矩阵  $F$  与真实标记的一致性, 最终使模型能够更好地对重要特征进行选取.

**2.4 模型优化求解** 综合式(7)、式(9)、式(12)和式(13)可得本文算法 WCFS 的目标函数, 如下所示:

$$\begin{aligned} \min_{W, F, Q_{ij}, Q_{ij}} & \frac{1}{2} \|XW - F\|_F^2 + \frac{\alpha}{2} \text{tr}(RW^T W) + \\ & \frac{\beta}{2} \text{tr}(WW^T L) + \frac{\gamma}{2} \|W\|_{2,1} + \eta \|Y - F\|_1 - \\ & \sum_i \sum_{x_j \in N_{x_i}} \lambda_{ij} I_{\lambda_{ij} \leq 0} \lg \frac{\exp(Q_{ij}/\tau)}{\sum_{x_k \in N_{x_i}, k \neq j} \exp(Q_{ik}/\tau) + \exp(Q_{ij}/\tau)} \\ \text{s.t. } & F = QYC \end{aligned} \quad (15)$$



目标函数含有四个变量  $W, F, Q_{ij}, Q_{ik}$ , 为了得到最优解, 对变量采用交替迭代最小化的方法进行优化.

Step 1. 固定其他变量, 更新  $W$ , 对变量  $W$  求梯度并使之等于0, 得到:

$$(X^T X + \beta L + \gamma D)W + \alpha WR = X^T F \quad (16)$$

其中,  $D$  是对角矩阵,

$$D_{ii} = \frac{1}{2\|W_i\|_2} \quad (1 \leq i \leq d) \quad (17)$$

此时, 式(16)是形如西尔维斯特  $AX + XB = C$  的问题<sup>[25]</sup>:

$$A = X^T X + \beta L + \gamma D$$

$$B = \alpha R, C = X^T F$$

Step 2. 固定其他变量, 更新变量  $E$ .

式(15)中含有  $l_1$  范数, 虽然它是凸函数, 但是非光滑的. 因此, 为了使式(15)可分离, 引入一个变量  $E$ , 令  $E = Y - F$ , 因此, 式(15)可以替换为:

$$\min_E \eta \|E\|_1 \quad (18)$$

$$s.t. E = Y - F, F = QYC$$

式(18)的增广拉格朗日函数为:

$$\min_E \eta \|E\|_1 + \frac{\mu_1}{2} \|E - Y + F\|_F^2 + \langle S_1, E - Y + F \rangle + \langle S_2, F - QYC \rangle + \frac{\mu_2}{2} \|F - QYC\|_F^2 \quad (19)$$

根据LADMAP方法<sup>[26]</sup>, 式(19)可重新写成:

$$\min_E \eta \|E\|_1 + \frac{\mu_1}{2} \left\| E - Y + F + \frac{S_1}{\mu_1} \right\|_F^2 + \frac{\mu_2}{2} \left\| F - QYC + \frac{S_2}{\mu_2} \right\|_F^2 \quad (20)$$

其中,  $S_1 \in R^{n \times q}, S_2 \in R^{n \times q}$  是拉格朗日乘子矩阵,  $\mu_1, \mu_2$  是惩罚系数. 于是有:

$$\min_E \eta \|E\|_1 + \frac{\mu_1}{2} \left\| E - Y + F + \frac{S_1}{\mu_1} \right\|_F^2 \quad (21)$$

根据参考文献[26], 上式有封闭解, 更新  $E$ :

$$E^{k+1} = S_{\eta/\mu_1} \left[ Y - (F)^k - (S_1/\mu_1)^k \right] \quad (22)$$

这里  $S_\epsilon[\cdot]$  表示软阈值算子:

$$S_\epsilon[e_{ij}] = \begin{cases} e_{ij} + \epsilon, & e_{ij} < -\epsilon \\ e_{ij} - \epsilon, & e_{ij} > \epsilon \\ 0, & -\epsilon \leq e_{ij} \leq \epsilon \end{cases} \quad (23)$$

其中,

$$e_{ij} = \left[ Y - (F)^k - (S_1/\mu_1)^k \right]_{ij}$$

Step 3. 固定其他变量, 结合式(15)、式(20), 更新  $F$ :

$$\min_F \frac{1}{2} \|XW - F\|_F^2 + \frac{\mu_1}{2} \left\| E - Y + F + \frac{S_1}{\mu_1} \right\|_F^2 + \frac{\mu_2}{2} \left\| F - QYC + \frac{S_2}{\mu_2} \right\|_F^2 \quad (24)$$

变量  $F$  可以用投影梯度下降(Projected Gradient Descent, PGD)方法求解, 由于标记空间的取值在  $[-1, 1]$ , 所以,

$$F_{t+1} = \min(1, \max(-1, F - \eta_t \nabla F)) \quad (25)$$

其中,  $\eta_t$  是  $t$  次迭代的步长.

$$\nabla F = \frac{XW - \mu_1 \left( E - Y + \frac{S_1}{\mu_1} \right) + \mu_2 \left( QYC - \frac{S_2}{\mu_2} \right)}{1 + \mu_1 + \mu_2} \quad (26)$$

然后, 通过以下步骤更新拉格朗日乘子矩阵  $S_1, S_2$  和惩罚系数  $\mu_1, \mu_2$ :

$$\begin{aligned} \mu_1^{k+1} &= \min(\mu_{\max}, \rho \mu_1^k) \\ \mu_2^{k+1} &= \min(\mu_{\max}, \rho \mu_2^k) \\ S_1^{k+1} &= S_1^k + \mu_1^{k+1} (E - Y + F) \\ S_2^{k+1} &= S_2^k + \mu_2^{k+1} (F - QYC) \end{aligned} \quad (27)$$

其中,  $\rho$  是一个正标量.

