

具有标签流形和动态图约束的多标签特征选择

武晓军¹, 陈怡丹², 郝耀军¹, 宋长伟³, 何德清⁴

(1.忻州师范学院, 山西 忻州 034000; 2.河南开放大学, 河南 郑州 450046; 3.河南农业大学信息与管理科学学院, 河南 郑州 450003; 4.华中科技大学计算机科学与技术学院, 湖北 武汉 430074)

摘要:将自适应动态图技术和标签流形集成到改进后的线性映射学习框架中,提出了具有标签流形和动态图约束的多标签特征选择算法。该算法基于特征自表示的改进矩阵分解技术,改进了线性映射模型,对特征和标签之间以及不同标签之间的相关性进行解耦。设计了一种具有拉普拉斯秩约束的自适应动态图技术,学习高质量的特征相似图。构建了基于标签相关性的标签流形,将标签信息充分的纳入模型的训练中。验证了自适应动态图技术可以有效的提高图矩阵的质量,以及所提算法在解决多标签特征选择问题上的有效性。

关键词:多标签学习;特征选择;流形学习;自适应学习;动态图学习

中图分类号:TP181 **文献标志码:**A

引用格式:武晓军,陈怡丹,郝耀军,等.具有标签流形和动态图约束的多标签特征选择[J].山东大学学报(理学版),2025,60(7):69-83.

Multi-label feature selection with label manifold and dynamic graph constraints

WU Xiaojun¹, CHEN Yidan², HAO Yaojun¹, SONG Changwei³, HE Deqing⁴

(1. Xinzhou Normal University, Xinzhou 034000, Shanxi, China; 2. Henan Open University, Zhengzhou 450046, Henan, China; 3. College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, Henan, China; 4. School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China)

Abstract: Multi-label feature selection algorithms with label manifolds and dynamic graph constraints are proposed by integrating adaptive dynamic graph technique and label manifold into an improved linear mapping learning framework. In this algorithm, an improved matrix decomposition technique based on feature self-representation improves the linear mapping model and decouples the correlation between features and labels as well as between different labels. An adaptive dynamic graph technique with Laplace rank constraints is designed to learn a high-quality feature similarity graph. A label manifold based on label relevance is constructed to fully incorporate label information into the training of the algorithm. Numerous experimental results verify that the adaptive dynamic graph technique can effectively improve the quality of the graph matrix and the effectiveness of the proposed algorithm in addressing the multi-label feature selection problem.

Key words: multi-label learning; feature selection; manifold learning; adaptive learning; dynamic graph learning

0 引言

多标签分类作为机器学习中的研究热点广泛应用于文本分类^[1-2]、生物医学^[3]和化学分析^[4]等领域。不同于单标签分类,多标签分类任务还须关注标签之间的复杂性和相关性。然而由于数据融合技术和物联网技术的高速发展,多标签数据的特征维数呈现指数增长趋势,大量冗余和不相关特征严重影响分类器的性能,增加分类器的计算负担^[5-7]。多标签特征选择作为维度简约的重要技术之一,目的是选择一个有区别的、信息性的特征子集,提高算法性能,降低计算成本^[8-9]。目前,多标签特征选择算法主要有过滤式、封装式和嵌入式。因为嵌入式算法在保证精度的同时降低了算法的时间复杂度,所以主要关注对嵌入式多标签

收稿日期:2024-05-14; 网络出版时间:2024-12-20 08:26:34

基金项目:国家自然科学基金资助项目(61902139); 山西省基础研究计划资助项目(202203021211116); 忻州师范学院基金资助项目(2021KY16)

第一作者:武晓军(1988—),男,讲师,硕士,研究方向为智能计算、图像处理、机器学习. E-mail:xiaojunwu1988@163.com

特征选择算法的研究。

许多特征选择算法中都引入基于固定图的流形学习^[10-11]指导特征选择^[12],然而固定的图矩阵导致模型的计算误差得不到修正而累加^[13-15]。此外,一些算法忽略了标签相关性,导致算法的性能不佳^[16-17]。

针对上述问题,本文提出了具有标签流形和动态图约束的多标签特征选择(multi-label feature selection with label manifold and dynamic graph constraints, LMDG)算法。该算法改进了传统的线性映射模型,对特征与标签以及不同标签之间的相关性进行解耦,利用基于图拉普拉斯秩约束的动态图学习技术纠正算法的计算误差,最后引入了标签流形,将标签相关性信息纳入特征选择过程。

1 相关工作

在多标签学习中,已有许多成熟的特征选择算法处理各领域的的数据。由文献[18]可知,现有算法通过3种策略挖掘标签矩阵中的监督信息:一阶策略,如二元关联^[19]对多标签数据的每个标签使用传统的单标签算法训练模型;二阶策略,如Huang等^[20]利用标签间的成对相关性挖掘每个标签的标签特定特征信息,提出了学习标签特定特征的多标签分类算法;高阶策略,许多基于图的特征选择算法通过引入流形学习挖掘标签间的高阶相关性信息^[21-23]。

在这些基于图的算法中,通常使用单一的图学习方式计算图矩阵,如余弦定理、布尔值、高斯热核函数和泊松相关系数。例如,Cai等^[24]在线性映射的基础上,结合特征流形和 $L_{2,1}$ 范数共同约束特征权重矩阵。Zhang等^[21]引入伪标签学习技术,通过实例流形和低维标签流形指导伪标签的学习。Gao等^[25]利用潜在表征学习设计了一种实例和标签共享结构的伪标签学习技术,引入实例流形约束伪标签。Hu等^[22]引入特征流形、低维标签流形和 $L_{2,1}$ 范数稀疏约束,对特征权重矩阵进行全面约束。Li等^[26]结合标签自表示学习和实例流形综合考虑标签的高阶相关性和低阶成对相关性,并使用 $L_{2,1}$ 范数约束自表示系数消除冗余信息和噪声。Li等^[23]将文献^[27]中的内积正则化作用于特征权重,设计了一种消除特征冗余信息的惩罚项,并结合标签低维流形,提出了鲁棒柔性稀疏正则化的多标签特征选择。

然而上述算法在引入流形学习时,使用的图矩阵是固定的,导致模型的计算误差因得不到修正而累加,从而限制算法的泛化性能^[15]。为了缓解这一问题,Li等^[15]设计了基于低维伪标签相似性的动态图矩阵,用于校正算法的计算误差。Zhang等^[16]在使用低维伪标签动态图的基础上,分别使用标签流形和特征流形约束伪标签和特征权重。Hu等^[14]将低维伪标签流形和 L_1 范数结合,共同约束特征权重矩阵,并利用实例流形约束伪标签矩阵。不同于上述算法,Zhang等^[13]以传统的特征图为基础,通过依图学图的方式,设计了一种新颖的动态图学习技术,然而由于缺乏对动态图的合理约束,导致模型的性能并不理想。Li等^[17]利用矩阵的非负分解技术,设计了一种新颖的动态图学习技术,并使用低冗余的稀疏约束指导特征选择。为了更准确的构建基于伪标签的动态图,Zhang等^[28]通过引入基于非负矩阵分解的潜在表征学习挖掘标签相关性的深层信息,用于约束伪标签。

上述算法中使用的传统线性映射技术无法考虑不同输出之间存在的耦合关系,使用的动态图学习技术缺乏对动态图的合理约束,文献[24-25]均忽略了标签相关性信息对算法性能的影响。针对这些问题,本文设计了一种基于特征自表示的改进矩阵分解技术,改进线性映射模型,并且设计了一种新颖的具有拉普拉斯秩约束的自适应动态图学习技术,用于修正模型的计算误差,引入了基于标签相关性的流形学习,提出了一种新颖的特征选择算法。

还有一些基于信息论的多标签特征选择算法。例如,Lin等^[29]考虑了特征依赖性和特征冗余性,提出了基于最大依赖性和最小冗余度的多标签特征选择。Lee等^[30]通过互信息量化特征和特征之间以及特征和标签(标签对)之间的信息量,提出了基于可扩展准则的大标签集多标签特征选择。然而特征和标签的组合可能呈指数级增长,导致特征选择不切实际。Hashemi等^[31]将特征选择过程建模为多标准决策过程。Zou等^[32]利用互信息量化特征-标签-标签和特征-特征-标签之间的交互作用描述另一变量背景下的波动相关性,实现了考虑3方变量交互作用的多标签特征选择。

2 模型建立

2.1 问题重述

假设 $D = \{X, Y\}$ 为任意多标签数据集, $X \in \mathbf{R}^{n \times d}$ 是实例矩阵, $Y \in \mathbf{R}^{n \times m}$ 是标签矩阵, n, d 和 m 分别表示实例、特征和标签的个数。若第 i 个实例属于第 j 类标签, 则 Y 的第 i 行、 j 列元素 $Y_{ij} = 1$; 反之, $Y_{ij} = 0$ 。多标签特征选择的目的是利用 $D = \{X, Y\}$ 构建一个特征选择函数 $F: f(D) \rightarrow F_S$, 从原始的 d 维特征中选择出最具代表性的 c 个特征构建特征子集 F_S 。

2.2 线性映射框架

由于传统的线性映射模型计算简单, 且具有较强的可解释性, 因此, 许多算法是基于传统的线性映射模型构建的。在本文中, 传统的线性映射模型定义为

$$F(\mathbf{W}, \mathbf{b}) = \min_{\mathbf{W}, \mathbf{b}} \|\mathbf{XW} + \mathbf{eb} - \mathbf{Y}\|_F^2, \quad (1)$$

式中: $F(\mathbf{W}, \mathbf{b})$ 是关于变量 \mathbf{W} 和 \mathbf{b} 的目标函数, \mathbf{e} 表示元素全为 1 的 n 维列向量, \mathbf{b} 是 m 维行向量, 表示线性映射的偏置向量, $\mathbf{W} \in \mathbf{R}^{d \times m}$ 是特征权重矩阵, 用于量化各特征的重要性。

然而, 式(1)仅是对单标签模型的简单扩展, 忽略了模型中不同标签之间的耦合关系。单一的特征权重矩阵 \mathbf{W} 无法同时准确的整合特征和标签之间, 以及不同标签之间的相关性。受文献[33]的启发, 设计了一个基于特征自表示的改进矩阵分解范式

$$\mathbf{X}_{*j} \approx \sum_{i=1}^d \mathbf{X}_{*i} U_{ij} = \sum_{i=1}^d \mathbf{X}_{*i} \sum_{l=1}^k A_{il} B_{lj}, \quad j=1, 2, \dots, m, \quad (2)$$

式中: \mathbf{X}_{*i} 和 \mathbf{X}_{*j} 分别是 \mathbf{X} 的第 i 和 j 行列向量, 表示实例的第 i 和 j 个特征。 $U_{ij} = \sum_{l=1}^k A_{il} B_{lj}$ 表示特征 \mathbf{X}_{*i} 和 \mathbf{X}_{*j} 之间的权值, 用于比较不同特征之间的重要性。 $\mathbf{A} \in \mathbf{R}^{d \times k}$ 是原始特征经线性变换后的 k 维潜在在变量空间, $\mathbf{B} \in \mathbf{R}^{k \times m}$ 表示 \mathbf{A} 与标签空间之间的亲和性。式(2)通过 \mathbf{A} 和 \mathbf{B} 对特征与标签以及不同标签之间的相关性进行解耦。

结合式(1)、(2), 将基于特征自表示的改进矩阵分解技术构建的线性映射模型定义为

$$F(\mathbf{b}, \mathbf{A}, \mathbf{B}) = \min_{\mathbf{A}, \mathbf{B}, \mathbf{b}} \|\mathbf{XAB} + \mathbf{eb} - \mathbf{Y}\|_F^2. \quad (3)$$

