

# 基于 $F$ 范数群组效应和谱聚类的无监督特征选择<sup>①</sup>



林清水<sup>1,2</sup>, 田鹏飞<sup>1</sup>, 张 旺<sup>1</sup>

<sup>1</sup>(辽宁工程技术大学 软件学院, 葫芦岛 125105)

<sup>2</sup>(辽宁工程技术大学 基础教学部, 葫芦岛 125105)

通信作者: 林清水, E-mail: 1961896290@qq.com

**摘 要:** 基于谱聚类的无监督特征选择主要涉及相关系数矩阵和聚类指示矩阵, 在以往的研究中, 学者们主要关注于相关系数矩阵, 并为此设计了一系列约束和改进, 但仅关注相关系数矩阵并不能充分学习到数据内在结构. 考虑群组效应, 本文向聚类指示矩阵施加  $F$  范数, 并结合谱聚类以使相关系数矩阵学习更为准确的聚类指示信息, 通过交替迭代法求解两个矩阵. 不同类型的真实数据集实验表明文中方法的有效性, 此外, 实验表明  $F$  范数还可以使方法更加鲁棒.

**关键词:** 无监督特征选择; 谱聚类; 群组效应;  $F$  范数; 降维

引用格式: 林清水, 田鹏飞, 张旺. 基于  $F$  范数群组效应和谱聚类的无监督特征选择. 计算机系统应用, 2024, 33(7): 201-212. <http://www.c-s-a.org.cn/1003-3254/9569.html>

## Unsupervised Feature Selection Based on $F$ -norm Group Effect and Spectral Clustering

LIN Qing-Shui<sup>1,2</sup>, TIAN Peng-Fei<sup>1</sup>, ZHANG Wang<sup>1</sup>

<sup>1</sup>(Software College, Liaoning Technical University, Huludao 125105, China)

<sup>2</sup>(Department of Basic Teaching, Liaoning Technical University, Huludao 125105, China)

**Abstract:** Unsupervised feature selection based on spectral clustering mainly involves the correlation coefficient matrix and the clustering indicator matrix. In previous studies, scholars have mainly focused on the correlation coefficient matrix, designing a series of constraints and improvements for it. However, focusing solely on the correlation coefficient matrix cannot fully learn the intrinsic structure of data. Considering the group effect, this study imposes the  $F$ -norm on the clustering indicator matrix and combines it with spectral clustering to make the correlation coefficient matrix learn more accurate clustering indicator information. The two matrices are solved through an alternating iteration method. Experiments on different types of real datasets show the effectiveness of the proposed method. In addition, experiments show that the  $F$ -norm can also make the method more robust.

**Key words:** unsupervised feature selection; spectral clustering; group effect;  $F$ -norm; dimension reduction

在许多现实生活应用中, 如人脸识别<sup>[1]</sup>、图像分类<sup>[2]</sup>、文字识别<sup>[3]</sup>、生物医学<sup>[4,5]</sup>等问题, 将观测对象表示为高维特征向量已成为一种必然趋势. 高维特征向量可以充分描述观测对象, 但也增加了数据可视化和可解释难度<sup>[6]</sup>. 在实际应用中, 并不是所有特征都是重要的和有鉴别性的, 高维特征向量中的大多数特征往往相互关联或冗余, 这些特征在聚类中会造成过拟

合、处理效率低等问题. 为解决这些问题, 最近的研究关注于从原始特征中选择出具有大量有效信息的特征. 这种方法被称为特征选择<sup>[7]</sup>.

相比于使用代数方法降低特征维度<sup>[8,9]</sup>, 特征选择直接从给定的特征集合中选择具有强分类信息且尽可能相互独立的特征组成特征子集, 从而实现更快的降维, 消除冗余. 而且, 由于选择的特征与原始特征有明

① 收稿时间: 2024-01-24; 修改时间: 2024-02-26; 采用时间: 2024-03-11; csa 在线出版时间: 2024-06-05

CNKI 网络首发时间: 2024-06-08

确的联系,所以选择出的特征具有更好的解释性,这有利于可视化和解释任务<sup>[10]</sup>。

根据算法中标签存在情形,特征选择可分为有监督<sup>[11]</sup>、半监督<sup>[12]</sup>、无监督<sup>[13-15]</sup>这3类。前两类算法的执行结果受限于标签信息的使用情况。在实际应用中,大量易获取的高维数据通常是无标签的,而为这些数据添加标签相当耗时且需要额外开销。因此无监督特征选择方法更为流行和实用。

无监督特征选择方法主要分为3类:过滤式、包裹式和嵌入式。过滤式方法通过相关统计量来选择重要特征,在模型训练过程中先进行特征选择再训练模型<sup>[16,17]</sup>,其特征选择和模型训练独立进行,不能充分考虑特征之间的组合关系。在包裹式方法中,特征选择被“包裹”在学习方法中,并以特征的分类性能作为评价标准<sup>[18]</sup>,但其计算开销较大,不利于大规模数据集求解。嵌入式方法在模型构建过程中同时进行特征选择<sup>[14]</sup>,在实践中具有较好的性能和相当低的计算复杂度。

为实现特征选择,无监督特征选择通常需要解决两个问题。首先,无监督特征选择需要可靠的标签,但无监督方法并没有真实标签作为参考。最近的研究方法通常将标签来源与聚类算法<sup>[15]</sup>结合,探索数据特征与伪标签信息的相关性来指导特征选择。其次,当数据维度过高时,运用有效求解方法寻找最优组合选择出有效特征是十分困难的,因为特征选择本质上属于组合优化问题,是NP难题<sup>[10]</sup>。在一些方法中,无监督特征选择方法的目标函数是凸的,以保证收敛到最优解。

研究表明高维数据并不是均匀线性分布<sup>[19]</sup>,其分布通常是稀疏的,高维数据包含了大量局部信息。谱图理论在获取数据局部信息方面展现出了良好性能<sup>[20]</sup>,因此,越来越多嵌入式无监督特征选择方法引入谱聚类以获得更好的聚类效果<sup>[21,22]</sup>。Cai等<sup>[13]</sup>提出MCFS(multi-cluster feature selection)通过谱聚类学习数据局部信息,然后对特征进行评分以选择有效特征。Li等<sup>[14]</sup>提出NDFS(nonnegative discriminative feature selection)引入谱聚类和非负约束,结合正交约束,得到更为离散的聚类信息。

在稀疏低秩子空间聚类问题中,为提高模型性能、避免关联矩阵过于稀疏,学者们提出了群组效应的概念<sup>[23,24]</sup>,即如果样本彼此接近,则其表示系数也是接近的,因此群组效应保证了同一子空间内的数据可以分组在一起。Lu等<sup>[23]</sup>提出LSR(least squares reg-

ression)考虑群组效应,对关联矩阵应用 $F$ 范数实现了子空间的有效分割。Wang等<sup>[25]</sup>提出SRSSC(structural reweight sparse subspace clustering)算法,利用群组效应和 $l_{2,1}$ 范数来确保相似的样本归为一类。

在方法学习过程中,通过对相关系数矩阵施加稀疏性约束<sup>[13,14,26,27]</sup>、正交约束<sup>[26,28,29]</sup>或不相关约束<sup>[30]</sup>等方式可以更有效地选择特征。Nie等<sup>[26]</sup>提出SOGFS(structured optimal graph feature selection)通过对相关系数矩阵施加 $l_{2,p}$ 范数和正交约束进行稀疏性特征选择。Du等<sup>[27]</sup>提出UGFS(unsupervised group feature selection)通过对相关系数矩阵施加 $l_{2,0}$ 约束进行组特征选择。Shang等<sup>[28]</sup>提出FSDSC(unsupervised feature selection via discrete spectral clustering)引入正交约束学习数据流行结构保留了数据几何信息。虽然这些研究方法可以有效地进行特征选择,但其在关注聚类指示矩阵方面存在不足,无法充分利用聚类指示矩阵的信息。

缺乏对聚类指示矩阵的关注,可能导致以下问题。首先,特征选择方法可能选择与聚类标签不相关但与其他特征高度相关的特征。这可能导致模型出现过拟合或低解释能力的问题。其次,特征选择方法忽略了聚类指示矩阵中的潜在结构和关联性,导致模型无法捕捉到指示向量之间的复杂关系。这可能导致聚类结果的不准确和方法泛化能力的下降。因此,本文提出对聚类指示矩阵关注不足的解决方法,即基于 $F$ 范数群组效应和谱聚类的无监督特征选择(unsupervised feature selection based on  $F$ -norm group effect and spectral clustering, FSGS)。考虑群组效应,对聚类指示矩阵施加 $F$ 范数约束,以强化聚类指示矩阵在聚类中的作用,通过交替迭代法求解优化问题,最终通过计算相关系数矩阵每行的 $l_2$ 范数选择出有效特征。在多个真实数据集上的实验证明了本文方法的有效性。