Step 4. 固定其他变量, 更新变量  $Q_{ij}$ , 对  $Q_{ij}$  求梯度得:

$$\nabla_{Q_{ij}} = -\lambda_{ij} \frac{1}{\tau} \sum_{x_k \in N_{ij}} P_{ik} \quad (28)$$

Step 5. 固定其他变量, 更新变量  $Q_{ik}$ , 对  $Q_{ik}$  求梯度得:

$$\nabla_{Q_{ik}} = \lambda_{ik} \frac{1}{\tau} P_{ik} \quad (29)$$

其中,

$$P_{ij} = I_{\lambda_{ij} \leq 0} \frac{\exp(Q_{ij}/\tau)}{\sum_{x_k \in N_{ij}} \exp(Q_{ik}/\tau) + \exp(Q_{ij}/\tau)} \quad (30)$$

具体的算法如下所示.

**算法** 基于弱监督对比学习的弱多标记特征选择

**输入**: 训练实例矩阵  $X \in R^{n \times d}$ , 训练标记矩阵  $Y \in R^{n \times q}$ , 参数  $\alpha, \beta, \gamma, \eta$ ;  
**输出**: 系数矩阵  $W^*$ .

1. 初始化:  $W_0, W_1 = \text{rand}(d, q), Q$ , 拉格朗日乘子矩

阵  $S_1 \in R^{n \times q}$ ,  $S_2 \in R^{n \times q}$ ;

2. 计算标记相似性  $C$ ;

3. 计算拉普拉斯矩阵  $L$ ;

开始循环:

4. 通过式(16)更新  $W$ ;

5. 通过式(22)更新  $E$ ;

6. 通过式(25)更新  $F$ ;

7. 通过式(27)更新  $S_1, S_2$ ;

8. 通过式(28)更新  $k$  近邻正实例对  $Q_{ij}$ ;

9. 通过式(29)更新  $k$  近邻负实例对  $Q_{ik}$ ;

10.  $t = t + 1$ ;

直到收敛;

11. 通过  $Y = X * W$  恢复缺失的标记矩阵;

12. 计算  $\|W_i\|, i \in [1, d]$  得到每个特征的重要度;

13. 对每个特征的重要度进行降序排序得到最终特征排序,且选取前 35% 特征为优质特征.

**2.5 时间复杂度分析** WCFS 算法中,初始化阶段构造标记相似矩阵  $C$ 、实例相似矩阵  $Q$  和拉普拉斯矩阵  $L$ , 计算复杂度分别为  $O(q^2)$ ,  $O(n^2)$  和  $O(q^3)$ . 接着,算法进入迭代循环阶段,每次迭代过程中,更新  $W$  的复杂度为  $O(q^2 + ndq + nd^2 + d^3)$ ,更新  $E$  的复杂度为  $O(nq)$ ,更新  $F$  的复杂度为  $O(n^2q + ndq + nq^2)$ ,更新  $Q_{ij}, Q_{ik}$  的复杂度为  $O(kn^2)$ . 因此,整个算法的复杂度为  $O(q^3 + n^2 + t(n^2q + ndq + nd^2 + nq^2 + d^3 + kn^2))$ , 其中,  $t$  为迭代次数,  $k$  为近邻实例个数.

### 3 实验及结果分析

**3.1 评价指标和对比算法** 选用常见的 10 种多标记数据集(表 1)来验证本文算法的有效性,并通过四种评价指标来综合衡量算法的性能.

(1) Macro - averaging AUC (Area under Curve):统计分类模型在各个类别标记的样本上的输出结果大于不含该对应类别标记样本的输出结果的情况.

$$AUC_{\text{macro-avg}}(f) = \frac{1}{q} \sum_{k=1}^q \frac{\left| \left\{ x_i, x_k \mid f(x_i, y_k) \geq f(x_k, y_k) \right\} \right|}{|Q_k| |\bar{Q}_k|} \quad (31)$$

其中,

$$x_i \in Q_k, Q_k = \{x_i \mid y_k \in Y_i, 1 \leq i \leq m\}$$

表 1 实验数据集

Table 1 Experimental datasets

序号	名称	类别	训练实例数	测试实例数	特征数	标记数
1	Arts	文本	2000	3000	462	26
2	Birds	音频	322	322	260	20
3	Computer	文本	2000	3000	681	33
4	Education	文本	2000	3000	550	33
5	Entertainment	文本	2000	3000	640	21
6	Image	文本	1000	1000	294	5
7	Recreation	文本	2000	3000	606	22
8	Science	文本	2000	3000	743	40
9	Slashdot	文本	2546	1236	1079	22
10	Yeast	生物	1499	918	103	14

$$x_k \in \bar{Q}_k, \bar{Q}_k = \{x_i \mid y_k \in \bar{Y}_i, 1 \leq i \leq m\}$$

(2) Ranking Loss (RL):考察未正确排序的标记对平均次数.

$$RL = \frac{1}{s} \sum_{i=1}^s \frac{\left| \left\{ (l_k, l_j) \mid f_k(x_i) \leq f_j(x_i), (l_k, l_j) \in Y_i \times \bar{Y}_i \right\} \right|}{|Y_i| |\bar{Y}_i|} \quad (32)$$

其中,  $\bar{Y}_i$  表示在标记集合  $L$  中  $Y_i$  的补集.

(3) Average Precision (AP):评估排名高于特定标记  $l_k \in Y_i$  的相关标记的平均分数.

$$AP = \frac{1}{s} \sum_{i=1}^s \frac{1}{|Y_i|} \sum_{l_k \in Y_i} \frac{|\mathcal{R}(x_i, l_k)|}{\text{rank}(x_i, l_k)} \quad (33)$$

其中,

$$\mathcal{R}(x_i, l_k) = \{l_j \mid \text{rank}(x_i, l_j) \leq \text{rank}(x_i, l_k), l_j \in Y_i\}$$

(4) One-Error (OE):考察在预测标记的排序序列中排名最靠前的标记不在样本真实标记集中的情况.

$$OE = \frac{1}{N} \sum_{i=1}^N \left[ \left[ \argmax_{y_i \in L} f(x_i, y_i) \right] \notin y_i' \right] \quad (34)$$

对于这四个性能评价指标, AUC 和 AP 取值越大越好, 而 RL 和 OE 取值越小越好. 为了有效验证本文算法的性能, 选择六种多标记特征选择算法进行对比实验.

