



智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

特征自表达和图正则化的鲁棒无监督特征选择

陈彤, 陈秀宏

引用本文:

陈彤,陈秀宏. 特征自表达和图正则化的鲁棒无监督特征选择[J]. 智能系统学报, 2022, 17(2): 286–294.

CHEN Tong, CHEN Xiuhong. Feature self-representation and graph regularization for robust unsupervised feature selection[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(2): 286–294.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202012043>

您可能感兴趣的其他文章

鲁棒的半监督多标签特征选择方法

A robust, semi-supervised, and multi-label feature selection method

智能系统学报. 2019, 14(4): 812–819 <https://dx.doi.org/10.11992/tis.201809017>

图正则化稀疏判别非负矩阵分解

Graph-regularized, sparse discriminant, non-negative matrix factorization

智能系统学报. 2019, 14(6): 1217–1224 <https://dx.doi.org/10.11992/tis.201811021>

基于低秩分解的鲁棒典型相关分析

Robust canonical correlation analysis based on low rank decomposition

智能系统学报. 2017, 12(4): 491–497 <https://dx.doi.org/10.11992/tis.201607024>

稀疏样本自表达子空间聚类算法

Sparse sample self-representation for subspace clustering

智能系统学报. 2016, 11(5): 696–702 <https://dx.doi.org/10.11992/tis.201601005>

面向自闭症辅助诊断的无监督模糊特征学习新方法

A novel unsupervised fuzzy feature learning method for computer-aided diagnosis of autism

智能系统学报. 2019, 14(5): 882–888 <https://dx.doi.org/10.11992/tis.201808005>

低秩分块矩阵的核近似

Kernel approximation of a low-rank block matrix

智能系统学报. 2019, 14(6): 1209–1216 <https://dx.doi.org/10.11992/tis.201904058>

微信公众平台



关注微信公众号，获取更多资讯信息

DOI: 10.11992/tis.202012043

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20210622.0900.002.html>

特征自表达和图正则化的鲁棒无监督特征选择

陈彤, 陈秀宏

(江南大学人工智能与计算机学院, 江苏 无锡 214122)

摘 要: 为了在揭示数据全局结构的同时保留其局部结构, 本文将特征自表达和图正则化统一到同一框架中, 给出了一种新的无监督特征选择 (unsupervised feature selection, UFS) 模型与方法。模型使用特征自表达, 用其余特征线性表示每一个特征, 以保持特征的局部结构; 用基于 $L_{2,1}$ 范数的图正则化项, 在保留数据的局部几何结构的同时可以降低噪声数据对特征选择的影响; 除此之外, 在权重矩阵上施加了低秩约束, 保留数据的全局结构。在 6 个不同的公开数据集上的实验表明, 所给算法明显优于其他 5 个对比算法, 表明了所提出的 UFS 框架的有效性。

关键词: 特征选择; 鲁棒; 图拉普拉斯; 特征自表达; 低秩约束; 无监督; $L_{2,1}$ 范数; 降维

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2022)02-0286-09

中文引用格式: 陈彤, 陈秀宏. 特征自表达和图正则化的鲁棒无监督特征选择 [J]. 智能系统学报, 2022, 17(2): 286–294.

英文引用格式: CHEN Tong, CHEN Xiuhong. Feature self-representation and graph regularization for robust unsupervised feature selection[J]. CAAI transactions on intelligent systems, 2022, 17(2): 286–294.

Feature self-representation and graph regularization for robust unsupervised feature selection

CHEN Tong, CHEN Xiuhong

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China)

Abstract: In order to reveal the global structure of data and preserve its local structure, this paper proposes a new unsupervised feature selection (UFS) method, which puts feature self-representation and graph regularization into the same framework. Specifically, the model uses the self-representation of the features to represent each feature through other features for preserving the local structure of the features. An $L_{2,1}$ -norm based graph regularization term is used to reduce the effect of noisy data on feature selection while preserving the local geometric structure. Furthermore, the model uses a low-rank constraint on the weight matrix to preserve the global structure. Experiments on six different public datasets show that the algorithm is clearly superior to the other five algorithms, which demonstrates the effectiveness of the proposed UFS framework.

Keywords: feature selection; robust; graph Laplacian; feature self-representation; low-rank constraint; unsupervised; $L_{2,1}$ -norm; dimension reduction

随着互联网的迅猛发展, 产生了大量的高维数据, 高维数据通常含有大量的冗余、噪声, 会降低学习算法的性能。基于这一问题, 特征选择和特征提取成为了降维的重要手段。特征提取通过将高维数据投影到低维子空间来减少数据的维度, 特征选择是直接选择原始数据的特征子集作为降维后的特征。本文, 从特征选择的角度展开研究。

特征选择方法包括有监督特征选择^[1]、半监督特征选择^[2]、无监督特征选择^[3-4]。有监督特征选择和半监督特征选择依赖于数据的标签信息, 然而在实际应用中, 类别标注的成本过高。因此, 这就要求应用无监督特征选择方法选择更具价值的特征。常见的无监督特征选择方法可以分为了三类: 过滤法^[3]、包裹法^[5]和嵌入法^[6]。过滤式特征选择方法独立于具体的学习算法, 运算效率较高。它的主要思想是对每一维特征赋予权重, 所得到的权重就代表着该特征的重要性, 然后依据权重进行排序, 把重要性相对较小的特征

收稿日期: 2020-12-28. 网络出版日期: 2021-06-22.

基金项目: 江苏省研究生科研与实践创新计划项目 (JNKY19_074).

通信作者: 陈秀宏. E-mail: xiuhongc@jiangnan.edu.cn.

