

# 基于双模糊学习的鲁棒无监督特征选择算法

高云龙<sup>1,2</sup>, 史曙光<sup>1</sup>, 赵志翔<sup>1</sup>, 曹超<sup>3</sup>, 潘金艳<sup>4\*</sup>

(1. 厦门大学萨本栋微米纳米科学技术研究院, 福建厦门 361102; 2. 厦门大学健康医疗大数据国家研究院, 福建厦门 361102;  
3. 自然资源部第三海洋研究所, 福建厦门 361005; 4. 集美大学海洋信息工程学院, 福建厦门 361021)

**摘要:** 由于维数灾难问题的存在, 如何在高维数据中摒弃冗余特征、保留关键信息成为一个关键问题. 无监督特征选择无需任何先验类别信息就能进行维数约减的特点引起更多研究者的关注. 尽管如此, 现有的无监督特征选择算法仍然存在两个问题需要解决: 模糊性是数据的普遍特性, 但现有大多数基于正则化回归的无监督特征选择算法忽略了数据中的模糊性, 导致特征子集不理想; 多数算法无法有效区分正常样本和噪声样本且易受噪声的影响. 针对这些问题, 本文提出了基于双模糊学习的鲁棒无监督特征选择 (Robust unsupervised Feature Selection with Double Fuzzy, DFRFS) 算法. DFRFS 算法在基于正则化回归的无监督特征选择中引入模糊隶属度, 允许数据在多个簇之间共享, 能够更好地反映数据的复杂结构和不确定性. 此外, DFRFS 算法通过鲁棒权重学习框架为样本赋予不同的权重, 从而在降低噪声样本影响的同时, 保留可靠样本的作用. 对比相关算法, 在合成数据集与真实数据集上的实验结果验证了 DFRFS 算法的有效性.

**关键词:** 无监督学习; 特征选择; 模糊学习; 稀疏学习; 维数约简

**基金项目:** 国家自然科学基金 (No.42076058); 福建省促进海洋与渔业产业高质量发展专项资金 (No.FJHYF-ZH-2023-05); 福建省自然科学基金 (No. 2020J01713, No. 2022J01061); 广东省基础与应用基础研究基金 (No.2024A1515011682)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2025)02-0604-19

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240682

## Robust Unsupervised Feature Selection with Double Fuzzy Learning

GAO Yun-long<sup>1,2</sup>, SHI Shu-guang<sup>1</sup>, ZHAO Zhi-xiang<sup>1</sup>, CAO Chao<sup>3</sup>, PAN Jin-yan<sup>4\*</sup>

(1. Pen-Tung Sah Institute of Micro-Nano Science and Technology, Xiamen University, Xiamen, Fujian 361102, China;  
2. National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, Fujian 361102, China;  
3. Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, Fujian 361005, China;  
4. School of Ocean Information Engineering, Jimei University, Xiamen, Fujian 361021, China)

**Abstract:** Due to the curse of dimensionality, effectively discarding redundant features while retaining critical information in high-dimensional data has become a key issue. Unsupervised feature selection, which performs dimensionality reduction without any prior class information, has attracted increasing attention. However, two common issues are ignored by existing unsupervised feature selection methods: Fuzziness is a common characteristic of data, but most existing unsupervised feature selection methods based on regularized regression ignore this aspect, resulting in suboptimal feature subsets; Most methods fail to effectively distinguish between normal and noisy samples and are susceptible to the noise. To tackle the mentioned issues, robust unsupervised feature selection with double fuzzy (DFRFS) learning is proposed. Specifically, DFRFS learning introduces fuzzy membership into unsupervised feature selection based on regularized regression, allowing data to be shared among multiple clusters, thereby better reflecting the complex structure and uncertainty of the data. Additionally, DFRFS learning assigns different weights to samples through the robust weight learning framework, thus suppressing the impact of noise while retaining the effect of normal samples. Experiments on toy and real-world datasets have demonstrated the effectiveness of the proposed method DFRFS learning.

**Key words:** unsupervised learning; feature selection; fuzzy learning; sparse learning; dimensionality reduction

**Foundation Item(s):** National Natural Science Foundation of China (No.42076058); Special Foundation of Fujian

Province to Promote High-quality Development of Marine and Fishery Industries (No.FJHYF-ZH-2023-05); The Natural Science Foundation of Fujian Province of China (No.2020J01713, No.2022J01061); The Guangdong Basic and Applied Basic Research Foundation (No.2024A1515011682)

## 1 引言

许多应用领域中的数据通常表示为高维向量,例如图像处理、模式识别和机器学习.高维数据增加了存储需求,其中的冗余特征和噪声信息不仅使数据处理更加复杂,而且影响后续学习算法的性能<sup>[1]</sup>.数据降维有助于解决维数灾难问题,提高相关学习算法的鲁棒性<sup>[2]</sup>.因此,研究数据降维算法具有重要的现实意义.特征选择与特征提取是目前常用的两类数据降维方法.特征选择是指从原始特征集中选择出最具代表性的特征子集;特征提取则是通过数学变换将原始特征集映射到一个新的低维空间中<sup>[3,4]</sup>.特征选择与特征提取都可以降低计算成本和数据存储需求.两者的区别在于特征选择仅保留原始特征中的子集,可解释性更好,而特征提取仍需对原始特征进行数学变换,变换后得到的特征物理意义不明确.因此,本文分析数据的特征选择.

一般来说,数据可以根据标注程度大致分为完全标记、部分标记和未标记.相应地,特征选择技术可以根据数据类标签信息的可用性分为三类,即有监督、半监督和无监督特征选择技术.有监督特征选择通常用于分类和回归任务,通过使用完全标记数据来评估特征的重要性和相关性;半监督特征选择旨在利用部分标记数据的少量注释工作来改善无监督技术的性能;无监督特征选择则利用未标记数据固有结构评估特征的重要性与相关性<sup>[4,5]</sup>.无监督特征选择缺乏标签信息来帮助特征选择,与有监督和半监督特征选择相比更具挑战性.基于此,本研究分析无监督特征选择算法.

无监督特征选择常用于聚类问题,其最终目标是根据聚类准确率等评判标准在所有特征空间中筛选一个具有代表性的最优特征子集,同时要求最终聚类结果达到一定的标准<sup>[6]</sup>.无监督特征选择算法可以进一步细分为过滤式、包裹式和嵌入式方法<sup>[7-9]</sup>.过滤式方法,例如最大方差(Max Variance, MV)法<sup>[10]</sup>,基于数据特征的方差大小评估每个特征,进而选择方差最大的重要特征.过滤式方法具有较高的计算效率,但通常忽略特征间的相关性,选择的特征冗余,从而导致学习算法性能下降.为了克服这一限制,研究人员设计了包裹式方法,以便为数据挖掘任务找到最佳的特征子集.包裹式方法逐渐增加或减少特征子集中的特征数量,并使用机器学习算法,例如聚类算法,来评估候选子集.包裹式方法比过滤式方法更准确,但需要多次试错过程,从而造成计算成本高昂,泛化能力有限.嵌入式方

法将特征选择集成到模型学习中,利用机器学习算法评估和权衡特征,得到特征权重<sup>[11]</sup>.在学习算法优化结束后,选择权重高的特征.嵌入式方法考虑特征全集而不是子集,通常比包裹式方法更高效,引起了更多研究人员的关注.

近年来,随着对数据流形结构的本质学习和聚类结构挖掘的深入研究,众多嵌入式方法被提出.例如:文献[12]基于谱分析与回归模型提出了多簇特征选择(Multi-Cluster Feature Selection, MCFS)算法,能够发现和保留多簇数据的相似性与流形结构;文献[13]基于判别分析和 $\ell_{2,1}$ 范数正则化提出了无监督判别特征选择(Unsupervised Discriminative Feature Selection, UDFS)算法,选择最具判别性的特征并保持数据流形结构.此外,考虑到非负且不相关的嵌入结果具有清晰的聚类结构指示能力,有学者通过对低维嵌入添加正交非负约束来提高聚类结果的可解释性.例如:文献[14]提出非负判别特征选择(Nonnegative Discriminative Feature Selection, NDFS)算法,将谱聚类与特征选择相结合,构建线性回归模型来引导特征选择;文献[15]提出了鲁棒无监督特征选择(Robust Unsupervised Feature Selection, RUFFS)算法,通过鲁棒非负矩阵分解与谱聚类的方法挖掘数据中的聚类信息,并采用线性回归与 $\ell_{2,1}$ 范数正则化进行特征选择;文献[16]提出了嵌入式无监督特征选择(Embedded Unsupervised Feature Selection, EUFS)算法,将稀疏学习嵌入到谱聚类中,确保所选择的特征能够有效保留数据中的结构信息;文献[17]提出了同步正交基聚类特征选择(Simultaneous Orthogonal basis Clustering Feature Selection, SOCFs)算法,将正交基聚类和特征选择结合起来,利用正交基构建目标矩阵来捕获投影数据的潜在簇中心,进而选择具有更强识别能力的特征.通过利用数据点的局部结构信息来引导特征选择,模型能够更加灵活地适应数据的局部特征,从而在描述数据流形结构时提高拟合能力和表征能力.基于此,文献[18]提出了结构化最优图特征选择(Structured Optimal Graph Feature Selection, SOGFS)算法,同时进行局部结构学习和特征选择.文献[19]提出了用于无监督特征选择的自适应图非相关回归(Uncorrelated Regression with Adaptive graph for unsupervised Feature Selection, URAFS)算法,将广义不相关约束、自适应局部结构学习和最大熵原理结合到统一的框架中.文献[20]引入了一种指数加权机制,以平衡特征权重,确保更多的信息特征能够有效参与局部结构

学习. 受局部结构学习和SOCFS算法的启发, 文献[21]采用自适应图学习和正交基聚类相结合的策略进行特征选择. 基于矩阵分解与自适应图学习技术, 文献[22]在SOCFS算法的基础上引入局部结构保持项以选择具有局部保持能力的特征. 通过在图学习中引入特殊图约束, 文献[23]提出了基于嵌入式图学习和约束的无监督特征选择 (unsupervised Feature Selection via Embedded Graph learning and Constraint, EGCFs) 算法, 将最大化类间散度矩阵与自适应图学习结合在一起进行图学习, 能够在保留数据本质结构的同时提取数据的判别信息. 为解决从原始数据构建的相似度矩阵质量不佳的问题, 文献[24]将相似度矩阵构造与特征选择整合到一个统一框架中. 文献[25]将对比学习融入自适应图学习的过程以获取更多图结构判别信息, 进一步提升对数据聚类特性的理解. 上述特征选择方法大多采用  $\ell_{2,1}$  范数, 而在稀疏学习领域,  $\ell_{2,p}$  范数因其更好的灵活性, 在选择更优稀疏模式方面表现出更好的性能. 文献[26]提出的稀疏主成分分析特征选择 (Sparse Principal Component Analysis for Feature Selection, SPCAFs) 算法将 PCA 形式化为重构误差最小化问题, 并采用  $\ell_{2,p}$  范数正则化项使得投影矩阵稀疏化, 学得的行稀疏和正交投影矩阵用于选择具有区分性的特征. 文献[27]提出的快速稀疏判别  $K$  均值 (Fast Sparse Discriminative  $K$ -means, FSDK) 算法将  $\ell_{2,p}$  范数与改进的  $K$ -means 算法高效结合, 能够快速在大规模数据上进行特征选择. 此外, 文献[28]提出的基于方差-协方差子空间距离的无监督特征选择 (Variance-Covariance Subspace Distance-based unsupervised Feature Selection, VCSDFS) 算法从子空间距离的角度出发, 有效地整合了特征的方差和协方差信息, 进而排除冗余特征并实现维度降低. 随着多视图数据的出现, 多视图无监督特征选择<sup>[29-31]</sup>相比于单视图无监督特征选择具有更强的表达能力和更高的鲁棒性, 这使得多视图无监督特征选择方法在高维数据处理中受到广泛关注.

模糊性在自然界和人类社会中广泛存在<sup>[32]</sup>. 然而, 目前大多数无监督特征选择算法仅考虑样本完全隶属于某一类的情况, 未能充分考虑数据与簇之间的模糊关系. 在过去几十年里, 研究人员提出了一些基于模糊的特征选择算法. 例如: 文献[33]提出了一种将神经模糊方法应用于无监督特征选择的算法, 引入结合加权距离的隶属函数计算变换空间中的隶属度, 并设计了相应的分层网络生成特征重要性排序; 文献[34]提出了一种基于模糊特征评估指数的混合无监督特征选择算法, 通过计算特征与聚类相关性的排名指数获取候选特征子集; 文献[35]使用模糊集从候选子集中删除特征, 显著降低数据维度的同时保留数据语义; 文献[36]

提出了无监督判别投影特征选择 (Unsupervised Discriminative Projection for Feature Selection, UDPFS) 算法, 将  $\ell_{2,1}$  范数正则化应用于投影矩阵, 并同时进行了模糊学习和稀疏学习来选择判别特征; 文献[37]通过模糊算法获取数据的伪标签, 并基于数据样本与聚类中心间的模糊距离进行特征选择. 上述算法将模糊学习与特征选择相结合, 凸显了模糊学习在特征选择中的优势.

现实世界的数据总是包含大量的噪声样本, 噪声样本会对每个数据簇的中心产生显著影响, 导致簇中心发生偏移, 从而降低模糊算法提取数据簇本质结构的能力. 然而, 上述基于模糊的特征选择算法未能区分正常样本与噪声样本. 考虑到上述分析, 本文提出了一种新颖的无监督特征选择算法. 该算法在基于正则化回归的无监督特征选择算法中引入模糊隶属度, 并通过鲁棒权重学习框架为样本赋予稀疏权重来解决现有无监督特征选择算法中存在的问题. 本文的主要贡献总结如下:

- (1) 在基于正则化回归的无监督特征选择算法中引入模糊隶属度, 允许数据在多个簇之间共享, 能够更好地反映数据的复杂结构和不确定性;
- (2) 基于样本的局部散布度, 引入模糊权重以实现数据簇局部结构的弹性收缩, 通过控制权重向量的稀疏性过滤并剔除噪声样本, 以实现整体的鲁棒性;
- (3) 提出迭代优化算法自适应更新模糊隶属度与样本权重, 在合成数据集与真实数据集上的实验结果表明了所提算法的优越性.

