

基于全局冗余最小的快速多标签特征选择方法

唐杰烽,张佳*,龙锦益

(暨南大学信息科学技术学院,广东广州510632)

摘要:为解决多标签学习维度灾难及过滤式特征选择方法易陷入局部最优的问题,提出一种基于全局冗余最小(global redundancy minimization, GRM)的快速多标签特征选择方法。通过K-means聚类和互信息计算从原始标签空间和特征空间中筛选出候选标签与候选特征子集;通过全局冗余最小化解决局部最优问题,得到特征冗余最小特征权重,保证输出的特征子集为最佳特征子集;采用集成学习策略增强特征选择稳定性。在14个多标签数据集上的试验结果表明所提方法相较其他方法在各分类指标上均有较优表现。

关键词:多标签学习;特征选择;维度灾难;全局冗余最小

中图分类号:TP181

文献标志码:A

引用格式:唐杰烽,张佳,龙锦益.基于全局冗余最小的快速多标签特征选择方法[J].山东大学学报(工学版),2025,55(6):21-34.

TANG Jiefeng, ZHANG Jia, LONG Jinyi. Fast multi-label feature selection method based on global redundancy minimization [J]. Journal of Shandong University (Engineering Science), 2025, 55(6):21-34.

Fast multi-label feature selection method based on global redundancy minimization

TANG Jiefeng, ZHANG Jia*, LONG Jinyi

(College of Information Science and Technology, Jinan University, Guangzhou 510632, Guangdong, China)

Abstract: To solve the problem of the curse of dimensionality in multi-label learning and the problem that filter feature selection methods were prone to fall into local optima, a fast multi-label feature selection method based on global redundancy minimization was proposed. Candidate labels and candidate feature subsets were selected from the original label space and feature space through K-means clustering and mutual information calculation; the local optima problem was solved by minimizing global redundancy, and the feature weight with minimum feature redundancy was obtained to ensure outputting the best feature subset; an ensemble learning strategy was used to enhance the stability of feature selection. The experimental results on 14 multi-label datasets showed that the proposed method had better performance than other methods in all classification indicators.

Keywords: multi-label learning; feature selection; curse of dimensionality; fast global redundancy minimization

0 引言

在传统监督学习任务中,每个样本被默认为只包含一种分类标签。这种假设往往与现实世界的真实情况不符,比如在图片分类任务中,一张草原风景图片同时包含“草地”、“花朵”、“蓝天”、“白云”等景色,使用传统单标签分类方法难以对这种

情况进行准确分类。多标签学习(multi-label learning, ML)^[1]应运而生,并在文本分类^[2]、视频分类^[3]、目标检测^[4]等领域取得出色成果。相较单标签学习,多标签学习输入输出空间维度更大。然而,过高特征维度将导致维度灾难问题^[5],使多标签学习任务效率低下,如何有效解决多标签学习任务维度灾难问题至关重要。

特征选择方法是解决维度灾难问题有效途径

收稿日期:2024-11-01

基金项目:国家自然科学基金资助项目(62106084, 62276115);广东省中医药信息化重点实验室资助项目(2021B1212040007)

第一作者简介:唐杰烽(1999—),男,广东佛山人,硕士研究生,主要研究方向为机器学习、数据挖掘。E-mail:jiefengtang@stu2022.jnu.edu.cn

*通信作者简介:张佳(1991—),男,湖南衡阳人,讲师,硕士生导师,博士,主要研究方向为机器学习、数据挖掘。E-mail:jiazhang@jnu.edu.cn

之一,它根据一定评价标准,从原始特征空间中去除不相关以及冗余特征,降低特征维度,提高分类性能^[6-7]。基于信息论的特征选择方法在多标签学习中应用广泛^[8-15],但该方法耗费大量时间且易陷入局部最优。

本研究提出一种基于全局冗余最小的(global redundancy minimization, GRM)快速多标签特征选择方法(简称 GRMfast 算法)以解决上述问题。首先,应用聚类算法并选取相同比例的标签获取候选标签,计算原始特征与候选标签互信息得到具有较大相关性的特征作为候选特征,减少不必要的熵计算,降低计算量;然后,基于候选标签子集与候选特征的相关性和候选特征之间的冗余性,通过全局冗余最小化解决局部最优问题,得到特征冗余最小特征权重,保证输出的特征子集为最佳特征子集;最后,考虑到候选标签可能无法完全覆盖整个原始标签空间导致学习性能下降,通过集成学习增强特征选择稳定性,将训练多个模型得到的特征选择结果集成投票得到最终特征选择结果。

1 相关工作

基于信息论的特征评价方法主要分为特征冗余最小化和新分类信息最大化2类^[16]。特征冗余最小化方法侧重于对特征冗余进行建模以提高性能。文献[17-19]使用多元互信息来测量多个标签下的条件冗余。一些研究者对标签信息利用也给予关注。文献[13]为提高特征子集分类性能,结合强相关标签增益和标签互助2个概念提出基于强相关标签增益和标签互助的多标签特征选择算法。文献[20]为提高分类准确率并避免过拟合,提出类标签特定互信息算法用于多标签特征选择(multi-label feature selection, MLFS)。新分类信息最大化是选取冗余度较小且有较强预测能力的特征子集。对于 MLFS,该类方法尝试扩展特征选择和多标签分类联合学习的标准^[16],或直接采用该标准实现特征选择^[21]。启发式搜索策略被广泛用于对特征进行排序^[9,10,22-24],但这类方法没有考虑到全局特征冗余,易陷入局部最优,可能无法找到最佳特征子集。此外,这种启发式搜索非常耗时,评价函数中涉及大量无效且重复的熵计算,降低了特征选择效率。另一种策略将特征选择过程视为优化问题。一些 MLFS 方法尝试使

用上述标准获得优化特征权重^[14-15]。然而这些方法利用标签信息的能力较差,且使用梯度下降策略的迭代优化可能导致大规模数据分析效率低下。

为解决效率问题,文献[10]通过放宽特征评价函数、避免重复熵计算以降低特征评价函数计算成本。文献[20]考虑待选特征和已选特征之间的条件冗余,以及待选特征与标签之间依赖关系,提出基于最大依赖性和最小冗余性的多标签特征选择算法;文献[11]通过约束凸优化构造一种基于互信息的多标签特征选择算法;文献[12]提出一种考虑优化全局相关性和冗余性的多标签特征选择方法。文献[25]忽略特征间冗余性影响,只考虑特征与标签之间的相关性,提出一种简单快速多标签特征选择算法。上述方法主要关注特征选择计算效率问题,未很好地兼顾分类性能。

2 GRMfast 算法

已知 $X \subseteq \mathbf{R}^d$ 为 d 维输入特征空间, $Y = \{y_1, y_2, \dots, y_m\}$ 为拥有 m 个标签的标签空间。给定一个多标签数据集 $D = \{(\mathbf{x}_i, \mathbf{y}_i) | 1 \leq i \leq n, \mathbf{x}_i \in X, \mathbf{y}_i \subseteq Y\}$, 其中包含 N 个样本, 1 个示例由 1 个 d 维特征向量 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \in X$ 表征, 标签向量 $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{im})$ ($1 \leq i \leq n$) 为与 \mathbf{x}_i 相关的标签子集。假设在一个多标签数据集 D 中, 原始特征空间中的特征集为 $F = \{f_1, f_2, \dots, f_d\}$, 训练数据矩阵为 $\mathbf{F} = [f_{11}, f_{12}, \dots, f_{1d}, f_{21}, f_{22}, \dots, f_{2d}, \dots, f_{n1}, f_{n2}, \dots, f_{nd}]$, f_i ($1 \leq i \leq d$) 是包含特征 f_i 信息的向量。每个实例与 m 个可能的标签所组成的标签集 $L = \{l_1, l_2, \dots, l_m\}$ 有关联。假设所有标签被聚类算法分成 c 组, 同一组中标签相似。GRMfast 旨在通过全局冗余最小的学习框架得到优化特征权重, 保证所选特征子集是全局最优的同时, 提高特征选择效率。GRMfast 算法的总体流程如下所示。

算法 GRMfast 算法

输入: 一个包含全部特征的特征集合 $F = \{f_1, f_2, \dots, f_d\}$ 和标签集 $L = \{l_1, l_2, \dots, l_m\}$ 的多标签数据集及标签选择比例参数 p 、特征选择比例参数 q 、聚类中心数 c 、集成模型个数 E ;

输出: 排名前 d 的特征;

1. 通过 K -means 算法将标签集 L 分成 c 组;
2. for $i = 1$ to E do
3. 通过从每组标签中随机选取相同比例 p

的标签生成标签子集 L_p ;

4. 通过求解式(1)得出候选特征子集 F_c ;
5. 计算 F_c 中特征和 L_p 中标签的互信息;
6. 计算 F_c 中特征之间的互信息;
7. 通过优化式(3)得到 $\mathbf{z}^{(i)}$;
8. end for
9. $\hat{\mathbf{Z}} = \frac{1}{E} \sum_{i=1}^E \mathbf{Z}^{(i)}$;
10. 通过对 $\|\hat{\mathbf{z}}_i\|_2 (i \in \{1, 2, \dots, d\})$ 降序排序得到排名前 d 的特征。