## 1 相关基础知识

给定数据矩阵 $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ ,其中 $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$ 表示第 $i$ 个样本,样本的原始特征数为 $d$ 维, $n$ 表示样本数量。特征选择方法选择的特征数量为 $r$ , $r \ll d$ 。

聚类指示矩阵 $\mathbf{G}=[\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n]^T \in \mathbb{R}^{n \times c}$ ,其中 $\mathbf{g}_i \in \mathbb{R}^{c \times 1}$ 表示第 $i$ 个样本对应的聚类指示向量, $c$ 表示数据集的簇类个数。

相关系数矩阵 $\mathbf{W}=[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d]^T \in \mathbb{R}^{d \times c}$ ,其中 $\mathbf{w}_i \in \mathbb{R}^{c \times 1}$ 表示样本第 $i$ 类特征与各簇类之间的相关系数。在

矩阵 $\mathbf{G}$ 和矩阵 $\mathbf{W}$ 中值越高表示相关性越强。

相似度矩阵使用 $\mathbf{S} \in \mathbb{R}^{m \times n}$ 表示。 $\mathbf{S}$ 中的值 $s_{ij}$ 表示样本 $\mathbf{x}_i$ 和样本 $\mathbf{x}_j$ 之间的相似度,一般采用热核函数<sup>[31]</sup>进行计算:

$$s_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{4t}} \quad (1)$$

其中, $e$ 是欧拉数, $t$ 是参数。

在实践中,有全连接法计算所有样本之间的相似度和 $k$ 近邻法,首先计算样本之间的欧氏距离,通过给定参数 $k$ ,选取距离当前样本最近的 $k$ 个样本作为邻居,然后计算其相似度,并令其余相似度为0两种方式。

给定矩阵 $\mathbf{H} \in \mathbb{R}^{m \times n}$ , $\mathbf{H}^T$ 表示矩阵的转置, $\text{tr}(\mathbf{H})$ 表示矩阵的迹, $\mathbf{H}^{-1}$ 表示矩阵的逆,矩阵的 $F$ 范数定义为:

$$\|\mathbf{H}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n h_{ij}^2} \quad (2)$$

给定向量 $\mathbf{h}_i \in \mathbb{R}^{m \times 1}$ ,其 $l_2$ 范数定义为:

$$\|\mathbf{h}_i\|_2 = \sqrt{\sum_{j=1}^n h_{ij}^2} \quad (3)$$

在本文中,大写字母表示矩阵,小写字母表示标量、索引或向量。

### 1.1 谱聚类

聚类指示矩阵需要充分反映出样本的判别信息。根据谱图理论<sup>[32]</sup>,如果两个样本点在数据集的亲和度图中具有高度相关性,则与它们对应的聚类标签也具有高度相关性。具体来说,数据点 $s_{ij}$ 表示的相似度越大,则指示向量 $\|\mathbf{g}_i - \mathbf{g}_j\|$ 之间的距离越小。因此,给定相似度矩阵 $\mathbf{S}$ ,其学习聚类指示矩阵 $\mathbf{G}$ 的目标函数为:

$$\begin{cases} \min \frac{1}{2} \sum_{i,j=1}^n s_{ij} \|\mathbf{g}_i - \mathbf{g}_j\|_F^2 \\ \text{s.t. } \mathbf{G}^T \mathbf{G} = \mathbf{I} \end{cases} \quad (4)$$

式(4)使 $\mathbf{S}$ 中相邻样本点之间的差异最小。根据谱理论<sup>[26]</sup>,式(4)的最终推导如下:

$$\begin{cases} \min \text{tr}(\mathbf{G}^T \mathbf{L} \mathbf{S} \mathbf{G}) \\ \text{s.t. } \mathbf{G}^T \mathbf{G} = \mathbf{I} \end{cases} \quad (5)$$

矩阵 $\mathbf{L} \mathbf{S}$ 是关于相似度矩阵 $\mathbf{S}$ 的拉普拉斯矩阵,其定义为 $\mathbf{L} \mathbf{S} = \mathbf{S} - \mathbf{D}$ ,矩阵 $\mathbf{D}$ 是关于相似度矩阵 $\mathbf{S}$ 的对角矩阵,其定义为 $d_{ii} = \sum_{j=1}^n s_{ij}$ 。

为了使指示矩阵获得更多的数据信息,对矩阵 $\mathbf{G}$ 施

加正交约束。正交约束 $\mathbf{G}^T \mathbf{G} = \mathbf{I}$ 促使同一簇类内的指示向量平方和为1。这避免了在最优化和迭代过程中产生平凡解,例如,某个聚类指示向量的值全为0。

### 1.2 $F$ 范数群组效应的无监督特征选择

相关系数矩阵 $\mathbf{W}$ 的作用是选择最有效的样本特征并尽可能准确地将数据矩阵 $\mathbf{X}$ 与聚类指示矩阵 $\mathbf{G}$ 相匹配。因此,对于矩阵 $\mathbf{W}$ 的学习经常采用最小化损失函数:

$$\begin{cases} \min \|\mathbf{X}^T \mathbf{W} - \mathbf{G}\|_F^2 \\ \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{cases} \quad (6)$$

其中, $\mathbf{X}^T \mathbf{W}$ 是一种线性降维的方法,通过对矩阵 $\mathbf{W}$ 施加正交约束实现原数据集的正交变换。通过正交变换可以将强相关的空间样本值变换成在新的表示下不相关或者相关性较弱的的数据,去除原高维特征中的冗余特征。对于给定的低维子空间,高维数据的正交投影能保留原始数据尽可能多的信息并使原始数据和相应投影之间差异最小化。

在之前的工作中,主要考虑矩阵 $\mathbf{W}$ 的稀疏性以实现鲁棒性的特征选择。但这种方法并不能直接描绘出样本与其聚类标签之间的关系。在以往研究中<sup>[14,26]</sup>,具有高度相关性的邻居样本被认为属于同一簇类。在稀疏低秩子空间聚类问题中,Lu等<sup>[23]</sup>和Hu等<sup>[24]</sup>进一步证明了数据点之间存在群组效应,这表明空间中相近的两样本点与其他同类样本点之间具有相似的亲和度。

基于群组效应的考虑,本文对式(6)添加关于聚类指示矩阵的正则项,并提出相应定理及证明:

$$\begin{cases} \min \|\mathbf{X}^T \mathbf{W} - \mathbf{G}\|_F^2 + \beta \|\mathbf{G}\|_F^2 \\ \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{cases} \quad (7)$$

其中, $\beta$ 是一个调整 $F$ 范数正则项影响的参数。

定理1. 给定数据集 $\mathbf{X} \in \mathbb{R}^{d \times n}$ 和与之对应的聚类指示矩阵 $\mathbf{G} \in \mathbb{R}^{n \times c}$ , $\forall i \neq j$ ,若 $\|\mathbf{x}_i - \mathbf{x}_j\| \rightarrow 0 \Rightarrow \|\mathbf{g}_i - \mathbf{g}_j\| \rightarrow 0$ ,则式(7)存在群组效应。

证明: 令 $\mathbf{x} \in \mathbb{R}^{d \times 1}$ 表示原数据集中的一个样本, $\mathbf{g} \in \mathbb{R}^{c \times 1}$ 表示矩阵 $\mathbf{G}$ 中一个指示向量。将式(7)写成向量形式:

$$\min \|\mathbf{W}^T \mathbf{x} - \mathbf{g}\|_F^2 + \beta \|\mathbf{g}\|_F^2 \quad (8)$$

令 $L(\mathbf{g}) = \|\mathbf{W}^T \mathbf{x} - \mathbf{g}\|_2^2 + \beta \|\mathbf{g}\|_2^2$ , $\mathbf{g}^*$ 是式(8)的最优解,故其最优解满足:

$$\frac{\partial L(\mathbf{g})}{\partial \mathbf{g}_k} \Big|_{\mathbf{g}=\mathbf{g}^*} = 0 \tag{9}$$

对  $L(\mathbf{g})$  关于向量  $\mathbf{g}$  求导:

$$(1+\beta)\mathbf{g} - \mathbf{W}^T \mathbf{x} = 0 \tag{10}$$

对于不同的  $\mathbf{x}_i$  和  $\mathbf{x}_j$ , 有:

$$(1+\beta)\mathbf{g}_i - \mathbf{W}^T \mathbf{x}_i = 0 \tag{11}$$

$$(1+\beta)\mathbf{g}_j - \mathbf{W}^T \mathbf{x}_j = 0 \tag{12}$$

由式 (11) 和式 (12) 可以得到:

$$\mathbf{g}_i - \mathbf{g}_j = \frac{1}{1+\beta} \mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j) \tag{13}$$

最终推导如下:

$$\|\mathbf{g}_i - \mathbf{g}_j\|_2 = a \|\mathbf{x}_i - \mathbf{x}_j\|_2 \tag{14}$$

其中,  $a$  为正数.