MDDM (Multi-Label Dimensionality Reduction via Dependence Maximization)<sup>[27]</sup>: 通过最大化特征和类标记之间的依赖关系, 将原始特征空间投影到更低维的特征空间, 其中包括两种投影策略, 即 MDDM<sub>sps</sub> 和 MDDM<sub>proj</sub>.

MIFS (Multi-Label Informed Feature Selection)<sup>[28]</sup>:利用标记相关性来选择跨多个标记的鉴别特征.

MLNB (Feature Selection for Multi-Label Naive Bayes Classification)<sup>[29]</sup>:利用贝叶斯分类器的多标记特征选择算法.

PMU (Feature Selection for Multi-Label Classification Using Multi-Variate Mutual Information)<sup>[30]</sup>:运用所选特征和标记集之间的互信息来处理特征选择问题.

LSGL (Learning Multi-Label Label-Specific Features via Global and Local Label Correlations)<sup>[31]</sup>:利用全局和局部标记相关性以及标记对应特征进行多标记分类.

本文算法  $\alpha, \beta, \gamma, \eta$  取值为  $\{2^{-6}, 2^{-5}, \dots, 2^0\}$ ,  $\tau$  取值为  $[0.01, 0.1]$ ; MDDM<sub>spc</sub> 和 MDDM<sub>proj</sub>, 参数  $\mu$  设置为 0.5; PMU, 采用等宽策略离散化连续型特征, 对于标记的缺失率, 设置为 30% 和 50%. 多标记分类器采用经典的 ML-KNN 分类器<sup>[32]</sup>, 参数  $Num$  和  $Smooth$  分别设置为 10 和 1.

**3.2 实验结果与分析** 表 2~9 中, “ $\uparrow$ ”表示该评价指标的取值越大越好, “ $\downarrow$ ”表示该评价指标的取值越小越好, 每张表的最后一列为实验结果的平均排名, 黑体字表示最优的实验结果. 为了达到特征选择的目的, 每个数据集取前 35% 个特征, 即 Arts, Birds, Computer, Education, Entertainment, Image, Recreation, Science, Slashdot, Yeast 数据集分别取经挑选后的前 162, 91, 238, 193, 224, 103, 212, 260, 378 和 36 个优质特征进行分类. 根据表 2~9 的结果可得出:

(1) 由表 2~5, 标记缺失率为 30% 时, WCFS

在 10 个数据集上的四个评价指标表现突出.  $AUC$  指标, WCFS 在 Entertainment, Image 和 Recreation 数据集上略低于 LSGL;  $AP$  和  $OE$  指标, WCFS 在 Arts, Computer, Recreation 数据集上略低于其他算法. 和其他算法相比, WCFS 的评价指标均取得最优.  $RL$  指标, 平均排名 WCFS 最优, LSGL 次之, 原因是 LSGL 利用了全局和局部以及标记对应特征的算法, 可有效预测缺失标记, 缺点是忽略了实例近邻中存在噪声的情况.

(2) 随着缺失率的不断提升, 所有的分类算法性能都呈下降趋势, 但 WCFS 和其他算法相比, 优势较明显, 尤其是在最大的数据集 Slashdot 的缺失率为 50% 时的优势更显著. 直观的解释是, 本文算法运用了弱监督对比学习, 让同类的实例更接近, 不同类的实例更远离, 实例越多效果越显著. 故本文算法在含有缺失率的大数据集多标记特征选择中具有较强的鲁棒性.

**3.3 参数敏感性分析** WCFS 有四个参数  $\alpha, \beta, \gamma, \eta$ .  $\alpha$  控制标记相关性对  $W$  的影响,  $\beta$  控制特征相似性对  $W$  的影响,  $\gamma$  调节经验误差和稀疏正则项之间的平衡,  $\eta$  控制标记一致性. 为了分析 WCFS 参数的敏感性, 在 Birds 数据集中, 标记缺失率为 30% 和 50% 时进行实验, 固定其中三个参数, 让最后一个参数在  $\{2^{-6}, 2^{-5}, \dots, 2^0\}$  变化, 得到的结果如图 2 和图 3 所示. 由图可见, WCFS 的性能指标对  $\gamma, \eta$  不敏感, 对  $\alpha, \beta$  敏感.  $\alpha, \beta$  增大时, 各项指标性能都增加, 这是因为  $\alpha, \beta$  控制标记相似性和类属属性. 但参数取值过大、权重过大时一定程度上会影响实验结果, 直观解释是, 由于真实数据集标记缺失, 通过标记相似性和类属属性可以帮助恢复缺失标记, 提高算法性能.

表 2 30% 标记缺失率下, 七种特征选择算法在 10 个数据集上的对比结果 ( $AUC \uparrow$ )

Table 2  $AUC (\uparrow)$  of seven feature selection algorithms on ten datasets with 30% missing labels

算法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.5802	0.6538	0.5858	0.6006	0.6279	0.7907	0.5796	0.5792	0.6314	0.5894	5.7000
MDDM <sub>proj</sub>	0.5894	0.6580	0.5900	0.5948	0.6140	0.7794	0.6096	0.5928	0.6326	0.5933	4.8000
MIFS	0.6066	0.7177	0.5971	0.6440	0.6442	0.7672	0.6136	0.6252	0.6240	0.6038	3.5000
MLNB	0.5878	0.6744	0.5845	0.6105	0.6241	0.7824	0.5784	0.5816	0.6258	0.5769	5.7000
PMU	0.5868	0.6848	0.5868	0.6251	0.6413	0.7746	0.6002	0.5893	0.6346	0.6066	4.3000
LSGL	0.6106	0.6232	0.5966	0.6287	<b>0.6505</b>	<b>0.8000</b>	<b>0.6147</b>	0.6289	0.6436	0.6188	2.3000
WCFS	<b>0.6162</b>	<b>0.7427</b>	<b>0.6009</b>	<b>0.6449</b>	0.6333	0.7971	0.6046	<b>0.6354</b>	<b>0.7092</b>	<b>0.6239</b>	<b>1.7000</b>