2.1 生成候选标签子集与候选特征

本研究从原始标签空间中筛选候选标签以降低优化目标函数复杂度。采用 K -means 聚类算法将标签分为 c 组,再从每组中随机选取相同比例 p 的标签,得到标签子集 L_p ,且 $L_p \subseteq L$ 。从原始特征空间中筛选比例为 q 的候选特征,在降低计算成本同时减少有效信息损失。候选特征子集 F_c 需与标签子集 L_p 较强相关, F_c 需满足

$$\max_{F_c} I(F_c, L_p), \text{ s.t. } F_c \subseteq F, \frac{|F_c|}{|F|} = q. \quad (1)$$

GRMfast 采用网格搜索策略寻找最优参数 p 和 q ,使得测试集上平均分类结果 M_{ACR} 最小。

$$M_{ACR}(\delta_{p,q}) = \sum_{i=1}^{30} (H_i(f, U) + R_i(f, U)), \quad (2)$$

式中, $\delta_{p,q}$ 为算法的参数组合, f 为分类器, U 为测试集, $H_i(f, U)$ 、 $R_i(f, U)$ 分别为输出首 30 个特征时对应的海明损失和排序损失。

2.2 基于全局冗余最小的学习框架

GRMfast 方法目标函数为

$$\min_{\mathbf{z}^T \mathbf{1} = 1, \mathbf{z} \geq 0} \frac{\mathbf{z}^T \mathbf{A} \mathbf{z}}{\mathbf{z}^T \mathbf{s}}, \quad (3)$$

式中: \mathbf{s} 为每个候选特征与标签子集的互信息, $\mathbf{s} = (s_1, s_2, \dots, s_{d_1})^T$, $\mathbf{s} \in \mathbf{R}^{n \times 1}$, d_1 为候选特征个数, s_{d_1} 为第 d_1 个候选特征与标签子集的互信息, $s_{d_1} = \sum_{j=1}^{d_1} I(f_{d_1}, l_j)$; \mathbf{A} 为特征冗余, $\mathbf{A} \in \mathbf{R}^{d_1 \times d_1}$; \mathbf{z} 为特征权重, $\mathbf{z} \in \mathbf{R}^{d_1}$ 。

公式(3)中 $\mathbf{z}^T \mathbf{A} \mathbf{z}$ 为全局特征冗余,分母 $\mathbf{z}^T \mathbf{s}$ 为特征权重 \mathbf{z} 和互信息 \mathbf{s} 之间的一致性,即特征权重 \mathbf{z} 与特征重要性的匹配程度,权重应优先分配给互信息高的重要特征,避免违背其初始贡献。

采用增广拉格朗日法 (augmented lagrangian multiplier, ALM)^[26] 求解公式(3)。该方法相较于其他优化方法求解过程较为简单,收敛速度较快,适

合处理高维多标签数据。ALM 收敛性在文献[26]中已有证明。

公式(3)转换为

$$\min_{\mathbf{z}^T \mathbf{1} = 1, \mathbf{z} \geq 0, \mathbf{z} = \mathbf{v}} \mathbf{v}^T \mathbf{A} \mathbf{v} - \boldsymbol{\lambda} \mathbf{v}^T \mathbf{s}, \quad (4)$$

式中, \mathbf{v} 为中间变量, $\boldsymbol{\lambda} = \frac{\mathbf{z}^T \mathbf{A} \mathbf{z}}{\mathbf{z}^T \mathbf{s}}$ 。

根据 ALM^[26] 定理,公式(4)等价于求解下述优化问题

$$\min_{\mathbf{z}^T \mathbf{1} = 1, \mathbf{z} \geq 0} \mathbf{v}^T \mathbf{A} \mathbf{v} - \boldsymbol{\lambda} \mathbf{v}^T \mathbf{s} + \frac{\mu}{2} \left\| \mathbf{z} - \mathbf{v} + \frac{1}{\mu} \boldsymbol{\beta} \right\|_F^2, \quad (5)$$

式中, μ 为惩罚参数, $\boldsymbol{\beta}$ 为拉格朗日乘子向量。

根据 \mathbf{v} 和 \mathbf{s} 使式(5)最小困难,可以每次只更新其中一个变量。当 \mathbf{s} 确定时,式(5)简化为一个关于 \mathbf{v} 的非约束优化问题。当 \mathbf{v} 确定时,式(5)简化为

$$\min_{\mathbf{z}^T \mathbf{1} = 1, \mathbf{z} \geq 0} \left\| \mathbf{z} - \mathbf{v} + \frac{1}{\mu} \boldsymbol{\beta} \right\|_F^2. \quad (6)$$

令 $\mathbf{w} = \mathbf{v} - \frac{1}{\mu} \boldsymbol{\beta}$, 则式(6)对应的拉格朗日函数为

$$\frac{1}{2} \|\mathbf{z} - \mathbf{w}\|_F^2 - \gamma \mathbf{z}^T \mathbf{1} - \boldsymbol{\varepsilon}^T \mathbf{z}, \quad (7)$$

其中 γ 和 $\boldsymbol{\varepsilon}$ 为拉格朗日系数。将式(7)最优解记作 \mathbf{z}^* , 则对应拉格朗日系数为 γ^* 和 $\boldsymbol{\varepsilon}^*$ 。根据 KKT 条件^[26] 得

$$\forall i, \mathbf{z}_i^* - \mathbf{w}_i - \gamma^* - \boldsymbol{\varepsilon}_i^* = 0. \quad (8)$$

式(8)两边同时乘以 $\mathbf{1}^T$ 并由约束 $\mathbf{z}^T \mathbf{1} = 1$, 可得

$$\gamma^* = \frac{1 - \mathbf{1}^T \mathbf{w} - \mathbf{1}^T \boldsymbol{\varepsilon}^*}{d_1}. \quad (9)$$

将式(9)代入式(8)可得

$$\mathbf{z}^* = \left(\mathbf{w} - \frac{\mathbf{1} \mathbf{1}^T}{d_1} \mathbf{w} + \frac{1}{d_1} \mathbf{1} - \frac{\mathbf{1}^T \boldsymbol{\varepsilon}^*}{d_1} \mathbf{1} \right) + \boldsymbol{\varepsilon}^*. \quad (10)$$

令 $\overline{\boldsymbol{\varepsilon}^*} = \frac{\mathbf{1}^T \boldsymbol{\varepsilon}^*}{d_1}$, $\mathbf{u} = \mathbf{w} - \frac{\mathbf{1} \mathbf{1}^T}{d_1} \mathbf{w} + \frac{1}{d_1} \mathbf{1}$, 则式(10)等价

于下式

$$\mathbf{z}^* = \mathbf{u} + \overline{\boldsymbol{\varepsilon}^*} - \overline{\boldsymbol{\varepsilon}^*} \mathbf{1}. \quad (11)$$

由 KKT 条件和式(11), 可得

$$\mathbf{u}_i + \boldsymbol{\varepsilon}_i^* - \overline{\boldsymbol{\varepsilon}^*} = (\mathbf{u}_i - \overline{\boldsymbol{\varepsilon}^*})_+. \quad (12)$$

由式(11)和(12)可知

$$\mathbf{z}_i = (\mathbf{u}_i - \overline{\boldsymbol{\varepsilon}^*})_+, \quad (13)$$

当确定 $\overline{\boldsymbol{\varepsilon}^*}$ 值后, 可得最优解。

由 KKT 条件和式(13)可知 $\boldsymbol{\varepsilon}_i^* = (\overline{\boldsymbol{\varepsilon}^*} - \mathbf{u}_i)_+$, 则有

$$\overline{\boldsymbol{\varepsilon}^*} = \frac{1}{d_1} \sum_{i=1}^{d_1} (\overline{\boldsymbol{\varepsilon}^*} - \mathbf{u}_i)_+. \quad (14)$$

定义如下函数:

$$f(\bar{\boldsymbol{\varepsilon}}) = \frac{1}{d_1} \sum_{i=1}^{d_1} (\bar{\boldsymbol{\varepsilon}}^* - \mathbf{u}_i)_+ - \bar{\boldsymbol{\varepsilon}}, \quad (15)$$

则 $\bar{\boldsymbol{\varepsilon}}^*$ 为 $f(\bar{\boldsymbol{\varepsilon}}) = 0$ 的根,通过牛顿法求得。将 $\bar{\boldsymbol{\varepsilon}}^*$ 代回式(13)可得最优解,即特征权重。

2.3 集成特征选择

假设训练 E 个模型用于特征选择,如果第 i 个模型($1 \leq i \leq E$)是由 $L_p^{(i)}$ 表示的随机候选标签子集执行,利用 $L_p^{(i)}$ 来搜索相应候选相关特征子集 $F_c^{(i)}$,可得特征选择结果 $\mathbf{Z}^{(i)}$

$$\mathbf{Z}^{(i)} = \arg \min_{\mathbf{z}} f(L_p^{(i)}, F_c^{(i)}). \quad (16)$$

当得到所有特征选择结果 $\{\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(E)}\}$ 后,利用集成投票方式得到最终特征选择结果 $\hat{\mathbf{Z}}$, $\hat{\mathbf{Z}} = \frac{1}{E} \sum_{i=1}^E \mathbf{Z}^{(i)}$ 。基于此,通过衡量 $\|\hat{\mathbf{z}}_i\|_2$ ($1 \leq i \leq d$)值来估计每个特征重要性,从而得到最终特征子集。