## 2 相关工作

在介绍相关工作之前, 首先对本文中所用到的符号进行介绍. 对于任意  $m$  维向量  $\mathbf{b} = [b_1, b_2, \dots, b_m] \in \mathbb{R}^m$ , 该向量的  $\ell_1$  范数和  $\ell_2$  范数分别表示为  $\|\mathbf{b}\|_1$  与  $\|\mathbf{b}\|_2$ . 对于给定矩阵  $\mathbf{W}$ ,  $\mathbf{w}^i$  和  $\mathbf{w}_j$  分别表示  $\mathbf{W}$  的第  $i$  行向量和第  $j$  列向量,  $w_{ij}$  表示  $\mathbf{W}$  的第  $i$  行的第  $j$  个元素, 该矩阵的  $\ell_{2,1}$  范数表示为  $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^d \|\mathbf{w}^i\|_2$ .  $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_m) \in \mathbb{R}^{m \times m}$  是对角矩阵, 其对角线上第  $i$  个元素为  $d_i$ .  $\mathbf{E} = \text{blkdiag}(U_1, U_2, \dots, U_c)$  是分块对角矩阵, 其对角块上第  $i$  个矩阵为  $U_i$ .  $\mathbf{b}^\dagger = [b_1^\dagger, b_2^\dagger, \dots, b_m^\dagger]$  表示对向量中的每个元素开方.  $\tilde{\mathbf{X}} = [\underbrace{\mathbf{X}, \mathbf{X}, \dots, \mathbf{X}}_c]$  表示将  $c$  个矩阵  $\mathbf{X}$  横向连接所得的关于  $\mathbf{X}$  的变换矩阵  $\tilde{\mathbf{X}}$ . 向量  $\mathbf{0}$  中的元素全为 0, 向量  $\mathbf{1}$  中的元素全为 1.  $\mathbf{A} \geq 0$  表示矩阵  $\mathbf{A}$  中所有元素都大于等于 0.  $(\bullet)_+ = \max(\bullet, 0)$  表示取  $\bullet$  与 0 之间的最大值. 为便于理解, 表 1 列出了本文常用的一些符号.

表 1 文中常用符号

符号	简介	符号	简介
$n$	样本总数	$d$	样本特征数
$c$	类别数	$m$	投影后子空间维度
$k$	参考样本个数	$r$	参考样本率
$\alpha$	正则化参数	$\beta$	正则化参数
$X \in \mathbb{R}^{d \times n}$	数据矩阵	$W \in \mathbb{R}^{d \times m}$	特征选择矩阵
$V \in \mathbb{R}^{m \times c}$	聚类中心矩阵	$U \in \mathbb{R}^{c \times n}$	模糊隶属度矩阵
$p \in \mathbb{R}^n$	样本权重向量	$D \in \mathbb{R}^{d \times d}$	对角矩阵
$\tilde{X} \in \mathbb{R}^{d \times (n \times c)}$	关于 $X$ 的变换矩阵	$\tilde{U} \in \mathbb{R}^{(n \times c) \times (n \times c)}$	关于 $U$ 的变换矩阵
$E \in \mathbb{R}^{c \times (n \times c)}$	关于 $U$ 的变换矩阵	$\tilde{P} \in \mathbb{R}^{(n \times c) \times (n \times c)}$	关于 $p$ 的变换矩阵

### 2.1 无监督特征选择算法

给定一个包含  $n$  个样本的数据矩阵  $X$ , 假设其目标矩阵为  $T \in \mathbb{R}^{m \times n}$ . 基于回归的特征选择问题一般可以解释为

$$\min_W \mathcal{L}(W^T X - T) + \beta \mathcal{R}(W) \quad (1)$$

其中,  $W \in \mathbb{R}^{d \times m}$  为特征选择矩阵, 也称投影矩阵;  $\mathcal{L}(W^T X - T)$  为损失函数, 衡量所选特征与目标矩阵的差异性;  $\mathcal{R}(W)$  是正则化项, 通常用来保证特征选择矩阵  $W$  的行稀疏性. 这种基于正则化回归的方法已被证明对于有监督和无监督方式的特征选择都是有效的<sup>[15]</sup>. 对于有监督特征选择算法, 类标签矩阵被识别为目标矩阵  $T$ . 而对于无监督特征选择算法, 由于缺乏类标签信息, 目标矩阵  $T$  不确定. 解决这一问题的常用方法是在特征选择过程中学习聚类中心来表示目标矩阵<sup>[13-15]</sup>. 一般而言, 式(1)的第一项基于数据的聚类信息采用线性回归模型来引导特征选择, 而第二项则是对矩阵  $W$  添加  $\ell_{2,1}$  正则化来确保其行稀疏的性质, 即转化为求解

$$\min_W \|W^T X - T\|_F^2 + \beta \|W\|_{2,1} \quad (2)$$

### 2.2 NDFS

NDFS 将非负谱分解与特征选择相结合, 利用非负谱分解来获取数据的目标矩阵  $F$ , 构建线性回归模型学习特征选择矩阵  $W$ , 并对投影矩阵添加  $\ell_{2,1}$  正则化来确保  $W$  具有特征选择能力. NDFS 算法数学表达式为

$$\begin{cases} \min_{F, W} \text{tr}(F^T L F) + \alpha \left( \|X^T W - F\|_F^2 + \beta \|W\|_{2,1} \right) \\ \text{s.t. } F^T F = I, F \geq 0 \end{cases} \quad (3)$$

式(3)中, NDFS 利用第一项保留数据的局部本质结构来获取数据的聚类信息, 而第二项则是式(2)中所描述的线性回归模型. 对于矩阵  $W$  添加的  $\ell_{2,1}$  正则化保证了其行稀疏的性质, 此时, 只有与目标矩阵  $F$  相关性较高的特征所对应的  $W$  的行向量具有较大的值, 而相关性较低的特征所对应的行向量则趋近于 0. 因此, 可以根据  $\|w\|_2$  的大小进行特征选择. 为了确保  $F$  具有伪标签

属性, NDFS 对其添加了正交非负约束, 将保留数据的本质结构和探索聚类结构结合在一起, 使模型更为简洁.

### 2.3 SOCFS

为对无标签数据进行有效的特征选择, SOCFS 将目标矩阵分解为两个子矩阵, 即编码矩阵和基矩阵, 并对两个子矩阵添加正交约束以达到正交基聚类的效果. 正交基聚类的目标函数和约束条件为

$$\begin{cases} \min_{W, B, E, F} \|W^T X - B E^T\|_F^2 + \beta \|W\|_{2,1} + \gamma \|F - E\|_F^2 \\ \text{s.t. } B^T B = I, E^T E = I, F \geq 0 \end{cases} \quad (4)$$

其中,  $B$  为基矩阵;  $E$  为编码矩阵. 正交约束  $B^T B = I$  保证  $B$  中的不同基向量相互独立. 正交约束  $E^T E = I$  以及非负约束  $F \geq 0$  保证  $E$  在理想状态下每一行都有唯一的非零元素, 因此,  $E$  可以独立地选择  $B$  中的基向量. 该算法通过正交基聚类来捕获投影后的数据点潜在的类簇中心, 继而指导投影矩阵  $W$  选择判别性的特征, 最终实现无监督地对数据集进行特征选择.

## 3 模型与优化

### 3.1 目标函数

模糊性是数据的普遍特性<sup>[32]</sup>, 在进行数据聚类分析时, 如果不考虑数据的模糊性, 可能会导致对数据真实类别特征的误解和偏差. 然而, 目前基于正则化回归的无监督特征选择算法往往忽视了数据的模糊性, 仅考虑样本完全隶属于某一类的情况, 未能充分考虑数据与类之间的模糊关系. 为了解决这一问题, 本文在基于回归的无监督特征选择算法的基础上, 引入了模糊隶属度

$$\begin{cases} \min_{W, U, V} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|W^T x_j - v_i\|_2^2 \\ \text{s.t. } u_j^T \mathbf{1} = 1, 0 \leq u_{ij} \leq 1, V^T V = I \end{cases} \quad (5)$$

其中,  $c$  为类别数;  $n$  为样本总数;  $x_j$  和  $v_i$  分别为第  $j$  个样本和第  $i$  个聚类中心;  $u_{ij}$  为第  $j$  个样本对第  $i$  类数据簇的隶属度;  $m$  为模糊因子, 通常情况下设置  $m = 2$ <sup>[38]</sup>. 正交约束  $V^T V = I$  使得  $V$  的列向量互相独立, 具有较好的类间可分性<sup>[17]</sup>. 模糊隶属度矩阵  $U$  是模糊聚类中的关键元素, 其能够在多个聚类中实现新颖的数据表示. 隶属度矩阵的列向量表示每个样本对不同簇的隶属度. 通过观察式(5)发现, 不同的样本在不同聚类中的  $\sum_{i=1}^c u_{ij}^m$  各不相同. 当且仅当  $u_{1j} = u_{2j} = \dots = u_{cj} = \frac{1}{c}$  且  $m = 2$  时, 基于 Jensen 不等式可得到式(6)的极小值

$$\begin{cases} \min \sum_{i=1}^c u_{ij}^2 \\ \text{s.t. } \mathbf{u}_j^T \mathbf{1} = 1, 0 \leq u_{ij} \leq 1 \end{cases} \quad (6)$$

这表明,通过在式(5)中引入模糊性,降低了不同数据簇交叠区域样本点的  $\sum_{i=1}^c u_{ij}^2$ ,从而减少了这些样本点所带来的损失.因此,这种灵活性使模糊聚类能够有效处理聚类结构不明确或重叠的场景.

考虑到模糊隶属度的拖尾与翘尾属性<sup>[39]</sup>,上述模糊加权只能降低但无法消除与整个数据簇偏移较大的样本点影响.即现有的基于模糊聚类的算法无法过滤和消除这些样本点的影响,这可能导致相应模型的功能退化.为了解决这个问题,本文受RWL-AN<sup>[40]</sup>启发,提出了基于双模糊学习的鲁棒无监督特征学习模型:

$$\begin{cases} \min_{\mathbf{W}, \mathbf{U}, \mathbf{V}, \mathbf{p}} \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2 + \alpha p_j^2 \\ \text{s.t. } \mathbf{p}^T \mathbf{1} = 1, \mathbf{u}_j^T \mathbf{1} = 1, 0 \leq p_j, u_{ij} \leq 1, \mathbf{V}^T \mathbf{V} = \mathbf{I} \end{cases} \quad (7)$$

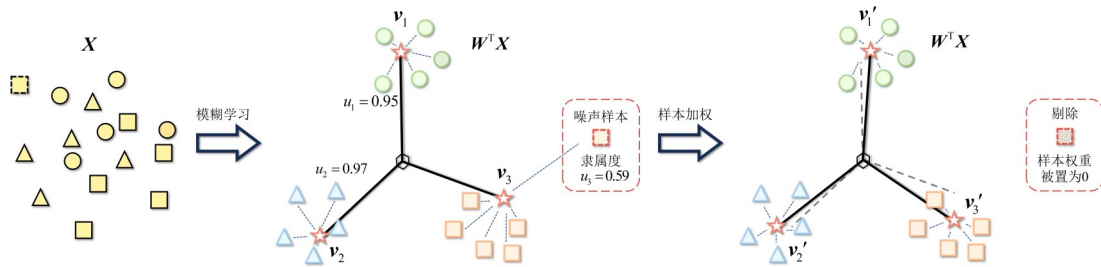
其中,  $\sum_{i=1}^c u_{ij}^2 \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2$  为在数据描述模型确定的情况下,即模糊隶属度  $\mathbf{U}$ 、聚类中心  $\mathbf{V}$  和投影空间  $\mathbf{W}$  确定时,样本点  $\mathbf{x}_j$  的局部散布度;  $\mathbf{p}$  为样本散布度的权重向量,其具有稀疏性与鲁棒性,且能够自适应更新;  $p_i$  为第  $i$  个样本局部散布度的权重;  $\alpha$  为正则化参数,可由预先确

定的参考样本数目  $k$  确定.从式(7)中不难看出,  $p_i$  非负且与  $\sum_{i=1}^c u_{ij}^2 \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2$  的大小成反比,散布度越大的样本被赋予趋近 0 或为 0 的权重.此外,参考样本数目  $k$  能够通过正则化参数  $\alpha$  对权重向量  $\mathbf{p}$  的稀疏程度进行调节.这意味着具有最小散布度的  $k$  个参考样本被用于模型训练,而散布度较大的被当作噪声样本剔除以实现数据簇局部结构的弹性收缩,从而确保全局的鲁棒性.

近年来,基于稀疏学习的特征选择算法因其优异的性能和可解释性受到了广泛关注<sup>[41]</sup>.作为稀疏学习方法之一,  $\ell_{2,1}$  范数具有显著的行稀疏性,能够有效剔除冗余与噪声特征<sup>[17]</sup>.因此,本文将  $\ell_{2,1}$  范数应用到特征选择矩阵  $\mathbf{W}$  上,得到最终的基于双模糊学习的鲁棒无监督特征选择 (Robust unsupervised Feature Selection with Double Fuzzy, DFRFS) 模型为

$$\begin{cases} \min_{\mathbf{W}, \mathbf{U}, \mathbf{V}, \mathbf{p}} \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2 + \alpha p_j^2 + \beta \|\mathbf{W}\|_{2,1} \\ \text{s.t. } \mathbf{p}^T \mathbf{1} = 1, \mathbf{u}_j^T \mathbf{1} = 1, 0 \leq p_j, u_{ij} \leq 1, \mathbf{V}^T \mathbf{V} = \mathbf{I} \end{cases} \quad (8)$$

其中,  $\ell_{2,1}$  范数正则项使得  $\mathbf{W}$  中无关特征对应的行向量元素趋近于 0.通过对  $\|\mathbf{w}^i\|_2$  进行排序可以选择最具判别信息的特征.本文将该算法命名为 DFRFS 算法,图 1 为该算法的示意图.



