

具有潜在表示和动态图约束的多标签特征选择

李坤¹, 刘婧², 齐赫¹

(1. 海口经济学院 中芯依智网络学院, 海口 570203; 2. 之江实验室, 杭州 311000)

摘要: 针对现有嵌入式方法忽略实例相关性的潜在表示对伪标记学习的影响以及固定的图矩阵导致计算误差随迭代的加深而不断增大的问题, 提出一种具有潜在表示和动态图约束的多标签特征选择方法. 该方法首先利用实例相关性的潜在表示构造伪标签矩阵, 并将其与线性映射和最小化伪标签与真实标签之间的Friedman范数距离相结合, 从而保证伪标签与真实标签之间具有较高的相似性. 其次, 利用伪标签的低维流形结构构建动态图, 以缓解固定图矩阵导致的随迭代深度增加计算误差的问题. 在12个数据集上与7种先进方法的对比实验结果表明, 该方法的整体分类性能优于现有先进方法, 能较好地处理多标记特征选择问题.

关键词: 多标签学习; 特征选择; 潜在表示; 动态图; 流形学习

中图分类号: TP181 **文献标志码:** A **文章编号:** 1671-5489(2024)05-1188-15

Multi-label Feature Selection with Latent Representation and Dynamic Graph Constraints

LI Kun¹, LIU Jing², QI He¹

(1. ZX-YZ School of Network Science, Haikou University of Economics, Haikou 570203, China;

2. Zhijiang Laboratory, Hangzhou 311000, China)

Abstract: Aiming at the problems that ignored by the existing embedded methods: the influence of the latent representation of instance correlation on pseudo-label learning, and the calculation error was caused by the fixed graph matrix, which increased with the deepening of iterations. We proposed a multi-label feature selection method with latent representation and dynamic graph constraints. Firstly, the proposed method used the latent representation of instance correlation to construct the pseudo-label matrix, and combined it with linear mapping and minimizing the Friedman norm distance between the pseudo-label and the ground-truth label to ensure a high similarity between pseudo-labels and the ground-truth labels. Secondly, the dynamic graph was constructed by using the low-dimensional manifold structure of pseudo-labels to alleviate the problem of increasing calculation error with iteration depth caused by a fixed graph matrix. The comparative experimental results with seven advance methods on 12 datasets show that the overall classification performance of the proposed method is superior to the existing advanced methods, and it can better deal with multi-label feature selection problems.

Keywords: multi-label learning; feature selection; latent representation; dynamic graph; manifold learning

多标签分类在图像分类^[1-2]、文本分类^[3-4]和生物信息学^[5]等领域应用广泛, 与传统的单标签分类任务不同, 多标签分类任务需考虑标签之间的相关性和复杂性. 在多标签分类中, 特征选择可以帮助提高分类器的准确性和效率. 因此, 多标签特征选择作为多标签分类问题中的一个重要预处理步骤, 已引起研究者的广泛关注^[6-7]. 随着机器学习技术的发展, 特征选择方法也在不断改进, 为多标签分类任务的应用提供了更完善和有效的解决方案. 目前, 已有多种方法解决多标签特征选择问题, 这些方法主要分为三类: 过滤式^[8-10]、封装式^[11-12]和嵌入式^[13-14]. 过滤式是一种基于特征之间相关性进行特征选择的方法, 常用的计算相关性方法包括互信息^[15-16]、卡方检验和 Pearson 相关系数^[17]等. 封装式是一种基于模型性能进行特征选择的方法, 它将特征选择视为一个搜索问题, 其中模型的性能用于评估函数. 嵌入式是将特征选择嵌入到分类模型的训练中, 使分类模型可以同时学习特征选择和分类任务.

本文主要研究嵌入式多标签特征选择方法的一个重要分支. 该分支基于线性映射, 即假设特征空间和真实标签空间(或伪标签空间)之间是线性相关的. 这类方法常利用特征和标签间线性映射的系数矩阵表示特征权重, 并采用 L_1 范数、 $L_{2,1}$ 范数等稀疏正则项约束模型的稀疏性. 此外, 由于流形学习^[18]的快速发展, 其在协同聚类算法^[19]、特征选择算法^[20-21]、维数简约^[22-24]等多领域应用广泛. 在特征选择算法中, 流形学习可为特征选择提供合适的图矩阵, 并指导特征权重矩阵的学习, 以确保映射空间与特征空间拓扑结构的一致性. 同时, 由于流形学习的引入, 利用 Laplace 变换探索特征流形结构或标签流形结构进行特征选择的方法得到了广泛关注^[25-26]. 例如, Hu 等^[27]使用特征流形和低维标签流形约束模型, 提出了 DRMFS (robust multi-label feature selection with dual-graph regularization), 在 DRMFS 中双图正则化约束的是特征权重矩阵而不是伪标签矩阵, 从而保证了特征权重矩阵的行和列分别拟合特征流形结构和低维标签流形结构. Huang 等^[28]为进一步挖掘伪标签与真实标签之间的相关性, 在特征选择模型中引入了 Hilbert-Schmidt 独立性准则^[29], 并结合特征流形和 $L_{2,1}$ 范数稀疏约束, 提出了 MRDM (multi-label feature selection via manifold regularization and dependence maximization). Li 等^[30]认为现有方法中的稀疏约束无法保证特征间的可区分性, 从而设计了一种鲁棒灵活稀疏正则化范数, 并结合低维标签流形, 提出了 RFSFS (multi-label feature selection via robust flexible sparse regularization).

但上述方法均忽略了潜在表示^[31]可以深度挖掘模型中的潜在监督信息. 因此, Gao 等^[32]在特征选择模型中引入了潜在表示, 并基于此构建了原始特征空间和原始标签空间的潜在共享空间作为模型的监督信息, 此外, 还利用特征流形保证潜在共享空间和原始特征空间流形结构的一致性, 再结合范数作为稀疏约束, 提出了 SSFS (multilabel feature selection with constrained latent structure shared term). 虽然 SSFS 通过引入潜在表示挖掘模型中的潜在监督信息, 但忽略了更深层的潜在信息, 即实例相关性的潜在表示对伪标签的影响. 上述方法均忽略了影响特征选择性能的一个因素, 即在流形学习中, 固定的图矩阵可能会随迭代的加深不断放大模型的计算误差. 为解决该问题, 研究者们相继提出了通过构建动态图代替固定图从而达到缓解模型误差累加的问题. 如 Li 等^[33]通过同时考虑潜在表示和动态图约束, 提出了 SLMDs (robust sparse and low-redundancy multi-label feature selection with dynamic local and global structure preservation). 虽然 SSFS^[32]和 SLMDs^[33]都考虑了潜在表示在特征选择中的应用, 但这些方法都仅考虑了特征或标签的潜在表示, 而忽略了实例相关性的潜在表示对伪标签学习的影响.

此外, 嵌入式多标签特征选择方法的另一个分支是基于信息论的方法, 这类方法常利用互信息量化特征重要性, 从而实现特征选择. 在这类方法中, 较经典的模型有 PMU (feature selection for multi-label classification using multivariate mutual information)^[34], MDMR (multi-label feature selection based on max-dependency and min-redundancy)^[35], FIMF (fast multi-label feature selection based on information-theoretic feature ranking)^[36]和 SCLS (multi-label feature selection based on scalable criterion for large label set)^[37]等. PMU 是利用多元互信息进行多标签分类的特征选择方法, 由于该方法只考虑了特征与标签之间的三维交互, 对高阶标签数据可能会丢失重要信息, 并且高阶多

元互信息的计算代价非常高. FIMF 是一种快速的特征选择方法, 该模型根据特征的重要性对特征进行排序, 以达到特征选择的目的. MDMR 通过互信息量化特征相关性和特征冗余性, 并约束特征相关性最大化和特征冗余性最小化. SCLS 是一种基于大标签集可扩展标准的特征选择方法, 但 SCLS 中的特性和标签的组合呈指数级增长, 使特性选择变得不切实际.

1 模型构建

特征选择的目标是利用给定的训练数据集 $D = \{\mathbf{X}, \mathbf{Y}\}$ 构建函数 $\Gamma: (\mathbf{X}, \mathbf{Y}) \rightarrow (\mathbf{X}^*, \mathbf{Y})$, 并通过函数 Γ 去除不相关和冗余特征, 选择出最具代表性的特征子集 \mathbf{X}^* , 从而提高分类器的性能, 其中 $\mathbf{X} \in \mathbb{R}^{n \times d}$ 为实例集, $\mathbf{Y} \in \mathbb{R}^{n \times m}$ 为真实标签集, n, d 和 m 分别表示实例数、特征数和标签数.

由于线性回归具有较强的可解释性, 且其损失函数便于求解, 因此本文用线性回归作为特征选择模型的基本框架:

$$\min_{\mathbf{W}, \mathbf{b}} \|\mathbf{X}\mathbf{W} + \mathbf{1}_n^T \mathbf{b} - \mathbf{Y}\|_{\text{F}}^2 + \gamma R(\mathbf{W}), \quad (1)$$

其中 $\mathbf{1}_n \in \mathbb{R}^n$ 为元素全为 1 的 n 维行向量; $\mathbf{b} \in \mathbb{R}^m$ 为 m 维行向量, 表示线性映射的偏置向量; $\mathbf{W} \in \mathbb{R}^{d \times m}$ 为权重矩阵, 可以量化特征重要性; $R(\mathbf{W})$ 为关于 \mathbf{W} 的惩罚函数, 用于约束 \mathbf{W} 的学习; γ 为关于 $R(\mathbf{W})$ 的惩罚参数.

考虑到真实标签的二值性与线性映射的平滑性不符, 因此本文引入伪标签作为真实标签的过渡变量拟合线性映射. 此外, 为保证伪标签包含真实标签的重要信息, 还需要约束伪标签和真实标签之间的相似性. 从而式(1)可重新表述为

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{V}, \mathbf{b}} \|\mathbf{X}\mathbf{W} + \mathbf{1}_n^T \mathbf{b} - \mathbf{V}\|_{\text{F}}^2 + \beta R(\mathbf{V}) + \gamma R(\mathbf{W}), \\ \text{s. t. } \mathbf{V} \rightarrow \mathbf{Y}, \quad \mathbf{V} \geq 0, \end{aligned} \quad (2)$$

其中: $\mathbf{V} \in \mathbb{R}^{n \times m}$ 为伪标签矩阵; $R(\mathbf{V})$ 为关于 \mathbf{V} 的惩罚函数, 用于约束 \mathbf{V} 的学习; β 为关于 $R(\mathbf{V})$ 的惩罚参数.

近年来, 由于潜在表示学习有利于许多数据挖掘和机器学习任务, 因此已引起研究者的广泛关注. 特别是在多标签特征选择任务中, 实例通常具有高维的特征和大规模的标签. 不同的实例具有不同的标签信息, 不同实例之间的潜在表示相互作用, 形成连接信息, 可用作伪标签学习的监督信息. 例如, SSFS^[32] 和 SLMDS^[34] 都倾向于使用实例和真实标签的潜在表示指导伪标签的学习, 但忽略了实例相关性的潜在表征对伪标签的影响.