2.4 时间复杂度分析

GRMfast算法时间复杂度主要由生成候选标签子集和候选特征子集、互信息计算、采用集成学习的特征选择4部分组成。为随机生成候选标签子集,需通过聚类生成若干组,时间复杂度为 $O(Tcm)$,其中 T 为聚类迭代次数, c 为聚类中心数, m 为原始标签空间标签数。为生成候选特征子集,需计算所有原始特征和候选标签子集的互信息,时间复杂度为 $O(dm_1)$,其中 d 为数据集的原始特征数, m_1 为候选标签子集中标签数量。为进一步计算候选特征和候选标签的互信息及候选特征之间的互信息,时间复杂度为 $O(d_1^2 + d_1 m_1)$,其中 d_1 为候选特征子集中的特征数量。在基于集成学习的特征选择方面,时间开销随集成模型个数线性增长,由于 E 为常数,因此时间复杂度为 $O(1)$ 。在每次训练1个新模型时,只需针对新添加标签和特征更新互信息,集成学习带来的时间开销对算法总时间开销而言可忽略不计。综上,GRMfast总时间复杂度为 $O(Tcm + d_1^2 + dm_1)$ 。

3 试验与分析

3.1 试验数据

试验数据主要来自公开多标签数据集网站 Multi-label Classification Dataset (<http://www.uco.es/kdis/mlresources/>)。试验所采用数据集中9个数据集来自文本领域,包括 Bibtex、4个 Stackex

数据集 (Chemistry、Philosophy、Cs、Cooking)、4个 Yahoo 数据集 (Computer、Entertainment、Science、Reference),提取特征为不同词在文本中频率,标签为文本类别信息。4个数据集来自多媒体领域,分别是 Birds、Core15K、Emotions 和 CAL500。Yeast数据集来自生物学领域。所选数据集详细信息如表1所示。

表1 试验数据集信息
Table 1 Description of the experimental dataset

数据集	样本数	特征数	标签数	训练集数	测试集数	数据类型
Birds	645	260	19	322	323	Audio
Chemistry	6 961	540	175	4 686	2 275	Text
Philosophy	3 971	842	233	2 623	1 348	Text
Core15K	5 000	499	374	4 500	500	Image
Yeast	2 417	103	14	1 499	918	Biology
Entertainment	5 000	640	21	2 000	3 000	Text
Reference	5 000	793	33	2 000	3 000	Text
Science	5 000	743	40	2 000	3 000	Text
Computer	5 000	681	33	2 000	3 000	Text
Emotions	593	72	6	391	202	Music
CAL500	502	68	174	251	251	Music
Bibtex	7 395	1 836	159	4 880	2 515	Text
Cs	9 270	635	274	6 115	3 155	Text
Cooking	10 491	577	400	6 990	3501	Text

3.2 试验设置

为与其他方法比较,本研究使用海明损失 L_H ,排序损失 L_R ,覆盖率 P_C ,平均准确率 A_p ,宏平均 $F1_{Ma}$,微平均 $F1_{Mi}$ 6个评价指标评估本研究算法的分类性能,完整评价指标定义查阅文献[27]。

试验采用 MDMR^[9]、FIMF^[10]、MICO^[11]、MIFS^[21]、MC-GM^[28]、MCLS^[29]、LRDG^[30]、MFS-ADGO^[31]8种多标签特征选择方法作为对比算法。实验采用多标签最近邻^[32]作为分类算法评估特征选择后的数据集,分类器中平滑因子设置为1,近邻数量设置为10。对比算法的参数设置按原文给出的推荐参数进行设置。GRMfast中参数 p 和 q 在 $\{10\%, 25\%, \dots, 100\%\}$ 内进行调参,聚类中心数 c 设置为5,集成模型个数 E 设置为70。

3.3 多标签特征选择算法分类性能比较

使用每种算法所选择的首50个特征所得结果的均值进行比较。9种多标签特征选择算法在6个评价指标的试验结果如表2~7所示。其中, L_H 、 L_R 、 P_C 3项指标越小,表示分类性能越优; A_p 、 $F1_{Ma}$ 、 $F1_{Mi}$ 指标数值越大,表示分类性能越优。黑体数字表示该算法在相应指标中优于其他算法。

表 2 不同算法的 L_H 均值对比
Table 2 Comparison mean results of different algorithms on L_H

数据集	L_H								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.049 5	0.065 3	0.064 9	0.049 5	0.053 7	0.062 5	0.055 9	0.060 1	0.066 9
Chemistry	0.012 1	0.012 3	0.012 2	0.012 2	0.012 2	0.012 1	0.012 2	0.012 2	0.012 1
Philosophy	0.009 0	0.009 2	0.009 4	0.009 0	0.009 2	0.009 4	0.009 5	0.009 5	0.009 5
Corel5K	0.009 4	0.009 4	0.009 4	0.009 5	0.009 5	0.009 4	0.009 4	0.009 4	0.009 5
Yeast	0.199 4	0.214 2	0.220 2	0.216 9	0.206 2	0.199 3	0.203 0	0.204 0	0.203 1
Entertainment	0.058 3	0.061 2	0.064 5	0.059 3	0.061 9	0.064 3	0.065 6	0.063 9	0.065 5
Reference	0.029 4	0.030 0	0.029 6	0.028 4	0.030 5	0.034 4	0.035 2	0.034 9	0.034 6
Science	0.033 5	0.034 6	0.034 5	0.033 6	0.035 5	0.035 7	0.035 9	0.035 5	0.035 5
Emotions	0.218 6	0.225 2	0.222 8	0.223 6	0.225 2	0.257 4	0.222 8	0.230 2	0.225 2
Computer	0.039 4	0.039 8	0.039 6	0.038 4	0.040 5	0.041 7	0.042 7	0.043 0	0.042 9
CAL500	0.139 4	0.141 6	0.140 0	0.141 7	0.139 8	0.142 1	0.142 0	0.141 8	0.141 7
Bibtex	0.013 5	0.014 0	0.014 6	0.013 7	0.014 1	0.015 1	0.014 9	0.015 1	0.013 8
Cs	0.008 7	0.008 9	0.009 0	0.008 8	0.008 8	0.009 1	0.009 0	0.009 0	0.009 1
Cooking	0.005 2	0.005 5	0.005 5	0.005 2	0.005 5	0.005 5	0.005 5	0.005 5	0.005 5

表 3 不同算法的 L_R 均值对比
Table 3 Comparison mean results of different algorithms on L_R

数据集	L_R								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.124 7	0.127 9	0.127 4	0.128 9	0.151 3	0.142 4	0.137 5	0.151 8	0.128 7
Chemistry	0.148 6	0.168 4	0.166 7	0.165 0	0.170 3	0.164 2	0.175 3	0.172 6	0.172 0
Philosophy	0.135 6	0.151 3	0.153 6	0.144 1	0.137 9	0.146 0	0.152 6	0.154 1	0.155 8
Corel5K	0.137 5	0.142 7	0.142 7	0.142 2	0.143 7	0.144 7	0.139 3	0.140 5	0.139 6
Yeast	0.175 6	0.199 6	0.192 5	0.186 9	0.178 6	0.173 8	0.176 3	0.182 2	0.178 6
Entertainment	0.122 2	0.136 0	0.139 8	0.127 0	0.131 4	0.142 2	0.141 0	0.130 7	0.137 1
Reference	0.089 8	0.094 9	0.092 7	0.095 2	0.093 1	0.097 4	0.104 0	0.104 1	0.102 3
Science	0.135 9	0.140 8	0.143 4	0.141 5	0.153 2	0.151 7	0.155 0	0.149 5	0.155 4
Emotions	0.172 3	0.215 5	0.231 1	0.181 1	0.169 7	0.194 7	0.178 2	0.195 6	0.192 9
Computer	0.090 5	0.097 8	0.095 0	0.092 8	0.098 6	0.099 1	0.098 8	0.101 3	0.101 3
CAL500	0.188 8	0.189 2	0.189 0	0.189 7	0.191 5	0.190 0	0.191 7	0.190 3	0.191 0
Bibtex	0.167 9	0.252 2	0.276 4	0.282 0	0.249 4	0.284 4	0.283 5	0.278 6	0.240 8
Cs	0.140 7	0.165 8	0.169 1	0.164 8	0.141 1	0.167 2	0.169 8	0.170 8	0.170 2
Cooking	0.169 9	0.206 1	0.209 0	0.195 5	0.216 2	0.189 0	0.213 8	0.215 1	0.210 7

表 4 不同算法的 P_C 均值对比
Table 4 Comparison mean results of different algorithms on P_C

数据集	P_C								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.169 7	0.167 2	0.169 0	0.180 2	0.187 6	0.204 5	0.182 8	0.200 3	0.169 0
Chemistry	0.246 8	0.269 2	0.266 8	0.265 9	0.297 7	0.263 5	0.277 1	0.273 8	0.275 1
Philosophy	0.243 2	0.262 2	0.265 2	0.253 2	0.246 2	0.258 8	0.263 9	0.263 8	0.266 4
Corel5K	0.316 4	0.324 4	0.324 3	0.322 3	0.322 1	0.324 0	0.318 1	0.321 3	0.319 4
Yeast	0.462 6	0.492 4	0.486 7	0.478 6	0.468 8	0.459 5	0.456 7	0.467 4	0.460 8
Entertainment	0.156 9	0.171 0	0.174 0	0.164 9	0.166 1	0.176 5	0.176 1	0.165 0	0.170 7
Reference	0.104 6	0.117 1	0.114 1	0.111 3	0.108 2	0.112 6	0.119 6	0.119 4	0.118 0
Science	0.171 7	0.177 2	0.179 7	0.178 8	0.189 4	0.187 9	0.191 2	0.185 6	0.192 1
Emotions	0.317 7	0.363 0	0.329 2	0.326 7	0.323 4	0.334 2	0.322 6	0.337 5	0.340 8
Computer	0.132 7	0.140 1	0.137 2	0.135 1	0.141 7	0.140 9	0.140 6	0.143 5	0.144 6
CAL500	0.754 2	0.746 4	0.744 6	0.744 8	0.747 2	0.744 9	0.747 4	0.747 8	0.747 2
Bibtex	0.275 3	0.377 9	0.411 9	0.425 2	0.383 1	0.432 4	0.427 2	0.430 1	0.363 6
Cs	0.259 9	0.294 1	0.297 3	0.294 7	0.262 1	0.294 5	0.300 4	0.301 3	0.298 9
Cooking	0.290 2	0.336 2	0.338 2	0.322 8	0.346 9	0.313 7	0.344 6	0.345 6	0.339 1

