

doi:10.13682/j.issn.2095-6533.2026.05.015

基于广义稀疏低秩回归的特征选择算法

支晓斌¹, 高攀², 孟玉²

(1. 西安邮电大学 理学院, 陕西 西安 710121; 2. 西安邮电大学 通信与信息工程学院, 陕西 西安 710121)

摘要: 针对基于广义稀疏回归的特征选择 (Generalized Sparse Regression based Feature Selection, GSRFS) 算法存在忽略类与类之间潜在相关性的问题, 提出一种基于广义稀疏低秩回归的特征选择 (Generalized Sparse Low-Rank Regression based Feature Selection, GSLRRFS) 算法。GSLRRFS 算法将广义形式的低秩正则化项引入 GSRFS 的目标函数, 以挖掘类与类之间潜在相关性信息提升算法特征选择性能。同时, 将相应优化问题转化为迭代重加权最小二乘问题, 利用最小平方 QR 分解算法进行高效求解。7 个真实数据上的对比实验结果表明, 在幂函数形式下选取 20% 特征比例, GSLRRFS 算法较 GSRFS 算法的平均分类精度提升 2.31 个百分点, 平均运行时间缩短为原来的约十分之一, 所提算法不仅能有效提升特征选择性能, 而且显著降低了运行时间。

关键词: 特征选择; 稀疏回归; 广义低秩; 迭代重加权; 最小平方 QR 分解

中图分类号: TP181

文献标志码: A

文章编号: 2095-6533(2026)05-0159-12

Generalized sparse low-rank regression based feature selection algorithm

ZHI Xiaobin¹, GAO Pan², MENG Yu²

(1. School of Science, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;

2. School of Communications and Information Engineering, Xi'an University of Posts and
Telecommunications, Xi'an 710121, China)

Abstract: The generalized sparse regression based feature selection (GSRFS) algorithm has the problem of ignoring potential correlations between classes. To address this issue, a generalized sparse low-rank regression based feature selection (GSLRRFS) algorithm is proposed. The GSLRRFS algorithm introduces a generalized form of low-rank regularization term into the objective function of GSRFS to explore potential correlation information between classes, which improves feature selection performance of the algorithm. Meanwhile, the optimization problem can be transformed into an iterative re-weighted least squares problem, which can be solved efficiently by the least square QR-factorization algorithm. Comparative experimental results on seven real datasets show that, under the power function form and with 20% feature selection ratio, the GSLRRFS algorithm improves average classification accuracy by 2.31 percentage points over the GSRFS algorithm, while reducing the average runtime to approximately onetenth of the original. The proposed algorithm can not only effectively enhance feature selection performance, but can also significantly decrease the computational time.

Keywords: feature selection; sparse regression; generalized low-rank; iterative re-weighted;

收稿日期: 2025-02-24

基金项目: 国家自然科学基金项目(62071378); 陕西省自然科学基金基础研究计划项目(2020JM-580); 陕西高校青年创新团队

引文格式: 支晓斌, 高攀, 孟玉. 基于广义稀疏低秩回归的特征选择算法[J]. 西安邮电大学学报, 2026, 31(5): 159-170.

ZHI Xiaobin, GAO Pan, MENG Yu. Generalized sparse low-rank regression based feature selection algorithm[J]. Journal of Xi'an University of Posts and Telecommunications, 2026, 31(5): 159-170.

least square QR-factorization

随着大数据和人工智能技术的蓬勃发展,图像处理^[1]、视频分析^[2]和生物信息学^[3]等领域产生了大量的高维数据,这些高维数据不可避免地含有较多的冗余及不相关特征,不仅会致使算法计算复杂度增加,还会干扰数据处理与分析方法的有效性和可靠性^[4-5]。特征选择^[6-7]作为经典的数据降维方法之一,在不改变数据原始结构的前提下,仅凭若干关键特征表征原始数据,能够有效剔除原始数据集中冗余和不相关的特征,保留重要特征,现已广泛应用于数据挖掘和机器学习中。

近年来,在特征选择领域,基于结构化稀疏诱导技术将稀疏正则化项与回归分析^[8]相结合得到的基于稀疏回归的特征选择(Sparse Regression based Feature Selection, SRFS)算法引发了广泛讨论^[9-12]。Nie等^[10]为实现鲁棒版本的SRFS,在回归损失函数和稀疏正则化项中联合使用 $l_{2,1}$ -范数,所得算法有效降低了数据中异常值和噪声对算法的影响。鉴于 l_p -范数($0 < p < 1$)可以得到比 l_1 -范数更稀疏的解,Wang等^[11]将 $l_{2,1}$ -范数扩展到 $l_{2,p}$ -范数($0 < p \leq 1$),提出了基于 $l_{2,p}$ -范数的SRFS算法。随后,Zhi等^[12]进一步采用不同的范数度量损失函数和稀疏正则化项,提出了 $l_{2,p}$ -SRFS模型的广义形式: $l_{2,\beta}$ -SRFS,并采用基于最小平方QR分解^[13](Least Square QR-Factorization, LSQR)的高效率算法进行求解。此外,Nie等^[14]使用一类函数构造回归损失项和稀疏正则化项,提出了基于广义稀疏回归的特征选择(Generalized Sparse Regression based Feature Selection, GSRFS)算法,并开发了具有简单高效性的迭代重加权(Iterative Re-Weighted, IRW)进行求解。GSRFS为SRFS提供了一个通用算法模型,但其与已有的SRFS相比,存在忽略类与类之间潜在相关性的问题,在数据存在类与类之间潜在相关性的实际应用场景中不能达到最优的特征选择性能。

在分类问题的实际应用中,数据中的不同类之间可能共享一组“潜在变量”或者“潜在特征”,即类与类之间存在潜在的相关性,挖掘这种相关性,可以有效降低模型复杂度,避免过拟合问题,从而提升模型的泛化性能^[15-18]。标准最小二乘回归相当于分别对数据的每个类进行回归,无法挖掘不同类之间的潜在相关性。为解决此问题,Anderson^[19]首次使用低秩来约束回归系数矩阵,提出了一种低秩回归(Low-Rank Regression, LRR)方法。此后,为

进一步提升LRR的性能,大量学者将稀疏正则化引入其框架,提出了众多基于稀疏低秩回归的特征选择(Sparse Low-Rank Regression based Feature Selection, SLRRFS)算法^[20-22]。SLRRFS算法分别从特征选择以及挖掘类与类之间潜在相关性两个方面,实现了算法的综合性能提升。然而,大多数SLRRFS算法以特定的范数为基础构造回归损失项、稀疏正则化项以及低秩正则化项,模型拓展性受限。因此,使用一类函数构造广义形式的SLRRFS算法框架,并设计统一的高效求解方式是一个值得讨论的问题。

鉴于GSRFS算法为广义形式的SRFS以及低秩在回归上的有效性,拟在GSRFS算法基础上,引入广义形式的低秩正则化项,提出一种基于广义稀疏低秩回归的特征选择(Generalized Sparse Low-Rank Regression based Feature Selection, GSLRRFS)算法。该算法利用稀疏正则化促使回归系数矩阵实现行稀疏,进行特征有效筛选;利用低秩正则化促使变换矩阵呈低秩结构,以挖掘类与类之间的潜在相关性,进一步提升算法性能。同时,为了进一步提高求解效率,GSLRRFS将原问题转化为迭代重加权最小二乘问题,采用LSQR算法进行高效求解。最后,使用对数函数和幂函数两种类型的函数形式进行实验,验证所提算法的性能。

1 预备知识

1.1 符号定义

设数据矩阵 $\mathbf{X} \in \mathbb{R}^{d \times n}$,对应的标签矩阵为 $\mathbf{Y} \in \mathbb{R}^{n \times c}$,回归系数矩阵为 $\mathbf{W} \in \mathbb{R}^{d \times c}$ 。其中, n 为样本个数, d 为样本的特征维数, c 为样本的类别数。

