

刘威,朱乙鑫,白润才,等.联合不相关回归和潜在表示的无监督特征选择[J].辽宁工程技术大学学报(自然科学版),2025,44(4):495-504.DOI:10.11956/j.issn.1008-0562.20240327
LIU Wei,ZHU Yixin,BAI Runcai,et al.Joint uncorrelated regression and latent representation for unsupervised feature selection[J].Journal of Liaoning Technical University(Natural Science),2025,44(4):495-504.DOI:10.11956/j.issn.1008-0562.20240327

联合不相关回归和潜在表示的无监督特征选择

刘威^{1,2,3},朱乙鑫^{1,2,3},白润才⁴,高琪^{1,2,3},李晓红⁵

(1. 辽宁工程技术大学理学院,辽宁阜新123000;2. 辽宁工程技术大学智能工程与数学研究院,辽宁阜新123000;3. 辽宁工程技术大学数学与系统科学研究所,辽宁阜新123000;4. 辽宁工程技术大学辽宁省高等学校矿产资源开发利用技术及装备研究院,辽宁阜新123000;5. 阜新蒙古族自治县职业教育中心,辽宁阜新123100)

摘要:针对基于图的无监督特征选择算法存在挖掘数据内在信息不充分,且易受噪声干扰难以获取更具有判别性特征的问题,提出一种基于广义不相关回归和潜在表示学习的无监督特征选择方法(unrelated regression and latent representation for unsupervised feature selection, URLUFS)。该方法将非负矩阵分解作用于广义不相关回归模型的投影矩阵,使投影矩阵实现非线性的维数约简并获得特征选择矩阵。在特征选择矩阵的基础上,引入自适应图学习来进一步挖掘数据的局部流形结构,并对特征选择矩阵施加范数约束以保持稀疏性。利用潜在表示对数据样本间的相互关系进行学习,引导回归模型中的伪标签矩阵,从而选择出更具有判别性的特征。在8个公开的数据集上进行了数值对比实验,实验结果表明:基于广义不相关回归和潜在表示学习的无监督特征选择算法明显优于其他8种无监督特征选择算法。

关键词:无监督特征选择;广义不相关回归;非负矩阵分解;潜在表示学习;自适应图学习

中图分类号:TP391

文献标志码:A

文章编号:1008-0562(2025)04-0495-10

Joint uncorrelated regression and latent representation for unsupervised feature selection

LIU Wei^{1,2,3}, ZHU Yixin^{1,2,3}, BAI Runcai⁴, GAO Qi^{1,2,3}, LI Xiaohong⁵

(1. School of Sciences, Liaoning Technical University, Fuxin 123000, China; 2. Institute of Intelligence Engineering and Mathematics, Liaoning Technical University, Fuxin 123000, China; 3. Institute of Mathematics and Systems Science, Liaoning Technical University, Fuxin 123000, China; 4. Liaoning Academy of Mineral Resources Development and Utilization Technology and Equipment, Liaoning Technical University, Fuxin 123000, China; 5. Fuxin Mongolian Autonomous County Vocational Education Center, Fuxin 123100, China)

Abstract:Addressing on the issues that most graph-based unsupervised feature selection algorithms often fail to fully mine the intrinsic information of data and are prone to noise interference, making it difficult to obtain more discriminative features, this study proposes an unsupervised feature selection based on generalized uncorrelated regression and latent representation learning (unrelated regression and latent representation for unsupervised feature selection, URLUFS). Non-negative matrix factorization is applied to the projection matrix of the generalized uncorrelated regression model to achieve nonlinear dimensionality reduction of the projection matrix and obtain the feature selection matrix. The latent representation is used to learn the mutual relationship between data samples, thereby guiding the pseudo-label matrix in the regression model. Adaptive graph learning is introduced to exploit the local manifold structure of the data, and norm constraints are imposed on the feature selection matrix to maintain sparsity, thereby selecting more discriminative features. Experiments are carried out on eight public datasets. The experimental results show that the proposed algorithm significantly outperforms the

收稿日期:2024-10-08 修回日期:2024-12-22 接受日期:2025-01-04 责任编辑:张雯

基金项目:国家自然科学基金项目(52374123);辽宁省教育厅高校基本科研项目(LJ212410147013;LJ212410147019);辽宁省“兴辽英才计划”项目(XLYC2211085)

作者简介:刘威(1977-)男,辽宁阜新新人,博士,副教授,主要从事机器学习、深度学习和矿业系统工程等方面的研究。

E-mail: lv8218218@126.com

other eight advanced unsupervised feature selection algorithms.

Keywords: unsupervised feature selection; generalized uncorrelated regression; nonnegative matrix factorization; latent representation learning; adaptive graph learning

0 引言

随着人工智能的发展,各行业数据量及其特征维度同步激增。这些高维数据虽能提供丰富的数据信息,但存在大量不相关、冗余和噪声特征,带来了巨大的计算压力,增加了过拟合的风险,导致模型泛化能力下降,出现“维度灾难”等问题^[1-3]。因此,有效降低高维数据的维数已成为当务之急。特征选择是一种数据降维的有效手段,核心是从高维特征中选择信息量最大且最具代表性的特征子集,从而缓解上述问题。特征选择根据有无标签划分为有监督特征选择、无监督特征选择和半监督特征选择^[4-6]。无监督特征选择可以分为3类:过滤式、包裹式和嵌入式^[7-9]。实际场景中,无标签或标签难以获取的数据居多,所以对无监督特征选择的研究具有重要的现实意义。

基于图的无监督特征选择算法,可以使用谱分析来探索样本数据的内在结构。HE等^[10]提出一种拉普拉斯得分算法(Laplacian score, LS),通过数据的局部结构计算特征的拉普拉斯得分,从而确定重要性。CAI等^[11]提出一种多簇特征选择算法(multi-cluster feature selection, MCFS),通过图拉普拉斯矩阵来表示特征之间的关系,并使用回归模型来避免独立评估每个特征的贡献。NIE等^[12]提出结构化最优图特征选择算法(structured optimal graph feature selection, SOGFS),通过精确连通分量来选择特征的同时捕获数据局部流行结构。ZHANG等^[13]提出一种基于自适应图学习和约束的无监督特征选择算法(unsupervised feature selection via adaptive graph learning and constraint, EGCFS),将相似矩阵的构造嵌入到优化过程中,并与最大化类间散度矩阵结合来进行特征选择。YUAN等^[14]提出具有自适应图的凸非负矩阵分解的无监督特征选择算法(convex non-negative matrix factorization with adaptive graph constraint, CNAFS)。与上述两种利用欧氏距离构造矩阵的方法不同,CNAFS采用最大熵原理进行图学习。ZHU等^[15]提出结合图学习和 $L_{2,0}$ 范数约束的无监督特征选择(graph learning unsupervised feature selection, GLUFS),设计了一种迭代方法来求解 $L_{2,0}$ 范数约束的模型,性能较好。这些基于图的无