表5 不同算法的 A_p 均值对比
Table 5 Comparison mean results of different algorithms on A_p

数据集	A_p								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.712 0	0.649 9	0.669 1	0.710 2	0.678 9	0.641 1	0.683 1	0.661 9	0.655 6
Chemistry	0.296 7	0.247 7	0.247 7	0.264 8	0.249 1	0.259 0	0.233 8	0.235 6	0.247 2
Philosophy	0.335 8	0.267 9	0.246 5	0.331 2	0.316 5	0.248 5	0.246 4	0.237 5	0.233 0
Core15K	0.225 9	0.215 8	0.216 3	0.217 0	0.187 3	0.196 7	0.224 6	0.217 4	0.228 7
Yeast	0.758 8	0.716 9	0.732 0	0.739 6	0.753 2	0.739 8	0.752 1	0.739 3	0.747 4
Entertainment	0.579 1	0.522 4	0.511 1	0.556 0	0.535 7	0.513 8	0.504 2	0.536 0	0.515 5
Reference	0.618 4	0.597 7	0.606 4	0.608 4	0.607 1	0.579 7	0.567 8	0.568 8	0.572 5
Science	0.476 1	0.458 2	0.454 2	0.469 3	0.401 1	0.402 1	0.393 5	0.409 0	0.396 1
Emotions	0.785 2	0.763 8	0.735 9	0.783 5	0.782 5	0.768 0	0.775 1	0.763 9	0.769 9
Computer	0.638 4	0.615 9	0.611 3	0.633 7	0.618 9	0.610 5	0.604 3	0.603 5	0.607 6
CAL500	0.486 0	0.480 5	0.482 5	0.479 5	0.477 1	0.482 9	0.477 4	0.476 3	0.476 3
Bibtex	0.411 6	0.244 3	0.195 9	0.244 7	0.248 5	0.192 7	0.180 8	0.183 1	0.260 7
Cs	0.332 8	0.255 6	0.221 3	0.276 7	0.335 6	0.217 7	0.223 5	0.219 7	0.216 3
Cooking	0.291 1	0.176 4	0.147 0	0.236 8	0.135 9	0.175 8	0.130 6	0.131 1	0.132 5

表6 不同算法的 $F1_{Ma}$ 均值对比
Table 6 Comparison mean results of different algorithms on $F1_{Ma}$

数据集	$F1_{Ma}$								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.176 7	0.123 5	0.106 7	0.152 1	0.079 9	0.078 2	0.086 0	0.061 2	0.089 5
Chemistry	0.011 8	0.002 5	0.002 3	0.008 0	0.004 5	0.011 9	0.000 0	0.003 9	0.005 2
Philosophy	0.024 6	0.006 9	0.001 9	0.022 0	0.022 4	0.012 4	0.003 2	0.002 1	0.002 1
Core15K	0.298 1	0.296 7	0.296 8	0.297 2	0.297 5	0.296 9	0.296 9	0.298 7	0.297 7
Yeast	0.368 0	0.253 6	0.275 8	0.291 6	0.356 0	0.299 0	0.345 2	0.333 6	0.329 4
Entertainment	0.139 6	0.086 4	0.064 2	0.087 4	0.099 7	0.029 9	0.065 4	0.079 0	0.0387
Reference	0.117 4	0.100 4	0.104 0	0.113 6	0.112 5	0.080 9	0.078 9	0.076 8	0.081 8
Science	0.119 2	0.105 7	0.097 3	0.101 1	0.010 5	0.003 5	0.005 1	0.014 6	0.001 6
Emotions	0.609 7	0.574 1	0.596 9	0.599 6	0.587 7	0.499 0	0.571 2	0.600 7	0.615 6
Computer	0.079 7	0.070 9	0.054 7	0.068 8	0.043 4	0.030 8	0.033 5	0.026 8	0.056 9
CAL500	0.063 1	0.057 7	0.056 5	0.057 6	0.055 6	0.065 2	0.063 1	0.057 0	0.056 9
Bibtex	0.084 3	0.019 5	0.005 0	0.031 4	0.019 2	0.014 9	0.021 8	0.008 9	0.034 5
Cs	0.030 8	0.008 9	0.002 1	0.016 4	0.038 1	0.008 8	0.005 6	0.004 5	0.004 5
Cooking	0.041 9	0.009 1	0.004 4	0.023 8	0.001 7	0.018 0	0.003 1	0.002 5	0.003 5

表7 不同算法的 $F1_{Mi}$ 均值对比
Table 7 Comparison mean results of different algorithms on $F1_{Mi}$

数据集	$F1_{Mi}$								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.550 6	0.350 8	0.388 3	0.545 5	0.475 0	0.326 7	0.468 3	0.410 3	0.413 0
Chemistry	0.055 3	0.025 7	0.020 7	0.041 4	0.019 0	0.025 0	0	0.011 0	0.025 4
Philosophy	0.172 3	0.090 0	0.040 7	0.166 0	0.145 6	0.046 3	0.006 6	0.011 9	0.004 0
Core15K	0.010 1	0	0	0.001 4	0.005 6	0.009 0	0.002 3	0.015 6	0.003 4
Yeast	0.645 5	0.569 3	0.602 3	0.607 9	0.629 4	0.591 1	0.620 8	0.621 5	0.609 4
Entertainment	0.283 8	0.215 8	0.121 7	0.250 5	0.201 4	0.114 2	0.097 8	0.156 7	0.062 3
Reference	0.393 0	0.336 1	0.345 8	0.376 9	0.297 6	0.251 6	0.325 1	0.280 6	0.335 8
Science	0.190 9	0.038 1	0.114 8	0.150 3	0.023 0	0.007 9	0.018 7	0.030 1	0.011 1
Emotions	0.640 4	0.604 9	0.629 1	0.632 3	0.624 5	0.532 9	0.618 6	0.614 1	0.636 5
Computer	0.337 4	0.339 7	0.388 0	0.415 8	0.378 7	0.373 8	0.380 8	0.367 4	0.389 5
CAL500	0.341 2	0.310 6	0.323 5	0.327 9	0.331 3	0.329 2	0.328 1	0.311 8	0.327 1
Bibtex	0.270 7	0.162 9	0.077 6	0.191 1	0.134 5	0.033 3	0.049 7	0.015 7	0.178 8
Cs	0.172 2	0.067 1	0.018 9	0.105 1	0.158 3	0.020 9	0.017 1	0.030 7	0.022 0
Cooking	0.177 5	0.083 5	0.037 6	0.154 7	0.022 4	0.051 2	0.009 5	0.016 0	0.012 0

由表 2~7 可知,本研究算法在各项评价指标上的表现总体优于其他比较算法。表 2 中,GRMfast 算法在 14 个数据集中的 11 个数据集上取得最佳结果,并且在所有数据集上都优于 MIFS。在数据集 Reference 和 Computer 上,GRMfast 算法性能仅次于 MICO。表 3 中,GRMfast 算法在 12 个数据集上取得最佳结果,在其他数据集上取得次优结果。表 4 中,除数据集 Birds、Yeast 和 CAL500,本文算法在其他数据集上都取得最佳或次优结果。表 5 中,本文算法在所有数据集上取得最佳结果。表 6 中,除数据集 Corel5K、Emotions、CAL500 和 Cs,GRMfast 算法在其他数据集上都取得最佳或次优结果。表 7 中,GRMfast 算法在 12 个数据集上取得最佳结果,在数据集 Corel5K 和 Computer 上性能效果欠佳。

使用 Friedman 检验^[33]进一步分析 GRMfast 与对比算法的分类性能差异。为拒绝原假设,Friedman 统计量 F_F 值需要大于或等于预先指定显著性水平相应临界表值,该值根据文献[34]中 F 分布表得到。本研究实验比较算法个数 $k=9$,数据集个数 $N=14$,查 F 分布表知在显著性水平 $\alpha=0.05$ 下临界值 $F(13,104)=1.8154$ 。每种指标 F_F 和相应临界表值如表 8 所示。由表 8 可知,每个指标以显

著性水平 $\alpha=0.05$ 时拒绝原假设,即所有算法性能不同。

表 8 每种指标的 Friedman 统计量 F_F 和相应临界表值
Table 8 Friedman statistics F_F and corresponding critical values for each indicator

评价指标	F_F	临界表值
L_H	7.253 0	
L_R	8.195 1	
P_C	4.373 7	
A_P	12.186 2	1.815 4
$F1_{Ma}$	7.571 9	
$F1_{Mi}$	8.178 9	