式 (14) 表明如果样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  属于同一类, 则与之相应的聚类指示向量之间是高度相关的, 它们之间差异几乎为零, 这将提高类内相似性和类间差异性并形成群组以便于聚类.

此外, 在应对噪声时, 聚类指示矩阵上施加  $F$  范数还可以提升鲁棒性<sup>[23]</sup>.

## 2 基于 $F$ 范数群组效应和谱聚类的无监督特征选择

### 2.1 模型构建

本文提出基于  $F$  范数群组效应和谱聚类的无监督特征选择 (unsupervised feature selection based on  $F$ -

norm group effect and spectral clustering, FSGS), 将聚类标签的学习和特征选择统一到一个联合框架, 同时学习聚类指示矩阵  $\mathbf{G}$  和相关系数矩阵  $\mathbf{W}$ . 通过对聚类指示矩阵施加  $F$  范数约束, 促使簇类形成群组, 结合谱聚类学习数据局部流形信息, 同时, 正交约束避免了平凡解, 最终进行特征选择降低样本维度. 具体目标函数描述如下:

$$\begin{cases} \min \|\mathbf{X}^T \mathbf{W} - \mathbf{G}\|_F^2 + \alpha \text{tr}(\mathbf{G}^T \mathbf{L} \mathbf{S} \mathbf{G}) + \beta \|\mathbf{G}\|_F^2 \\ \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}, \mathbf{G}^T \mathbf{G} = \mathbf{I} \end{cases} \tag{15}$$

其中  $\alpha$ 、 $\beta$  是正则化参数, 分别用于调整谱聚类和矩阵  $\mathbf{G}$  对目标函数的影响.

如图 1 所示, 图 1(a)–图 1(c) 分别表示通过主成分分析方法<sup>[8]</sup>对 Jaffe 数据集的原始所有特征、SOGFS 方法选择的有效特征和 FSGS 方法选择的有效特征分别降至 2 维时的图示. 两个方法选择的特征维度为 140. 通过图示可以明显看到, FSGS 方法将原始图像中的绝大部分簇类进行了组别的划分, 形成群组, 而 SOGFS 方法处理后的簇类则不具备群组效应.

### 2.2 模型求解

在方法 MCFS 中, 聚类指示矩阵  $\mathbf{G}$  首先进行学习, 然后再学习相关系数矩阵  $\mathbf{W}$ , 这是一个两步走的策略. 与之不同的是, 最近 FSDSC、UGFS 等方法广泛采用联合学习框架实现更精确的特征选择. 这是因为联合学习在调节损失函数时更具有灵活性. 即使最优化方法的计算复杂度增加, 实验结果依然展现出优越性. 受此启发, 本文提出的 FSGS 也将采用联合学习框架学习矩阵  $\mathbf{W}$  和  $\mathbf{G}$ .

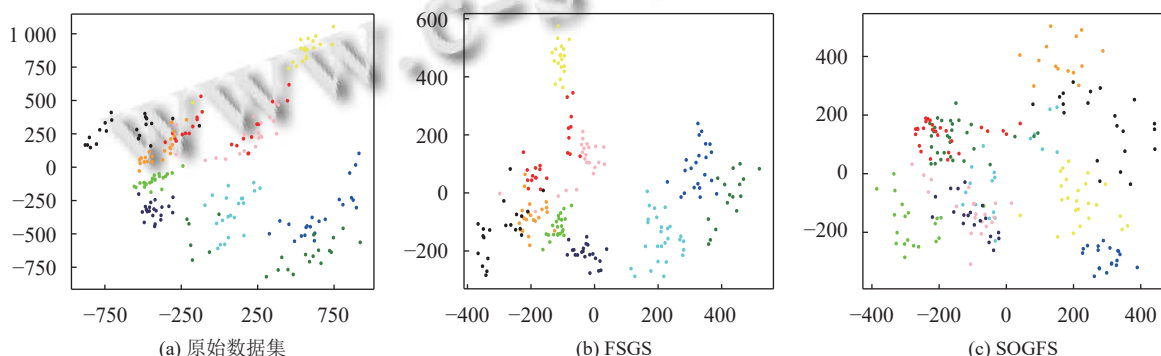


图 1 SOGFS 与 FSGS 聚类效果和原数据集对比

目标函数需要求解矩阵  $\mathbf{W}$  和  $\mathbf{G}$ . 由于矩阵  $\mathbf{W}$  和  $\mathbf{G}$  都具有正交约束, 因此目标函数是非凸的. 为了最优化目标函数, 本文采用交替迭代法求解目标函数.

在以往的方法中, 矩阵  $\mathbf{G}$  通常采用特征值分解法求解. 经过实践分析, 在矩阵维度较高时, 特征值分解法求解速度慢且并不容易取得好效果. 因此本文采

用代数方法求解矩阵 $\mathbf{G}$ ,代数方法不仅可以降低方法的计算复杂度而且可以简化相关系数矩阵的计算过程.

首先固定矩阵 $\mathbf{G}$ ,定义关于矩阵 $\mathbf{W}$ 的函数:

$$\mathcal{F}(\mathbf{G}, \mathbf{W}) = \|\mathbf{X}^T \mathbf{W} - \mathbf{G}\|_F^2 + \frac{1}{2} \mu \|\mathbf{W}^T \mathbf{W} - \mathbf{I}\|_F^2 \quad (16)$$

对 $\mathcal{F}(\mathbf{G}, \mathbf{W})$ 关于矩阵 $\mathbf{W}$ 求导并令其为0:

$$\frac{\partial \mathcal{F}}{\partial \mathbf{W}} = 2\mathbf{X}\mathbf{X}^T \mathbf{W} - 2\mathbf{X}\mathbf{G} + 2\mu(\mathbf{W}\mathbf{W}^T - \mathbf{I})\mathbf{W} = 0 \quad (17)$$

得到关于矩阵 $\mathbf{W}$ 的更新表达式:

$$\mathbf{W} = [\mathbf{X}\mathbf{X}^T + \mu(\mathbf{W}\mathbf{W}^T - \mathbf{I})]^{-1} \mathbf{X}\mathbf{G} \quad (18)$$

通过分析可知式(18)的时间复杂度为 $O(nd^2 + cd^2 + cnd + d^3)$ ,对于高维数据其时间复杂度非常之大.当 $d \gg n$ 时,采用如下方法,求解式(15)中的矩阵 $\mathbf{W}$ 等价于求解式(6),对于任意的矩阵 $\mathbf{H}$ 有运算法则,  $\|\mathbf{H}\|_F^2 = \text{tr}(\mathbf{H}^T \mathbf{H})$ ,对式(6)应用该法则可得:

$$\begin{cases} \min \text{tr}(\mathbf{W}^T \mathbf{A}^T \mathbf{W} - 2\mathbf{W}^T \mathbf{B}) \\ \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{cases} \quad (19)$$

其中,  $\mathbf{A} = \mathbf{X}\mathbf{X}^T$ ,  $\mathbf{B} = \mathbf{X}\mathbf{G}$ . 式(19)是一个典型的 Stiefel 流形上的二次优化问题 (quadratic problem of Stiefel manifold, QPSM), 可采用 Nie 等<sup>[33]</sup>提出的 GPI (generalized power iteration) 方法进行求解,其时间复杂度为 $O(nd^2 + cd^2)$ ,具体过程如算法1所示.

#### 算法 1. GPI 方法

输入: 矩阵 $\mathbf{A} \in \mathbb{R}^{d \times d}$ , 矩阵 $\mathbf{B} \in \mathbb{R}^{d \times c}$ .

输出: 矩阵 $\mathbf{W}$ .

- 1) 初始化: 初始化矩阵 $\mathbf{W} \in [0,1]^{d \times c}$ , s.t.  $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ ; 初始化参数 $v$ , 使得 $\tilde{\mathbf{A}} = v\mathbf{I} - \mathbf{A}$ 是一个正定矩阵.
- 2) 循环:
- 3) Step 1: 更新矩阵 $\mathbf{R}$ :  $\mathbf{R} \leftarrow \tilde{\mathbf{A}}\mathbf{W} + 2\mathbf{B}$ ;
- 4) Step 2: 使用压缩奇异值分解法求解矩阵 $\mathbf{R}$ , 得到 $\mathbf{U}\Sigma\mathbf{V}^T = \mathbf{R}$ ;
- 5) Step 3: 更新矩阵 $\mathbf{W}$ :  $\mathbf{W} \leftarrow \mathbf{U}\mathbf{V}^T$ ;
- 6) 收敛或达到最大迭代次数时结束循环.

