智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

联合不相关回归和非负谱分析的无监督特征选择

朱星宇, 陈秀宏

引用本文:
朱星宇,陈秀宏. 联合不相关回归和非负谱分析的无监督特征选择[J]. 智能系统学报, 2022, 17(2): 303–313.
ZHU Xingyu,CHEN Xiuhong. Joint uncorrelated regression and non–negative spectral analysis for unsupervised feature selection[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(2): 303–313.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202012033

DOI: 10.11992/tis.202012033
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20211015.0035.002.html

联合不相关回归和非负谱分析的无监督特征选择

朱星宇[1], 陈秀宏[2]

（1. 江南大学 人工智能与计算机学院, 江苏 无锡 214122; 2. 江南大学 江苏省媒体设计与软件技术重点实验室, 江苏 无锡 214122）

摘　要: 在无标签高维数据普遍存在的数据挖掘和模式识别任务中, 无监督特征选择是必不可少的预处理步骤。然而现有的大多数特征选择方法忽略了数据特征之间的相关性, 选择出具有高冗余、低判别性的特征。本文提出一种基于联合不相关回归和非负谱分析的无监督特征选择方法 (joint uncorrelated regression and nonnegative spectral analysis for unsupervised feature selection), 在选择不相关且具有判别性特征的同时, 自适应动态确定数据之间的相似性关系, 从而能获得更准确的数据结构和标签信息。而且, 模型中广义不相关约束能够避免平凡解, 所以此方法具有不相关回归和非负谱聚类两种特征选择方法的优点。本文还设计出一种求解模型的高效算法, 并在多个数据集上进行了大量实验与分析, 验证模型的优越性。

关键词: 不相关回归; 非负谱分析; 冗余特征; 局部结构学习; 无监督学习; 自适应图; 特征选择; 判别性特征

中图分类号: TP391.4　　**文献标志码:** A　　**文章编号:** 1673−4785(2022)02−0303−11

中文引用格式: 朱星宇, 陈秀宏. 联合不相关回归和非负谱分析的无监督特征选择 [J]. 智能系统学报, 2022, 17(2): 303–313.

英文引用格式: ZHU Xingyu, CHEN Xiuhong. Joint uncorrelated regression and non-negative spectral analysis for unsupervised feature selection[J]. CAAI transactions on intelligent systems, 2022, 17(2): 303–313.

Joint uncorrelated regression and non-negative spectral analysis for unsupervised feature selection

ZHU Xingyu[1], CHEN Xiuhong[2]

(1. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; 2. Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China)

Abstract: Unsupervised feature selection is an essential preprocessing step in the data mining and pattern recognition tasks of unlabeled high-dimensional data. However, most existing feature selection methods ignore the correlation between data features and select features with high redundancy and low discrimination. This paper proposes an unsupervised feature selection method based on joint uncorrelated regression and non-negative spectral analysis (Joint uncorrelated regression and nonnegative spectral analysis for unsupervised feature selection). It adaptively and dynamically determines the similarity relationship between data while selecting uncorrelated and discriminant features, so that more accurate data structure and label information can be obtained. Moreover, the generalized uncorrelated constraints in the model can avoid trivial solutions, so this method has the advantages of two feature selection methods: uncorrelated regression and non-negative spectral clustering. An efficient algorithm for solving the model is also designed, and a large number of experiments and analyses are carried out on multiple data sets to verify the superiority of the model.

Keywords: uncorrelated regression; non-negative spectral analysis; redundant features; local structure learning; unsupervised learning; adaptive graph; feature selection; discriminant feature

收稿日期: 2020−12−18.　**网络出版日期:** 2021−10−15.
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.

在信息时代背景下产生了海量的高维数据, 然而这些数据通常含有大量冗余特征和噪声, 降

低了算法的性能。因此，如何从高维数据中选出最有效的特征即特征选择已成为一项研究热点，特征选择旨在通过去除冗余、不相关和有噪声的特征，找到一组简洁且具有良好泛化能力的特征，由此产生的方法已在机器学习[1]、数据挖掘[2]和生物信息学[3]等领域得到了广泛的应用。基于是否使用标签，特征选择方法可分为有监督学习[4]和无监督学习[5]。

有监督学习依赖于数据的标签信息，并利用它直接指导学习过程。然而，从未标记数据中提取最有判别性的信息是一个具有挑战性的问题，由于无监督特征选择可以根据原始数据的潜在属性来确定特征的重要性，因此近年来越来越受到研究人员的关注。本文主要研究无监督特征选择方法。

基于谱分析的无监督特征选择因其出色的性能引起了广泛关注，本节将回顾一些比较经典的谱特征选择方法。无监督的判别特征选择 ($L_{2,1}$-norm regularized discriminative feature selection for unsupervised learning)[6] 联合使用判别分析和 $L_{2,1}$ 范数进行无监督特征选择，它虽然可以选择判别性的特征，但却忽略了数据间的内在结构。Nie 等[7] 提出了灵活流形嵌入 (flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction) 作为降维的一般框架，非负判别特征选择 (unsupervised feature selection using nonnegative spectral analysis)[8] 则在 FME 框架中结合特征选择和谱聚类学习，它具有鲁棒非负矩阵分解、局部学习和鲁棒特征学习的优势。联合嵌入学习和稀疏回归 (joint embedding learning and sparse regression: a framework for unsupervised feature selection)[9] 是基于 FME 和 $L_{2,1}$ 范数的特征选择方法，它致力于嵌入学习高维样本的低维表示。最近，Nie 等[10] 提出了结构最优图特征选择 (unsupervised feature selection with structured graph optimization)，该方法同时执行特征选择和局部结构学习，同时它可以自适应地确定数据间的相似性矩阵，但过于严格的正交约束选择出的特征会丢失一定程度的判别性，从而降低他们在聚类或分类任务中的性能。自适应图的广义不相关回归 (generalized uncorrelated regression with adaptive graph for unsupervised feature selection)[11] 通过最大熵来自适应地构造相似图，进而同时进行特征选择和谱聚类，但它在学习相似图时采用标签矩阵学习，没有考虑原始数据的类簇结构，这显然不是很合理。