给定任意矩阵 $\mathbf{M} \in \mathbb{R}^{d \times n}$, $\mathbf{m}^i$ 和 $\mathbf{m}_j$ 分别表示 $\mathbf{M}$ 的第 i 行和第 j 列,则 $\mathbf{M}$ 的 $l_{2,1}$ -范数^[10]定义为

$$\|\mathbf{M}\|_{2,1} = \sum_{i=1}^d \left(\sum_{j=1}^n |\mathbf{m}_{ij}|^2 \right)^{\frac{1}{2}} = \sum_{i=1}^d \|\mathbf{m}^i\|_2 \quad (1)$$

在优化低秩回归模型时,矩阵低秩约束的凸近似一般采用核范数实现, $\mathbf{M}$ 的核范数定义为

$$\|\mathbf{M}\|_* = \text{Tr}((\mathbf{M}\mathbf{M}^T)^{\frac{1}{2}}) = \sum_{i=1}^d \sigma_i(\mathbf{M}) \quad (2)$$

式中: $\text{Tr}(\cdot)$ 表示矩阵的迹; $\sigma_i(\mathbf{M})$ 表示矩阵 $\mathbf{M}$ 的第 i 个奇异值。

矩阵 $\mathbf{M}$ 的矩阵范数和核范数的广义形式^[14]分别定义为

$$\|M\|_{2,f} = \sum_{i=1}^d f(\|m^i\|_2^2) \quad (3)$$

$$\|M\|_*^f = \text{Tr}(f(MM^T)) = \sum_{i=1}^d f(\sigma_i^2(M)) \quad (4)$$

式中: $f(\cdot)$ 表示单调不减的可微凹函数。假设 MM^T 的特征分解为 $MM^T = Q\Lambda Q^T$, 其中 Q 为由 MM^T 特征向量构成的正交矩阵, $\Lambda \in \mathbb{R}^{d \times d}$ 为由 MM^T 特征值构成的对角矩阵, 则 $f(MM^T) = Qf(\Lambda)Q^T$ 。此时, $f(\Lambda) \in \mathbb{R}^{d \times d}$ 为对角矩阵, 其对角线上第 i 个元素定义为 $f(\lambda_i)$, λ_i 为 MM^T 的第 i 个特征值。由于 $\lambda_i(MM^T) = \sigma_i^2(M)$, 则 $f(MM^T)$ 的定义通过对 MM^T 的特征值(或 M 的奇异值)定义来实现。

假设 $f(x) = x^{1/2}$, 将 M 代入式(3)和式(4), 则分别有

$$\|M\|_{2,f} =$$

$$\sum_{i=1}^d (\|m^i\|_2^2)^{\frac{1}{2}} = \sum_{i=1}^d \|m^i\|_2 = \|M\|_{2,1} \quad (5)$$

$$\|M\|_*^f = \sum_{i=1}^d (\sigma_i^2(M))^{\frac{1}{2}} = \sum_{i=1}^d \sigma_i(M) = \|M\|_* \quad (6)$$

即 $\|M\|_{2,f}$ 和 $\|M\|_*^f$ 分别退化为 $\|M\|_{2,1}$ 和 $\|M\|_*$ 。

1.2 低秩回归

传统 LRR 模型可以学习回归系数矩阵 W 的低秩结构, 捕捉类与类之间潜在相关性, 使不同类别共享低维特征空间, 从而提升回归的性能。LRR 模型可描述为优化问题

$$\begin{aligned} \min_W \|X^T W - Y\|_F^2 \\ \text{s. t. rank}(W) \leq r. \end{aligned} \quad (7)$$

式中: $\text{rank}(W)$ 表示矩阵 W 的秩; $r < \min(d, c)$ 。

鉴于式(7)中的优化问题是 N-P (Non-Deterministic Polynomial-Time) 难问题, 在实际的 LRR 学习过程中, 可以采用以下两种形式的低秩优化策略。

1) 显式地将 W 分解为如图 1 所示的两个秩为 r 的矩阵的乘积, 即 $W = UV^T$, 此时 LRR 模型变为

$$\min_{U, V} \|X^T UV^T - Y\|_F^2 \quad (8)$$

式中: $U \in \mathbb{R}^{d \times r}$; $V \in \mathbb{R}^{c \times r}$; $r < \min(d, c)$ 。

根据线性代数中矩阵秩运算的性质可知

$$\text{rank}(UV^T) \leq \min\{\text{rank}(U), \text{rank}(V)\} \quad (9)$$

所以式(8)中 r 越小, $\text{rank}(W)$ 越小。

式(8)可看作先用矩阵 U 对原数据 X 进行降维, 得到潜在变量 $X^T U$, 然后对潜在变量 $X^T U$ 和标签矩阵 Y 用 V^T 进行回归^[15]。此时, 高维数据被降维, 所以过拟合问题将不存在。此外, 这种降维潜在变量 $X^T U$ 与标签矩阵 Y 直接产生了关系, $X^T U$

的每一列即为由所有响应变量(Y 的列)共享的“潜在变量”^[15]或者“潜在特征”^[16-17]构成, 标签矩阵因为共享“潜在变量”而变得相关。 V 的第 i 行 v^i 度量了第 i 个类对每个潜在变量的依赖性程度, v^i 的第 j 个分量的绝对值(或者模长)越大, 表明第 i 个类对第 j 个潜在变量的依赖性越大。

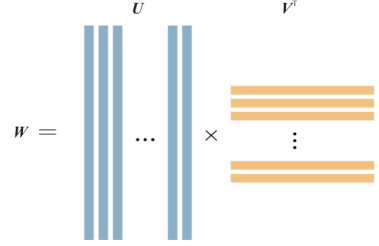


图 1 W 分解示意图

2) 低秩化策略是用凸松弛方法, 即通过最小化 $\text{rank}(W)$ 的凸近似 $\|W\|_*$ 来实现 W 的低秩^[23]。此时 LRR 的优化模型为

$$\min_W \|X^T W - Y\|_F^2 + \gamma \|W\|_* \quad (10)$$

对优化式(10)所得低秩变换矩阵 W 进行奇异值分解可得

$$W = U_f \Sigma_f V_f^T \quad (11)$$

式中: $U_f \in \mathbb{R}^{d \times d}$; $\Sigma_f \in \mathbb{R}^{d \times c}$; $V_f \in \mathbb{R}^{c \times c}$ 。

令 $U_1 = U_f \Sigma_f^{1/2}$, $V_1 = V_f \Sigma_f^{1/2}$, 可得

$$W = U_1 V_1^T \quad (12)$$

取 U_1 和 V_1 的前 r 列构成 U 和 V , 得到

$$W = UV^T \quad (13)$$

综上可知, 两种低秩化策略在目的是等效的。

1.3 基于广义稀疏回归的特征选择

现有的大多数 SRFS 算法都是基于特定的模型, 不具有普适性。为此, 广义形式的 SRFS 模型^[14]表示为

$$\min_W \|X^T W - Y\|_{2,f} + \gamma \|W\|_{2,h} \quad (14)$$

式中: 第一项为回归损失项, 第二项为稀疏正则化项, 均采用式(3)中的广义度量形式, 并且 $f(\cdot)$ 和 $h(\cdot)$ 为不同的单调不减可微凹函数。

GSRRFS 将已有的多个 SRFS 模型作为特例, 通过有收敛性保证的 IRW 算法求解, 其主要意义在为实际应用提供了更多的可供选择的特征选择算法模型。

2 GSLRRFS 算法

2.1 模型建立

鉴于 GSRRFS 只考虑回归系数矩阵 W 的行稀疏性, 忽略了类与类之间的潜在相关性, 而低秩在

捕捉类与类之间潜在相关性信息具有独特的效果。因此,在 GSRFS 基础上,GSLRRFS 算法引入广义形式的低秩正则化项,表示为

$$\min_{\mathbf{W}} J(\mathbf{W}) = \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}\|_{2,f} + \gamma_1 \|\mathbf{W}\|_{2,h} + \gamma_2 \|\mathbf{W}\|_*^l \quad (15)$$