2.3 自适应动态图学习

图矩阵的学习常用于约束模型中伪标签矩阵或特征权重矩阵, 直接或间接的影响模型的性能。换言之, 一个高质量的动态图可以通过修正模型的计算误差达到提高模型性能的目的。然而这些作用是现有固定图和低质量动态图所无法提供的。

受文献[13]的启发, 通过依图学图的方式学习动态图。然而文献[13]缺乏对动态图的合理约束, 导致动态图的质量并不理想。为此, 本文引入图拉普拉斯秩约束对文献[13]中的动态图学习技术进行改进, 并称之为自适应动态图学习技术, 定义为

$$\min_{\mathbf{S}} \|\mathbf{S} - \mathbf{Z}\|_1, \quad \text{s.t.} \quad 0 \leq S_{ij}, \quad \sum_j S_{ij} = 1, \quad S_{ii} = 0, \quad \text{rank}(\mathbf{L}_S) = d - c, \quad (4)$$

式中: $\mathbf{S} \in \mathbf{R}^{d \times d}$ 是动态图矩阵, S_{ij} 是 \mathbf{S} 的第 i 行 j 列元素, 表示特征 \mathbf{X}_{*i} 和 \mathbf{X}_{*j} 的相似性, $\mathbf{L}_S = \mathbf{P} - \frac{\mathbf{S}^T + \mathbf{S}}{2}$ 是关于 \mathbf{S} 的拉普拉斯矩阵, $\mathbf{P} \in \mathbf{R}^{d \times d}$ 是对角矩阵, $P_{ii} = \sum_j S_{ij}$; $\text{rank}(\mathbf{L}_S) = d - c$ 表示约束 \mathbf{L}_S 的秩为 $d - c$ 。 \mathbf{Z} 是传统的固定图矩阵, 即

$$\min_{\mathbf{Z}} \|\mathbf{X}_{*i} - \mathbf{X}_{*j}\|_2^2 Z_{ij} + \mu \sum_j Z_{ij}^2, \quad \text{s.t.} \quad 0 \leq Z_{ij}, \quad \sum_j Z_{ij} = 1, \quad Z_{ii} = 0, \quad (5)$$

式中: Z_{ij} 是 \mathbf{Z} 的第 i 行、 j 列元素, μ 是平衡参数。

2.4 标签流形

由式(2)可知, \mathbf{B} 中应包含尽可能多的标签相关性信息。考虑到流形学习的约束 2 变量间流形结构一致性, 实现信息传递。具体而言, 若标签 \mathbf{Y}_{*i} 和 \mathbf{Y}_{*j} 较为相似, 则 \mathbf{B}_{*i} 和 \mathbf{B}_{*j} 也较为相似, 从而

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \| \mathbf{B}_{*i} - \mathbf{B}_{*j} \|^2 (\mathbf{Z}_Y)_{ij} = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\mathbf{B}_{*i} - \mathbf{B}_{*j}) (\mathbf{B}_{*i} - \mathbf{B}_{*j})^T (\mathbf{Z}_Y)_{ij} \\ & = \sum_{j=1}^m \mathbf{B}_{*j}^T \mathbf{B}_{*j} \mathbf{Q}_{ii} - \sum_{i=1}^m \sum_{j=1}^m \mathbf{B}_{*j}^T \mathbf{B}_{*i} (\mathbf{Z}_Y)_{ij} = \text{Tr}(\mathbf{B}(\mathbf{Q} - \mathbf{Z}_Y)\mathbf{B}^T) = \text{Tr}(\mathbf{B}\mathbf{L}_Y\mathbf{B}^T), \end{aligned} \quad (6)$$

式中: $\text{Tr}(\cdot)$ 表示矩阵的迹, \mathbf{B}_{*i} 和 \mathbf{B}_{*j} 分别表示 \mathbf{B} 的第 i 列和第 j 列向量, $\mathbf{L}_Y = \mathbf{Q} - \mathbf{Z}_Y$ 是关于 \mathbf{Z}_Y 的拉普拉斯矩阵, $\mathbf{Q} \in \mathbf{R}^{m \times m}$ 是对角矩阵, $Q_{ii} = \sum_j (\mathbf{Z}_Y)_{ij}$; $\mathbf{Z}_Y \in \mathbf{R}^{m \times m}$ 是标签相似矩阵, 由高斯热核函数计算得到。

2.5 目标函数

结合式(3)、(4)、(6), 得到所提算法的初步目标函数

$$\begin{aligned} F(\mathbf{b}, \mathbf{A}, \mathbf{B}, \mathbf{S}) &= \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{b}} \| \mathbf{XAB} + \mathbf{eb} - \mathbf{Y} \|^2_F + \alpha \| \mathbf{S} - \mathbf{Z} \|_1 + \gamma \text{Tr}(\mathbf{B}\mathbf{L}_Y\mathbf{B}^T), \\ \text{s.t. } & 0 \leq S_{ij}, \quad \sum_j S_{ij} = 1, \quad S_{ii} = 0, \quad \text{rank}(\mathbf{L}_S) = d - c, \end{aligned} \quad (7)$$

式中 α 和 γ 是正则化参数。

式(7)中要保证 \mathbf{S} 的非负性, 则对变量 \mathbf{A} 和 \mathbf{B} 施加非负约束。此外, 拉普拉斯秩约束问题的求解过于复杂且无法直接求解, 根据 Ky Fan 定理^[34], 将秩约束问题转化为 $\text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{L}_S \mathbf{A} \mathbf{B})$ 纳入目标函数, 直接求解。因此, 将 LMDG 的最终目标函数表述为

$$\begin{aligned} F(\mathbf{b}, \mathbf{A}, \mathbf{B}, \mathbf{S}) &= \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{b}} \| \mathbf{XAB} + \mathbf{eb} - \mathbf{Y} \|^2_F + \alpha \| \mathbf{S} - \mathbf{Z} \|_1 + \beta \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{L}_S \mathbf{A} \mathbf{B}) + \gamma \text{Tr}(\mathbf{B}\mathbf{L}_Y\mathbf{B}^T), \\ \text{s.t. } & 0 \leq S_{ij}, \quad \sum_j S_{ij} = 1, \quad S_{ii} = 0, \quad \{\mathbf{A}, \mathbf{B}\} \geq 0, \end{aligned} \quad (8)$$

式中 β 是正则化参数。

3 模型求解

3.1 优化求解和算法设计

当 \mathbf{A} 、 \mathbf{B} 和 \mathbf{S} 固定时, 目标函数(8)是关于 \mathbf{b} 的凸函数。因此, 求式(8)关于 \mathbf{b} 的导函数, 并令导函数等于 0, 即得式(8)关于 \mathbf{b} 的更新公式

$$\mathbf{b} = \frac{\mathbf{e}^T \mathbf{Y} - \mathbf{e}^T \mathbf{XAB}}{n}. \quad (9)$$

当 \mathbf{b} 和 \mathbf{S} 固定时, 利用非负矩阵分解技术^[35]求解 \mathbf{A} 和 \mathbf{B} 。将约束条件 $\{\mathbf{A}, \mathbf{B}\} \geq 0$ 积分到目标函数中, 即得式(8)关于 \mathbf{A} 和 \mathbf{B} 的拉格朗日函数

$$L(\mathbf{A}, \mathbf{B}) = \text{Tr}((\mathbf{XAB} - \mathbf{Y})^T \mathbf{H} (\mathbf{XAB} - \mathbf{Y})) + \beta \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{L}_S \mathbf{A} \mathbf{B}) + \gamma \text{Tr}(\mathbf{B}\mathbf{L}_Y\mathbf{B}^T) - \text{Tr}(\Phi \mathbf{A}^T) - \text{Tr}(\Psi \mathbf{B}^T), \quad (10)$$

式中: $L(\mathbf{A}, \mathbf{B})$ 是关于 \mathbf{A} 和 \mathbf{B} 拉格朗日函数, $\mathbf{H} = \mathbf{I} - \frac{\mathbf{e}^T \mathbf{e}}{n}$ 是中心矩阵, \mathbf{I} 是 n 维单位矩阵, $\Phi \in \mathbf{R}^{d \times k}$ 和 $\Psi \in \mathbf{R}^{k \times m}$ 是拉格朗日乘子。根据非负矩阵分解技术, 求出 $L(\mathbf{A}, \mathbf{B})$ 关于 \mathbf{A} 和 \mathbf{B} 的偏导函数

$$\begin{cases} \frac{\partial L(\mathbf{A}, \mathbf{B})}{\partial \mathbf{A}} = 2\mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{A} \mathbf{B} \mathbf{B}^T - 2\mathbf{X}^T \mathbf{H} \mathbf{Y} \mathbf{B}^T + \beta \mathbf{L}_S \mathbf{A} \mathbf{B} \mathbf{B}^T - \Phi, \\ \frac{\partial L(\mathbf{A}, \mathbf{B})}{\partial \mathbf{B}} = 2\mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{A} \mathbf{B} + 2\mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y} + \beta \mathbf{A}^T \mathbf{L}_S \mathbf{A} \mathbf{B} + \gamma \mathbf{B} \mathbf{L}_Y - \Psi. \end{cases} \quad (11)$$

由 Karush-Kuhn-Tucker 定理^[36]的互补松弛条件可知, $\Phi_{ij} A_{ij} = 0$ 、 $\Psi_{ij} B_{ij} = 0$ 。再根据非负矩阵分解技术的计算规则, 即可得到目标函数关于 \mathbf{A} 和 \mathbf{B} 的更新公式

$$\begin{cases} A_{ij}^{t+1} \leftarrow A_{ij}^t \frac{(\mathbf{X}^T \mathbf{H} \mathbf{Y} \mathbf{B}^T + \beta \mathbf{S} \mathbf{A} \mathbf{B} \mathbf{B}^T)_{ij}^t}{(\mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{A} \mathbf{B} \mathbf{B}^T + \beta \mathbf{P} \mathbf{A} \mathbf{B} \mathbf{B}^T)_{ij}^{t+1}}, \\ B_{ij}^{t+1} \leftarrow B_{ij}^t \frac{(\mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y} + \beta \mathbf{A}^T \mathbf{S} \mathbf{A} \mathbf{B} + \gamma \mathbf{B} \mathbf{Z}_Y)_{ij}^t}{(\mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{A} \mathbf{B} + \beta \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B} + \gamma \mathbf{B} \mathbf{Q})_{ij}^t}. \end{cases} \quad (12)$$

当 \mathbf{A} 、 \mathbf{B} 和 \mathbf{b} 固定时, 式(8)关于 \mathbf{S} 的目标函数为

$$\begin{aligned}
F(\mathbf{S}) &= \min \alpha \|\mathbf{S}-\mathbf{Z}\|_1 + \beta \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{L}_s \mathbf{A} \mathbf{B}) \\
&= \min \sum_{i,j} (\alpha |S_{ij}-Z_{ij}| + \beta \|\mathbf{A}\mathbf{B}\|_{i^*} - \|\mathbf{A}\mathbf{B}\|_{j^*} \|^2_2 S_{ij}) \\
\text{s.t. } & 0 \leq S_{ij}, \quad \sum_j S_{ij} = 1, \quad S_{ii} = 0.
\end{aligned} \tag{13}$$

由于 \mathbf{S} 是对称矩阵,对于不同的 j ,式(13)是独立,因此,对式(13)进行拆分求解,即

$$F(\mathbf{S}_{*j}) = \min \frac{\alpha}{2} \mathbf{S}_{*j}^T \mathbf{M} \mathbf{S}_{*j} - \mathbf{S}_{*j}^T \mathbf{G}_{*j}, \quad \text{s.t. } 0 \leq S_{ij}, \quad \sum_j S_{ij} = 1, \quad S_{ii} = 0, \tag{14}$$