因此,通过 Bonferroni-Dunn 后续检验^[35]进一步区别各算法的分类性能。Bonferroni-Dunn 通过计算平均序值差别的临界值域 C_α ^[33]

$$C_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (17)$$

式中, q_α 为 Bonferroni-Dunn 检验的临界表值, α 为显著性水平。查表知显著性水平 $\alpha=0.05$ 、 $k=9$ 时对应的 $q_\alpha=2.724$,求得 $C_\alpha=2.8196$ 。GRMfast 算法和其他对比算法在不同评价指标上的 C_α 值如图 1 所示。在一个 C_α 内的算法用一条线连接起来,任何不连接的方法都被认为与 GRMfast 算法有显著不同的性能。

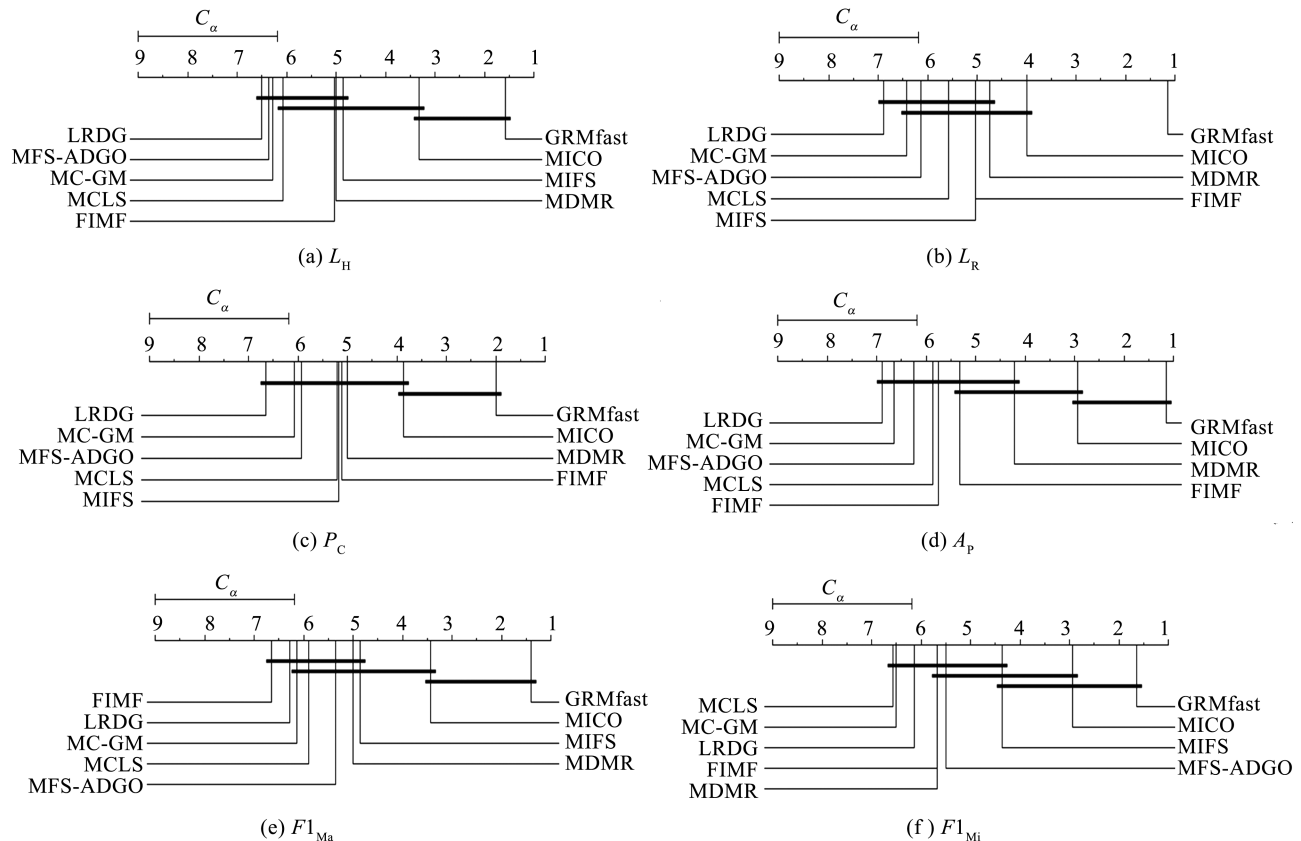


图 1 GRMfast 算法与其他 MLFS 算法的 Bonferroni-Dunn 检验结果 ($C_\alpha=2.8196$, $\alpha=0.05$)

Fig.1 Comparison of GRMfast against other MLFS algorithms with the Bonferroni-Dunn test($C_\alpha=2.8196$, $\alpha=0.05$)

由图1可知,GRMfast在9个算法中总体排第1位,GRMfast算法显著优于FIMF、MDMR、MCLS、MIFS、MC-GM、LRDG、MFS-ADGO。尽管GRMfast与MICO在各指标上的表现无显著差异,但

GRMfast与其他方法相比具有显著差异,存在明显优势。

本研究以Computer数据集为例展示特征选择数量对分类性能影响,试验结果如图2所示。

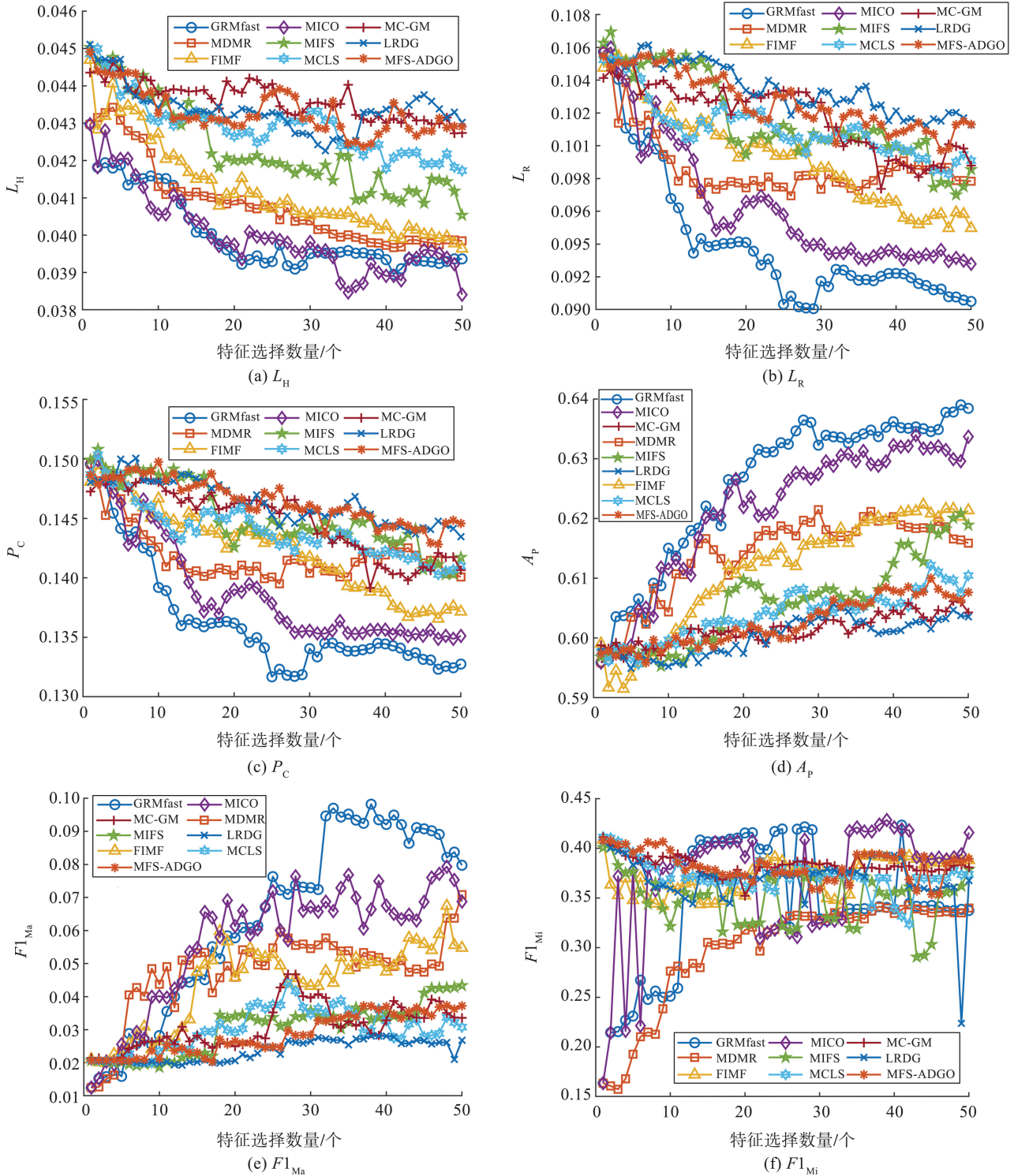


图2 Computer数据集上特征选择数量对各指标的影响
Fig.2 Influence of the number of selected feature on the Computer dataset on each indicator