尽管上述无监督特征选择方法已经在各种应用中取得了良好的性能，但仍有以下不足：1) 上述基于谱特征选择的方法构造的相似图是从原始数据得到且是静态的，而现实世界的数据始终包含大量噪声，这使得静态相似矩阵很不可靠，进而破坏局部流形结构；2) 最优相似矩阵应具有精确的 c 个连通分量（c 是类簇数），它能更准确地揭示数据的局部邻域结构；3) 通过传统正交约束选择出的特征虽达到了低冗余的目的，但会丢失一些判别性的特征，这些特征在分类任务中往往能起到关键作用。

综合上述分析，本文联合广义不相关回归、结构化最优图和非负谱分析，提出一个联合不相关回归和非负谱分析的无监督特征选择模型。模型中对相似矩阵增加包含精确的 c 个连通分量的约束，可以用来动态学习结构化最优图，进而更准确地揭示数据之间的局部结构信息；同时通过非负谱聚类可以学习到更真实准确的聚类指标；利用广义不相关约束的正则回归方法选择不相关和判别性的特征，有效避免不相关约束导致的平凡解。在不同数据集上的实验表明所提出的方法是有效的，且该模型在无监督二维图像的特征选择领域，如：医学图像分类、人脸识别、模式识别等计算机视觉任务方面有着广泛的应用价值。

1 相关工作

给定输入数据集 $\boldsymbol{X}=[\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_n]\in\mathbf{R}^{d\times n}$，$\boldsymbol{x}_i$ 表示第 i 个样本，并且这些数据点属于 c 个不同的类簇。记 $\mathrm{tr}(\boldsymbol{M})$ 为矩阵的迹；$\boldsymbol{M}^{\mathrm{T}}$ 为矩阵 $\boldsymbol{M}$ 的转置；$\|\boldsymbol{M}\|_F$ 为矩阵 $\boldsymbol{M}$ 的 F 范数；$\boldsymbol{m}^i$ 表示矩阵 $\boldsymbol{M}$ 的第 i 行向量；$\boldsymbol{m}_j$ 表示矩阵 $\boldsymbol{M}$ 的第 j 列向量；m_{ij} 表示矩阵 $\boldsymbol{M}$ 的第 i 行，第 j 列元素；给定矩阵 $\boldsymbol{M}\in\mathbf{R}^{d\times c}$，它的 $L_{2,1}$ 范数定义为 $\|\boldsymbol{M}\|_{2,1}=\sum_{i=1}^{d}\sqrt{\sum_{j=1}^{c}m_{ij}^2}=\sum_{i=1}^{d}\left\|\boldsymbol{m}^i\right\|_2$，其中 $\left\|\boldsymbol{m}^i\right\|_2$ 表示向量 $\boldsymbol{m}^i$ 的 L_2 范数；$\boldsymbol{I}$ 为单位矩阵；$\boldsymbol{1}=[1\ 1\ \cdots\ 1]^{\mathrm{T}}\in\mathbf{R}^{n\times 1}$。中心矩阵 $\boldsymbol{H}$ 定义为 $\boldsymbol{H}=\boldsymbol{I}-\dfrac{1}{n}\boldsymbol{1}\boldsymbol{1}^{\mathrm{T}}$。

1.1 广义不相关回归模型

无监督特征选择也可以表示为回归模型，从而正则化回归模型也广泛应用在无监督特征选择中。但是，已有工作忽略了特征的不相关性。尤其是当只需少量特征时，需要选择最具判别性的特征。Li 等[11] 提出的以下广义不相关回归模型 (generalized uncorrelated regression model, GURM) 可以获得具有判别性且不相关的流形结构：

$$\min_{W,F,b}\left\|X^{\mathrm{T}}W+\mathbf{1}b^{\mathrm{T}}-F\right\|_F^2+\beta\|W\|_{2,1}\quad \text{s.t.}\quad W^{\mathrm{T}}S_t^G W=I \tag{1}$$

式中: $W\in\mathbf{R}^{d\times c}$是回归矩阵, $b\in\mathbf{R}^{c\times1}$是偏差项, $F\in\mathbf{R}^{n\times c}$是学习的嵌入聚类结构; 正则化项$\|W\|_{2,1}$可使得$W$的行是稀疏的, 进而选择重要的特征; β是正则化参数, β越大, W行越稀疏; $S_t^G=S_t+\beta G$为广义散度矩阵, G为$d\times d$对角矩阵, 其对角元素定义为

$$G_{ii}=\frac{1}{2\sqrt{\|w^i\|_2^2+\varepsilon}},(\varepsilon>0,i=1,2,\cdots,d) \tag{2}$$

