

# 联合低秩重构和投影重构的稳健特征选择方法

仪双燕<sup>1</sup>, 梁永生<sup>2</sup>, 陆晶晶<sup>3</sup>, 柳伟<sup>4</sup>, 胡涛<sup>5</sup>, 何震宇<sup>6</sup>

(1. 深圳信息职业技术学院软件学院, 广东 深圳 518000; 2. 哈尔滨工业大学(深圳)电子与信息工程学院, 广东 深圳 518000;  
3. 深圳国微福芯技术有限公司, 广东 深圳 518000; 4. 深圳信息职业技术学院计算机学院, 广东 深圳 518000;  
5. 深圳信息职业技术学院信息技术研究所, 广东 深圳 518000; 6. 哈尔滨工业大学(深圳)计算机科学与技术学院, 广东 深圳 518000)

**摘 要:** 针对当前特征选择方法依然受噪声影响以及无法将聚类效果和重构效果有效统一的问题, 提出了一种稳健的特征选择方法。从干净数据和重构数据作差的思路着手, 将低秩重构数据和投影重构数据作差构建稳健的重构误差项, 并提出从学习到的干净数据上选择特征用于聚类。将干净数据的学习和特征选择技能进行联合学习, 相互促进, 从而提升方法在有噪数据上的稳健性, 并且将重构效果和聚类效果进行有效统一。在 5 个数据集上与几种图嵌入角度的特征选择以及 PCA 重构角度的特征选择方法进行聚类实验对比, 实验结果表明, 除 LUNG 噪声数据集外, 所提方法在 2 种评价指标 (ACC 和 NMI) 下都优于对比特征选择方法。

**关键词:** 重构; 低秩; 投影; 稀疏; 特征选择

中图分类号: TN92

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2023061

## Robust feature selection method via joint low-rank reconstruction and projection reconstruction

YI Shuangyan<sup>1</sup>, LIANG Yongsheng<sup>2</sup>, LU Jingjing<sup>3</sup>, LIU Wei<sup>4</sup>, HU Tao<sup>5</sup>, HE Zhenyu<sup>6</sup>

1. School of Software Engineering, Shenzhen Institute of Information Technology, Shenzhen 518000, China  
2. School of Electronics and Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518000, China  
3. Shenzhen Guowei Fuxin Technology Co., Ltd., Shenzhen 518000, China  
4. School of Computer Sciences, Shenzhen Institute of Information Technology, Shenzhen 518000, China  
5. Institute of Information Technology, Shenzhen Institute of Information Technology, Shenzhen 518000, China  
6. School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518000, China

**Abstract:** Aiming at the problem that current feature selection methods were still affected by noise and cannot effectively unify clustering and reconstruction effects, a robust feature selection method was proposed. A robust reconstruction error term was built by making the difference between low-rank reconstruction and projection reconstruction. After that, the features for clustering were selected from the reconstructed data instead of the original data. The learning of clean data and feature selection technique are allowed for joint learning and promote each other, thereby improving the robustness of the method on noisy data, and effectively unifying reconstruction and clustering. Compared with several kinds of graph embedding feature selection and reconstruction feature selection methods on five datasets, the experimental results showed that, except for the LUNG noise dataset, the proposed method outperforms the comparative feature selection method under both evaluation indicators (ACC and NMI).

**Keywords:** reconstruction, low-rank, projection, sparsity, feature selection

收稿日期: 2022-11-06; 修回日期: 2023-02-02

通信作者: 梁永生, liangys@sziit.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61906124, No.62031013); 中国博士后科学基金资助项目 (No.2018M630158); 广东省自然科学基金资助项目 (No.2022A1515011447)

**Foundation Items:** The National Natural Science Foundation of China (No.61906124, No.62031013), The Chinese Postdoctoral Science Foundation (No.2018M630158), Guangdong Natural Science Foundation (No.2022A1515011447)

## 0 引言

特征选择是数据处理中的一个重要任务。在现实场景中,干净且有价值的特征通常是有限的,为了挖掘这部分重要的特征子集来进行下一步的学习,就需要对数据中存在的冗余和噪声进行处理。

根据数据标签在算法的目标函数中是否被使用,可以将特征选择方法分为有监督、半监督、无监督的特征选择方法。一般而言,有监督特征选择方法选择出的特征更具可分性,但是需要标签支撑。然而,标签并不容易获取,标签质量也会影响算法性能。因此,无监督特征选择方法因其简单有效而显得尤其重要。

早期的无监督特征选择方法<sup>[1-4]</sup>按照特定的评分准则给单个特征的重要程度打分,再根据需求选出分数最高的特征子集。通过这种方式选出的特征具有很强的个体代表性,但是如果数据的相关性很强,通过分数的高低选择出来的特征仍存在冗余,而且这种方式具有片面性,完全忽略了特征的组合效应,即一些得分很低的特征和得分较高的特征组合起来可能有很好的代表性。为了降低数据的离散性对特征选择造成不确定性的影响,文献[5-9]致力于探索联合特征选择。通用的做法是通过正则项对投影矩阵进行 $\ell_{2,1}$ 范数稀疏约束,强制投影矩阵的载荷向量<sup>[10]</sup>的元素值一致性接近0,保留下来的就是区分度较高的特征的组合系数(载荷向量)。联合特征选择是综合考虑所有特征相关性之后选择出的有效特征,而且能通过组合系数的大小反映特征的重要程度,具有较好的特征解释性。因此,联合特征选择备受关注。

联合特征选择功能可以嵌入不同的特征抽取算法中而形成不同的分支,其中一个分支是基于图嵌入的无监督特征选择方法<sup>[3,6,11-18]</sup>。这种方法借助数据的流形结构来找到一些局部保持能力强的特征子集。例如,无监督特征选择(UDFS)方法<sup>[13]</sup>考虑数据的流形结构,认为高维特征应该分布在某个低维流形附近,所以利用原始高维数据的局部判别信息和特征相关性进行无监督特征选择,在正则项引入 $\ell_{2,1}$ 范数约束联合评估特征的重要性。虽然UDFS能够保证选择出的联合特征维持了原始数据结构,但是这种方法高度依赖原始数据 $k$ 近邻的划分,而噪声的干扰会破坏数据的局部流形结构<sup>[15-16]</sup>,这时UDFS会找到一个不理想的邻域关系,进而影响特征选择的效果。结构化

最优图特征选择(SOGFS)方法<sup>[17]</sup>通过构造动态图的方式来更新邻域关系,以保证数据的最优局部结构。然而,当原始数据含有较严重的噪声时,SOGFS也不能准确构建最优的图结构。综上可知,特征选择的可靠性主要受噪声的影响。为此,将联合特征选择功能嵌入主成分分析(PCA, principal component analysis)重构方法<sup>[19]</sup>,即基于PCA重构角度的特征选择方法<sup>[20-22]</sup>相继被提出。

基于PCA重构角度的特征选择方法旨在使用不同的稀疏约束来提升模型对噪声的稳健性,但是由于重构目标是原始数据,当原始数据含有噪声时,使用原始数据和重构后的数据作差构建的重构误差项会使模型学习到不恰当的稀疏投影矩阵,进而影响特征选择的准确性。此外,当前的特征选择方法通常从原始数据上选择特征用于聚类,导致模型忽略重构效果而只追求聚类效果。对于重构角度的特征选择方法而言,重构和聚类效果的有效统一是非常必要的。