监督特征选择算法采用投影矩阵来保留局部结构,而原始数据的投影矩阵中包含大量的噪声,会导致算法性能降低。

非负矩阵分解^[16](non-negative matrix factorization, NMF)作为降维和特征提取的方法被广泛应用于数据分析。盛超等^[17]提出基于子空间学习和伪标签回归的无监督特征选择,ZHOU等^[18]提出基于软标签引导非负矩阵分解的无监督特征选择,两者均利用NMF,将子空间学习与回归模型结合进行特征选择。后者与广义不相关回归模型相结合,构建软标签进一步引导回归模型,提高整体模型性能。CUI等^[19]提出一种子空间学习引导凸NMF的无监督特征选择,将凸NMF和多子空间结构学习相结合来选择特征。潜在表示学习作为一种新兴的方法,用来捕获数据的潜在信息,也就是样本间的聚类结构,它可以通过对称非负矩阵分解来实现^[20]。TANG等^[21]提出通过潜在表示学习和流形正则化的无监督特征选择(latent representation learning and graph based manifold regularization, LRLMR)。该算法将潜在表示学习应用到无监督特征选择上,与流形正则化相结合进行特征选择。XIE等^[22]提出无监督特征选择的图和潜在表示的联合学习(joint learning of graph and latent representation for unsupervised feature selection, JGLUFS),将自适应图的构造集成于基于潜在表示学习的特征选择方法中,减少外部条件对图质量的影响。潜在表示学习更注重样本间的信息,同时基于潜在表示的算法大都与普通的回归模型相结合,导致在特征选择过程中无法考虑特征之间的不相关性。

针对上述问题并深入挖掘更具判别性的特征,本文提出一种有效的特征选择算法,融合非负矩阵分解与广义不相关回归模型,使投影矩阵实现非线性的维度约简并且获得更加稀疏表达的特征选择矩阵。构建基于特征选择子空间的自适应图学习,利用稀疏特征选择矩阵保留数据的局部结构,降低原始投影矩阵噪声干扰,从而提高算法性能。将潜在表示学习和广义不相关回归模型相结合,利用潜在表示学习对广义不相关回归模型中的伪标签矩阵进行引导,兼顾了特征之间的不相关性。

1 模型设计

1.1 符号说明

符号说明见表1。 $\mathbf{X}=(x_1, x_2, \dots, x_n)^T$, 其中 $x_i \in \mathbf{R}^d$ 表示数据的第 i 行, 即第 i 个样本。定义 $\mathbf{W}$ 的 $\ell_{2,1}$ 范数为

$$\|\mathbf{W}\|_{2,1} = \sum_{i=1}^d \sqrt{\sum_{j=1}^c w_{ij}^2} = \sum_{i=1}^d \|\mathbf{w}_i\|_2。$$

$\mathbf{W}$ 的 Frobenius 范数为

$$\|\mathbf{W}\|_F = \sqrt{\sum_{i=1}^d \|\mathbf{w}_i\|_2^2} = \sqrt{\text{Tr}(\mathbf{W}^T \mathbf{W})},$$

式中: $\text{Tr}(\cdot)$ 表示求矩阵的迹; $\|\cdot\|_F^2$ 为 Frobenius 范数的平方。

表1 符号说明

Tab.1 notation description

符号	说明
n	样本数
d	特征数
l	子空间维度
c	聚类簇数
$\mathbf{x}_i$	第 i 个数据样本
$\mathbf{X} \in \mathbf{R}^{n \times d}$	数据矩阵
$\mathbf{W} \in \mathbf{R}^{d \times c}$	投影矩阵
$\mathbf{V} \in \mathbf{R}^{d \times l}$	特征选择矩阵
$\mathbf{G} \in \mathbf{R}^{l \times c}$	系数矩阵
$\mathbf{b} \in \mathbf{R}^{c \times 1}$	偏差
$\mathbf{H} \in \mathbf{R}^{n \times c}$	伪标签矩阵
$\mathbf{L} \in \mathbf{R}^{n \times n}$	拉普拉斯矩阵
$\mathbf{D} \in \mathbf{R}^{n \times n}$	度矩阵
$\mathbf{S} \in \mathbf{R}^{n \times n}$	相似矩阵
$\mathbf{A} \in \mathbf{R}^{n \times n}$	潜在相似矩阵

1.2 基于非负矩阵分解的广义不相关回归

广义回归模型考虑特征的不相关性, 输出兼具判别性与独立性特征^[23]。通过学习投影子空间与伪标签空间的线性映射关系, 构建目标函数为

$$\begin{cases} \min_{\mathbf{W}, \mathbf{b}, \mathbf{H}} \|\mathbf{XW} + \mathbf{1b}^T - \mathbf{H}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} \\ \text{s.t. } \mathbf{H} \geq 0 \end{cases}, \quad (1)$$

式中: $\mathbf{b} \in \mathbf{R}^{c \times 1}$ 为偏差; $\mathbf{1} \in \mathbf{R}^n$ 为全 1 列向量; $\mathbf{H} \in \mathbf{R}^{n \times c}$ 是伪标签矩阵; λ 为正则化参数; c 为聚类簇数。

NMF 是一种矩阵分解技术, 通过寻找特征选择矩阵 $\mathbf{V} \in \mathbf{R}^{d \times l}$ 和系数矩阵 $\mathbf{G} \in \mathbf{R}^{l \times c}$ 来表示投影矩阵 $\mathbf{W}$ (l 为子空间维度)。对 NMF 施加 Frobenius 范数约束, NMF 模型的目标函数为

$$\begin{cases} \min_{\mathbf{V}, \mathbf{G}} \|\mathbf{W} - \mathbf{VG}\|_F^2 \\ \text{s.t. } \mathbf{V}, \mathbf{G} \geq 0 \end{cases}。 \quad (2)$$

从原始数据学习得到的投影矩阵含有较多的噪声。NMF 在非负性约束下, 尽可能保持数据信息不变。因此, 将式 (1) 和式 (2) 结合, 使 $\mathbf{W}$ 实现非线性的维数约简并且获得更加稀疏的表达^[24], 从而减少噪声影响, 见图 1。同时, 对获得的特征选择矩阵 $\mathbf{V}$ 施加 $\ell_{2,1}$ 范数, 使选择出来的特征更具有判别力, 构建目标函数为

$$\begin{cases} \min_{\mathbf{V}, \mathbf{G}, \mathbf{b}, \mathbf{H}} \|\mathbf{XVG} + \mathbf{1b}^T - \mathbf{H}\|_F^2 + \lambda \|\mathbf{V}\|_{2,1} \\ \text{s.t. } \mathbf{V}, \mathbf{G}, \mathbf{H}, \geq 0, \mathbf{V}^T \mathbf{V} = \mathbf{I}_l \end{cases}。 \quad (3)$$