这里, $\varepsilon>0$是一个很小的值, 可避免因W的行稀疏而导致的分母为 0 的情形; $S_t=XHX^{\mathrm{T}}$是样本数据的总散度矩阵。一般地, 当样本数量小于数据维度时, S_t是奇异的, 而S_t^G总是正定的, 这样约束条件$W^{\mathrm{T}}S_t^G W=I$可使数据在投影子空间中是统计不相关的, 这样可以很好地保留数据的全局结构。当β很小时, 可使投影后样本之间高度不相关, 因此, 约束条件$W^{\mathrm{T}}S_t^G W=I$在描述样本的不相关性时要优于正交约束$W^{\mathrm{T}}W=I$和传统不相关约束$W^{\mathrm{T}}S_t W=I$。

1.2　局部结构学习

研究表明, 相似性矩阵可用来描述数据间的局部结构。然而, 大多数构造相似性矩阵的方法并没有考虑原始数据中的冗余特征和噪声, 从而导致所学习到的局部结构不够准确, 最终影响特征选择的结果。本节采用一种自适应确定相似度矩阵的方法[12], 可以同时执行特征选择和局部结构学习。

两个数据样本相邻的概率可以用来描述它们之间的相似度, 记相似度矩阵为$S\in\mathbf{R}^{n\times n}$, 其中元素$s_{ij}$表示相似性图中与样本$x_i$和$x_j$相对应的结点之间相连的概率。样本$x_i$和$x_j$越接近, 则对应的连接概率$s_{ij}$越大, 它与对应结点之间的距离成反比。相似度矩阵$S\in\mathbf{R}^{n\times n}$可通过以下优化问题得到:

$$\min\sum_{i,j=1}^{n}\left\|x_i-x_j\right\|_2^2 s_{ij}+\alpha\|S\|_F^2\quad \text{s.t.}\sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant0 \tag{3}$$

这里, 约束条件$\sum_{j=1}^{n}s_{ij}=1$使得相似矩阵S的第i行元素之和等于 1, 表示与x_i对应的总连接概率为 1, s_{ij}衡量样本x_i和x_j之间的局部相似度; $\alpha\geqslant0$是正则化参数, 可以自适应来确定[13]; 正则化项$\|S\|_F^2$可以避免出现平凡解。

1.3　结构化最优图学习

如果学习得到的相似矩阵恰好包含c个连通分量时(即它仅含有c个对角分块), 则每个样本的近邻是最佳的。但是, 问题 (3)(式 (3)) 的解一般不具备此性质, 大多数情况下其解只包含 1 个连通分量[12]。

受文献 [14] 的启发, 对拉普拉斯矩阵$L_S=D_S-\dfrac{S+S^{\mathrm{T}}}{2}$的秩进行约束可使得矩阵$S$具有$c$个连通分量。这时, 问题 (3) 转化为

$$\min\sum_{i,j=1}^{n}\left\|x_i-x_j\right\|_2^2 s_{ij}+\alpha\|S\|_F^2\quad \text{s.t.}\quad \forall i,\ \sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant0,\mathrm{rank}(L_S)=n-c \tag{4}$$

然而, 问题 (4)(式 (4)) 很难直接求解, 因为秩约束是一个复杂的非线性约束。假设半正定矩阵L_S的n个非负特征值依次由小到大排列$\sigma_1(L_S)\leqslant\sigma_2(L_S)\leqslant\cdots\leqslant\sigma_n(L_S)$, 则问题 (4) 可转化为

$$\min\sum_{i,j=1}^{n}\left\|x_i-x_j\right\|_2^2 s_{ij}+\alpha\|S\|_F^2+2\lambda\sum_{i=1}^{c}\sigma_i(L_S)\quad \text{s.t.}\quad \forall i,\sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant0 \tag{5}$$

其中, λ是一个充分大的正数, 使得$\sum_{i=1}^{c}\sigma_i(L_S)=0$且矩阵$L_S$的秩等于$n-c$。根据 KyFan's 理论[15]:

$$\sum_{i=1}^{c}\sigma_i(L_S)=\min_{F\in R^{n\times c},F^{\mathrm{T}}F=I}\mathrm{tr}(F^{\mathrm{T}}L_S F) \tag{6}$$

这里$F\in\mathbf{R}^{n\times c}$是类指标矩阵。故问题 (5)(式 (5)) 可重写为

$$\min\sum_{i,j=1}^{n}\left\|x_i-x_j\right\|_2^2 s_{ij}+\alpha\|S\|_F^2+2\lambda\cdot\mathrm{tr}(F^{\mathrm{T}}L_S F)\quad \text{s.t.}\quad \forall i,\sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant0,\quad F^{\mathrm{T}}F=I,F\geqslant0 \tag{7}$$

正交非负约束可使得F的每一行只有一个元素大于 0, 其他元素都等于 0。从而得到的类指标矩阵F更准确, 且获得更具判别性的信息。此外, 求解问题 (7)(式 (7)) 得到的S包含精确的c个连通分量, 能够捕获更准确的局部结构信息。

2　联合不相关回归和非负谱分析模型 (JURNFS)

2.1　模型建立

根据流形学习的理论, 希望原始空间中样本的近邻关系在低维投影空间中得到保持, 为此, 本文讨论在低维空间内学习最优图。设$W\in\mathbf{R}^{d\times c}$是投影矩阵, 则由 (7) 可得到以下模型:

$$\min \sum_{i,j=1}^{n} \left\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_j \right\|_2^2 s_{ij} + \alpha \|\boldsymbol{S}\|_F^2 + 2\lambda \cdot \mathrm{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}_S\boldsymbol{F}) \\ \text{s.t.} \quad \forall i, \sum_{j=1}^{n} s_{ij} = 1, s_{ij} \geqslant 0, \quad \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} = \boldsymbol{I}, \boldsymbol{F} \geqslant 0 \tag{8}$$