为了在噪声环境下学习一个比较准确的投影矩阵,需要构建更加准确的重构误差项,来确保特征选择模型的稳健性和准确性。为此,本文提出了联合低秩重构和投影重构的稳健特征选择方法,从干净数据和重构数据作差的思路着手,采用低秩PCA方法中的低秩重构项作为原始数据的干净数据,再以这些干净数据而不是原始数据为基准和投影重构项作差,这样设计的重构误差项具有较好的抗噪性,从干净数据上选择的特征得到的聚类效果更好。本文具体贡献如下。

1) 提出将低秩重构项和投影重构项作差的方式构建重构误差项。低秩重构项是通过低秩PCA思想学习的干净数据,投影重构项是通过投影PCA思想学习的干净数据,将两者作差构建的重构误差项使所提特征选择方法可以适用于有噪声的数据集。

2) 已有的特征选择方法根据投影矩阵 $\mathbf{Q}$ 从原始数据 $\mathbf{X}$ 上选择特征后用于聚类,导致忽略重构效果而只追求聚类效果。本文根据学习到的投影矩阵 $\mathbf{Q}$ 从干净的数据 $\mathbf{L}$ 上选择特征后用于聚类,不仅重构效果好,而且聚类效果好,从而将重构和聚类进行有效统一。

## 1 预备知识

本节首先介绍一些符号的意义,然后简单介绍投影重构PCA方法、低秩重构PCA方法以及自动加权稀疏PCA(AW-SPCA)方法,以便于理解本

文提出的模型。

给定数据矩阵  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ ，其中， $\mathbf{x}_j \in \mathbb{R}^m$  为样本， $n$  为样本的数量， $m$  为特征个数。设  $x_{ij}$  为  $\mathbf{X}$  中第  $i$  行第  $j$  列的元素，并定义  $\mathbf{X}$  的  $\ell_1$  范数和  $\ell_{2,1}$  范数分别为  $\|\mathbf{X}\|_1 = \sum_{i=1}^m \sum_{j=1}^n |x_{ij}|$  和  $\|\mathbf{X}\|_{2,1} = \sum_{j=1}^n \|\mathbf{x}_j\|_2$ 。

PCA 重构方法可分为投影重构 PCA<sup>[20-22]</sup>和低秩重构 PCA<sup>[23-27]</sup>。投影重构 PCA 方法<sup>[19]</sup>旨在学习高维数据的低维子空间，使原始数据和低维数据之间的能量损失最小，即需要学习一个使重构误差最小的投影矩阵  $\mathbf{P}$ ，通过投影降低原始数据的维度。具体模型为

$$\begin{aligned} \min_{\mathbf{P}} & \|\mathbf{X} - \mathbf{P}\mathbf{P}^T\mathbf{X}\|_F^2 \\ \text{s.t.} & \mathbf{P}^T\mathbf{P} = \mathbf{I} \end{aligned} \quad (1)$$

其中，辅助投影矩阵  $\mathbf{P} \in \mathbb{R}^{m \times d}$  为列正交投影矩阵，投影后的数据为  $\mathbf{P}^T\mathbf{X} \in \mathbb{R}^{d \times n}$ ， $\mathbf{I}$  为单位矩阵。该方法能够处理一定程度的高斯噪声，但在其他噪声下失效。为了提升投影重构 PCA 方法在其他噪声下的稳健性，基于  $\ell_1$  范数、 $\ell_{2,1}$  范数和  $\ell_{2,p}$  范数的投影重构 PCA 方法<sup>[28-35]</sup>被提出。

低秩重构 PCA 方法<sup>[23-27,36]</sup>旨在通过寻求数据低秩结构来提升对离群点等噪声的稳健性，可以直接学习具有低秩特性的干净数据。具体地，文献[23]提出将原始数据划分为干净数据和噪声数据两部分，具体模型为

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{E}} & \|\mathbf{L}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} & \mathbf{X} = \mathbf{L} + \mathbf{E} \end{aligned} \quad (2)$$

其中， $\mathbf{L} \in \mathbb{R}^{m \times n}$  为低秩重构后的干净数据矩阵； $\mathbf{E} \in \mathbb{R}^{m \times n}$  为稀疏噪声矩阵；正则参数  $\lambda > 0$ ，用于平衡  $\|\mathbf{L}\|_*$  和  $\|\mathbf{E}\|_1$ 。

上述 PCA 方法都不具有明确的特征解释性。由于投影重构 PCA 方法可以借助稀疏正则项使之具有特征选择能力，因此双投影矩阵的稀疏 PCA 方法<sup>[20-21]</sup>一度成为研究热点。基于  $\ell_1$  范数的稀疏 PCA 方法<sup>[20]</sup>从原理上阐述了特征选择项如何被添加进投影重构方法，使之具有初步特征选择能力。基于  $\ell_{2,1}$  范数的稀疏 PCA 方法<sup>[21]</sup>通过对载荷向量施加块稀疏约束，相当于使载荷向量列一致稀疏，实现特征选择的功能。AW-SPCA<sup>[22]</sup>使用  $\ell_{2,1}$  范数同时约束重构误差矩阵和投影矩阵，使模型可以通过自适应权重的大小抑制噪声对重构误差的影响，进而

学习更准确的投影矩阵，具体模型为

$$\begin{aligned} \min_{\mathbf{P}, \mathbf{Q}} & \|\mathbf{X} - \mathbf{P}\mathbf{Q}\mathbf{X}\|_{2,1} + \lambda \|\mathbf{Q}\|_{2,1} \\ \text{s.t.} & \mathbf{P}^T\mathbf{P} = \mathbf{I} \end{aligned} \quad (3)$$

其中， $\mathbf{Q} \in \mathbb{R}^{m \times d}$  为投影矩阵。双投影矩阵可以有效缓解正交性与稀疏性的不兼容问题。

然而，AW-SPCA 模型的不足之处在于投影重构项  $\mathbf{P}\mathbf{Q}\mathbf{X}$  的重构目标是原始数据  $\mathbf{X}$ ，当原始数据  $\mathbf{X}$  含有噪声时，会影响投影矩阵的准确性。为此，需要设计一种对噪声更加稳健的重构误差项，以学习到更加准确的稀疏投影矩阵。

## 2 方法

目前，大多基于 PCA 重构角度的特征选择方法所涉及的重构误差是将原始数据  $\mathbf{X}$  和重构后的数据  $\mathbf{X}^{\text{rec}}$  作差，然后进行  $\ell_{2,1}$  范数约束得到  $\|\mathbf{X} - \mathbf{X}^{\text{rec}}\|_{2,1}$ 。这种重构误差项的设计针对相对干净的数据而言比较有效。然而，原始数据往往并不是干净数据，通常含有噪声。此时，在原始数据有噪声干扰的情况下，将原始数据和重构后的数据作差构建的重构误差项并不能较好地适用于有噪声的数据集。

具体地，采用  $\ell_{2,1}$  范数约束的重构误差项  $\|\mathbf{X} - \mathbf{X}^{\text{rec}}\|_{2,1}$  来提升特征选择模型对噪声的稳健性。换句话说，模型希望重构误差项  $\|\mathbf{X} - \mathbf{X}^{\text{rec}}\|_{2,1}$  的值尽可能小，或者重构后的数据  $\mathbf{X}^{\text{rec}}$  是干净数据。但是当原始数据  $\mathbf{X}$  包含离群点等噪声时，不失一般性，本文使用  $\mathbf{X}^{\text{noise}}$  来表示离群点噪声的数据，那么显然对应离群点噪声的数据的重构误差值  $\|\mathbf{X}^{\text{noise}} - (\mathbf{X}^{\text{noise}})^{\text{rec}}\|_{2,1}$  反而很大。因此，将原始数据和重构后的数据作差构建的重构误差项  $\|\mathbf{X} - \mathbf{X}^{\text{rec}}\|_{2,1}$  不能很好地适用离群点噪声数据。