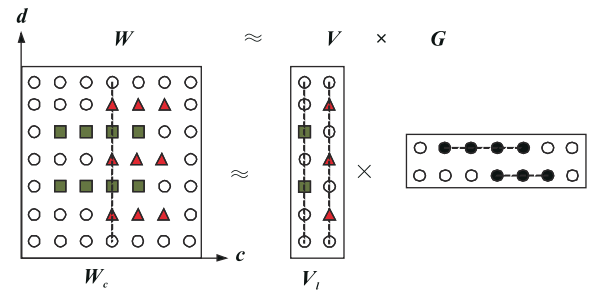


图1 非负矩阵分解

Fig.1 nonnegative matrix factorization

对 $\mathbf{V}$ 施加正交约束, 确保 $\mathbf{V}$ 中的每一行或者每一列至多含有一个非零元素, 从而能够获得更具代表性和判别性的特征选择矩阵。对式 (3) 求导, 可得偏差 $\mathbf{b}$ 的最优解为

$$\mathbf{b} = \frac{1}{n} (\mathbf{H} - \mathbf{XVG})^T \mathbf{1}。 \quad (4)$$

将式 (4) 代入式 (3), 得到目标函数为

$$\min_{\mathbf{V}, \mathbf{G}, \mathbf{H}} \|\mathbf{C}_n (\mathbf{XVG} - \mathbf{H})\|_F^2 + \lambda \|\mathbf{V}\|_{2,1}, \quad (5)$$

式中, $\mathbf{C}_n = \mathbf{I}_n - n^{-1} \times \mathbf{1}_{n \times n}$, $\mathbf{I}_n \in \mathbf{R}^{n \times n}$ 为单位矩阵。

1.3 基于特征选择子空间的自适应图学习

将相似度矩阵融入优化过程, 构建自适应学习机制, 可有效适配样本的局部流形结构, 较好地刻画数据间的真实结构^[25]。局部结构保留理论假设原始空间中两个数据样本 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 之间的距离非常近, 则它在投影子空间中的映射点之间的距离也应该非常近。该理论可以定义为

$$\sum_{i,j=1}^n \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|_2^2 s_{ij} = \text{Tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W}),$$

式中, s_{ij} 为样本点 $\mathbf{x}_i$, $\mathbf{x}_j$ 之间的相似度。

大部分图学习都是基于投影空间进行的, 并且投影空间容易受到噪声影响, 因此通过对特征选择子空间进行局部结构保留, 并对相似度矩阵施加 Frobenius 范数约束, 以达到自适应的效果,

提高特征选择算法的性能。具体表示为

$$\begin{cases} \min_{V, S, \varepsilon} \text{Tr}(V^T X^T L X V) + \varepsilon \|S\|_F^2 \\ \text{s.t.} \sum_{j=1}^n s_{ij} = 1, s_{ij} \geq 0 \end{cases}, \quad (6)$$

式中, L 为拉普拉斯矩阵, $L = D - S$, $S = (s_{ij})_{n \times n} \in \mathbf{R}^{n \times n}$ 为样本的相似矩阵, $D \in \mathbf{R}^{n \times n}$ 为度矩阵, 其对角元素为

$$d_{ii} = \sum_{j=1}^n s_{ij}$$

1.4 基于潜在表示的伪标签学习

传统无监督特征选择方法通常假设样本分布相同并且相互独立, 但这与现实数据特征存在偏差, 因为样本之间存在相互关联^[26]。潜在表示学习虽然能够捕获样本间的关联, 但缺少对特征之间信息的挖掘。因此, 本文将潜在表示的思想嵌入广义不相关回归中来指导伪标签矩阵的学习, 即 URLUFS 算法, 可表示为

$$\min_{H \geq 0} \|A - HH^T\|_F^2, \quad (7)$$

A 中元素为

$$A_{ij} = \exp\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{-2\sigma^2}\right)。$$

综上, 为获取更具有判别力的特征, URLUFS 算法将广义不相关回归作为模型的基础, 式 (5) 对矩阵 W 进行非负矩阵分解得到更加稀疏表述的 V 。式 (7) 对广义不相关回归模型中的 H 使用潜在表示学习进行引导。式 (6) 将原始数据矩阵的局部结构保留在特征选择子空间中, 加入自适应图学习部分, 实现 URLUFS 算法的自适应效果。得到最终的目标函数为

$$\begin{cases} \min_{V, G, H, S, \gamma} \|C_n(XVG - H)\|_F^2 + \alpha \|A - HH^T\|_F^2 + \\ \beta \text{Tr}(V^T X^T L X V) + \varepsilon \|S\|_F^2 + \lambda \|V\|_{2,1} \\ \text{s.t.} \quad V, G, H, \geq 0, V^T V = I, \sum_{j=1}^n s_{ij} = 1, s_{ij} \geq 0 \end{cases}, \quad (8)$$

式中, α 、 β 、 ε 和 λ 为平衡参数。

2 URLUFS 算法

2.1 模型求解

由于所提出的目标函数式 (8), 包含了 V 、 G 、 H 和 S , 是非凸函数且很难同时对变量进行优化求解。因此, 采用交替迭代乘子法进行求解, 即固定其余变量, 更新目标变量, 求解过程如下。

(1) 固定 G 、 H 和 S , 求解 V

为了便于计算, 首先将稀疏正则化项 $\|V\|_{2,1}$ 转换为矩阵形式 $\text{Tr}(V^T O V)$, 定义对角权重矩阵 $O \in \mathbf{R}^{d \times d}$, 它的对角元素定义为

$$O_{ii} = \frac{1}{2\|V_i\|_2} \quad (i=1, \dots, d),$$

式中, V_i 表示 V 中的第 i 行。

其次通过定义两个拉格朗日乘子 δ 、 γ 约束 V , 得到拉格朗日函数。此时关于 V 的函数可以写为

$$\mathcal{L}(V) = \|C_n(XVG - H)\|_F^2 + \beta \text{Tr}(V^T X^T L X V) + \lambda \text{Tr}(V^T O V) + \gamma \|V^T V - I\|_F^2 + \text{Tr}(\delta V^T)。$$

令 $\mathcal{L}(V)$ 对 V 的导数为 0, 求解可得

$$X^T C_n X V G G^T - X^T C_n H G^T + \gamma V V^T V - \gamma V + \lambda O V + \beta X^T (D - S) X V - \delta = 0。$$

使用 KKT 条件 (Karush-Kuhn-Tucker)^[27], 令 δ 、 V 中逐元素满足 $\delta_{ij} V_{ij} = 0$, 得到 V 中各元素的更新式为

$$V_{ij} = V'_{ij} \frac{(X^T C_n H G^T + \gamma V + \beta X^T S X V)_{ij}}{(X^T Z + \lambda O V + \gamma V V^T V + \beta X^T D X V)_{ij}}, \quad (9)$$

式中: V'_{ij} 为上一次迭代的值; $Z = C_n X V G G^T$ 。

(2) 固定 V 、 H 和 S , 求解 G

通过定义拉格朗日乘子 ζ 约束 G , 得到关于 G 的拉格朗日函数, 并令导数为 0, 即

$$V^T X^T C_n X V G - V^T X^T C_n H - \zeta = 0。$$

使用 KKT 条件, 令 ζ 、 G 中逐元素满足 $\zeta_{ij} G_{ij} = 0$, 得到 G 中各元素的更新式为

$$G_{ij} = G'_{ij} \frac{(V^T X^T C_n H)_{ij}}{(V^T X^T C_n X V G)_{ij}}, \quad (10)$$