将模型 (8)(式 (8)) 与稀疏回归模型 (1) 相结合，得到本文联合不相关回归和非负谱分析模型：

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{b},\boldsymbol{S}} \left\| \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} + \boldsymbol{1}\boldsymbol{b}^{\mathrm{T}} - \boldsymbol{F} \right\|_F^2 + \sum_{i,j=1}^{n} \left\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_j \right\|_2 s_{ij} + \\ \alpha \|\boldsymbol{S}\|_F^2 + \beta \|\boldsymbol{W}\|_{2,1} + 2\lambda \cdot \mathrm{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}_S\boldsymbol{F}) \\ \text{s.t.} \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} = \boldsymbol{I}, \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} = \boldsymbol{I}, \boldsymbol{F} \geqslant 0, \\ \forall i, \sum_{j=1}^{n} s_{ij} = 1, s_{ij} \geqslant 0 \tag{9}$$

其中，α、β、λ是正则化参数。在式 (9) 中，第 1 项、第 4 项和第 5 项刻画了无监督不相关回归和非负谱分析，用于学习稀疏投影矩阵和预测标签矩阵，且$L_{2,1}$范数可使得$\boldsymbol{W}$保持行稀疏，从而能够选择出更具有价值和判别性的特征；第 2 项和第 3 项用于学习数据的局部结构，确保原始空间内数据的相似结构在投影子空间内得到保持。这里第 2 项未采用L_2范数距离的平方，这是考虑到若原始数据中有噪声，平方会扩大噪声对模型学习与分类的影响。

令问题 (9)(式 (9)) 关于$\boldsymbol{b}$的拉格朗日函数为

$$\Omega_b(\boldsymbol{b}) = \left\| \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} + \boldsymbol{1}\boldsymbol{b}^{\mathrm{T}} - \boldsymbol{F} \right\|_F^2 + \Gamma(\boldsymbol{W},\boldsymbol{F},\boldsymbol{S}) \tag{10}$$

其中，$\Gamma(\boldsymbol{W},\boldsymbol{F},\boldsymbol{S})$表示式 (9) 中依赖于$\boldsymbol{W}$、$\boldsymbol{F}$、$\boldsymbol{S}$但又独立于$\boldsymbol{b}$的项。取$\Omega_b(\boldsymbol{b})$关于$\boldsymbol{b}$的导数并令其等于 0，得

$$\boldsymbol{b} = \frac{1}{n}(\boldsymbol{F}^{\mathrm{T}} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X})\boldsymbol{1} \tag{11}$$

在实际应用中，数据结构总是多模态的，为了在多模态数据上获得更好的性能，可以研究图的局部性。假设$\boldsymbol{S}$的每一行有一个具体的α_i[13]，再结合式 (11)，问题 (9) 可重写为

$$\min_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{S}} \left\| \boldsymbol{H}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} - \boldsymbol{F}) \right\|_F^2 + \sum_{i,j=1}^{n} (\left\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_j \right\|_2 s_{ij} + \\ \alpha_i s_{ij}^2) + \beta \|\boldsymbol{W}\|_{2,1} + 2\lambda \cdot \mathrm{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}_S\boldsymbol{F}) \\ \text{s.t.} \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} = \boldsymbol{I}, \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} = \boldsymbol{I}, \boldsymbol{F} \geqslant 0, \\ \forall i, \sum_{j=1}^{n} s_{ij} = 1, s_{ij} \geqslant 0 \tag{12}$$

此时，参数α_i可以控制每个样本自适应近邻的数量[12]。

2.2 模型求解

由于式 (12) 中的目标函数是凸的，故它有全局最优解，但直接求其全局解比较困难。本节给出一种交替优化方法来迭代求解它。

1) 固定$\boldsymbol{F}$和$\boldsymbol{S}$，更新$\boldsymbol{W}$

此时，问题 (12)(式 (12)) 等价于以下问题：

$$\min_{\boldsymbol{W}} \left\| \boldsymbol{H}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} - \boldsymbol{F}) \right\|_F^2 + \sum_{i,j=1}^{n} \left\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_j \right\|_2 s_{ij} + \beta \|\boldsymbol{W}\|_{2,1} \\ \text{s.t.} \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} = \boldsymbol{I} \tag{13}$$

令$d_{ij} = \dfrac{1}{2\left\| \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_i - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}_j \right\|_2}$，并定义矩阵$\tilde{\boldsymbol{S}}$中的元素为$\tilde{s}_{ij} = d_{ij}s_{ij}$，则问题 (13)(式 (13)) 可等价于以下问题：

$$\min_{\boldsymbol{W}} \mathrm{tr}(\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} - 2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{F} + \boldsymbol{F}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{F}) + \\ \mathrm{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}) \\ \text{s.t.} \quad \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} = \boldsymbol{I} \tag{14}$$

式中：图拉普拉斯矩阵$\tilde{\boldsymbol{L}} = \tilde{\boldsymbol{D}} - \tilde{\boldsymbol{S}}$，对角矩阵$\tilde{\boldsymbol{D}}$的对角元素为$\tilde{D}_{ii} = \sum_{j=1}^{n} \tilde{S}_{ij}$。再由约束$\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} = \boldsymbol{I}$以及固定的$\boldsymbol{F}$，问题 (14)(式 (14)) 又可转化为

$$\min_{\boldsymbol{W}} \mathrm{tr}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W} - 2\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{F}) \\ \text{s.t.} \quad \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{S}_t + \beta\boldsymbol{G})\boldsymbol{W} = \boldsymbol{I} \tag{15}$$