基于上述分析，本文从重构误差项的构造着手，首先从被噪声污染的原始数据  $\mathbf{X}$  中找到隐藏的干净数据，再以这部分干净数据作为重构目标的基准来设计重构误差项。将低秩重构项和投影重构项作差，一方面保证了重构误差项的有效性，另一方面可以通过投影重构项引入特征选择能力。具体模型为

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{E}, \mathbf{P}, \mathbf{Q}} & \|\mathbf{L}\|_* + \alpha \|\mathbf{E}\|_1 + \\ & \gamma (\|\mathbf{Q}\|_{2,1} + \beta \|\mathbf{L} - \mathbf{P}\mathbf{Q}\mathbf{X}\|_1) \\ \text{s.t.} & \mathbf{X} = \mathbf{L} + \mathbf{E}, \mathbf{P}^T\mathbf{P} = \mathbf{I} \end{aligned} \quad (4)$$

其中,  $L \in \mathbb{R}^{m \times n}$  为低秩重构后的干净数据矩阵,  $E \in \mathbb{R}^{m \times n}$  为稀疏噪声矩阵,  $P \in \mathbb{R}^{m \times d}$  和  $Q \in \mathbb{R}^{d \times m}$  分别为辅助投影矩阵和投影矩阵, 投影后的数据为  $Q^T X \in \mathbb{R}^{d \times n}$ ,  $\alpha > 0$  用来控制  $E$  的稀疏程度,  $\beta > 0$  用来控制重构误差项的稀疏程度,  $\gamma > 0$  用来平衡  $\|L\|_* + \alpha \|E\|_1$  和  $\|Q\|_{2,1} + \beta \|L - PQX\|_1$  的重要程度。

所提方法的核心部分为重构误差项  $\|L - PQX\|_1$ , 该项将低秩重构项和投影重构项完美连接。可以看到式(4)所示模型的目标函数中前两项的作用在于通过低秩重构项找到干净数据  $L$ , 后两项则是通过投影重构项在干净数据上找到鉴别特征。本文认为鉴别特征就是原始数据中那些可以最大程度还原干净数据  $L$  的特征子集。因为  $Q$  的每一列代表不同特征的权重向量, 所以通过列稀疏可以选择一致特征, 然后通过最小化低秩重构数据  $L$  与投影重构数据  $PQX$  的误差来保证选取可靠的一致特征。

### 3 优化

#### 3.1 优化过程

式(4)所示模型包含了  $\ell_{2,1}$  范数和 4 个不同的变量, 难以直接处理, 所以本文采用固定其他变量来优化一个变量的迭代算法进行求解。

首先, 将式(4)所示模型转化为

$$\begin{aligned} \min_{L, E, P, Q, W} & \|L\|_* + \alpha \|E\|_1 + \gamma (\|Q\|_{2,1} + \beta \|W\|_1) \\ \text{s.t.} & X = L + E, P^T P = I, W = L - PQX \end{aligned} \quad (5)$$

然后, 将式(5)转化为不含约束条件的拉格朗日函数, 即

$$\begin{aligned} \Gamma(L, E, P, Q, W) &= \|L\|_* + \alpha \|E\|_1 + \\ & \gamma (\|Q\|_{2,1} + \beta \|W\|_1) + \text{tr}(Y_1^T (X - L - E)) + \\ & \text{tr}(Y_2^T (W - L + PQX)) + \\ & \frac{\mu}{2} (\|X - L - E\|_F^2 + \|W - L + PQX\|_F^2) \\ \text{s.t.} & P^T P = I \end{aligned} \quad (6)$$

式(6)的求解可分为以下几个步骤。

**步骤 1** 固定其他变量, 优化  $L$ , 此时式(6)变为

$$\begin{aligned} L^* &= \arg \min_L \frac{1}{2\mu} \|L\|_* + \\ & \frac{1}{2} \left\| L - \frac{X - E + W + PQX}{2} - \frac{Y_1 + Y_2}{2\mu} \right\|_F^2 \end{aligned} \quad (7)$$

式(7)的解可以由奇异值阈值收缩算子得到, 即

$$L^* = \Theta_{\frac{1}{2\mu}} \left( \frac{X - E + W + PQX}{2} + \frac{Y_1 + Y_2}{2\mu} \right) \quad (8)$$

**步骤 2** 固定其他变量, 优化  $E$ , 此时式(6)变为

$$E^* = \arg \min_E \frac{\alpha}{\mu} \|E\|_1 + \frac{1}{2} \left\| E - \left( X - L + \frac{Y_1}{\mu} \right) \right\|_F^2 \quad (9)$$

式(9)具有闭式解, 其解为

$$\begin{aligned} E^* &= \max \left( 0, X - L + \frac{Y_1}{\mu} - \frac{\alpha}{\mu} \right) + \\ & \min \left( 0, X - L + \frac{Y_1}{\mu} + \frac{\alpha}{\mu} \right) \end{aligned} \quad (10)$$

**步骤 3** 固定其他变量, 优化  $P$ , 此时式(6)变为

$$\begin{aligned} P^* &= \arg \min_P \frac{\mu}{2} \left\| PQX - \left( L - W - \frac{Y_2}{\mu} \right) \right\|_F^2 \\ \text{s.t.} & P^T P = I \end{aligned} \quad (11)$$

式(11)是一个正交普鲁克问题, 其解为

$$P^* = VU^T \quad (12)$$

其中,  $U$  和  $V$  分别是对  $QX \left( L - W - \frac{Y_2}{\mu} \right)^T$  做奇异值分解得到的左奇异矩阵和右奇异矩阵。

**步骤 4** 固定其他变量, 优化  $Q$ , 此时式(6)变为

$$\begin{aligned} Q^* &= \arg \min_Q \gamma \|Q\|_{2,1} + \\ & \frac{\mu}{2} \left\| W - L + PQX + \frac{Y_2}{\mu} \right\|_F^2 \end{aligned} \quad (13)$$

对  $Q$  求偏导可得  $\mu P^T \left( L - W - \frac{Y_2}{\mu} \right) = \mu QXX^T + 2\gamma QG$ 。其中,  $G = \text{diag} \left( \frac{1}{2\|q_1\|_2}, \frac{1}{2\|q_2\|_2}, \dots, \frac{1}{2\|q_m\|_2} \right)$ ,  $q_i$  是  $Q$  的列向量。因此,

$$Q^* = \mu P^T \left( L - W - \frac{Y_2}{\mu} \right) X^T (2\gamma G + \mu XX^T)^{-1} \quad (14)$$

**步骤 5** 固定其他变量, 优化  $W$ , 此时式(6)变为

$$\begin{aligned} W^* &= \arg \min_W \frac{\gamma\beta}{\mu} \|W\|_1 + \\ & \frac{1}{2} \left\| W - \left( L - PQX - \frac{Y_2}{\mu} \right) \right\|_F^2 \end{aligned} \quad (15)$$

式(15)的解为

$$\begin{aligned} W^* &= \max\left(0, L - PQX - \frac{Y_2}{\mu} - \frac{\gamma\beta}{\mu}\right) + \\ &\min\left(0, L - PQX - \frac{Y_2}{\mu} + \frac{\gamma\beta}{\mu}\right) \end{aligned} \quad (16)$$