注:DFRFS 首先学习模糊隶属度以发掘数据的本质结构,然后通过样本加权消除噪声样本的影响.

图 1 DFRFS 算法的示意图

### 3.2 模型求解

对于式(8),需要更新的变量包括  $\mathbf{U}$ 、 $\mathbf{p}$ 、 $\mathbf{W}$  和  $\mathbf{V}$ .因此,本文引入交替迭代更新算法来解决该问题.本文提出算法的迭代更新规则满足凸优化中收敛的一般条件<sup>[42]</sup>.

#### 3.2.1 更新 $\mathbf{U}$

固定其他参数,式(8)关于  $\mathbf{U}$  的子问题为

$$\begin{cases} \min_{\mathbf{U}} \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2 \\ \text{s.t. } \mathbf{u}_j^T \mathbf{1} = 1, 0 \leq u_{ij} \leq 1 \end{cases} \quad (9)$$

利用 Lagrange 乘法<sup>[43]</sup>,可以得到模糊隶属度的最优解为

$$u_{ij} = \frac{1}{\sum_{q=1}^c \frac{\|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2}{\|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_q\|_2^2}} \quad (10)$$

#### 3.2.2 更新 $\mathbf{p}$

固定其他参数,令  $g(\mathbf{x}_j) = \sum_{i=1}^c u_{ij}^m \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2$ , 式(8)

关于  $\mathbf{p}$  的子问题为

$$\begin{cases} \min_{\mathbf{p}} \sum_{j=1}^n p_j g(\mathbf{x}_j) + \alpha p_j^2 \\ \text{s.t. } \mathbf{p} \geq 0, \mathbf{p}^T \mathbf{1} = 1 \end{cases} \quad (11)$$

用  $g_j$  表示  $g(\mathbf{x}_j)$ ,  $\mathbf{g} = [g_1, g_2, \dots, g_n]$ , 式(11)可以改

写为

$$\begin{cases} \min_p \frac{1}{2} \left\| p + \frac{g}{2\alpha} \right\|_2^2 \\ \text{s.t. } p \geq 0, p^T \mathbf{1} = 1 \end{cases} \quad (12)$$

利用拉格朗日乘子法,建立拉格朗日函数:

$$\mathcal{L}(p, \lambda, \sigma) = \frac{1}{2} \left\| p + \frac{g}{2\alpha} \right\|_2^2 - \lambda(p^T \mathbf{1} - 1) - \sigma^T p \quad (13)$$

其中,  $\lambda$  和  $\sigma$  为拉格朗日乘子. 式(13)满足 KKT 条件 (Karush - Kuhn - Tucker conditions) 的最优解为

$$p_i = \left( \lambda - \frac{g_i}{2\alpha} \right)_+ \quad (14)$$

为确定  $\lambda$  与  $\alpha$ , 假设  $g_1 \leq g_2 \leq \dots \leq g_n$ . 因此, 根据  $p_i$  与  $g_i$  的负相关关系可得  $p_1 \geq p_2 \geq \dots \geq p_n$ . 当只考虑具有最小重构误差的  $k$  个参考样本时, 可得

$$\begin{cases} p_k > 0 \Rightarrow \lambda - \frac{g_k}{2\alpha} > 0 \\ p_{k+1} = 0 \Rightarrow \lambda - \frac{g_{k+1}}{2\alpha} \leq 0 \end{cases} \quad (15)$$

结合式(15)与  $p^T \mathbf{1} = 1$ , 可得

$$\sum_{i=1}^k \left( \lambda - \frac{g_i}{2\alpha} \right) = 1 \Rightarrow \lambda = \frac{1}{k} + \frac{1}{2\alpha k} \sum_{i=1}^k g_i \quad (16)$$

根据式(15)与式(16), 可得

$$\begin{cases} \frac{1}{k} > \frac{g_k}{2\alpha} - \frac{1}{2\alpha k} \sum_{i=1}^k g_i \\ \frac{1}{k} \leq \frac{g_{k+1}}{2\alpha} - \frac{1}{2\alpha k} \sum_{i=1}^k g_i \end{cases} \Rightarrow \frac{k}{2} g_k - \frac{1}{2} \sum_{i=1}^k g_i < \alpha \leq \frac{k}{2} g_{k+1} - \frac{1}{2} \sum_{i=1}^k g_i \quad (17)$$

为了获得  $k$  个参考样本的具体权重, 取

$$\alpha = \frac{k}{2} g_{k+1} - \frac{1}{2} \sum_{j=1}^k g_j \quad (18)$$

结合式(16)与式(18), 式(14)可以改写为

$$p_i = \left( \frac{g_{k+1} - g_i}{kg_{k+1} - \sum_{j=1}^k g_j} \right)_+ \quad (19)$$

在缺乏先验知识的情况下, 参考样本数目  $k$  的设置存在一定难度. 同时, 对于样本总数不同的数据集,  $k$  的取值会发生改变. 为了进一步简化对参数的设置, 参考文献[44]的做法, 令

$$k = \lceil n \times r \rceil \quad (20)$$

其中,  $r$  为参考样本率且  $r \in (0, 1]$ ;  $\lceil \cdot \rceil$  为对  $\cdot$  向上取整. 根据式(20), 模型对参考样本数目  $k$  的选择可以转换为对参考样本率  $r$  的设定. 经过此变换, 对于任意大小的数据集, 参数的取值范围均限定在  $(0, 1]$ , 从而降低了参

数设定的难度.

### 3.2.3 更新 $W$

固定其他参数, 式(8)关于  $W$  的子问题为

$$\min_W \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \left\| W^T x_j - v_i \right\|_2^2 + \beta \|W\|_{2,1} \quad (21)$$

令  $\tilde{X} = [\underbrace{X, X, \dots, X}_c]$ ,  $E = \text{blkdiag}(u^1, u^2, \dots, u^c)$ ,  $\tilde{U} = \text{diag}(u_{11}, u_{12}, \dots, u_{1n}, u_{21}, u_{22}, \dots, u_{2n}, \dots, u_{c1}, u_{c2}, \dots, u_{cn})$ ,  $\tilde{P} = \text{diag}(\underbrace{p, p, \dots, p}_c)$ . 经过数学变换, 式(21)可以改写为

$$\min_W \left\| W^T \tilde{X} \tilde{U} \tilde{P} - V E \tilde{P} \right\|_F^2 + \beta \|W\|_{2,1} \quad (22)$$

对式(22)中的  $W$  求导并将其设为 0, 可得

$$\begin{aligned} 2\tilde{X} \tilde{U} \tilde{P} \tilde{P}^T \tilde{U}^T \tilde{X}^T W - 2\tilde{X} \tilde{U} \tilde{P} \tilde{P}^T E^T V^T + 2\beta D W &= 0 \\ \Rightarrow W &= (\tilde{X} \tilde{U} \tilde{P} \tilde{P}^T \tilde{U}^T \tilde{X}^T + \beta D)^{-1} \tilde{X} \tilde{U} \tilde{P} \tilde{P}^T E^T V^T \end{aligned} \quad (23)$$

其中,  $D$  为对角元素为  $\frac{1}{2\|w^i\|_2}$  的对角矩阵.  $\|W\|_{2,1}$  的导数被计算为  $2DW^{[17]}$ .

### 3.2.4 更新 $V$

固定其他参数, 式(8)关于  $V$  的子问题为

$$\begin{cases} \min_V \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \left\| W^T x_j - v_i \right\|_2^2 \\ \text{s.t. } V^T V = I \end{cases} \quad (24)$$

与更新  $W$  时的矩阵变换相同, 式(24)可以改写为

$$\begin{cases} \min_V \left\| W^T \tilde{X} \tilde{U} \tilde{P} - V E \tilde{P} \right\|_F^2 \\ \text{s.t. } V^T V = I \end{cases} \Rightarrow \begin{cases} \min_V \left\| \tilde{P}^T E^T V^T - \tilde{P}^T \tilde{U}^T \tilde{X}^T W \right\|_F^2 \\ \text{s.t. } V^T V = I \end{cases} \quad (25)$$

参考文献[45]的工作, 有以下定理.

**定理 1** 假设有两个矩阵  $M \in \mathbb{R}^{n \times m}$  和  $Q \in \mathbb{R}^{n \times d}$ . 对于优化问题:

$$\begin{cases} \min_{\tilde{T}} \left\| M \tilde{T} - Q \right\|_F^2 \\ \text{s.t. } \tilde{T} \tilde{T}^T = I \end{cases} \quad (26)$$

有解析解:

$$\tilde{T} = L I_{m,d} R^T \quad (27)$$

其中,  $L \in \mathbb{R}^{m \times m}$  和  $R \in \mathbb{R}^{d \times d}$  分别为对矩阵  $M^T Q$  进行奇异值分解 (Singular Value Decomposition, SVD) 分解得来的左、右特征向量.

分别令  $M = \tilde{P}^T E^T$ ,  $Q = \tilde{P}^T \tilde{U}^T \tilde{X}^T W$ ,  $\tilde{T} = V^T$ . 根据定理 1, 式(25)的最优解为

$$V = R_V I_{m,c} L_V^T \quad (28)$$

其中,  $L_V \in \mathbb{R}^{c \times c}$ ,  $R_V \in \mathbb{R}^{m \times m}$  分别为对  $E \tilde{P} \tilde{P}^T \tilde{U}^T \tilde{X}^T W$  进行 SVD 分解得来的左、右特征向量.

算法 1 给出了求解式(8)的具体算法.

---

**算法 1** DFRFS
 

---

输入:数据矩阵  $X$ ,类别数  $c$ ,超参数  $r, \beta$

输出:将  $\|\mathbf{w}^i\|_2$  按降序排列选择特征

初始化: $t=0, \mathbf{D}_t=\mathbf{I}, \mathbf{W}_t, \mathbf{V}_t$

REPEAT

  根据式(10)更新  $U_t$

  根据式(19)更新  $\mathbf{p}_t$

  根据式(23)更新  $\mathbf{W}_t$

  根据式(28)更新  $\mathbf{V}_t$

$t=t+1$

UNTIL  $\|\Delta J(\mathbf{W}_t, \mathbf{U}_t, \mathbf{V}_t, \mathbf{p}_t)\| \leq \varepsilon$  或  $t \geq T$

---

本文所提算法的主要计算负担在于迭代更新过程中,根据算法 1,共需更新 4 个矩阵.在每次迭代中,更新  $U$  的计算复杂度为  $O(dnc + nc^2)$ ,更新  $\mathbf{p}$  的计算复杂度为  $O(dnc + nc^2 + n^2r)$ ,更新  $\mathbf{W}$  的计算复杂度为  $O(d^3 + d^2nc + dn^2c^2)$ ,更新  $\mathbf{V}$  的计算复杂度为  $O(dnc^2 + n^2c^3)$ .一般地,  $d \gg c, n \gg c$ ,且样本参考率  $r \in (0, 1]$ ,因此,本文所提算法一次迭代的计算复杂度为  $O(d^3 + d^2nc + dn^2c^2)$ .

### 3.3 收敛性证明

通过执行算法 1,式(8)中的目标函数将会单调递减直至收敛.为了证明算法的收敛性,需要引入以下定理.

**定理 2** 对于任意两个非零向量  $\mathbf{u}$  和  $\mathbf{v}$ ,不等式(29)成立.文献[41]给出了定理 2 的详细证明.

$$\sqrt{\mathbf{u}} - \frac{\mathbf{u}}{2\sqrt{\mathbf{u}}} \leq \sqrt{\mathbf{v}} - \frac{\mathbf{v}}{2\sqrt{\mathbf{v}}} \quad (29)$$

为表述方便,定义式(8)中的目标函数为

$$J(\mathbf{W}, \mathbf{U}, \mathbf{V}, \mathbf{p}) = \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \|\mathbf{W}^T \mathbf{x}_j - \mathbf{v}_i\|_2^2 + \alpha p_j^2 + \beta \|\mathbf{W}\|_{2,1} \quad (30)$$

本文将通过如下定理证明算法 1 的收敛性.