由图2可知,GRMfast算法分类性能先随特征选择数量增加而提高,然后逐渐趋于稳定甚至下降。GRMfast在6个评价指标上的表现与对比算法

差异明显。整体而言,GRMfast算法在6个评价指标上相对对比算法能取得较优结果,表明GRMfast能从多标签数据集中选取有效特征从而提高分类

性能。

3.4 多标签特征选择算法运行时间比较

9 种算法运行时间如表 9 所示,其中“—”表示运行时间超过 1 000 s。

由表 9 可知,GRMfast 相较其他算法在 14 个数据集上运行时间最短。MDMR、FIMF、MICO 及 GRMfast 均为基于信息论的特征选择方法。MDMR、FIMF 基于启发式搜索策略,在评价函数中涉及大量无效或重复熵计算,面对多标签数据高维度特性,算法时间复杂度呈指数增长,特征

选择效率降低。MICO 采用优化策略进行特征选择,但其中涉及较复杂的矩阵运算,特征选择效率不及 GRMfast。基于稀疏学习的特征选择方法 MIFS、MC-GM 通过引入稀疏性约束减少特征选择数量,使得模型更简单,但数据稀疏性低时,特征选择效率比 GRMfast 低。基于流形学习的特征选择方法 MCLS、LRDG、MFS-ADGO 在处理高维数据时涉及构造邻域图和计算拉普拉斯矩阵需大量计算资源,导致特征选择效率比 GRMfast 低。

表 9 不同算法的运行时间

Table 9 Running time comparison of different algorithms

数据集	运行时间/s								
	GRM	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Birds	0.15	37.65	0.66	1.39	0.81	1.23	0.16	0.18	0.87
Chemistry	4.10	—	602.33	33.38	303.84	62.68	53.34	65.36	508.83
Philosophy	4.59	—	697.77	83.87	215.23	83.92	15.01	30.76	164.68
Corel5K	25.30	—	—	42.18	990.45	87.69	175.21	83.32	690.52
Yeast	0.09	27.21	0.25	0.95	5.22	4.01	1.65	2.08	15.95
Entertainment	0.61	344.17	4.32	18.41	13.59	4.48	8.79	9.18	70.12
Reference	0.57	671.31	13.71	56.24	19.78	4.79	8.67	15.36	74.03
Science	3.83	765.26	15.54	46.30	5.04	4.91	4.60	9.57	86.36
Emotions	0.08	1.21	0.09	0.33	0.26	0.17	0.17	0.18	1.02
Computer	3.58	—	5.22	22.66	15.30	3.99	4.45	19.09	46.68
CAL500	0.14	17.86	3.63	0.38	5.93	1.47	0.30	0.27	0.85
Bibtex	24.84	—	—	—	—	49.25	69.58	191.71	977.06
Cs	16.44	—	—	73.19	510.17	315.84	135.84	143.34	—
Cooking	13.41	—	—	104.57	—	476.52	223.53	201.67	—

采用 Friedman 检验进行显著性测试。假设所有算法的运行效率都相同,由表 9 求得运行时间 Friedman 统计量 $F_F = 27.099$ $7 > F(13, 104) = 1.8154$,以显著性水平 $\alpha = 0.05$ 拒绝原假设。由于

所有算法的运行效率不同,进行 Bonferroni-Dunn 后续检验进一步区分各算法运行效率差异,检验结果如图 3 所示。

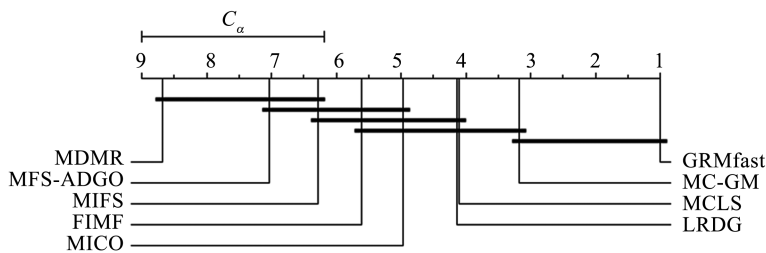


图 3 GRM 算法与其他 MLFS 算法的 Bonferroni-Dunn 检验结果($C_\alpha = 2.8196$, $\alpha = 0.05$)

Fig.3 Comparison of GRMfast against other MLFS algorithms with the Bonferroni-Dunn test($C_\alpha = 2.8196$, $\alpha = 0.05$)

由图 3 可知,GRMfast 运行时间和 MDMR、MFS-ADGO、MIFS、FIMF、MICO、LRDG、MCLS 7 个算法有显著差异,GRMfast 运行时间和 MC-GM 在一定误差范围内,在统计上无显著差异。总体来看,GRMfast 运行时间在 9 个算法中平均排名第一。

为展示 GRMfast 在极端场景下的表现和可扩展性,本研究从 Extreme Classification Repository (<http://manikvarma.org/downloads/XC/XMLRep->

ository.html) 选取 2 个大规模数据集 Delicious 和 Mediamill。Delicious 数据集包含网页文本数据,总共 16 105 个实例和 983 个类标签,每个实例由 500 个特征表示。Mediamill 数据集包含视频标注数据,总共 43 907 个实例和 101 个类标签,每个实例由 120 个特征表示。9 种算法在 Delicious、Mediamill 数据集上运行所需时间如表 10 所示,其中“—”表示运行时间超过 1 000 s。

表10 不同多标签特征选择算法在数据集 Delicious 和 Mediamill 上的运行时间
Table 10 Running timecomparison of different MLFS algorithms on the Delicious and Mediamill datasets

数据集	运行时间/s								
	GRMfast	MDMR	FIMF	MICO	MIFS	MCLS	MC-GM	LRDG	MFS-ADGO
Delicious	102.60	—	—	210.60	—	—	238.78	973.56	—
Mediamill	225.72	—	—	634.14	550.29	—	—	—	—

由表10可知,MDMR、FIMF 算法在处理大规模多标签数据时表现不佳;MIFS、MC-GM、LRDG 相比其他算法运行较快,但显然 GRMfast 的运行时间比它们更短;MICO 算法仅在 Delicious 数据集上运行时间较快,在 Mediamill 数据集上所需时间较长。