式中:第一项回归损失项和第二项稀疏正则化项都采用式(3)的广义形式进行度量;第三项为低秩正则化项促使 $\mathbf{W}$ 能够挖掘类与类之间的潜在相关性,采用式(4)的广义形式进行度量; γ_1 和 γ_2 为两个非负正则化参数,当 $\gamma_2 = 0$ 时,GSLRRFS 可退化为 GSRFS 模型。

将式(15)进行光滑近似处理,并将回归损失项、稀疏正则化项和低秩正则化项分别用单调不减的可微凹函数 $f(\cdot)$ 、 $h(\cdot)$ 和 $r(\cdot)$ 表示,且三者可以采取不同的函数形式,则式(15)可进一步表示为

$$\min_{\mathbf{W}} J(\mathbf{W}) = \sum_{i=1}^n f(\|\mathbf{x}_i^T \mathbf{W} - \mathbf{y}^i\|_2^2 + \epsilon) + \gamma_1 \sum_{i=1}^d h(\|\mathbf{w}^i\|_2^2 + \epsilon) + \gamma_2 \text{Tr}(r(\mathbf{W}\mathbf{W}^T + \epsilon \mathbf{I}_d)) \quad (16)$$

式中: ϵ 表示充分小的正数; $\mathbf{I}_d$ 为单位矩阵。

2.2 模型求解

与 GSRFS 相比,式(16)增加了低秩正则化项,这使得求解的难度加大,但该问题仍可以使用 IRW 算法求解。令 $u_i = \|\mathbf{x}_i^T \mathbf{W} - \mathbf{y}^i\|_2^2 + \epsilon$, $v_i = \|\mathbf{w}^i\|_2^2 + \epsilon$, $g(\mathbf{W}) = \mathbf{W}\mathbf{W}^T + \epsilon \mathbf{I}_d$, 则式(16)可表示为

$$\min_{\mathbf{W}} \sum_{i=1}^n f(u_i) + \gamma_1 \sum_{i=1}^d h(v_i) + \gamma_2 \text{Tr}(r(g(\mathbf{W}))) \quad (17)$$

式(17)对 $\mathbf{W}$ 进行求导,并令其导数为 0,根据链式法则^[24]可得

$$\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} = \sum_{i=1}^n \frac{\partial f(u_i)}{\partial u_i} \frac{\partial u_i}{\partial \mathbf{W}} + \gamma_1 \sum_{i=1}^d \frac{\partial h(v_i)}{\partial v_i} \frac{\partial v_i}{\partial \mathbf{W}} + \gamma_2 \text{Tr} \left(\frac{\partial r(g(\mathbf{W}))}{\partial g(\mathbf{W})} \frac{\partial g(\mathbf{W})}{\partial \mathbf{W}} \right) = 0 \quad (18)$$

定义对角矩阵 $\mathbf{D}_1 \in \mathbb{R}^{n \times n}$, $\mathbf{D}_2 \in \mathbb{R}^{d \times d}$, 令其对角元素分别为

$$\mathbf{D}_1(i, i) = \frac{\partial f(u_i)}{\partial u_i} = f'(u_i) \quad (19)$$

$$\mathbf{D}_2(i, i) = \frac{\partial h(v_i)}{\partial v_i} = h'(v_i) \quad (20)$$

定义矩阵 $\mathbf{D}_3 \in \mathbb{R}^{d \times d}$ 为

$$\mathbf{D}_3 = \frac{\partial r(g(\mathbf{W}))}{\partial g(\mathbf{W})} = r'(g(\mathbf{W})) \quad (21)$$

则式(18)可简化为

$$\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}} = \sum_{i=1}^n \frac{\mathbf{D}_1(i, i) \partial u_i}{\partial \mathbf{W}} + \gamma_1 \sum_{i=1}^d \frac{\mathbf{D}_2(i, i) \partial v_i}{\partial \mathbf{W}} +$$

$$2\gamma_2 \mathbf{D}_3 \mathbf{W} = 0 \quad (22)$$

式(22)可进一步表示为

$$\mathbf{X} \mathbf{D}_1 \mathbf{X}^T \mathbf{W} - \mathbf{X} \mathbf{D}_1 \mathbf{Y} + \lambda_1 \mathbf{D}_2 \mathbf{W} + \lambda_2 \mathbf{D}_3 \mathbf{W} = \mathbf{0} \quad (23)$$

此外,当 $\mathbf{D}_1$, $\mathbf{D}_2$ 和 $\mathbf{D}_3$ 固定时,存在问题

$$\min_{\mathbf{W}} \sum_{i=1}^n \mathbf{D}_1(i, i) u_i + \gamma_1 \sum_{i=1}^d \mathbf{D}_2(i, i) v_i + \gamma_2 \text{Tr}(\mathbf{D}_3 g(\mathbf{W})) \quad (24)$$

此时,利用式(24)对 $\mathbf{W}$ 求导,并令导数为 0 可得式(23)。因此,式(17)等价于求解式(24)。

整理式(23)得

$$\mathbf{W} = (\mathbf{X} \mathbf{D}_1 \mathbf{X}^T + \gamma_1 \mathbf{D}_2 + \gamma_2 \mathbf{D}_3)^{-1} \mathbf{X} \mathbf{D}_1 \mathbf{Y} \quad (25)$$

利用式(25)对 $\mathbf{W}$ 进行求解,计算 $(\mathbf{X} \mathbf{D}_1 \mathbf{X}^T + \gamma_1 \mathbf{D}_2 + \gamma_2 \mathbf{D}_3)^{-1}$, 主要计算复杂度为 $O(dn^2 + nd^2 + d^3)$, 当维数 d 较大时,会导致算法计算复杂度偏高。为高效求解,分别令

$$\mathbf{X} \mathbf{D}_1 \mathbf{X}^T + \gamma_1 \mathbf{D}_1 + \gamma_2 \mathbf{D}_2 = \mathbf{A} \quad (26)$$

$$\mathbf{X} \mathbf{D}_1 \mathbf{Y} = \mathbf{B} \quad (27)$$

则 $\mathbf{W}$ 可通过优化下述最小二乘问题得到

$$\min_{\mathbf{W}} \|\mathbf{A} \mathbf{W} - \mathbf{B}\|_2^2 = 0 \quad (28)$$

对于式(28),可采用 LSQR 算法进行高效求解。此时,主要计算复杂度为 $O(nd^2 + dnc)$ 。

总结上述讨论可知,求解式(17)可以采用如下迭代方法求解:当 $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$ 固定时,用 LSQR 算法求解式(28)以更新回归系数矩阵 $\mathbf{W}$;当回归系数矩阵 $\mathbf{W}$ 固定时,计算式(19)、式(20)和式(21)以更新矩阵 $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_3$ 。GSLRRFS 算法伪代码如下。

算法:GSLRRFS 算法

输入:数据矩阵 $\mathbf{X} \in \mathbb{R}^{d \times n}$;标签矩阵 $\mathbf{Y} \in \mathbb{R}^{n \times c}$;正则化参数 γ_1 和 γ_2 ;选择特征的比例 $s\%$;算法最大迭代次数 $J_{\max}$

输出:低维矩阵数据 $\mathbf{X}^L$

1: 初始化 $t=0$, 回归系数矩阵 $\mathbf{W}^{(0)} = \mathbf{I}$

2: while $t > J_{\max}$ and $\frac{\|\mathbf{W}^{(t-1)}\|_2 - \|\mathbf{W}^{(t)}\|_2}{\|\mathbf{W}^{(t-1)}\|_2} < 10^{-5}$

3: 计算式(19)、式(20)和式(21), 得到对角矩阵 $\mathbf{D}_1^{(t)}$, $\mathbf{D}_2^{(t)}$, $\mathbf{D}_3^{(t)}$

4: 计算式(26)和式(27), 得到矩阵 $\mathbf{A}$ 和 $\mathbf{B}$

5: 运行 LSQR 算法求解式(28)得到 $\mathbf{W}^{(t)}$

6: $t = t + 1$

7: end while

8: 计算矩阵 $\mathbf{W}^{(t)}$ 每一行的 l_2 -范数, 然后对其进行降序排序并选择对应的前 $s\%$ 特征组成新的数据矩阵 $\mathbf{X}^L$

2.3 收敛性分析

为证明 GSLRRFS 算法的收敛性,给出下面引理 1 和引理 2,并依据式(4)将式(17)进一步表示为

$$\min_{\mathbf{W}} \sum_{i=1}^n f(u_i) + \gamma_1 \sum_{i=1}^d h(v_i) + \gamma_2 \sum_{i=1}^d r(\lambda_i) \quad (29)$$

式中: λ_i 为 $g(\mathbf{W})$ 的第 i 个特征值, $g(\mathbf{W}) = \mathbf{W}\mathbf{W}^T + \epsilon \mathbf{I}$.