**定理 3** 在每次迭代过程中,算法 1 将会单调减小式(30)的值,直至收敛.

**证明** 算法 1 通过求解式(22)来更新  $\mathbf{W}$ ,相当于求解式(31):

$$\min_{\mathbf{W}} \left\| \mathbf{W}^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \text{tr}(\mathbf{W}^T \mathbf{D}_t \mathbf{W}) \quad (31)$$

因此,不等式(32)成立.

$$\begin{aligned} & \left\| \mathbf{W}_{t+1}^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \text{tr}(\mathbf{W}_{t+1}^T \mathbf{D}_t \mathbf{W}_{t+1}) \leq \\ & \left\| \mathbf{W}_t^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \text{tr}(\mathbf{W}_t^T \mathbf{D}_t \mathbf{W}_{t+1}) \end{aligned} \quad (32)$$

式(32)等价于:

$$\left\| \mathbf{W}_{t+1}^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \sum_i \frac{\|\mathbf{w}_{t+1}^i\|_2^2}{2\|\mathbf{w}_t^i\|_2} \quad (33)$$

$$\leq \left\| \mathbf{W}_t^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \sum_i \frac{\|\mathbf{w}_t^i\|_2^2}{2\|\mathbf{w}_t^i\|_2}$$

根据定理 2,  $\sum_i \|\mathbf{w}_{t+1}^i\|_2 - \sum_i \frac{\|\mathbf{w}_{t+1}^i\|_2^2}{2\|\mathbf{w}_t^i\|_2} \leq \sum_i \|\mathbf{w}_t^i\|_2 - \sum_i \frac{\|\mathbf{w}_t^i\|_2^2}{2\|\mathbf{w}_t^i\|_2}$  成立,即

$$\|\mathbf{W}_{t+1}\|_{2,1} - \sum_i \frac{\|\mathbf{w}_{t+1}^i\|_2^2}{2\|\mathbf{w}_t^i\|_2} \leq \|\mathbf{W}_t\|_{2,1} - \sum_i \frac{\|\mathbf{w}_t^i\|_2^2}{2\|\mathbf{w}_t^i\|_2} \quad (34)$$

结合式(33)与式(34),有不等式(35)成立.

$$\begin{aligned} & \left\| \mathbf{W}_{t+1}^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \|\mathbf{W}_{t+1}\|_{2,1} \\ & \leq \left\| \mathbf{W}_t^T \tilde{X} \tilde{U}_t \tilde{\mathbf{P}}_t - \mathbf{V}_t \mathbf{E}_t \tilde{\mathbf{P}}_t \right\|_F^2 + \beta \|\mathbf{W}_t\|_{2,1} \end{aligned} \quad (35)$$

即

$$J(\mathbf{W}_{t+1}, \mathbf{U}_t, \mathbf{V}_t, \mathbf{p}_t) \leq J(\mathbf{W}_t, \mathbf{U}_t, \mathbf{V}_t, \mathbf{p}_t) \quad (36)$$

而由于  $\mathbf{U}, \mathbf{p}$  和  $\mathbf{V}$  分别由式(9)、式(11)和式(24)求解而来,因此有不等式(37)成立.

$$J(\mathbf{W}_t, \mathbf{U}_{t+1}, \mathbf{V}_{t+1}, \mathbf{p}_{t+1}) \leq J(\mathbf{W}_t, \mathbf{U}_t, \mathbf{V}_t, \mathbf{p}_t) \quad (37)$$

结合式(36)与式(37),可得

$$J(\mathbf{W}_{t+1}, \mathbf{U}_{t+1}, \mathbf{V}_{t+1}, \mathbf{p}_{t+1}) \leq J(\mathbf{W}_t, \mathbf{U}_t, \mathbf{V}_t, \mathbf{p}_t) \quad (38)$$

因此,依据算法 1 的迭代更新规则,目标函数式(8)的取值将在每次迭代过程中单调下降,直至算法收敛.证毕.

## 4 实验结果与分析

为了展示本文所提出的无监督特征选择算法的优势,本文在合成数据集和真实数据集上进行了多种实验.首先,简要介绍了一些相关算法以及评价指标.接着,对合成数据集进行了可视化实验.最后,简要介绍了现实世界的数据集,并通过各种比较验证了本文所提 DFRFS 算法的优越性.此外,本文还进行了参数敏感性分析、消融实验,以及收敛性验证和运行时间对比.

### 4.1 对比算法与评价指标

为验证 DFRFS 的有效性,本文采用了基准算法 AllFea 以及 MV<sup>[10]</sup>、MCFS<sup>[12]</sup>、UDFS<sup>[13]</sup>、NDFS<sup>[14]</sup>、SOCFS<sup>[17]</sup>、EGCFS<sup>[23]</sup>、SPCAFS<sup>[26]</sup>、VCSDFS<sup>[28]</sup>、基于模糊  $K$  均值聚类 and 稀疏投影的无监督特征选择 (unsupervised feature selection via Fuzzy  $K$ -Means and Sparse Projection, FKMSP)<sup>[37]</sup> 和 FSDK<sup>[27]</sup> 十种相关无监督特征选择算法与 DFRFS 进行对比实验.首先,使用这些算法

在不同的数据集上进行特征选择,然后利用K-Means算法进行聚类,从而评估这些特征选择算法的性能优劣.相关对比算法的介绍如下.

(1) AllFea: 基准算法. 使用K-Means算法对所有特征进行聚类作为基准.

(2) MV: MV通过计算每个特征的方差来评估其重要性,并选择具有MV的特征.

(3) MCFS: 受流形学习和 $\ell_1$ 正则化模型的启发,MCFS联合谱聚类和稀疏回归问题选择回归系数较大的特征,从而更好地保留数据的簇结构.

(4) UDFS: 在输入数据的类别标签可以通过线性分类器进行预测的前提下,UDFS将判别分析和 $\ell_{2,1}$ 正则化结合到无监督特征选择的联合框架中,以此选择最具判别性的特征子集.

(5) NDFS: NDFS同时考虑局部判别信息与特征间的相关性. 在学习输入数据的类别标签时,通过非负谱聚类进行特征选择,并施加 $\ell_{2,1}$ 正则化约束以减少冗余或噪声特征.

(6) SOCFS: SOCFS在特征选择框架中引入基矩阵与编码矩阵,通过正交基聚类学习输入数据的潜在聚类信息,进而引导投影矩阵选择具有区分性的特征.

(7) EGCFS: EGCFS将最大化类间散布矩阵和自适应图结构的思想融合到统一框架中,通过自适应图学习优化相似度矩阵,以选择不相关但具有区分性的特征.

(8) SPCAFS: SPCAFS通过将PCA形式化为重构误差最小化问题,并采用 $\ell_{2,p}$ 范数正则化项使得投影矩阵稀疏化,学得行稀疏和正交投影矩阵用于选择具有区分性的特征.

(9) VCSDFS: VCSDFS基于方差-协方差子空间距离整合特征的方差和协方差信息,进而排除冗余特征实现特征选择.

(10) FKMSF: FKMSF通过模糊算法获取数据的伪标签,并基于数据样本与聚类中心间的模糊距离进行特征选择.

(11) FSDK: FSDK结合 $\ell_{2,p}$ 范数与改进K-Means算法用于优化稀疏回归问题,能够在大规模数据上高效地执行特征选择.

为了评估实验结果,本文使用准确率(Accuracy, ACC)、归一化互信息(Normalized Mutual Information, NMI)和纯度(Purity)作为评价指标. 三个指标的数值范围均为 $[0, 1]$ ,其数值越大表示算法特征选择性能越好.

ACC用于测量通过算法获得的实际标签与预测标签之间的准确性,其表达式为

$$ACC = \frac{\sum_{i=1}^n \delta(\text{map}(f_i), l_i)}{N} \quad (39)$$

其中, $l_i$ 为 $x_i$ 的真实标签; $f_i$ 为 $x_i$ 的聚类结果;对于 $\delta(x, y)$ 函数,如果 $x=y$ ,则 $\delta(x, y)=1$ ,否则 $\delta(x, y)=0$ ;  $\text{map}(\cdot)$ 为将聚类标签置换以匹配真实标签的最佳排列映射函数,使用Kuhn-Munkres算法<sup>[46]</sup>进行映射.

NMI是计算相同数据的两个标签之间相似度的标准化度量,其表达式为

$$NMI = \frac{I(C, Z)}{\sqrt{H(C)H(Z)}} \quad (40)$$

其中, $C$ 和 $Z$ 分别为真实标签和聚类结果; $I(C, Z)$ 为 $C$ 和 $Z$ 之间的互信息; $H(\bullet)$ 为熵.

Purity是衡量一个簇包含一个类中数据的程度,其表达式为

$$Purity = \frac{1}{n} \sum_k \max_j |Z_k \cap C_j| \quad (41)$$

其中, $\{Z_1, Z_2, \dots, Z_k\}$ 为经聚类算法处理后簇的划分; $\{C_1, C_2, \dots, C_k\}$ 为真实簇的划分.

为了保证对比实验的公平性,各部分实验中本文算法与相关算法均在相同条件下进行比较. 实验的运行环境如下所示:PC为HP,CPU为Intel Core i7-14700F 2.10 GHz, RAM为32 GB,应用软件为Matlab R2023b. 本文算法使用Matlab编程语言实现,并可从<https://github.com/gaoyl-group/DFRFS>下载.

## 4.2 合成数据集实验

为了验证DFRFS是否能够捕获数据真实结构并选择具有判别性的特征,本文在三个合成数据集上进行可视化实验. 在这些合成数据集的前两个维度,数据分别服从three-Gaussian、two-moon与three-ring分布. 而对于其他八个维度,则选择服从正态分布的随机数据作为噪声特征. 由于在上述合成数据集中,数据中的判别信息仅包含在前两个维度,因此能否准确选择出前两个维度对于特征选择算法而言至关重要.

SOCFS算法同时执行正交基聚类和特征选择,EGCFS算法则将自适应图学习、最大化类间散布矩阵与特征选择结合到一个统一的框架下来选择判别特征. 在无监督特征选择领域,SOCFS和EGCFS性能优异,且在数学表示形式上与本文所提算法有一定的相似性. 因此,在本节中使用SOCFS和EGCFS作为对比算法,特征选择的数目设置为2,实验结果如图2所示,不同算法选择的二维特征在括号中标明.

由图2可知,对于包含噪声的three-Gaussian数据集,EGCFS与SOCFS无法选择出全部的具有判别信息的特征. 对于数据分布更为复杂的two-moon、three-ring数据集,EGCFS无法选择出判别特征,SOCFS只能选择

出部分判别特征. 与 SOCFS 和 EGCFS 不同, 本文提出的 DFRFS 可以选择出上述三个合成数据集集中的全部判别特征. 这是由于 DFRFS 在无监督特征选择过程中,

通过模糊学习更准确地捕捉了数据的复杂结构, 并为样本赋予不同的权重, 使数据描述更加精确, 从而显著提升了算法的鲁棒性.

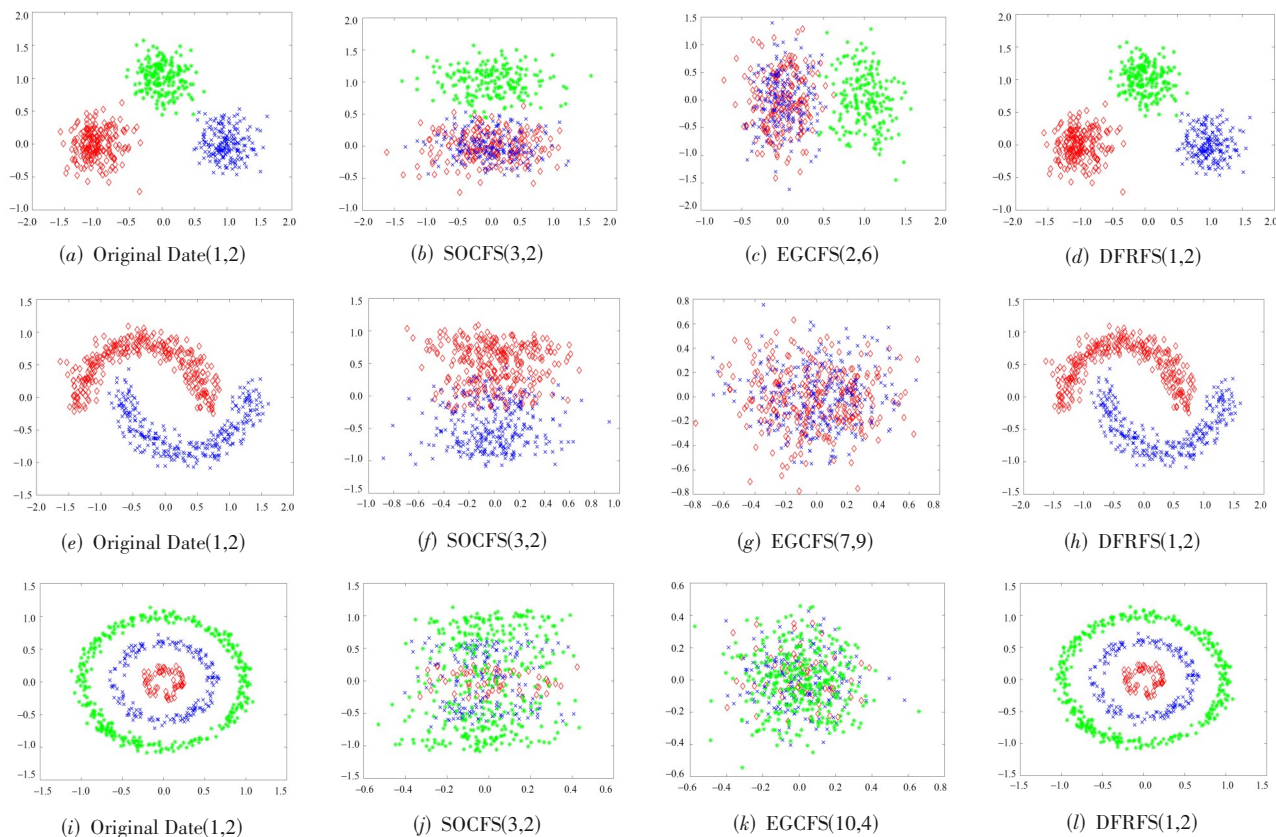


图2 不同算法在合成数据集上的特征选择结果

### 4.3 真实数据集介绍与实验参数设置

本文在 12 个真实数据集上进行了实验, 这些数据集涵盖了人脸图像、生物信息、医学、数字图像、时间序列、物理等多个领域. 数据集的详细介绍及实验中选择的特征数量如表 2 所示.

为确保对比实验的公平, 本文采用相同的策略设置所有对比算法的参数. 参考文献[17]工作中的实验设置, 对于 MCFS、NDFS、UDFS 和 EGCFS, 设置近邻点个数为 5. 参考文献[27]的工作, 对于 FSDK, 设置  $\ell_{2,p}$  范数参数  $p$  的寻优范围为  $\{10^{-3}, 10^{-2}, 10^{-1}, 0.5, 1, 1.5, 2\}$ . 对于其他正则化参数, 本文采用网格搜索策略进行调整, 调整的范围为  $\{10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$ . 对于 DFRFS, 本文设置正则化参数  $\beta$  的寻优范围为  $\{10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$ , 样本参考率  $r$  的寻优范围为  $\{0.1, 0.4, 0.7, 0.8, 0.9, 0.95, 0.98\}$ . 此外, 投影空间维度  $m$  设置为数据类别数  $c$ .

对原始数据进行特征选择后, 本文使用  $K$ -Means 算法对选择的特征进行聚类来评估特征选择算法的性

能. 由于  $K$ -Means 算法对于初始点的选择敏感, 参考文献[12]的做法, 使用随机初始点重复实验 20 次, 记录结果的平均值与方差.