如果两个实例 X_{i^*} 和 X_{j^*} 之间较相似, 则认为它们之间的潜在表示也相似, 从而存在一条连接路径, 反之则没有连接路径. 一般地, 连接信息的潜在表示可由一个对称的非负矩阵分解模型形成, 再结合特征选择模型, 可使用该模型将实例相似矩阵 \mathbf{S}_X 分解为低维潜在空间中的非负伪标签矩阵 \mathbf{V} 和 \mathbf{V}^T 的乘积, 用公式表示为

$$\begin{aligned} \min_{\mathbf{V}} \|\mathbf{S}_X - \mathbf{V}\mathbf{V}^T\|_{\text{F}}^2, \\ \text{s. t. } \mathbf{V} \geq 0, \end{aligned} \quad (3)$$

其中 \mathbf{S}_X 为实例相似矩阵, 其元素表示对应两实例间的相似度.

令 $R(\mathbf{V}) = \|\mathbf{S}_X - \mathbf{V}\mathbf{V}^T\|_{\text{F}}^2$, 则可将式(2)更新为

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{V}, \mathbf{b}} \|\mathbf{X}\mathbf{W} + \mathbf{1}_n^T \mathbf{b} - \mathbf{V}\|_{\text{F}}^2 + \beta \|\mathbf{S}_X - \mathbf{V}\mathbf{V}^T\|_{\text{F}}^2 + \gamma R(\mathbf{W}), \\ \text{s. t. } \mathbf{V} \rightarrow \mathbf{Y}, \quad \mathbf{V} \geq 0. \end{aligned} \quad (4)$$

此外, \mathbf{W} 作为特征权值矩阵, 在模型中起至关重要的作用, 特别是在后续的特征选择过程中, \mathbf{W} 的优劣会直接影响模型的性能. 由于特征流形可以保证 \mathbf{W} 具有类似于原始特征空间的流形结构, 因此近年来越来越多的模型使用特征流形约束 \mathbf{W} 的学习. 但这些使用特征流形的方法却忽略了一个关键问题, 即使用固定的 Laplace 图矩阵构造特征流形模型, 而忽略了图矩阵在特征选择中的动态变化. 并且不同方式学习到的图矩阵对后续的更新策略有不同的影响, 从而导致模型的局限性.

为解决该问题, 本文利用伪标签的低维流形结构构造一个动态 Laplace 图矩阵, 并利用该动态图

约束特征权值矩阵的学习. 根据问题假设: 若伪标签中 \mathbf{V}_{*i} 和 \mathbf{V}_{*j} 较相似, 则权重矩阵中 \mathbf{W}_{*i} 和 \mathbf{W}_{*j} 也应该较相似, 从而有

$$\sum_{i=1}^m \sum_{j=1}^m \|\mathbf{W}_{*i} - \mathbf{W}_{*j}\|_2^2 (\mathbf{S}_{\mathbf{V}^T})_{ij} = \text{tr}(\mathbf{W}\mathbf{L}_{\mathbf{V}^T}\mathbf{W}^T), \tag{5}$$

其中: $\text{tr}(\cdot)$ 表示矩阵的迹; \mathbf{W}_{*i} 和 \mathbf{W}_{*j} 分别表示 \mathbf{W} 的第 i 列和第 j 列向量; $\mathbf{L}_{\mathbf{V}^T} \in \mathbb{R}^{m \times m}$ 表示关于 $\mathbf{S}_{\mathbf{V}^T}$ 的 Laplace 矩阵, $\mathbf{L}_{\mathbf{V}^T} = \mathbf{P}_{\mathbf{V}^T} - \mathbf{S}_{\mathbf{V}^T}$, $\mathbf{P}_{\mathbf{V}^T} \in \mathbb{R}^{m \times m}$ 为对角矩阵, $(\mathbf{P}_{\mathbf{V}^T})_{ii} = \sum_{j=1}^m (\mathbf{S}_{\mathbf{V}^T})_{ij}$, $\mathbf{S}_{\mathbf{V}^T} \in \mathbb{R}^{m \times m}$ 为伪标签矩阵 \mathbf{V} 的低维相似矩阵, $(\mathbf{S}_{\mathbf{V}^T})_{ij}$ 为第 i 行、第 j 列元素, 表示 \mathbf{V}_{*i} 和 \mathbf{V}_{*j} 的相似度. 本文使用高斯核函数^[38]计算 $\mathbf{S}_{\mathbf{V}^T}$:

$$(\mathbf{S}_{\mathbf{V}^T})_{ij} = \begin{cases} \exp\left\{-\frac{\|\mathbf{V}_{*i} - \mathbf{V}_{*j}\|_2^2}{\sigma}\right\}, & \mathbf{V}_{*i} \in N_K(\mathbf{V}_{*j}) \text{ 或 } \mathbf{V}_{*j} \in N_K(\mathbf{V}_{*i}), \\ 0, & \text{其他,} \end{cases} \tag{6}$$

其中: $\sigma \in \mathbb{R}$ 表示高斯核函数的参数; \mathbf{V}_{*i} 和 \mathbf{V}_{*j} 分别表示 \mathbf{V} 的第 i 列和第 j 列向量; $N_K(\cdot)$ 表示矩阵的 K 近邻, K 是近邻参数.

令 $R(\mathbf{W}) = \text{tr}(\mathbf{W}\mathbf{L}_{\mathbf{V}^T}\mathbf{W}^T)$, 则可将式(4)更新为

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{V}, \mathbf{b}} & \|\mathbf{X}\mathbf{W} + \mathbf{1}_n^T \mathbf{b} - \mathbf{V}\|_{\mathbb{F}}^2 + \beta \|\mathbf{S}_X - \mathbf{V}\mathbf{V}^T\|_{\mathbb{F}}^2 + \gamma \text{tr}(\mathbf{W}\mathbf{L}_{\mathbf{V}^T}\mathbf{W}^T), \\ \text{s. t. } & \mathbf{V} \rightarrow \mathbf{Y}, \quad \mathbf{V} \geq 0. \end{aligned} \tag{7}$$

为保证 \mathbf{V} 的非负性, 需要约束 \mathbf{W} 也是非负的, 从而构建 LRLDG 模型的最终目标函数为

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{V}, \mathbf{b}} & \|\mathbf{X}\mathbf{W} + \mathbf{1}_n^T \mathbf{b} - \mathbf{V}\|_{\mathbb{F}}^2 + \alpha \|\mathbf{Y} - \mathbf{V}\|_{\mathbb{F}}^2 + \beta \|\mathbf{S}_X - \mathbf{V}\mathbf{V}^T\|_{\mathbb{F}}^2 + \gamma \text{tr}(\mathbf{W}\mathbf{L}_{\mathbf{V}^T}\mathbf{W}^T), \\ \text{s. t. } & \mathbf{V} \geq 0, \quad \mathbf{W} \geq 0, \end{aligned} \tag{8}$$

其中 α 为正则项参数.

2 优化求解

本文设计一个简单而有效的优化方案求解目标函数(8). 考虑到目标函数(8)中的两个变量有非负约束, 因此使用非负矩阵分解(nonnegative matrix factorization, NMF)^[39]和交替迭代的方法求解目标函数(8).

首先, 固定 \mathbf{W} 和 \mathbf{V} 求解 \mathbf{b} . 当 \mathbf{W} 和 \mathbf{V} 固定时, 式(8)是关于 \mathbf{b} 的凸函数, 所以求出式(8)关于 \mathbf{b} 的导函数, 并令导函数为 0, 则可得 \mathbf{b} 的更新公式为

$$\mathbf{b} = \frac{1}{n}(\mathbf{1}_n \mathbf{V} - \mathbf{1}_n \mathbf{X}\mathbf{W}). \tag{9}$$

其次, 固定 \mathbf{b} 求解 \mathbf{W} 和 \mathbf{V} . 由于对任意矩阵 \mathbf{M} , 存在 $\|\mathbf{M}\|_{\mathbb{F}}^2 = \text{tr}(\mathbf{M}\mathbf{M}^T)$, 所以可将式(9)代入目标函数(8), 并使用矩阵迹的形式重新表述目标函数(8):

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{V}, \mathbf{b}} & \{ \text{tr}((\mathbf{X}\mathbf{W} - \mathbf{V})^T \mathbf{H}(\mathbf{X}\mathbf{W} - \mathbf{V})) + \alpha \text{tr}((\mathbf{Y} - \mathbf{V})^T (\mathbf{Y} - \mathbf{V})) + \\ & \beta \text{tr}((\mathbf{S}_X - \mathbf{V}\mathbf{V}^T)^T (\mathbf{S}_X - \mathbf{V}\mathbf{V}^T)) + \gamma \text{tr}(\mathbf{W}\mathbf{L}_{\mathbf{V}^T}\mathbf{W}^T) \}, \\ \text{s. t. } & \mathbf{V} \geq 0, \quad \mathbf{W} \geq 0, \end{aligned} \tag{10}$$

其中 $\mathbf{H} = \mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$ 是中心矩阵, $\mathbf{I} \in \mathbb{R}^{n \times n}$ 是 n 维单位矩阵.

考虑到变量 \mathbf{W} 和 \mathbf{V} 都具有非负约束, 因此在求解过程中, 需要将变量 \mathbf{W} 和 \mathbf{V} 的非负约束整合到目标函数中, 然后才能继续进行 NMF 过程. 构造以下 Lagrange 公式:

$$\begin{aligned} L(\mathbf{W}, \mathbf{V}) = & \text{tr}((\mathbf{X}\mathbf{W} - \mathbf{V})^T \mathbf{H}(\mathbf{X}\mathbf{W} - \mathbf{V})) + \alpha \text{tr}((\mathbf{Y} - \mathbf{V})^T (\mathbf{Y} - \mathbf{V})) + \\ & \beta \text{tr}((\mathbf{S}_X - \mathbf{V}\mathbf{V}^T)^T (\mathbf{S}_X - \mathbf{V}\mathbf{V}^T)) + \gamma \text{tr}(\mathbf{W}\mathbf{L}_{\mathbf{V}^T}\mathbf{W}^T) - \text{tr}(\mathbf{\Phi}\mathbf{V}^T) - \text{tr}(\mathbf{\Psi}\mathbf{W}^T), \end{aligned} \tag{11}$$

其中 $L(\mathbf{W}, \mathbf{V})$ 表示关于 \mathbf{W} 和 \mathbf{V} 的 Lagrange 函数, $\mathbf{\Phi} \in \mathbb{R}^{n \times m}$ 和 $\mathbf{\Psi} \in \mathbb{R}^{d \times m}$ 均为 Lagrange 乘子.