过滤掉,将重要性较大的特征作为原始数据的表示。例如,最大方差法^[7]以数据方差作为评价标准,将特征按重要性排序进行选择。除此之外,拉普拉斯评分法^[3](LapScore)也是常见的无监督的过滤式特征选择方法。包裹式特征选择方法从初始特征集合中不断地选择特征子集,训练学习器,根据学习器的性能进行评价,直到选择出最佳的子集。LVW(Las Vegas Wrapper)^[8]是一个典型的包裹式特征选择方法,它在拉斯维加斯(Las Vegas method)框架下使用随机策略来进行子集搜索,并以最终分类器的误差作为特征子集评级准则。嵌入式特征选择在学习器训练过程中自动地进行特征选择,其分类效果通常较好,同时该方法可以实现对多个特征的选择。嵌入式选择最常用的是 L_1 正则化和 L_2 正则化^[6],当正则化项增大到一定程度时,所有的特征系数都会趋于0,在这个过程中,会有一部分特征的系数先变为0,也就实现了特征选择过程。除此之外,近年来又研究出了更加有效且高效的无监督特征选择方法,例如Li等^[9]将局部几何一致性和冗余最小化结合到同一框架中,利用局部几何结构的一致性来提高聚类精度,并在此过程中进行特征选择。Liu等^[10]提出了一种嵌入式无监督特征选择方法,该方法通过局部线性嵌入(locally linear embedding, LLE)算法得到特征权重矩阵,并使用 L_1 -范数来描述重构误差最小化。大量的实验结果证明,以上这些方法均是有效的。

对于特征选择来说,保持局部几何数据结构显然比保持全局结构更为重要^[11]。最常用的局部几何结构保持方法有以下3种:基于 L_2 范数的局部线性嵌入(LLE)^[12]、局部保留投影(locality preserving projections, LPP)^[13]以及局部切空间对齐(local tangent space alignment, LTSA)^[14]。而传统的局部保留方法是基于 L_2 范数,很容易受到噪声和冗余数据的影响,其次, L_2 范数已经被应用于许多正则化项中,这使得传统方法对离群值十分敏感,导致特征选择效果不理想。

基于以上提出的问题,本文通过特征自表达、图正则化以及低秩约束,提出了一个鲁棒无监督特征选择模型,同时保留了数据的局部几何结构和全局结构。并在6个公开数据集上进行实验,且与5个对比算法进行对比,证明了该模型的有效性。

1 相关工作

1.1 符号元素定义

对于任意矩阵 $X \in \mathbf{R}^{m \times n}$, x_{ij} 表示的是该矩阵

第 i 行第 j 列的元素。 $\text{tr}(X)$ 表示矩阵 X 的迹, X^{-1} 表示矩阵 X 的逆。矩阵 X 的 $L_{2,1}$ 范数被定义为

$$\|X\|_{2,1} = \sum_{i=1}^m \|x_i\|_2 = \sum_{i=1}^m \sqrt{\sum_{j=1}^n x_{ij}^2}$$

矩阵 X 的 F 范数被定义为

$$\|X\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2}$$

1.2 特征自表达

特征自表达已经被广泛地应用于无监督特征选择之中。例如Zhu等^[15]提出了一种用于无监督特征选择的正则化自表达(regularized self-representation, RSR)模型。在RSR中, $X = [x^1 x^2 \cdots x^n] = [x_1 x_2 \cdots x_d] \in \mathbf{R}^{n \times d}$ 表示特征矩阵,其中样本数和特征数分别是 n 和 d , X 的每一行代表一个样本,每一列代表一个特征维度。样本的特征自表达定义为

$$x_i \approx \sum_{j=1}^d x_j w_{ji}, \quad i = 1, 2, \dots, d \quad (1)$$

式中:权重矩阵 $W \in \mathbf{R}^{d \times d}$ 的元素 w_{ji} 表示第 i 个特征 x_i 和第 j 个特征 x_j 之间的权重。在式(1)中,每一个特征的特征表示项都是由其余重要的特征组成的,而不重要的特征应该从特征表示项中移除。

特征表示系数可通过以下模型来求解

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

式中: λ 为一个平衡参数,基于 $L_{2,1}$ 范数的正则化项可得到行稀疏的权重矩阵 W ,从而实现特征选择。

从式(1)、(2)可以看出,所有训练样本的每一个特征(即式(1)左侧的 x_i)都是由其他特征的线性组合所表示的,相应的权重向量是式(2)中 W 的第 i 列 w_i 。显然, w_i 中的值越大,其对应的特征在 x_i 的表示中所占的比重越多。除此之外,如果 W 的某一行的元素全为0,则相应的特征不出现在特征自表达中,即所有参与特征自表达的特征应该是重要的,而那些不重要的特征将通过 $\|W\|_{2,1}$ 来去除。

1.3 局部保留投影(LPP)

局部保留投影(LPP)^[13]通过获得线性投影来最优地保持数据的邻域结构,即样本之间的某种非线性关系在降维后仍然保留着这种关系。假设 W 是将样本数据投影到子空间的投影矩阵,则可以通过优化以下目标函数得到 W 的最优解:

$$\min_W \sum_{i,j=1}^n \|W^T x^i - W^T x^j\|^2 s_{ij} \quad (3)$$

式中: W 是投影矩阵; s 是相似矩阵,关于 s 的元素定义如下:

$$s_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|_2}{2\sigma^2}\right), & \mathbf{x}^i \text{ 是 } \mathbf{x}^j \text{ 的 } k \text{ 近邻或 } \mathbf{x}^j \text{ 是 } \mathbf{x}^i \text{ 的 } k \text{ 近邻} \\ 0, & \text{其他} \end{cases} \quad (4)$$

$N_k(\mathbf{x}^i)$ 是 \mathbf{x}^i 的 k 个近邻的集合, σ 是宽度参数。

2 基于特征自表达和图正则化模型

2.1 模型建立

为了给出基于特征自表达的鲁棒联合图正则化无监督特征选择 (feature self-representation and graph regularization, FSRGR) 模型, 首先引入以下图正则化特征自表达无监督特征选择一般模型:

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{XW}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} + \beta \varphi(\mathbf{W}) \quad (5)$$

式中: $\varphi(\mathbf{W})$ 是图正则化项, 用于保留数据的局部几何结构。 λ 和 β 是用于平衡 $L_{2,1}$ 范数和正则化项的两个参数。