### 4.4 真实数据集实验结果与分析

为了全面比较 12 种不同算法的性能, 本文在图 3~图 5 中展示了不同算法在 12 个真实数据集上具有不同数目特征时的 ACC、NMI 与 Purity 值的曲线, 其中横轴表示选择的特征数目, 纵轴表示 ACC 或 NMI 或 Purity. 此外, 表 3~表 5 展示了这些算法在 12 个真实数据集上具有最佳参数时的 ACC、NMI 与 Purity 值, 最优结果以粗体表示, 次优结果以下划线表示, 最后一行是不同对比算法在所有数据集上的平均结果.

由图 3~图 5 可知, 大多数对比算法的特征选择结果都优于使用所有特征的基线结果. 由此可见, 对高维数据进行特征选择十分重要, 因为原始高维数据总是包含大量无关、冗余和噪声特征, 这些特征会严重影响聚类结果. 无监督特征选择算法可以评估每个特征的重要性, 选择具有区分性的特征, 从而提高算法的性

表2 真实数据集描述

数据名称	样本数目	特征数目	类别数	特征选择数目	数据类型
Ionosphere	351	34	2	[2,4,6,8,10,12]	物理
Dermatology	366	34	6	[2,4,6,8,10,12]	医学
Control	600	60	6	[5,10,15,20,25,30]	时间序列
USPS49	1 532	256	2	[20,40,60,80,100,120]	数字图像
LungD	73	325	7	[50,70,90,110,130,150]	医学
Jaffe	213	676	10	[50,80,110,140,170,200]	人脸图像
Yale	165	1 024	15	[50,100,150,200,250,300]	人脸图像
ORL	400	1 024	40	[50,100,150,200,250,300]	人脸图像
SRBCTML	83	2 308	4	[50,100,150,200,250,300]	生物信息
WarpAR10P	130	2 400	10	[50,100,150,200,250,300]	人脸图像
LEUML	72	3 571	2	[50,100,150,200,250,300]	医学
TOX_171	171	5 748	4	[50,100,150,200,250,300]	生物信息

能. 本文提出的DFRFS在大多数数据集上优于对比算法, 原因在于DFRFS引入模糊隶属度, 允许数据在多个簇之间共享, 能够更好地反映数据的复杂结构和不确定性. 此外, DFRFS为样本赋予不同的权重, 能够在降低噪声样本影响的同时保留可靠样本. 因此, DFRFS可以从所有特征中选出最具判别信息的特征.

观察表3和表4, DFRFS在大部分数据集上都取得了最优的聚类ACC和NMI. 尽管在ORL、TOX\_171和WarpAR10P等数据集上没有取得最优的聚类ACC或NMI, 但DFRFS的实验结果也与其他对比算法相当. 由表5可知, DFRFS在Purity指标上整体优于其他对比算法. 此外, 本文提出的DFRFS算法在所有数据集上取得了最优的平均聚类ACC、NMI和Purity, 这进一步证明了DFRFS的有效性和优越性.

为了说明DFRFS与其他对比算法之间的统计性差异, 本文进行了Friedman-Nemenyi检验<sup>[47]</sup>. 图6分别展示了表3~表5中实验结果的统计性差异. 图的顶部线条为绘制不同算法平均排名的坐标轴, 坐标越靠近右侧, 性能越优. 此外, 图的上方展示了临界差异值(Critical Difference). 由图6可知, DFRFS显著优于MCFS、VCSDFS、SPCAFS、MV、UDFS等算法.

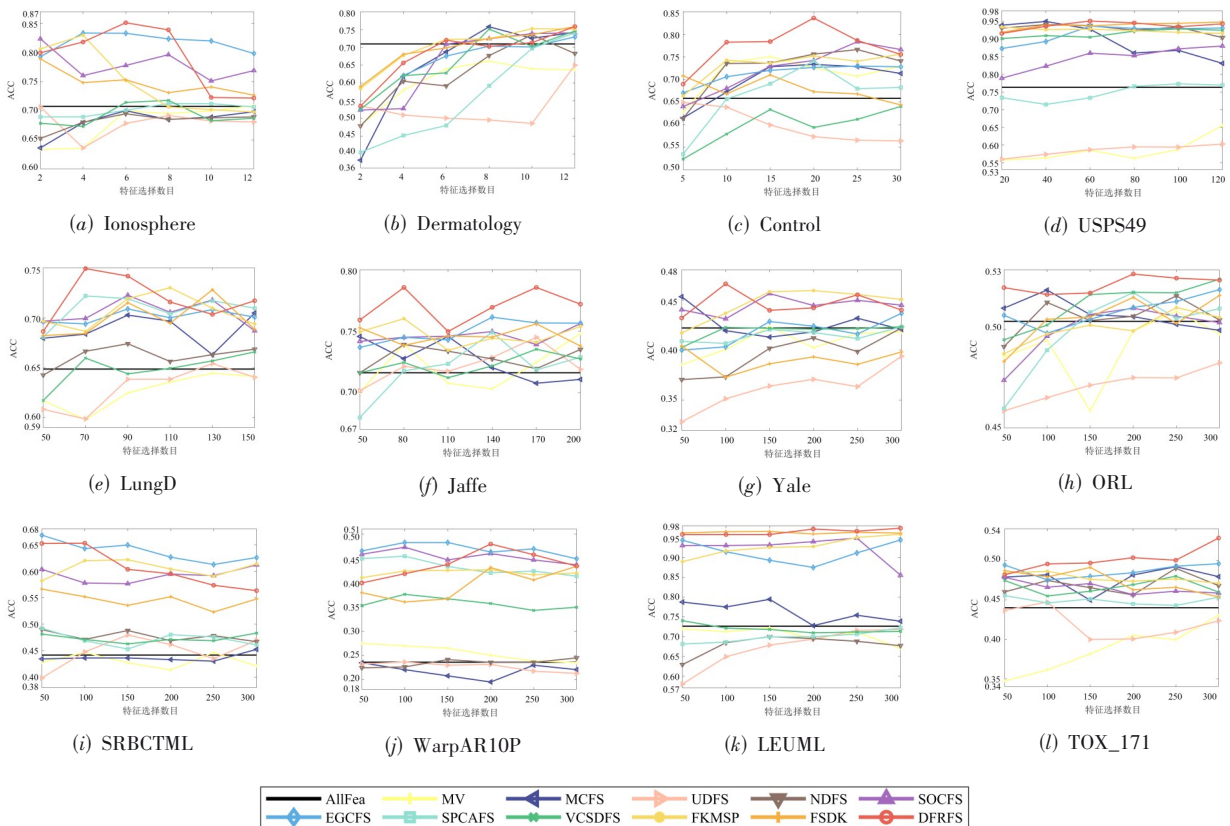


图3 所有算法在真实数据集上使用不同特征选择数目时的ACC

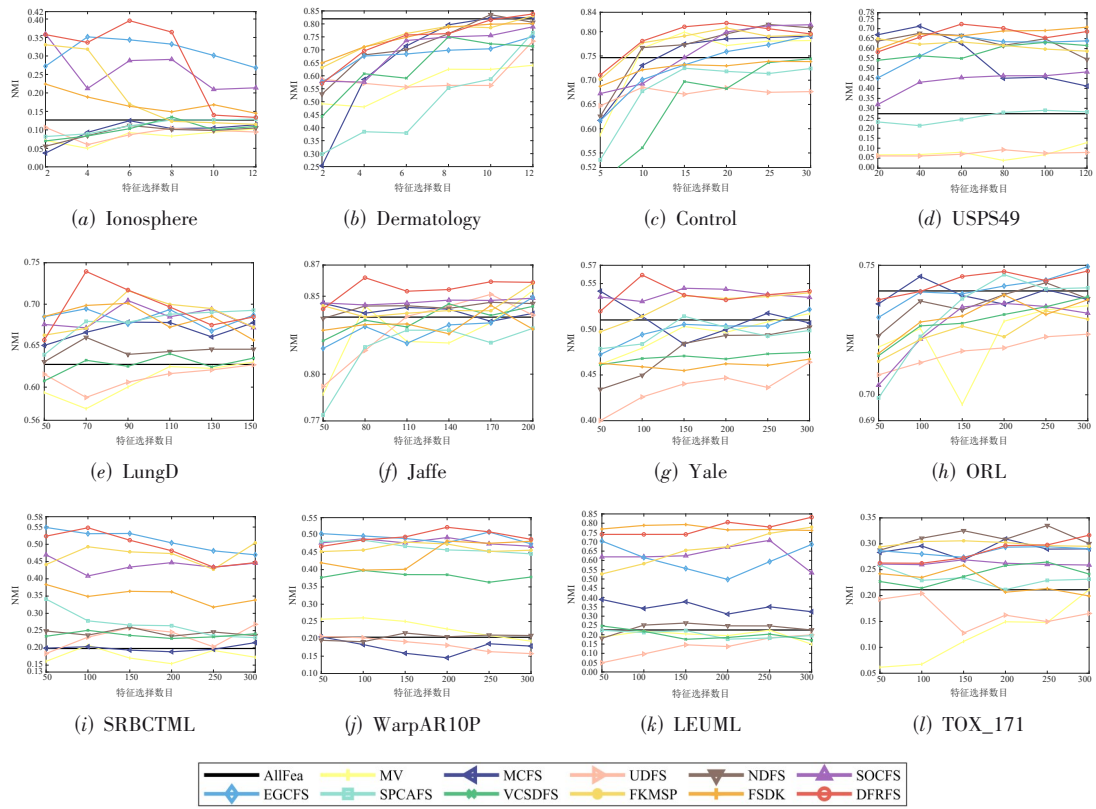


图 4 所有算法在真实数据集上使用不同特征选择数目时的 NMI

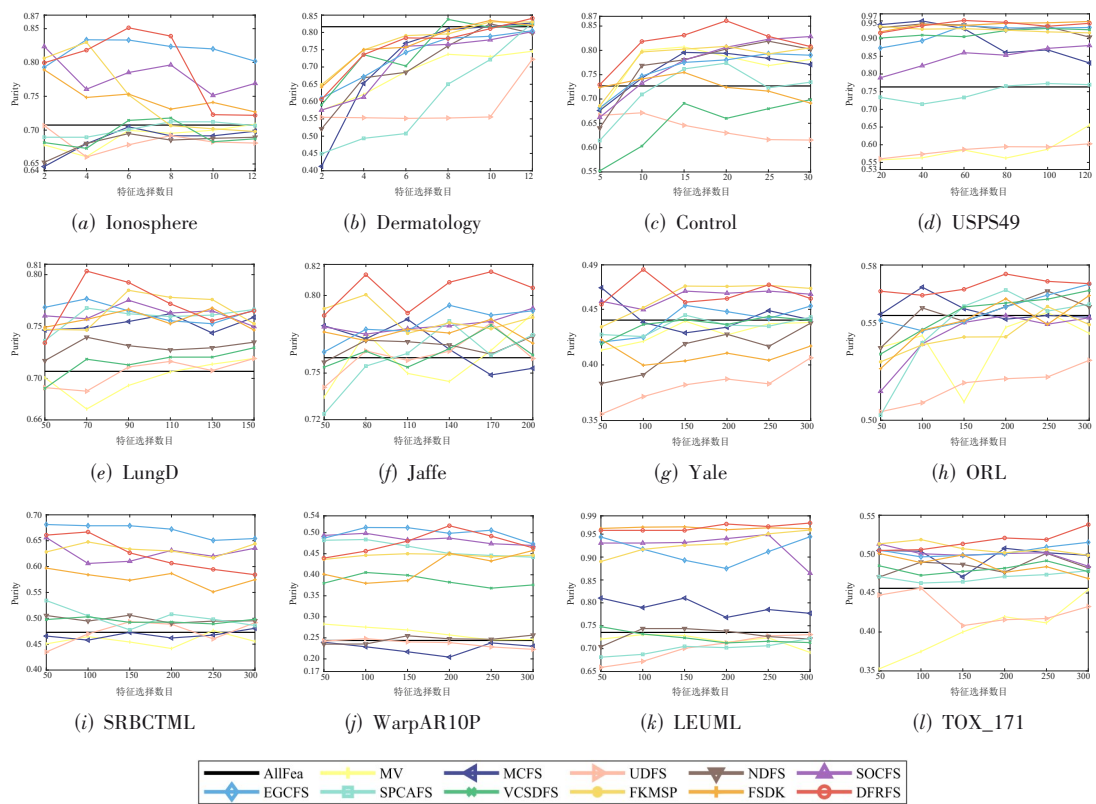


图 5 所有算法在真实数据集上使用不同特征选择数目时的 Purity

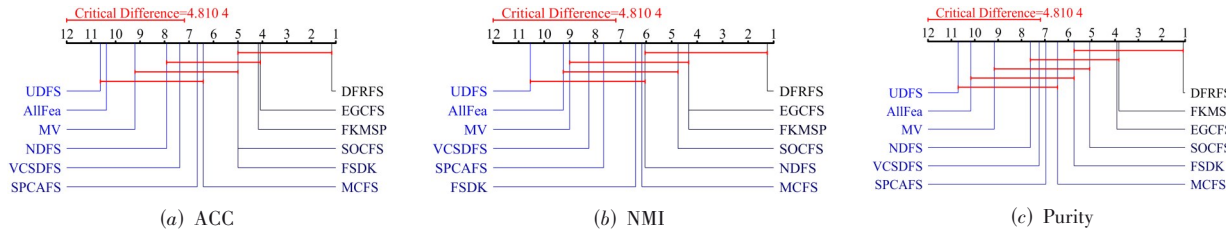


图6 不同算法在 ACC、NMI 和 Purity 上的 Nemenyi 检验(显著性水平为 0.05)

表3 所有算法在真实数据集上的实验结果(ACC±标准差)