下面进行 NMF 过程. 首先, 分别计算式(11)关于 \mathbf{W} 和 \mathbf{V} 的导函数:

$$\begin{cases} \frac{\partial L(\mathbf{W})}{\partial \mathbf{W}} = 2\mathbf{X}^T \mathbf{H}\mathbf{X}\mathbf{W} - 2\mathbf{X}^T \mathbf{H}\mathbf{V} + 2\gamma \mathbf{W}\mathbf{L}_{\mathbf{V}^T} - \boldsymbol{\Psi}, \\ \frac{\partial L(\mathbf{V})}{\partial \mathbf{V}} = 2\mathbf{H}\mathbf{V} - 2\mathbf{H}\mathbf{X}\mathbf{W} + 2\alpha \mathbf{V} - 2\alpha \mathbf{Y} + 4\beta \mathbf{V}\mathbf{V}^T \mathbf{V} - 4\beta \mathbf{S}_x \mathbf{V} + 4\mathbf{V}\mathbf{G} - \boldsymbol{\Phi}, \end{cases} \quad (12)$$

其中

$$\mathbf{G} = \text{Diag}\left(-\frac{1}{2\sigma} \mathbf{S}_{\mathbf{V}^T} \circ (\text{diag}(\mathbf{W}^T \mathbf{W}) \mathbf{E}^T + \mathbf{E} \text{diag}(\mathbf{W}^T \mathbf{W})^T - 2\mathbf{W}^T \mathbf{W})\right), \quad (13)$$

$\text{Diag}(\cdot)$ 和 $\text{diag}(\cdot)$ 分别表示向量对角矩阵化和矩阵对角向量化, \mathbf{E} 表示元素全为 1 的 m 维列向量.

其次, 通过 KKT(Karush-Kuhn-Tucker) 定理^[40] 的互补松弛条件可得: $\Phi_{ij} V_{ij} = 0$, $\Psi_{ij} W_{ij} = 0$. 从而可得

$$\begin{cases} (2\mathbf{X}^T \mathbf{H}\mathbf{X}\mathbf{W} - 2\mathbf{X}^T \mathbf{H}\mathbf{V} + 2\gamma \mathbf{W}\mathbf{L}_{\mathbf{V}^T}) \circ \boldsymbol{\Psi} = 0, \\ (2\mathbf{H}\mathbf{V} - 2\mathbf{H}\mathbf{X}\mathbf{W} + 2\alpha \mathbf{V} - 2\alpha \mathbf{Y} + 4\beta \mathbf{V}\mathbf{V}^T \mathbf{V} - 4\beta \mathbf{S}_x \mathbf{V} + 4\mathbf{V}\mathbf{G}) \circ \boldsymbol{\Phi} = 0, \end{cases} \quad (14)$$

其中 \circ 表示两矩阵的 Hadamard 积.

最后, 根据式(14)和 NMF 的计算规则可得 \mathbf{W} 和 \mathbf{V} 的更新公式:

$$\begin{cases} \mathbf{V} \leftarrow \frac{\mathbf{H}\mathbf{X}\mathbf{W} + \alpha \mathbf{Y} + 2\beta \mathbf{S}_x \mathbf{V} - 2\gamma \mathbf{V} \text{diag}(\mathbf{U}\mathbf{E})}{\mathbf{H}\mathbf{V} + \alpha \mathbf{V} + 2\beta \mathbf{V}\mathbf{V}^T \mathbf{V} - 2\gamma \mathbf{V}\mathbf{U}}, \\ \mathbf{W} \leftarrow \frac{\mathbf{X}^T \mathbf{H}\mathbf{V} + \gamma \mathbf{W}\mathbf{S}_{\mathbf{V}^T}}{\mathbf{X}^T \mathbf{H}\mathbf{X}\mathbf{W} + \gamma \mathbf{W}\mathbf{P}_{\mathbf{V}^T}}, \end{cases} \quad (15)$$

其中

$$\mathbf{U} = -\frac{1}{2\sigma} \mathbf{S}_{\mathbf{V}^T} \circ (\text{diag}(\mathbf{W}^T \mathbf{W}) \mathbf{E}^T + \mathbf{E} \text{diag}(\mathbf{W}^T \mathbf{W})^T - 2\mathbf{W}^T \mathbf{W}). \quad (16)$$

至此已完成了对目标函数(8)的求解. 此外, 本文还设计了如下一种新的具有收敛性的交替迭代算法对模型进行优化求解.

算法 1 LRLDG 算法.

输入: 训练数据集 $D = \{\mathbf{X}, \mathbf{Y}\}$, 特征选择数量 k , 超参数 α, β 和 γ ;

输出: 特征选择结果 \mathbf{F}_s ;

1) 初始化中心矩阵 $\mathbf{H} = \mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$;

2) 随机初始化 \mathbf{W} 和 \mathbf{V} ;

3) 初始化迭代次数 $t = 0$;

重复:

根据式(6)计算 $\mathbf{P}_{\mathbf{V}^T}$ 和 $\mathbf{S}_{\mathbf{V}^T}$;

更新 \mathbf{U} : $\mathbf{U} = -\frac{1}{2\sigma} \mathbf{S}_{\mathbf{V}^T} \circ (\text{diag}(\mathbf{W}^T \mathbf{W}) \mathbf{E}^T + \mathbf{E} \text{diag}(\mathbf{W}^T \mathbf{W})^T - 2\mathbf{W}^T \mathbf{W})$;

更新 \mathbf{W} 和 \mathbf{V} :
$$\begin{cases} \mathbf{V} \leftarrow \frac{\mathbf{H}\mathbf{X}\mathbf{W} + \alpha \mathbf{Y} + 2\beta \mathbf{S}_x \mathbf{V} - 2\gamma \mathbf{V} \text{diag}(\mathbf{U}\mathbf{E})}{\mathbf{H}\mathbf{V} + \alpha \mathbf{V} + 2\beta \mathbf{V}\mathbf{V}^T \mathbf{V} - 2\gamma \mathbf{V}\mathbf{U}}, \\ \mathbf{W} \leftarrow \frac{\mathbf{X}^T \mathbf{H}\mathbf{V} + \gamma \mathbf{W}\mathbf{S}_{\mathbf{V}^T}}{\mathbf{X}^T \mathbf{H}\mathbf{X}\mathbf{W} + \gamma \mathbf{W}\mathbf{P}_{\mathbf{V}^T}}; \end{cases}$$

更新 \mathbf{b} : $\mathbf{b} = \frac{1}{n} (\mathbf{1}_n \mathbf{V} - \mathbf{1}_n \mathbf{X}\mathbf{W})$;

更新 t : $t = t + 1$;

直到收敛为止;

4) 计算并排序 $\|\mathbf{W}_{i*}\|_2$, 找出前 k 个最大值的索引赋给 \mathbf{F}_s .

3 算法收敛性证明

下面给出 LRLDG 算法的收敛性证明.

定义 1^[41] 如果 $Z(\mathbf{W}, \mathbf{W}')$ 和 $F(\mathbf{W})$ 满足以下条件:

$$Z(\mathbf{W}, \mathbf{W}) = F(\mathbf{W}), \quad Z(\mathbf{W}, \mathbf{W}') \geq F(\mathbf{W}), \tag{17}$$

则 $Z(\mathbf{W}, \mathbf{W}')$ 是 $F(\mathbf{W})$ 的一个辅助函数, 其中 \mathbf{W}' 是 \mathbf{W} 在上一步迭代中的更新值.

引理 1^[41] 如果定理 1 成立, 则 $F(\mathbf{W})$ 基于以下形式单调递减:

$$\mathbf{W}^{t+1} = \underset{\mathbf{W}}{\operatorname{argmin}} Z(\mathbf{W}, \mathbf{W}'), \tag{18}$$

其中 $\underset{\mathbf{W}}{\operatorname{argmin}} Z(\mathbf{W}, \mathbf{W}')$ 表示第 t 次迭代的最优解.

命题 1^[42] 对任何矩阵 $\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{d \times d}$, $\mathbf{W}, \mathbf{W}' \in \mathbb{R}^{d \times m}$. 如果 $\mathbf{A} \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{d \times d}$ 都是对称矩阵, 则以下不等式成立:

$$\sum_{i=1}^n \sum_{j=1}^m \frac{(\mathbf{B}\mathbf{W}'\mathbf{A})_{ij}}{W'_{ij}} \geq \operatorname{tr}(\mathbf{W}^T \mathbf{B}\mathbf{W}\mathbf{A}). \tag{19}$$

根据命题 1, 易得如下引理.

引理 2 如果命题 1 中的 $\mathbf{A} \in \mathbb{R}^{m \times m}$ 是一个 m 维单位矩阵, 则以下不等式成立:

$$\sum_{i=1}^n \sum_{j=1}^m \frac{(\mathbf{B}\mathbf{W}')_{ij}}{W'_{ij}} \geq \operatorname{tr}(\mathbf{W}^T \mathbf{B}\mathbf{W}). \tag{20}$$

考虑到 LRLDG 算法关于 \mathbf{W} 和 \mathbf{V} 的收敛性证明类似, 所以本文仅给出关于 \mathbf{W} 的收敛性证明. 结合式(8), 可将定义 1 中的 $F(\mathbf{W})$ 改写为

$$F(\mathbf{W}) = \frac{1}{2} \operatorname{tr}(\mathbf{W}^T \boldsymbol{\Omega}\mathbf{W}) - \frac{1}{2} \operatorname{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{H}\mathbf{V}) + \frac{1}{2} \gamma \operatorname{tr}(\mathbf{W}\mathbf{P}_{\mathbf{V}^T} \mathbf{W}^T) - \frac{1}{2} \gamma \operatorname{tr}(\mathbf{W}\mathbf{S}_{\mathbf{V}^T} \mathbf{W}^T), \tag{21}$$

其中 $\boldsymbol{\Omega} = \mathbf{X}^T \mathbf{H}\mathbf{X}$ 是对称矩阵. 根据式(21), 可给出辅助函数 $Z(\mathbf{W}, \mathbf{W}')$ 的具体形式为

$$\begin{aligned} Z(\mathbf{W}, \mathbf{W}') &= \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m \frac{(\boldsymbol{\Omega}\mathbf{W}')_{ij} W'_{ij}}{W'_{ij}} - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m (\mathbf{X}^T \mathbf{H}\mathbf{V})_{ij} W_{ij} + \\ &\quad \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m (\gamma \mathbf{W}\mathbf{P}_{\mathbf{V}^T})_{ij} W_{ij} - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m (\gamma \mathbf{W}\mathbf{S}_{\mathbf{V}^T})_{ij} W_{ij}. \end{aligned} \tag{22}$$

由引理 1 可知, 若 $Z(\mathbf{W}, \mathbf{W}')$ 是 $F(\mathbf{W})$ 的一个辅助函数, 则 LRLDG 算法关于 \mathbf{W} 是收敛的. 因此, 只需证明 $Z(\mathbf{W}, \mathbf{W}')$ 是 $F(\mathbf{W})$ 的一个辅助函数即可.