引理 1 对于任意一个可微凹函数, 以下不等式恒成立

$$f(\tilde{x}) - f(x) \leq f'(x)(\tilde{x} - x) \quad (30)$$

引理 2 设 $\tilde{\mathbf{W}}$ 为式(24)的最优解, $g(\mathbf{W})$ 和 $g(\tilde{\mathbf{W}})$ 的特征值分解分别为 $\mathbf{W}\mathbf{W}^T + \epsilon \mathbf{I} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$ 和 $\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T + \epsilon \mathbf{I} = \tilde{\mathbf{P}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{P}}^T$, 其中 $\mathbf{\Lambda} \in \mathbb{R}^{d \times d}$ 和 $\tilde{\mathbf{\Lambda}} \in \mathbb{R}^{d \times d}$ 分别是 $g(\mathbf{W})$ 和 $g(\tilde{\mathbf{W}})$ 特征值采用降序进行排列所构成的对角阵, $\mathbf{P} \in \mathbb{R}^{d \times d}$ 和 $\tilde{\mathbf{P}} \in \mathbb{R}^{d \times d}$ 为相应特征向量构成的正交阵, 则分别有

$$\text{Tr}(\mathbf{D}_3 g(\mathbf{W})) = \sum_{i=1}^d r'(\lambda_i) \lambda_i \quad (31)$$

$$\sum_{i=1}^d r'(\lambda_i) \tilde{\lambda}_i \leq \text{Tr}(\mathbf{D}_3 g(\tilde{\mathbf{W}})) \quad (32)$$

式中: $r'(\lambda_i)$ 为 $\mathbf{D}_3$ 的第 i 个特征值。

证明 因为 $g(\mathbf{W}) = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$, 根据式(21)和矩阵函数的定义^[25]可知, $\mathbf{D}_3 = \mathbf{P}\mathbf{r}'(\mathbf{\Lambda})\mathbf{P}^T$, 则有

$$\begin{aligned} \text{Tr}(\mathbf{D}_3 g(\mathbf{W})) &= \\ \text{Tr}(\mathbf{P}\mathbf{r}'(\mathbf{\Lambda})\mathbf{P}^T \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T) &= \sum_{i=1}^d r'(\lambda_i) \lambda_i \end{aligned} \quad (33)$$

此外, 由于 $g(\mathbf{W})$ 和 $g(\tilde{\mathbf{W}})$ 的特征值都按降序排列, 即

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \quad (34)$$

$$\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \dots \geq \tilde{\lambda}_d \quad (35)$$

又因为 $r(\cdot)$ 是单调不减的可微凹函数, 所以分别有

$$r'(\lambda_1) \leq r'(\lambda_2) \leq \dots \leq r'(\lambda_d) \quad (36)$$

$$r'(\tilde{\lambda}_1) \leq r'(\tilde{\lambda}_2) \leq \dots \leq r'(\tilde{\lambda}_d) \quad (37)$$

根据 Von Neumann 迹不等式^[26]可知以下不等式成立

$$\sum_{i=1}^d r'(\lambda_i) \tilde{\lambda}_i \leq \text{Tr}(\mathbf{D}_3 g(\tilde{\mathbf{W}})) \quad (38)$$

定理 1 假设函数 $f(\cdot)$, $h(\cdot)$, $r(\cdot)$ 全为单调不减的可微凹函数, 则由算法 1 产生的优化模型式(16)的目标函数值是单调递减有下界且收敛的。

证明 假设 $\tilde{\mathbf{W}}$ 为式(29)取得最优解时 $\mathbf{W}$ 的取值, 则 $\tilde{u}_i = \|\mathbf{x}_i^T \tilde{\mathbf{W}} - \mathbf{y}^i\|_2^2 + \epsilon$, $\tilde{v}_i = \|\tilde{\mathbf{w}}^i\|_2^2 + \epsilon$, $\tilde{\lambda}_i$ 分别是式(29)取得最优解 u_i, v_i, λ_i 的取值。由于函数 $f(\cdot)$, $h(\cdot)$, $r(\cdot)$ 为单调不减的可微的凹函数。根据引理 1, 可以得到

$$f(\tilde{u}_i) + \gamma_1 h(\tilde{v}_i) + \gamma_2 r(\tilde{\lambda}_i) -$$

$$f(u_i) - \gamma_1 h(v_i) - \gamma_2 r(\lambda_i) \leq f'(u_i)(\tilde{u}_i - u_i) + \gamma_1 h'(v_i)(\tilde{v}_i - v_i) + \gamma_2 r'(\lambda_i)(\tilde{\lambda}_i - \lambda_i) \quad (39)$$

整理可得

$$\begin{aligned} f(\tilde{u}_i) + \gamma_1 h(\tilde{v}_i) + \gamma_2 r(\tilde{\lambda}_i) - \\ f'(u_i)\tilde{u}_i - \gamma_1 h'(v_i)\tilde{v}_i - \gamma_2 r'(\lambda_i)\tilde{\lambda}_i \leq \\ f(u_i) + \gamma_1 h(v_i) + \gamma_2 r(\lambda_i) - \\ f'(u_i)u_i - \gamma_1 h'(v_i)v_i - \gamma_2 r'(\lambda_i)\lambda_i \end{aligned} \quad (40)$$

将所有的 i 相加可得

$$\begin{aligned} \sum_{i=1}^n f(\tilde{u}_i) + \gamma_1 \sum_{i=1}^d h(\tilde{v}_i) + \gamma_2 \sum_{i=1}^d r(\tilde{\lambda}_i) - \\ \sum_{i=1}^n f'(u_i)\tilde{u}_i - \gamma_1 \sum_{i=1}^d h'(v_i)\tilde{v}_i - \gamma_2 \sum_{i=1}^d r'(\lambda_i)\tilde{\lambda}_i \leq \\ \sum_{i=1}^n f(u_i) + \gamma_1 \sum_{i=1}^d h(v_i) + \gamma_2 \sum_{i=1}^d r(\lambda_i) - \\ \sum_{i=1}^n f'(u_i)u_i - \gamma_1 \sum_{i=1}^d h'(v_i)v_i - \gamma_2 \sum_{i=1}^d r'(\lambda_i)\lambda_i \end{aligned} \quad (41)$$

依据式(19)和式(20), 可将式(41)进一步表示为

$$\begin{aligned} \sum_{i=1}^n f(\tilde{u}_i) + \gamma_1 \sum_{i=1}^d h(\tilde{v}_i) + \gamma_2 \sum_{i=1}^d r(\tilde{\lambda}_i) - \\ \sum_{i=1}^n \mathbf{D}_1(i, i)\tilde{u}_i - \gamma_1 \sum_{i=1}^d \mathbf{D}_2(i, i)\tilde{v}_i - \gamma_2 \sum_{i=1}^d r'(\lambda_i)\tilde{\lambda}_i \leq \\ \sum_{i=1}^n f(u_i) + \gamma_1 \sum_{i=1}^d h(v_i) + \gamma_2 \sum_{i=1}^d r(\lambda_i) - \\ \sum_{i=1}^n \mathbf{D}_1(i, i)u_i - \gamma_1 \sum_{i=1}^d \mathbf{D}_2(i, i)v_i - \\ \gamma_2 \sum_{i=1}^d r'(\lambda_i)\lambda_i \end{aligned} \quad (42)$$