表 3 30% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果( $RL \downarrow$ )Table 3  $RL (\downarrow)$  of seven feature selection algorithms on ten datasets with 30% missing labels

算 法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.1600	0.1512	0.0953	0.0958	0.1257	0.2105	0.1956	0.1471	0.1893	0.1971	5.7000
MDDM <sub>proj</sub>	0.1565	0.1567	0.0950	0.0979	0.1279	0.2248	0.1914	0.1423	0.1814	0.1955	5.2000
MIFS	0.1510	0.1479	0.0935	0.0915	0.1220	0.2402	<b>0.1844</b>	0.1349	0.1888	0.1869	3.2000
MLNB	0.1570	0.1508	0.0972	0.0958	0.1279	0.2201	0.2031	0.1465	0.1797	0.2037	5.7000
PMU	0.1564	0.1423	0.0953	0.0948	0.1243	0.2263	0.1907	0.1418	0.1838	0.1830	4.1000
LSGL	<b>0.1490</b>	0.1702	<b>0.0912</b>	0.0924	<b>0.1219</b>	0.2032	0.1846	0.1350	0.1764	0.1820	2.4000
WCFS	0.1515	<b>0.1339</b>	0.0926	<b>0.0889</b>	0.1238	<b>0.1994</b>	0.1913	<b>0.1325</b>	<b>0.1591</b>	<b>0.1794</b>	<b>1.7000</b>

表 4 30% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果( $AP \uparrow$ )Table 4  $AP (\uparrow)$  of seven feature selection algorithms on ten datasets with 30% missing labels

算 法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.4741	0.6333	0.6135	0.5295	0.5522	0.7548	0.4383	0.4350	0.4478	0.7274	5.8000
MDDM <sub>proj</sub>	0.4799	0.6412	0.6245	0.5151	0.5571	0.7413	0.4618	0.4504	0.4588	0.7237	4.7000
MIFS	0.5066	0.6724	0.6242	0.5560	0.5652	0.7211	<b>0.4899</b>	0.4801	0.4250	0.7384	3.3000
MLNB	0.4787	0.6421	0.6219	0.5323	0.5531	0.7483	0.4287	0.4256	0.4558	0.7175	5.5000
PMU	0.4847	0.6542	0.6165	0.5359	0.5601	0.7409	0.4605	0.4419	0.4326	0.7381	4.7000
LSGL	0.5162	0.6315	<b>0.6289</b>	0.5545	0.5624	0.7581	0.4862	0.4764	0.4687	0.7428	2.7000
WCFS	<b>0.5174</b>	<b>0.6967</b>	0.6268	<b>0.5617</b>	<b>0.5656</b>	<b>0.7595</b>	0.4820	<b>0.4831</b>	<b>0.5321</b>	<b>0.7441</b>	<b>1.3000</b>

表 5 30% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果( $OE \downarrow$ )Table 5  $OE (\downarrow)$  of seven feature selection algorithms on ten datasets with 30% missing labels

算 法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.6817	0.4830	0.4737	0.6190	0.6153	0.3730	0.7243	0.6993	0.7055	0.2495	5.4000
MDDM <sub>proj</sub>	0.6770	0.4520	0.4500	0.6357	0.6020	0.3930	0.6917	0.6797	0.6934	0.2691	5.1000
MIFS	0.6263	0.3777	0.4517	0.5750	0.5903	0.4340	<b>0.6540</b>	0.6487	0.7783	0.2593	3.8000
MLNB	0.6767	0.4768	0.4497	0.6133	0.6013	0.3770	0.7320	0.7167	0.7128	0.2603	5.1000
PMU	0.6673	0.4458	0.4667	0.6043	0.5883	0.3910	0.6963	0.6980	0.7330	0.2625	4.6000
LSGL	<b>0.6123</b>	0.4706	<b>0.4457</b>	0.5690	0.5947	0.3780	0.6570	0.6457	0.6739	0.2495	2.6000
WCFS	0.6150	<b>0.3591</b>	0.4477	<b>0.5683</b>	<b>0.5837</b>	<b>0.3570</b>	0.6670	<b>0.6437</b>	<b>0.5979</b>	<b>0.2407</b>	<b>1.4000</b>

表 6 50% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果( $AUC \uparrow$ )Table 6  $AUC (\uparrow)$  of seven feature selection algorithms on ten datasets with 50% missing labels

算 法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.5712	0.6670	0.5786	0.5816	0.6182	0.7557	0.5905	0.5721	0.6321	0.5612	5.3000
MDDM <sub>proj</sub>	0.5832	0.6644	0.5743	0.5835	0.6153	0.7734	0.6053	0.5858	0.6232	0.5615	4.9000
MIFS	0.6055	0.7025	0.5822	0.6177	0.6421	0.7461	0.6085	0.5991	0.6245	0.5973	3.4000
MLNB	0.5876	0.6342	0.5222	0.6003	0.6291	0.7679	0.5848	0.5754	0.6414	0.5503	4.9000
PMU	0.5766	0.6583	0.5750	0.6148	0.6234	0.7510	0.5757	0.5722	0.6208	0.6068	5.3000
LSGL	<b>0.6149</b>	0.6030	<b>0.5919</b>	0.6209	<b>0.6450</b>	0.7769	<b>0.6092</b>	0.6019	0.6269	0.5996	2.4000
WCFS	0.6135	<b>0.7238</b>	0.5860	<b>0.6265</b>	0.6302	<b>0.7820</b>	0.6052	<b>0.6000</b>	<b>0.7311</b>	<b>0.6156</b>	<b>1.8000</b>

**3.4 Holm 检验** 为了更好地对比算法性能,进一步引入 Holm 检验. 选择本文算法为控制函数. 为了一致性,让 10 个数据集在标记缺失率为 30% 的情况下继续进行实验. 各算法  $AUC$  指标的

Holm 检验结果如表 10 所示.