单位:%

数据集	AllFea	MV	MCFS	UDFS	NDFS	SOCFS	EGCFS	SPCAFS	VCSDFS	FKMSP	FSDK	DFRFS
Ionosphere	70.75±	70.87±	70.01±	70.66±	69.49±	82.36±	<u>83.36±</u>	71.23±	71.79±	82.99±	78.92±	<u>85.13±</u>
	0.14	0.13	5.96	0.00	0.58	2.54	<u>2.72</u>	0.00	0.00	0.28	0.00	<u>0.95</u>
Dermatology	70.96±	66.15±	75.79±	65.00±	73.55±	74.10±	72.95±	74.13±	75.14±	75.26±	<u>75.86±</u>	<u>75.87±</u>
	12.00	5.88	7.83	6.94	11.00	8.45	9.55	9.11	9.26	7.39	<u>6.20</u>	<u>6.85</u>
Control	65.85±	75.09±	73.36±	64.93±	76.64±	<u>78.37±</u>	72.98±	74.13±	64.05±	75.63±	71.03±	<u>83.64±</u>
	7.06	9.64	7.88	5.44	6.67	<u>9.83</u>	5.96	6.58	7.52	10.37	6.48	<u>10.42</u>
USPS 49	76.34±	65.55±	<u>95.00±</u>	60.27±	94.13±	88.05±	93.69±	77.32±	92.89±	93.41±	94.84±	<u>95.13±</u>
	2.40	1.60	<u>0.06</u>	0.99	0.02	0.00	0.03	0.58	0.00	0.05	0.00	<u>0.03</u>
LungD	64.86±	64.45±	70.48±	65.41±	67.40±	72.26±	70.89±	72.19±	66.58±	<u>73.01±</u>	72.81±	<u>74.93±</u>
	6.58	7.24	6.73	6.48	6.94	7.12	7.46	9.25	7.74	<u>8.98</u>	6.38	<u>8.64</u>
Jaffe	71.62±	74.81±	74.58±	74.51±	73.92±	75.63±	<u>76.15±</u>	74.93±	73.54±	76.03±	75.61±	<u>78.57±</u>
	7.82	10.10	9.16	12.23	8.83	8.09	<u>10.12</u>	7.07	7.57	9.37	7.64	<u>8.67</u>
Yale	42.21±	42.36±	45.33±	39.39±	42.21±	45.64±	43.67±	42.33±	42.33±	<u>45.94±</u>	40.42±	<u>46.61±</u>
	4.05	4.31	5.08	2.61	4.78	2.94	4.73	3.26	4.61	<u>3.82</u>	4.46	<u>4.53</u>
ORL	50.39±	50.91±	51.98±	48.29±	51.68±	51.06±	52.00±	51.85±	<u>52.52±</u>	51.09±	51.71±	<u>52.79±</u>
	2.97	2.73	3.39	2.43	3.28	3.28	2.65	2.81	<u>2.46</u>	2.71	3.06	<u>2.58</u>
SRBCTML	44.16±	44.64±	45.24±	47.95±	48.98±	61.20±	<u>66.81±</u>	49.22±	48.31±	62.17±	56.63±	<u>65.30±</u>
	6.72	6.17	7.21	6.27	4.55	10.02	<u>9.17</u>	5.32	4.83	11.88	9.57	<u>6.42</u>
Warp ARIOP	23.62±	27.50±	23.50±	24.50±	24.50±	47.27±	<u>48.23±</u>	45.42±	37.58±	42.62±	43.35±	<u>47.92±</u>
	3.15	2.66	2.88	4.16	4.16	3.06	<u>4.29</u>	3.55	3.12	3.02	5.37	<u>4.51</u>
LEUML	72.64±	72.22±	79.44±	71.81±	70.00±	94.93±	94.44±	72.22±	74.03±	95.90±	<u>96.60±</u>	<u>97.43±</u>
	9.39	6.40	13.53	9.95	15.43	0.82	0.00	6.48	6.39	1.31	<u>0.71</u>	<u>1.37</u>
TOX_171	44.01±	43.04±	49.21±	44.74±	49.01±	47.87±	<u>49.59±</u>	45.56±	48.01±	48.63±	49.12±	<u>52.84±</u>
	4.57	3.76	5.03	4.21	3.41	1.59	<u>3.62</u>	1.35	4.02	2.92	0.83	<u>3.55</u>
Mean	58.12	58.13	62.83	56.46	61.79	68.23	<u>68.73</u>	62.54	62.23	68.56	67.24	<u>71.35</u>

### 4.5 参数敏感性分析

目标函数式(8)中主要的参数为 $\alpha$ 和 $\beta$ ,其中 $\alpha$ 用于避免简单解与过拟合,而 $\beta$ 用于控制投影矩阵的行稀疏性.由式(18)与式(20)可以看出 $\alpha$ 由参考样本数目 $k$ 决定,参考样本数目 $k$ 的设定可以转换为对参考样本率 $r$ 的设定.因此,本文所提出的算法主要考虑 $r$ 和 $\beta$ 两个参数.

固定参数 $r=0.9$ ,DFRFS在不同 $\beta$ 和特征选择数目下的聚类结果如图7所示.由图7可知,DFRFS对参数 $\beta$ 敏感,并且当 $\beta$ 较小时获得的结果更好.这是因为 $\beta$ 用于控制投影矩阵的行稀疏性,而过大的 $\beta$ 会导致投影矩阵过于稀疏,从而影响对关键特征的选择.

固定参数 $\beta=0.01$ ,DFRFS在不同 $r$ 和特征选择数

目下的聚类结果如图8所示.由图8可知,DFRFS在多数数据集上对参数 $r$ 不敏感.对于LungD数据集,DFRFS对参数 $r$ 敏感.这是因为 $r$ 用于控制参考样本数目,而参考样本数目过多或过少会导致噪声污染或训练样本不足等问题,从而影响算法的性能.

### 4.6 消融实验

本文算法在基于正则化回归的无监督特征选择中引入了模糊隶属度和稀疏权重,以更好地学习数据的复杂结构并降低噪声样本的影响.为了验证模糊隶属度 $U$ 和权重 $p$ 的有效性,本文进行了消融实验.消融后的模型如表6所示,其中DFRFS是本文提出的模型,Model 1去除了模糊隶属度 $U$ ,Model 2

表4 所有算法在真实数据集上的实验结果(NMI±标准差)

单位:%

数据集	AllFea	MV	MCFS	UDFS	NDFS	SOCFS	EGCFS	SPCAFS	VCSDFS	FKMSP	FSDK	DFRFS
Ionosphere	12.67±	10.26±	12.48±	10.67±	11.20±	<u>35.95±</u>	35.17±	12.92±	13.49±	33.03±	22.40±	<u>39.54±</u>
	0.24	0.13	8.18	0.00	0.53	<u>3.40</u>	2.89	0.00	5.36	0.56	0.00	2.23
Dermatology	81.95±	64.05±	82.49±	73.32±	<u>83.55±</u>	78.88±	75.12±	76.57±	75.03±	83.10±	80.13±	<u>83.83±</u>
	6.60	2.55	5.75	3.61	<u>6.60</u>	5.60	8.87	4.67	3.13	5.87	3.99	3.17
Control	74.65±	79.90±	79.09±	68.51±	<u>81.50±</u>	81.44±	79.23±	72.47±	74.42±	80.77±	73.88±	<u>81.78±</u>
	1.96	2.66	3.99	2.39	<u>2.41</u>	4.60	5.61	3.49	4.00	3.54	2.50	3.76
USPS 49	27.30±	12.80±	<u>71.26±</u>	9.22±	67.67±	48.29±	66.23±	29.00±	63.12±	64.92±	70.75±	<u>72.30±</u>
	3.58	1.37	<u>0.26</u>	0.04	0.09	0.32	0.14	0.79	0.00	0.20	0.00	<u>0.20</u>
LungD	62.74±	63.03±	67.84±	62.69±	66.00±	70.44±	69.45±	69.26±	64.03±	<u>71.63±</u>	70.11±	<u>73.95±</u>
	5.07	5.70	7.56	6.94	4.37	6.34	5.55	4.32	5.13	<u>6.14</u>	4.65	5.74
Jaffe	83.66±	85.36±	84.55±	85.12±	84.61±	84.87±	84.93±	82.94±	84.51±	<u>85.83±</u>	84.41±	<u>86.19±</u>
	4.25	4.67	4.41	5.93	4.13	3.75	3.71	5.25	4.10	<u>2.38</u>	3.86	4.49
Yale	51.02±	51.71±	54.17±	46.40±	50.23±	<u>54.49±</u>	52.17±	51.43±	47.47±	53.99±	46.73±	<u>55.93±</u>
	3.63	5.08	3.89	2.56	2.77	<u>2.11</u>	2.88	3.29	3.01	1.87	2.57	3.08
ORL	74.01±	73.44±	74.57±	72.34±	74.33±	73.53±	<u>74.95±</u>	74.64±	73.79±	73.24±	73.86±	<u>74.78±</u>
	1.49	1.45	1.77	1.03	1.52	1.49	<u>1.52</u>	1.30	1.27	1.46	1.51	<u>1.45</u>
SRBCTML	19.82±	20.77±	21.52±	26.82±	26.00±	46.96±	<u>54.87±</u>	34.12±	25.08±	50.42±	38.36±	<u>54.79±</u>
	8.76	11.55	10.05	12.66	4.47	14.39	<u>10.18</u>	9.24	6.53	13.56	15.10	<u>7.35</u>
Warp AR10P	20.47±	26.07±	20.57±	21.69±	21.69±	49.26±	<u>50.85±</u>	48.49±	39.78±	47.93±	48.23±	<u>52.22±</u>
	4.19	4.78	3.48	3.81	3.81	2.64	<u>5.40</u>	2.56	3.52	3.40	3.98	3.12
LEUML	22.56±	22.48±	38.98±	19.96±	26.30±	70.77±	70.41±	22.42±	24.74±	77.88±	<u>79.30±</u>	<u>83.23±</u>
	13.59	25.05	23.47	13.90	18.00	4.46	10.59	2.74	6.65	4.71	<u>4.93</u>	<u>8.02</u>
TOX_171	21.12±	21.13±	30.93±	20.42±	33.52±	26.91±	29.44±	25.86±	26.47±	30.58±	25.84±	<u>31.66±</u>
	7.61	6.36	5.20	5.19	<u>4.22</u>	1.97	3.94	3.15	4.31	2.60	2.84	<u>4.62</u>
Mean	46.00	44.25	53.20	43.10	52.22	60.15	61.90	50.01	50.99	<u>62.78</u>	59.50	<u>65.85</u>

去除了权重  $p$ , Model 3 则同时去除了模糊隶属度  $U$  和权重  $p$ .

实验数据集和参数设置与对比实验相同,消融实验结果如图9所示.由图9可知,Model 1和Model 2分别在DFRFS的基础上去除了模糊隶属度  $U$  和权重  $p$  后,性能均有所下降.此外,对于同时去除模糊隶属度  $U$  和权重  $p$  的基于正则化回归的模型 Model 3, Model 1和Model 2在此基础上分别引入了权重  $p$  和模糊隶属度  $U$ ,在绝大多数数据集上均表现出性能提升.DFRFS在同时引入模糊隶属度  $U$  和权重  $p$  后,实验效果进一步改善,这更好地验证了模糊隶属度  $U$  和权重  $p$  在无监督特征选择过程中的有效性.

#### 4.7 特征选择可视化结果

参考文献[22]的工作,本文选择Yale数据集的人脸图像进行特征选择可视化.在实验中,随机选择Yale数据集中的人脸样本,并将其重塑为  $32 \times 32$  灰度人脸图像,之后对每个样本选择  $\{50, 100, 150, 200, 250, 300\}$  个特征.在图10中,最左侧图像为重塑后的人脸图像.右侧图像为在该人脸图像上选择不同数目特征的可视化结果,从左往右选择的特征数量依次增

加.右侧图像第一行中,选定特征在图像上显示为白色像素,未选定的特征保持其原始值,从该行的可视化结果中可以看出,所选择的特征倾向于捕捉面部的重要部分,例如眉毛、眼睛、鼻子和嘴巴;右侧图像第二行中,未选定的特征在图像上显示为白色像素,选定特征保持其原始值,从该行的可视化结果中可以看出,所选择的特征集中在面部轮廓部分,如脸颊、下巴等.图10中的特征选择可视化结果表明本文所提算法可以选择相关且有区别的特征.

#### 4.8 收敛性分析与运行时间对比

本文通过实验分析验证算法的收敛性,固定参数  $r = 0.95, \beta = 100$  时,25次迭代中目标函数的取值情况如图11所示.图11的收敛曲线表明了本文提出的算法是收敛的,验证了3.3节中的收敛性证明.此外,由图11可知,本文所提出的算法收敛速度较快,通常在10次迭代内实现收敛.