当 $\mathbf{D}_1, \mathbf{D}_2$ 和 $\mathbf{D}_3$ 固定时, 由式(24)可得

$$\begin{aligned} \sum_{i=1}^n \mathbf{D}_1(i, i)\tilde{u}_i + \gamma_1 \sum_{i=1}^d \mathbf{D}_2(i, i)\tilde{v}_i + \\ \gamma_2 \text{Tr}(\mathbf{D}_3 g(\tilde{\mathbf{W}})) \leq \sum_{i=1}^n \mathbf{D}_1(i, i)u_i + \\ \gamma_1 \sum_{i=1}^d \mathbf{D}_2(i, i)v_i + \gamma_2 \text{Tr}(\mathbf{D}_3 g(\mathbf{W})) \end{aligned} \quad (43)$$

由引理 2 可得

$$\gamma_2 \text{Tr}(\mathbf{D}_3 g(\mathbf{W})) = \gamma_2 \sum_{i=1}^d r'(\lambda_i) \lambda_i \quad (44)$$

$$\gamma_2 \sum_{i=1}^d r'(\lambda_i) \tilde{\lambda}_i \leq \gamma_2 \text{Tr}(\mathbf{D}_3 g(\tilde{\mathbf{W}})) \quad (45)$$

将式(42)、式(43)、式(44)和式(45)相加可得

$$\begin{aligned} \sum_{i=1}^n f(\tilde{u}_i) + \gamma_1 \sum_{i=1}^d h(\tilde{v}_i) + \gamma_2 \sum_{i=1}^d r(\tilde{\lambda}_i) \leq \\ \sum_{i=1}^n f(u_i) + \gamma_1 \sum_{i=1}^d h(v_i) + \gamma_2 \sum_{i=1}^d r(\lambda_i) \end{aligned} \quad (46)$$

即为

$$\sum_{i=1}^n f(\tilde{u}_i) + \gamma_1 \sum_{i=1}^d h(\tilde{v}_i) + \gamma_2 \text{Tr}(r(\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T + \epsilon \mathbf{I})) \leq$$

$$\sum_{i=1}^n f(u_i) + \gamma_1 \sum_{i=1}^d h(v_i) + \gamma_2 \text{Tr}(r(\mathbf{W}\mathbf{W}^T + \epsilon \mathbf{I})) \quad (47)$$

综上可证,当 $f(\cdot), h(\cdot), r(\cdot)$ 全部为单调不减的可微凹函数时,优化模型式(16)的值单调递减,又由于其值非负有下界,所以 GSLRRFS 算法收敛。

3 实验结果及分析

为了验证 GSLRRFS 算法的有效性,首先在人工合成数据集上与 $l_{2,1}$ -SRFS^[10], $l_{2,p}$ -SRFS^[11], GSRFS^[14] 和 GRLSR^[27] (Generalized and Robust Least Squares Regression) 等 4 种 SRFS 算法进行分类对比实验,验证 GSLRRFS 算法引入低秩正则化项可以有效挖掘类与类之间的潜在相关性,提升算法分类精度。之后,在 7 个真实数据集上除了与上述算法对比之外,还加入拉普拉斯分数^[28] (Laplacian Score, LS)和 SPCAFS^[29] (Sparse Principal Component Analysis for Feature Selection) 两种无监督特征选择算法,3 种经典的非稀疏特征选择算法:RF (ReliefF)^[30]、TR (Trace ratio)^[31] 和最大相关最小冗余 (Minimal-Redundancy-Maximal-Relevance Criterion, mRMR) 算法^[32],以及判别特征选择算法:S²DFS^[7] (Subspace Sparsity Discriminative Feature Selection) 算法;此外,对 GSLRRFS 算法的运行效率、收敛性和参数敏感性进行了实验分析,以全面评估 GSLRRFS 算法性能。

3.1 实验设置

考虑 GSLRRFS 算法为广义形式,采用对数函数和幂函数两类具体形式的函数对其进行验证,给出 log-GSLRRFS 和 l -GSLRRFS 算法。

log-GSLRRFS 算法回归损失项和稀疏正则化项均对数函数,低秩正则化项采用幂函数,函数形式分别为 $f(x) = \log_{\alpha+1}(\alpha x + 1)$, $h(x) = \log_{\beta+1}(\beta x + 1)$, $r(x) = x^{p/2}$ 此时目标函数为

$$\min_{\mathbf{W}} \sum_{i=1}^n \log_{\alpha+1}(\alpha \|\mathbf{x}_i^T \mathbf{W} - \mathbf{y}^i\|_2^2 + 1) + \gamma_1 \sum_{i=1}^d \log_{\beta+1}(\beta \|\mathbf{w}^i\|_2^2 + 1) + \gamma_2 \text{Tr}((\mathbf{W}\mathbf{W}^T + 10^{-6} \times \mathbf{I})^{\frac{p}{2}}) \quad (48)$$

式中: $\alpha > 0; \beta > 0; 0 < p \leq 1$ 。

l -GSLRRFS 算法回归损失项、稀疏正则化项和低秩正则化项均为幂函数,函数形式分别为 $f(x) = x^{\alpha/2}$, $h(x) = x^{\beta/2}$, $r(x) = x^{p/2}$ 形式,此时目标函数为

$$\min_{\mathbf{W}} \sum_{i=1}^n (\|\mathbf{x}_i^T \mathbf{W} - \mathbf{y}^i\|_2^2 + 10^{-6})^{\frac{\alpha}{2}} +$$

$$\gamma_1 \sum_{i=1}^d (\|\mathbf{w}^i\|_2^2 + 10^{-6})^{\frac{\beta}{2}} + \gamma_2 \text{Tr}((\mathbf{W}\mathbf{W}^T + 10^{-6} \times \mathbf{I})^{\frac{p}{2}}) \quad (49)$$

式中: $\alpha > 0; 0 < \beta < 2; 0 < p \leq 1$ 。

为确保实验评估的公平性,GSRFS 算法的回归损失项和稀疏正则化项分别采用幂函数 $f(x) = x^{\alpha/2}$ 和 $h(x) = x^{\beta/2}$ 的形式进行试验,使其与 l -GSLRRFS 算法中的度量形式一致。

所有实验均在配置为 Windows 10 操作系统、搭载 Intel Core i5-8500 3.0 GHz CPU 以及 16GiB 内存的环境中进行。实验过程中,先对所有数据进行标准化和中心化预处理,再使用五重交叉验证法随机抽取样本,其中 80% 的样本作为训练集,20% 的样本作为测试集,利用 K-最近邻 (K-Nearest Neighbor, KNN) 算法作为分类器,选取算法各参数组合下五重交叉验证的最优平均分类精度作为最终的分类结果。各算法参数设置: $l_{2,1}$ -SRFS^[10]、 $l_{2,p}$ -SRFS^[11]、GRLSR^[27]、SPCAFS^[29] 和 GSRFS^[14] 算法中正则化参数 γ 取值范围集合为 $\{10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$; $l_{2,p}$ -SRFS、SPCAFS 和 GSLRRFS 算法中参数 p 取值范围集合为 $\{0.25, 0.5, 0.75, 1\}$; GSLRRFS 算法中正则化参数 γ_1 和 γ_2 的取值范围集合分别为 $\{10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$ 和 $\{0, 10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$; GSRFS 和 GSLRRFS 算法中参数 α 和 β 集合分别设置为 $\{0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3\}$ 和 $\{0.25, 0.5, 0.75, 1, 1.5\}$ 。

3.2 人工合成数据实验

为验证加入 GSLRRFS 算法中低秩正则化项的有效性,构建人工合成数据集 $\mathbf{X}_1$ 。数据集 $\mathbf{X}_1$ 包含 220 个样本,每个样本具有 1 024 维特征,类数为 20,其中后 5 个类是由前 15 类经不同的线性组合方式所构成,因此 $\mathbf{X}_1$ 具有线性相关的类。实验分别选取前 20%、40%、60% 以及 80% 特征,不同特征选择算法在数据集 $\mathbf{X}_1$ 上的分类精度如表 1 所示。