首先, 根据矩阵迹的计算规则可得

$$\begin{aligned} \sum_{i=1}^d \sum_{j=1}^m (\mathbf{X}^T \mathbf{H}\mathbf{V})_{ij} W_{ij} &= \operatorname{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{H}\mathbf{V}), \quad \sum_{i=1}^d \sum_{j=1}^m (\gamma \mathbf{W}\mathbf{P}_{\mathbf{V}^T})_{ij} W_{ij} = \gamma \operatorname{tr}(\mathbf{W}\mathbf{P}_{\mathbf{V}^T} \mathbf{W}^T), \\ \sum_{i=1}^d \sum_{j=1}^m (\gamma \mathbf{W}\mathbf{S}_{\mathbf{V}^T})_{ij} W_{ij} &= \gamma \operatorname{tr}(\mathbf{W}\mathbf{S}_{\mathbf{V}^T} \mathbf{W}^T). \end{aligned} \tag{23}$$

其次, 根据引理 2 可得

$$\sum_{i=1}^d \sum_{j=1}^m \frac{(\boldsymbol{\Omega}\mathbf{W}')_{ij} W'_{ij}}{W'_{ij}} \geq \operatorname{tr}(\mathbf{W}^T \boldsymbol{\Omega}\mathbf{W}). \tag{24}$$

结合式(23)和式(24), 可知 $Z(\mathbf{W}, \mathbf{W}') \geq F(\mathbf{W})$ 成立. 此外, 当 $\mathbf{W} = \mathbf{W}'$ 时, 根据矩阵迹的计算规则, 式(24)可重写为

$$\sum_{i=1}^d \sum_{j=1}^m \frac{(\boldsymbol{\Omega}\mathbf{W}')_{ij} W_{ij}}{W_{ij}} = \sum_{i=1}^d \sum_{j=1}^m (\boldsymbol{\Omega}\mathbf{W})_{ij} W_{ij} = \operatorname{tr}(\mathbf{W}^T \boldsymbol{\Omega}\mathbf{W}). \tag{25}$$

结合式(23)和式(25), 可知当 $\mathbf{W} = \mathbf{W}'$ 时, $Z(\mathbf{W}, \mathbf{W}) = F(\mathbf{W})$ 成立.

由定理 1 可知 $Z(\mathbf{W}, \mathbf{W}')$ 是 $F(\mathbf{W})$ 的一个辅助函数, 再由引理 1 可知, LRLDG 算法关于 \mathbf{W} 是收敛的. 此外, 满足 $\frac{\partial Z(\mathbf{W}, \mathbf{W}')}{\partial \mathbf{W}} = 0$ 的 \mathbf{W}^* 是 \mathbf{W} 的全局最优解. 综上所述, LRLDG 算法的收敛性得证.

4 实 验

下面将 LRLDG 算法与 7 个现有的先进算法 PMU^[34], SCLS^[37], FIMF^[36], MDMR^[35], SSFS^[32], RFSFS^[30] 和 SLMDS^[33] 进行对比实验, 并利用统计检验的方式对比所有特征选择方法的整体性能.

此外,为深度剖析本文方法的性能,还设计了针对 LRLDG 算法的一系列实验,包括参数敏感性实验和收敛性实验等.

4.1 数据集

对比实验中,采用从木兰图书馆^[43]获取的 12 个来自 5 个不同领域的经典多标签数据集,其中包括音乐数据集 Emotion,生物数据集 Yeast 和 Genbase,声音数据集 Birds 和 CAL500,图像数据集 Scene,Image 和 Flags,文本数据集 Enron,Computer,Recreation 和 Social. 表 1 列出了所有数据集的具体参数.

表 1 各数据集参数
Table 1 Parameters of each dataset

数据集	实例数	特征数	标签数	训练集	测试集	领域
Emotion	593	72	6	391	202	音乐
Yeast	2 417	103	14	1 500	917	生物
Genbase	662	1 185	27	463	199	生物
Birds	645	260	19	322	323	声音
CAL500	502	68	174	335	167	声音
Scene	2 407	294	6	1 211	1 196	图像
Image	600	294	5	400	200	图像
Flags	194	19	7	129	65	图像
Enron	1 702	1 001	53	1 123	579	文本
Computer	5 000	681	33	2 000	3 000	文本
Recreation	5 000	606	22	2 000	3 000	文本
Social	5 000	1 047	39	2 000	3 000	文本

4.2 实验设置

实验选用 ML-KNN^[44]作为多标签分类算法的代表对各特征选择算法的性能进行评价,并将 ML-KNN 的参数设置为 $S=1, K=10$. 选用 5 个多标签评价指标 AP(average precision), MA(macro- F_1), MI(micro- F_1), HL(Hamming loss)和 RL(ranking loss)^[45]量化并对比各特征选择算法的性能.

为使数据集适用于 FIMF 和 SCLS 等算法,采用等宽区间对数据进行离散化处理^[46]. 对于 SSFS, RFSFS 和 LRLDG 算法,将其近邻参数设为 5. 此外,由于 SSFS, RFSFS, SLMDS 和 LRLDG 算法存在随机初始化变量的问题,为保证实验结果的准确性,因此本文选取各算法 10 次实验结果的均值作为最终实验结果. 最后,采用网格搜索策略^[47]对各算法的正则化参数进行调整,并将调整范围设为 $[0.001, 0.01, 0.1, 1, 10, 100, 1 000]$,将在各数据集上选择的特征数量设为 2%~20%,步长为 2%,其中由于数据集 Flags 的原始特征数为 19,所以在数据集 Flags 上选择的特征数量设为 2~18.

为保证实验环境的公平性,所有实验均在 Windows11 系统,ADM 锐龙 495 5 3600 6 核处理器 3.59 GHz,内存 16.00 GB 环境中完成,并使用 MATLAB R2016a 实现.

4.3 实验结果分析

下面给出并分析所有特征选择方法在各评价指标上的对比实验结果.

表 2~表 6 列出了各特征选择方法在其最优参数下的最优实验结果,其中 SSFS, RFSFS, SLMDS 和 LRLDG 算法给出的是 10 次重复实验结果的均值. 为方便观察,在表 2~表 6 中,“()”中给出各实验结果的简单排序结果,并在表中最后一行给出各特征选择方法在同一指标下所有实验结果的均值,用于观察各特征选择方法的整体性能. 由表 2~表 6 可见, LRLDG 算法在所有指标下的整体性能均为最优,表明与对比方法相比, LRLDG 算法在处理多标签特征选择问题上有一定的优越性.

表 2 和表 3 分别列出了所有方法在 HL 和 RL 指标上的实验结果,其值越小表示算法的性能越好. 由表 2 可见, LRLDG 算法在数据集 Image 上的性能相对于次优算法 SLMDS 的性能提升了 3.93%,在数据集 Recreation 上的性能相对于次优算法 SSFS 的性能提升了 4.03%. 由表 3 可见,相对于性能次优的 MDMR 算法, LRLDG 算法在数据集 Flags 上的性能提升了 4.3%.

表 2 各数据集上不同算法的 HL 对比

Table 2 HL comparison of different algorithms on each dataset

数据集及均值	PMU	SCLS	FIMF	MDMR	SSFS	RFSFS	SLMDS	LRLDG
Yeast	0.209 7(5)	0.214 8(6)	0.220 9(7)	0.208 4(3)	0.220 9(7)	0.209 1(4)	0.202 2(2)	0.200 6(1)
Emotion	0.278 1(8)	0.242 6(4)	0.253 3(5)	0.259 1(7)	0.253 3(5)	0.224 8(3)	0.223 1(2)	0.219 5(1)
Birds	0.050 4(7)	0.050 2(6)	0.049 5(4)	0.050 5(8)	0.049 5(4)	0.047 9(3)	0.046 2(1)	0.047 5(2)
Scene	0.115 0(3)	0.132 5(6)	0.169 3(8)	0.132 5(6)	0.129 8(5)	0.126 0(4)	0.103 5(2)	0.100 8(1)
Image	0.231 0(8)	0.227 0(6)	0.230 0(7)	0.217 0(4)	0.218 8(5)	0.210 8(3)	0.203 4(2)	0.195 4(1)
Flags	0.593 4(4)	0.633 0(6)	0.593 4(4)	0.593 4(4)	0.641 3(7)	0.646 2(8)	0.588 6(2)	0.571 0(1)
Genbase	0.004 8(4)	0.032 0(6)	0.042 8(7)	0.002 6(1)	0.042 8(7)	0.002 9(2)	0.004 8(4)	0.003 5(3)
CAL500	0.966 0(6)	0.965 3(4)	0.966 4(8)	0.966 0(6)	0.965 8(5)	0.963 8(1)	0.963 9(2)	0.964 0(3)
Enron	0.050 0(5)	0.051 7(6)	0.052 6(8)	0.049 8(4)	0.049 5(3)	0.052 2(7)	0.049 3(2)	0.048 4(1)
Computer	0.040 3(5)	0.040 4(6)	0.040 4(6)	0.039 8(4)	0.040 9(8)	0.039 4(3)	0.039 3(2)	0.038 7(1)
Recreation	0.061 3(6)	0.063 5(7)	0.063 7(8)	0.060 9(5)	0.059 6(2)	0.060 0(3)	0.060 8(4)	0.057 2(1)
Social	0.026 2(8)	0.025 3(6)	0.025 6(7)	0.024 3(3)	0.023 5(2)	0.024 6(4)	0.024 9(5)	0.023 2(1)
均值	0.218 9	0.223 2	0.225 7	0.217 0	0.224 6	0.217 3	0.209 2	0.205 8

表 3 各数据集上不同算法的 RL 对比

Table 3 RL comparison of different algorithms on each dataset

数据集及均值	PMU	SCLS	FIMF	MDMR	SSFS	RFSFS	SLMDS	LRLDG
Yeast	0.188 0(5)	0.193 5(6)	0.199 7(7)	0.184 5(3)	0.199 7(7)	0.185 1(4)	0.177 2(2)	0.174 1(1)
Emotion	0.257 0(8)	0.199 1(4)	0.218 1(5)	0.225 0(7)	0.218 1(5)	0.198 1(3)	0.187 7(2)	0.187 2(1)
Birds	0.258 5(7)	0.155 3(1)	0.212 3(5)	0.259 1(8)	0.212 3(5)	0.207 0(4)	0.168 2(2)	0.184 5(3)
Scene	0.134 0(3)	0.165 2(7)	0.240 9(8)	0.137 1(4)	0.148 6(6)	0.142 2(5)	0.099 1(2)	0.098 4(1)
Image	0.284 6(8)	0.250 0(5)	0.269 2(7)	0.260 4(6)	0.228 0(4)	0.218 9(3)	0.208 6(2)	0.201 2(1)
Flags	0.197 7(5)	0.248 2(8)	0.182 3(3)	0.181 3(2)	0.207 8(6)	0.210 2(7)	0.183 3(4)	0.173 5(1)
Genbase	0.008 2(5)	0.049 9(6)	0.080 4(7)	0.006 2(1)	0.080 4(7)	0.007 9(4)	0.007 1(3)	0.006 4(2)
CAL500	0.181 9(4)	0.183 4(7)	0.183 5(8)	0.181 9(4)	0.183 2(6)	0.181 7(2)	0.181 7(2)	0.181 7(2)
Enron	0.094 2(5)	0.095 8(7)	0.099 7(8)	0.088 7(2)	0.089 9(3)	0.095 4(6)	0.091 9(4)	0.088 1(1)
Computer	0.096 5(8)	0.091 8(4)	0.093 6(6)	0.093 6(6)	0.093 6(6)	0.091 2(2)	0.091 6(3)	0.088 0(1)
Recreation	0.189 3(6)	0.192 8(7)	0.196 0(8)	0.184 2(5)	0.176 0(3)	0.172 3(2)	0.183 1(4)	0.171 0(1)
Social	0.065 3(5)	0.067 3(7)	0.071 8(8)	0.063 7(3)	0.060 9(1)	0.064 0(4)	0.067 1(6)	0.063 1(2)
均值	0.162 9	0.157 7	0.170 6	0.155 5	0.158 2	0.147 8	0.137 2	0.134 8