式中, G'_{ij} 为上一次迭代的值。

(3) 固定 V 、 G 和 S , 求解 H

通过定义拉格朗日乘子 θ 约束 H , 得到关于 H 的拉格朗日函数, 令导数为 0, 即

$$C_n H - C_n X V G + \alpha H H^T H - \alpha A H - \theta = 0。$$

使用 KKT 条件, 令 θ 、 H 中逐元素满足 $\theta_{ij} H_{ij} = 0$, 得到 H 中各元素的更新式为

$$H_{ij} = H'_{ij} \frac{(C_n X V G + \alpha A H)_{ij}}{(C_n H + \alpha H H^T H)_{ij}}, \quad (11)$$

式中, H'_{ij} 为上一次迭代的值。

(4) 固定 V 、 G 和 H , 求解 S

当 V 、 G 和 H 固定时, 可以得到关于 S 的目标函数并转换成向量形式, 有

$$\begin{cases} \min_s \sum_{i,j=1}^n (\beta \|V^T x_i - V^T x_j\|_2^2 s_{ij} + \varepsilon s_{ij}^2) \\ \text{s.t.} \quad \sum_{j=1}^n s_{ij} = 1, s_{ij} \geq 0 \end{cases} \quad (11)$$

令 $d_{ij} = \|V^T x_i - V^T x_j\|_2^2$, 进一步将上式转化为向量形式, 有

$$\min_{s, \mathbf{1}^T \mathbf{s} = 1, s_i \geq 0} \|\mathbf{s}_i + \frac{\beta}{2\varepsilon} \mathbf{d}_i\|_2^2. \quad (12)$$

引入拉格朗日乘子对上式进行求解, 得到的拉格朗日函数为

$$\mathcal{L}(s_i, \omega, \varphi_i) = \left\| \mathbf{s}_i + \frac{\beta}{2\varepsilon} \mathbf{d}_i \right\|_2^2 - \omega (\mathbf{s}_i^T \mathbf{1} - 1) - \varphi_i^T \mathbf{s}_i,$$

式中, $\omega \geq 0$ 和 $\varphi_i \geq 0$ 为拉格朗日乘子。

根据 KKT 条件和互补松弛条件, 最优解应为

$$s_{ij} = \left(-\frac{\beta d_{ij}}{2\varepsilon} + \omega \right)_+.$$

综上所述, URLUFS 模型的求解流程如下。

算法 1 URLUFS 算法。

输入: 原始数据矩阵 $\mathbf{X}$ 以及平衡参数。

输出: 通过 $\|\mathbf{V}\|_2$ 计算所有 d 个特征的评分, 并按降序排列形成新的数据矩阵 $\mathbf{X}_{\text{new}}$ 。

- 1: 初始化矩阵 $\mathbf{V}$ 、 $\mathbf{G}$ 、 $\mathbf{H}$ 和 $\mathbf{C}_n$;
- 2: 构造相似矩阵 $\mathbf{S}$ 和潜在表示矩阵 $\mathbf{A}$;
- 3: **while** not convergence **do**;
- 4: 根据式 (9) 更新 $\mathbf{V}$;
- 5: 根据式 (10) 更新 $\mathbf{G}$;
- 6: 根据式 (11) 更新 $\mathbf{H}$;
- 7: 根据式 (12) 更新 $\mathbf{S}$;
- 8: 根据目标函数式 (8) 计算函数数值;
- 9: **end while**
- 10: 通过 $\|\mathbf{V}\|_2$ 计算所有 d 个特征的评分, 并按降序排列形成新的数据矩阵 $\mathbf{X}_{\text{new}}$ 。

2.2 算法分析

URLUFS 算法可根据算法 1 的流程划分为 4 个子过程: 更新矩阵 $\mathbf{V}$ 、 $\mathbf{G}$ 、 $\mathbf{S}$ 和 $\mathbf{H}$ 。根据式 (9) 更新矩阵 $\mathbf{V}$ 的时间成本为 $O(n^2 d + n^2 c + d^2 l + nd^2)$ 。根据式 (10) 更新矩阵 $\mathbf{G}$ 的时间成本为 $O(n^2 d + n^2 c + ndl)$ 。根据式 (11) 更新矩阵 $\mathbf{H}$ 的时间成本为 $O(n^2 d + n^2 c + ndl)$ 。根据式 (12) 更新矩阵 $\mathbf{S}$ 的时间成本为 $O(n^2)$ 。由于 $c \leq n$ 和 $l \leq d$, 所以总计算复杂度为 $O((n^2 d + n^2 c + d^2 l + nd^2)T)$, T 为迭代次数。

URLUFS 算法目标函数的收敛性证明可分为 4 部分。由于 s_i 的闭式解形式, 用式 (12) 更新可容易证明 $\mathbf{S}$ 是非增的。其次, 根据文献[18]和文献

[30]可容易证明, 其余 3 个变量在式 (9) ~ 式 (11) 下更新也是非增的。

定理 1 函数 $F(h)$ 在以下更新公式下单调递减。

$$h^{(t+1)} = \arg \min_h J(h, h^{(t)}),$$

式中: t 为迭代次数; $J(h, h')$ 是 $F(h)$ 的辅助函数, 满足

$$J(h, h') \geq F(h), J(h, h) = F(h),$$

证明 $F(h^{(t+1)}) \leq J(h^{(t+1)}, h^{(t)}) \leq J(h^{(t)}, h^{(t)}) = F(h^{(t)})$ 。

当 $\mathbf{V}$ 、 $\mathbf{H}$ 和 $\mathbf{S}$ 不变时, 只保留目标函数中包含 $\mathbf{G}$ 的项, 可得

$$F(\mathbf{G}) = \|\mathbf{C}_n(\mathbf{X}\mathbf{V}\mathbf{G} - \mathbf{H})\|_F^2. \quad (13)$$

对式 (13) 逐元素求一阶和二阶导数, 即

$$\begin{cases} F'_{ij}(\mathbf{G}) = (\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V} \mathbf{G} - \mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{H})_{ij} \\ F''_{ij}(\mathbf{G}) = (\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V})_{ij} \end{cases}. \quad (14)$$

引理 1 引入 F_{ij} 的辅助函数为

$$J(\mathbf{G}, \mathbf{G}_{ij}^{(t)}) = F_{ij}(\mathbf{G}_{ij}^{(t)}) + F'_{ij}(\mathbf{G}_{ij}^{(t)}) (\mathbf{G} - \mathbf{G}_{ij}^{(t)}) + \frac{(\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V} \mathbf{G})_{ij}}{\mathbf{G}_{ij}^{(t)}} (\mathbf{G} - \mathbf{G}_{ij}^{(t)})^2. \quad (15)$$

证明: $F_{ij}(\mathbf{G})$ 的泰勒展开为

$$F_{ij}(\mathbf{G}) = F_{ij}(\mathbf{G}_{ij}^{(t)}) + F'_{ij}(\mathbf{G}_{ij}^{(t)}) (\mathbf{G} - \mathbf{G}_{ij}^{(t)}) + \frac{(\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V})_{ij}}{\mathbf{G}_{ij}^{(t)}} (\mathbf{G} - \mathbf{G}_{ij}^{(t)})^2. \quad (16)$$