式中: $\mathbf{M} \in \mathbf{R}^{d \times d}$ 是对角矩阵, $M_{ii} = \frac{1}{2|S_{ij}-Z_{ij}| + \varepsilon}$, ε 是一个非负常数, $\mathbf{G}_{*j} = \frac{\alpha}{2} \mathbf{M} \mathbf{S}_{*j} - \frac{\beta}{2} \mathbf{V}_{*j}$; $V_{ij} = \|\mathbf{A}\mathbf{B}\|_{i^*} - \|\mathbf{A}\mathbf{B}\|_{j^*} \|^2_2$ 。

将式(14)的约束条件积分到目标函数中,即可得到关于式(14)的拉格朗日函数

$$L(\mathbf{S}_{*j}) = \frac{\alpha}{2} \mathbf{S}_{*j}^T \mathbf{M} \mathbf{S}_{*j} - \mathbf{S}_{*j}^T \mathbf{G}_{*j} - \lambda (\mathbf{S}_{*j}^T \mathbf{I}_d - 1) - \boldsymbol{\Omega}_{*j}^T \mathbf{S}_{*j}, \tag{15}$$

式中: \mathbf{I}_d 是元素全为 1 的 d 维列向量, $\lambda \geq 0$, $\boldsymbol{\Omega}_{ij} \geq 0$ 是拉格朗日乘子。

对每个 i ,求出式(15)关于 S_{ij} 的导函数,并令导函数等于 0,即

$$\alpha M_{ii} S_{ij} - G_{ij} - \lambda - \Omega_{ij} = 0. \tag{16}$$

由 Karush-Kuhn-Tucker 定理的互补松弛条件可知, $\Omega_{ij} S_{ij} = 0$ 。又因为 $S_{ij} \geq 0$,所以 $\Omega_{ij} = 0$ 。得到目标函数关于 \mathbf{S} 的更新公式

$$S_{ij} = \max\left(0, \frac{G_{ij} + \lambda}{\alpha M_{ii}}\right). \tag{17}$$

考虑到式(17)中的 λ 未知,根据约束条件 $\sum_j S_{ij} = 1$ 和 \mathbf{S} 是对称矩阵,设计了一个关于 λ 的分段单调递增函数 $f(\lambda)$,用于求解 λ :

$$f(\lambda) = \sum_i \max\left(0, \frac{G_{ij} + \lambda}{\alpha M_{ii}}\right) - 1. \tag{18}$$

由于式(18)是分段单调递增的,因此利用牛顿迭代获取式(18)的最优解 $\hat{\lambda}$,将 $\lambda = \hat{\lambda}$ 带入到式(17)中,即可得到 \mathbf{S} 的最优解。

综上,根据上述优化求解结果,结合交替迭代策略,设计了一种可证明收敛性的优化算法,如算法 1 所示。

算法 1 具有标签流形和动态图约束的多标签特征选择算法。

输入 多标签数据集,正则化参数 α 、 β 和 γ ;

输出 特征排序结果 \mathbf{F}_s 。

- (1) 初始化中心矩阵 \mathbf{H} ;
- (2) 根据高斯热核函数计算 \mathbf{Z}_γ 和 \mathbf{Q} ;
- (3) 根据式(5)计算 \mathbf{Z} ;
- (4) 初始化迭代次数 $t=0$;
- (5) 随机初始化 \mathbf{A}^t 、 \mathbf{B}^t 和 \mathbf{S}^t ;
- (6) 计算 $\mathbf{S}^t = \frac{1}{2}[\mathbf{S}^t + (\mathbf{S}^t)^T]$;
- (7) 重复:
- (8) 根据式(12)更新 \mathbf{A}^{t+1} 和 \mathbf{B}^{t+1} ;
- (9) 使用牛顿迭代获取式(18)的最优解 $\hat{\lambda}$;
- (10) 令 $\lambda = \hat{\lambda}$,根据式(17)更新 \mathbf{S}^{t+1} ;
- (11) 根据式(9)更新 \mathbf{b} ;
- (12) $t=t+1$;
- (13) 直到收敛为止;
- (14) 计算 $\|\mathbf{A}\mathbf{B}\|_i$, $i=1,2,\dots,d$,并从大到小排序,将排序结果记录在 \mathbf{F}_s 中。

3.2 收敛性证明

由式(8)可知, $F(\mathbf{b}, \mathbf{A}, \mathbf{B}, \mathbf{S})$ 是关于 \mathbf{b} 和 \mathbf{S} 的凸函数, 因此, 在算法 1 的第 t 次迭代中, 有

$$F(\mathbf{b}^{t+1}, \mathbf{A}^t, \mathbf{B}^t, \mathbf{S}^{t+1}) \leq F(\mathbf{b}^t, \mathbf{A}^t, \mathbf{B}^t, \mathbf{S}^t). \quad (19)$$

$F(\mathbf{b}, \mathbf{A}, \mathbf{B}, \mathbf{S})$ 是关于 \mathbf{A} 和 \mathbf{B} 的非凸函数, 因此, 在证明算法 1 关于 \mathbf{A} 和 \mathbf{B} 的收敛性之前, 须要了解一些相关概念。

定义 1^[37] 若有 $g_1(\mathbf{A}, \mathbf{A}') \geq g_2(\mathbf{A})$ 和 $g_1(\mathbf{A}, \mathbf{A}) = g_2(\mathbf{A})$ 成立, 则 $g_1(\mathbf{A}, \mathbf{A}')$ 是 $g_2(\mathbf{A})$ 的一个辅助函数。其中, \mathbf{A}' 是 \mathbf{A} 在上一步迭代中的更新值。

引理 1^[37] 若 $g_1(\mathbf{A}, \mathbf{A}')$ 是 $g_2(\mathbf{A})$ 的一个辅助函数, 则 $g_2(\mathbf{A})$ 以 $\mathbf{A}^{t+1} = \arg \min_{\mathbf{A}} g_1(\mathbf{A}, \mathbf{A}')$ 的形式单调递减, 其中, \mathbf{A}^{t+1} 是第 t 次迭代的最优解。

命题 1^[38] 若 $\mathbf{M}_1 \in \mathbf{R}^{k \times k}$ 和 $\mathbf{M}_2 \in \mathbf{R}^{d \times d}$ 是对称矩阵, 则对任意矩阵 $\mathbf{A} \in \mathbf{R}^{d \times k}$ 和 $\mathbf{A}' \in \mathbf{R}^{d \times k}$, 有如下不等式成立:

$$\sum_{i,j} \frac{(\mathbf{M}_2 \mathbf{A}' \mathbf{M}_1)_{ij} A_{ij}^2}{A'_{ij}} \geq \text{Tr}(\mathbf{A}^T \mathbf{M}_2 \mathbf{A} \mathbf{M}_1). \quad (20)$$

以算法 1 关于 \mathbf{A} 的收敛性为例。根据式(8), 将 $g_2(\mathbf{A})$ 定义为

$$g_2(\mathbf{A}) = \frac{1}{2} \text{Tr}(\mathbf{A}^T \mathbf{D}_1 \mathbf{A} \mathbf{D}_2) - \text{Tr}(\mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y} \mathbf{B}^T) + \frac{1}{2} \beta \text{Tr}(\mathbf{A}^T \mathbf{L}_s \mathbf{A} \mathbf{D}_2), \quad (21)$$

式中 $\mathbf{D}_1 = \mathbf{X}^T \mathbf{H} \mathbf{X}$ 和 $\mathbf{D}_2 = \mathbf{B} \mathbf{B}^T$ 均是对称矩阵。

根据式(21), 将 $g_1(\mathbf{A}, \mathbf{A}')$ 定义为

$$g_1(\mathbf{A}, \mathbf{A}') = \frac{1}{2} \sum_{i,j} \frac{(\mathbf{D}_1 \mathbf{A}' \mathbf{D}_2)_{ij} A_{ij}^2}{A'_{ij}} - \sum_{i,j} (\mathbf{X}^T \mathbf{H} \mathbf{Y} \mathbf{B}^T)_{ij} A_{ij} + \frac{1}{2} \beta \sum_{i,j} \frac{(\mathbf{L}_s \mathbf{A}' \mathbf{D}_2)_{ij} A_{ij}^2}{A'_{ij}}. \quad (22)$$

由命题 1 可知, 对于式(21)、(22), 有

$$\begin{cases} \frac{1}{2} \sum_{i,j} \frac{(\mathbf{D}_1 \mathbf{A}' \mathbf{D}_2)_{ij} A_{ij}^2}{A'_{ij}} \geq \text{Tr}(\mathbf{A}^T \mathbf{D}_1 \mathbf{A} \mathbf{D}_2), \\ \frac{1}{2} \beta \sum_{i,j} \frac{(\mathbf{L}_s \mathbf{A}' \mathbf{D}_2)_{ij} A_{ij}^2}{A'_{ij}} \geq \text{Tr}(\mathbf{A}^T \mathbf{L}_s \mathbf{A} \mathbf{D}_2). \end{cases} \quad (23)$$

由矩阵迹的计算机制可知, 对于式(21)、(22), 有

$$\sum_{i,j} (\mathbf{X}^T \mathbf{H} \mathbf{Y} \mathbf{B}^T)_{ij} A_{ij} = \text{Tr}(\mathbf{A}^T \mathbf{X}^T \mathbf{H} \mathbf{Y} \mathbf{B}^T). \quad (24)$$

结合式(23)、(24), 得到

$$g_1(\mathbf{A}, \mathbf{A}') \geq g_2(\mathbf{A}). \quad (25)$$

当 $\mathbf{A}' = \mathbf{A}$ 时, 式(23)重新表述为

$$\begin{cases} \frac{1}{2} \sum_{i,j} \frac{(\mathbf{D}_1 \mathbf{A} \mathbf{D}_2)_{ij} A_{ij}^2}{A_{ij}} = \text{Tr}(\mathbf{A}^T \mathbf{D}_1 \mathbf{A} \mathbf{D}_2), \\ \frac{1}{2} \beta \sum_{i,j} \frac{(\mathbf{L}_s \mathbf{A} \mathbf{D}_2)_{ij} A_{ij}^2}{A_{ij}} = \text{Tr}(\mathbf{A}^T \mathbf{L}_s \mathbf{A} \mathbf{D}_2). \end{cases} \quad (26)$$

由式(24)、(26)可知, 当 $\mathbf{A}' = \mathbf{A}$ 时, $g_1(\mathbf{A}, \mathbf{A}) = g_2(\mathbf{A})$ 。再结合式(25), 即可证明 $g_1(\mathbf{A}, \mathbf{A}')$ 是 $g_2(\mathbf{A})$ 的一个辅助函数。最后, 根据引理 1, 证明算法 1 关于 \mathbf{A} 是收敛的。此外, 由于算法 1 关于 \mathbf{A} 和 \mathbf{B} 的收敛性证明类似。