步骤6 更新  $Y_1$ ,  $Y_2$  和  $\mu$ , 即

$$\begin{aligned} Y_1 &= Y_1 + \mu(X - L - E) \\ Y_2 &= Y_2 + \mu(W - L + PQX) \\ \mu &= \min(\rho\mu, \mu_{\max}) \end{aligned} \quad (17)$$

详细算法如算法1所示。

算法1 式(4)所示模型的优化算法

输入 标准化后的数据矩阵  $X$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,

降维后的维度  $d=100$

初始化  $\mu_0 = 0.1$ ,  $\rho = 1.1$ ,  $\mu_{\max} = 10^6$ ,

$P^{m \times d} = Q^T = I^{m \times d}$ ,  $E = W = Y_1 = Y_2 = \mathbf{0}$ ,  $\varepsilon = 10^{-4}$

当算法不收敛时执行以下步骤

1) 更新  $L$ ,  $L^* = \Theta_{\frac{1}{2\mu}}\left(\frac{X - E + W + PQX}{2} + \frac{Y_1 + Y_2}{2\mu}\right)$

2) 更新  $E$ ,  $E^* = \max\left(0, X - L + \frac{Y_1}{\mu} - \frac{\alpha}{\mu}\right) + \min\left(0, X - L + \frac{Y_1}{\mu} + \frac{\alpha}{\mu}\right)$

3) 更新  $P$ ,  $P^* = VU^T$

4) 更新  $Q$ ,  $Q^* = \mu P^T \left(L - W - \frac{Y_2}{\mu}\right) X^T \cdot (\gamma G + \mu XX^T)^{-1}$

5) 更新  $W$ ,  $W^* = \max\left(0, L - PQX - \frac{Y_2}{\mu} - \frac{\gamma\beta}{\mu}\right) + \min\left(0, L - PQX - \frac{Y_2}{\mu} + \frac{\gamma\beta}{\mu}\right)$

6) 更新  $Y_1$  和  $Y_2$ ,  $\mu$

$$\begin{aligned} Y_1 &= Y_1 + \mu(X - L - E) \\ Y_2 &= Y_2 + \mu(W - L + PQX) \\ \mu &= \min(\rho\mu, \mu_{\max}) \end{aligned}$$

7) 检查收敛条件

$$\|X - L - E\|_{\infty} < \varepsilon, \quad \|W - L + PQX\|_{\infty} < \varepsilon$$

结束循环

输出  $L$ ,  $E$ ,  $P$ ,  $Q$

### 3.2 算法复杂度分析

算法1中,  $L$  的计算复杂度主要集中于奇异值

分解,  $L$  中奇异值分解的计算复杂度为  $O(mn^2)$ 。 $E$  的计算复杂度为  $O(mn)$ 。 $P$  的计算复杂度也主要集中于奇异值分解, 为  $O(km^2)$ ,  $k$  是降维后的维数。 $Q$  的计算复杂度主要集中于  $(\gamma G + \mu XX^T)^{-1}$  涉及的逆运算, 为  $O(m^3)$ 。 $W$  的计算复杂度为  $O(mn)$ 。因此, 在一次迭代中, 计算复杂度不会超过  $O(m^3)$ 。如果算法1需要  $t$  次迭代, 则总的计算复杂度为  $O(tm^3)$ 。

## 4 实验

本节通过重构和聚类实验验证所提方法的有效性。

### 4.1 重构和聚类实验

本节旨在对所提方法进行评估。评估实验主要是测试所提方法以及7种特征选择对比方法在5个不同数据集上的聚类效果。对比方法分别为基于灵活优化图(FOG, flexible optimal graph)的无监督特征选择方法<sup>[18]</sup>、UDFS<sup>[13]</sup>、稳健的无监督特征选择(RUFS, robust unsupervised feature selection)方法<sup>[15]</sup>、SOGFS<sup>[17]</sup>、无监督软标签特征选择(USFS, unsupervised soft-label feature selection)方法<sup>[11]</sup>、AW-SPCA<sup>[22]</sup>、稀疏主成分分析(SPCA, sparse principal component analysis)方法<sup>[21]</sup>以及BaseLine。为了公平对比, 上述所有方法在5个基准数据集上进行评估, 包括ORL<sup>[12]</sup>、COIL20<sup>[37]</sup>、MNIST<sup>[38]</sup>、USPS<sup>[39]</sup>和LUNG<sup>[40]</sup>。为了测试所提方法对噪声的稳健性, 本文在ORL和MNIST数据集上, 从每个类中随机抽取20%的样本替换为离群点噪声。在COIL20、USPS和LUNG数据集上, 从每个类中随机抽取20%的样本, 并对抽到的每个样本随机添加40%的椒盐噪声, 椒盐噪声示例如图1所示。不同数据集的基本信息如表1所示。对实验效果的评估, 本文采用2种广泛使用的评价指标<sup>[18,41-42]</sup>, 即聚类准确率(ACC, accuracy)和归一化互信息(NMI, normalized mutual information)。ACC评估聚类结果和真实标签的一致程度, 计算式为

$$ACC = \frac{\sum_{i=1}^n \delta(\text{map}(f_i), g_i)}{n} \quad (18)$$

其中,  $n$  为样本数量;  $f_i$  为样本  $x_i$  的聚类标签;  $g_i$  为真实的类标签;  $\text{map}(f_i)$  为排列映射函数, 目的是映射每一个聚类标签  $f_i$  到样本真实的标签。当  $x \neq y$  时,  $\delta(x, y) = 0$ ; 否则,  $\delta(x, y) = 1$ 。

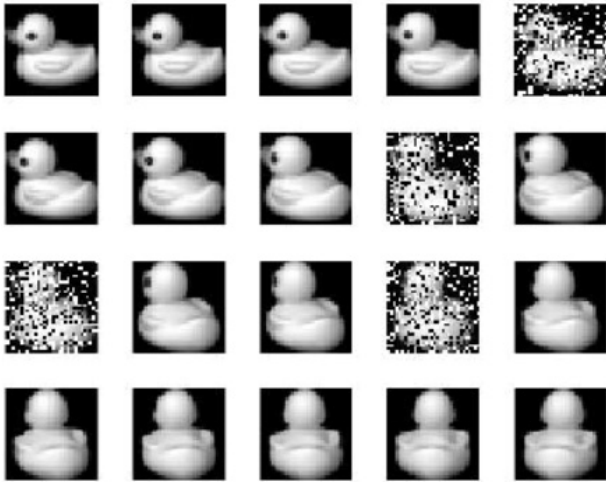


图 1 椒盐噪声示例

表 1 不同数据集的基本信息

数据集	样本数	特征数	类别数
ORL	400	1 024	40
COIL20	1 440	1 024	20
MNIST	2 000	784	10
USPS	9 298	256	10
LUNG	203	3 312	5

NMI 用于评估聚类结果和真实标签的关联程度，计算式为

$$NMI = \frac{MI(C, C')}{\max(H(C), H(C'))} \quad (19)$$

其中,  $C$  和  $C'$  分别表示真正的类别信息和算法得到的聚类信息,  $H(C)$  和  $H(C')$  分别表示  $C$  和  $C'$  的熵, 互信息 MI 的计算式为

$$MI(C, C') = \sum_{c_i \in C, c'_j \in C'} p(c_i, c'_j) \log \frac{p(c_i, c'_j)}{p(c_i)p(c'_j)} \quad (20)$$

其中,  $n$  为样本数量,  $p(c_i)$  和  $p(c'_j)$  分别为样本属于聚类  $c_i$  和  $c'_j$  的概率,  $p(c_i, c'_j)$  为样本既属于聚类  $c_i$  又属于聚类  $c'_j$  的概率. ACC 和 NMI 的取值为 0~1, 其值越大, 说明聚类效果越好.