表 1 不同特征选择算法在数据集 $\mathbf{X}_1$ 上的分类精度 %

算法	20%特征	40%特征	60%特征	80%特征
$l_{2,1}$ -SRFS	61.87	61.43	61.89	60.11
$l_{2,p}$ -SRFS	66.46	64.14	62.37	61.87
GRLSR	64.58	62.81	61.89	61.00
GSRFS	61.85	61.01	61.89	61.88
log-GSLRRFS	66.45	65.55	64.64	64.19
l -GSLRRFS	66.92	65.56	65.06	64.67

由表 1 可以看出,log-GSLRRFS 和 l -GSLR-RFS 分类精度明显优于其他 4 种算法。特别是,当选择的特征比例为 60%和 80%时, l -GSLRRFS 相较于其他 4 种对比算法中的最佳分类精度提高了约 2.7 个百分点。因此,加入低秩可以提高算法的分类精度。

3.3 真实数据实验

为了进一步评估 GSLRRFS 算法性能,在 7 个真实数据集上,除了将其与 10 种特征选择算法进行比较外,还对 l -GSLRRFS 算法的收敛性和参数敏感性进行实验分析。真实数据集包括 1 个基因数据集:COLON^[33]和 6 个图像数据集:COIL20, USPS^[33], UMIST^[34], Yale^[35]、ORL^[36]和 YaleB。

表 3 特征比例为 20%的不同特征选择算法的分类精度对比												%
数据集	LS	SPCAFS	mRMR	RF	TR	S ² DFS	$l_{2,1}$ -SRFS	$l_{2,p}$ -SRFS	GRLSR	GSRFS	log-GSLRRFS	l -GSLRRFS
COLON	64.74	76.03	77.44	79.10	74.23	72.82	78.97	77.44	79.10	80.90	79.10	83.85
COIL20	91.74	99.17	99.79	99.86	98.62	99.65	98.82	99.31	99.58	98.68	99.51	99.65
USPS	83.05	93.84	93.85	92.24	89.04	91.04	96.27	—	96.03	96.15	95.20	96.15
UMIST	96.22	98.80	98.63	98.45	98.80	98.63	98.80	99.15	99.15	98.97	99.32	99.32
Yale	62.49	63.13	61.88	60.70	65.02	63.63	64.27	64.23	66.09	68.00	66.71	68.00
ORL	87.75	91.25	90.50	93.00	92.50	92.25	94.00	95.00	94.75	94.50	95.00	95.75
YaleB	31.41	79.07	80.88	83.88	81.17	72.41	76.32	—	85.24	75.82	86.10	86.47
平均精度	73.91	85.90	86.14	86.75	85.63	84.35	86.78	87.03	88.56	87.57	88.71	89.88

表中,“—”表示对应的算法实验迭代过长,无法得到分类结果。由表 3 可以得出,GSLRRFS 算法的平均分类精度均优于其他算法。其中,log-GSLRRFS 和 l -GSLRRFS 算法比 GSRFS 算法分别高出 1.14 和 2.31 个百分点。特别是在 YaleB 数据集上,log-GSLRRFS 和 l -GSLRRFS 算法的分类精度比 GSRFS 算法分别高出 10.28 和 10.65 个百分点。

表 4 特征比例为 20%的不同特征选择算法的运行时间对比												s
数据集	LS	SPCAFS	mRMR	RF	TR	S ² DFS	$l_{2,1}$ -SRFS	$l_{2,p}$ -SRFS	GRLSR	GSRFS	log-GSLRRFS	l -GSLRRFS
COLON	0.01	144.24	19.38	0.59	0.08	187.93	0.06	0.48	2.58	12.04	1.90	1.69
COIL20	0.54	27.83	6.42	12.98	0.61	23.00	1.38	0.59	1.65	6.95	1.11	0.97
USPS	10.90	9.77	11.62	92.38	11.92	0.90	157.58	—	184.85	34.92	3.68	3.23
UMIST	0.25	12.79	1.77	1.58	0.05	16.50	0.12	0.10	0.25	1.64	0.33	0.18
Yale	0.01	64.19	2.85	0.37	0.02	14.83	0.03	0.23	0.28	2.41	0.62	0.40
ORL	0.03	68.43	3.72	1.23	0.04	16.46	0.12	0.37	1.03	3.04	1.30	0.79
YaleB	0.65	30.48	9.20	27.10	0.65	19.36	5.33	—	4.58	11.37	2.18	1.84
平均时间	1.77	51.10	7.85	19.46	1.91	39.85	23.52	0.35	27.89	10.34	1.59	1.30

3.3.3 收敛性分析

为验证 GSLRRFS 算法的收敛性,将 l -GSLRRFS 算法参数 $\alpha,\beta,p,\gamma_1,\gamma_2$ 分别固定为 0.25,0.75,0.25, $10^2,10^2$,不同数据集上目标函数值的收敛曲线如图 2

各真实数据集相关特性如表 2 所示。

表 2 真实数据集统计特性			
数据集	样本数	维数	类数
COLON	62	2 000	2
COIL20	1 440	1 024	20
USPS	9 298	256	10
UMIST	574	644	20
Yale	165	1 024	15
ORL	400	1 024	40
YaleB	2 432	4 096	38

3.3.1 分类精度对比

实验中各算法特征选择的比例分别为 20%, 40%,60%和 80%,下面给出特征选择为 20%时不同算法的分类精度,结果如表 3 所示。

3.3.2 运行时间对比

为验证 GSLRRFS 算法的高效性,不同特征选择算法选取 20%特征时的运行时间对比如表 4 所示。由表 4 可知,在 7 个真实数据集上,GSLRRFS 算法运行效率相较于 GSRFS 算法有了明显提升,且平均运行时间缩减为原来的约十分之一。特别是在 USPS 数据集上,除 S²DFS 算法外,GSLRRFS 算法运行效率均快于其他算法。

所示。从图 2 中可以看出,目标函数值随着迭代次数的增加呈递减趋势。除在 USPS 数据集上目标函数于第 6 次迭代后收敛外,其他数据集均在第 2 次迭代后收敛,因此 GSLRRFS 算法收敛速度较快。