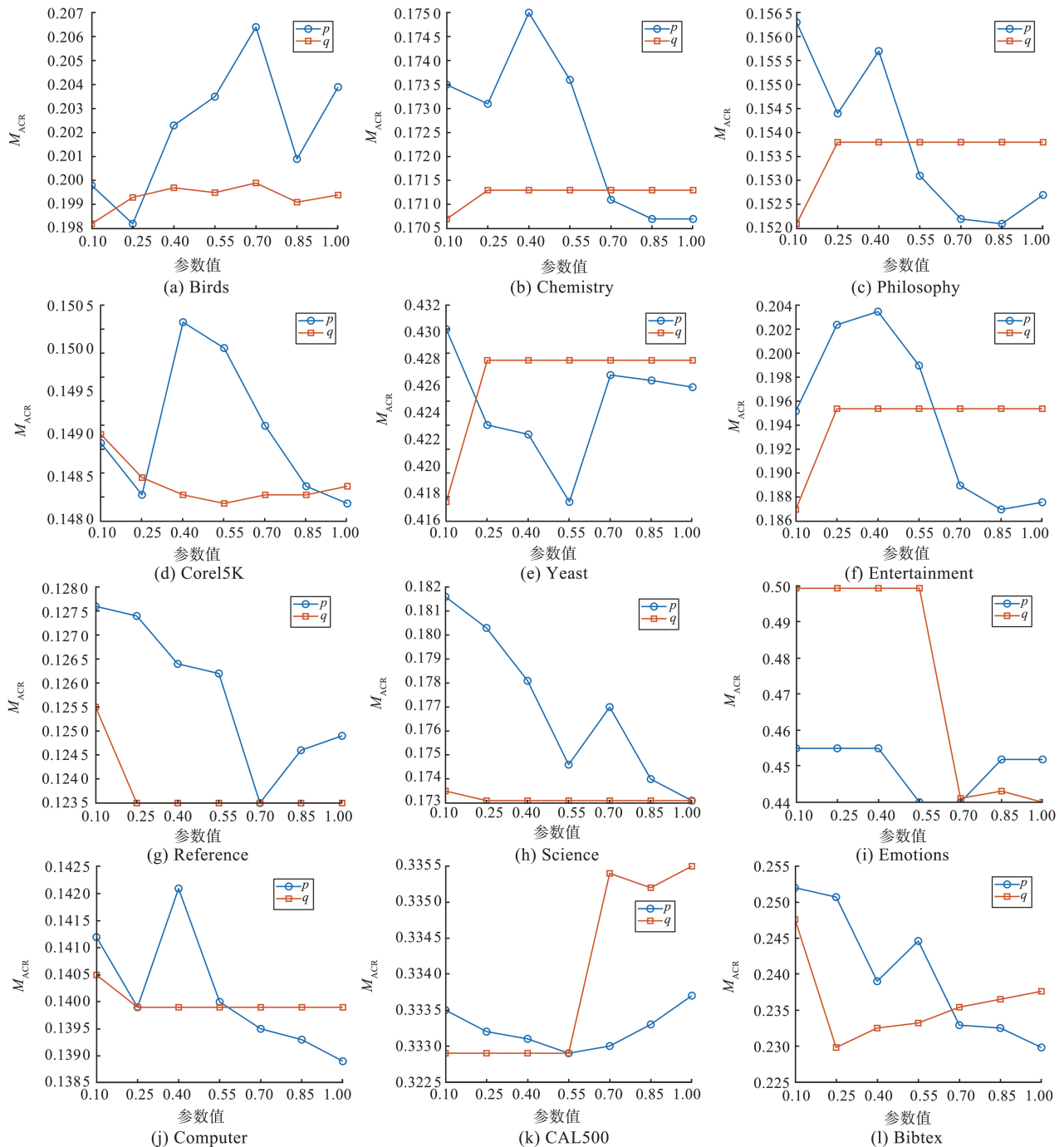
3.5 参数敏感性分析

本节分别讨论 GRMfast 中候选标签和候选

特征比例及集成模型数量对算法分类性能影响。调节其中 1 个参数时,其他参数保持最佳设置。

3.5.1 标签选择比例和特征选择比例对算法分类性能的影响

不同标签选择比例 p 和特征选择比例 q 组合所得 M_{ACR} 结果如图 4 所示。



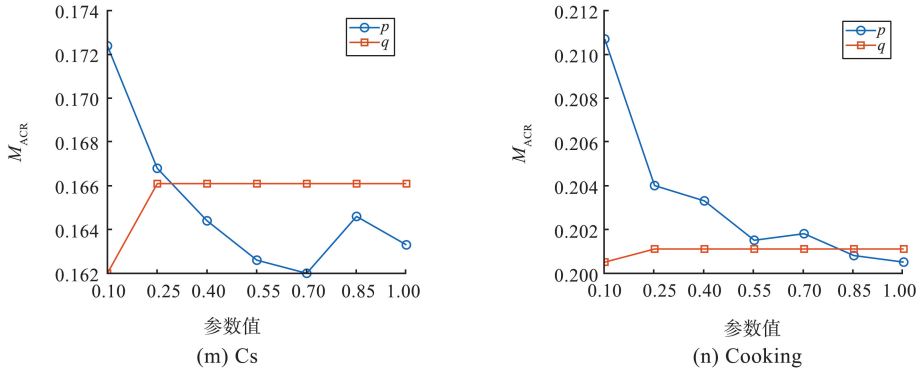


图 4 GRMfast 在所有数据集上的参数分析结果
Fig.4 Parameter analysis of GRMfast on all datasets

从图 4 可知,当改变参数 p 或 q 时,ACR 值变化显著,说明 GRMfast 对这 2 个参数敏感。由 2.1 节式(2)得出不同数据集的最优标签选择比例和特征选择比例,如表 11 所示。

表 11 不同数据集的最优标签选择比例 p 和特征选择比例 q 设置

Table 11 Best parameters setting for selecting labels and features for different datasets

数据集	p	q	数据集	p	q
Birds	0.25	0.10	Science	1.00	0.25
Chemistry	0.85	0.10	Emotions	0.55	1.00
Philosophy	0.85	0.10	Computer	1.00	0.25
Corel5K	1.00	0.55	CAL500	0.55	0.55
Yeast	0.55	0.10	Bibtex	1.00	0.25
Entertainment	0.85	0.10	Cs	0.70	0.10
Reference	0.70	0.25	Cooking	1.00	0.10

表 12 不同集成模型数量在 GRMfast 上的结果比较

Table 12 Comparison results on GRMfast for different numbers of ensembles

数据集	E	L_H	L_R	P_C	A_p	$F1_{Ma}$	$F1_{Mi}$	时间/s
Chemistry	1	0.012 1	0.156 7	0.255 3	0.289 6	0.013 2	0.051 2	3.220
	10	0.012 1	0.149 0	0.247 9	0.311 4	0.010 6	0.053 2	3.940
	30	0.012 2	0.164 5	0.267 5	0.257 7	0.001 6	0.008 2	4.010
	50	0.012 2	0.149 9	0.246 4	0.294 2	0.013 5	0.040 3	4.080
	70	0.012 1	0.148 6	0.246 8	0.296 7	0.011 8	0.055 3	4.100
	90	0.012 2	0.153 1	0.253 2	0.289 6	0.008 2	0.039 6	5.720
Yeast	1	0.201 5	0.179 4	0.469 4	0.747 5	0.337 4	0.628 9	0.072
	10	0.201 4	0.184 0	0.473 6	0.744 0	0.337 0	0.630 8	0.074
	30	0.203 2	0.183 3	0.468 5	0.744 6	0.336 5	0.631 2	0.075
	50	0.203 3	0.176 7	0.464 1	0.752 1	0.331 7	0.625 8	0.078
	70	0.199 4	0.175 6	0.462 6	0.758 8	0.368 0	0.645 5	0.088
	90	0.200 9	0.179 3	0.465 3	0.748 2	0.331 0	0.615 4	0.110
Philosophy	1	0.009 1	0.137 3	0.244 8	0.318 2	0.025 0	0.164 6	3.950
	10	0.009 0	0.136 8	0.243 3	0.339 9	0.024 7	0.188 7	4.250
	30	0.009 2	0.135 7	0.242 3	0.324 4	0.021 5	0.153 2	4.340
	50	0.009 1	0.135 9	0.243 2	0.334 7	0.027 4	0.201 6	4.380
	70	0.009 0	0.135 6	0.243 2	0.335 8	0.024 6	0.172 3	4.590
	90	0.009 1	0.145 7	0.255 1	0.287 0	0.016 3	0.148 4	4.700

由表 11 可知,多数数据集从原始标签集中选取 p 比例标签提升学习性能。在一些数据集比如 Corel5K、Reference、Science 上 q 取值较大,从而获得更好学习性能;在其他数据集上, q 取值为 0.1,说明只要从原始特征空间中提取少量比例特征即可获得良好分类性能。总体而言,GRMfast 可以通过从原始标签集和特征集中选取一定比例的标签和特征从而减少无效标签和特征之间的计算。

3.5.2 集成模型数量对算法性能影响

GRMfast 采用集成学习以补偿候选标签可能无法完全覆盖整个原始标签空间所带来的性能下降。为展示不同集成模型数量 E 对算法性能影响,使用 Chemistry、Yeast、Philosophy 数据集进行实验,结果如表 12 所示。加粗字体表示该指标在对应集成数取得最优值。

由表12可知,一定数量的集成模型能提升算法分类性能,且特征选择所需时间随集成模型数量增加而增加。如果集成模型数量不足,模型可能没有足够学习能力捕捉数据中复杂模式和关系,导致模型过于简单,无法很好地泛化到新数据上;集成模型数量增至70时,算法较其他集成次数取得较好结果;当集成模型数量过多时,模型可能会过度拟合

训练数据,导致算法分类性能下降。

3.6 收敛性分析

目标函数在 Birds、Emotions、Science 和 Bibtex4 个数据集上迭代次数如图5(a)~(d)所示。由图5可知,GRMfast 在4个数据集上可以在4次迭代内快速收敛,第5次迭代时达到稳定水平,说明 GRMfast 面对不同规模的数据集能够快速收敛。

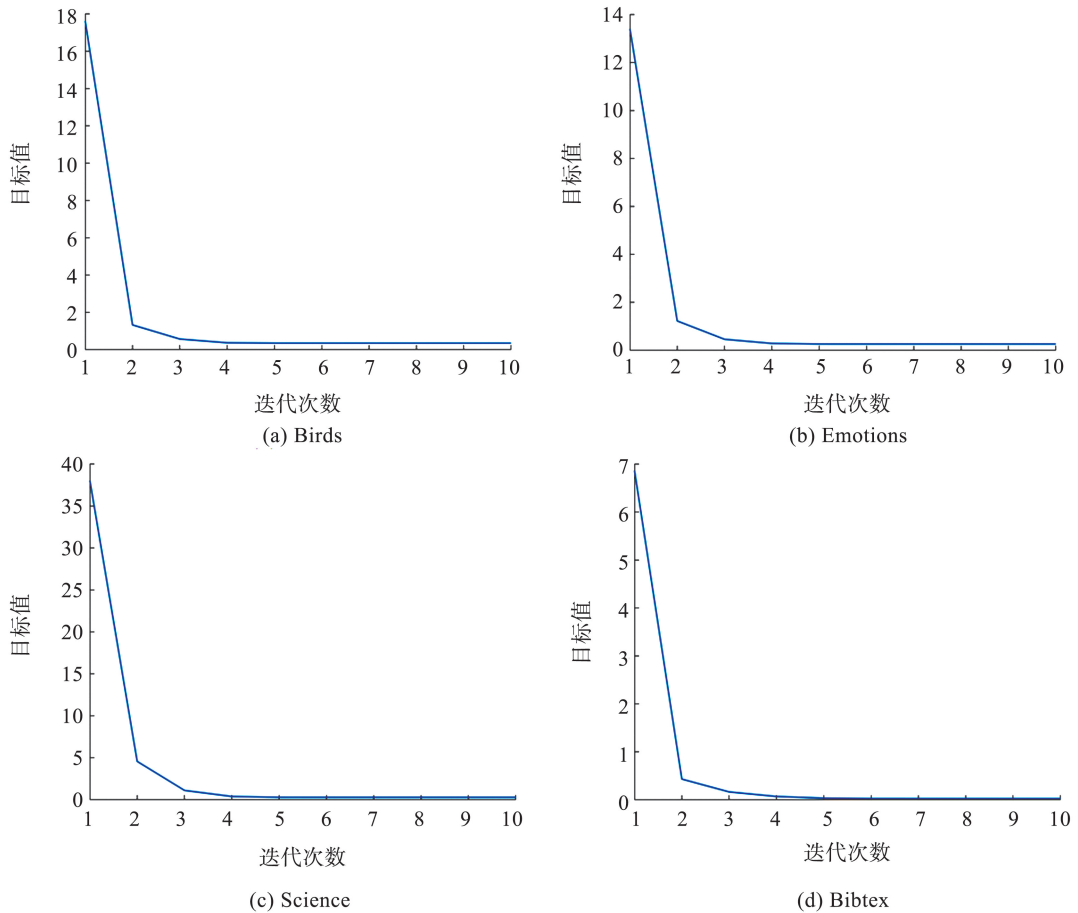


图5 GRMfast 的收敛曲线图
Fig.5 Convergence graph of GRMfast

4 结论

本研究提出一种基于全局冗余最小的快速多标签特征选择方法,与其他8种算法在14个数据集上进行试验对比。试验结果表明,本研究算法在各项评价指标上的表现总体优于其他比较算法,所提方法在兼顾分类性能的同时,执行效率较高。