2.1.1 基于 L_2 范数图正则化项的无监督特征选择

局部线性嵌入 (LLE)^[12]、局部保留投影 (LPP)^[13] 和局部切线空间对齐 (LTSA)^[14] 等均是基于 L_2 范数的图正则化方法。将 LPP 思想应用于模型 (5) 中, 则可得到以下模型:

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{XW}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} + \beta \sum_{i,j=1}^n \|\mathbf{W}^T \mathbf{x}^i - \mathbf{W}^T \mathbf{x}^j\|^2 s_{ij} \quad (6)$$

式 (6) 也可以表示为

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{XW}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} + \beta \text{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W}) \quad (7)$$

式中: \mathbf{L} 是一个拉普拉斯矩阵, 且 $\mathbf{L} = \mathbf{D} - \mathbf{S}$, \mathbf{D} 是一个对角矩阵, 其对角元素为 $D_{ii} = \sum_{j=1}^n s_{ij}$, \mathbf{S} 由式 (4) 定义。

2.1.2 基于 $L_{2,1}$ 范数图正则化项的无监督特征选择

在式 (7) 中, 最后一项图正则化项用于保留数据的局部几何结构, 然而, 基于 L_2 范数的正则化函数容易受到噪声数据的影响。为减少噪声的影响, 下面介绍一种基于 $L_{2,1}$ 范数的图正则化方法, 使得模型对噪声数据更加鲁棒。

由于拉普拉斯矩阵 \mathbf{L} 是实对称的, 故设矩阵 \mathbf{L} 特征分解为

$$\mathbf{L} = \mathbf{U} \mathbf{V} \mathbf{U}^T \quad (8)$$

于是, 式 (7) 中的图正则化项可以改写为

$$\text{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W}) = \text{tr}(\mathbf{W}^T \mathbf{X}^T \mathbf{U} \mathbf{V} \mathbf{U}^T \mathbf{X} \mathbf{W}) = \|\mathbf{V}^{\frac{1}{2}} \mathbf{U}^T \mathbf{X} \mathbf{W}\|_F^2 = \|\mathbf{G} \mathbf{W}\|_F^2 \quad (9)$$

其中 $\mathbf{G} = \mathbf{V}^{\frac{1}{2}} \mathbf{U}^T \mathbf{X}$ 。为了使模型对噪声数据更加鲁棒, 使用 $L_{2,1}$ 范数代替 F 范数, 可以得到基于 $L_{2,1}$ 范数的图正则化项:

$$\varphi(\mathbf{W}) = \|\mathbf{G} \mathbf{W}\|_{2,1} \quad (10)$$

于是式 (7) 可改写为

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{XW}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} + \beta \|\mathbf{G} \mathbf{W}\|_{2,1} \quad (11)$$

2.1.3 目标函数

式 (11) 通过特征自表达和图正则化来保留数据的局部几何结构, 但该模型却缺少了全局结构信息, 无法准确地捕捉特征的全局结构信息。因此本文采用低秩约束来保留数据的全局结构。区别于传统的特征自表达, 此处将特征自表达以及图正则化项与低秩约束相结合, 分别考虑了局部特征相关性和全局特征相关性, 从而得到局部特征自表达和全局特征自表达。具体来说, 就是对 \mathbf{W} 施加一个低秩假设, 即 $\mathbf{W} = \mathbf{A} \mathbf{B}$, 所得目标函数为

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{X} \mathbf{A} \mathbf{B}\|_F^2 + \lambda \|\mathbf{A} \mathbf{B}\|_{2,1} + \beta \|\mathbf{G} \mathbf{A} \mathbf{B}\|_{2,1} \quad (12)$$

式中: $\mathbf{A} \in \mathbf{R}^{d \times k}$ 为投影矩阵; $\mathbf{B} \in \mathbf{R}^{k \times d}$ 为恢复矩阵。 λ 和 β 均是平衡参数。

该模型利用特征自表达保留了特征的局部结构, 并采用了低秩约束 (即用 \mathbf{A} 和 \mathbf{B} 代替 \mathbf{W}) 来保留样本以及特征的全局结构。此外, 对传统 LPP 模型中的拉普拉斯矩阵进行特征分解, 建立基于 $L_{2,1}$ 范数的图正则化项, 增强其鲁棒性和稀疏性, 并保留了数据的局部几何结构。

2.2 优化算法

虽然式 (12) 中的目标函数是凸的, 但很难联合地求解所有变量。这里使用交替迭代优化策略去优化它, 即交替地迭代每一个变量, 直至算法收敛。

首先, 式 (12) 可以改写为

$$\min_{\mathbf{A}, \mathbf{B}} \text{tr}(\mathbf{X}^T \mathbf{X} - 2 \mathbf{X}^T \mathbf{X} \mathbf{A} \mathbf{B} + \mathbf{B}^T \mathbf{A}^T \mathbf{X}^T \mathbf{X} \mathbf{A} \mathbf{B}) + \lambda \text{tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B}) + \beta \text{tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{G}^T \mathbf{D} \mathbf{G} \mathbf{A} \mathbf{B}) \quad (13)$$

式中: \mathbf{P} 为对角矩阵, 将其第 i 个对角元素定义为

$$p_{ii} = \frac{1}{2\|(\mathbf{A} \mathbf{B})^i\|_2}, i = 1, 2, \dots, n \quad (14)$$

\mathbf{D} 为对角矩阵, 其第 i 个对角元素定义为

$$d_{ii} = \frac{1}{2\|(\mathbf{G} \mathbf{A} \mathbf{B})^i\|_2}, i = 1, 2, \dots, n \quad (15)$$

式中: $(\mathbf{A} \mathbf{B})^i$ 是矩阵 $\mathbf{A} \mathbf{B}$ 的第 i 行; $(\mathbf{G} \mathbf{A} \mathbf{B})^i$ 是矩阵 $\mathbf{G} \mathbf{A} \mathbf{B}$ 的第 i 行。

对式 (13) 中目标函数关于 \mathbf{B} 求导数并设令其等于 0, 可以得到:

$$\mathbf{B} = (\mathbf{A}^T \mathbf{S}_1 \mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}^T \mathbf{X} \quad (16)$$