该问题将不相关性表现为流形结构，且具有闭式解。问题 (15)(式 (15)) 可表示为

$$\min_{\bar{\boldsymbol{W}}} \mathrm{tr}(\bar{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{A}\bar{\boldsymbol{W}} - 2\bar{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{B}) \\ \text{s.t.} \quad \bar{\boldsymbol{W}}^{\mathrm{T}}\bar{\boldsymbol{W}} = \boldsymbol{I} \tag{16}$$

其中，

$$\bar{\boldsymbol{W}} = \sqrt{\boldsymbol{S}_t + \beta\boldsymbol{G}}\boldsymbol{W}, \quad \boldsymbol{A} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{X}\tilde{\boldsymbol{L}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}, \\ \boldsymbol{B} = \boldsymbol{C}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{H}\boldsymbol{F}, \quad \boldsymbol{C} = \left(\sqrt{\boldsymbol{S}_t + \beta\boldsymbol{G}}\right)^{-1} \tag{17}$$

问题 (16)(式 (16)) 可以利用广义幂迭代方法[16]求解，详细过程见算法 1。

算法 1 求解问题 (14)

输入 数据矩阵$\boldsymbol{X} \in \mathbf{R}^{d \times n}$，总散度矩阵$\boldsymbol{S}_t \in \mathbf{R}^{d \times d}$，类指标矩阵$\boldsymbol{F} \in \mathbf{R}^{n \times c}$，对角矩阵$\boldsymbol{G} \in \mathbf{R}^{d \times d}$，图拉普拉斯矩阵$\tilde{\boldsymbol{L}} \in \mathbf{R}^{n \times n}$。

输出 投影矩阵$\bar{\boldsymbol{W}} \in \mathbf{R}^{d \times c}$。

初始化 满足$\bar{\boldsymbol{W}}^{\mathrm{T}}\bar{\boldsymbol{W}} = \boldsymbol{I}$的随机矩阵$\bar{\boldsymbol{W}} \in \mathbf{R}^{d \times c}$，根据等式 (17) 计算$\boldsymbol{A} \in \mathbf{R}^{d \times d}$和$\boldsymbol{B} \in \mathbf{R}^{d \times c}$，定义随机$v$，通过幂方法使得$\tilde{\boldsymbol{A}} = v\boldsymbol{I} - \boldsymbol{A} \in \mathbf{R}^{d \times d}$为正定矩阵。

重复：

①更新$\mathbf{R} \leftarrow 2\tilde{\boldsymbol{A}}\bar{\boldsymbol{W}} + 2\boldsymbol{B}$；

②设矩阵$\boldsymbol{R}$的满 SVD 分解为$\mathbf{R} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$，则更新$\bar{\boldsymbol{W}} \leftarrow \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$；

直到收敛。

当求得式 (16) 的最优解之后，问题 (14) 的最优解$\boldsymbol{W}$可通过下述公式获得：

$$\boldsymbol{W} = \boldsymbol{C}\bar{\boldsymbol{W}} \tag{18}$$

2) 固定$\boldsymbol{W}$和$\boldsymbol{S}$，更新$\boldsymbol{F}$

此时，问题 (12) 等价于：

$$\min_{F}\operatorname{tr}(F^{\mathrm{T}}EF)-2\operatorname{tr}(F^{\mathrm{T}}M)\quad \text{s.t.}\quad F^{\mathrm{T}}F=I,F\geqslant 0 \tag{19}$$

其中, $E=H+2\lambda L_S$, $M=HX^{\mathrm{T}}W$。

此问题可采用乘性更新规则[17]来求解。首先考虑它的松弛问题:

$$\min_{F}\operatorname{tr}(F^{\mathrm{T}}EF)-2\operatorname{tr}(F^{\mathrm{T}}M)+\frac{\gamma}{2}\left\|F^{\mathrm{T}}F-I\right\|_F^2\quad \text{s.t.}\quad F\geqslant 0 \tag{20}$$

其中, γ为充分大的正数, 由 KKT 最优性条件得

$$2EF-2M+\gamma F(F^{\mathrm{T}}F-I_c)-\Phi=0 \tag{21}$$

$$\varphi_{ij}F_{ij}=0 \tag{22}$$

这里, φ_{ij}为对应于约束$F_{ij}\geqslant 0$的非负拉格朗日乘子。于是有以下更新规则:

$$F_{ij}\leftarrow F_{ij}\frac{(M+\gamma F)_{ij}}{(EF+\gamma FF^{\mathrm{T}}F)_{ij}} \tag{23}$$

再对 F 的每一列归一化使得它满足条件$(F^{\mathrm{T}}F)_{ii}=1, i=1,2,\cdots,n$。

3) 固定W和F, 更新S

此时, 由于$2\operatorname{tr}(F^{\mathrm{T}}L_SF)=\sum_{i,j=1}^{n}\left\|f^i-f^j\right\|_2^2 s_{ij}$, 故问题 (12) 等价于以下问题:

$$\begin{gathered}\min_{S}\sum_{i,j=1}^{n}\left(\left\|W^{\mathrm{T}}x_i-W^{\mathrm{T}}x_j\right\|_2 s_{ij}+\alpha_i s_{ij}^2\right)+\lambda\sum_{i,j=1}^{n}\left\|f^i-f^j\right\|_2^2 s_{ij}\\ \text{s.t.}\quad \forall i,\sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant 0\end{gathered} \tag{24}$$