然后固定矩阵 $\mathbf{W}$ ,定义关于矩阵 $\mathbf{G}$ 的函数:

$$\begin{aligned} \mathcal{L}(\mathbf{G}, \mathbf{W}) = & \|\mathbf{X}^T \mathbf{W} - \mathbf{G}\|_F^2 + \alpha \text{tr}(\mathbf{G}^T \mathbf{L} \mathbf{S} \mathbf{G}) \\ & + \beta \|\mathbf{G}\|_F^2 + \frac{1}{2} \gamma \|\mathbf{G}^T \mathbf{G} - \mathbf{I}\|_F^2 \end{aligned} \quad (20)$$

对 $\mathcal{L}(\mathbf{G}, \mathbf{W})$ 关于矩阵 $\mathbf{G}$ 求导并令其为0:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{G}} = & -2(\mathbf{X}^T \mathbf{W} - \mathbf{G}) + 2\alpha \mathbf{L} \mathbf{S} \mathbf{G} \\ & + 2\beta \mathbf{G} + 2\gamma(\mathbf{G}\mathbf{G}^T - \mathbf{I})\mathbf{G} = 0 \end{aligned} \quad (21)$$

得到关于矩阵 $\mathbf{G}$ 的更新表达式:

$$\mathbf{G} = [(1 + \beta - \gamma)\mathbf{I} + \alpha \mathbf{L} \mathbf{S} + \gamma \mathbf{G}\mathbf{G}^T]^{-1} \mathbf{X}^T \mathbf{W} \quad (22)$$

基于以上分析, FSGS 步骤如算法2所示.

#### 算法 2. FSGS

输入: 数据矩阵 $\mathbf{X} \in \mathbb{R}^{d \times n}$ , 选择的特征数量 $r$ , 正则化参数 $\alpha, \beta$ , 聚类数 $c$ .  
输出: 将所有 $d$ 维特征根据 $\|\mathbf{w}_i\|_2$  ( $i=1,2,\dots,d$ )进行降序排列, 选择数据集 $\mathbf{X}$ 的前 $r$ 个有效特征进行聚类.

- 1) 初始化: 初始化矩阵 $\mathbf{W} \in [0,1]^{d \times c}$ , s.t.  $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ ; 初始化矩阵 $\mathbf{G} \in [0,1]^{n \times c}$ , s.t.  $\mathbf{G}^T \mathbf{G} = \mathbf{I}$ ; 初始化迭代次数 $t=0$ .
- 2) 计算相似度矩阵 $\mathbf{S}$ , 拉普拉斯矩阵 $\mathbf{L}\mathbf{S}$ .
- 3) 循环:
- 4) Step 1: 更新矩阵 $\mathbf{W}$ : 使用算法1或式(18)更新矩阵 $\mathbf{W}$ ;
- 5) Step 2: 更新矩阵 $\mathbf{G}$ :  
 $\mathbf{G}^{t+1} = [(1 + \beta - \gamma)\mathbf{I} + \alpha \mathbf{L} \mathbf{S} + \gamma \mathbf{G}^t (\mathbf{G}^t)^T]^{-1} \mathbf{X}^T \mathbf{W}^t$ ;
- 6) Step 3: 更新迭代次数 $t: t=t+1$ ;
- 7) 收敛或达到最大迭代次数时结束循环.

### 2.3 复杂度分析

根据算法2的步骤,分析每步的时间复杂度.

Step 1 中矩阵 $\mathbf{W}$ 采用 GPI 方法更新的时间复杂度为 $O(d^2n + d^2c)$ .

Step 2 中 $\mathbf{G}^T \mathbf{G}$ 的时间复杂度为 $O(cn^2)$ ,  $\mathbf{X}^T \mathbf{W}$ 的时间复杂度为 $O(cdn)$ , 矩阵求逆的时间复杂度为 $O(n^3)$ , 总的时间复杂度为 $O(cn^2 + cnd + n^3)$ .

综上所述, 算法2第 $t$ 次迭代时整体时间复杂度为 $O(n^3 + nd^2 + cn^2)$ . 在算法执行过程中, 设置最大迭代次数为50次.

## 3 实验及结果分析

实验条件为 Windows 10 系统, 12 GB 内存, Intel(R) Core(TM) i5-7200U @ 2.50 GHz, Python 3.8.

### 3.1 数据集和对照方法

本文实验在9个真实数据集上进行, 包括人脸数据集 (Jaffe<sup>[34]</sup>, Msra25<sup>[35]</sup>, Yale<sup>[13]</sup>, ORL<sup>[13]</sup>), 物品数据集 COIL-20<sup>[36]</sup>、生物数据集 (Lung<sup>[37]</sup>, Glioma<sup>[38]</sup>)、字符数据集 (binary alphabet (BA)<sup>[39]</sup>, USPS<sup>[40]</sup>). 表1展示了这些数据集的详细信息.

为了验证本文方法的精度, 对比如下无监督特征选择方法.

1) Baseline: 作为测试结果的基准线, 不进行特征选择, 采用 K-means 方法对所有样本特征进行聚类.

2) SMR (smooth representation clustering)<sup>[24]</sup>: 具有群组效应的子空间聚类算法, 通过应用图拉普拉斯正

则化约束学习自表示矩阵, 不进行特征选择.

3) LS<sup>[16]</sup>: 通过保留局部流行结构的能力去给每个特征进行评分, 选择分数最高的前 $r$ 个特征作为选择的特征子集.

表1 数据集信息

数据集	样本数	特征数	类别数
Jaffe	213	676	10
Msra25	1799	256	12
Yale	165	1024	15
ORL	400	1024	40
COIL-20	1440	1024	20
Lung	203	3312	5
Glioma	50	4434	4
BA	1404	320	36
USPS	1854	256	10

4) MCFS<sup>[13]</sup>: 采用两步走策略, 首先通过归一化割方法获得聚类指示矩阵 $G$ , 然后应用 $l_1$ 范数约束去学习相关系数矩阵 $W$ , 最后通过计算每个特征的MCFS分数并降序排列选择出值最高的 $r$ 个特征.

5) NDFS<sup>[14]</sup>: 采用联合学习框架, 通过对矩阵 $W$ 施加 $l_{2,1}$ 范数、对矩阵 $G$ 采用非负约束学习数据结构. 最后通过计算矩阵 $W$ 每行的 $l_2$ 范数并进行降序排列选择出值最大的 $r$ 行特征.

6) SOGFS<sup>[26]</sup>: 同时进行特征选择和局部结构学习, 可以自适应学习相似度矩阵. 通过对矩阵 $W$ 施加 $l_{2,p}$ 范数和正交约束学习数据结构. 最后通过计算矩阵 $W$ 每行的 $l_2$ 范数并进行降序排列选择出值最大的 $r$ 行特征.

7) FSDSC<sup>[28]</sup>: 通过正交回归模型学习矩阵 $W$ , 采用离散化的聚类指示矩阵提供清晰的指示信息, 引入对角特征权重矩阵进行特征评估, 最后将选择最大的 $r$ 个对角元素对应的特征用于聚类.

### 3.2 评估方法

在应用特征选择方法产生特征标签后, 再使用 K-means 聚类算法对选择的特征进行聚类. 由于 K-means 算法的聚类结果与初始化有关, 所以对每个方法采用相同的随机种子进行 20 次 K-means 聚类, 然后记录平均结果用于分析.

在所选特征聚类形成簇类标签之后, 将簇类标签使用 Kuhn-Munkres 算法<sup>[41]</sup>与真实标签进行最佳匹配. 对匹配后的标签使用准确率 (accuracy, ACC) 和归一化互信息 (normalized mutual information, NMI) 进行评估. 在这两个指标中, 值越高方法越优.

### 3.3 参数设置

特征选择算法在执行时需要提供选择的特征数量, Jaffe、Msra25、BA 和 USPS 选择 50–200 个特征, 间隔 30; 其他数据集选择 50–300 个特征, 间隔 50. 对于需要调节参数的模型, 参数取值范围设置为 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$ , 其中 UDFS 和 FSGS 调节的参数为 $\alpha$ 和 $\beta$ , SOGFS 和 FSDSC 调节的参数为 $\alpha$ . 本文分别采用全连接法和 $k$ 近邻法计算相似度矩阵参与方法的特征选择过程, 然后选择最优的结果进行分析, 根据之前的方法<sup>[11,14,15]</sup>,  $k$ 设置为 5.

### 3.4 实验结果及分析

图 2 和图 3 分别为 8 种无监督聚类方法在 9 个数据集上选择特征个数与 ACC 和 NMI 的关系.

实验表明了特征选择的可行性和重要性, 大多数特征选择方法在选择少量特征的情况下可以达到甚至超过 K-means 方法的结果. 在大多数数据集的 ACC 和 NMI 结果中, 本文方法都超过了其他方法的结果.