综上所述, 算法 1 的收敛性得证。

4 实验

为了分析所提算法的性能和有效性, 将 LMDG 算法与 9 种先进的算法进行对比, 包括: ①基于最大依赖性和最小冗余度的多标签特征选择 (multi-label feature selection based on max-dependency and min-redundancy, MDMR)^[29]; ②基于可扩展准则的大标签集多标签特征选择 (multi-label feature selection based

on scalable criterion for large label set, SCLS)^[30];③基于多准则决策的多标签特征选择(Multi-label feature selection using multi-criteria decision making, MFS-MCDM)^[31];④具有约束潜在结构共享项的多标签特征选择(multilabel feature selection with constrained latent structure shared term, SSFS)^[25];⑤鲁棒灵活稀疏正则化的多标签特征选择(multi-label feature selection via robust flexible sparse regularization, RFSFS)^[23];⑥具有双图正则化的鲁棒多标签特征选择(robust multi-label feature selection with dual-graph regularization, DRMFS)^[22];⑦动态子空间双图正则化多标签特征选择(dynamic subspace dual-graph regularized multi-label feature selection, DSMFS)^[14];⑧基于动态图流形正则化的稀疏多标签特征选择(sparse multi-label feature selection via dynamic graph manifold regularization, DMMFS)^[13];⑨具有动态局部和全局结构保护的鲁棒稀疏低冗余多标签特征选择(robust sparse and low-redundancy multi-label feature selection with dynamic local and global structure preservation, SLMDS)^[17]。实验在 10 个多标签数据集上进行,并使用一种多标签学习的懒惰学习算法(a lazy learning approach to multi-label learning, ML-KNN)^[39]作为多标签分类器的代表,评估各算法的指标为汉明损失、排序损失、错误率、覆盖率和平均精度^[16]。

4.1 数据集

实验中,所有数据集均来自木兰图书馆(<http://mulan.sourceforge.net/dataset-mlc.html>),包含生物、音乐、声音、图像、文本应用领域。Genbase 是一个用于蛋白质功能分类的数据集,每个实例和标签分别表示一个蛋白质和一个类。Yeast 数据集中包含 2 417 个酵母菌基因,具有微阵列表达和系统发育谱。此外,每个酵母都带有 14 个标签的子集(例如,能量、转录等)。Scene 数据集来自于图像领域,包含 2 407 个具有 294 个特征的实例,且每个实例对应 6 个类标签。考虑到这些数据集涵盖多个领域,且常用于多标签学习,因此,本文选用这些数据集作为实验数据。表 1 中列出了各数据集的详细信息。

表 1 数据集的详细信息
Table 1 The details of datasets

数据集	实例数	特征数	标签数	领域	训练集个数	测试集个数
Yeast	2 417	103	14	生物	1 500	917
Genbase	662	1 185	27	生物	463	199
Emotion	593	72	6	音乐	391	202
CAL500	502	68	174	声音	335	167
Scene	2 407	294	6	图像	1 211	1 196
Image	600	294	5	图像	400	200
Flags	194	19	7	图像	129	65
Enron	1 702	1 001	53	文本	1 123	579
Computer	5 000	681	33	文本	2 000	3 000
Recreation	5 000	606	22	文本	2 000	3 000

4.2 实验细节

实验中,将分类器 ML-KNN 的平滑参数设置为 1,近邻参数设置为 10。将所有实验数据使用等宽区间^[40]进行离散化处理。对于基于图的算法,将 Image 数据集上的标签近邻参数设置为 4,将其它数据上的近邻参数设置为 5;并使用各算法默认的图学习技术计算图矩阵。

为了实验的公平性,使用网格搜索策略^[41]调整各实验算法的正则化参数,并将搜索范围设置为{0.001, 0.01, 0.1, 1, 10, 100, 1 000},将 Flags 数据集上的所选特征数设置为 2~18,将其它实验数据集上的所选特征数设置为原始特征数的 2%~20%,步长为 2%。

最后,将所有实验均在微软 Windows11 系统,处理器为 ADM (Advanced Micro Devices) 锐龙 495 5 3600 6,主频率为 3.59 GHz,内存为 16.00 GB 环境中,使用 Matlab R2016a 运行所有算法。

4.3 对比实验

表 2—6 为各算法的汉明损失、排序损失、错误率、覆盖率和平均精度 5 个评价指标 10 次运行的最佳结果的“平均值(性能排序)”。在表 2—5 中,LMDG 算法的评价指标越小,表示 LMDG 算法越优于其它算法,表 6 中,LMDG 算法的评价指标越大,说明 LMDG 算法的性能越好。

为了比较各算法的整体性能,表 2—6 的最后一行为各算法在所有数据集上的结果的均值,用粗体和下划线分别标出,当前指标下某以算法在对应数据集上取得最佳结果和次佳结果。可以看出,LMDG 算法至

少在 90% 的数据集上的结果被粗体或下划线标记。

具体而言,在表 2、3、5 中,LMDG 算法在所有数据集上的结果均被粗体或下划线标记。根据表 3 展示的结果可知,在 Image 数据集上,相较其它算法,LMDG 算法的排序损失降低了 5.87%~28%。表 4 中,相较其它算法,LMDG 算法的错误率降低了 2.37%~29.27%;尤其在 Scene 数据集上,相对与 MDMR、SCLS、MFS-MCDM、SSFS、RFSFS、DRMFS、DSMFS、DMMFS 和 SLMDS 算法,LMDG 算法的错误率分别降低了 28.87%、29.82%、12.09%、29.10%、25.91%、2.31%、3.68%、2.79% 和 3.86%。从表 6 可以看出,虽然 LMDG 算法在 Genbase 数据集上的平均精度略低于 MDMR、DRMFS、DSMFS、DMMFS 和 SLMDS 算法,但 LMDG 算法在所有数据集上平均精度的均值仍大于其他算法。

表 2 各数据集上不同算法的汉明损失

Table 2 Hamming loss of different algorithms on each dataset

算法	MDMR	SCLS	MFS-MCDM	SSFS	RFSFS	DRMFS	DSMFS	DMMFS	SLMDS	LMDG
Yeast	0.208 4(6)	0.214 8(9)	0.210 2(8)	0.220 9(10)	0.209 1(7)	<u>0.199 1(2)</u>	0.200 6(3)	0.202 8(5)	0.202 2(4)	0.197 7(1)
Emotion	0.259 1(10)	0.242 6(8)	0.220 3(3)	0.253 3(9)	0.224 8(6)	0.222 0(4)	0.225 4(7)	<u>0.218 7(2)</u>	0.223 1(5)	0.214 4(1)
Scene	0.132 5(9.5)	0.132 5(9.5)	0.105 5(6)	0.129 8(8)	0.126 0(7)	0.102 5(3)	0.104 4(5)	<u>0.101 3(2)</u>	0.103 5(4)	0.100 9(1)
Image	0.217 0(8)	0.227 0(10)	0.204 0(5)	0.218 8(9)	0.210 8(7)	0.206 0(6)	0.198 6(3)	<u>0.196 0(2)</u>	0.203 4(4)	0.189 4(1)
Flags	0.593 4(6)	0.633 0(8)	0.580 2(3)	0.641 3(9)	0.646 2(10)	<u>0.578 0(2)</u>	0.583 3(4)	0.600 0(7)	0.588 6(5)	0.569 2(1)
Genbase	<u>0.002 6(4)</u>	0.032 0(8)	0.034 2(9)	0.042 8(10)	0.002 9(6)	0.002 4(1.5)	<u>0.002 6(4)</u>	0.002 4(1.5)	0.004 8(7)	<u>0.002 6(4)</u>
CAL500	0.966 0(10)	0.965 3(8)	0.964 0(7)	0.965 8(9)	0.963 8(5)	0.963 6(4)	<u>0.963 5(2.5)</u>	0.962 5(1)	0.963 9(6)	<u>0.963 5(2.5)</u>
Enron	0.049 8(7)	0.051 7(8)	0.053 6(10)	0.049 5(6)	0.052 2(9)	<u>0.048 2(2)</u>	0.048 3(3)	0.048 7(4)	0.049 3(5)	0.047 9(1)
Computer	0.039 8(7)	0.040 4(8)	0.040 8(9)	0.040 9(10)	0.039 4(6)	0.039 0(3)	<u>0.038 7(2)</u>	0.039 1(4)	0.039 3(5)	0.038 2(1)
Recreation	0.060 9(8)	0.063 5(9)	0.064 5(10)	0.059 6(5)	0.060 0(6)	0.057 6(3)	0.056 8(1)	0.057 7(4)	0.060 7(7)	<u>0.056 9(2)</u>
均值	0.253 0	0.260 3	0.247 7	0.262 3	0.253 5	<u>0.241 8</u>	0.242 2	0.242 9	0.243 9	0.238 1

表 3 各数据集上不同算法的排序损失

Table 3 Ranking loss of different algorithms on each dataset

算法	MDMR	SCLS	MFS-MCDM	SSFS	RFSFS	DRMFS	DSMFS	DMMFS	SLMDS	LMDG
Yeast	0.184 5(6)	0.193 5(9)	0.186 7(8)	0.199 7(10)	0.185 1(7)	<u>0.172 3(2)</u>	0.174 5(3)	0.174 6(4)	0.177 2(5)	0.171 9(1)
Emotion	0.225 0(10)	0.199 1(8)	0.194 6(6)	0.218 1(9)	0.198 1(7)	0.185 3(3)	0.188 4(5)	<u>0.183 9(2)</u>	0.187 7(4)	0.179 8(1)
Scene	0.137 1(7)	0.165 2(10)	0.111 1(6)	0.148 6(9)	0.142 2(8)	0.098 3(3)	0.100 0(5)	0.097 2(1)	0.099 1(4)	<u>0.098 2(2)</u>
Image	0.260 4(10)	0.250 0(9)	0.206 3(4)	0.228 0(8)	0.218 9(7)	0.212 1(6)	0.199 6(3)	<u>0.199 2(2)</u>	0.208 6(5)	0.187 5(1)
Flags	0.181 3(3)	0.248 2(10)	0.182 3(4.5)	0.207 8(8)	0.210 2(9)	0.182 3(4.5)	<u>0.177 1(2)</u>	0.202 1(7)	0.183 3(6)	0.173 4(1)
Genbase	0.006 2(2.5)	0.049 9(9)	0.048 0(8)	0.080 4(10)	0.007 9(7)	0.006 2(2.5)	0.006 2(2.5)	0.006 2(2.5)	0.007 1(6)	<u>0.006 7(5)</u>
CAL500	0.181 9(6)	0.183 4(10)	0.182 0(7)	0.183 2(9)	0.181 7(3.5)	0.181 8(5)	<u>0.181 3(2)</u>	0.182 3(8)	0.181 7(3.5)	0.181 2(1)
Enron	0.088 7(4)	0.095 8(9)	0.100 6(10)	0.089 9(6)	0.095 4(8)	0.088 4(3)	<u>0.087 5(2)</u>	0.089 5(5)	0.091 9(7)	0.087 1(1)
Computer	0.093 6(8.5)	0.091 8(7)	0.096 4(10)	0.093 6(8.5)	0.091 2(5)	0.090 1(3)	0.086 2(1)	0.090 2(4)	0.091 6(6)	<u>0.086 3(2)</u>
Recreation	0.184 2(8)	0.192 8(9)	0.202 9(10)	0.176 0(6)	0.172 3(5)	0.171 8(4)	0.164 6(1)	0.166 9(3)	0.181 4(7)	<u>0.165 2(2)</u>
均值	0.154 3	0.167 0	0.151 1	0.162 5	0.150 3	0.138 9	<u>0.136 5</u>	0.139 2	0.141 0	0.133 7