其中, $\mathbf{S}_1 = \mathbf{X}^T \mathbf{X} + \lambda \mathbf{P} + \beta \mathbf{G}^T \mathbf{D} \mathbf{G}$ 。

将式 (16) 代入到式 (13) 中, 式 (13) 可以变为以下形式:

$$\min_{\mathbf{A}} \text{tr}(\mathbf{X}^T \mathbf{X} - \mathbf{X}^T \mathbf{X} \mathbf{A} (\mathbf{A}^T \mathbf{S}_1 \mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}^T \mathbf{X}) \quad (17)$$

它等价于以下问题:

$$\max_{\mathbf{A}} \text{tr}((\mathbf{A}^T \mathbf{S}_1 \mathbf{A})^{-1} \mathbf{A}^T \mathbf{S}_2 \mathbf{A}) \quad (18)$$

其中 $S_2 = X^T X X^T X$ 。

对式 (18) 关于 A 求导并令其等于零可得:

$$S_2 A = S_1 A ((A^T S_1 A)^{-1} A^T S_2 A) \quad (19)$$

如设广义特征方程 $S_2 a = \lambda S_1 a$ 的前 k 个最大特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_k$, 对应的特征向量分别为 a_1, a_2, \dots, a_k , 并令 $A = [a_1 \ a_2 \ \dots \ a_k]$, 则有

$$S_2 A = S_1 A \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) \quad (20)$$

此时, 式 (18) 中 $\text{tr}((A^T S_1 A)^{-1} A^T S_2 A)$ 达到最大。

这样, 通过求解式 (18) 得到 A , 再式 (16) 得到 B 。此过程可迭代进行, 直到终止条件满足。算法给出了更新 A 和 B 的具体步骤。

算法 FSRGR

输入 数据矩阵 $X \in \mathbb{R}^{n \times d}$, $L \in \mathbb{R}^{n \times n}$, k, λ, β 。

输出 $A \in \mathbb{R}^{d \times k}$, 计算 $\|A^i\|_2 (i = 1, 2, \dots, d)$ 并按照降序排列, 然后从 X 中选择与前 k 个最大值所对应的特征。

对 L 进行特征分解: $L = UVU^T$;

计算 G : $G = V^{\frac{1}{2}} U^T X$;

初始化: 对角矩阵 $P = I \in \mathbb{R}^{d \times d}$, $D = I \in \mathbb{R}^{n \times n}$, 投影矩阵 A , 恢复矩阵 B , 迭代次数 $t=0$ 。

重复:

1) 求解问题 (18) 更新 A ;

2) 根据式 (16) 更新 B ;

3) 根据式 (14) 更新 $P = \frac{1}{2\|(AB)^i\|_2}$;

4) 根据式 (15) 更新 $D = \frac{1}{2\|(GAB)^i\|_2}$;

5) $t = t + 1$;

直到收敛。

2.3 复杂度分析

在每次迭代中, FSRGR 算法的时间主要花费在式 (10) 中 $X^T X + \lambda P + \beta G^T D G$ 和 $(A^T S_1 A)^{-1} A^T X^T X$ 以及特征分解的计算上, 相应的复杂度为 $\max\{O(n^2 d), O(nd^2)\}$, $O(nd^2)$ 及 $O(n^3)$, 其中 n 和 d 代表样本数及特征数。在本次实验中, 所提出的算法一般在 10 次迭代内收敛, 所以算法的时间复杂度为 $\max\{O(n^2 d), O(nd^2), O(n^3)\}$ 。

3 实验与分析

将在 6 个公开数据集上对 FSRGR 算法进行评估, 并且与 5 个特征选择算法进行比较。以验证该算法的有效性。

3.1 数据集

实验中使用的 6 个公开数据集中, 包括 4 个人脸数据集 YALE^[16]、ORL^[17]、MSRA^[18] 以及 AR^[19], 2 个手写数字数据集 USPS^[20] 和 LEAVES^[21]。这些数据集的具体信息如表 1 所示, 数据集的部分图像如图 1 所示。并且对于每一个数据集, 均从

每类中随机选取 10 个样本组成训练集, 剩余的样本组成测试集进行实验。

表 1 数据集具体信息

数据集	样本数	类别数	特征数
Yale	165	15	1024
ORL	400	10	1024
MSRA	1799	12	256
AR	3120	120	1024
USPS	2007	10	256
LEAVES	400	10	1024



(a) YALE 数据集



(b) ORL 数据集



(c) MSRA 数据集



(d) AR 数据集



(e) USPS 数据集



(f) LEAVES 数据集

图 1 部分数据集的图片

Fig. 1 Visualization of some data sets with different noise

3.2 对比算法

为了验证所研究算法的有效性, 考虑以下对比方法:

1) NDFS^[22]: 联合学习数据的局部几何结构和基于 $L_{2,1}$ 范数正则化项的稀疏线性回归。

2) JELSR^[4]: 同时考虑了拉普拉斯正则化器和权重矩阵对特征的得分进行了排序。

3) SOGFS^[23]: 从低维空间中学习样本的全局结构来选择重要特征。

4) L1-UFS^[24]: 将特征自表达以及基于 L_1 范数的图拉普拉斯正则化融入到同一框架中。

5) LRLMR^[25]: 在低维子空间中进行学习, 使得对噪声更加鲁棒; 通过图的流形正则化项来保持原始数据的局部结构。

3.3 重构图像

图 2 展示了 6 种不同的算法 (NDFS、JELSR、SOGFS、L1-UFS、LRLMR、FSRGR) 关于 AR 数据集得到的部分重构图像。以图 2(g) 为例, 共展示了 7 张图像, 从左至右分别选取了 200、250、300、350、400、450、500 个特征进行图像的重构。从

图 2 中可以看出,同其他 5 个对比算法相比,FSRGR 算法对于图像的重构效果更好。而在 6 种算法中,JELSR 算法的重构效果最差,可以明显看出,随着所选特征数量的增加,该算法所选择的特征均为不重要的面部特征,而重要的人脸的五官特征并没有被选择。