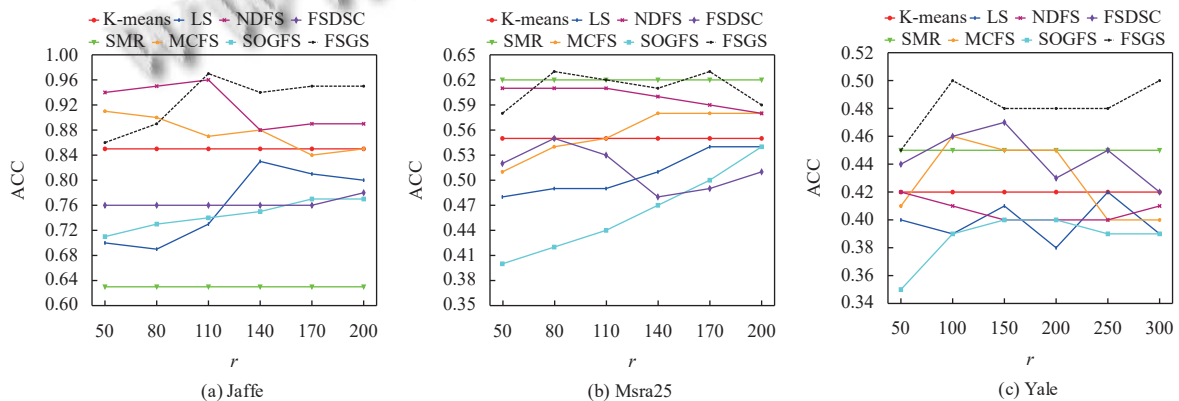


图2 各方法在 9 个数据集上选择不同特征数时的 ACC 结果

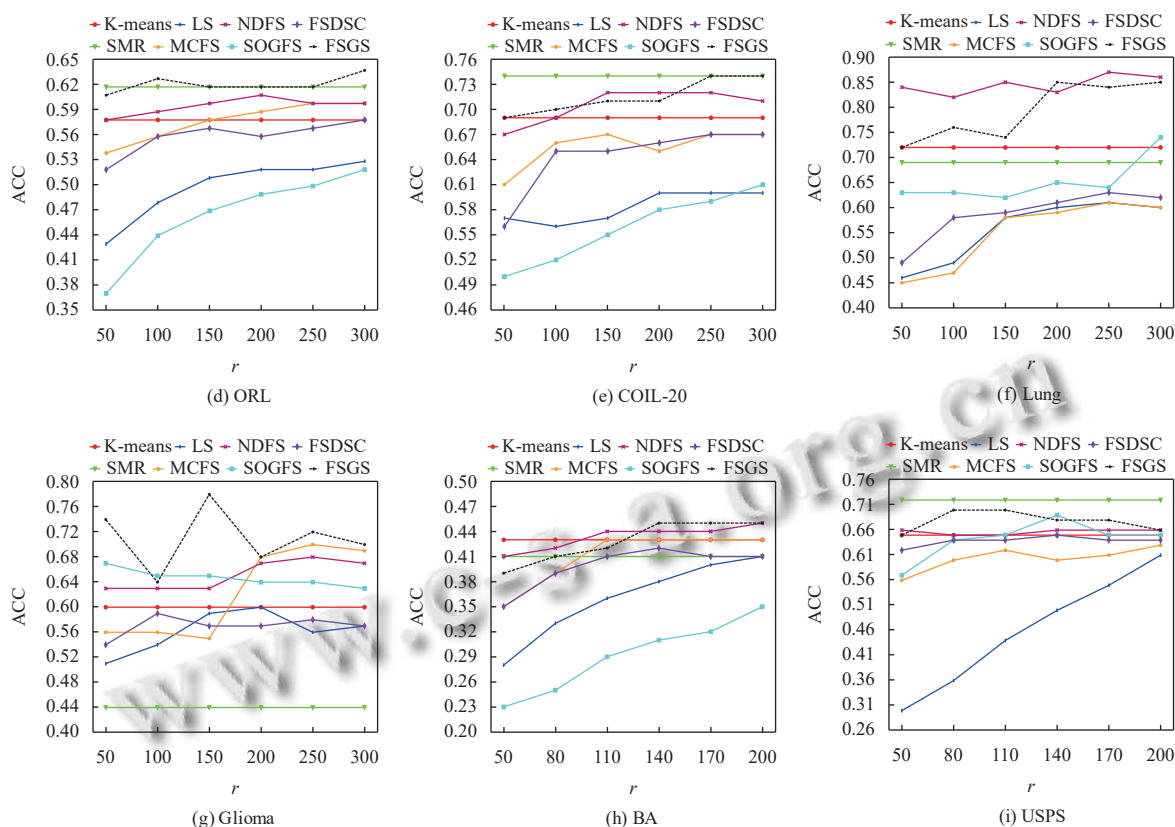


图2 各方法在9个数据集上选择不同特征数时的ACC结果(续)

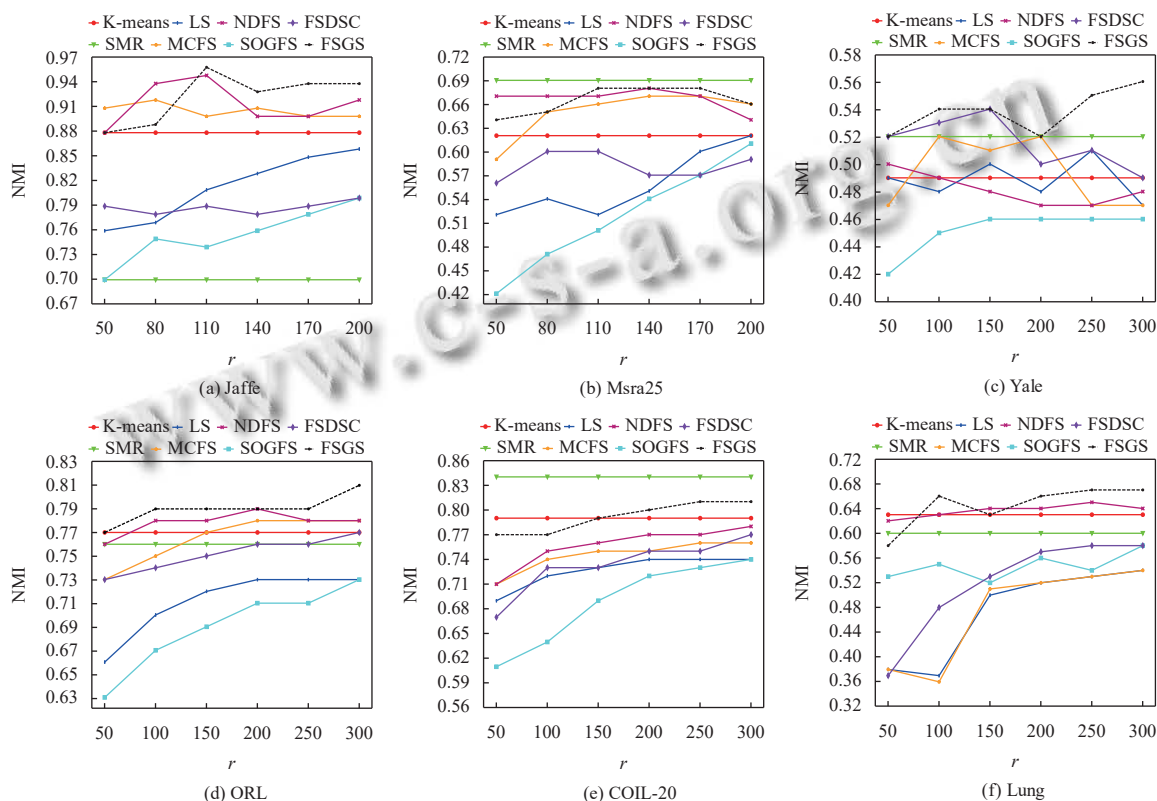


图3 各方法在9个数据集上选择不同特征数时的NMI结果

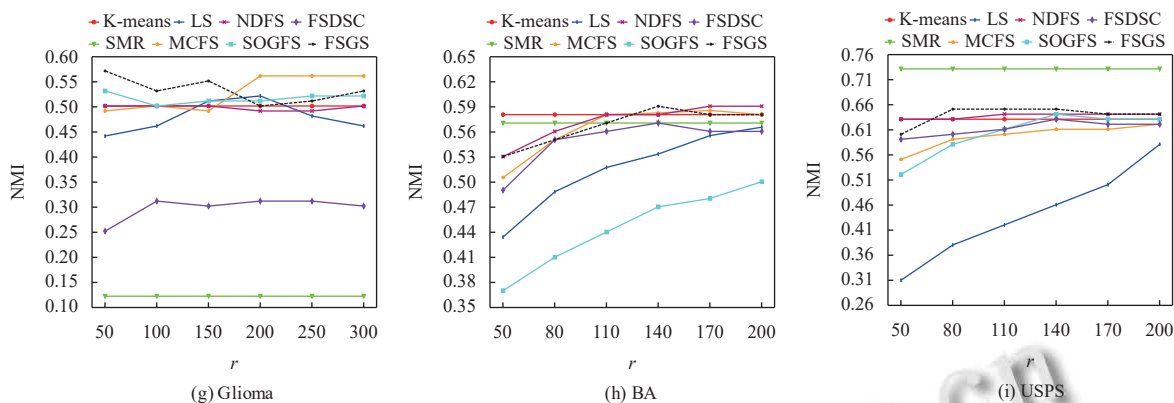


图3 各方法在9个数据集上选择不同特征数时的NMI结果(续)

通过图2和图3可以发现诸如FSGS、UDFS、SOGFS和FSDSC等嵌入式方法性能要明显优于LS和MCFS等过滤式方法,这表明嵌入式方法的优越性能.同时FSGS、UDFS、FSDSC和UGFS等采用的联合学习方法聚类结果要好于MCFS采用的两步走策略.