从式 (15) 和式 (16) 可以看出 $F_{ij}(\mathbf{G}) \leq J(\mathbf{G}, \mathbf{G}^{(t)})$ 等价于式 (17)。

$$(\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V})_{ij} \leq \frac{(\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V} \mathbf{G})_{ij}}{\mathbf{G}_{ij}^{(t)}}. \quad (17)$$

由式 (17) 可知以下公式是有效的:

$$\begin{aligned} \mathbf{G}_{ij}^{(t)} (\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V})_{ij} &\leq \mathbf{G}_{in}^{(t)} \sum_{n=1}^m (\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V})_{nj}, \\ \mathbf{G}_{in}^{(t)} \sum_{n=1}^m (\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V})_{nj} &= (\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V} \mathbf{G})_{ij}. \end{aligned}$$

显然, $F_{ij}(\mathbf{G}) = J(\mathbf{G}, \mathbf{G})$ 且 $F_{ij}(\mathbf{G}) \leq J(\mathbf{G}, \mathbf{G}^{(t)})$ 成立。通过将定理 1 与式 (13) 和式 (14) 相结合, 可得

$$\mathbf{G}_{ij}^{(t+1)} = \arg \min(\mathbf{G}, \mathbf{G}^t) = \mathbf{G}_{ij}^{(t)} \frac{(\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{H})_{ij}}{(\mathbf{V}^T \mathbf{X}^T \mathbf{C}_n \mathbf{X} \mathbf{V} \mathbf{G})_{ij}}. \quad (18)$$

可以发现, 式 (18) 是 $\mathbf{G}$ 的更新公式。所以随着 $\mathbf{G}$ 的最优更新, 目标值单调递减。目标函数其余变量的收敛性证明与 $\mathbf{G}$ 类似, 不再重复。

3 实验与分析

3.1 实验方案

(1) 评价指标。聚类精度 (clustering accuracy, ACC) 和归一化互信息 (normalized mutual information, NMI) 被用作标准评估指标^[28-29]。ACC 和 NMI 的值越高表明性能越好。

ACC 定义为

$$V_{\text{ACC}} = \frac{1}{n} \sum_{i=1}^n \delta(p_i, f_{\text{map}}(m_i)),$$

式中: p_i 为理想标签; $f_{\text{map}}(\cdot)$ 为最佳映射函数; m_i 为预测标签; $\delta(a, b)$ 为指示函数, 如果 $a=b$, 则 $\delta(a, b)=1$, 否则 $\delta(a, b)=0$ 。

NMI 定义为

$$V_{\text{NMI}} = \frac{I(w, m)}{\sqrt{H(w)H(m)}},$$

式中: w 为理想类标签; m 为预测类标签; $I(\cdot)$ 为互信息; $H(\cdot)$ 为信息熵。

(2) 数据集。为了评估 URLUFS 的性能, 在实验中, 使用 8 个测试数据集。这些数据集可以分为 3 种类型: 人脸图像、物体图像和生物数据。数据集信息见表 2。

表 2 数据集信息

Tab.2 detail introduction to datasets

数据集	样本个数	特征个数	聚类个数	数据类型
JAFFE	213	256	10	人脸图像
COIL20	1 440	1 024	20	物体图像
ORL	400	1 024	40	人脸图像
Yale	165	1 024	15	人脸图像
warpAR10P	130	2 400	10	人脸图像
warpPIE10P	210	2 420	10	人脸图像
lung	203	3 312	5	生物数据
TOX-171	174	5 748	4	生物数据

(3) 对比算法。URLUFS 与 8 种无监督特征选择算法 LS^[10]、MCFS^[11]、SOGFS^[12]、EGCFS^[13]、CNAFS^[14]、GLUFS^[15]、LRLMR^[21] 和 JGLUFS^[22] 进行比较。基线 (Baseline) 为选择所有特征进行聚类。

(4) 参数设置。首先, 对于需要构建局部图的方法, 统一将邻域大小固定为 5, 距离函数的高斯核宽度设为 1。其次, 为了确保每种方法都能获得满意的性能, 所有算法的平衡参数的调整范围为 $\{10^{-6}, 10^{-4}, 10^{-2}, 1, 10^2, 10^4, 10^6\}$ 。最后, 采用 K-means 算法进行聚类。考虑到 K-means 算法对初始

化较为敏感, 重复 20 次 K-means 聚类以避免算法中的偶然性, 并使用聚类结果计算 V_{ACC} 和 V_{NMI} 。

3.2 聚类性能分析

图 2 和图 3 为 8 种算法在不同数据集选择不同特征数时对应的 2 种评价指标值, 其中 AllFea 表示基线。由图 2 和图 3 可以看出, URLUFS 算法选择不同特征数量的 V_{ACC} 和 V_{NMI} , 整体优于对比算法。与选用所有特征进行 K-means 聚类的基线相比, URLUFS 算法的性能均提升 3% 左右, 在部分数据集上, 提升尤为明显, 如数据集 warpPIE10P 和数据集 warpAR10P。在数据集 JAFFE 上, 相较于对比算法和 Baseline, URLUFS 算法在所选特征数量范围内的 V_{ACC} 和 V_{NMI} 均提升了 10% 左右。同时, URLUFS 算法与 GLUFS 算法在 JAFFE 数据集上的 V_{ACC} 和 V_{NMI} 明显高于其他算法。URLUFS 算法的性能始终优于基于 $\ell_{2,0}$ 范数约束的 GLUFS 算法, 体现出在稀疏表达的特征选择矩阵上进行特征选择的有效性。在数据集 ORL 上, URLUFS 算法在选择特征数量范围内分类准确率均超过基线值, 实现约 5% 的性能提升。因此, 对比其他算法可以看出, URLUFS 算法在特征选择矩阵上进行特征选择而非在投影矩阵上进行, 可以有效过滤原始数据投影矩阵中的大量噪声, 提高算法的鲁棒性。在数据集 PIE10P 和数据集 AR10P 上, 使用自适应图学习保存数据流形结构的 SOGFS 算法和 CNAFS 算法等虽取得了不错的结果, 但基于特征选择子空间进行自适应图学习的 URLUFS 算法性能更加出色, 更精准地挖掘数据中的隐藏结构, 从而更完整地保留原始数据的流形信息。

在生物数据集 lung 上, URLUFS 算法前期 NMI 值相较于 LRLMR 算法没有明显的提升。但随着选择特征数量的增加, URLUFS 算法的性能逐步提升, 最终超越 LRLMR 算法。在生物数据集 TOX-171 上, LRLMR 和 JGLUFS 算法均采用潜在表示学习进行特征选择并取得较好的结果, 证明潜在表示学习是一种有效的方法。URLUFS 算法在生物数据集上的曲线最高点始终高于其他对比算法, 体现出潜在表示学习和广义不相关回归相结合兼顾了特征之间的不相关性。综上, URLUFS 算法可以在保存数据结构的同时发现数据的隐藏信息, 减少数据噪声的影响, 考虑特征之间的不相关性, 从而选择出更具有判别力的特征。