为了公平对比, 所有方法的模型参数都从  $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$  中选取. 特征选择的比例参数统一从  $\{0.1, 0.2, 0.3, \dots, 0.9, 1.0\}$  中选取. 所有方法在目标函数值收敛时停止迭代, 然后获得一个聚类结果. 所有方法对每个数据集重复 20 次的 k-means 聚类实验, 以聚类平均值作为最终的结果.

本文测试了不同方法在不同数据集及其噪声数据集上的聚类 ACC 和 NMI, 结果如表 2 和表 3 所示, 其中, —表示未添加噪声, 20%表示从数据集中选取 20%的样本添加噪声, 加粗数据表示最优结果. 从表 2 和表 3 可以看出, 所提方法在除 LUNG 噪声数据集外的其他数据集上表现出优越的性能. 此外, 为了直观展示聚类结果和特征个数之间的关系, 不同方法在不同数据集上的聚类 ACC 如图 2 所示. 从图 2 可以看出, 每个特征选择方法的聚类 ACC 通常随着所选特征数的增加先增大后减小. 这说明特征选择方法的聚类结果和所选特征数密切相关. 综合表 2、表 3 和图 2 可以看出, 所提方法在噪声数据集上相对次优方法得到较大的提高, 尤其是聚类 ACC 在 COIL20 噪声数据集上提高了 13.37%, 在 ORL 噪声数据集上提高了 1.46%, 在 MNIST 噪声数据集上提高了 3.02%; 在干净数据集上的提高相对较小, 提高范围为 1.44%~3.65%.

表 2 不同方法在不同数据集上的聚类 ACC

数据集	噪声	BaseLine	UDFS	RUFS	SOGFS	AW-SPCA	SPCA	USFS	FOG	所提方法
ORL	—	0.518 1	0.525 9	0.540 4	0.533 3	0.524 3	0.535 9	0.537 9	0.534 5	<b>0.554 8</b>
	20%	0.359 1	0.374 5	0.377 1	0.368 4	0.372 1	0.367 3	0.378 9	0.429 3	<b>0.443 9</b>
MNIST	—	0.423 2	0.460 3	0.451 4	0.456 0	0.492 4	0.457 2	0.522 6	0.501 7	<b>0.559 1</b>
	20%	0.379 5	0.385 8	0.400 1	0.390 2	0.424 1	0.425 4	0.419 2	0.386 5	<b>0.455 6</b>
COIL20	—	0.539 3	0.563 6	0.584 3	0.567 9	0.574 2	0.584 3	0.613 3	0.586 9	<b>0.631 0</b>
	20%	0.400 2	0.423 8	0.450 0	0.435 5	0.494 7	0.469 7	0.485 8	0.455 9	<b>0.628 4</b>
USPS	—	0.566 8	0.591 0	0.588 7	0.589 9	0.655 9	0.624 7	0.651 6	0.613 3	<b>0.671 1</b>
	20%	0.496 1	0.519 6	0.523 1	0.528 6	0.533 9	0.528 6	0.536 4	0.529 7	<b>0.541 1</b>
LUNG	—	0.730 3	0.718 0	0.759 9	0.714 0	0.743 3	0.744 8	0.761 3	0.752 5	<b>0.788 7</b>
	20%	0.568 2	0.625 6	0.651 7	0.616 3	0.641 6	0.641 1	<b>0.657 1</b>	0.636 9	0.645 3

表 3 不同方法在不同数据集上的 NMI

数据集	噪声	BaseLine	UDFS	RUFS	AW-SPCA	SPCA	USFS	FOG	所提方法
ORL	—	0.718 7	0.727 4	0.737 3	0.727 5	0.731 2	0.732 5	0.731 4	<b>0.746 8</b>
	20%	0.573 0	0.589 0	0.590 0	0.588 0	0.589 0	0.595 0	0.581 2	<b>0.595 3</b>
MNIST	—	0.338 3	0.391 2	0.377 0	0.421 5	0.382 9	0.449 2	0.435 9	<b>0.496 5</b>
	20%	0.282 0	0.291 0	0.304 0	0.332 0	0.344 0	0.324 0	0.290 6	<b>0.367 2</b>
COIL20	—	0.709 3	0.721 5	0.735 1	0.737 5	0.723 4	0.747 3	0.729 1	<b>0.751 4</b>
	20%	0.522 0	0.546 0	0.571 0	0.613 0	0.597 0	0.606 0	0.563 6	<b>0.747 8</b>
USPS	—	0.546 0	0.552 7	0.559 0	0.611 6	0.579 3	0.610 7	0.573 0	<b>0.614 2</b>
	20%	0.446 0	0.464 0	0.458 0	0.482 0	0.470 0	0.474 0	0.461 6	<b>0.484 2</b>
LUNG	—	0.521 4	0.518 2	0.543 5	0.538 7	0.539 1	0.545 8	0.539 4	<b>0.560 1</b>
	20%	0.267 0	0.337 0	0.344 0	0.327 0	0.327 0	<b>0.351 0</b>	0.334 5	0.344 7