Holm 检验按降序排序,按照重要性测试假设. 表 10 中, Holm 检验的参数  $p_6 = 0.0016$ , 小于对应的 Holm 值, 拒绝假设;  $p_5 = 0.0016$ , 小于对



表7 50% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果

Table 7  $RL(\downarrow)$  of seven feature selection algorithms on ten datasets with 50% missing labels

算法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.1610	0.1621	0.0956	0.1024	0.1283	0.2348	0.1962	0.1478	0.1882	0.2046	5.4000
MDDM <sub>proj</sub>	0.1578	0.1627	0.0959	0.1033	0.1280	0.2188	0.1875	0.1453	0.1929	0.2052	5.2000
MIFS	0.1524	0.1478	0.0964	0.0958	0.1217	0.2350	<b>0.1845</b>	0.1404	0.1900	0.1897	3.2000
MLNB	0.1558	0.1636	0.1066	0.0986	0.1257	0.2273	0.2020	0.1468	0.1766	0.2039	5.1000
PMU	0.1572	0.1611	0.1008	0.0970	0.1262	0.2466	0.2007	0.1459	0.1928	0.1855	5.0000
LSGL	<b>0.1492</b>	0.1568	0.0945	0.0966	<b>0.1206</b>	0.2211	0.1864	<b>0.1355</b>	0.1917	0.1886	2.4000
WCFS	0.1510	<b>0.1463</b>	<b>0.0943</b>	<b>0.0948</b>	0.1240	<b>0.2073</b>	0.1936	0.1372	<b>0.1589</b>	<b>0.1830</b>	<b>1.7000</b>

表8 50% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果( $AP\uparrow$ )Table 8  $AP(\uparrow)$  of seven feature selection algorithms on ten datasets with 50% missing labels

算法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.4790	0.6242	0.6180	0.5019	0.5446	0.7320	0.4470	0.4203	0.4614	0.7164	5.5000
MDDM <sub>proj</sub>	0.4842	0.6312	0.6168	0.5060	0.5441	0.7446	0.4678	0.4413	0.4141	0.7154	5.2000
MIFS	0.4985	0.6681	0.6192	0.5363	0.5598	0.7267	<b>0.4815</b>	0.4455	0.4298	0.7322	3.4000
MLNB	0.4938	0.6307	0.5966	0.5376	0.5480	0.7339	0.4370	0.4168	0.4969	0.7168	4.7000
PMU	0.4769	0.6352	0.6155	0.5309	0.5476	0.7105	0.4260	0.4235	0.4410	0.7322	5.4000
LSGL	<b>0.5077</b>	0.6422	0.6250	0.5373	<b>0.5738</b>	0.7486	0.4766	0.4577	0.4370	0.7368	2.3000
WCFS	0.5054	<b>0.6699</b>	<b>0.6273</b>	<b>0.5492</b>	0.5603	<b>0.7545</b>	0.4640	<b>0.4589</b>	<b>0.5306</b>	<b>0.7403</b>	<b>1.5000</b>

表9 50% 标记缺失率下,七种特征选择算法在 10 个数据集上的对比结果( $OE\downarrow$ )Table 9  $OE(\downarrow)$  of seven feature selection algorithms on ten datasets with 50% missing labels

算法	Arts	Birds	Computer	Education	Entertainment	Image	Recreation	Science	Slashdot	Yeast	平均排名
MDDM <sub>spc</sub>	0.6723	0.5046	0.4603	0.6607	0.6183	0.4030	0.7033	0.7250	0.6788	0.2614	5.5000
MDDM <sub>proj</sub>	0.6660	0.4706	0.4593	0.6450	0.6240	0.3930	0.6810	0.6940	0.7654	0.2647	5.0000
MIFS	0.6360	0.4149	0.4573	0.6013	<b>0.5703</b>	0.4230	0.6697	0.6793	0.7613	<b>0.2407</b>	3.0000
MLNB	0.6433	0.4799	0.4797	0.5940	0.6130	0.4080	0.7270	0.7257	0.6343	0.2571	4.8000
PMU	0.6810	0.4799	0.4610	0.6037	0.6123	0.4450	0.7357	0.7180	0.7039	0.2538	5.4000
LSGL	0.6323	0.4551	0.4543	0.6050	0.5753	0.3840	<b>0.6677</b>	0.6760	0.7152	0.2462	2.7000
WCFS	<b>0.6307</b>	<b>0.4118</b>	<b>0.4460</b>	<b>0.5863</b>	0.5947	<b>0.3760</b>	0.6813	<b>0.6690</b>	<b>0.5987</b>	0.2440	<b>1.6000</b>

应的 Holm 值,拒绝假设;以此类推; $p_2 = 0.0578$ ,大于对应的 Holm 值,接受假设.可以看出,本文算法与 MDDM<sub>proj</sub>, MLNB, MDDM<sub>spc</sub>, PMU 有显著性差异,与 MIFS, LSGL 无显著性差异.

表 11 为各算法  $RL$  指标的 Holm 检验结果,采用类似表 10 的分析方法,可以看出,本文算法与 MDDM<sub>proj</sub>, MLNB, MDDM<sub>spc</sub>, PMU 有显著性差异,与 MIFS, LSGL 无显著性差异.

表 12 和表 13 分别给出各算法  $AP$  和  $OE$  指标的 Holm 检验结果,可以看出,本文算法性能与 MDDM<sub>proj</sub>, MDDM<sub>spc</sub>, MLNB, PMU, MIFS 有显著性差异,与 LSGL 无显著性差异.

为了进一步观察所选特征的数量如何影响各

算法的分类性能,图 4 为标记缺失率 50% 下 Slashdot 数据集上的预测分类情况.由图可见,选择的特征数量不同,分类效果也不同.另外,随着特征数目的增加,各项评价指标都不是单调增加或者单调递减的,证明不是所有的特征都对分类起作用,只有某些特征对分类效果的影响较大.