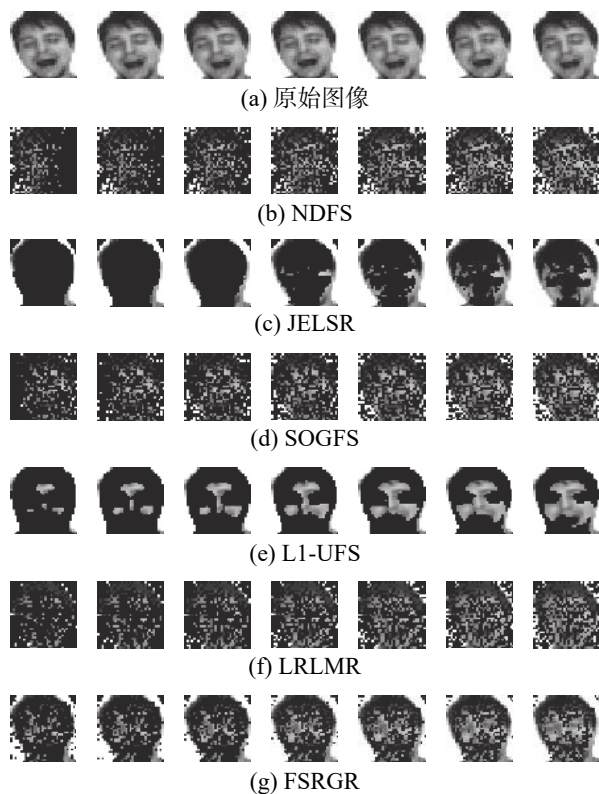


图 2 不同算法对部分 AR 数据集的重构图像

Fig. 2 Reconstructed images of partial AR data sets by different algorithms

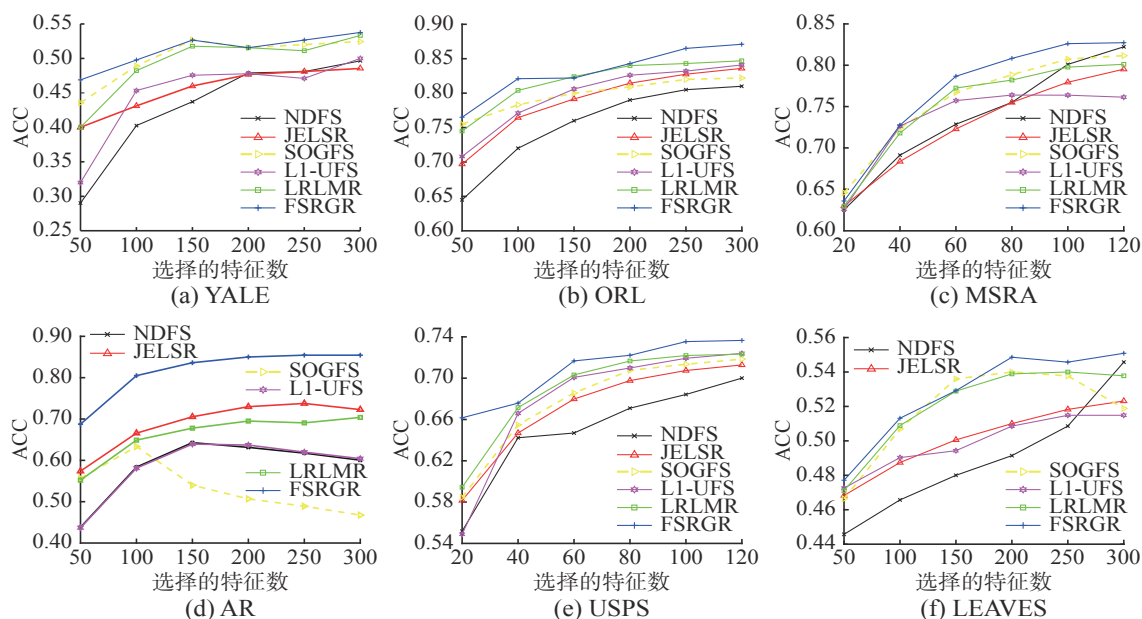


图 3 6 种算法在 6 种数据集上的分类精度

Fig. 3 Classification accuracies of six methods on six data sets

3.4 分类精度和聚类精度

将在不同数据集上比较 FSRGR 算法同其他 5 种对比算法,分别评估它们的聚类及分类性能。具体的结果见图 3 和图 4。其中图 3 展示了所有算法在 6 个数据集上关于分类精度的对比结果。如图 3 所示,其横轴表示所选择的特征数,纵轴代表不同的特征数所对应的分类精度 (ACC)。由于不同数据集的特征数通常是不相同的,因此对于 YALE、ORL、AR 和 LEAVES 数据集,我们所选择的特征数的范围为 {50, 100, 150, 200, 250, 300}, 而对于 MSRA 及 USPS 数据集,将所选择的特征数的范围设置为 {20, 40, 60, 80, 100, 120}。并且对于每一个算法,为了呈现出最真实的比较结果,均将其参数调至为最佳参数,即所呈现出的实验结果均为其最好的结果。首先从图 3 可以看出,在大多数情况下,本文所提出的 FSRGR 算法均优于其他对比算法。尤其在 AR 数据集上,相比其他 5 种对比算法,FSRGR 所取得的分类效果明显更加优越。还可以看出,在 5 种对比算法中,分类性能最好的是 LRLMR 算法,分类效果最差的是 NDFS 算法。其次,图 4 比较了 FSRGR 算法同其他对比算法在不同数据集上的聚类性能。从图中可以看出,随着所选特征数的增加,通过 FSRGR 算法所得到的聚类效果可以稳定地优于其他对比算法。这是因为本文所提出的方法既保留了数据的局部几何结构,又保留了数据的全局结构,并且通过对权重矩阵施加低秩约束,使得所学习到的投影向量的数量不受限制,从而可以得到足够多的投影向量。