表7总结了所有对比算法的计算复杂度.本文所提算法DFRFS的计算复杂度与类别数  $c$  相关.当类别数  $c$  较小时,DFRFS的计算复杂度与其他无监督特征选择算法相似.当类别数  $c$  较大时,DFRFS的计算效率会

表 5 所有算法在真实数据集上的实验结果(Purity±标准差)

单位:%

数据集	AllFea	MV	MCFS	UDFS	NDFS	SOCFS	EGCFS	SPCAFS	VCSDFS	FKMSP	FSDK	DFRFS
Ionosphere	70.75±	70.87±	70.48±	70.66±	69.49±	82.36±	<u>83.36±</u>	71.23±	71.79±	82.99±	78.92±	85.13±
	0.14	0.13	5.27	0.00	0.58	2.54	<u>2.72</u>	0.00	0.00	0.28	0.00	<b>0.95</b>
Dermatology	81.65±	74.55±	82.75±	72.23±	82.49±	80.22±	80.48±	82.80±	<u>83.77±</u>	83.05±	83.46±	<b>84.08±</b>
	5.75	4.59	4.09	4.37	5.68	3.61	1.72	3.55	<u>6.19</u>	2.48	3.00	<b>2.04</b>
Control	72.68±	80.70±	79.58±	67.16±	81.90±	<u>82.85±</u>	79.33±	77.38±	69.86±	80.83±	75.47±	<b>86.09±</b>
	5.03	6.34	2.73	4.10	4.41	<u>7.18</u>	5.50	4.08	5.56	5.49	3.57	<b>7.20</b>
USPS 49	76.34±	65.55±	<u>95.00±</u>	60.27±	94.13±	88.05±	93.69±	77.32±	92.89±	93.41±	94.84±	95.13±
	2.40	1.60	<u>0.06</u>	0.99	0.02	0.00	0.03	0.58	0.00	0.05	0.00	<b>0.03</b>
LungD	70.68±	71.92±	76.23±	71.92±	73.97±	77.53±	77.67±	76.78±	72.95±	<u>78.49±</u>	76.78±	<b>80.34±</b>
	4.67	4.84	4.38	6.23	3.87	5.28	5.11	5.38	4.49	<u>4.71</u>	4.52	<b>4.83</b>
Jaffe	75.99±	78.83±	78.47±	78.22±	77.37±	79.18±	79.37±	78.36±	78.08±	<u>80.05±</u>	78.47±	<b>81.53±</b>
	6.42	8.16	7.63	9.96	7.39	6.57	7.35	5.85	6.11	<u>6.99</u>	6.03	<b>6.73</b>
Yale	44.03±	43.94±	46.94±	40.64±	43.76±	46.64±	45.36±	44.48±	44.36±	<u>47.15±</u>	42.27±	<b>48.58±</b>
	3.55	3.58	4.49	2.26	4.11	2.80	3.35	3.98	3.85	<u>3.04</u>	3.92	<b>3.90</b>
ORL	55.40±	55.56±	56.86±	53.09±	56.66±	55.39±	<u>57.01±</u>	56.71±	56.69±	55.86±	56.42±	57.54±
	2.59	2.55	3.11	2.41	2.62	2.74	<u>2.28</u>	2.52	2.05	2.73	2.68	<b>2.04</b>
SRBCTML	47.29±	47.65±	48.07±	49.34±	50.60±	65.60±	<b>68.13±</b>	53.43±	50.30±	64.76±	59.70±	<u>66.69±</u>
	5.74	4.20	5.94	5.88	2.95	9.77	<b>7.59</b>	5.38	4.21	9.44	8.52	<u>4.89</u>
Warp AR10P	24.38±	28.27±	23.96±	25.65±	25.65±	49.85±	<u>51.23±</u>	48.38±	40.54±	45.00±	45.77±	51.65±
	3.01	2.65	3.05	3.92	3.92	2.65	<u>3.59</u>	2.96	2.62	3.62	4.76	<b>3.20</b>
LEUML	73.54±	73.06±	81.04±	73.06±	74.38±	94.93±	94.44±	72.22±	74.79±	95.90±	<u>96.60±</u>	<b>97.43±</b>
	8.15	7.94	10.89	8.51	11.21	0.82	0.00	6.48	3.72	1.31	<u>0.71</u>	<b>1.37</b>
TOX_171	45.61±	45.44±	50.76±	45.64±	50.12±	51.32±	51.52±	47.87±	49.15±	<u>51.87±</u>	50.09±	53.80±
	4.40	3.57	4.48	3.83	2.81	0.93	2.39	1.72	3.50	<u>2.22</u>	0.90	<b>2.60</b>
Mean	61.53	61.36	65.85	58.99	65.04	71.16	<u>71.80</u>	65.58	66.27	71.61	69.90	<b>74.00</b>

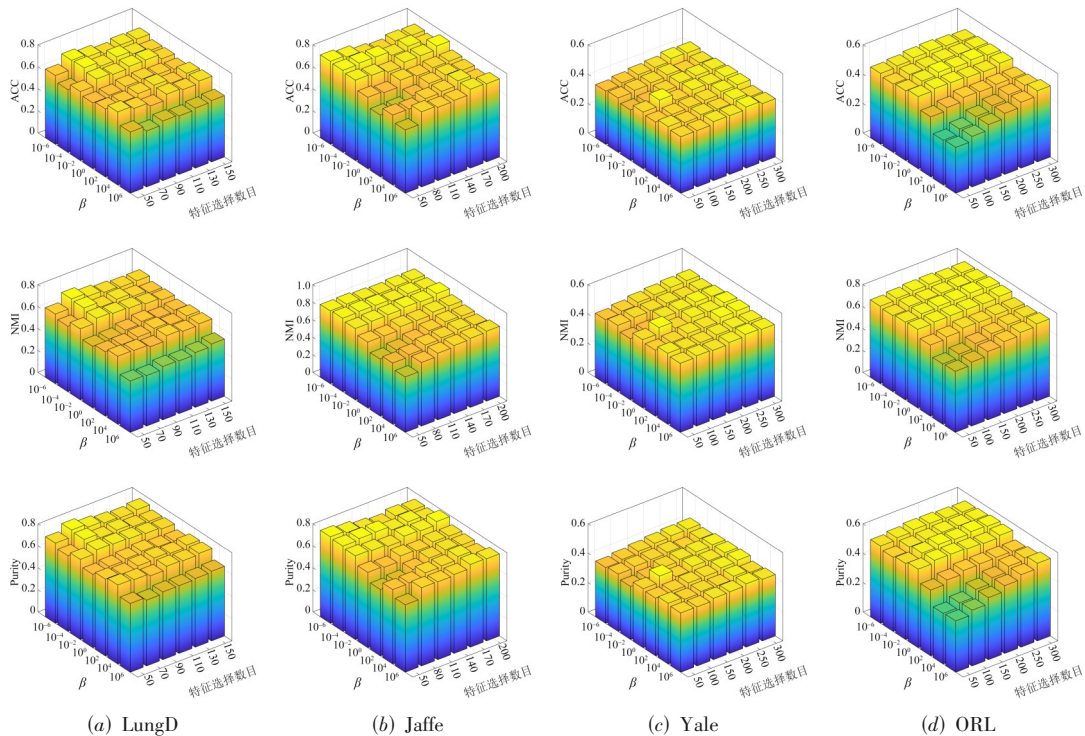


图 7  $r=0.9$ 时,不同 $\beta$ 和特征选择数目下的DFRFS实验结果

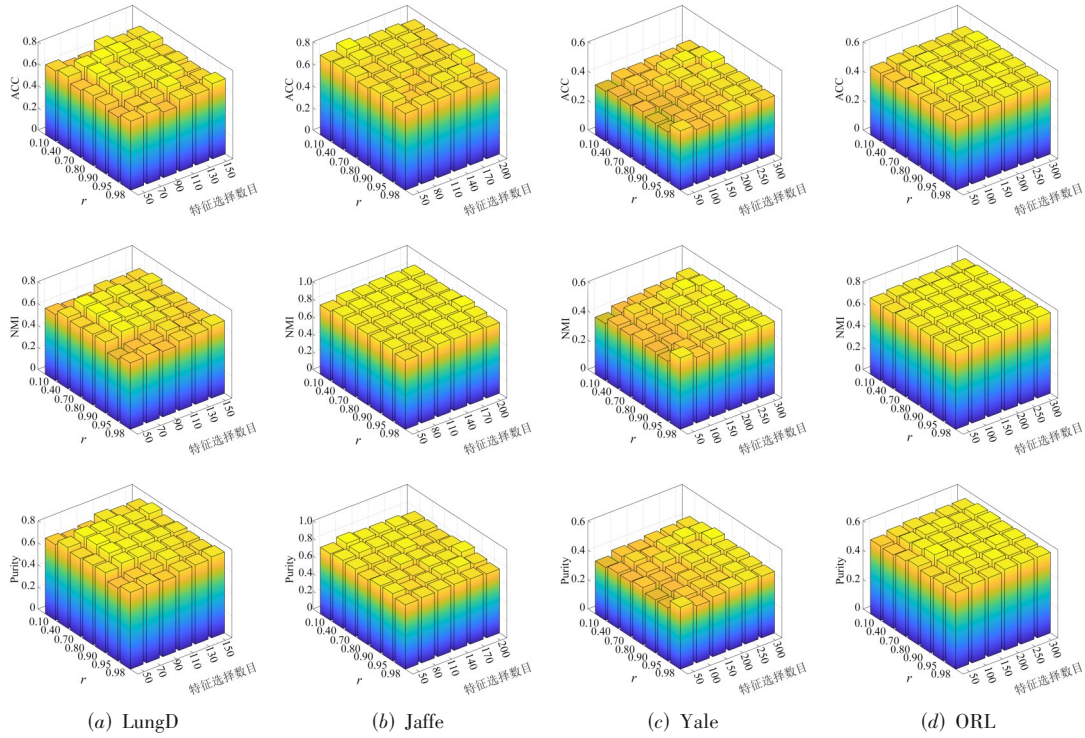


图8  $\beta=0.01$ 时,不同 $r$ 和特征选择数目下的DFRFS实验结果

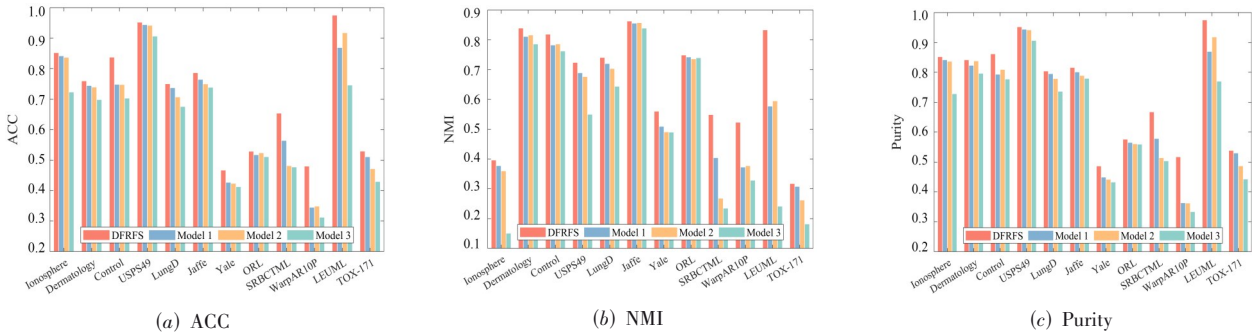


图9 不同消融模型和DFRFS在真实数据集上的实验结果

表6 消融实验对比模型介绍

模型	目标函数及约束
DFRFS	$\begin{cases} \min_{W,U,V,p} \sum_{j=1}^n p_j \sum_{i=1}^c u_{ij}^2 \ W^T x_j - v_i\ _2^2 + \alpha p_j^2 + \beta \ W\ _{2,1} \\ \text{s.t. } p^T \mathbf{1} = 1, u_j^T \mathbf{1} = 1, 0 \leq p_j, u_{ij} \leq 1, V^T V = I \end{cases}$
Model 1	$\begin{cases} \min_{W,U,V,p} \sum_{i=1}^c \sum_{j \in C_i} p_j \ W^T x_j - v_i\ _2^2 + \alpha p_j^2 + \beta \ W\ _{2,1} \\ \text{s.t. } p^T \mathbf{1} = 1, 0 \leq p_j \leq 1, V^T V = I \end{cases}$
Model 2	$\begin{cases} \min_{W,U,V} \sum_{j=1}^n \sum_{i=1}^c u_{ij}^2 \ W^T x_j - v_i\ _2^2 + \beta \ W\ _{2,1} \\ \text{s.t. } u_j^T \mathbf{1} = 1, 0 \leq u_{ij} \leq 1, V^T V = I \end{cases}$
Model 3	$\begin{cases} \min_{W,V} \sum_{i=1}^c \sum_{j \in C_i} \ W^T x_j - v_i\ _2^2 + \beta \ W\ _{2,1} \\ \text{s.t. } V^T V = I \end{cases}$



注:最左侧图像为Yale数据集中人脸图像,右侧图片为选择不同数量特征的可视化结果.

图10 特征选择可视化

变得较低. 对于上述问题,有相关的研究工作<sup>[48]</sup>进行探讨和处理. 未来值得研究的方向是进一步提高所提算法的效率. 图12比较了每种算法在所有真实数据集上的总运行时间. 考虑到算法的迭代次数以及现实世界中类别数 $c$ 通常远小于特征数目 $d$ 、样本总数 $n$ ,由图12可知,DFRFS的时间成本与EGCFS,SPCAFS等取

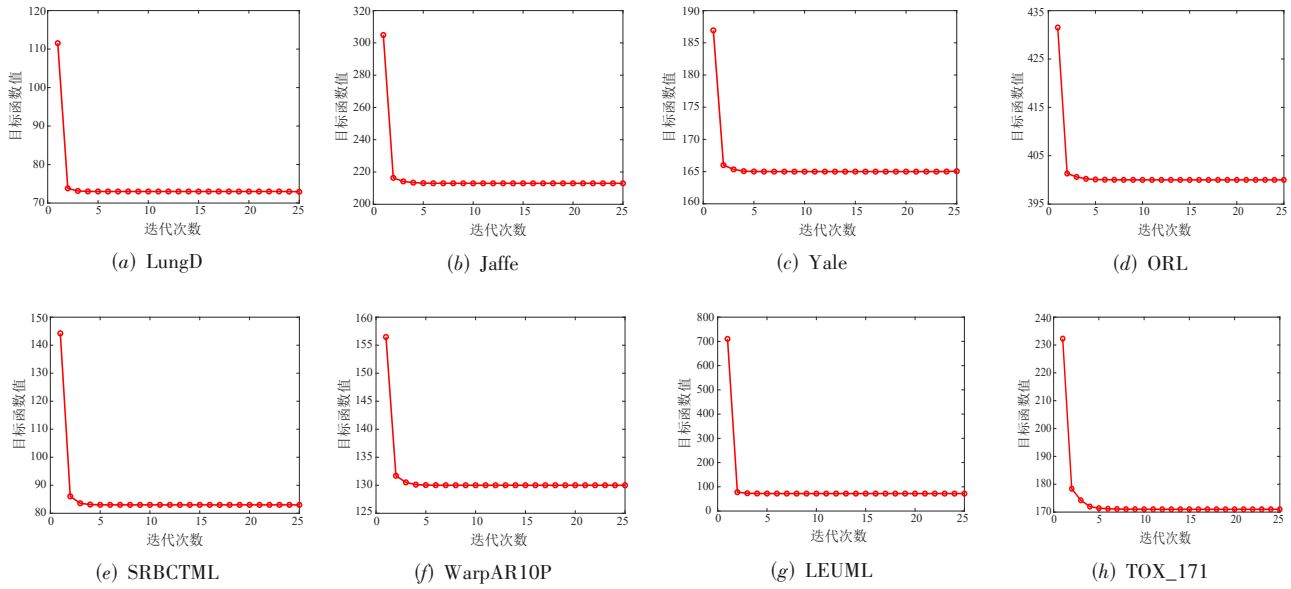


图 11 DFRFS 在真实数据集上的收敛曲线

得优良实验结果的算法相当. 此外,表 8 对比了 DFRFS 与在真实数据集上取得次优平均结果的 EGCFS 在不同数据集上的运行时间. 由表 8 可知,对于类别数较大的

ORL 数据集等,DFRFS 运行时间较长. 对于类别数目较小的 LEUML、TOX\_171 数据集等,DFRFS 计算较为高效,也更适合处理这类数据.