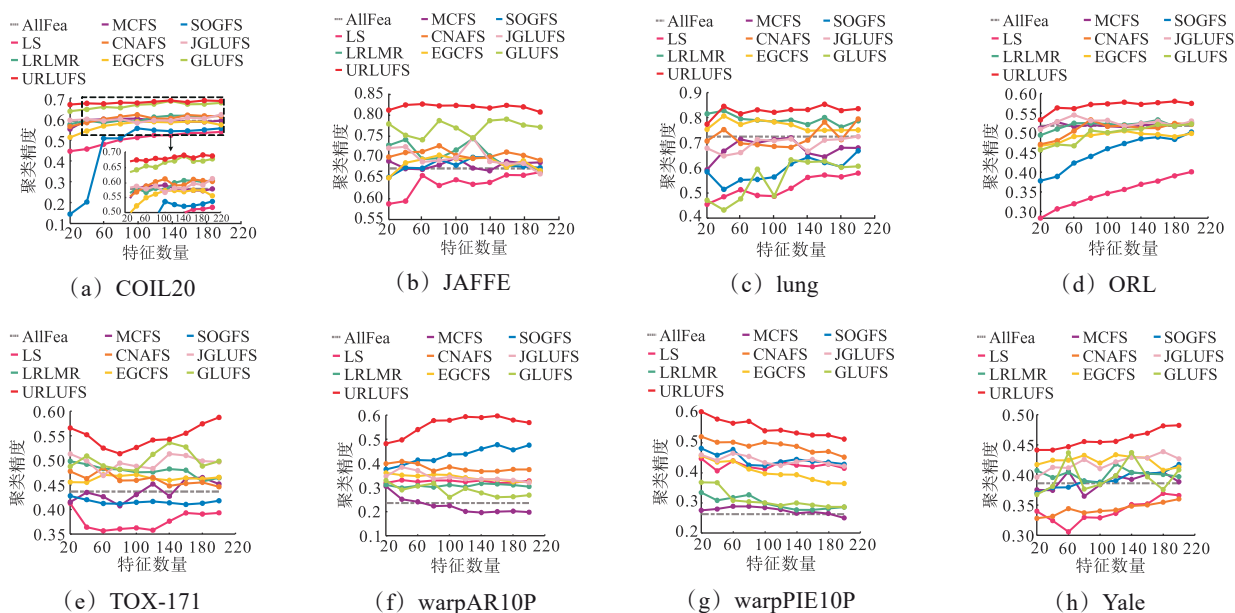


图2 在不同数据集上不同特征数量的聚类精度

Fig.2 ACC with different number of features on different datasets

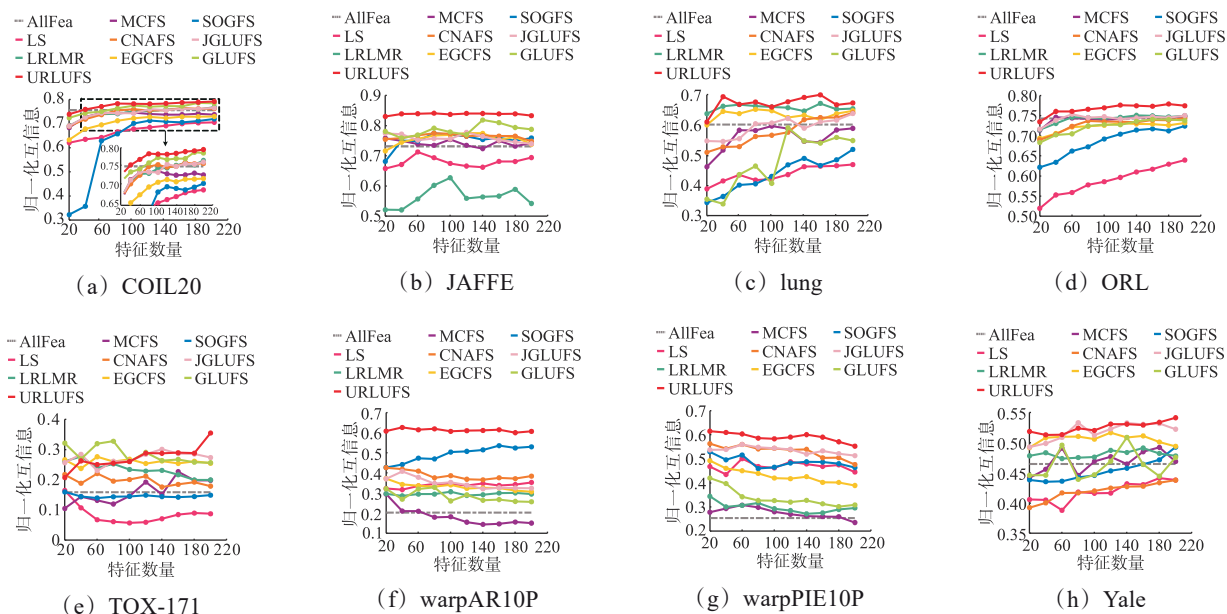


图3 在不同数据集上不同特征数量的归一化互信息

Fig.3 NMI with different number of features on different datasets

3.3 参数敏感度分析

本节实验中需要调整的参数为 α 、 β 、 γ 和 λ ,通过固定 γ 和 λ ,调节 α 和 β 来测试本文所提算法对 α 和 β 的敏感性。同理,对参数 γ 和 λ 的敏感性进行测试。并选取COIL20图像数据集和lung生物数据集进行实验,结果见图4。从图4(a)~图4(d)可以看出,当参数变化时,本文算法在COIL20数

据集上的 V_{ACC} 和 V_{NMI} 虽有小幅波动,但整体保持稳定。从图4(e)~图4(h)可以看出,由于lung数据集是样本量少、特征多的生物数据集,这些特征中可能存在一些重复特征或者非区分性特征,这增加了特征选择的难度,特征选择的结果波动较大。但是在各个参数范围内,URLUFS算法的性能基本可保持稳定。