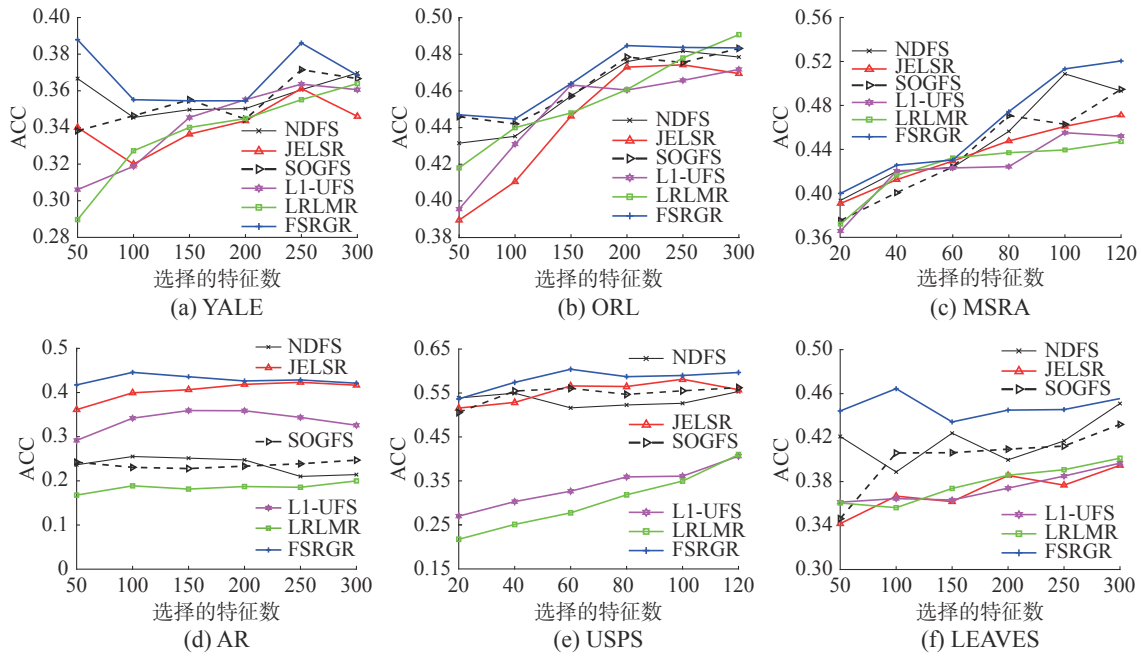


图 4 6 种算法在 6 种数据集上的聚类精度

Fig. 4 Clustering accuracies of six methods on six data sets

3.5 运行时间

为了比较不同算法的运行时间,在 YALE、MSRA、ORL 和 USPS 数据集上进行了聚类实验,所有算法均在 MATLAB 中实现。具体的实验结果位于表 2 中。

从表 2 可以看出,在所有方法中运行时间最短的是 LRLMR 算法,本文所提出的 FSRGR 算法同 L1-UFS、LRLMR 及 NDFS 算法相比,运行时间较长。而 JELSR 与 SOGFS 算法对于不同的数据集,运行时间并不稳定。

表 2 不同方法的运行时间

Table 2 Running time of different methods

数据集	FSRGR	JELSR	L1-UFS	LRLMR	NDFS	SOGFS
YALE	1.2050±0.0087	0.4319±0.0022	0.8064±0.0975	0.3943±0.1185	0.1782±0.098	2.1252±0.2921
MSRA	1.4480±0.0015	3.0495±0.0032	0.2080±0.0413	0.2080±0.0895	1.0388±0.0895	1.4380±0.1261
ORL	1.1572±0.0024	0.5262±0.0023	1.0063±0.0516	0.5111±0.1063	0.2211±0.0982	2.2545±0.0073
USPS	1.1509±0.0075	3.7314±0.0039	0.1315±0.0558	0.0995±0.0467	1.3163±0.0741	0.1380±0.0030

3.6 收敛性分析

通过计算不同迭代次数对应的目标函数值来观察 FSRGR 算法的收敛曲线。并且最大迭代次数设置为 10。由于 FSRGR 算法在不同的数据集上有相似的收敛性,因此在收敛性分析的部分只展示了在 YALE 数据集上的实验结果。从图 5 中可以发现: 1) 所提出的算法优化了式 (12) 所提出的目标函数,单调地减少了目标函数值直至算法收敛。2) 所提出的算法需要经过数次迭代来有效地达到收敛。

3.7 参数设置

重点研究参数 β 和 λ 的取值对于聚类精度的

影响。首先参数 β 是用于平衡 $\|X - XAB\|_F^2$ 和 $\|GAB\|_{2,1}$, 其次是参数 λ , 用于控制 AB 的稀疏性。并且这两个参数的取值范围均被设置为 $\{10^{-6}, 10^{-4}, 10^{-2}, 1, 10^2, 10^4, 10^6\}$ 。实验采用了 K-means 算法, 呈现出了不同的参数组合在不同的数据集上对于聚类精度的影响。具体的实验结果见图 6。

从图 6 可以看出, 本文所提出的方法对参数的设置较为敏感, 不同的参数组合具有不同的聚类结果。举例说明, 对于人脸数据集 YALE, 当 $\lambda = 10^{-2}, \beta = 10^{-6}$ 时, FSRGR 的聚类精度达到了 36.89% 的最好效果; 而对于手写字数据集 USPS, 当 $\lambda = 10^4, \beta = 10^{-4}$ 时, FSRGR 的聚类精度达到了

57.70%的最好效果。

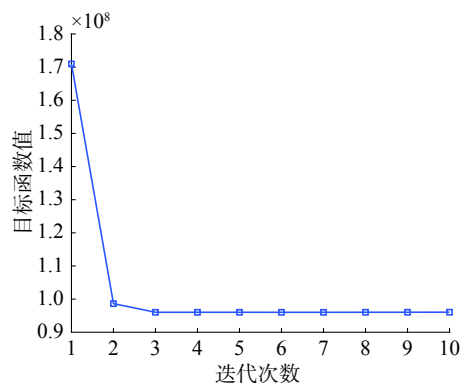


图5 FSRGR在YALE数据集上的收敛性
Fig. 5 Convergence of FSRGR on YALE data sets

3.8 低秩约束的影响

式(12)中不同秩($k \in \{3, 6, 9, 12, 15\}$)在数据集AR、ISOLET、PIE、LUNG、USPS、YALE上对聚类精度的影响。图7中给出了FSRGR在不同数据集上满秩和低秩下的聚类精度。其中横轴代表所选定的秩的范围,纵轴代表不同的秩所对应的分类精度。