表 4 各数据集上不同算法的错误率

Table 4 One-error of different algorithms on each dataset

算法	MDMR	SCLS	MFS-MCDM	SSFS	RFSFS	DRMFS	DSMFS	DMMFS	SLMDS	LMDG
Yeast	0.246 5(8)	0.247 5(9)	0.242 1(7)	0.248 6(10)	0.241 4(6)	0.225 1(3)	<u>0.223 3(2)</u>	0.226 8(4)	0.232 1(5)	0.220 1(1)
Emotion	0.376 2(10)	0.321 8(8)	<u>0.277 2(2)</u>	0.341 6(9)	0.302 0(5.5)	0.302 0(5.5)	0.303 0(7)	0.272 3(1)	0.290 1(4)	0.284 2(3)
Scene	0.367 9(8)	0.372 9(10)	0.297 7(6)	0.369 1(9)	0.353 2(7)	<u>0.267 9(2)</u>	0.271 7(4)	0.269 2(3)	0.272 2(5)	0.261 7(1)
Image	0.465 0(9)	0.485 0(10)	0.385 0(5)	0.411 0(8)	0.398 0(7)	0.370 0(4)	<u>0.363 0(2)</u>	0.365 0(3)	0.390 0(6)	0.342 0(1)
Flags	0.156 3(8.5)	0.203 1(10)	0.140 6(5.5)	0.150 0(7)	0.140 6(5.5)	<u>0.109 4(2)</u>	0.118 8(3)	0.156 3(8.5)	0.125 0(4)	0.106 3(1)
Genbase	0(3)	0.392 0(8)	0.422 1(9)	0.532 7(10)	0.003 0(7)	0(3)	0(3)	0(3)	0(3)	<u>0.002 0(6)</u>
CAL500	0.089 8(9)	0.089 8(9)	0.083 8(6.5)	0.089 8(9)	<u>0.079 0(3)</u>	<u>0.079 0(3)</u>	0.077 8(1)	0.083 8(6.5)	0.082 6(5)	<u>0.079 0(3)</u>
Enron	0.253 9(7)	0.283 2(8)	0.319 5(10)	0.241 1(5)	0.284 6(9)	0.224 5(1)	0.226 3(3)	0.231 4(4)	0.244 2(6)	<u>0.225 2(2)</u>
Computer	0.446 0(7)	0.449 3(8)	0.449 7(9)	0.461 0(10)	0.431 0(5)	<u>0.426 5(2.5)</u>	<u>0.426 5(2.5)</u>	0.428 3(4)	0.438 6(6)	0.421 8(1)
Recreation	0.668 0(8)	0.712 3(9)	0.745 3(10)	0.629 9(5)	0.632 1(6)	0.597 3(4)	0.566 7(1)	0.582 0(3)	0.660 7(7)	<u>0.573 9(2)</u>
均值	0.307 0	0.355 7	0.336 3	0.347 5	0.286 5	0.260 2	<u>0.257 7</u>	0.261 5	0.273 6	0.251 6

表 5 各数据集上不同算法的覆盖率

Table 5 Coverage of different algorithms on each dataset

算法	MDMR	SCLS	MFS-MCDM	SSFS	RFSFS	DRMFS	DSMFS	DMMFS	SLMDS	LMDG
Yeast	6.542 0(6)	6.677 2(9)	6.620 5(8)	6.726 3(10)	6.576 7(7)	6.385 0(2)	6.416 1(3)	6.430 8(4)	6.465 2(5)	6.373 8(1)
Emotion	2.183 2(10)	2.099 0(8)	2.029 7(6)	2.163 4(9)	2.056 4(7)	1.980 2(2.5)	2.017 8(5)	1.980 2(2.5)	1.988 1(4)	1.955 4(1)
Scene	0.786 0(7)	0.933 1(10)	0.659 7(6)	0.844 5(9)	0.814 4(8)	0.595 2(3)	0.605 0(5)	0.583 6(1)	0.599 2(4)	0.592 5(2)
Image	1.325 0(10)	1.270 0(9)	1.080 0(3)	1.179 0(8)	1.163 0(7)	1.130 0(6)	1.084 0(4)	1.075 0(2)	1.118 0(5)	1.033 0(1)
Flags	3.630 8(2)	4.046 2(10)	3.676 9(5)	3.840 0(9)	3.769 2(8)	3.707 7(7)	3.646 2(3)	3.661 5(4)	3.683 1(6)	3.624 6(1)
Genbase	0.572 9(2.5)	1.708 5(9)	1.618 1(8)	2.577 9(10)	0.580 9(6)	0.572 9(2.5)	0.572 9(2.5)	0.572 9(2.5)	0.584 9(7)	0.577 9(5)
CAL500	129.209 6(8)	129.335 3(9.5)	128.485 0(7)	129.335 3(9.5)	127.730 0(5)	127.410 0(3.5)	127.410 0(3.5)	127.330 0(1)	127.780 0(6)	127.358 1(2)
Enron	12.806 6(5)	13.412 8(9)	13.621 8(10)	12.840 8(6)	13.376 0(8)	12.726 0(3)	12.709 0(2)	12.756 0(4)	13.091 0(7)	12.564 8(1)
Computer	4.474 0(8)	4.411 3(7)	4.613 0(10)	4.485 3(9)	4.385 5(5)	4.363 4(3)	4.177 8(1)	4.379 7(4)	4.403 5(6)	4.181 1(2)
Recreation	4.912 0(8)	5.058 3(9)	5.305 3(10)	4.754 9(6)	4.679 7(5)	4.672 6(4)	4.533 3(1)	4.582 0(3)	4.871 9(7)	4.543 3(2)
均值	16.644 2	16.895 2	16.771 0	16.874 7	16.513 2	16.354 3	16.317 2	16.335 2	16.458 5	16.280 4

表 6 各数据集上不同算法的平均精度

Table 6 Average precision of different algorithms on each dataset

算法	MDMR	SCLS	MFS-MCDM	SSFS	RFSFS	DRMFS	DSMFS	DMMFS	SLMDS	LMDG
Yeast	0.741 6(6)	0.730 7(9)	0.737 4(8)	0.722 3(10)	0.741 1(7)	0.758 2(2)	0.756 6(3)	0.755 8(4)	0.751 4(5)	0.759 1(1)
Emotion	0.734 9(10)	0.764 0(8)	0.787 1(2)	0.749 1(9)	0.766 5(7)	0.779 2(5)	0.777 7(6)	0.785 4(3)	0.779 4(4)	0.790 6(1)
Scene	0.777 3(8)	0.760 8(10)	0.817 5(6)	0.769 4(9)	0.779 9(7)	0.836 4(3)	0.833 6(4)	0.837 1(2)	0.833 3(5)	0.838 8(1)
Image	0.697 9(9)	0.694 4(10)	0.744 9(6)	0.734 8(8)	0.737 6(7)	0.752 8(4)	0.761 3(3)	0.761 9(2)	0.749 5(5)	0.773 6(1)
Flags	0.846 2(4)	0.802 4(10)	0.842 5(6.5)	0.837 2(9)	0.841 1(8)	0.844 4(5)	0.849 5(3)	0.842 5(6.5)	0.856 4(2)	0.861 9(1)
Genbase	0.993 9(2.5)	0.734 2(8)	0.707 1(9)	0.619 1(10)	0.990 4(7)	0.993 9(2.5)	0.993 9(2.5)	0.993 9(2.5)	0.992 3(5)	0.991 7(6)
CAL500	0.496 5(7)	0.493 7(10)	0.497 1(6)	0.494 4(9)	0.498 6(3)	0.497 6(4)	0.499 0(1)	0.495 8(8)	0.497 3(5)	0.498 7(2)
Enron	0.657 1(7)	0.637 6(8)	0.619 1(10)	0.660 7(5)	0.633 1(9)	0.672 0(2)	0.670 0(3)	0.667 9(4)	0.658 2(6)	0.673 6(1)
Computer	0.632 5(6)	0.631 8(8)	0.621 5(9)	0.620 4(10)	0.639 7(5)	0.641 7(4)	0.645 6(2)	0.643 6(3)	0.632 3(7)	0.649 0(1)
Recreation	0.480 7(8)	0.447 1(9)	0.424 1(10)	0.506 0(5)	0.505 3(6)	0.527 3(4)	0.551 5(1)	0.539 7(3)	0.486 5(7)	0.545 3(2)
均值	0.705 9	0.669 7	0.679 8	0.671 3	0.713 3	0.730 4	0.733 9	0.732 4	0.723 7	0.738 2

为了比较各算法在当前指标下评价结果随所选特征数增加的变化情况,在图 1—5 中分别描绘了各算法的汉明损失、排序损失、错误率、覆盖率和平均精度变化曲线。可以看出,在图 1—4 中,LMDG 算法在 Yeast、Emotion、Image 和 Enron 数据集上的曲线始终出现在其它算法曲线的下方;在图 5 中,LMDG 算法在所有数据集上的曲线都出现在其它算法曲线的上方。特别地,由图 1(a)可以看出,在 Flags 数据集上,当所选特征数为 6 时,LMDG 算法的汉明损失曲线达到最低点,且明显低于所有对比算法的曲线,这表明 LMDG 算法可以更早的得到高质量的特征子集。

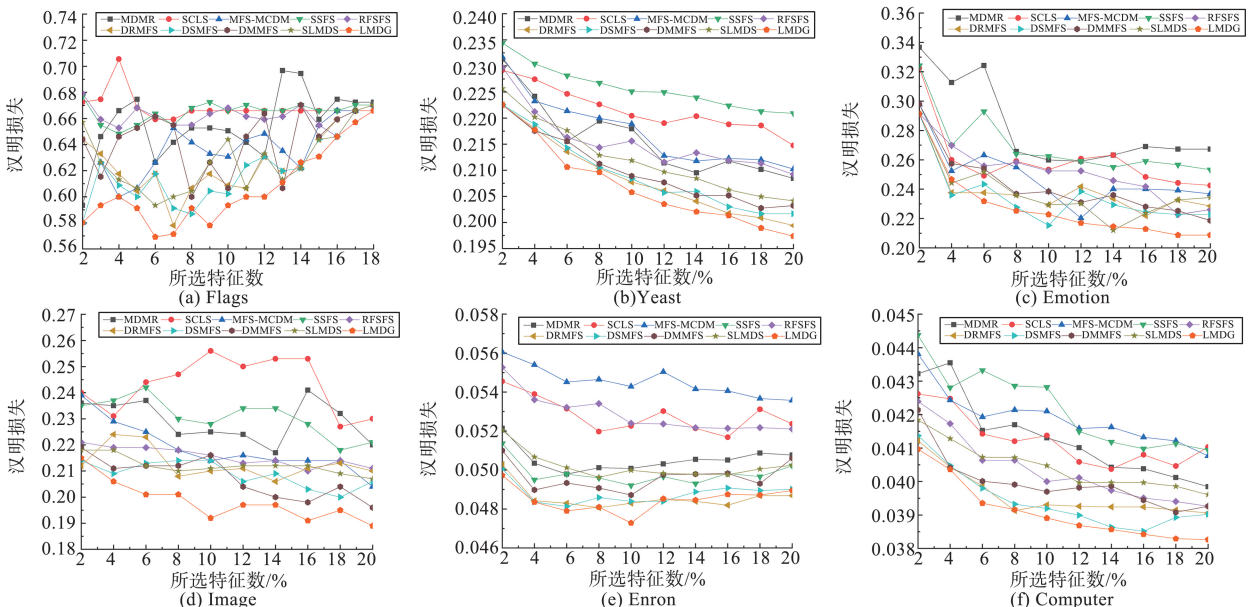


图 1 随所选特征数的增加,各算法的汉明损失比较

Fig.1 Comparison of the Hamming loss of each algorithm as the number of selected features increases