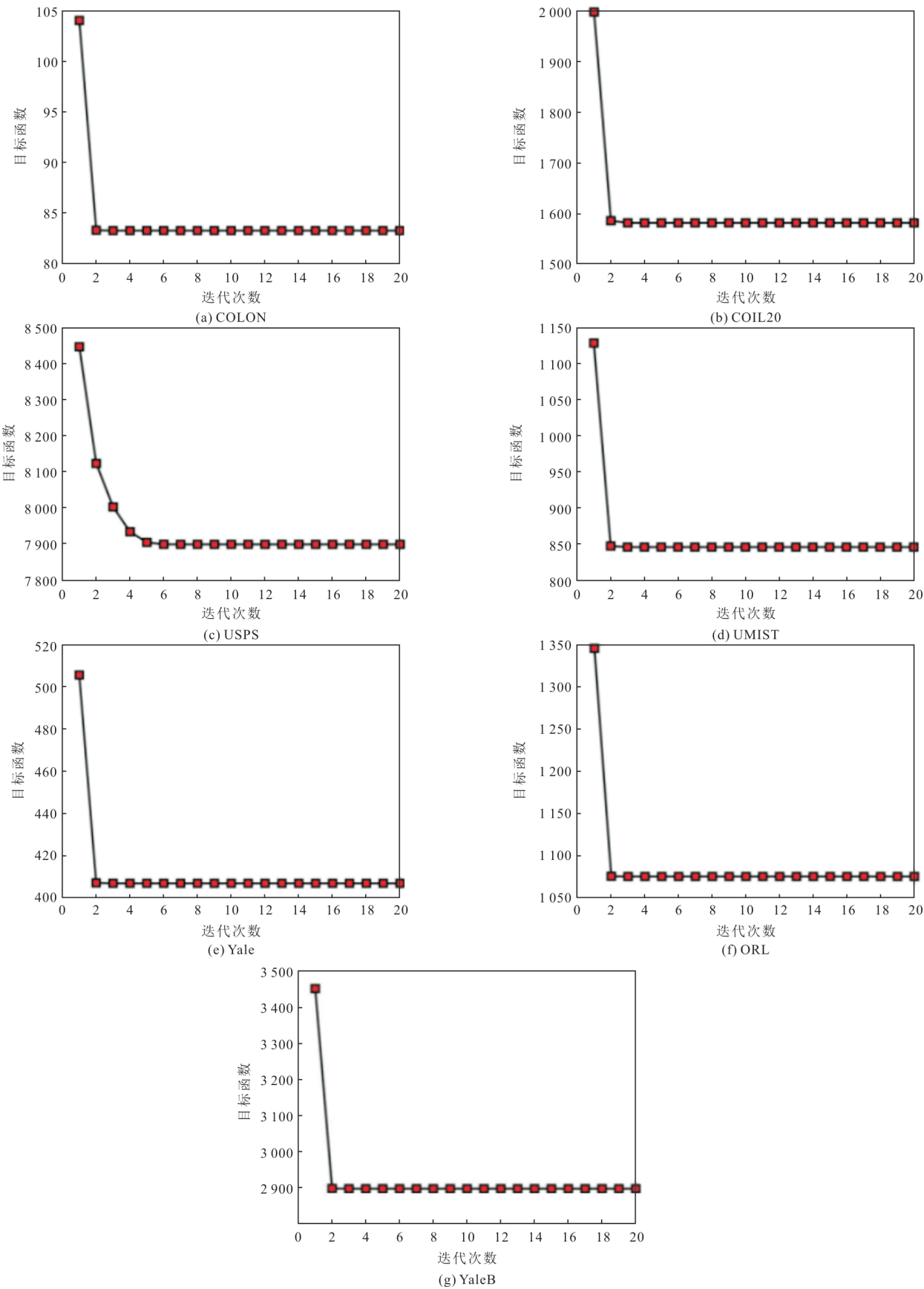


图 2 目标函数收敛曲线

3.3.4 参数敏感性分析

GSLRRFS 算法中含有多个参数,为验证选取

参数范围的有效性,以 l -GSLRRFS 为例,分析其 α , β , γ_1 , γ_2 , p 共计 5 个参数对算法分类精度的影响。

对于参数 α , 实验中固定 $\gamma_1=1, \gamma_2=1, \beta=1, p=1$, α 的取值范围为集合 $\{0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3\}$, 随着 α 取值的变化, 算法的分类精度如图 3 所示。

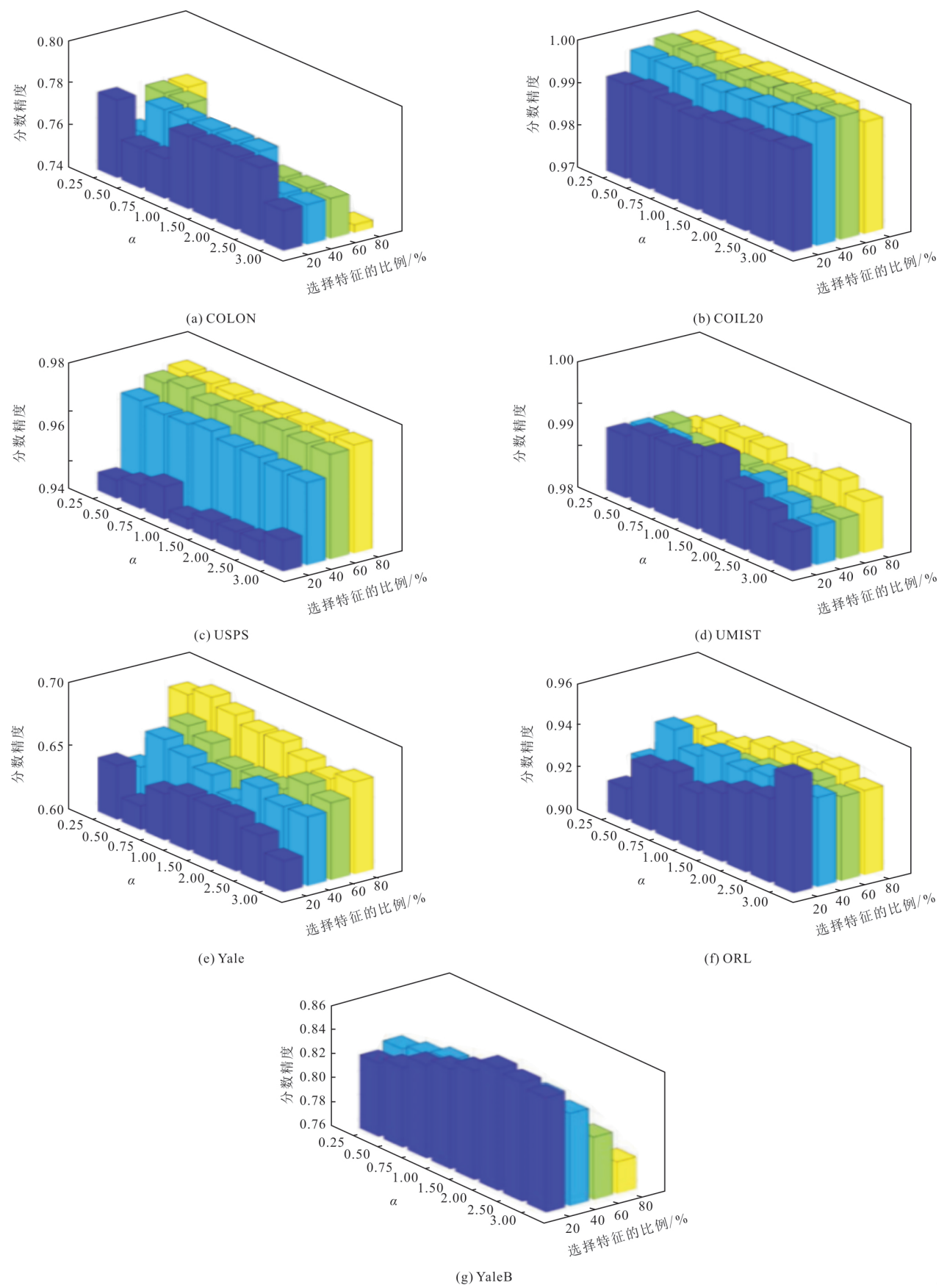


图 3 参数对分类精度的影响

从图3可以看出, α 的取值的变化对分类精度具有一定影响,且在大部分数据集上, α 取区间 $[1.5, 2]$ 内的值,算法可实现较好的分类精度。类似地,亦分析了其余参数对算法分类精度的影响,从而给出其他4个参数的经验取值及范围: $\beta=0.75, \gamma_1 \in [10^4, 10^6], \gamma_2 \in [10^4, 10^6], p=1$ 。

4 结语

已有的 GSRFS 给出了 SRFS 的统一形式,为实际应用提供了更多可供选择的模型,但该算法和已有的其他 SRFS 算法都忽略了数据类与类之间的潜在相关性,当数据的类与类确实存在潜在相关关系时,这些算法将不能实现最优的特征选择。鉴于此,在已有的 GSRFS 的模型基础上引入广义低秩正则化项,进而提出了一种基于广义稀疏低秩回归的特征选择算法:GSLRRFS。GSLRRFS 利用对回归系数矩阵低秩化实现对类与类之间潜在相关性信息的捕捉,从而实现更优的特征选择。同时,GSLRRFS 优化问题可以转化为迭代重加权最小二乘问题,从而可采用最小平方 QR 分解算法高效率求解。实验结果表明,相较于 GSRFS 和其他相关算法,GSLRRFS 在选择 20% 比例的特征时分类精度有一定的提升,表明 GSLRRFS 具有更优的特征选择能力。此外,GSLRRFS 的平均运行时间缩减为 GSRFS 的约十分之一,这表明所提出的基于迭代重加权最小二乘的求解框架是有效的。GSLRRFS 是 SRFS 的更一般化框架。在实际应用中,针对不同特点的数据选择合适的 GSLRRFS 具体形式是重要的问题,这一问题将构成未来的研究工作。