从图7中可以观察到,低秩约束下的聚类性能在大多数情况下都优于满秩。例如,在数据集AR和USPS上,同全秩的情况下相比,低秩约束下所得到的平均分类精度分别高出了12.43%和17.88%。这说明在特征选择中,对高维数据进行低秩约束是有效的。

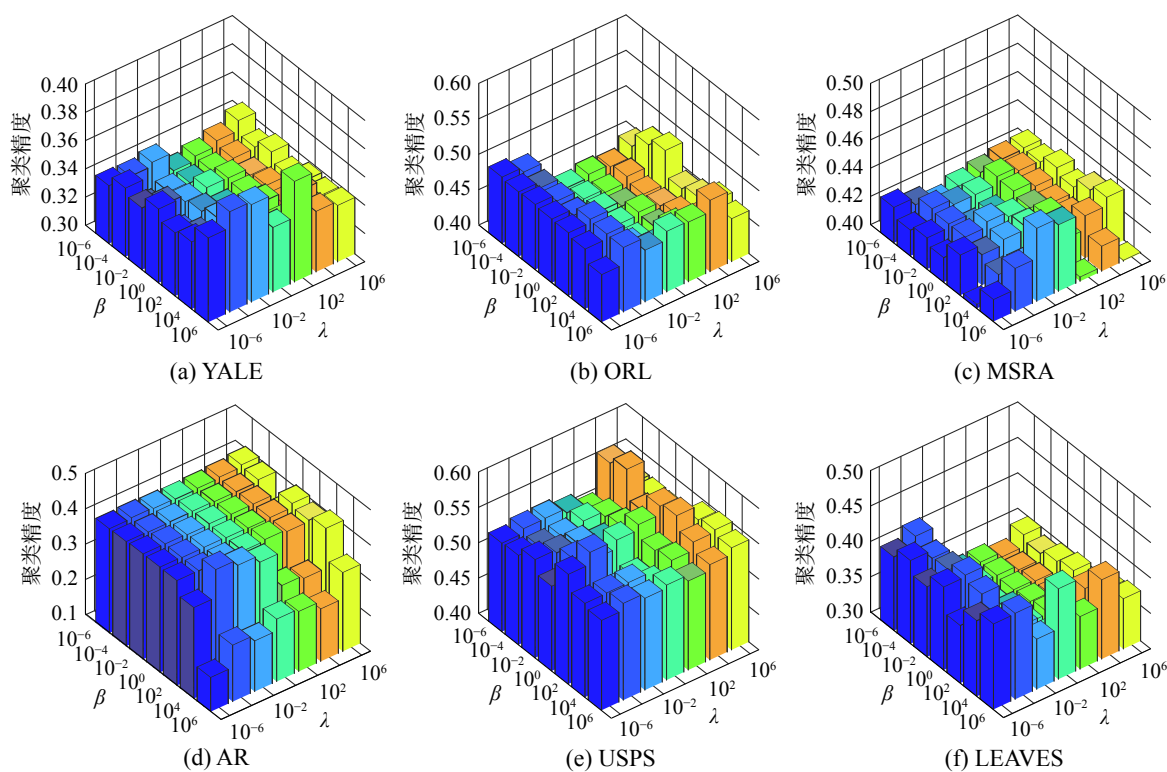
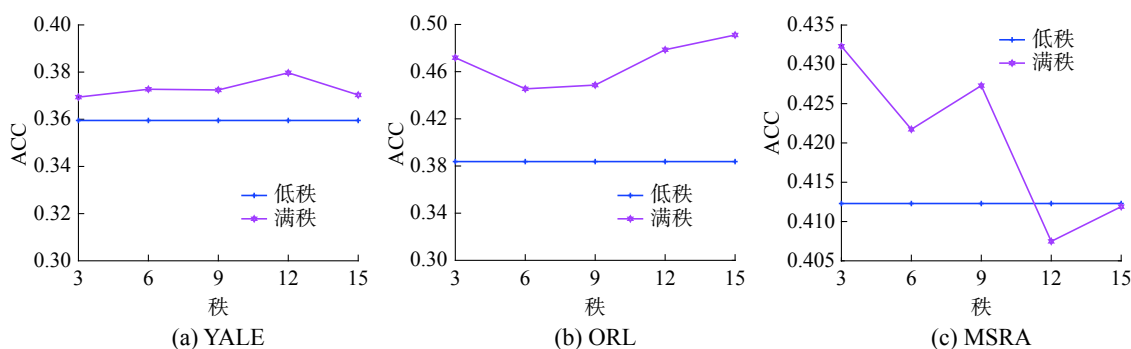


图6 不同 β 和 λ 组合下的FSRGR聚类精度

Fig. 6 Clustering accuracy of FSRGR algorithm with different combination of β and λ