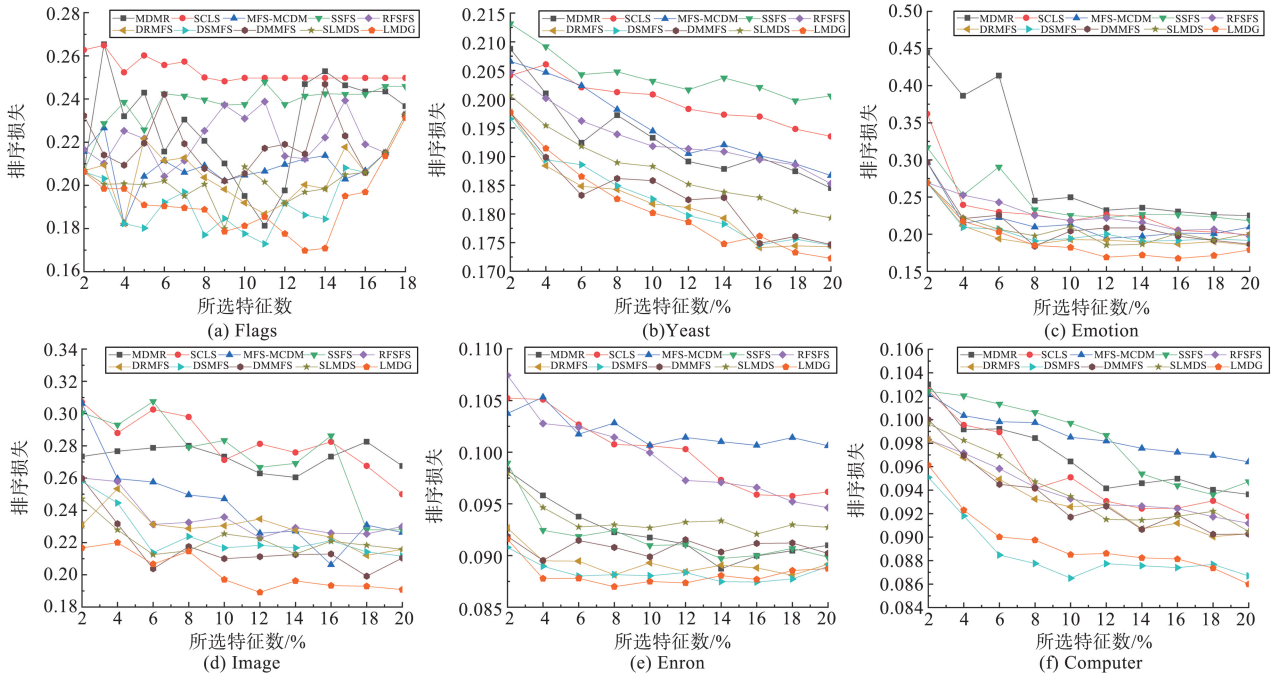


图 2 随所选特征数的增加,各算法的排序损失比较

Fig.2 Comparison of the Ranking loss of each algorithm as the number of selected features increases

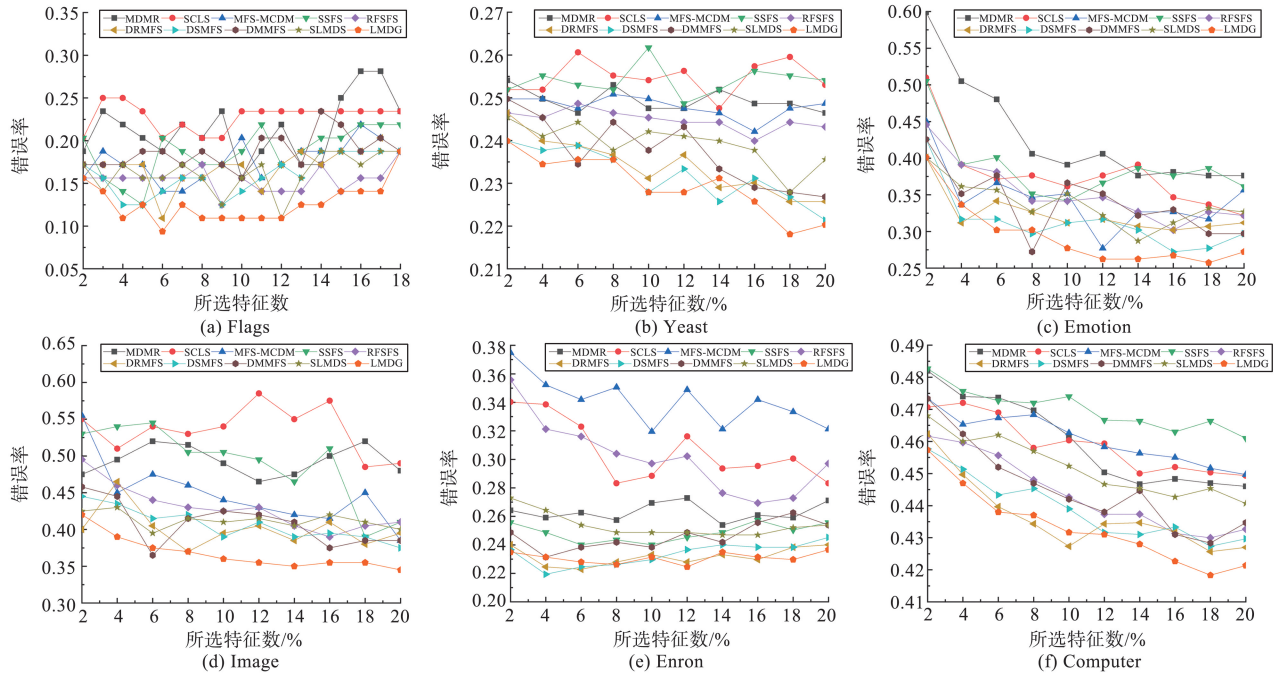
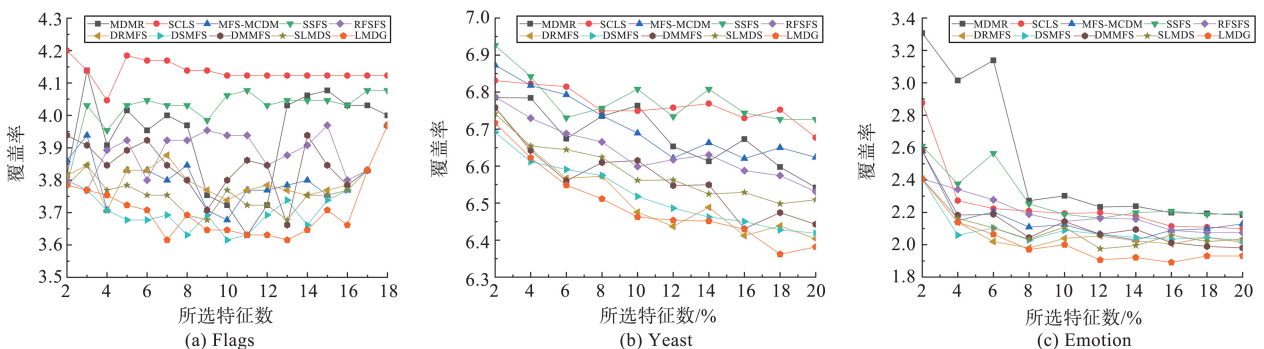


图 3 随所选特征数的增加,各算法的错误率比较

Fig.3 Comparison of the One-error of each algorithm as the number of selected features increases



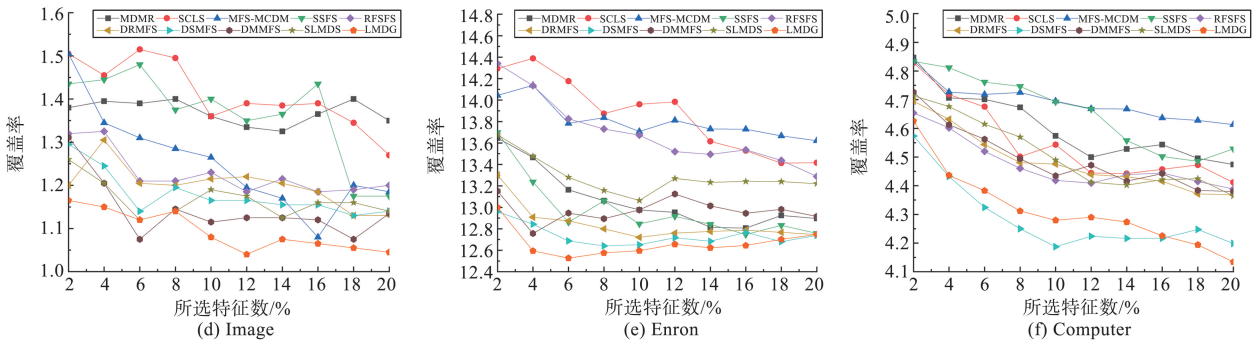


图4 随所选特征数的增加,各算法的覆盖率比较

Fig.4 Comparison of the Coverage of each algorithm as the number of selected features increases

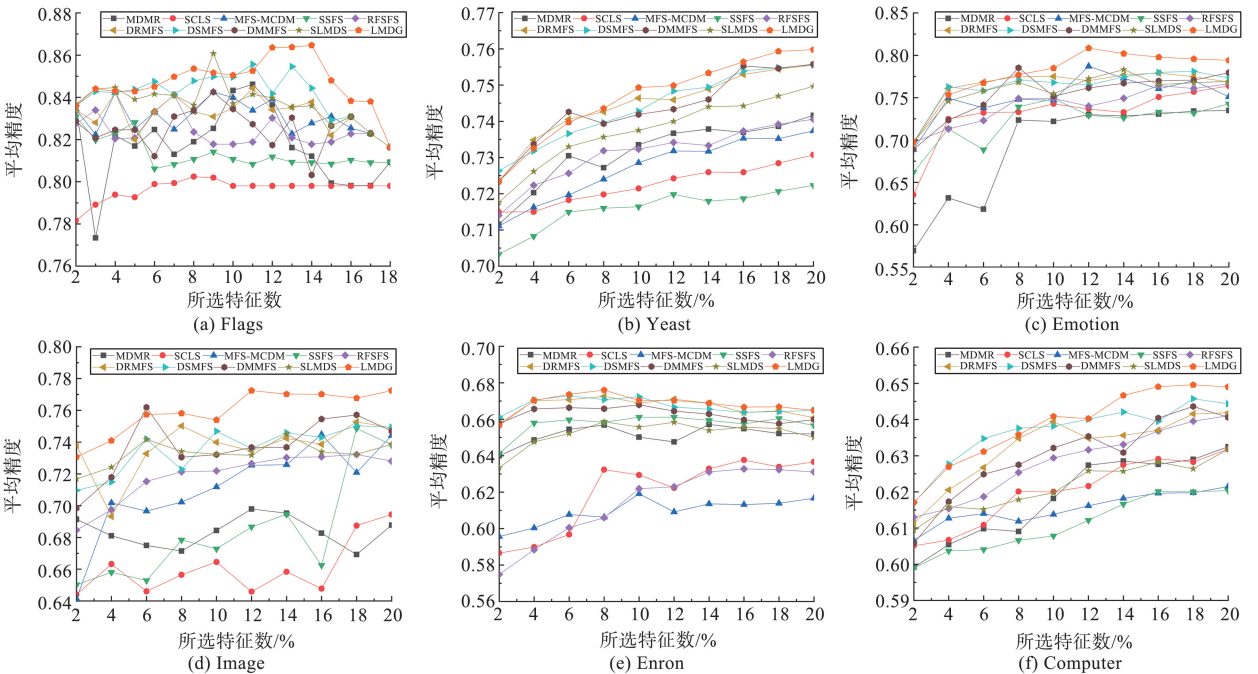


图5 随所选特征数的增加,各算法的平均精度比较

Fig.5 Comparison of the Average precision of each algorithm as the number of selected features increases

综上,LMDG 算法虽然在 Genbase 和 Emotion 数据集上的结果欠佳,但在其它数据集上 LMDG 算法均优于对比算法,尤其在 Image 和 Flags 数据集上,LMDG 算法的优势更为明显。