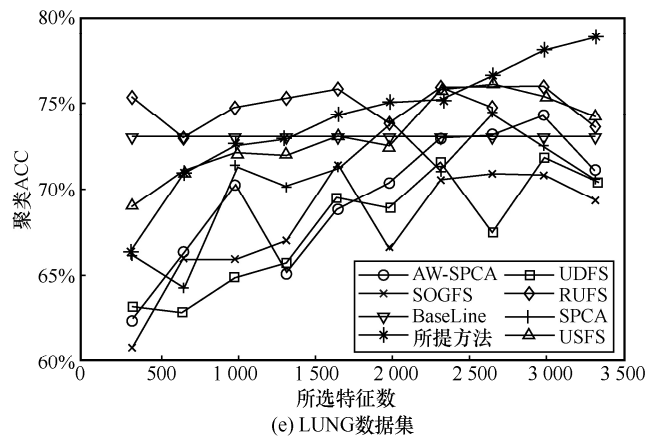
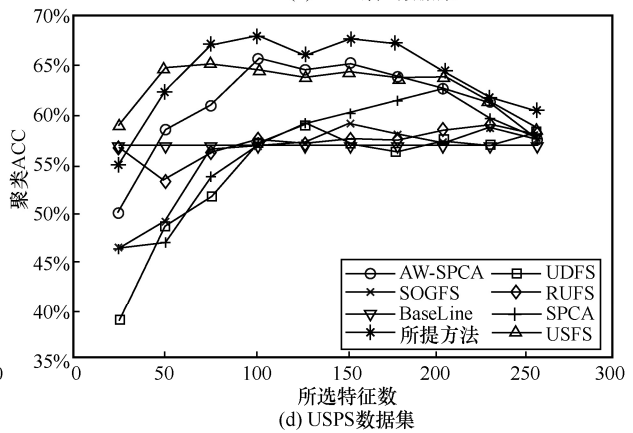
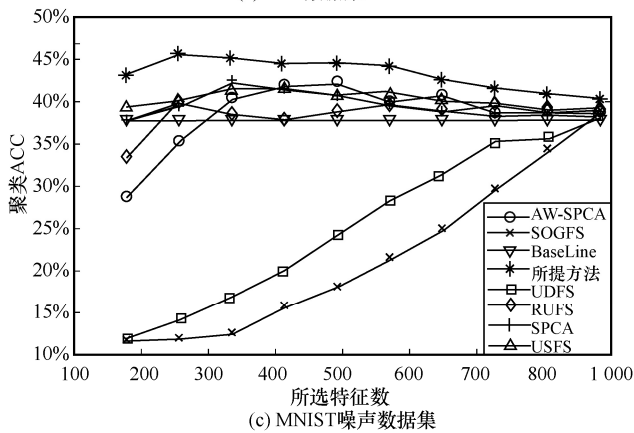
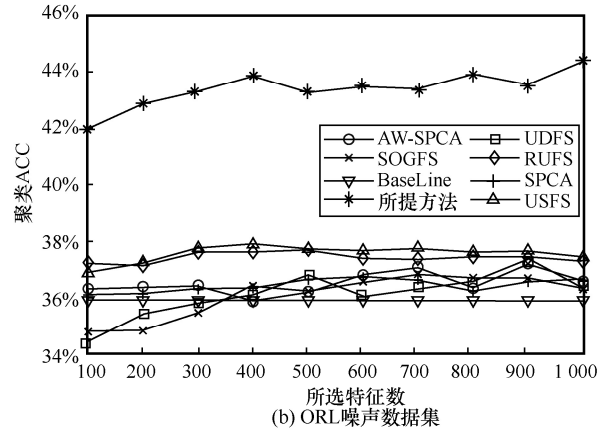
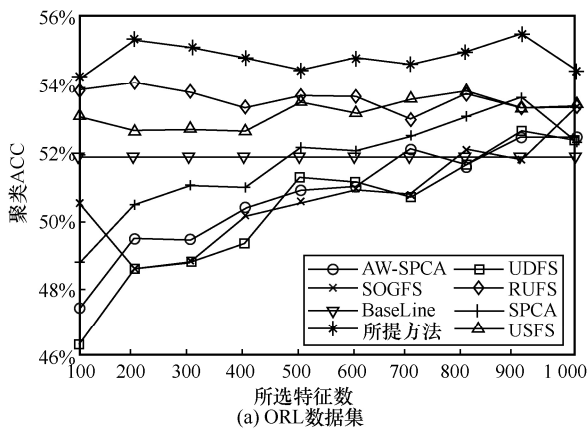


图 2 不同方法在不同数据集上的聚类准确率

通过表 2、表 3 和图 2 可以看到，对比方法中比较具有竞争力的是 USFS 方法。USFS 方法在数据的低维子空间中，借助聚类中心动态构建软标签矩阵，指导鉴别特征子集的认识。由于 USFS 方法同时考虑了噪声和数据局部结构，因此能较好地选取有效特征。

对比 PCA 重构角度的特征选择方法 AW-SPCA 和 SPCA，所提方法在所有数据集上表现出绝对的优越性。SPCA 和 AW-SPCA 都是使用原始数据作为重构的标准。虽然 AW-SPCA 方法可以赋予被噪声污染的样本相应的权重，再通过权重去控制这些样本在重构误差项中的参与程度，但是模型本质上还是在重构可能含有噪声的原始数据。不合理的重构目标导致模型学习出一个不准确的投影矩阵，进而影响聚类效果。相比之下，所提方法通过对参与重构误差项的样本做低秩处理，去除噪声的影响，这样可以保证重构目标是干净数据。所提方法在干净数据和投影重构数据之间作差，从原理上保证了学习到的投影矩阵更准确，最终聚类效果得以提升。

为了展示所提方法在最优聚类结果下的重构效果，本文将 COIL20 噪声数据集上的重构结果进行可视化，如图 3 所示。

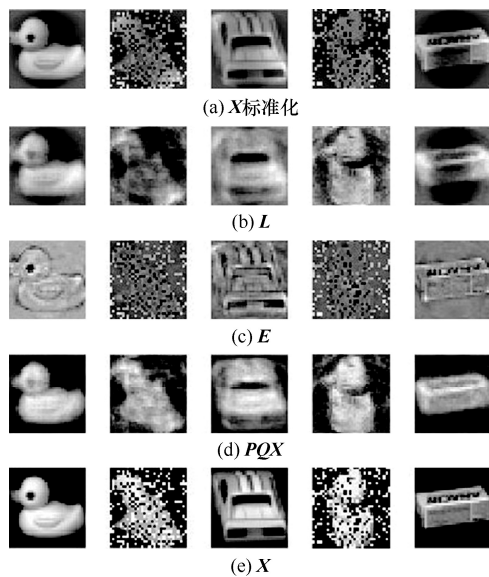


图 3 所提方法在 COIL20 噪声数据集上的重构效果可视化

图 3(a)表示对原始数据  $X$  标准化后的数据，图 3(b)和图 3(c)分别表示学习到的  $L$  和  $E$ ，图 3(d)表示使用模型学习到的  $P$  和  $Q$  对原始数据  $X$  的重构效果，图 3(e)表示标准化前的原始数据  $X$ 。从图 3 可以看出，所提方法可以将椒盐噪声较好地分

离出来。所提方法根据学习到的  $Q$  从  $L$  上选择特征后用于聚类，因此所提方法不仅重构效果好，而且聚类效果也好。

### 4.2 参数分析

所提模型中有 3 个超参数，即  $\alpha$ 、 $\beta$ 、 $\gamma$ 。本节实验令 3 个参数都在区间  $[10^{-3}, 10^3]$  遍历。值得注意的是，模型中参数  $\alpha$  的值越大，则  $E$  越稀疏。参数  $\alpha$  的调整经验是在一个理论值附近寻找，该理论值为  $\frac{1}{\sqrt{\max(m,n)}}$ 。因此，先遍历参数  $\alpha$ ，然后在理论值附近进行微调。同理，参数  $\beta$  的值越大， $L - PQX$  则越稀疏。

所提方法的最优参数具体如下，在 ORL 数据集上， $\alpha = 0.1$ ， $\beta = 10^3$ ， $\gamma = 10$ 。在 ORL 噪声数据集上， $\alpha = 0.0375$ ， $\beta = 10$ ， $\gamma = 0.001$ 。在 COIL20 数据集上， $\alpha = 0.1$ ， $\beta = 0.01$ ， $\gamma = 1000$ 。在 COIL20 噪声数据集上， $\alpha = 1$ ， $\beta = 10$ ， $\gamma = 1000$ 。在 MNIST 数据集上， $\alpha = 0.1$ ， $\beta = 0.001$ ， $\gamma = 100$ 。在 MNIST 噪声数据集上， $\alpha = 1$ ， $\beta = 1$ ， $\gamma = 1000$ 。在 USPS 数据集上， $\alpha = 0.1$ ， $\beta = 1$ ， $\gamma = 100$ 。在 USPS 噪声数据集上， $\alpha = 1$ ， $\beta = 0.1$ ， $\gamma = 1000$ 。在 LUNG 数据集上， $\alpha = 10$ ， $\beta = 100$ ， $\gamma = 10^{-3}$ 。在 LUNG 噪声数据集上， $\alpha = 1$ ， $\beta = 1000$ ， $\gamma = 1000$ 。