图2和图3表明具有群组效应的SMR方法在大多数数据集上可以取得较好的效果,但其似乎受数据集影响较大.例如,虽然其在COIL-20和USPS数据集上得到了所有方法中的最好效果,但在Glioma和Jaffe数据集上却取得了最差效果.相反,本文方法同样

考虑了群组效应,且选择少量特征进行聚类,但在所有数据集上都具有稳定的表现,且结果优于或基本等于大多数算法的结果,这表明了本文考虑群组效应的有效性.

通过图2和图3可以发现考虑聚类指示矩阵的NDFS方法和本文方法在大多数数据集中都取得了最好效果,这说明了在特征选择方法中充分考虑聚类指示矩阵的必要性.而NDFS方法对于聚类指示矩阵在迭代过程中全为正时将会失效,相比之下本文方法的泛化能力更强.

图4和图5分别表示FSGS在每个数据集上选择200个特征时ACC与NMI随不同参数 $\alpha$ 和 $\beta$ 变化的结果.

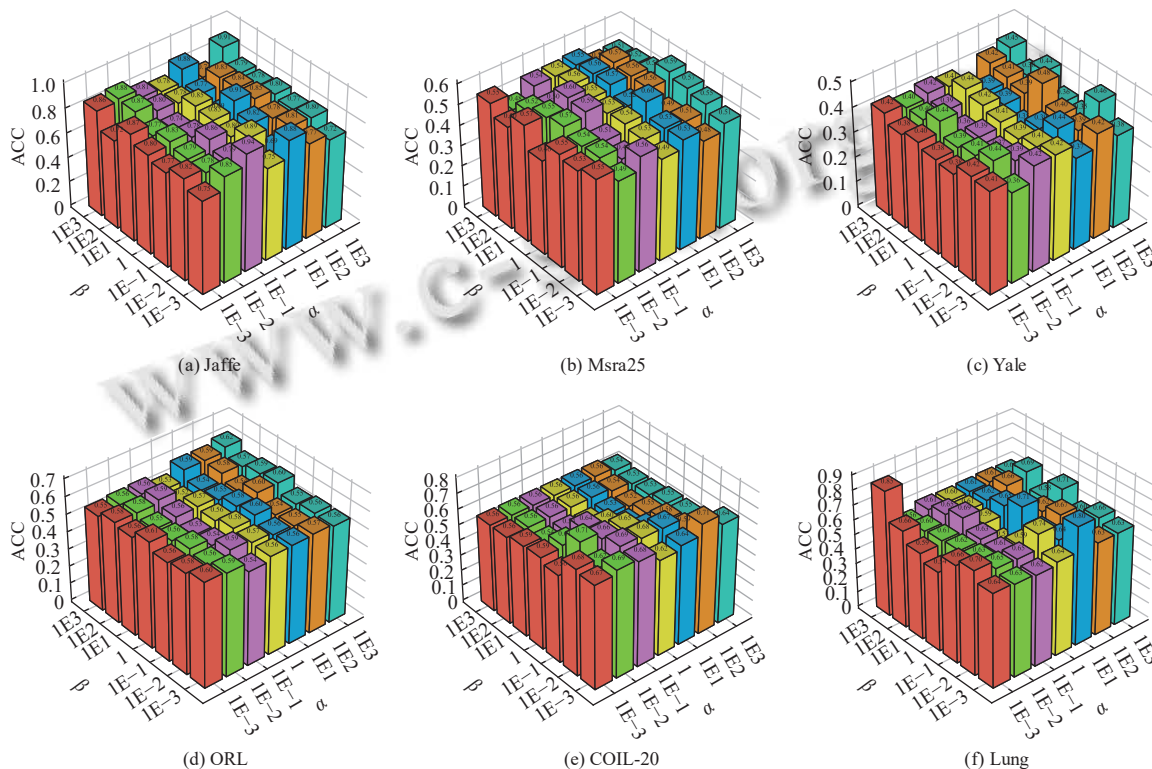


图4 FSGS在9个数据集上选择不同参数 $\alpha$ 和 $\beta$ 时的ACC结果



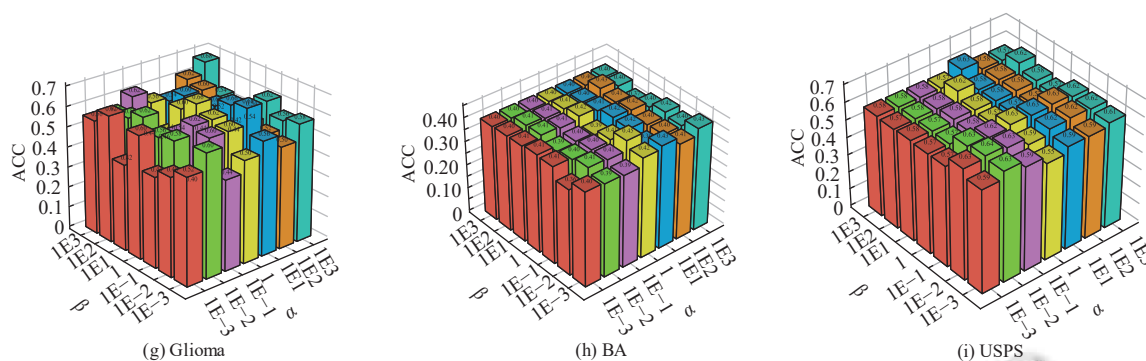


图4 FSGS在9个数据集上选择不同参数 $\alpha$ 和 $\beta$ 时的ACC结果(续)