为了分析各算法间的显著性差异,使用显著性水平为 0.1 的 Bonferroni-Dunn 检验^[42]对各算法的性能进行分析,并将检验结果以平均秩图的形式在图 6 中展示。

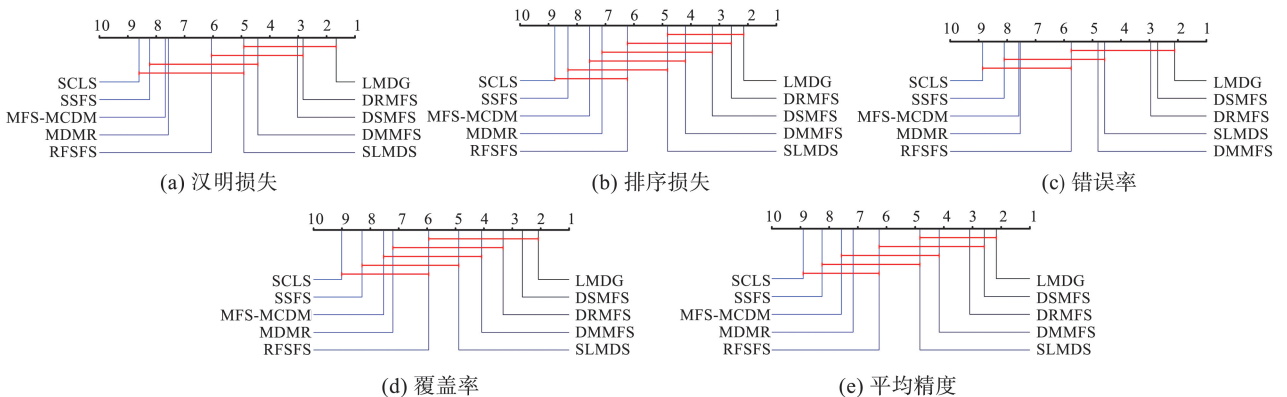


图6 Bonferroni-Dunn 检验结果的平均秩图形式

Fig.6 Bonferroni-Dunn test results in mean rank graph form

在图 6 中,横轴表示各算法在当前指标下整体性能的排序结果,从左到右,算法的整体性能越来越好;并使

用红线连接不存在显著性差异的算法,其中红线的长度表示 Bonferroni-Dunn 检验的临界值,这里临界值为 3.953 7。可以看出,虽然 LMDG 算法与 DSMFS、DRMFS、DMMFS、SLMDS 算法之间不存在显著性差异,但 LMDG 算法的排序位置始终处于横轴的最右端。

为了进一步对比分析各算法间的性能差异,使用显著性水平为 0.05 的 Friedman 检验^[43]作为后验技术,对各算法的性能进行分析。在本文中,Friedman 检验的临界值为 1.998;汉明损失、排序损失、错误率、覆盖率和平均精度各指标的 Friedman 检验值分别为 21.208、18.472、17.865、21.465 和 18.176。各指标的 Friedman 检验值均大于临界值,这表明虽然 LMDG 算法与 DSMFS、DRMFS、DMMFS、SLMDS 算法之间不存在显著性差异,但 LMDG 算法的性能仍优于 DSMFS、DRMFS、DMMFS、SLMDS 算法。

4.4 消融实验

为了分析 LMDG 算法中各主要成分的有效性,在 Image、Yeast 和 Emotion 数据集上进行了一系列消融实验。实验中,实验设置和对比实验的实验设置相同。此外,针对 LMDG 算法中的各主要成分,对 LMDG 算法进行衍生变换,具体如下:

(1) LMDG-1 算法。在 LMDG 算法中使用传统的线性映射模型,标函数定义为

$$F(\mathbf{b}, \mathbf{W}, \mathbf{S}) = \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{b}} \|\mathbf{X}\mathbf{W} + \mathbf{e}\mathbf{b} - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{S} - \mathbf{Z}\|_1 + \beta \text{Tr}(\mathbf{W}^T \mathbf{L}_S \mathbf{W}) + \gamma \text{Tr}(\mathbf{W} \mathbf{L}_Y \mathbf{W}^T), \quad (27)$$

$$\text{s.t. } 0 \leq S_{ij}, \sum_j S_{ij} = 1, S_{ii} = 0, \mathbf{W} \geq 0,$$

式中: $\mathbf{W} \in \mathbf{R}^{d \times m}$ 是系数矩阵,用于量化特征重要性。

(2) LMDG-2 算法。在 LMDG 算法中使用高斯热核函数捕获固定的特征相似图,标函数定义为

$$F(\mathbf{b}, \mathbf{A}, \mathbf{B}, \mathbf{S}) = \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{b}} \|\mathbf{X}\mathbf{A}\mathbf{B} + \mathbf{e}\mathbf{b} - \mathbf{Y}\|_F^2 + \beta \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{L}_X \mathbf{A}\mathbf{B}) + \gamma \text{Tr}(\mathbf{B} \mathbf{L}_Y \mathbf{B}^T), \quad (28)$$

$$\text{s.t. } \{\mathbf{A}, \mathbf{B}\} \geq 0,$$

式中: $\mathbf{L}_X = \mathbf{P}_X - \mathbf{Z}_X$ 是关于特征相似图的拉普拉斯矩阵, $\mathbf{P}_X \in \mathbf{R}^{d \times d}$ 是对角矩阵, $(\mathbf{P}_X)_{ii} = \sum_j (\mathbf{Z}_X)_{ij}$; $\mathbf{Z}_X \in \mathbf{R}^{d \times d}$ 是使用高斯热核函数捕获的特征相似图。

(3) LMDG-3 算法。在 LMDG 算法中不使用标签流形,标函数定义为

$$F(\mathbf{b}, \mathbf{A}, \mathbf{B}, \mathbf{S}) = \min_{\mathbf{A}, \mathbf{B}, \mathbf{S}, \mathbf{b}} \|\mathbf{X}\mathbf{A}\mathbf{B} + \mathbf{e}\mathbf{b} - \mathbf{Y}\|_F^2 + \alpha \|\mathbf{S} - \mathbf{Z}\|_1 + \beta \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{L}_S \mathbf{A}\mathbf{B}), \quad (29)$$

$$\text{s.t. } 0 \leq S_{ij}, \sum_j S_{ij} = 1, S_{ii} = 0, \{\mathbf{A}, \mathbf{B}\} \geq 0.$$

所有实验结果均记录在表 7 中。为了比较方便,在表 7 中使用粗体标记出最佳结果。在 15 组实验中,有 14 组 LMDG 算法的实验结果被粗体标记,表明 LMDG 算法中的各指标是有效的。

表 7 在 3 个数据集上的消融实验结果

Table 7 Results of ablation experiments on the three datasets

数据集	算法	汉明损失	排序损失	错误率	覆盖率	平均精度
Image	LMDG-1	0.192 5	0.198 6	0.354 6	1.083 0	0.763 5
	LMDG-2	0.202 2	0.205 6	0.353 6	1.075 0	0.764 6
	LMDG-3	0.212 3	0.204 6	0.368 3	1.125 0	0.755 2
	LMDG	0.189 4	0.187 5	0.342 0	1.033 0	0.773 6
Yeast	LMDG-1	0.198 2	0.174 6	0.223 1	6.375 0	0.757 4
	LMDG-2	0.202 2	0.177 2	0.232 5	6.402 0	0.748 2
	LMDG-3	0.208 4	0.178 9	0.232 1	6.415 2	0.745 8
	LMDG	0.197 7	0.171 9	0.220 1	6.373 8	0.759 1
Emotion	LMDG-1	0.215 6	0.180 4	0.271 0	1.980 2	0.787 7
	LMDG-2	0.223 2	0.183 2	0.277 2	1.983 4	0.789 2
	LMDG-3	0.218 4	0.184 4	0.282 3	2.029 7	0.779 1
	LMDG	0.214 4	0.179 8	0.284 2	1.955 4	0.790 6

4.5 参数敏感性实验

本小节在 Scene 数据集上分析了 LMDG 算法中各参数的有效性和敏感性。实验中,以平均精度作为评价指标,所选特征数设置为 $\{14, 28, 42, 56, 70, 84, 98\}$; 目标参数在 $\{0.001, 0.01, 0.1, 1, 10, 100, 1\ 000\}$ 范围内搜索,

并固定其它 2 个参数为 1,观察目标参数的有效性和 LMDG 算法对该参数的敏感性。实验结果如图 7 所示,在图 7 的(a)—(c)中为目标参数为 α 、 β 和 γ 上的实验结果。根据图 7 可知,随着所选特征数量的增加,LMDG 算法的平均精度先增加再趋于稳定,这说明使用 LMDG 算法解决特征选择问题是有效性;当所选特征数较小时 LMDG 算法对参数 β 较为敏感,但随着所选特征数的增加,LMDG 算法对所有参数的变化均不敏感。

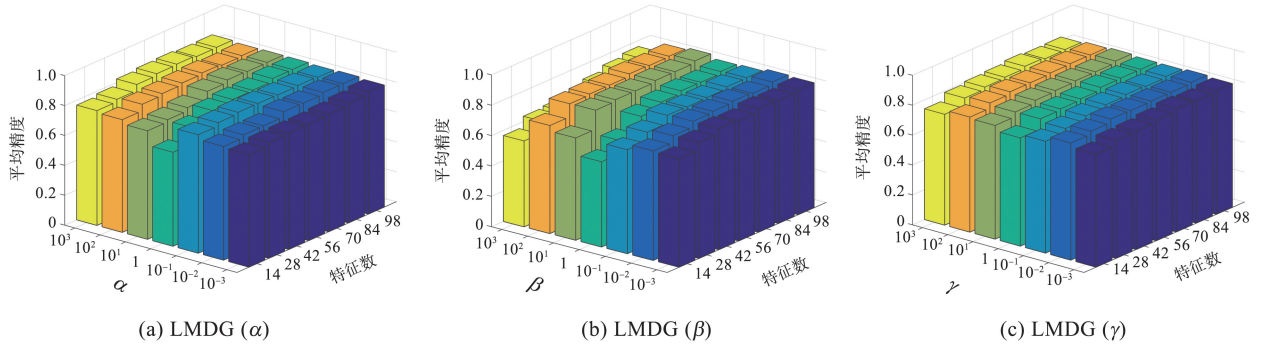


图 7 LMDG 算法在 Scene 数据集上的参数敏感性实验结果

Fig.7 Experimental results of parameter sensitivity of LMDG algorithm on Scene dataset

4.6 收敛性实验和时间复杂度分析

为了验证 3.2 节中算法 1 的收敛性理论证明的准确性,在 Emotion、Image、Enron 和 Computer 数据集上进行了收敛性实验。实验中,设置算法 1 的最大迭代次数为 50,观察目标函数值随迭代的增加的变化情况。实验结果如图 8 所示,可以看出算法 1 在各数据集上均是收敛的,30 次迭代内达到收敛条件,这一现象也在其它数据集上出现。