参 考 文 献

- [1] 项颂阳,许章华,张艺伟,等. 高光谱图像分类的 Relief-RFE 特征选择算法构建与应用[J]. 光谱学与光谱分析, 2022, 42(10): 3283-3290.
Xiang Songyang, Xu Zhanghua, Zhang Yiwei, et al. Construction and application of Relief-RFE feature selection algorithm for hyperspectral image classification[J]. Spectroscopy and Spectral Analysis, 2022, 42(10): 3283-3290.
- [2] Luo T J, Hou C P, Nie F P, et al. Semi-supervised feature selection via insensitive sparse regression with application to video semantic recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(10): 1943-1956.
- [3] 张俐. 基因数据的交互依赖特征选择算法[J]. 电子科技大学学报, 2022, 51(5): 754-759.
Zhang Li. An algorithm for cross-dependent feature selection of genetic data[J]. Journal of University of Electronic Science and Technology of China, 2022, 51(5): 754-759.
- [4] 高慧敏,王云鹤,卞闯,等. 基于混合进化算法的特征选择方法研究[J]. 电子学报, 2023, 51(6): 1619-1636.
Gao Huimin, Wang Yunhe, Bian Chuang, et al. Research on feature selection based on hybrid evolutionary algorithm[J]. Acta Electronica Sinica, 2023, 51(6): 1619-1636.
- [5] Li X P, Wang Y D, Ruiz R. A survey on sparse learning models for feature selection[J]. IEEE Transactions on Cybernetics, 2022, 52(3): 1642-1660.
- [6] 施启军,潘峰,龙福海,等. 特征选择方法研究综述[J]. 微电子学与计算机, 2022, 39(3): 1-8.
Shi Qijun, Pan Feng, Long Fuhai, et al. A review of feature selection methods [J]. Microelectronics & Computer, 2022, 39(3): 1-8.
- [7] Nie F P, Wang Z, Tian L, et al. Subspace sparse discriminative feature selection[J]. IEEE Transactions on Cybernetics, 2022, 52(6): 4221-4233.
- [8] Maulud D, Abdulazeez A M. A review on linear regression comprehensive in machine learning[J]. Journal of Applied Science and Technology Trends, 2020, 1(2): 140-147.
- [9] 支晓斌,武少茹. 基于最小平方 QR 分解的改进鲁棒特征选择[J]. 西安邮电大学学报, 2019, 24(6): 35-41.
Zhi Xiaobin, Wu Shaoru. Improved robust feature selection based on least square QR-factorization [J]. Journal of Xi'an University of Posts and Telecommunications, 2019, 24(6): 35-41.
- [10] Nie F, Huang H, Cai X, et al. Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization [C]//24th International Conference on Neural Information Processing Systems. Vancouver: ACM, 2010: 1813-1821.
- [11] Wang L P, Chen S C, Wang Y P. A unified algorithm for mixed ℓ_1 -minimizations and its application in feature selection[J]. Computational Optimization and Applica-

- tions, 2014, 58(2): 409-421.
- [12] Zhi X, Liu J, Wu S, et al. A generalized ℓ_1 -norm regression based feature selection algorithm[J]. Journal of Applied Statistics, 2023, 50(3): 703-723.
- [13] Paige C C, Saunders M A. LSQR: An algorithm for sparse linear equations and sparse least squares[J]. ACM Transactions on Mathematical Software, 1982, 8(1): 43-71.
- [14] Nie F P, Hu Z X, Wang X Q, et al. Iteratively re-weighted method for sparsity-inducing norms [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7045-7055.
- [15] Reinsel G C, Velu R P. Multivariate reduced-rank regression: Theory and applications[M]. New York: Springer, 1998: 17-18.
- [16] Argyriou A, Evgeniou T, Pontil M. Multi-task feature learning [C]//20th International Conference on Neural Information Processing Systems. [S. l.]: ACM, 2006: 41-48.
- [17] Amit Y, Fink M, Srebro N, et al. Uncovering shared structures in multiclass classification[C]//24th International Conference on Machine Learning. Corvalis: ACM, 2007: 17-24.
- [18] Shalev-Shwartz S, Wexler Y, Shashua A. Shareboost: Efficient multiclass learning with feature sharing[C]//Advances in Neural Information Processing Systems 24 (NIPS 2011). New York: Curran Associates, Inc, 2011: 1179-1187.
- [19] Anderson T W. Estimating linear restrictions on regression coefficients for multivariate normal distributions[J]. The Annals of Mathematical Statistics, 1951, 22(3): 327-351.
- [20] Cai X, Ding C, Nie F P, et al. On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions[C]//19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago: ACM, 2013: 1124-1132.
- [21] 杨利锋, 林大华, 邓振云, 等. 低秩特征选择多输出回归算法[J]. 计算机工程与应用, 2017, 53(20): 116-121.
- Yang Lifeng, Lin Dahua, Deng Zhenyun, et al. Low-rank feature selection for multiple-output regression algorithm[J]. Computer Engineering and Applications, 2017, 53(20): 116-121.
- [22] 胡荣耀, 刘星毅, 程德波, 等. 基于稀疏学习的低秩属性选择算法[J]. 计算机工程与应用, 2017, 53(10): 132-138.
- Hu Rongyao, Liu Xingyi, Cheng Debo, et al. Low-rank feature selection algorithm based on sparse learning [J]. Computer Engineering and Applications, 2017, 53(10): 132-138.
- [23] Wright J, Ganesh A, Rao S, et al. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization[C]//Advances in Neural Information Processing Systems. Vancouver: Curran Associates, Inc., 2009: 2080-2088.
- [24] Petersen K B, Pedersen M S. The matrix cookbook [J]. Technical University of Denmark, 2008, 7(15): 510.
- [25] Frommer A, Simoncini V. Matrix functions[M]//Model Order Reduction: Theory, Research Aspects and Applications. Berlin: Springer, 2008: 275-303.
- [26] Marshall A W, Olkin I, Arnold B C. Inequalities: Theory of majorization and its applications[M]. New York: Springer, 2011: 341.
- [27] Wang J Y, Xie F Y, Nie F P, et al. Generalized and robust least squares regression[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(5): 7006-7020.
- [28] He X F, Cai D, Niyogi P. Laplacian score for feature selection[C]//19th International Conference on Neural Information Processing Systems. [S. l.]: ACM, 2005: 507-514.
- [29] Li Z X, Nie F P, Bian J T, et al. Sparse PCA via ℓ_2 , p -norm regularization for unsupervised feature selection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 5322-5328.
- [30] Kononenko I. Estimating attributes: Analysis and extensions of RELIEF[C]//Machine Learning: ECML-94. Berlin: Springer, 1994: 171-182.
- [31] Nie F P, Xiang S M, Jia Y Q, et al. Trace ratio criterion for feature selection[C]//23rd National Conference on Artificial Intelligence. [S. l.]: ACM, 2008: 671-676.
- [32] 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.
- [33] Tao H, Hou C P, Nie F P, et al. Effective discriminative feature selection with nontrivial solution[J]. IEEE Transactions on Neural Networks and Learning

Systems, 2016, 27(4): 796-808.

[34] Xiong H L, Swamy M N S, Ahmad M O. Two-dimensional FLD for face recognition[J]. Pattern Recognition, 2005, 38(7): 1121-1124.

[35] Pei X B, Chen C B, Guan Y. Joint sparse representation and embedding propagation learning: A framework for graph-based semisupervised learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(12): 2949-2960.

[36] Chen X H, Lu Y. Dynamic graph regularization and label relaxation-based sparse matrix regression for two-dimensional feature selection[J]. IEEE Access, 2020, 8: 62855-62870.



高攀(2002—),女,陕西杨凌人,西安邮电大学硕士研究生,主要研究方向为高维数据分析方法以及在图像和视频分析中的应用。E-mail: 3316509194@qq.com



孟玉(1998—),女,陕西西安人,西安邮电大学硕士研究生,主要研究方向为高维数据分析方法以及在图像和视频分析中的应用。E-mail: meng_yu8@163.com

[作者简介]

[责任编辑:祝 剑]



支晓斌(1976—),男,内蒙古巴彦淖尔人,博士,西安邮电大学副教授,主要研究方向为高维数据分析方法以及在图像和视频分析中的应用。E-mail:xbzhi@xupt.edu.cn