从调参经验可知，当数据集中含有噪声时，最优参数  $\alpha$  集中在理论值  $\frac{1}{\sqrt{\max(m,n)}}$  附近，取值较小，最优参数  $\gamma$  的值不会太大。当数据集不含有噪声时，如 USPS 和 LUNG 数据集，最优参数  $\alpha$  取值相对较大，最优参数  $\gamma$  的值比较大，弱化低秩项的作用，强化投影项的作用。为了让  $PQX$  接近  $L$ ，参数  $\beta$  的值应该较大。

为了直观展示  $\alpha$ 、 $\beta$ 、 $\gamma$  对聚类结果的影响，可以通过固定其中一个参数，变化其他 2 个参数来观察具体影响。例如，图 4 展示了在 ORL 数据集上， $\alpha$ 、 $\beta$ 、 $\gamma$  对聚类结果的影响。从图 4 可以看出，当固定  $\alpha = 0.1$  时，实验结果对  $\gamma$  比较敏感；当固定  $\gamma = 10$  时，实验结果对  $\alpha$  比较敏感。当固定  $\beta = 1000$  时，实验结果对  $\alpha$  和  $\gamma$  都敏感。

### 4.3 收敛分析

所提方法的收敛曲线如图 5 所示。从图 5 可以看到，所提方法的目标函数值经过 50 次迭代左右收敛。

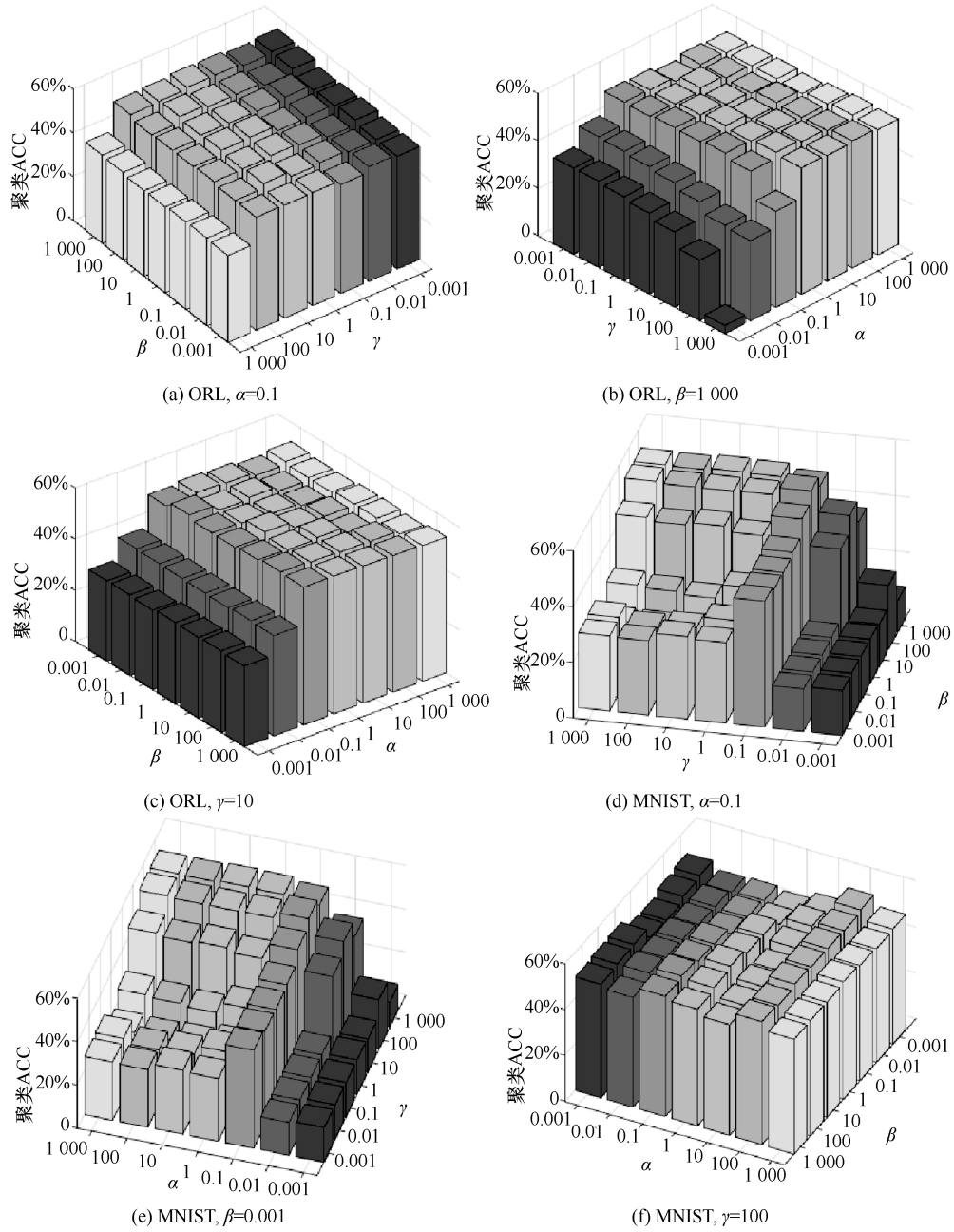


图 4 不同参数下的聚类准确率

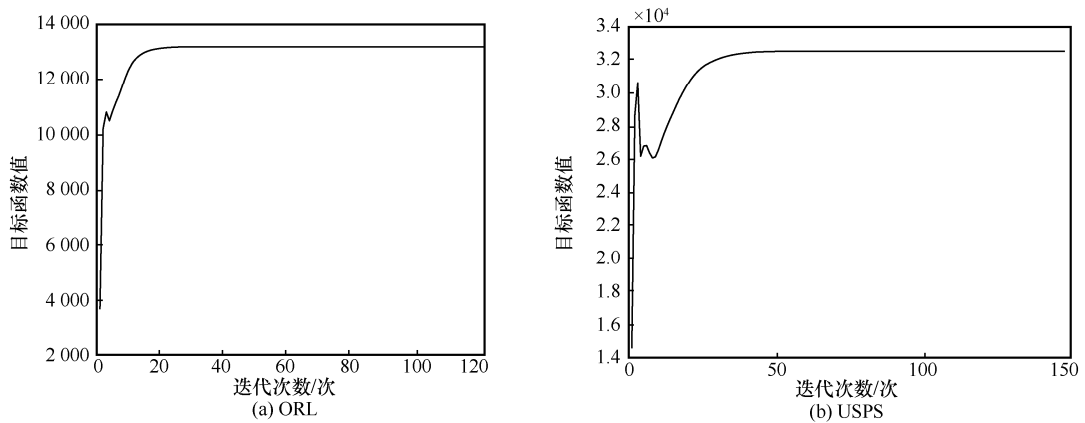


图 5 所提方法的收敛曲线

## 5 结束语

本文提出一种联合低秩重构和投影重构的稳健特征选择方法,该方法基于低秩重构和投影重构构建有效的重构误差项,重构实验结果和聚类实验结果表明,所提方法从干净数据上选择特征用于聚类可以获得较好的效果。更重要的是,聚类和重构效果达到了一致性。