其中, f^i和f^j分别为F的第 i 行和第 j 行。该问题可分解为以下 n 个独立的子问题:

$$\begin{gathered}\min_{s_{ij}}\sum_{j=1}^{n}\left(\left\|W^{\mathrm{T}}x_i-W^{\mathrm{T}}x_j\right\|_2 s_{ij}+\alpha_i s_{ij}^2\right)+\lambda\sum_{j=1}^{n}\left\|f^i-f^j\right\|_2^2 s_{ij}\\ \text{s.t.}\quad \sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant 0\end{gathered} \tag{25}$$

令 $e_{ij}=\left\|W^{\mathrm{T}}x_i-W^{\mathrm{T}}x_j\right\|_2$, $f_{ij}\left\|f^i-f^j\right\|_2^2$, 则式 (25) 转化为

$$\begin{gathered}\min_{s^i}\left\|s^i+\frac{m^i}{2a_i}\right\|_2^2\\ \text{s.t.}\quad \sum_{j=1}^{n}s_{ij}=1,s_{ij}\geqslant 0\end{gathered} \tag{26}$$

其中, $m^i=(e_{il}+\lambda f_{il},\cdots,e_{in}+\lambda f_{in})$。

该问题可利用文献 [18] 中的方法求解, 并可自适应地确定式 (9) 中参数α[13], 进而获得具有精确 k 个非零分量的最优解s^i。

以上 3 步可迭代地进行, 直到目标函数收敛或满足终止条件, 整个过程概括为算法 2。

算法 2　JURNFS

输入　数据矩阵$X\in\mathbf{R}^{d\times n}$, 聚类数量 c, 总散度矩阵$S_t\in\mathbf{R}^{d\times d}$, 正则化参数$\alpha$和$\beta$及足够大的参数$\lambda$。

输出　计算$\left\|w^i\right\|_2(i=1,2,\cdots,d)$并按照降序排序, 然后选择前$f$个排好序的特征作为特征选择的结果。

初始化　通过 K-means 聚类初始化矩阵F, 定义随机矩阵W, 通过求解问题 (3) 初始化相似矩阵S, 并计算图拉普拉斯矩阵$L_S\in\mathbf{R}^{n\times n}$和$\tilde{L}\in\mathbf{R}^{n\times n}$, 令$G=I_{d\times d}$。

重复:

①由算法 1 求解式 (16) 得$\bar{W}$, 并由式 (18) 更新W;

②根据式 (2) 更新G;

③由式 (23) 更新F;

④求解 (26) 更新S;

⑤更新$L_S=D_S-\dfrac{S+S^{\mathrm{T}}}{2}$;

⑥更新矩阵$\tilde{S}$及对应的拉普拉斯矩阵$\tilde{L}=\tilde{D}-\tilde{S}$;

直到收敛。

2.3　时间复杂度分析

在以上算法中, 主要的计算复杂度为步骤①中的奇异值分解和矩阵求逆, 故本算法时间复杂度最高为$O(n^2d)$, 假设算法迭代T次, 则该部分的时间复杂度为$O(Tn^2d)$, 从而整个算法的时间复杂度为$O(Tn^2d)$。

3　实验及分析

3.1　实验方案

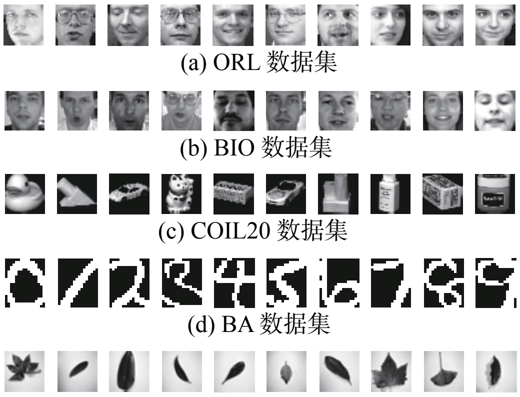
在本节中, 通过进行大量实验以充分验证本文所提出方法的有效性和优越性。在展示结果之前, 首先提供一些详细的实验方案。

1) 数据集: 实验中使用了 6 个公共数据集, 包括 2 个人脸数据集 ORL[19] 和 BIO[20], 1 个物体数据集 COIL20[21], 1 个手写字数据集 BA[22], 1 个树叶数据集 LEAVES[23] 以及 1 个生物学数据集 LUNG[24]。数据集的详细信息见表 1 及图 1 所示。

表 1　数据集信息

Table 1　Detail introduction to datasets

数据集	样本数	特征数	类别数	选择特征数
ORL	400	1 024	40	50、100、…、300
BIO	1 460	1 024	22	50、100、…、300
COIL20	1 440	1 024	20	20、40、…、120
BA	1 404	320	36	20、40、…、120
LEAVES	400	1 024	10	50、100、…、300
LUNG	203	3 312	5	100、150、…、350



(a) ORL 数据集

(b) BIO 数据集

(c) COIL20 数据集

(d) BA 数据集

(e) LEAVES 数据集

图 1　部分数据集的图片

Fig. 1　Visualization of some datasets

2) 对比算法：实验中将本文所提出的联合不相关回归和非负谱分析模型 (joint uncorrelated regression and nonnegative spectral analysis for unsupervised feature selection, JURNFS) 与 5 个特征选择方法进行了比较，分别是：无监督判别特征选择 ($L_{2,1}$-norm regularized discriminative feature selection for unsupervised learning, UDFS)[6]、非负判别性特征选择 (unsupervised feature selection using nonnegative spectral analysis,NDFS)[8]、联合嵌入学习和稀疏回归 (joint embedding learning and sparse regression: A framework for unsupervised feature selection,JELSR)[9]、最优结构图的特征选择 (unsupervised feature selection with structured graph optimization,SOGFS)[10] 和用于无监督特征选择的自适应图的广义不相关回归 (generalized uncorrelated regression with adaptive graph for unsupervised feature selection,URAFS)[11]。