总之,WCFS 的分类性能优于其他算法,而且,随着所选特征数量的增加,其优势更加凸显.

为了评估 WCFS 在弱多标记数据中的鲁棒性,继续选用 Slashdot 数据集,掩盖标记的比例,以 10% 为区间,将 Slashdot 数据集中获得的标记从 10% 改变到 90%.为了一致性,仍然选择前 35% 的特征数进行比较.图 5 给出了不同比例获

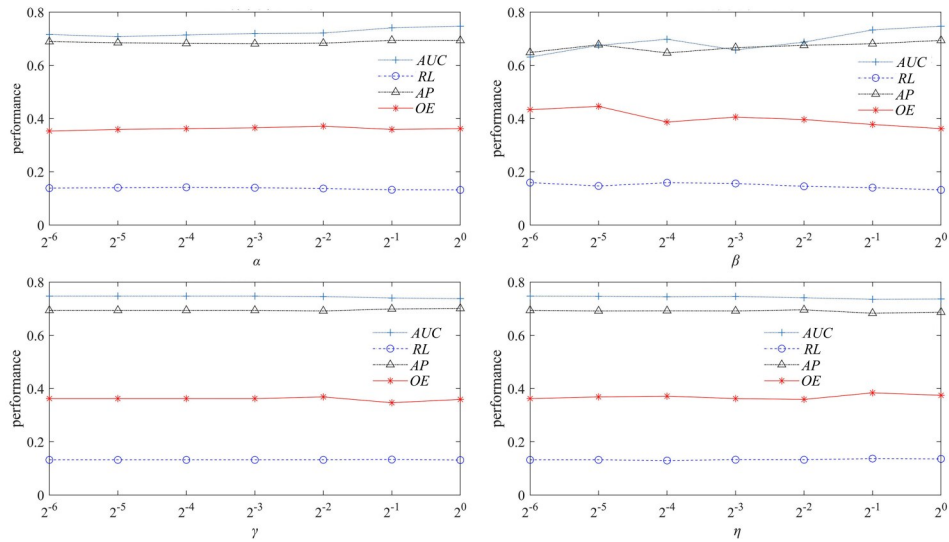


图2 WCFS在Birds数据集上标记缺失率为30%的参数敏感度分析

Fig. 2 Parameter sensitivity of WCFS on Birds dataset with 30% missing labels

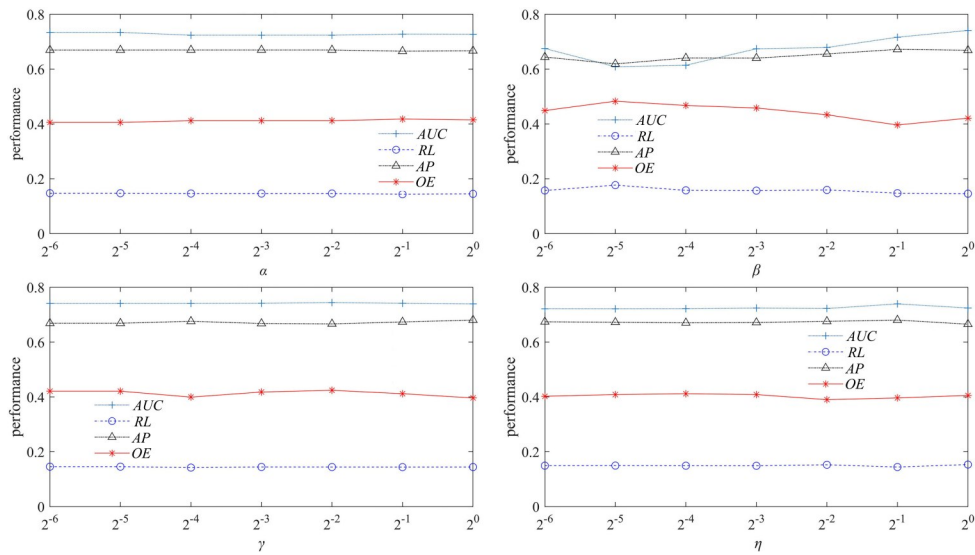


图3 WCFS在Birds数据集上标记缺失率为50%的参数敏感度分析

Fig. 3 Parameter sensitivity of WCFS on Birds dataset with 50% missing labels

表10 AUC指标的Holm检验结果

Table 10 Holm test results of subset AUC index

算法	$i$	$p$	Holm ( $\alpha/i$ )	假设 ( $\alpha=0.05$ )
MDDMspc	6	0.0016	0.0083	拒绝
MLNB	5	0.0016	0.0100	拒绝
MDDMproj	4	0.0114	0.0125	拒绝
PMU	3	0.0114	0.0167	拒绝
MIFS	2	0.0578	0.0250	接受
LSGL	1	0.2059	0.0500	接受

表11 RL指标的Holm检验结果

Table 11 Holm test results of subset RL index

算法	$i$	$p$	Holm ( $\alpha/i$ )	假设 ( $\alpha=0.05$ )
MDDMproj	6	0.0016	0.0083	拒绝
MLNB	5	0.0016	0.0100	拒绝
MDDMspc	4	0.0016	0.0125	拒绝
PMU	3	0.0016	0.0167	拒绝
MIFS	2	0.2059	0.0250	接受
LSGL	1	0.5271	0.0500	接受

得标记数目的不同指标的曲线,证明 WCFS 在不同性能指标和不同比例的不完整标记信息下,均

表 12  $AP$  指标的 Holm 检验结果

Table 12 Holm test results of subset  $AP$  index

算法	$i$	$p$	Holm ( $\alpha/i$ )	假设 ( $\alpha=0.05$ )
MDDM <sub>spc</sub>	6	0.0016	0.0083	拒绝
MDDM <sub>proj</sub>	5	0.0016	0.0100	拒绝
MLNB	4	0.0016	0.0125	拒绝
PMU	3	0.0016	0.0167	拒绝
MIFS	2	0.0114	0.0250	拒绝
LSGL	1	0.0578	0.0500	接受