### 参考文献:

- [1] JENNIFER G D, BRODLEY C E. Feature selection for unsupervised learning[J]. *Journal of Machine Learning Research*, 2004, 5: 845-889.
- [2] HE X F, CAI D, PARTHA N. Laplacian score for feature selection[C]//*Advances in Neural Information Processing Systems*. Massachusetts: MIT Press, 2006: 507-514.
- [3] ZHAO Z, LIU H. Spectral feature selection for supervised and unsupervised learning[C]//*Proceedings of the 24th International Conference on Machine Learning*. New York: ACM Press, 2007: 1151-1157.
- [4] NIE F P, XIANG S M, JIA Y Q, et al. Trace ratio criterion for feature selection[C]//*Proceedings of the 23rd National Conference on Artificial Intelligence*. New York: ACM Press, 2008: 671-676.
- [5] ZHAO Z, WANG L, LIU H. Efficient spectral feature selection with minimum redundancy[C]//*Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. New York: ACM Press, 2010: 673-678.
- [6] LI Z C, YANG Y, LIU J, et al. Unsupervised feature selection using nonnegative spectral analysis[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 26(1): 1026-1032.
- [7] ZHU P F, ZUO W M, ZHANG L, et al. Unsupervised feature selection by regularized self-representation[J]. *Pattern Recognition*, 2015, 48(2): 438-446.
- [8] LEI C, ZHU X F. Unsupervised feature selection via local structure learning and sparse learning[J]. *Multimedia Tools and Applications*, 2018, 77(22): 29605-29622.
- [9] ARGYRIOU A, EVGENIOU T, PONTIL M. Convex multi-task feature learning[J]. *Machine Learning*, 2008, 73(3): 243-272.
- [10] 刘文慧. PCA 与 PLS 用于高维数据分类的比较性研究[C]//2011 年中国卫生统计学年会论文集. [S.l.:s.n.], 2011: 422-424.  
LIU W H. Comparative study of PCA and PLS for high-dimensional data classification[C]//*Proceedings of the 2011 China Annual Conference of Health Statistics*. [S.l.:s.n.], 2011: 422-424.
- [11] WANG F, ZHU L, LI J, et al. Unsupervised soft-label feature selection[J]. *Knowledge-Based Systems*, 2021: doi.org/10.1016/j.knsys.2021.106847.
- [12] 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*. New York: ACM Press, 2010: 25-28.
- [13] YANG Y, SHEN H T, MA Z G, et al.  $l_{2,1}$ -norm regularized discriminative feature selection for unsupervised learning[C]//*Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*. New York: ACM Press, 2011: 1589-1594.
- [14] SHI Y, MIAO J Y, WANG Z Y, et al. Feature selection with  $l_{2,1,2}$  regularization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(10): 4967-4982.
- [15] QIAN M, ZHAI C. Robust unsupervised feature selection[C]//*Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. New York: ACM Press, 2013: 1621-1627.
- [16] SHI L, DU L, SHEN Y D. Robust spectral learning for unsupervised feature selection[C]//*Proceedings of 2014 IEEE International Conference on Data Mining*. Piscataway: IEEE Press, 2015: 977-982.
- [17] 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*. New York: ACM Press, 2016: 1302-1308.
- [18] CHEN H, NIE F P, WANG R, et al. Unsupervised feature selection with flexible optimal graph[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022: doi.org/10.1109/TNNLS.2022.3186171.
- [19] TURK M A, PENTLAND A P. Face recognition using eigenfaces[C]//*Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 1991: 586-591.
- [20] ZOU H, HASTIE T, TIBSHIRANI R. Sparse principal component analysis[J]. *Journal of Computational and Graphical Statistics*, 2006, 15(2): 265-286.
- [21] SEGHOUEANE A K, SHOKOUHI N, KOCH I. Sparse principal component analysis with preserved sparsity pattern[J]. *IEEE Transactions on Image Processing*, 2019, 28(7): 3274-3285.
- [22] YI S Y, HE Z Y, JING X Y, et al. Adaptive weighted sparse principal component analysis for robust unsupervised feature selection[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(6): 2153-2163.
- [23] CANDES E J, LI X, MA Y. Robust principal component analysis?[J]. *Journal of the ACM*, 2011, 58(3): 1-39.
- [24] WRIGHT J, PENG Y, MA Y, et al. Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization[C]//*Proceedings of Neural Information Processing Systems*. Massachusetts: MIT Press, 2009: 1-9.
- [25] LIU G C, LIN Z C, YU Y. Robust subspace segmentation by low-rank representation[C]//*Proceedings of the 27th International Conference on International Conference on Machine Learning*. New York: ACM Press, 2010: 663-670.
- [26] LIU G C, LIN Z C, YAN S C, et al. Robust recovery of subspace structures by low-rank representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 171-184.
- [27] SHAHID N, KALOFOLIAS V, BRESSON X, et al. Robust principal component analysis on graphs[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2016: 2812-2820.
- [28] NIE F P, YUAN J J, HUANG H. Optimal mean robust principal component analysis[C]//*Proceedings of the 31st International Conference on International Conference on Machine Learning*. New York: ACM Press, 2014: 1062-1070.
- [29] SHI X S, NIE F P, LAI Z H, et al. Robust principal component analysis via optimal mean by joint  $l_{2,1}$  and Schatten  $p$ -norms minimization[J]. *Neurocomputing*, 2018, 283: 205-213.
- [30] LUO M N, NIE F P, CHANG X J, et al. Avoiding optimal mean  $l_{2,1}$ -norm maximization-based robust PCA for reconstruction[J]. *Neural Computation*, 2017, 29(4): 1124-1150.

- [31] LUO M N, NIE F P, CHANG X J, et al. Avoiding optimal mean robust PCA/2DPCA with non-greedy  $l_1$ -norm maximization[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: ACM Press, 2016: 1802-1808.
- [32] LIAO S L, LI J, LIU Y, et al. Robust formulation for PCA: avoiding mean calculation with  $l_{2,p}$ -norm maximization[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018: 3604-3610.
- [33] NIE F P, TIAN L, HUANG H, et al. Non-greedy  $l_{2_1}$ -norm maximization for principal component analysis[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 30: 5277-5286.
- [34] WANG Q Q, GAO Q X, GAO X B, et al.  $l_{2,p}$ -norm based PCA for image recognition[J]. IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society, 2018, 27(3): 1336-1346.
- [35] 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.
- [36] QW A, QXG A, GAN S B, et al. Double robust principal component analysis[J]. Neurocomputing, 2020, 391: 119-128.
- [37] NENE S A, NAYAR S K, MURASE H, et al. Columbia object image library[R]. 1996.
- [38] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [39] HULL J J. A database for handwritten text recognition research[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(5): 550-554.
- [40] SINGH D, FEBBO P G, ROSS K, et al. Gene expression correlates of clinical prostate cancer behavior[J]. Cancer Cell, 2002, 1(2): 203-209.
- [41] 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.

- [42] HU H J, WANG R, NIE F P, et al. Fast unsupervised feature selection with anchor graph and  $l_{2,1}$ -norm regularization[J]. Multimedia Tools and Applications, 2018, 77(17): 22099-22113.

#### [作者简介]



仪双燕（1987-），女，山东菏泽人，博士，深圳信息职业技术学院讲师，主要研究方向为模式识别、机器学习。

梁永生（1971-），男，黑龙江肇东人，博士，哈尔滨工业大学教授，主要研究方向为通信信源、信道、网络协同优化编码。

陆晶晶（1996-），女，广西南宁人，深圳国微福芯技术有限公司工程师，主要研究方向为模式识别。

柳伟（1973-），男，湖南长沙人，博士，深圳信息职业技术学院教授，主要研究方向为人工智能、视觉媒体处理。

胡涛（1979-），男，湖北黄冈人，博士，深圳信息职业技术学院高级工程师，主要研究方向为图像处理、机器视觉等。

何震宇（1978-），男，江西抚州人，博士，哈尔滨工业大学（深圳）教授、博士生导师，主要研究方向为人工智能、计算机视觉等。