3) 评价指标：为了验证所选特征的优劣性，本文利用两个度量标准来衡量每种算法的性能，一个是识别精度 (ACC)[25]，定义：

$$\mathrm{ACC}=\sum_{i=1}^{n}\delta(y_i,\hat{y}_i)/n$$

式中：y_i是每一个样本的真实标签；$\hat{y}_i$是对应的预测标签；n 是测试样本的数量；函数$\delta(\cdot,\cdot)$衡量两个输入参数之间的关系，如果两个参数相等则等于 1，否则等于 0，聚类精度越高，说明算法的效果越好；另一个评价指标是标准化互信 (NMI)[26]，定义：

$$\mathrm{NMI}=\mathrm{MI}(C,C')/\max(\varGamma(C),\varGamma(C')),$$

式中：C是真实标签的集合；C'是预测标签的集合；$\mathrm{MI}(C,C')$是互信息指标；$\varGamma(C)$、$\varGamma(C')$分别是C和C'的熵。NMI 越大，表示算法性能越稳定。

4) 参数设置：本文通过网格搜索来确定每个算法的最优参数，所有算法的参数都是从$\{10^{-4}, 10^{-3}, 10^{-2}, \cdots, 10^{2}, 10^{3}, 10^{4}\}$中选取。此外，在聚类实验中，本文采用流行的 K-means 聚类方法，用于对具有选定特征形成的新数据进行聚类，每个数据集的选择特征数在表 1 中列出。为减少 K-means 中随机起点触发的偶然性影响，对每种算法进行 10 次聚类，并计算平均值。对于分类实验，本文使用 1-最近邻 (1-NN) 分类器对选定特征的测试图像数据进行分类。为了减少偏差，所有实验均重复运行 10 次以随机选择训练样本，最后计算平均分类结果。另外，K-means 聚类方法中的 K 值设置为 c，并且将每种方法中子空间的维数也都设置为 c，除 SOGFS 的子空间维数设置为 $d/2$。在 UDFS、NDFS、JELSR、SOGFS 和 JURNFS 中，最近邻 k 的数量设置为 5。

3.2　聚类性能与分析

本节将不同方法的聚类精度进行比较，实验中将每个数据集的所有样本都用作训练集。首先，每个算法在每个数据集上进行学习以选择重要且具有判别性的特征，不同数据集所选特征数也不一样，具体细节见表 1；然后，本文使用 K-means 聚类算法对由这些选择出的特征所形成的新样本进行聚类实验。图 2 给出了 6 个算法在 6 个数据集上的聚类精度，从图 2 可以看出，在大多数情况下，本文所提出的 JURNFS 相比其他算法都取得了比较好的效果，这证明了 JURNFS 的优越性。

此外，1) 从图 2(a) 和 (b) 可知，JURNFS 在人脸数据集上的效果远远领先于其他 5 个算法，并且 JURNFS 在 ORL 和 BIO 上相对于其他算法分别平均提升了约 3.28% 和 4.75%，这是因为 JURNFS 通过学习数据的局部流形结构选择出了人脸上更具有判别性的特征，这些特征能够显著地代表原数据，因而获得了较高的聚类精度。2) 所有算法在 COIL20 和 BA 数据集上的聚类效果都比较接近，这可能是由于这两个数据集里的样本特征区分度信息不是很高，所有算法学习到了近似的局部结构，而 JURNFS 中采用的广义不相关约束在保留低冗余度特征的同时维持了对判别性部分特征的关注（如 COIL20 中物体的轮廓商标部分；BA 中文字笔画拐弯部分），所以 JURNFS 的聚类效果仍能够优于其他算法。3) 在 LEAVES 和 LUNG 数据集上，所有算法的聚类效果随着所选特征数而波动。这也证明并不是选择的特征数越多，聚类效果越好。因为特征越多，冗余的特征也可能越多，因此有必要进行特征选择。而 JURNFS 采用低维流形结构化最优图学习和广义不相关回归联合学习的方式保留了判别性局部结构信息（如 LEAVES 中叶的根茎部，LUNG 中的病灶部分），因此在处理非人脸数据集时仍能具有较好的性能。此外，可以发现 LUNG 数据集的

数据量较大, 标签类数只有 5, 因此每一类标签的训练数据集较大, 这也可能是 JURNFS 优于其他算法的原因, 因为在这种情况下 JURNFS 可以选择更准确、更判别性的特征。总之, 通过特征选择获得的精炼的数据包含更多有价值的信息, 而 JURNFS 通过广义不相关回归以及结构化最优图, 所选择的特征更具判别性及有效性, 因而可以获得更好的性能。