表 4 各数据集上不同算法的 AP 对比

Table 4 AP comparison of different algorithms on each dataset

数据集及均值	PMU	SCLS	FIMF	MDMR	SSFS	RFSFS	SLMDS	LRLDG
Yeast	0.737 1(5)	0.730 7(6)	0.722 3(7)	0.741 6(3)	0.722 3(7)	0.741 1(4)	0.751 4(2)	0.756 4(1)
Emotion	0.714 3(8)	0.764 0(4)	0.749 1(5)	0.734 9(7)	0.749 1(5)	0.766 5(3)	0.779 4(2)	0.780 5(1)
Birds	0.443 4(7)	0.574 0(2)	0.514 3(5)	0.414 0(8)	0.514 3(5)	0.514 5(4)	0.577 4(1)	0.543 9(3)
Scene	0.799 1(3)	0.760 8(7)	0.655 4(8)	0.777 3(5)	0.769 4(6)	0.779 9(4)	0.833 3(2)	0.838 1(1)
Image	0.675 5(8)	0.694 4(6)	0.679 8(7)	0.697 9(5)	0.734 8(4)	0.737 6(3)	0.749 5(2)	0.761 4(1)
Flags	0.841 1(4)	0.802 4(8)	0.841 0(6)	0.846 2(3)	0.837 2(7)	0.841 1(4)	0.856 4(2)	0.860 6(1)
Genbase	0.990 3(5)	0.734 2(6)	0.619 1(7)	0.993 9(1)	0.619 1(7)	0.990 4(4)	0.992 3(3)	0.992 5(2)
CAL500	0.494 3(6)	0.493 7(7)	0.491 5(8)	0.496 5(4)	0.494 4(5)	0.498 6(1)	0.497 3(3)	0.498 3(2)
Enron	0.649 5(5)	0.637 6(6)	0.631 0(8)	0.657 1(4)	0.660 7(2)	0.633 1(7)	0.658 2(3)	0.669 8(1)
Computer	0.625 5(6)	0.631 8(5)	0.620 4(7)	0.632 5(3)	0.620 4(7)	0.639 7(2)	0.632 3(4)	0.641 7(1)
Recreation	0.463 9(6)	0.447 1(7)	0.432 0(8)	0.480 7(5)	0.506 0(2)	0.505 3(3)	0.484 0(4)	0.537 0(1)
Social	0.702 2(6)	0.699 1(7)	0.686 4(8)	0.713 4(3)	0.727 9(1)	0.713 3(4)	0.703 3(5)	0.727 6(2)
均值	0.678 0	0.664 2	0.636 9	0.682 2	0.663 0	0.696 8	0.709 6	0.717 3

表 4~表 6 分别给出了各算法在 AP, MA 和 MI 指标上的最优结果, 其值越大表示算法的性能越

好. 在表 4~表 6 的 36 组对比实验中, LRLDG 算法获得了 25 个最优和 5 个次优的结果. 表明 LRLDG 算法相对于其他对比算法有一定的优越性. 在数据集 Recreation 上, LRLDG 算法在 AP, MA 和 MI 指标上的性能相对次优算法 SSFS 的性能分别提升了 6.13%, 26.01% 和 30.45%. 由表 5 可见, LRLDG 算法在数据集 Computer 上的性能比次优算法 RFSFS 的性能提升了 11.74%.

表 5 各数据集上不同算法的 MA 对比

Table 5 MA comparison of different algorithms on each dataset

数据集及均值	PMU	SCLS	FIMF	MDMR	SSFS	RFSFS	SLMDS	LRLDG
Yeast	0.312 0(5)	0.279 3(6)	0.245 0(7)	0.317 4(3)	0.245 0(7)	0.313 3(4)	0.347 3(2)	0.357 9(1)
Emotion	0.385 6(8)	0.493 1(4)	0.484 9(5)	0.455 7(7)	0.484 9(5)	0.594 6(3)	0.595 6(2)	0.607 6(1)
Birds	0.038 1(4)	0.074 6(7)	0.067 9(5)	0.020 1(8)	0.067 9(5)	0.081 6(3)	0.104 0(1)	0.099 9(2)
Scene	0.617 4(3)	0.554 1(6)	0.291 2(8)	0.552 3(7)	0.567 5(4)	0.561 7(5)	0.682 0(2)	0.694 0(1)
Image	0.331 1(7)	0.360 3(5)	0.278 7(8)	0.378 3(4)	0.350 1(6)	0.391 0(3)	0.429 6(2)	0.489 3(1)
Flags	0.937 1(3)	0.906 9(6)	0.937 1(3)	0.937 1(3)	0.902 9(7)	0.901 8(8)	0.937 0(5)	0.949 1(1)
Genbase	0.609 1(5)	0.243 3(6)	0.179 2(7)	0.737 2(1)	0.179 2(7)	0.692 4(2)	0.637 1(4)	0.681 6(3)
CAL500	0.933 1(4)	0.933 0(6)	0.932 3(8)	0.933 1(4)	0.932 9(7)	0.934 3(1)	0.934 1(2)	0.934 0(3)
Enron	0.094 9(5)	0.090 9(6)	0.076 7(8)	0.102 4(3)	0.102 1(4)	0.087 0(7)	0.105 7(2)	0.113 4(1)
Computer	0.060 1(6)	0.083 0(5)	0.040 6(7)	0.087 0(3)	0.040 6(7)	0.098 8(2)	0.086 0(4)	0.110 4(1)
Recreation	0.077 5(6)	0.054 2(7)	0.038 8(8)	0.095 8(4)	0.119 2(2)	0.110 1(3)	0.090 4(5)	0.150 2(1)
Social	0.132 9(5)	0.116 9(7)	0.101 9(8)	0.145 6(3)	0.158 6(1)	0.138 3(4)	0.124 5(6)	0.155 7(2)
均值	0.377 4	0.349 1	0.306 2	0.396 8	0.345 9	0.408 7	0.422 8	0.445 3

表 6 各数据集上不同算法的 MI 对比

Table 6 MI comparison of different algorithms on each dataset

数据集及均值	PMU	SCLS	FIMF	MDMR	SSFS	RFSFS	SLMDS	LRLDG
Yeast	0.609 6(3)	0.583 6(6)	0.560 1(7)	0.605 7(5)	0.560 1(7)	0.608 4(4)	0.629 5(2)	0.634 2(1)
Emotion	0.450 2(8)	0.584 7(4)	0.550 1(5)	0.529 2(7)	0.550 1(5)	0.628 5(2)	0.626 6(3)	0.636 9(1)
Birds	0.104 3(7)	0.172 8(4)	0.155 6(5)	0.057 1(8)	0.155 6(5)	0.196 7(2)	0.280 9(1)	0.196 7(2)
Scene	0.622 9(3)	0.557 9(6)	0.307 6(8)	0.557 5(7)	0.573 5(4)	0.570 6(5)	0.681 2(2)	0.690 6(1)
Image	0.344 3(7)	0.381 5(5)	0.286 6(8)	0.401 0(4)	0.355 6(6)	0.427 1(3)	0.462 5(2)	0.509 2(1)
Flags	0.937 5(4)	0.907 5(6)	0.937 5(4)	0.937 5(4)	0.902 6(8)	0.904 4(7)	0.938 7(2)	0.949 9(1)
Genbase	0.944 7(5)	0.518 7(6)	0.237 2(7)	0.970 8(1)	0.237 2(7)	0.967 1(2)	0.946 5(4)	0.960 0(3)
CAL500	0.937 8(6)	0.938 1(4)	0.937 4(8)	0.937 8(6)	0.937 8(6)	0.938 9(1)	0.938 8(2)	0.938 7(3)
Enron	0.519 3(4)	0.496 2(6)	0.487 8(8)	0.509 9(5)	0.529 6(2)	0.490 2(7)	0.529 5(3)	0.544 5(1)
Computer	0.406 3(5)	0.380 1(6)	0.378 7(7)	0.422 6(3)	0.378 7(7)	0.429 0(2)	0.417 9(4)	0.433 2(1)
Recreation	0.128 4(6)	0.079 4(7)	0.056 8(8)	0.167 7(4)	0.206 6(2)	0.189 9(3)	0.162 1(5)	0.269 5(1)
Social	0.499 8(3)	0.469 3(7)	0.460 9(8)	0.498 1(4)	0.527 0(2)	0.485 9(5)	0.483 2(6)	0.531 9(1)
均值	0.542 1	0.525 8	0.446 4	0.549 6	0.492 9	0.569 7	0.591 5	0.607 9

各特征选择算法在不同数据集上取不同特征数量时的性能变化情况如图 1~图 5 所示, 其中横轴表示选择的特征数量, 纵轴表示各算法在相应指标上的实验结果. 由图 1~图 5 可见, 大多数情况下, LRLDG 算法的性能曲线始终位于最优或次优的位置, 进一步证实了 LRLDG 算法的整体性能优于其他算法. 由图 1 和图 2 可见, LRLDG 算法的性能曲线均在其他算法性能曲线的下方, 说明在各实验数据集上, LRLDG 算法在 HL 和 RL 指标上的性能均优于其他算法. 尤其是在数据集 Yeast, Computer, Recreation, Image 和 Flags 上, LRLDG 算法的性能优势更明显. 由图 3~图 5 可见, 多数情况下, LRLDG 算法在 AP, MA 和 MI 指标上的性能曲线先明显上升再趋于稳定. 尤其是在图 5(G)中, 由于数据集 Flags 的特征仅有 19 个, LRLDG 算法在 AP 指标的性能先持续上升, 当选择的特征数为 15 时 LRLDG 算法的性能达到最优, 然后再快速下降. 表明 LRLDG 算法在处理特多标签征选择问题上有效且有意义. 虽然 LRLDG 算法在数据集 Birds, CAL500 和 Genbase 上的性能略不如 SLMDS 和 RFSFS 等算法, 但在其他数据集上的性能和整体性能上都表明了 LRLGD 算法的有效性和优越性.