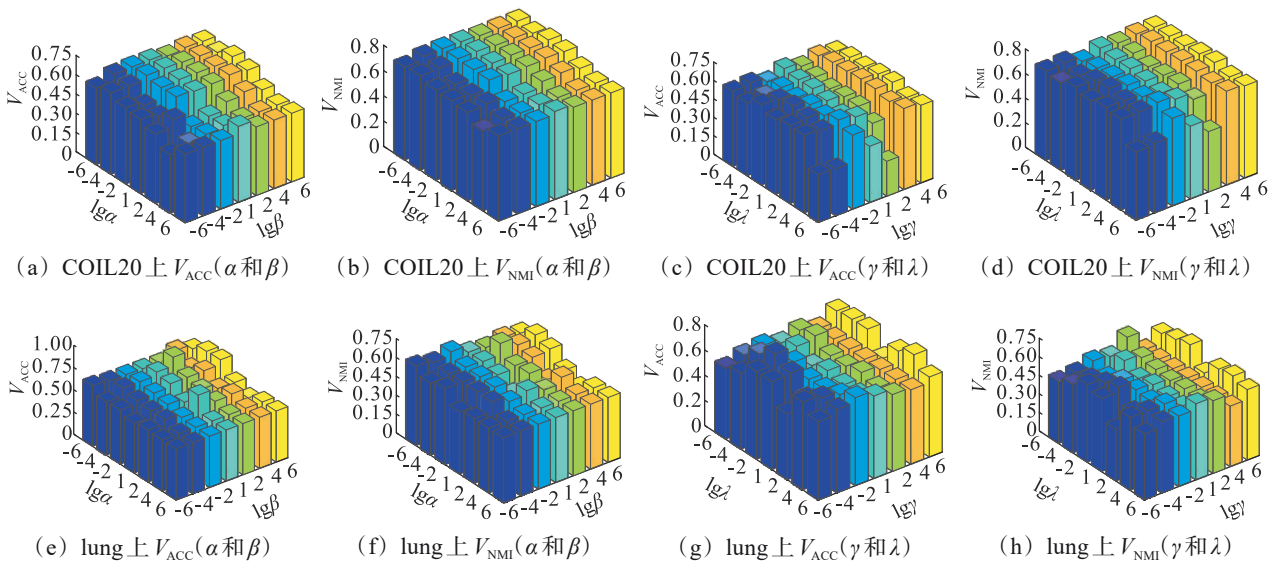
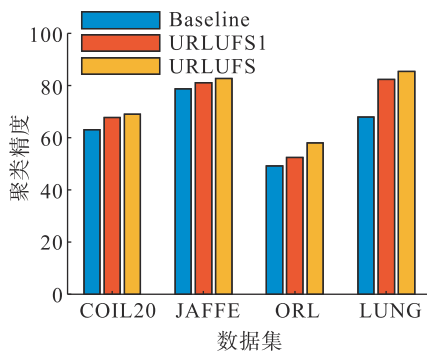


图4 URLUFS在不同数据集上的参数敏感度

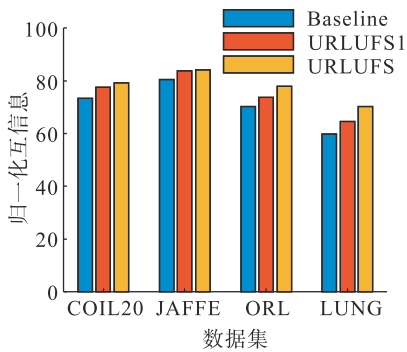
Fig.4 parameter sensitivity of URLUFS on different datasets

3.4 消融实验分析

对URLUFS算法的非负矩阵分解模块和基于潜在表示的伪标签学习模块的有效性进行分析。将广义回归模型作为基线算法，将平衡参数设为0，得到1个特殊情况URLUFS1。消融实验结果见图5。



(a) ACC



(b) NMI

图5 URLUFS算法的消融实验结果

Fig.5 ablation experiment results of URLUFS

由图5可以看出，URLUFS算法在所有数据集上的性能最优，验证了该算法的有效性。通过对比基线算法和URLUFS1算法可知，将非负矩阵分解作用在投影矩阵上，并以得到的特征选择矩阵进行自适应图学习能够提高模型的性能。对比URLUFS和URLUFS1算法可知，在广义回归模型考虑特征不相关性的同时，基于潜在表示的伪标签学习能够较好地兼顾样本之间的信息，以此来提升算法性能。因此，兼顾各个模块的URLUFS算法在无监督特征选择上有着较为出色的表现。

3.5 收敛性与运行时间分析

不同算法在COIL20数据集上的运行时间见表3，所提算法与所有对比算法相比，能够较快地完成特征选择任务。

表3 在COIL20数据集上不同算法的运行时间

Tab.3 running time of different algorithms on COIL20

dataset	
算法	运行时间/s
LS	0.171
MCFS	9.546
SOGFS	73.750
LRLMR	8.655
EGCFS	98.139
CNAFS	52.597
GLUFS	10.965
JGLUFS	9.982
URLUFS	13.693

URLUFS算法在不同数据集上的收敛曲线见图6。图6中，在大部分数据集上目标函数值均快

速减小,迭代20次时达到稳定。在数据集 TOX-171 上目标函数值收敛速度稍慢,但在30次迭代时也趋于收敛,验证了 URLUFS 算法的最大迭代次数取30次为宜。

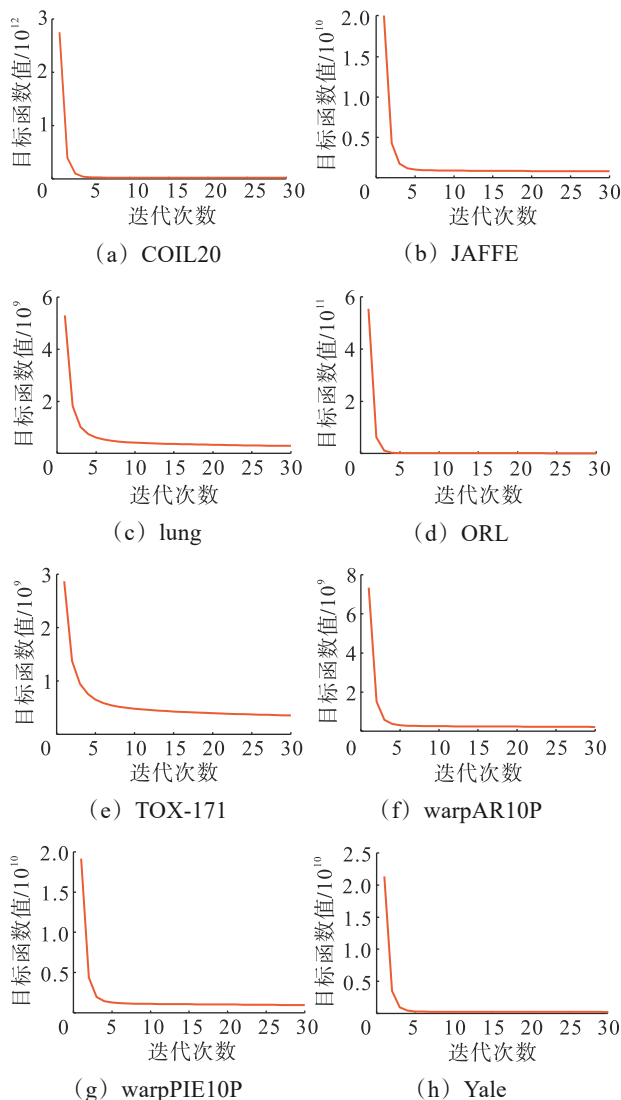


图6 URLUFS算法在不同数据集上的收敛曲线

Fig.6 convergence curves of URLUFS on different datasets

4 结论

为了获取更具判别性的特征,本文提出了一种无监督特征选择方法,将非负矩阵分解融入不相关回归模型中,使投影矩阵实现维数约简,并利用潜在表示对数据潜在信息的学习来引导伪标签矩阵。同时,为了保持数据的局部信息,结合基于特征选择子空间的自适应图来学习数据的局部流形结构。通过在真实数据集上进行大量对比实验,证明了该算法的有效性和优越性,得出如下结论。