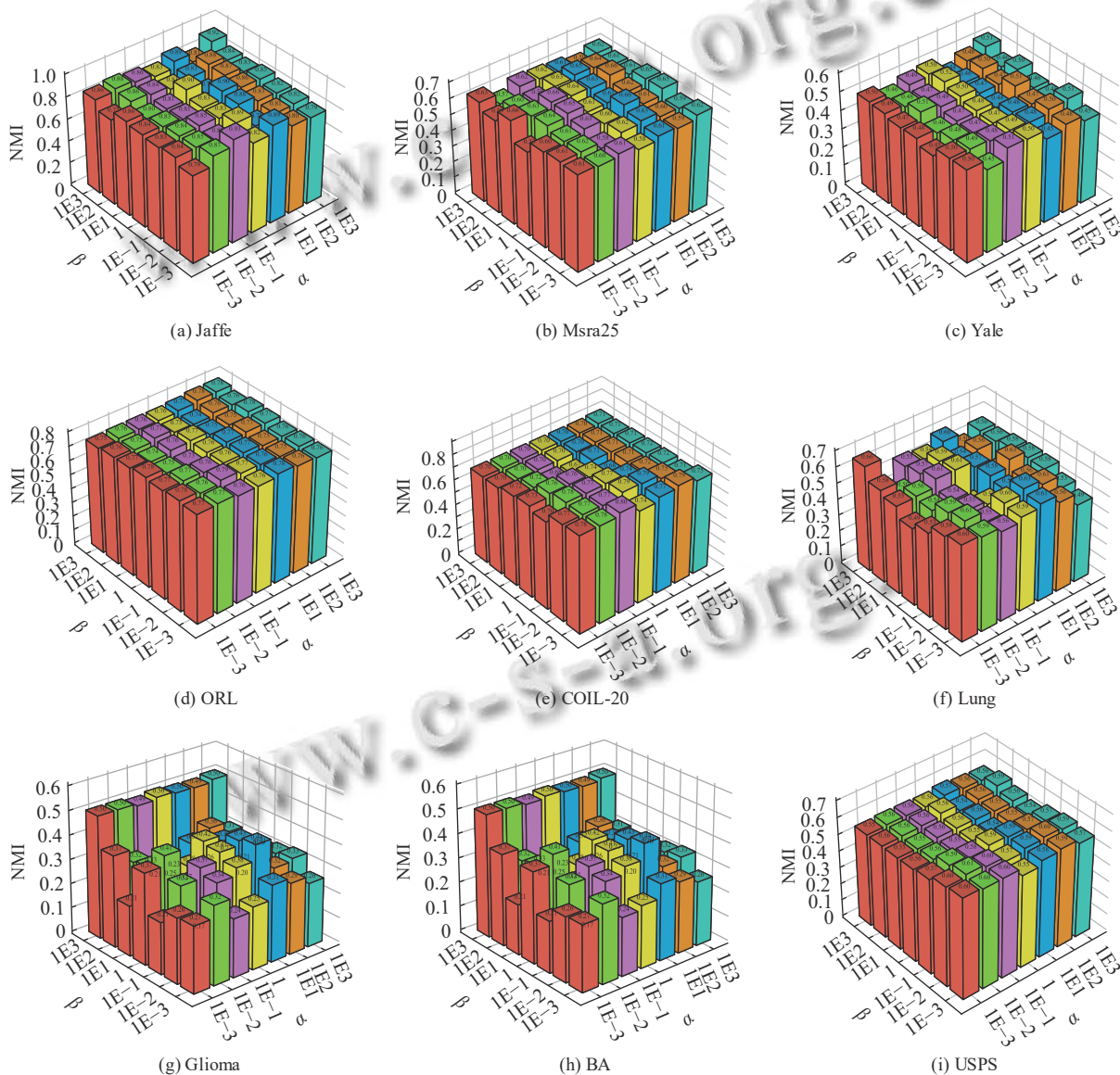


图5 FSGS在9个数据集上选择不同参数 $\alpha$ 和 $\beta$ 时的NMI结果

通过图4和图5可以分析参数 $\alpha$ 和 $\beta$ 对聚类效果的影响. 总的来说, 相比于NMI, ACC似乎更易受到

参数影响, 在大多数数据集中当参数 $\alpha$ 和 $\beta$ 同时取较大值或其中一个取较大值时, 可以获得最佳的聚类效

果. 特别是在数据集 Glioma 的 NMI 评价中, 在  $\beta$  取最大值时, 聚类效果明显优于其他参数下的效果. 这表明结合谱聚类和群组效应可以学习到更具判别性的伪标签信息.

表 2 和表 3 分别表示各个方法在 9 个数据集上最高 ACC 和 NMI 对比, 其中黑体数字表示最高值, 下划线数字表示次高值.

表 2 各方法在 9 个数据集上最高 ACC 值对比 (%)

数据集	Baseline	SMR	LS	MCFS	NDFS	SOGFS	FSDSC	FSGS
Jaffe	85.19	63.38	82.63	91.22	<b>93.90</b>	77.39	77.51	<b>96.71</b>
Msr25	54.71	<b>61.70</b>	54.10	58.44	61.40	53.94	54.67	<b>62.98</b>
Yale	41.58	45.45	41.88	45.76	42.15	39.67	<b>46.64</b>	<b>49.70</b>
ORL	58.46	<b>61.75</b>	52.73	59.95	60.74	51.53	57.64	<b>64.50</b>
COIL-20	68.63	<b>74.24</b>	59.73	67.28	<b>71.81</b>	60.66	67.32	<b>73.82</b>
Lung	71.90	68.97	60.74	61.03	<b>87.49</b>	73.82	62.68	<b>85.22</b>
Glioma	60.40	44.00	58.90	<b>68.50</b>	67.80	66.80	58.80	<b>78.00</b>
BA	42.63	40.95	41.25	43.04	<b>44.66</b>	35.32	41.64	<b>45.37</b>
USPS	64.53	<b>72.82</b>	61.03	63.40	66.13	68.67	64.91	<b>70.06</b>

由表 2 可以看出, 本文方法在 6 个数据集上都取得了最优的聚类结果. 在人脸数据集 Jaffe、Yale 和 ORL 上, FSGS 在聚类准确度上相比于次优结果平均提升 4.67%; 在高维数据集 Lung 和 Glioma 上 FSGS 的聚类结果位于首位或次位; 在多类别数据集 ORL 和 BA 上 FSGS 则取得最高的聚类结果. 这表明了 FSGS 考虑群组效应的鲁棒性.

表 3 各方法在 9 个数据集上最高 NMI 值对比 (%)

数据集	Baseline	SMR	LS	MCFS	NDFS	SOGFS	FSDSC	FSGS
Jaffe	88.12	70.65	85.62	91.55	<b>92.26</b>	80.21	80.42	<b>96.40</b>
Msr25	62.48	<b>68.57</b>	61.97	67.40	66.82	61.12	60.48	<b>68.00</b>
Yale	49.20	52.28	51.16	52.30	49.77	46.18	<b>54.37</b>	<b>56.00</b>
ORL	77.27	77.33	73.03	78.39	<b>78.65</b>	72.63	76.70	<b>80.64</b>
COIL-20	79.01	<b>84.72</b>	74.25	76.41	78.05	74.13	76.66	<b>80.84</b>
Lung	63.73	59.72	53.55	53.97	<b>64.82</b>	57.91	58.32	<b>67.30</b>
Glioma	50.37	18.21	52.11	<b>56.09</b>	49.84	52.70	31.50	<b>56.97</b>
BA	58.16	57.07	56.49	58.52	<b>59.11</b>	49.98	56.53	<b>58.92</b>
USPS	62.73	<b>73.35</b>	58.06	61.83	64.50	64.12	62.71	<b>65.13</b>

表 4 表示各个方法在 9 个数据集上的平均运行时间对比.

表 4 各方法在 9 个数据集上的平均运行时间对比 (s)

数据集	SMR	LS	MCFS	NDFS	SOGFS	FSDSC	FSGS
Jaffe	0.76	1.60	2.44	4.57	59.92	11.16	10.01
Msr25	123.12	15.50	18.92	31.65	228.53	499.45	36.33
Yale	0.67	2.08	3.34	10.74	91.98	32.02	19.65
ORL	6.21	7.91	12.31	17.21	109.69	63.24	41.75
COIL-20	64.07	9.07	14.69	41.61	186.22	534.13	26.77
Lung	2.26	1.20	1.24	155.54	2402.90	214.81	119.41
Glioma	0.81	0.75	1.31	278.08	5371.21	384.72	252.36
BA	55.05	15.88	17.48	22.99	108.75	762.98	26.59
USPS	145.55	14.38	21.68	22.11	189.24	587.66	97.40

由表 4 中数据可以看出, 相比于 SMR、LS 和 MCFS 等非迭代方法, 其他采用迭代策略的方法运行时间明显上升. 在面临高维数据集 Lung 和 Glioma 时, 运行时间达到了最高, 这表明在高维数据集中选择最优组合是具有挑战性的. 其中, 采用的特征值分解法的 SOGFS 在面临高维数据时会造成极大的时间消耗. 本文方法的运行时间基本与 NDFS 方法一致且在可以接受的范围内, 并且本文方法也取得了较好的聚类效果.

综上所述, 本文方法可在不同类型的多个数据集上取得较好的效果, 泛化能力强, 运行效率高. 结果表明, 本文方法在人脸识别、生物医学等领域具有一定的应用价值, 而且相较于其他方法, 本文方法在生物医学等高维数据集上具有最短的运行时间, 因此本文方法是合理有效的无监督特征选择方法.

## 4 结束语

本文提出基于  $F$  范数群组效应和谱聚类的无监督特征选择 (FSGS), 利用群组效应结合谱聚类促使学习更为准确的簇类结构以进行最优的特征子集选择. 为了求解非凸正交约束最优化函数, 将正交约束作为目标函数的一部分并使用矩阵代数方法求解, 通过交替迭代法不断优化变量. 最后通过实验验证本文方法的有效性, 实验表明本文方法在大部分数据集上都取得最佳结果, 相比之前方法得到明显提升. 如何更好地学习聚类指示矩阵指导进行特征选择是今后研究方向之一, 此外, 方法的运行效率也需要考虑使用更好的方法进行优化.