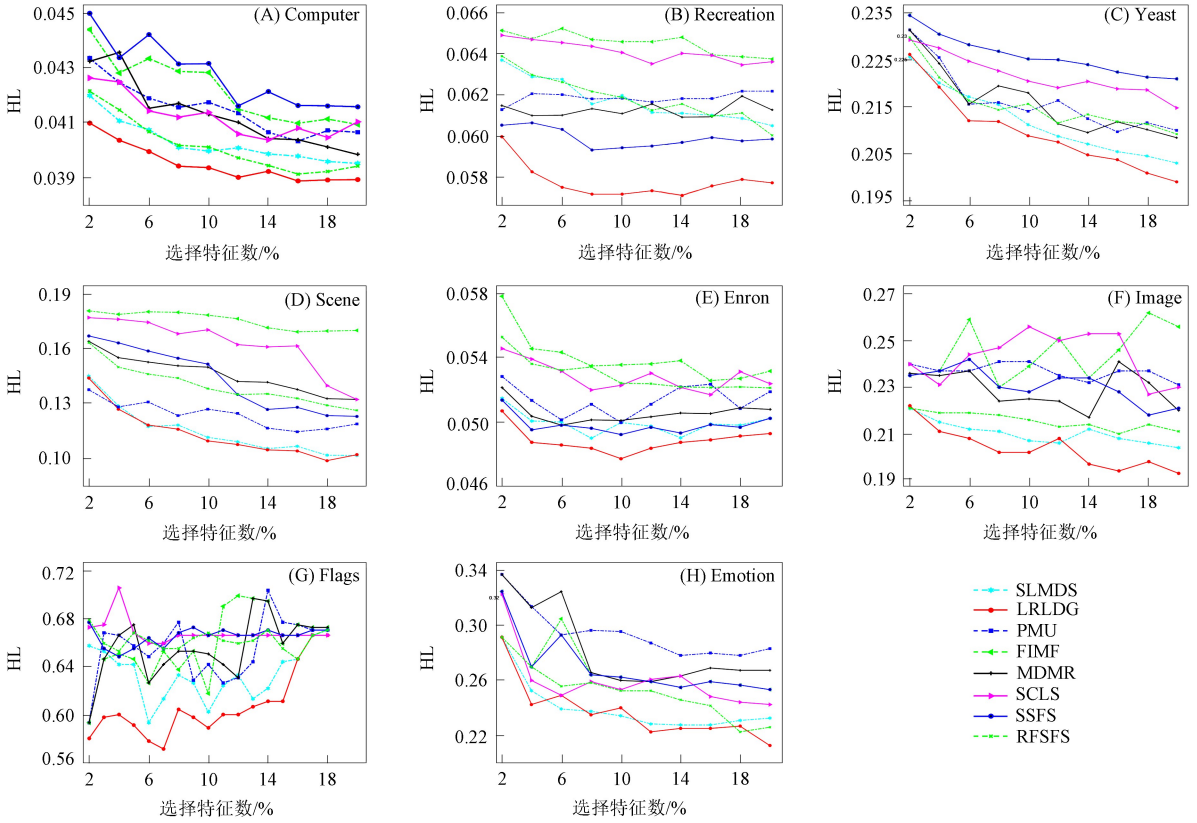


图 1 不同算法在 8 个数据集上的 HL 比较

Fig. 1 HL comparison of different algorithms on eight datasets

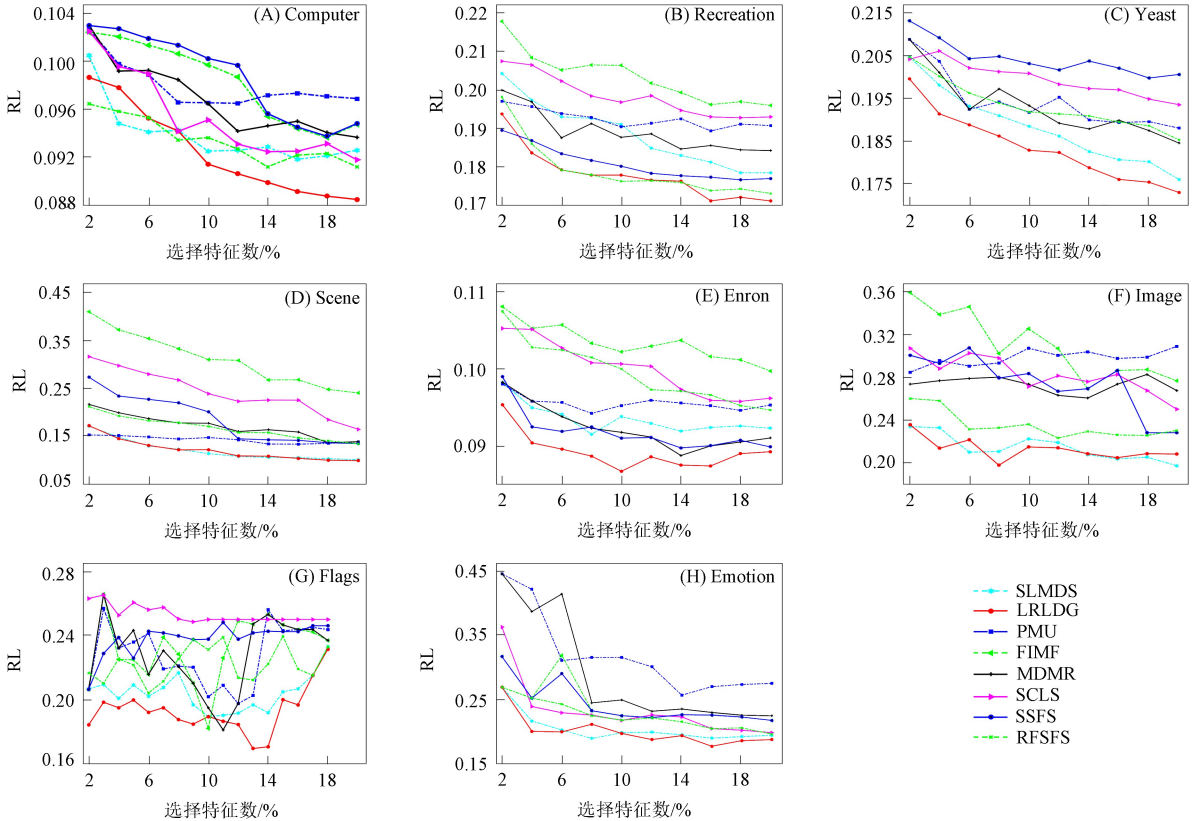


图 2 不同算法在 8 个数据集上的 RL 比较

Fig. 2 RL comparison of different algorithms on eight datasets

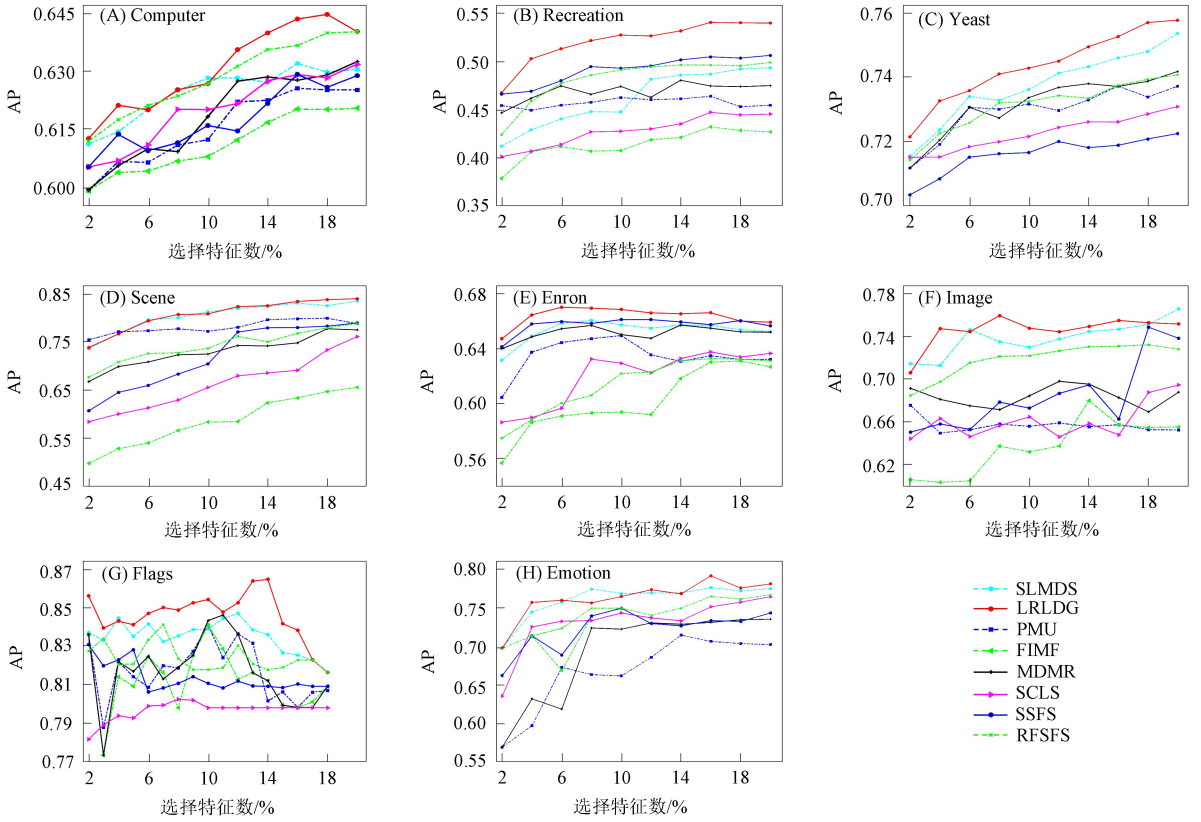


图 3 不同算法在 8 个数据集上的 AP 比较

Fig. 3 AP comparison of different algorithms on eight datasets

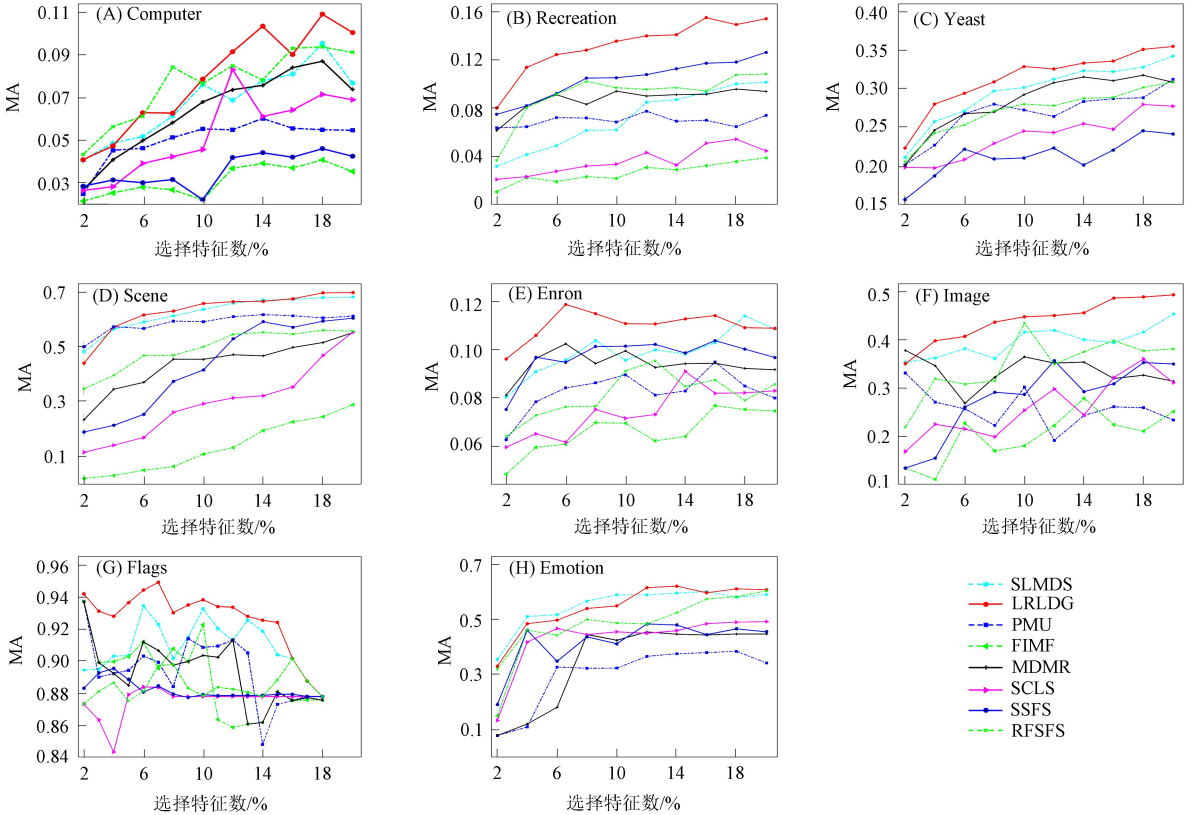


图 4 不同算法在 8 个数据集上的 MA 比较

Fig. 4 MA comparison of different algorithms on eight datasets

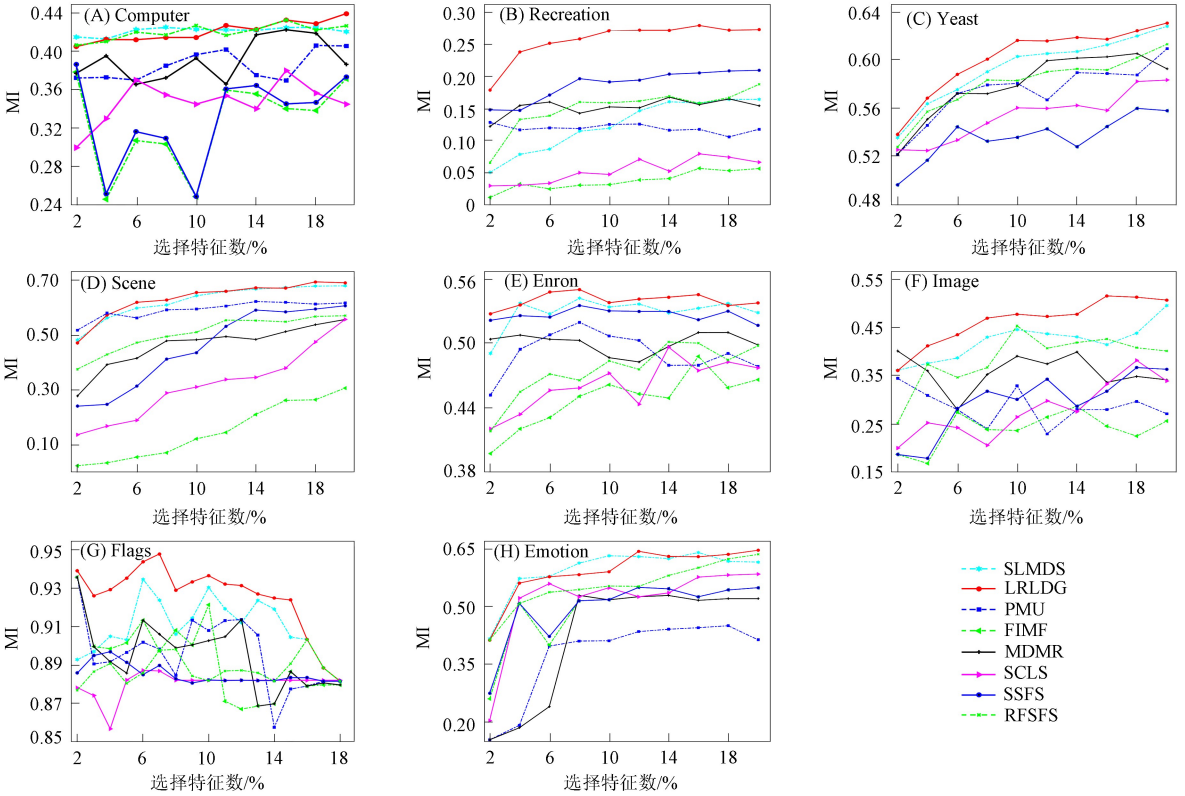


图 5 不同算法在 8 个数据集上的 MI 比较

Fig. 5 MI comparison of different algorithms on eight datasets

4.4 参数敏感性分析

下面通过参数敏感性实验观察并分析 LRLDG 算法中各参数对其性能的影响, 实验选用 AP 指标, 并使用数据集 Scene. 实验中, 将 LRLDG 算法中的一个正则化参数在 $[0.001, 0.01, 0.1, 1, 10, 100, 1000]$ 内调整, 并固定另外两个正则化参数为 1, 以便于观察 LRLDG 算法的性能随参数的变化情况. 此外, 本文还将实验数据集 Scene 的特征选择范围调整为 $[14, 28, 42, 56, 70, 84, 98]$, 以便于观察选取不同特征数量时各正则化参数对 LRLDG 算法的性能变化情况. 实验结果如图 6 所示.