(1) 将投影矩阵进行非负矩阵分解并进行自

适应图学习,能有效地降低投影矩阵中的噪声干扰。

(2) 基于特征选择矩阵的自适应图学习,能够挖掘数据内在结构,选择出更具判别性的特征。

(3) 数值实验结果表明,融合广义不相关矩阵和潜在表示学习的算法可实现更优的性能表现。

鉴于本文算法在处理高维数据方面的优势,未来研究将聚焦生物高维数据场景,进一步优化算法以提升稳定性和实用性。

参考文献(References):

- [1] JAHANI M S, AGHAMOLLAIE G, EFTEKHARI M, et al. Unsupervised feature selection guided by orthogonal representation of feature space[J]. Neurocomputing, 2023, 516: 61-76.
- [2] WANG R, ZHANG C Y, BIAN J T, et al. Sparse and flexible projections for unsupervised feature selection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(6): 6362-6375.
- [3] SOLORIO-FERNÁNDEZ S, CARRASCO-CHOA J A, MARTÍNEZ-TRINIDAD J F. Filter unsupervised spectral feature selection method for mixed data based on a new feature correlation measure[J]. Neurocomputing, 2024, 571: 127111.
- [4] SHANG R H, KONG J R, WANG L J, et al. Unsupervised feature selection via discrete spectral clustering and feature weights[J]. Neurocomputing, 2023, 517: 106-117.
- [5] 朱星宇,陈秀宏.联合不相关回归和非负谱分析的无监督特征选择[J].智能系统学报,2022,17(2):303-313.
ZHU Xingyu, CHEN Xiuhong. Joint uncorrelated regression and non-negative spectral analysis for unsupervised feature selection[J]. CAAI Transactions on Intelligent Systems, 2022, 17(2): 303-313.
- [6] WANG C C, WANG J, GU Z C, et al. Unsupervised feature selection by learning exponential weights[J]. Pattern Recognition, 2024, 148: 110183.
- [7] YOU M B, YUAN A H, HE D J, et al. Unsupervised feature selection via neural networks and self-expression with adaptive graph constraint[J]. Pattern Recognition, 2023, 135: 109173.
- [8] 马天,李凡卉,席润韬,等.基于生成对抗网络结合Transformer的半监督图像增强方法[J].西安科技大学学报,2023,43(6):1207-1218.
MA Tian, LI Fanhui, XI Runtao, et al. Semi-supervised image enhancement method based on generative adversarial network combined with transformer[J]. Journal of Xi'an University of Science and Technology, 2023, 43(6): 1207-1218.
- [9] ZHOU Q, WANG Q Q, GAO Q X, et al. Unsupervised discriminative feature selection via contrastive graph learning[J]. IEEE Transactions on Image Processing, 2024, 33: 972-986.
- [10] HE X F, CAI D, NIYOGI P. Laplacian score for feature selection [J]. Advances in Neural Information Processing Systems, 2005, 18: 507-514.
- [11] CAI D, ZHANG C Y, HE X F. Unsupervised feature selection for multi-cluster data[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data

- Mining. July 25-28, 2010, Washington. New York: ACM, 2010: 333-342.
- [12] NIE F P, ZHU W, LI X L. Unsupervised feature selection with structured graph optimization[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16). Washington: AAAI, 2016, 30(1): 1302-1308.
- [13] ZHANG R, ZHANG Y X, LI X L. Unsupervised feature selection via adaptive graph learning and constraint[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(3): 1355-1362.
- [14] YUAN A H, YOU M B, HE D J, et al. Convex non-negative matrix factorization with adaptive graph for unsupervised feature selection[J]. IEEE Transactions on Cybernetics, 2022, 52(6): 5522-5534.
- [15] ZHU P C, HOU X, TANG K K, et al. Unsupervised feature selection through combining graph learning and $\ell_{2,0}$ -norm constraint[J]. Information Sciences, 2023, 622: 68-82.
- [16] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[C]// Proceedings of the 14th International Conference on Neural Information Processing Systems. Cambridge, Massachusetts: MIT Press, 2000, 13:556-562.
- [17] 盛超, 宋鹏, 郑文明, 等. 基于子空间学习和伪标签回归的无监督特征选择[J]. 信号处理, 2021, 37(9): 1701-1708.
SHENG Chao, SONG Peng, ZHENG Wenming, et al. Subspace learning and virtual label regression based unsupervised feature selection[J]. Journal of Signal Processing, 2021, 37(9): 1701-1708.
- [18] ZHOU S X, SONG P, SONG Z H, et al. Soft-label guided non-negative matrix factorization for unsupervised feature selection[J]. Expert Systems with Applications, 2023, 216: 119468.
- [19] CUI G S, LI X L, DONG Y S. Subspace clustering guided convex nonnegative matrix factorization[J]. Neurocomputing, 2018, 292: 38-48.
- [20] HE Z S, XIE S L, ZDUNEK R, et al. Symmetric nonnegative matrix factorization: algorithms and applications to probabilistic clustering[J]. IEEE Transactions on Neural Networks, 2011, 22(12): 2117-2131.
- [21] 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.
- [22] XIE X J, CAO Z W, SUN F X. Joint learning of graph and latent representation for unsupervised feature selection[J]. Applied Intelligence, 2023, 53(21): 25282-25295.
- [23] LI X L, ZHANG H, ZHANG R, et al. Generalized uncorrelated regression with adaptive graph for unsupervised feature selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(5): 1587-1595.
- [24] 李乐, 章毓晋. 非负矩阵分解算法综述[J]. 电子学报, 2008, 36(4): 737-743.
LI Le, ZHANG Yujin. A survey on algorithms of non-negative matrix factorization[J]. Acta Electronica Sinica, 2008, 36(4): 737-743.
- [25] 曹浪财, 林晓昌, 苏思行. 基于矩阵分解和自适应图的无监督特征选择[J]. 系统工程与电子技术, 2021, 43(8): 2197-2208.
CAO Langcai, LIN Xiaochang, SU Sixing. Unsupervised feature selection based on matrix factorization and adaptive graph[J]. Systems Engineering and Electronics, 2021, 43(8): 2197-2208.
- [26] SHANG R H, WANG L J, SHANG F H, et al. Dual space latent representation learning for unsupervised feature selection[J]. Pattern Recognition, 2021, 114: 107873.
- [27] SABERI-MOVAHED F, ROSTAMI M, BERAHMAND K, et al. Dual Regularized Unsupervised Feature Selection Based on Matrix Factorization and Minimum Redundancy with application in gene selection[J]. Knowledge-Based Systems, 2022, 256: 109884.
- [28] PAPADIMITRIOU C H, STEIGLITZ K. Combinatorial optimization: algorithms and complexity[M]. Mineola, New York: Dover Publications, INC, 1998:30-35.
- [29] STREHL A, GHOSH J. Cluster ensembles-a knowledge reuse framework for combining multiple partitions[J]. Journal of Machine Learning Research, 2002, 3: 583-617.
- [30] SHANG R H, XU K M, JIAO L C. Subspace learning for unsupervised feature selection via adaptive structure learning and rank approximation[J]. Neurocomputing, 2020, 413: 72-84.