在未来工作中,可进一步研究标签选择和标签嵌入在快速多标签特征选择中的应用以及深度学习在表征学习的应用,进而设计能更好兼顾模型性能与计算成本的特征选择方法。

参考文献:

- [1] ZHANG M L, ZHOU Z H. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [2] 马坤, 刘筱云, 李乐平, 等. 用于意图识别的自适应多标签信息学习模型[J]. 山东大学学报(工学版), 2024, 54(1): 45-51.
MA Kun, LIU Xiaoyun, LI Leping, et al. Adaptive label information learning for intention detection [J]. Journal of Shandong University (Engineering Science), 2024, 54(1): 45-51.
- [3] 李云, 卢志翔, 刘姝伊, 等. 基于深度多模态关联学习的短视频多标签分类研究[J]. 数据分析与知识发现, 2024, 8(7): 77-88.

- LI Yun, LU Zhixiang, LIU Shuyi, et al. Research on micro-video multi-label classification based on deep multimodal association learning [J]. *Data Analysis and Knowledge Discovery*, 2024, 8(7): 77-88.
- [4] 张建贺, 姜晓燕. 结合双路网络和多标签分类的弱监督行人搜索 [J]. *计算机工程与应用*, 2023, 59(9): 159-166.
- ZHANG Jianhe, JIANG Xiaoyan. Weakly supervised person search combining dual-path network and multi-label classification [J]. *Computer Engineering and Applications*, 2023, 59(9): 159-166.
- [5] 周慧颖, 汪廷华, 张代俐. 多标签特征选择研究进展 [J]. *计算机工程与应用*, 2022, 58(15): 52-67.
- ZHOU Huiying, WANG Tinghua, ZHANG Daili. Research progress of multi-label feature selection [J]. *Computer Engineering and Applications*, 2022, 58(15): 52-67.
- [6] 李永豪, 胡亮, 高万夫. 基于稀疏系数矩阵重构的多标记特征选择 [J]. *计算机学报*, 2022, 45(9): 1827-1841.
- LI Yonghao, HU Liang, GAO Wanfu. Multi-label feature selection based on sparse coefficient matrix reconstruction [J]. *Chinese Journal of Computers*, 2022, 45(9): 1827-1841.
- [7] 胡军, 王海峰. 基于加权信息粒化的多标记数据特征选择算法 [J]. *智能系统学报*, 2023, 18(3): 619-628.
- HU Jun, WANG Haifeng. Feature selection algorithm of multi-labeled data based on weighted information granulation [J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(3): 619-628.
- [8] LEE J, KIM D W. Feature selection for multi-label classification using multivariate mutual information [J]. *Pattern Recognition Letters*, 2013, 34(3): 349-357.
- [9] LIN Y J, HU Q H, LIU J H, et al. Multi-label feature selection based on max-dependency and min-redundancy [J]. *Neurocomputing*, 2015, 168: 92-103.
- [10] LEE J, KIM D W. Fast multi-label feature selection based on information-theoretic feature ranking [J]. *Pattern Recognition*, 2015, 48(9): 2761-2771.
- [11] SUN Z Q, ZHANG J, DAI L, et al. Mutual information based multi-label feature selection via constrained convex optimization [J]. *Neurocomputing*, 2019, 329: 447-456.
- [12] ZHANG J, LIN Y D, JIANG M, et al. Multi-label feature selection via global relevance and redundancy optimization [C]//*Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. Yokohama, Japan: IEEE, 2020: 2512-2518.
- [13] DAI J H, HUANG W Y, ZHANG C C, et al. Multi-label feature selection by strongly relevant label gain and label mutual aid [J]. *Pattern Recognition*, 2024, 145: 109945.
- [14] BUGATA P, DROTAR P. On some aspects of minimum redundancy maximum relevance feature selection [J]. *Science China Information Sciences*, 2020, 63(1): 112103.
- [15] HASHEMI A, DOWLATSHAHI M B, NAZAMABADI-POUR H. Minimum redundancy maximum relevance ensemble feature selection: a bi-objective Pareto-based approach [J]. *Journal of Soft Computing and Information Technology*, 2023, 12(1): 20-28.
- [16] ZHOU H F, WANG X Q, ZHU R R. Feature selection based on mutual information with correlation coefficient [J]. *Applied Intelligence*, 2022, 52(5): 5457-5474.
- [17] WANG X J, ZHOU Y C. Multi-label feature selection with conditional mutual information [J]. *Computational Intelligence and Neuroscience*, 2022(1): 9243893.
- [18] ZHANG P, LIU G X, SONG J Z. MFSJMI: multi-label feature selection considering join mutual information and interaction weight [J]. *Pattern Recognition*, 2023, 138: 109378.
- [19] 张俐, 王枫. 基于最大相关最小冗余联合互信息的多标签特征选择算法 [J]. *通信学报*, 2018, 39(5): 111-122.
- ZHANG Li, WANG Cong. Multi-label feature selection algorithm based on joint mutual information of max-relevance and min-redundancy [J]. *Journal on Communications*, 2018, 39(5): 111-122.
- [20] RAKESH D K, JANA P K. A general framework for class label specific mutual information feature selection method [J]. *IEEE Transactions on Information Theory*, 2022, 68(12): 7996-8014.
- [21] JIAN L, LI J, SHU K, et al. Multi-label informed feature selection [C]//*Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. New York, USA: AAAI, 2016: 1627-1633.
- [22] PENG H C, LONG F H, DING C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1226-1238.
- [23] SHIRZAD M B, KEYVANPOUR M R. A feature selection method based on minimum redundancy maximum relevance for learning to rank [C]//*2015 AI & Robotics (IRANOPEN)*. Qazvin, Iran: IEEE, 2015: 1-5.
- [24] AGHAEIPOOR F, JAVIDI M M. A hybrid fuzzy feature selection algorithm for high-dimensional regression problems: an mRMR-based framework [J]. *Expert Systems with Applications*, 2020, 162: 113859.

- [25] 徐洪峰, 孙振强. 多标签学习中基于互信息的快速特征选择方法[J]. 计算机应用, 2019, 39(10): 2815-2821.
XU Hongfeng, SUN Zhenqiang. Fast feature selection method based on mutual information in multi-label learning[J]. Journal of Computer Applications, 2019, 39(10): 2815-2821.
- [26] BERTSEKAS D P. Constrained optimization and lagrange multiplier methods[M]. Amsterdam: Elsevier, 1982.
- [27] WU X Z, ZHOU Z H. A unified view of multi-label performance measures [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: ACM, 2017: 3780-3788.
- [28] MA J H, CHIU B C Y, CHOW T W S. Multilabel classification with group-based mapping: a framework with local feature selection and local label correlation[J]. IEEE Transactions on Cybernetics, 2022, 52(6): 4596-4610.
- [29] HUANG R, JIANG W D, SUN G L. Manifold-based constraint Laplacian score for multi-label feature selection [J]. Pattern Recognition Letters, 2018, 112: 346-352.
- [30] ZHANG Y, HUO W, TANG J. Multi-label feature selection via latent representation learning and dynamic graph constraints [J]. Pattern Recognition, 2024, 151: 110411.
- [31] SUN Z Z, XIE H, LIU J H, et al. Multi-label feature selection via adaptive dual-graph optimization[J]. Expert Systems with Applications, 2024, 243: 122884.
- [32] ZHANG M L, ZHOU Z H. ML-KNN: a lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [33] FRIEDMAN M. A comparison of alternative tests of significance for the problem of m rankings [J]. The Annals of Mathematical Statistics, 1940, 11(1): 86-92.
- [34] SHESKIN D J. Handbook of parametric and nonparametric statistical procedures, fifth edition [M]. New York: Chapman and Hall/CRC, 2020.
- [35] DEMŠAR J. Statistical comparisons of classifiers over multiple data sets[J]. The Journal of Machine Learning Research, 2006, 7: 1-30.

(编辑:管昕昉)

(上接第20页)

- [33] BACHTIAR F A, ARWANI I, PRASETYO R A B, et al. Distraction detection in driving using pose extraction and machine learning techniques [C]//Proceedings of the 2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Purwokerto, Indonesia: IEEE, 2023: 279-284.
- [34] PUTRA N C B, YUNIARNO E M, RACHMADI R F. Driver visual distraction detection based on face mesh feature using deep learning [C]// 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA). Surabaya, Indonesia: IEEE, 2023: 6-11.
- [35] CAO Y, XU J R, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops. Seoul: IEEE, 2019: 1971-1980.
- [36] LI X, WANG W H, WU L J, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: ACM, 2020: 21002-21012.
- [37] YANG Z, LIU S H, HU H, et al. RepPoints: point set representation for object detection [C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9656-9665.
- [38] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2999-3007.

(编辑:管昕昉)