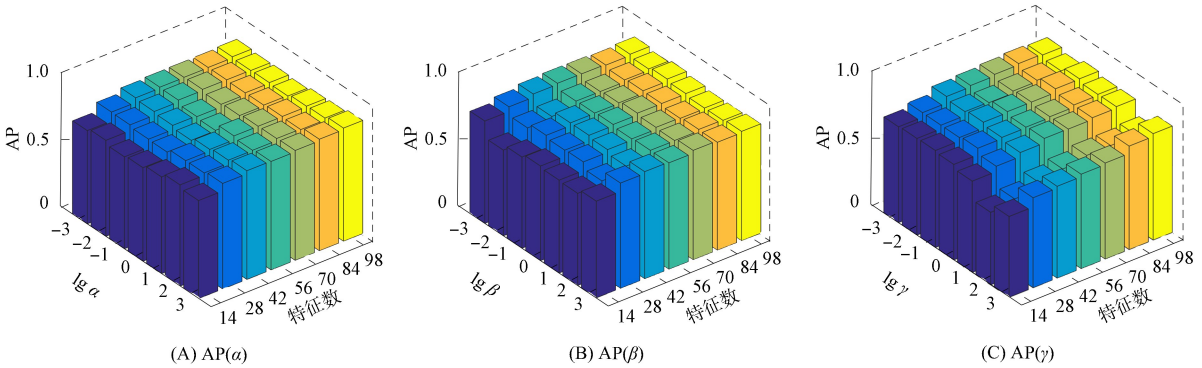


图 6 各参数对 LRLDG 算法性能的影响

Fig. 6 Effect of each parameter on performance of LRLDG algorithm

由图 6 可见, 对于所有参数, LRLDG 算法在 AP 指标上的性能会随着特征数量的增加而上升, 从而进一步证实了 LRLDG 算法对解决特征选择问题有效. 在图 6(C)中, 虽然当参数 $\gamma = 10^2$ 时, LRLDG 算法的性能波动略大, 但由图 6(A), (B)可见, 随着参数 α 和 β 的增大, LRLDG 算法的性能也逐渐提升, 且趋于稳定.

4.5 收敛性与时间复杂度分析

为进一步证实 LRLDG 算法的收敛性,下面在数据集 Birds, Emotion, Flags 和 Image 上设计针对 LRLDG 算法的收敛性实验,设 LRLDG 算法中各正则项参数的值为 1,并将迭代次数设为 50,观察 LRLDG 算法的目标函数值变化情况,实验结果如图 7 所示,其中横轴表示迭代次数,纵轴表示目标函数值.由图 7 可见,LRLDG 算法均能在 10 次迭代内收敛.

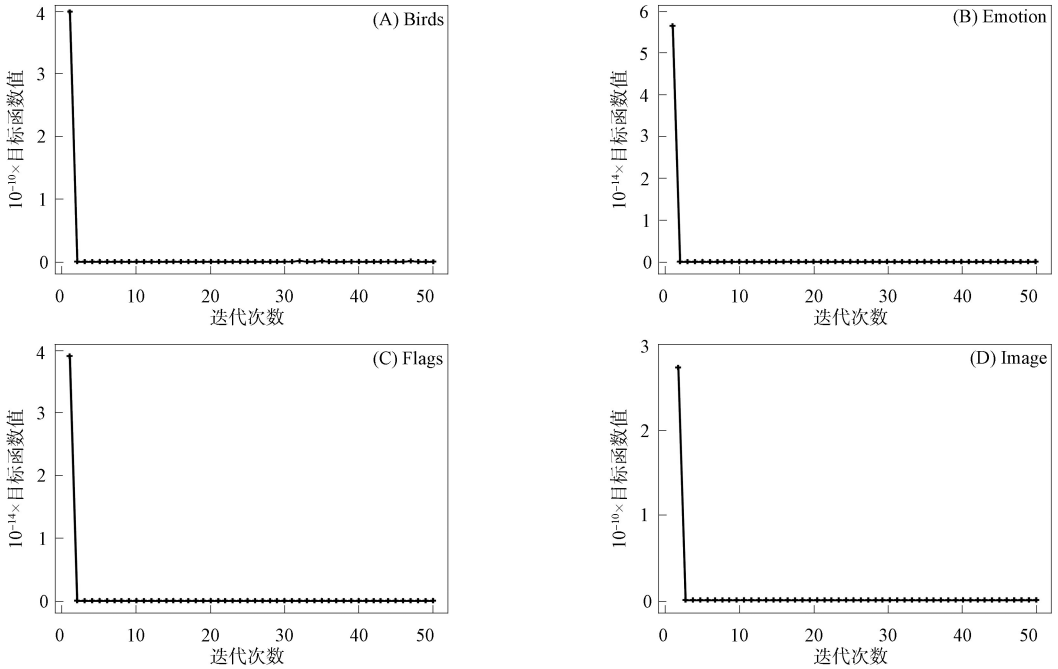


图 7 LRLDG 算法在数据集 Birds, Emotion, Flags 和 Image 上的收敛情况

Fig. 7 Convergence of LRLDG algorithm on Birds, Emotion, Flags and Image datasets

一般情况下在多标签数据集中 $n \geq m$, $d \geq m$. LRLDG 算法主要更新 W, V 和 U . 其中在每次迭代中更新 W 的时间复杂度为 $O(d^2n)$, 更新 V 的时间复杂度为 $O(dn^2)$, 更新 U 的时间复杂度为 $O(m^2)$, 所以在 t 次迭代中 LRLDG 算法的总时间复杂度为 $O(tdn^2 + td^2n + tm^2)$. LRLDG 算法可以在迭代次数 t 较小时达到收敛, 所以 LRLDG 算法的总时间复杂度为 $O(dn^2 + d^2n + m^2)$.