优于其他所有方法,显示了 WCFS 处理弱多标记数据的能力。

表 13  $OE$  指标的 Holm 检验结果

Table 13 Holm test results of subset  $OE$  index

算法	$i$	$p$	Holm ( $\alpha/i$ )	假设 ( $\alpha=0.05$ )
MDDM <sub>spc</sub>	6	0.0016	0.0083	拒绝
MDDM <sub>proj</sub>	5	0.0016	0.0100	拒绝
MLNB	4	0.0016	0.0125	拒绝
PMU	3	0.0016	0.0167	拒绝
MIFS	2	0.0114	0.0250	拒绝
LSGL	1	0.2059	0.0500	接受

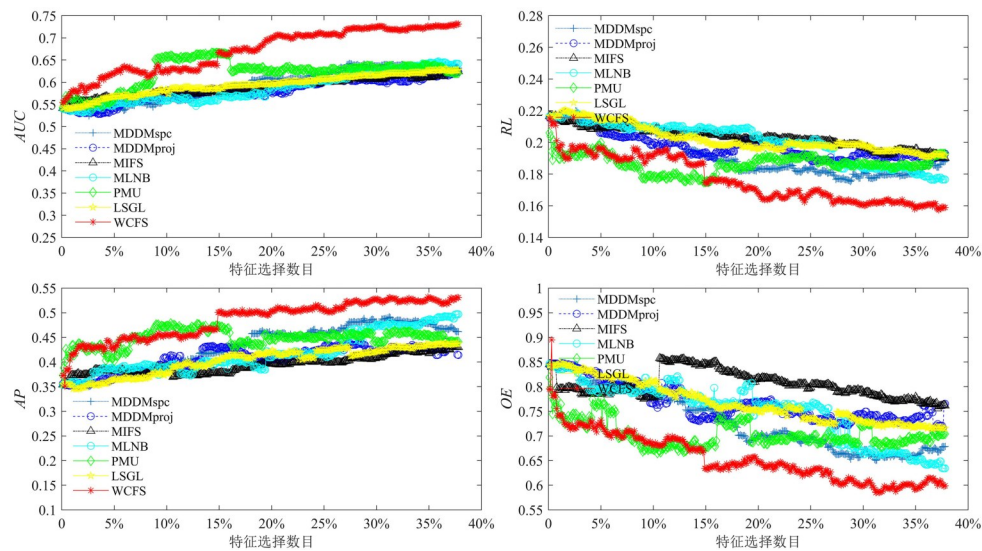


图 4 在 Slashdot 数据集上标记缺失率 50% 时的预测分类情况

Fig. 4 Predictive classification situation on Slashdot dataset with 50% missing labels

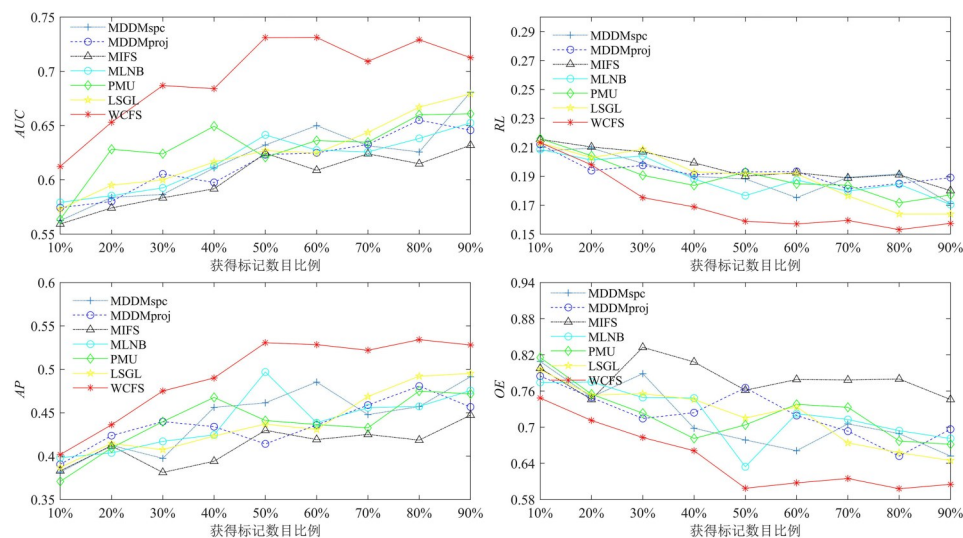


图 5 在 Slashdot 数据集上,  $AUC$ ,  $RL$ ,  $AP$  和  $OE$  在不同比例获得标记数目的效果

Fig. 5  $AUC$ ,  $RL$ ,  $AP$  and  $OE$  with different ratios of observed labels on Slashdot dataset

## 4 结论

现有的多标记学习算法大多依赖具有大量标记的数据,但是人工标注数据成本高昂、数据包含噪声、标记空间不完备,这些问题制约了多标记学习算法的性能. 本文立足于大规模弱标记数据,提出一种新型的基于弱监督对比学习的弱多标记特征选择算法,可以恢复缺失的标记空间,并充分挖掘不同类之间的对比模式,完善 $k$ 近邻机制,削弱噪声数据对训练过程产生的负面影响,提高标记一致性. 对比实验验证了算法的有效性和鲁棒性. 针对各种噪声,如何引入自监督学习模式并实现有效的多标记学习算法是未来进一步研究的重点课题.