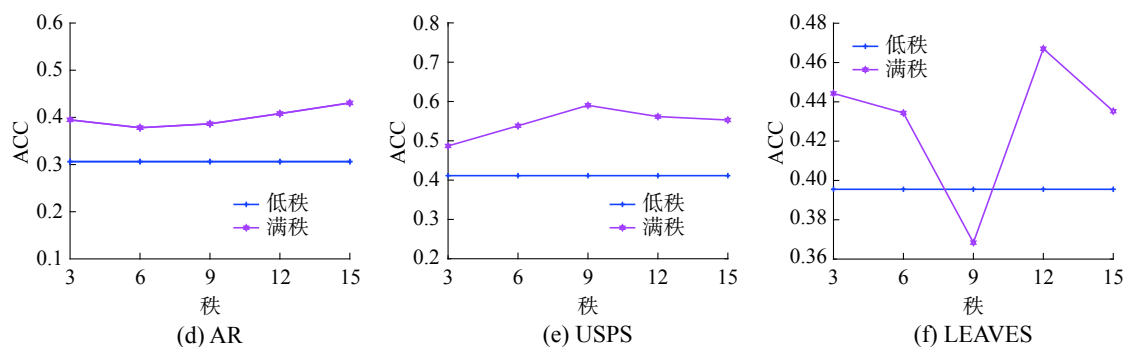


图7 FSRGR在不同秩上的ACC结果

Fig. 7 ACC results of FSRGR at different number of ranks

4 结束语

本文提出了一种新的鲁棒无监督特征选择模型,它将特征自表达、低秩约束、图正则化以及 $L_{2,1}$ 范数正则化项统一在同一框架中,将传统的基于 L_2 范数的图正则化项通过特征分解改为基于 $L_{2,1}$ 范数的正则化项,并对权重矩阵施加了一个低秩约束。该方法能够降低噪声数据对特征选择的影响,并且在保留了数据局部几何结构的同时又很好地保留了数据的全局结构。理论分析和实验结果均表明,与现有的特征选择方法相比,所给出的方法具有更好的效果。

在未来的研究中,将对所提出的框架进行扩展,以对不完整的高维数据进行特征选择,因为在许多实际应用中经常会发现不完整的数据集。

参考文献:

- [1] NIE Feiping, HUANG Heng, CAI Xiao, et al. Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization [C]//Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver, Canada, 2010: 1813–1821.
- [2] ZHAO Jidong, LU Ke, HE Xiaofei. Locality sensitive semi-supervised feature selection[J]. *Neurocomputing*, 2008, 71(10/12): 1842–1849.
- [3] HE Xiaofei, CAI Deng, NIYOGI P. Laplacian score for feature selection[C]//Proceedings of the 18th International Conference on Neural Information Processing Systems. Vancouver, Canada, 2005: 507–514.
- [4] HOU Chenping, NIE Feiping, LI Xuelong, et al. Joint embedding learning and sparse regression: a framework for unsupervised feature selection[J]. *IEEE transactions on cybernetics*, 2014, 44(6): 793–804.
- [5] TABAKHI S, MORADI P, AKHLAGHIAN F. An unsupervised feature selection algorithm based on ant colony optimization[J]. *Engineering applications of artificial intelligence*, 2014, 32: 112–123.
- [6] CAI Deng, ZHANG Chiyuan, HE Xiaofei. Unsupervised feature selection for multi-cluster data[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2010: 333–342.
- [7] ADHIKARY J R, MURTY M N. Feature selection for unsupervised learning[C]//Proceedings of 19th International Conference on Neural Information Processing. Doha, Qatar, 2012: 382–389.
- [8] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016: 1–425.
- [9] LI Hao, WANG Yongli, LI Yanchao, et al. Joint local structure preservation and redundancy minimization for unsupervised feature selection[J]. *Applied intelligence*, 2020, 50(12): 4394–4411.
- [10] LIU Yanfang, YE Dongyi, LI Wenbin, et al. Robust neighborhood embedding for unsupervised feature selection[J]. *Knowledge-based systems*, 2020, 193: 105462.
- [11] LIU Xinwang, WANG Lei, ZHANG Jian, et al. Global and local structure preservation for feature selection[J]. *IEEE transactions on neural networks and learning systems*, 2014, 25(6): 1083–1095.
- [12] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290(5500): 2323–2326.
- [13] HE Xiaofei, NIYOGI P. Locality preserving projections [C]//Proceedings of the 16th International Conference on Neural Information Processing Systems. Whistler, Canada, 2003: 153–160.
- [14] ZHANG Tianhao, YANG Jie, ZHAO Deli, et al. Linear local tangent space alignment and application to face recognition[J]. *Neurocomputing*, 2007, 70(7/8/9): 1547–1553.
- [15] ZHU Pengfei, ZUO Wangmeng, ZHANG Lei, et al. Unsupervised feature selection by regularized self-representation[J]. *Pattern recognition*, 2015, 48(2): 438–446.
- [16] BELHUMEUR P N, HESPAHNA J P, KRIEGMAN D

- J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection[J]. *IEEE transactions on pattern analysis and machine intelligence*, 1997, 19(7): 711–720.
- [17] SAMARIA F S, HARTER A C. Parameterisation of a stochastic model for human face identification[C]//Proceedings of 1994 IEEE Workshop on Applications of Computer Vision. Sarasota, USA, 1994: 138–142.
- [18] NIE Feiping, HUANG Heng. Subspace clustering via new low-rank model with discrete group structure constraint[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York, USA, 2016.
- [19] MARTÍNEZ A M, BENAVENTE R. The AR face database: CVC Technical Report #24[R]. Barcelona, Computer Vision Center, 1998.
- [20] HULL J J. A database for handwritten text recognition research[J]. *IEEE transactions on pattern analysis and machine intelligence*, 1994, 16(5): 550–554.
- [21] DU Jixiang, WANG Xiaofeng, ZHANG Guojun. Leaf shape based plant species recognition[J]. *Applied mathematics and computation*, 2007, 185(2): 883–893.
- [22] LI Zechao, YANG Yi, LIU Jing, et al. Unsupervised feature selection using nonnegative spectral analysis[C]//Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. Toronto, Canada, 2012: 1026–1032.
- [23] NIE Feiping, ZHU Wei, LI Xuelong. Unsupervised feature selection with structured graph optimization[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, Arizona, 2016: 1302–1308.
- [24] TANG Chang, ZHU Xinzhong, CHEN Jiajia, et al. Robust graph regularized unsupervised feature selection[J]. *Expert systems with applications*, 2018, 96: 64–76.
- [25] TANG Chang, BIAN Meiru, LIU Xinwang, et al. Unsupervised feature selection via latent representation learning and manifold regularization[J]. *Neural networks*, 2019, 117: 163–178.

作者简介:



陈彤, 女, 硕士研究生, 主要研究方向为数字图像处理和模式识别。



陈秀宏, 男, 教授, 博士后, 主要研究方向为数字图像处理和模式识别、优化理论与方法。发表学术论文 120 余篇。