综上所述,针对现有嵌入式方法忽略实例相关性的潜在表示对伪标记学习的影响以及固定的图矩阵导致计算误差随迭代的加深而不断增大的问题,本文设计了一种新的具有潜在表示学习和动态图约束的多标签特征选择方法 LRLDG. 该方法首先利用实例相关性的潜在表示学习指导伪标签的学习,并利用伪标签的低维流形结构构造动态图矩阵,以缓解由固定的图矩阵导致的计算误差随迭代次数的增加不断增大的问题. 其次, LRLDG 算法通过最小化伪标签和真实标签之间的 Friedman 范数距离,并结合标签流形学习保证伪标签和真实标签的高度相关性. 最后,在 12 个基准多标签数据集上将 LRLDG 算法与 7 种先进的现有方法进行了对比实验,在 7 个常用评价指标上的实验结果均验证了 LRLDG 算法良好的特征选择性能.

参 考 文 献

- [1] WU J, SHENG V S, ZHANG J, et al. Multi-label Active Learning Algorithms for Image Classification: Overview and Future Promise [J]. ACM Computing Surveys (CSUR), 2020, 53(2): 1-35.
- [2] YU W J, CHEN Z D, LUO X, et al. Delta: A Deep Dual-Stream Network for Multi-label Image Classification [J]. Pattern Recognition, 2019, 91: 322-331.
- [3] WANG B Y, HU X G, LI P P, et al. Cognitive Structure Learning Model for Hierarchical Multi-label Text Classification [J]. Knowledge-Based Systems, 2021, 218: 106876-1-106876-13.
- [4] XIONG J, YU L, NIU X, et al. Xrr: Extreme Multi-label Text Classification with Candidate Retrieving and Deep

- Ranking [J]. *Information Sciences*, 2023, 622: 115-132.
- [5] GUO Y M, CHUNG F L, LI G Z, et al. Multi-label Bioinformatics Data Classification with Ensemble Embedded Feature Selection [J]. *IEEE Access*, 2019, 7: 103863-103875.
- [6] 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): e1240-1-e1240-29.
- [7] YIN H, YANG S Q, SONG X Y, et al. Deep Fusion of Multimodal Features for Social Media Retweet Time Prediction [J]. *World Wide Web*, 2021, 24: 1027-1044.
- [8] DING C C, ZHAO M, LIN J, et al. Multi-objective Iterative Optimization Algorithm Based Optimal Wavelet Filter Selection for Multi-fault Diagnosis of Rolling Element Bearings [J]. *ISA transactions*, 2019, 88: 199-215.
- [9] LABANI M, MORADI P, AHMADIZAR F, et al. A Novel Multivariate Filter Method for Feature Selection in Text Classification Problems [J]. *Engineering Applications of Artificial Intelligence*, 2018, 70: 25-37.
- [10] YAO C, LIU Y F, JIANG B, et al. Lle Score: A New Filter-Based Unsupervised Feature Selection Method Based on Nonlinear Manifold Embedding and Its Application to Image Recognition [J]. *IEEE Transactions on Image Processing*, 2017, 26(11): 5257-5269.
- [11] GONZÁLEZ J, ORTEGA J, DAMAS M, et al. A New Multi-objective Wrapper Method for Feature Selection-Accuracy and Stability Analysis for BCI [J]. *Neurocomputing*, 2019, 333: 407-418.
- [12] JADHA V S, HE H M, JENKINS K. Information Gain Directed Genetic Algorithm Wrapper Feature Selection for Credit Rating [J]. *Applied Soft Computing*, 2018, 69: 541-553.
- [13] MALDONADO S, LÓPEZ J. Dealing with High-Dimensional Class-Imbalanced Datasets: Embedded Feature Selection for SVM Classification [J]. *Applied Soft Computing*, 2018, 67: 94-105.
- [14] KONG Y C, YU T W. A Graph-Embedded Deep Feedforward Network for Disease Outcome Classification and Feature Selection Using Gene Expression Data [J]. *Bioinformatics*, 2018, 34(21): 3727-3737.
- [15] DONG H B, SUN J, LI T, et al. A Multi-objective Algorithm for Multi-label Filter Feature Selection Problem [J]. *Applied Intelligence*, 2020, 50: 3748-3774.
- [16] HU L, GAO L B, LI Y H, et al. Feature-Specific Mutual Information Variation for Multi-label Feature Selection [J]. *Information Sciences*, 2022, 593: 449-471.
- [17] 马盈仓, 张要, 张宁, 等. 基于流形学习与 $L_{2,1}$ 范数的无监督多标签特征选择 [J]. *纺织高校基础科学学报*, 2021, 34(3): 102-111. (MA Y C, ZHANG Y, ZHANG N, et al. Unsupervised Multi-label Feature Selection Based on Manifold Learning and $L_{2,1}$ Norm [J]. *Basic Sciences Journal of Textile Universities*, 2021, 34(3): 102-111.)
- [18] CHEN J, ZHONG M, LI J X, et al. Effective Deep Attributed Network Representation Learning with Topology Adapted Smoothing [J]. *IEEE Transactions on Cybernetics*, 2021, 52(7): 5935-5946.
- [19] HU X C, SHEN Y H, PEDRYCZ W, et al. Identification of Fuzzy Rule-Based Models with Collaborative Fuzzy Clustering [J]. *IEEE Transactions on Cybernetics*, 2021, 52(7): 6406-6419.
- [20] 张要, 马盈仓, 杨小飞, 等. 结合流形结构与柔性嵌入的多标签特征选择 [J]. *山东大学学报(理学版)*, 2021, 56(7): 91-102. (ZHANG Y, MA Y C, YANG X F, et al. Multi-label Feature Selection Based on Manifold Structure and Flexible Embedding [J]. *Journal of Shandong University (Natural Science)*, 2021, 56(7): 91-102.)
- [21] 张要, 马盈仓, 朱恒东, 等. 结合流形学习与逻辑回归的多标签特征选择 [J]. *计算机工程*, 2022, 48(3): 90-99. (ZHANG Y, MA Y C, ZHU H D, et al. Multi-label Feature Selection Combining Manifold Learning and Logistic Regression [J]. *Computer Engineering*, 2022, 48(3): 90-99.)
- [22] LIU K Y, YANG X B, FUJITA H, et al. An Efficient Selector for Multi-granularity Attribute Reduction [J]. *Information Sciences*, 2019, 505: 457-472.
- [23] CHEN Y, LIU K Y, SONG J J, et al. Attribute Group for Attribute Reduction [J]. *Information Sciences*, 2020, 535: 64-80.
- [24] JING Y G, LI T R, FUJITA H, et al. An Incremental Attribute Reduction Approach Based on Knowledge Granularity with a Multi-granulation View [J]. *Information Sciences*, 2017, 411: 23-38.
- [25] KAWANO S. Semi-supervised Logistic Discrimination via Labeled Data and Unlabeled Data from Different Sampling

- Distributions [J]. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2013, 6: 472-481.
- [26] KAWANO S, MISUMI T, KONISHI S. Semi-supervised Logistic Discrimination via Graph-Based Regularization [J]. *Neural Processing Letters*, 2012, 36: 203-216.
- [27] HU J C, LI Y H, GAO W F, et al. Robust Multi-label Feature Selection with Dual-Graph Regularization [J]. *Knowledge-Based Systems*, 2020, 203: 106-126.
- [28] HUANG R, WU Z J. Multi-label Feature Selection via Manifold Regularization and Dependence Maximization [J]. *Pattern Recognition*, 2021, 120: 108149-1-108149-12.
- [29] GRETTON A, BOUSQUET O, SMOLA A, et al. Measuring Statistical Dependence with Hilbert-Schmidt Norms [C]//*Proceedings of the International Conference on Algorithmic Learning Theory*. Berlin: Springer, 2005: 63-77.
- [30] LI Y H, HU L, GAO W F. Multi-label Feature Selection via Robust Flexible Sparse Regularization [J]. *Pattern Recognition*, 2023, 134: 109074-1-109074-15.
- [31] TANG C, BIAN M R, LIU X W, et al. Unsupervised Feature Selection via Latent Representation Learning and Manifold Regularization [J]. *Neural Networks*, 2019, 117: 163-178.
- [32] GAO W F, LI Y H, HU L. Multilabel Feature Selection with Constrained Latent Structure Shared Term [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(3): 1253-1262.
- [33] LI Y H, HU L, GAO W F. Robust Sparse and Low-Redundancy Multi-label Feature Selection with Dynamic Local and Global Structure Preservation [J]. *Pattern Recognition*, 2023, 134: 109120-1-109120-16.
- [34] LEE J, KIM D W. Feature Selection for Multi-label Classification Using Multivariate Mutual Information [J]. *Pattern Recognition Letters*, 2013, 34(3): 349-357.
- [35] 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.
- [36] LEE J, KIM D W. Fast Multi-label Feature Selection Based on Information-Theoretic Feature Ranking [J]. *Pattern Recognition*, 2015, 48(9): 2761-2771.
- [37] LEE J, KIM D W. Scfs: Multi-label Feature Selection Based on Scalable Criterion for Large Label Set [J]. *Pattern Recognition*, 2017, 66: 342-352.
- [38] CAI D, ZHANG C, HE X. Unsupervised Feature Selection for Multi-cluster Data [C]//*Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2010: 333-342.
- [39] LEE D D, SEUNG H S. Learning the Parts of Objects by Non-negative Matrix Factorization [J]. *Nature*, 1999, 401: 788-791.
- [40] HUANG J, NIE F, HUANG H. A New Simplex Sparse Learning Model to Measure Data Similarity for Clustering [C]//*Proceedings of the 24th International Joint Conference on Artificial Intelligence*. New York: ACM, 2015: 3569-3575.
- [41] LEE D D, SEUNG S H. Algorithms for Non-negative Matrix Factorization [J]. *Advances in Neural Information Processing Systems*, 2001, 13: 556-562.
- [42] DING C H Q, LI T, JORDAN M I. Convex and Semi-nonnegative Matrix Factorizations [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 32(1): 45-55.
- [43] TSOUMAKAS G, SPYROMITROS-XIOUFIS E, VILCEK J, et al. Mulan: A Java Library for Multi-label Learning [J]. *The Journal of Machine Learning Research*, 2011, 12: 2411-2414.
- [44] ZHANG M L, ZHOU Z H. ML-KNN: A Lazy Learning Approach to Multi-label Learning [J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [45] ZHANG Y, MA Y C. Non-negative Multi-label Feature Selection with Dynamic Graph Constraints [J]. *Knowledge-Based Systems*, 2022, 238: 107924-1-107924-14.
- [46] DOUGHERTY J, KOHAVI R, SAHAMI M. Supervised and Unsupervised Discretization of Continuous Features [M]. San Mateo: Morgan Kaufmann, 1995: 194-202.
- [47] XUE G T, ZHONG M, LI J X, et al. Dynamic Network Embedding Survey [J]. *Neurocomputing*, 2022, 472: 212-223.