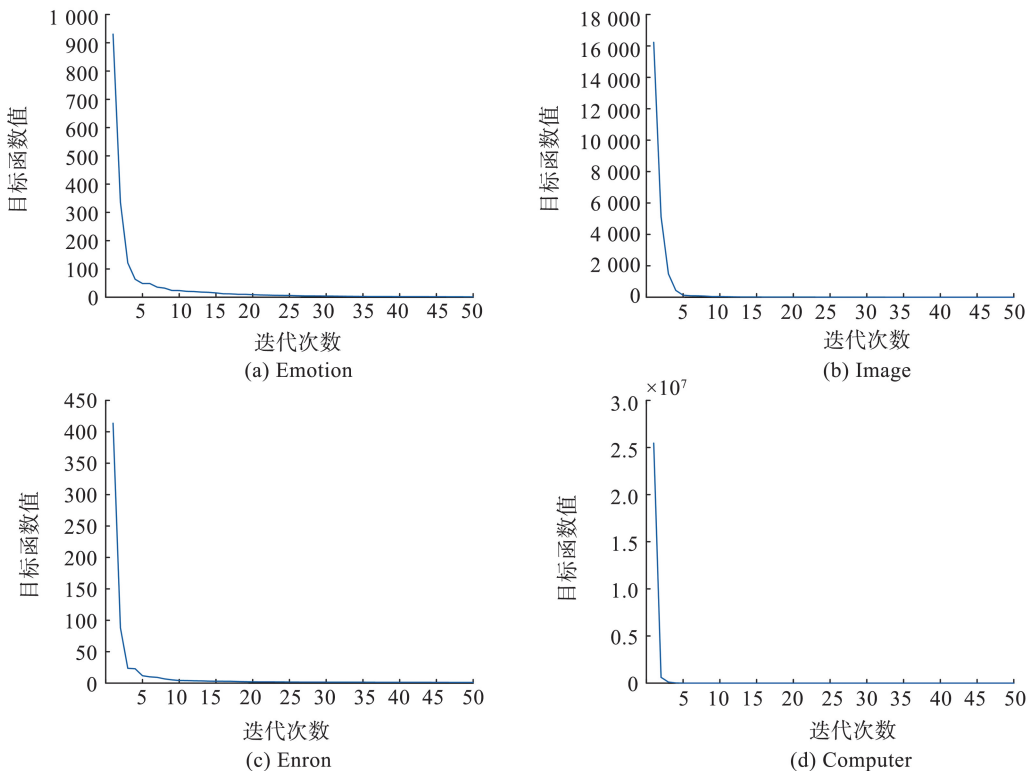


图 8 LMDG 算法在 Emotion、Image、Enron 和 Computer 数据集上的收敛性实验结果

Fig.8 Convergence experimental results of LMDG algorithm on Emotion, Image, Enron and Computer datasets

此外,为了更好的分析算法 1 的时间成本,对算法 1 的时间复杂度进行分析。在多标签数据集中: $n>d$ 、 $n>m$,算法 1 中, $b, \lambda, S \in \mathbf{R}^{d \times d}$, $A \in \mathbf{R}^{d \times k}$, $B \in \mathbf{R}^{k \times m}$ 。由式(9)、(12)、(17)和(18)可知,算法 1 在每次迭代中更新 b, λ, S, A 和 B 的时间复杂度分别为 $O(dm)$ 、 $O(d^2)$ 、 $O(d^2)$ 、 $O(n^2k)$ 和 $O(nd^2k)$ 。此外,由图 8 可知,算法 1 达到收敛的迭代次数一般较小,因此算法 1 的总时间复杂度为 $O(n^2k+nd^2k+d^2+dm)$ 。

5 结论

本文设计了一种基于拉普拉斯秩约束的自适应动态图学习技术,基于特征自表示的改进矩阵分解技术改进了传统的线性映射模型。结合标签流形,提出了具有标签流形和动态图约束的多标签特征选择算法。大量实验结果表明,与9种先进算法相比,LMDG算法具有以下竞争优势:相较现有的动态图学习技术,LMDG算法中设计的自适应动态图学习技术可以更好地捕获特征相似性,更好修正算法的计算误差;改进后的线性映射模型可以更好的解耦特征与特征之间以及特征与标签之间的相关性;各种实验结果验证了LMDG算法可以更好的解决多标签特征选择问题。

然而,实验中也发现了LMDG算法的缺陷,LMDG算法忽略了多标签学习中的标签冗余问题。因此,后续工作将关注标签冗余问题,通过设计合理的惩罚项(如内积正则项、余弦相似性、因果关系等)消除标签冗余,提高算法的计算效率,并将所提算法应用到实际领域中。

参考文献:

- [1] WANG Boyan, HU Xuegang, LI Peipei, et al. Cognitive structure learning model for hierarchical multi-label text classification [J]. Knowledge-Based Systems, 2021, 218:106876.
- [2] XIONG Jie, YU Li, NIU Xi, et al. XRR: extreme multi-label text classification with candidate retrieving and deep ranking[J]. Information Sciences, 2023, 622:115-132.
- [3] KOMEILI M, LOUIS W, ARMANFARD N, et al. Feature selection for nonstationary data: application to human recognition using medical biometrics[J]. IEEE Transactions on Cybernetics, 2017, 48(5):1446-1459.
- [4] JANET J P, KULIK H J. Resolving transition metal chemical space: feature selection for machine learning and structure-property relationships[J]. The Journal of Physical Chemistry A, 2017, 121(46):8939-8954.
- [5] BERMINGHAM M L, PONG-WONG R, SPILIOPOULOU A, et al. Application of high-dimensional feature selection: evaluation for genomic prediction in man[J]. Scientific Reports, 2015, 5:10312.
- [6] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The elements of statistical learning: data mining, inference, and prediction[J]. The Mathematical Intelligencer, 2005, 27(2):83-85.
- [7] SUN Xin, LIU Yanheng, LI Jin, et al. Using cooperative game theory to optimize the feature selection problem [J]. Neurocomputing, 2012, 97:86-93.
- [8] KASHEF S, NEZAMABADI-POUR H, NIKPOUR B. Multilabel feature selection: a comprehensive review and guiding experiments[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(2):e1240.
- [9] YIN Hui, YANG Shuiqiao, SONG Xiangyu, et al. Deep fusion of multimodal features for social media retweet time prediction [J]. World Wide Web, 2021, 24:1027-1044.
- [10] SHI Jianbo, MALIK J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):888-905.
- [11] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500):2323-2326.
- [12] ZHANG Yao, MA Yingcang, YANG Xiaofei. Multi-label feature selection based on logistic regression and manifold learning [J]. Applied Intelligence, 2022, 52(8):9256-9273.
- [13] ZHANG Yao, MA Yingcang. Sparse multi-label feature selection via dynamic graph manifold regularization [J]. International Journal of Machine Learning and Cybernetics, 2023, 14(3):1021-1036.
- [14] HU Juncheng, LI Yonghao, XU Gaochao, et al. Dynamic subspace dual-graph regularized multi-label feature selection [J]. Neurocomputing, 2022, 467:184-196.
- [15] 李永豪,胡亮,张平,等. 基于动态图拉普拉斯的多标签特征选择[J]. 通信学报,2020,41(12):47-59.
LI Yonghao, HU Liang, ZHANG Ping, et al. Multi-label feature selection based on dynamic graph Laplacian [J]. Journal on Communications, 2020, 41(12):47-59.
- [16] ZHANG Yao, MA Yingcang. Non-negative multi-label feature selection with dynamic graph constraints [J]. Knowledge-Based Systems, 2022, 238:107924.
- [17] LI Yonghao, HU Liang, GAO Wanfu. Robust sparse and low-redundancy multi-label feature selection with dynamic local and global structure preservation [J]. Pattern Recognition, 2023, 134:109120.
- [18] ZHANG Minling, ZHOU Zhihua. A review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 26(8):1819-1837.

- [19] BOUTELL M R, LUO Jiebo, SHEN Xipeng, et al. Learning multi-label scene classification[J]. *Pattern Recognition*, 2004, 37(9):1757-1771.
- [20] HUANG Jun, LI Guorong, HUANG Qingming, et al. Learning label specific features for multi-label classification[C]// *Proceedings of the 2015 IEEE International Conference on Data Mining*. Atlantic City, New Jersey, IEEE, 2015:181-190.
- [21] ZHANG Jia, LUO Zhiming, LI Candong, et al. Manifold regularized discriminative feature selection for multi-label learning[J]. *Pattern Recognition*, 2019, 95:136-150.
- [22] HU Juncheng, LI Yonghao, GAO Wanfu, et al. Robust multi-label feature selection with dual-graph regularization[J]. *Knowledge-Based Systems*, 2020, 203:106126.
- [23] LI Yonghao, HU Liang, GAO Wanfu. Multi-label feature selection via robust flexible sparse regularization[J]. *Pattern Recognition*, 2023, 134:109074.
- [24] CAI Zhiling, ZHU William. Multi-label feature selection via feature manifold learning and sparsity regularization[J]. *International Journal of Machine Learning and Cybernetics*, 2018, 9:1321-1334.
- [25] GAO Wanfu, LI Yonghao, HU Liang. Multilabel feature selection with constrained latent structure shared term[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(3):1253-1262.
- [26] LI Yonghao, HU Liang, GAO Wanfu. Label correlations variation for robust multi-label feature selection[J]. *Information Sciences*, 2022, 609:1075-1097.
- [27] HAN Jiuqi, SUN Zhengya, HAO Hongwei. Selecting feature subset with sparsity and low redundancy for unsupervised learning[J]. *Knowledge-Based Systems*, 2015, 86:210-223.
- [28] ZHANG Yao, HUO Wei, TANG Jun. Multi-label feature selection via latent representation learning and dynamic graph constraints[J]. *Pattern Recognition*, 2024, 151:110411.
- [29] LIN Yaojin, HU Qinghua, LIU Jinghua, et al. Multi-label feature selection based on max-dependency and min-redundancy[J]. *Neurocomputing*, 2015, 168:92-103.
- [30] LEE J, KIM D W. SCLS: multi-label feature selection based on scalable criterion for large label set[J]. *Pattern Recognition*, 2017, 66:342-352.
- [31] HASHEMI A, DOWLATSHAHI M B, NEZAMABADI-POUR H. MFS-MCDM: multi-label feature selection using multi-criteria decision making[J]. *Knowledge-Based Systems*, 2020, 206:106365.
- [32] ZOU Yizhang, HU Xuegang, LI Peipei. Gradient-based multi-label feature selection considering three-way variable interaction[J]. *Pattern Recognition*, 2024, 145:109900.
- [33] XU Wei, GONG Yihong. Document clustering by concept factorization[C]// *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Sheffield: ACM, 2004:202-209.
- [34] FAN K. On a theorem of Weyl concerning eigenvalues of linear transformations I [J]. *Proceedings of the National Academy of Sciences*, 1949, 35(11):652-655.
- [35] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 1999, 401(6755):788-791.
- [36] HUANG Jin, NIE Feiping, HUANG Heng. A new simplex sparse learning model to measure data similarity for clustering[C]// *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. Burlington: Morgan Kaufmann, 2015:3569-3575.
- [37] LEE D D, SEUNG S H. Algorithms for non-negative matrix factorization[J]. *Advances in Neural Information Processing Systems*, 2001, 13:556-562.
- [38] DING C H Q, Li Tao, JORDAN M I. Convex and semi-nonnegative matrix factorizations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 32(1):45-55.
- [39] ZHANG Minling, ZHOU Zhihua. ML-KNN: a lazy learning approach to multi-label learning[J]. *Pattern Recognition*, 2007, 40(7):2038-2048.
- [40] DOUGHERTY J, KOHAVI R, SAHAMI M. Supervised and unsupervised discretization of continuous features [M]. Massachusetts: Morgan Kaufmann, 1995.
- [41] XUE Guotong, ZHONG Ming, LI Jianxin, et al. Dynamic network embedding survey[J]. *Neurocomputing*, 2022, 472:212-223.
- [42] DUNN O J. Multiple comparisons among means[J]. *Journal of the American statistical association*, 1961, 56(293):52-64.
- [43] FRIEDMAN M. A comparison of alternative tests of significance for the problem of rankings[J]. *The Annals of Mathematical Statistics*, 1940, 11(1):86-92.