表 7 不同算法的计算复杂度

算法	MV	MCFS	UDFS	NDFS	SOCFS	EGCFS
复杂度	$O(dn)$	$O(d^3 + d^2n + n^2m)$	$O(d^3 + n^2c)$	$O(d^3 + n^2c)$	$O(n^3 + nc^2)$	$O(d^3 + n^3)$
算法	SPCAFS	VCSDFS	FSDK	FKMSP	DFRFS	
复杂度	$O(d^3 + d^2n)$	$O(d^2)$	$O(d^3 + d^2n)$	$O(d^3 + d^2n)$	$O(d^3 + d^2nc + dn^2c^2)$	

表 8 DFRFS 与 EGCFS 在真实数据集上的运行时间

单位:s

算法	Ionosphere	Dermatology	Control	USPS49	LungD	Jaffe
DFRFS	8.06	97.52	383.31	285.41	8.95	149.07
EGCFS	15.79	22.33	54.52	457.81	8.84	37.78
算法	Yale	ORL	SRBCTML	WarpAR10P	LEUML	TOX_171
DFRFS	265.35	11 853.31	50.32	174.08	146.89	476.60
EGCFS	72.90	139.03	498.57	665.13	2 191.18	11 083.32

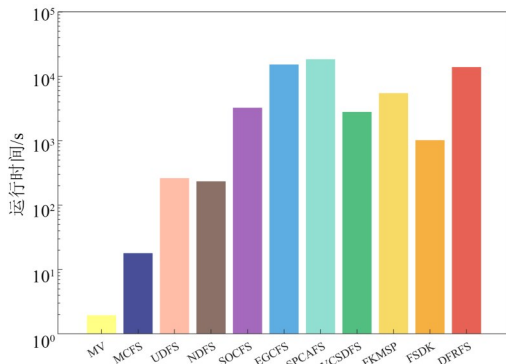


图 12 不同算法在真实数据集上的运行时间对比

### 5 结束语

针对现有的无监督特征选择算法忽视数据模糊性以及易受噪声影响的问题,本文结合模糊学习与鲁棒权重提出一种新颖的鲁棒无监督特征选择算法 DFRFS. 与其他基于正则化回归的无监督特征选择算法不同,DFRFS 考虑到了数据中的模糊性,通过引入模糊隶属度,更好地学习数据的真实结构. 此外,DFRFS 通过鲁棒权重学习框架为样本赋予不同的权重,在降低噪声样本影响的同时,保留了可靠样本的作用. 本文开发了一种简单而有效的迭代优化算法来进行求解,并在合成数据集与真实数据集上进行大量实验与分

析. 结果表明:DFRFS性能优于其他无监督特征选择算法. 具体实验过程中,矩阵变换需要较大时间开销,如何在处理大规模数据集时降低计算成本是我们未来的研究方向.

#### 参考文献

- [1] SASSI HIDRI M, ALI ZOGHLAMI M, BEN AYED R. Speeding up the large-scale consensus fuzzy clustering for handling Big Data[J]. *Fuzzy Sets and Systems*, 2018, 348: 50-74.
- [2] 刘艺, 曹建军, 刁兴春, 等. 特征选择稳定性研究综述[J]. *软件学报*, 2018, 29(9): 2559-2579.  
LIU Y, CAO J J, DIAO X C, et al. Survey on stability of feature selection[J]. *Journal of Software*, 2018, 29(9): 2559-2579. (in Chinese)
- [3] 张俐, 陈小波. 动态加权条件互信息的特征选择算法[J]. *电子与信息学报*, 2021, 43(10): 3028-3034.  
ZHANG L, CHEN X B. Feature selection algorithm for dynamically weighted conditional mutual information[J]. *Journal of Electronics & Information Technology*, 2021, 43(10): 3028-3034. (in Chinese)
- [4] 江兵兵, 何文达, 吴兴宇, 等. 基于自适应图学习的半监督特征选择[J]. *电子学报*, 2022, 50(7): 1643-1652.  
JIANG B B, HE W D, WU X Y, et al. Semi-supervised feature selection with adaptive graph learning[J]. *Acta Electronica Sinica*, 2022, 50(7): 1643-1652. (in Chinese)
- [5] 任永功, 王玉玲, 刘洋, 等. 基于用户相关性的动态网络媒体数据无监督特征选择算法[J]. *计算机学报*, 2018, 41(7): 1517-1535.  
REN Y G, WANG Y L, LIU Y, et al. Unsupervised feature selection algorithm for dynamic network media data based on user correlation[J]. *Chinese Journal of Computers*, 2018, 41(7): 1517-1535. (in Chinese)
- [6] SOLORIO-FERNÁNDEZ S, CARRASCO-OCHOA J A, MARTÍNEZ-TRINIDAD J F. A review of unsupervised feature selection methods[J]. *Artificial Intelligence Review*, 2020, 53(2): 907-948.
- [7] LUO M N, NIE F P, CHANG X J, et al. Adaptive unsupervised feature selection with structure regularization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(4): 944-956.
- [8] WANG A G, AN N, CHEN G L, et al. Accelerating wrapper-based feature selection with K-nearest-neighbor[J]. *Knowledge-Based Systems*, 2015, 83: 81-91.
- [9] AMBUSAIIDI M A, HE X J, NANDA P, et al. Building an intrusion detection system using a filter-based feature selection algorithm[J]. *IEEE Transactions on Computers*, 2016, 65(10): 2986-2998.
- [10] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in python[J]. *Journal of Machine Learning Research*, 2011, 12: 2825-2830.
- [11] 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 H M, WANG Y H, BIAN C, et al. Research on feature selection based on hybrid evolutionary algorithm[J]. *Acta Electronica Sinica*, 2023, 51(6): 1619-1636. (in Chinese)
- [12] CAI D, ZHANG C Y, HEX F, et al. 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.
- [13] YANG Y, SHEN H T, MA Z G, et al.  $\ell_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning[C]//*Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence-Volume Two*. New York: ACM, 2011: 1589-1594.
- [14] LI Z C, YANG Y, LIU J, et al. Unsupervised feature selection using nonnegative spectral analysis[C]//*Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. New York: ACM, 2012: 1026-1032.
- [15] QIAN M J, ZHAI C X, QIAN M J, et al. Robust unsupervised feature selection[C]//*Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. New York: ACM, 2013: 1621-1627.
- [16] WANG S H, TANG J L, LIU H, et al. Embedded unsupervised feature selection[C]//*Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. New York: ACM, 2015: 470-476.
- [17] HAN D, KIM J. Unsupervised simultaneous orthogonal basis clustering feature selection[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2015: 5016-5023.
- [18] NIE F P, ZHU W, LI X L, et al. Unsupervised feature selection with structured graph optimization[C]//*Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. New York: ACM, 2016: 1302-1308.
- [19] 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.
- [20] WANG C C, WANG J, GU Z C, et al. Unsupervised feature selection by learning exponential weights[J]. *Pattern Recognition*, 2024, 148: 110183.
- [21] LIN X C, GUAN J W, CHEN B L, et al. Unsupervised

- feature selection *via* orthogonal basis clustering and local structure preserving[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(11): 6881-6892.
- [22] LUO C, ZHENG J, LI T R, et al. Orthogonally constrained matrix factorization for robust unsupervised feature selection with local preserving[J]. *Information Sciences*, 2022, 586: 662-675.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] LI Z X, NIE F P, BIAN J T, et al. Sparse pca via  $\ell_{2,p}$ -norm regularization for unsupervised feature selection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(4): 5322-5328.
- [27] NIE F P, MA Z Y, WANG J Y, et al. Fast sparse discriminative K-means for unsupervised feature selection[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(7): 9943-9957.
- [28] KARAMI S, SABERI-MOVAHED F, TIWARI P, et al. Unsupervised feature selection based on variance-covariance subspace distance[J]. *Neural Networks*, 2023, 166: 188-203.
- [29] CAO Z W, XIE X J. Structure learning with consensus label information for multi-view unsupervised feature selection[J]. *Expert Systems with Applications*, 2024, 238: 121893.
- [30] CAO Z W, XIE X J, LI Y Q. Multi-view unsupervised feature selection with consensus partition and diverse graph[J]. *Information Sciences*, 2024, 661: 120178.
- [31] ZHOU S X, SONG P. Consistency-exclusivity guided unsupervised multi-view feature selection[J]. *Neurocomputing*, 2024, 569: 127119.
- [32] ZADEH L A. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338-353.
- [33] PAL S K, DE R K, BASAK J. Unsupervised feature evaluation: A neuro-fuzzy approach[J]. *IEEE Transactions on Neural Networks*, 2000, 11(2): 366-376.
- [34] LI Y, LU B L, WU Z F. A hybrid method of unsupervised feature selection based on ranking[C]//The 18th International Conference on Pattern Recognition (ICPR'06). Piscataway: IEEE, 2006: 687-690.
- [35] MAC PARTHALÁIN N, JENSEN R. Unsupervised fuzzy-rough set-based dimensionality reduction[J]. *Information Sciences*, 2013, 229: 106-121.
- [36] WANG R, BIAN J T, NIE F P, et al. Unsupervised discriminative projection for feature selection[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(2): 942-953.
- [37] JIANG K, ZHU L, SUN Q D. Unsupervised Feature Selection via Fuzzy K-means and Sparse Projection[M]// *Lecture Notes in Computer Science*. Singapore: Springer Nature Singapore, 2024: 224-239.
- [38] PAL N R, BEZDEK J C. On cluster validity for the fuzzy c-means model[J]. *IEEE Transactions on Fuzzy Systems*, 1995, 3(3): 370-379.
- [39] 潘金艳, 高朋, 高云龙, 等. 基于可靠性的鲁棒模糊聚类[J]. *控制理论与应用*, 2021, 38(4): 516-528.
- PAN J Y, GAO P, GAO Y L, et al. Reliability-based of robust fuzzy clustering[J]. *Control Theory & Applications*, 2021, 38(4): 516-528. (in Chinese)
- [40] ZHANG R, TONG H H. Robust principal component analysis with adaptive neighbors[J]. *Advances in Neural Information Processing Systems* 32, 2019, 32: 6961-6969.
- [41] NIE F P, HUANG H, CAI X, et al. Efficient and robust feature selection via joint  $\ell_{2,1}$ -norms minimization[C]// *Proceedings of the 24th International Conference on Neural Information Processing Systems-Volume 2*. New York: ACM, 2010: 1813-1821.
- [42] AFONSO M V, BIOUCAS-DIAS J M, FIGUEIREDO M A T. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems[J]. *IEEE Transactions on Image Processing*, 2011, 20(3): 681-695.
- [43] BEZDEK J C, EHRlich R, FULL W. FCM: The fuzzy c-means clustering algorithm[J]. *Computers & Geosciences*, 1984, 10(2/3): 191-203.
- [44] NIE F P, WU D Y, WANG R, et al. Truncated robust principle component analysis with a general optimization framework[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(2): 1081-1097.
- [45] VIKLANDS T. Algorithms for the Weighted Orthogonal Procrustes Problem and Other Least Squares Problems[D]. Umea: Department of Computing Science, Umea University, 2006.
- [46] KUHN H W. The Hungarian method for the assignment problem[J]. *Naval Research Logistics Quarterly*, 1955, 2(1/2): 83-97.

[47] DEMŠAR J. Statistical comparisons of classifiers over multiple data sets[J]. Journal of Machine Learning Research, 2006, 7: 1-30.

[48] ZHANG R X, HANG S J, SUN Z S, et al. Anchor-based fast spectral ensemble clustering[J]. Information Fusion, 2025, 113: 102587.

### 作者简介



**高云龙** 男,1979年8月出生于山西省忻州市.现为厦门大学副教授.2005年获得兰州大学计算机专业硕士学位.2011年获得西安交通大学控制科学与工程专业博士学位.主要研究方向为模式识别和时间序列分析.

E-mail: gaoyl@xmu.edu.cn



**曹超** 男,1982年1月出生于安徽省芜湖市.现为自然资源部第三海洋研究所研究员.主要研究方向为基于机器学习的海岸侵蚀机制与风险评价.

E-mail: caochao@tio.org.cn



**史曙光** 男,1999年7月出生于河南省周口市.现为厦门大学硕士研究生.主要研究方向为机器学习与维数约简.

E-mail: 23220231151771@stu.xmu.edu.cn



**潘金艳** 女,1978年7月出生于新疆维吾尔自治区石河子市.现为集美大学副教授.主要研究方向为最优化方法和数据挖掘.

E-mail: jypan@jmu.edu.cn



**赵志翔** 男,2000年11月出生于内蒙古自治区包头市.现为厦门大学硕士研究生.主要研究方向为机器学习和数据聚类.

E-mail: zhaozhixiang@stu.xmu.edu.cn