### 参考文献

- [1] 李志欣,卓亚琦,张灿龙,等. 多标记学习研究综述. 计算机应用研究, 2014, 31(6): 1601—1605. (Li Z X, Zhuo Y Q, Zhang C L, et al. Survey on multi-label learning. Application Research of Computers, 2014, 31(6): 1601—1605.)
- [2] Jabreel M, Moreno A. A deep learning - based approach for multi - label emotion classification in tweets. Applied Sciences, 2019, 9(6): 1123.
- [3] Zhang M L, Zhou Z H. A review on multi - label learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819—1837.
- [4] Liu H W, Li X L, Zhang S C. Learning instance correlation functions for multilabel classification. IEEE Transactions on Cybernetics, 2017, 47(2): 499—510.
- [5] 马宏亮,万建武,王洪元. 一种嵌入样本流形结构与标记相关性的多标记降维算法. 南京大学学报(自然科学), 2019, 55(1): 92—101. (Ma H L, Wan J W, Wang H Y. A multi-label dimensionality reduction algorithm embedded sample manifold structure and label correlation. Journal of Nanjing University (Natural Science), 2019, 55(1): 92—101.)
- [6] 程玉胜,陈飞,庞淑芳. 标记倾向性的粗糙互信息 $k$ 特征核选择. 南京大学学报(自然科学), 2020, 56(1): 19—29. (Cheng Y S, Chen F, Pang S F.  $k$ -Kernel feature selection of tendentious labels based on rough mutual information. Journal of Nanjing University (Natural Science), 2020, 56(1): 19—29.)
- [7] Naula P, Airola A, Salakoski T, et al. Multi-label learning under feature extraction budgets. Pattern Recognition Letters, 2014(40): 56—65.
- [8] 孙林,黄苗苗,徐久成. 基于邻域粗糙集和Relief的弱标记特征选择方法. 计算机科学, 2022, 49(4): 152—160. (Sun L, Huang M M, Xu J C. Weak label feature selection method based on neighborhood rough sets and Relief. Computer Science, 2022, 49(4): 152—160.)
- [9] 钱文彬,黄琴,王映龙,等. 多标记不完备数据的特征选择算法. 计算机科学与探索, 2019, 13(10): 1768—1780. (Qian W B, Huang Q, Wang Y L, et al. Feature selection algorithm in multi-label incomplete data. Journal of Frontiers of Computer Science and Technology, 2019, 13(10): 1768—1780.)
- [10] Lin Y J, Hu Q H, Liu J H, et al. Multi-label feature selection based on max - dependency and min - redundancy. Neurocomputing, 2015(168): 92—103.
- [11] Gharroudi O, Elghazel H, Aussem A. A comparison of multi - label feature selection methods using the random forest paradigm//The 27<sup>th</sup> Canadian Conference on Artificial Intelligence. Springer Berlin Heidelberg, 2014: 95—106.
- [12] Zhang J, Luo Z M, Li C D, et al. Manifold regularized discriminative feature selection for multi-label learning. Pattern Recognition, 2019(95): 136—150.
- [13] Deriu J, Lucchi A, De Luca V, et al. Leveraging large amounts of weakly supervised data for multi-language sentiment classification//Proceedings of the 26<sup>th</sup> International Conference on World Wide Web. Perth, Australia: ACM, 2017: 1045—1052.
- [14] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations. 2020, arXiv:2002.05709.
- [15] Van Den Oord A, Li Y Z, Vinyals O. Representation learning with contrastive predictive coding. 2018, arXiv:1807.03748.
- [16] Zhang M L, Wu L. Lift: Multi-label learning with label-specific features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 107—120.



- [17] 吴磊,张敏灵. 基于类属属性的多标记学习算法. 软件学报, 2014, 25(9): 1992—2001. (Wu L, Zhang M L. Research of label-specific features on multi-label learning algorithm. Journal of Software, 2014, 25(9): 1992—2001.)
- [18] Huang J, Li G R, Huang Q M, et al. Learning label-specific features and class-dependent labels for multi-label classification. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3309—3323.
- [19] Zhang J, Li C D, Cao D L, et al. Multi-label learning with label-specific features by resolving label correlations. Knowledge-Based Systems, 2018(159): 148—157.
- [20] Zhu P F, Xu Q, Hu Q H, et al. Multi-label feature selection with missing labels. Pattern Recognition, 2018(74): 488—502.
- [21] Wang C X, Lin Y J, Liu J H. Feature selection for multi-label learning with missing labels. Applied Intelligence, 2019, 49(8): 3027—3042.
- [22] Vapnik V, Sterin A M. On structural risk minimization or overall risk in a problem of pattern recognition. Automation and Remote Control, 1977, 10(3): 1495—1503.
- [23] Chapelle O, Scholkopf B, Zien A. Semi-supervised learning. IEEE Transactions on Neural Networks, 2009, 20(3): 542.
- [24] 李娜,潘志松,周星宇. 基于多标记重要性排序的分类器链算法. 模式识别与人工智能, 2016, 29(6): 567—575. (Li N, Pan Z S, Zhou X Y. Classifier chain algorithm based on multi-label importance rank. Pattern Recognition and Artificial Intelligence, 2016, 29(6): 567—575.)
- [25] Wu B Y, Liu Z L, Wang S F, et al. Multi-label learning with missing labels//2014 22<sup>nd</sup> International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014: 1964—1968.
- [26] Lin Z C, Liu R S, Su Z X. Linearized alternating direction method with adaptive penalty for low-rank representation//Proceedings of the 24<sup>th</sup> International Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc., 2011: 612—620.
- [27] Zhang Y, Zhou Z H. Multilabel dimensionality reduction via dependence maximization. ACM Transactions on Knowledge Discovery from Data, 2010, 4(3): 14.
- [28] Jian L, Li J D, Shu K, et al. Multi-label informed feature selection//Proceedings of the 25<sup>th</sup> International Joint Conference on Artificial Intelligence. New York, NY, USA: AAAI Press, 2016: 1627—1633.
- [29] Zhang M L, Peña J M, Robles V. Feature selection for multi-label naive bayes classification. Information Sciences, 2009, 179(19): 3218—3229.
- [30] Lee J, Kim D W. Feature selection for multi-label classification using multivariate mutual information. Pattern Recognition Letters, 2013, 34(3): 349—357.
- [31] Zhao D W, Gao Q W, Lu Y X, et al. Learning multi-label label-specific features via global and local label correlations. Soft Computing, 2022, 26(5): 2225—2239.
- [32] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 2007, 40(7): 2038—2048.

(责任编辑 杨可盛)