## 参考文献

- Xu YS, Chen S, Li J, *et al.* Fast subspace clustering by learning projective block diagonal representation. *Pattern Recognition*, 2023, 135: 109152. [doi: 10.1016/j.patcog.2022.109152]
- Beiranvand F, Mehrdad V, Dowlatshahi MB. Unsupervised feature selection for image classification: A bipartite matching-based principal component analysis approach. *Knowledge-based Systems*, 2022, 250: 109085. [doi: 10.1016/j.knosys.2022.109085]
- Akhtar Z, Lee JW, Attique Khan M, *et al.* Optical character recognition (OCR) using partial least square (PLS) based feature reduction: An application to artificial intelligence for biometric identification. *Journal of Enterprise Information*

- Management, 2023, 36(3): 767–789. [doi: [10.1108/JEIM-02-2020-0076](https://doi.org/10.1108/JEIM-02-2020-0076)]
- 4 Bojaraj L, Jaikumar R. Hierarchical clustering fuzzy features subset classifier with ant colony optimization for lung image classification. In: Singhal P, Verma A, Srivastava PK, *et al.* eds. Image Processing and Intelligent Computing Systems. Boca Raton: CRC Press, 2023. 49–62.
  - 5 Li D, Liang HN, Qin P, *et al.* A self-training subspace clustering algorithm based on adaptive confidence for gene expression data. *Frontiers in Genetics*, 2023, 14: 1132370. [doi: [10.3389/fgene.2023.1132370](https://doi.org/10.3389/fgene.2023.1132370)]
  - 6 Mousavian Anaraki SA, Haeri A, Moslehi F. A hybrid reciprocal model of PCA and K-means with an innovative approach of considering sub-datasets for the improvement of K-means initialization and step-by-step labeling to create clusters with high interpretability. *Pattern Analysis and Applications*, 2021, 24(3): 1387–1402. [doi: [10.1007/s10044-021-00977-x](https://doi.org/10.1007/s10044-021-00977-x)]
  - 7 Zhu PC, Hou X, Tang KK, *et al.* Unsupervised feature selection through combining graph learning and  $\ell_{2,0}$ -norm constraint. *Information Sciences*, 2023, 622: 68–82. [doi: [10.1016/j.ins.2022.11.156](https://doi.org/10.1016/j.ins.2022.11.156)]
  - 8 Turk MA, Pentland AP. Face recognition using eigenfaces. *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Maui: IEEE, 1991. 586–591.
  - 9 Luo C, Zheng J, Li TR, *et al.* Orthogonally constrained matrix factorization for robust unsupervised feature selection with local preserving. *Information Sciences*, 2022, 586: 662–675. [doi: [10.1016/j.ins.2021.11.068](https://doi.org/10.1016/j.ins.2021.11.068)]
  - 10 Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 2003, 3: 1157–1182.
  - 11 Zhou Y, Zhang WJ, Kang JH, *et al.* A problem-specific non-dominated sorting genetic algorithm for supervised feature selection. *Information Sciences*, 2021, 547: 841–859. [doi: [10.1016/j.ins.2020.08.083](https://doi.org/10.1016/j.ins.2020.08.083)]
  - 12 Chen XJ, Chen RJ, Wu QY, *et al.* Semisupervised feature selection via structured manifold learning. *IEEE Transactions on Cybernetics*, 2022, 52(7): 5756–5766. [doi: [10.1109/TCYB.2021.3052847](https://doi.org/10.1109/TCYB.2021.3052847)]
  - 13 Cai D, Zhang CY, He XF. Unsupervised feature selection for multi-cluster data. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington: ACM, 2010. 333–342.
  - 14 Li ZC, Yang Y, Liu J, *et al.* Unsupervised feature selection using nonnegative spectral analysis. *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. Toronto: AAAI, 2012. 1026–1032.
  - 15 Wang R, Zhang CY, Bian JT, *et al.* Sparse and flexible projections for unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(6): 6362–6375.
  - 16 He XF, Cai D, Niyogi P. Laplacian score for feature selection. *Proceedings of the 18th International Conference on Neural Information Processing Systems*. Vancouver: MIT Press, 2005. 507–514.
  - 17 Bommert A, Welchowski T, Schmid M, *et al.* Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. *Briefings in Bioinformatics*, 2022, 23(1): bbab354. [doi: [10.1093/bib/bbab354](https://doi.org/10.1093/bib/bbab354)]
  - 18 孙雅芝, 江峰, 杨志勇. 基于粒度粗糙熵与改进蜂群算法的特征选择. *计算机系统应用*, 2023, 32(6): 121–129. [doi: [10.15888/j.cnki.csa.009108](https://doi.org/10.15888/j.cnki.csa.009108)]
  - 19 Fang XZ, Xu Y, Li XL, *et al.* Locality and similarity preserving embedding for feature selection. *Neurocomputing*, 2014, 128: 304–315. [doi: [10.1016/j.neucom.2013.08.040](https://doi.org/10.1016/j.neucom.2013.08.040)]
  - 20 Zhang R, Zhang YX, Li XL. Unsupervised feature selection via adaptive graph learning and constraint. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(3): 1355–1362. [doi: [10.1109/TNNLS.2020.3042330](https://doi.org/10.1109/TNNLS.2020.3042330)]
  - 21 Luo MN, Nie FP, Chang XJ, *et al.* Adaptive unsupervised feature selection with structure regularization. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(4): 944–956. [doi: [10.1109/TNNLS.2017.2650978](https://doi.org/10.1109/TNNLS.2017.2650978)]
  - 22 Zhang H, Zhang R, Nie FP, *et al.* An efficient framework for unsupervised feature selection. *Neurocomputing*, 2019, 366: 194–207. [doi: [10.1016/j.neucom.2019.07.020](https://doi.org/10.1016/j.neucom.2019.07.020)]
  - 23 Lu CY, Min H, Zhao ZQ, *et al.* Robust and efficient subspace segmentation via least squares regression. *Proceedings of the 12th European Conference on Computer Vision*. Florence: Springer, 2012. 347–360.
  - 24 Hu H, Lin ZC, Feng JJ, *et al.* Smooth representation clustering. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014. 3834–3841.
  - 25 Wang P, Han B, Li J, *et al.* Structural reweight sparse subspace clustering. *Neural Processing Letters*, 2019, 49(3): 965–977. [doi: [10.1007/s11063-018-9859-8](https://doi.org/10.1007/s11063-018-9859-8)]
  - 26 Nie FP, Zhu W, Li XL. Structured graph optimization for unsupervised feature selection. *IEEE Transactions on*

- Knowledge and Data Engineering, 2019, 33(3): 1210–1222.
- 27 Du XZ, Nie FP, Wang WQ, *et al.* Exploiting combination effect for unsupervised feature selection by  $\ell_{2,0}$  norm. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(1): 201–214. [doi: [10.1109/TNNLS.2018.2837100](https://doi.org/10.1109/TNNLS.2018.2837100)]
- 28 Shang RH, Kong JR, Wang LJ, *et al.* Unsupervised feature selection via discrete spectral clustering and feature weights. Neurocomputing, 2023, 517: 106–117. [doi: [10.1016/j.neucom.2022.10.053](https://doi.org/10.1016/j.neucom.2022.10.053)]
- 29 Yang ZG, Wang JK, Li Q, *et al.* Graph optimization for unsupervised dimensionality reduction with probabilistic neighbors. Applied Intelligence, 2023, 53(2): 2348–2361. [doi: [10.1007/s10489-022-03534-z](https://doi.org/10.1007/s10489-022-03534-z)]
- 30 Nie FP, Wang XQ, Huang H. Clustering and projected clustering with adaptive neighbors. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014. 977–986.
- 31 Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Vancouver: MIT Press, 2001. 585–591.
- 32 Zhu PF, Zhu WC, Hu QH, *et al.* Subspace clustering guided unsupervised feature selection. Pattern Recognition, 2017, 66: 364–374. [doi: [10.1016/j.patcog.2017.01.016](https://doi.org/10.1016/j.patcog.2017.01.016)]
- 33 Nie FP, Zhang R, Li XL. A generalized power iteration method for solving quadratic problem on the stiefel manifold. Science China Information Sciences, 2017, 60(11): 112101. [doi: [10.1007/s11432-016-9021-9](https://doi.org/10.1007/s11432-016-9021-9)]
- 34 Lyons MJ, Budynek J, Akamatsu S. Automatic classification of single facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(12): 1357–1362. [doi: [10.1109/34.817413](https://doi.org/10.1109/34.817413)]
- 35 He XF, Yan SC, Hu YX, *et al.* Learning a locality preserving subspace for visual recognition. Proceedings of the 9th IEEE International Conference on Computer Vision. Nice: IEEE, 2003. 385–392.
- 36 Nene SA, Nayar SK, Murase H. Columbia object image library (COIL-20). Technical Report, CUCS-006-96, New York: Columbia University, 1996.
- 37 Singh D, Febbo PG, Ross K, *et al.* Gene expression correlates of clinical prostate cancer behavior. Cancer Cell, 2002, 1(2): 203–209. [doi: [10.1016/S1535-6108\(02\)00030-2](https://doi.org/10.1016/S1535-6108(02)00030-2)]
- 38 Li JD, Cheng KW, Wang SH, *et al.* Feature selection: A data perspective. ACM Computing Surveys, 2017, 50(6): 94.
- 39 Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711–720.
- 40 Hull JJ. A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(5): 550–554. [doi: [10.1109/34.291440](https://doi.org/10.1109/34.291440)]
- 41 Kuhn HW. The Hungarian method for the assignment problem. Naval Research Logistics Quarterly, 1955, 2(1–2): 83–97. [doi: [10.1002/nav.3800020109](https://doi.org/10.1002/nav.3800020109)]

(校对责编: 张重毅)