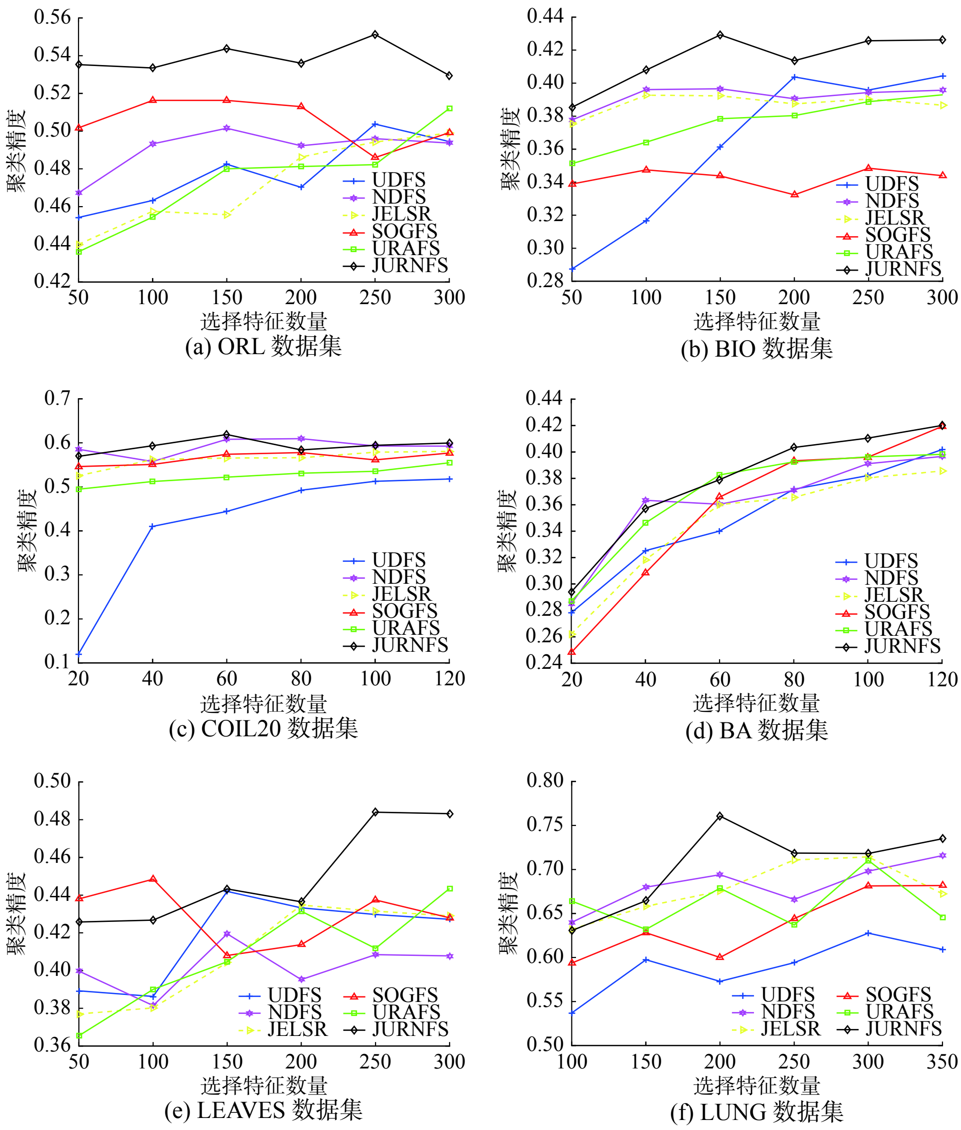


图 2　6 个算法在 6 个数据集上的聚类精度

Fig. 2　Clustering accuracies of six methods on six different datasets

为了进一步说明 JURNFS 的优越性, 表 2 给出了在 6 个不同的数据集上 6 种不同算法的标准偏差的最优 NMI 值, 其中最优值以黑体加粗。一般而言, NMI 值越高, 特征选择的性能越好。显然, 与其他特征选择算法相比, JURNFS 的 NMI 相对较高, 这也表明 JURNFS 具有更好的算法性能。

表 2　不同数据集上不同方法的最佳 NMI (标准偏差)

Table 2　Best NMI with standard deviation of different methods on different data sets　　%

NMI±STD	ORL	BIO	COIL20	BA	LEAVES	LUNG
UDFS	70.04±0.99	55.33±1.06	63.30±2.25	55.90±0.63	51.28±4.14	46.43±2.83
NDFS	71.26±0.88	52.27±1.02	74.00±1.07	55.68±0.93	48.49±2.95	51.83±8.62
JELSR	70.91±1.50	50.98±1.12	71.89±1.50	54.56±1.05	49.53±2.49	52.82±5.72
SOGFS	72.76±1.14	47.98±1.61	70.65±2.82	57.08±1.32	43.49±1.68	49.78±6.60
URAFS	71.69±2.30	52.58±0.85	70.57±1.83	55.93±1.28	48.43±2.95	51.79±5.51
JURNFS	**75.08**±1.64	**55.78**±1.70	**74.04**±1.40	**57.73**±1.05	**53.85**±3.25	**53.97**±5.86

3.3　分类性能与分析

本节将不同方法的分类精度进行比较，首先从不同数据集上的每个类中随机选取适当数量的样本作为训练样本，其余的作为测试样本。对于 ORL、LEAVES 和 LUNG 数据集，随机选取 5 个图像样本作为每类的训练样本，剩余的作为测试样本。对于 BIO、BA 和 COIL20 数据集，随机选取 10 个图像样本作为每个类的训练样本，剩余的作为测试样本。此外，不同的数据集在实验中选取的特征数量也不同 (见表 1)。

图 3 描述了不同特征选择方法的分类精度随所选特征数量的增加而变化的情况。由该图可以发现 JURNFS 在所有数据集上都能取得很好的分类效果，这得益于 JURNFS 联合使用广义不相关回归和非负谱分析，在保留全局结构的同时自适应进行局部结构学习，很好地保留了特征的判别性信息，这些判别性特征在后续进行分类任务时起到了关键作用。

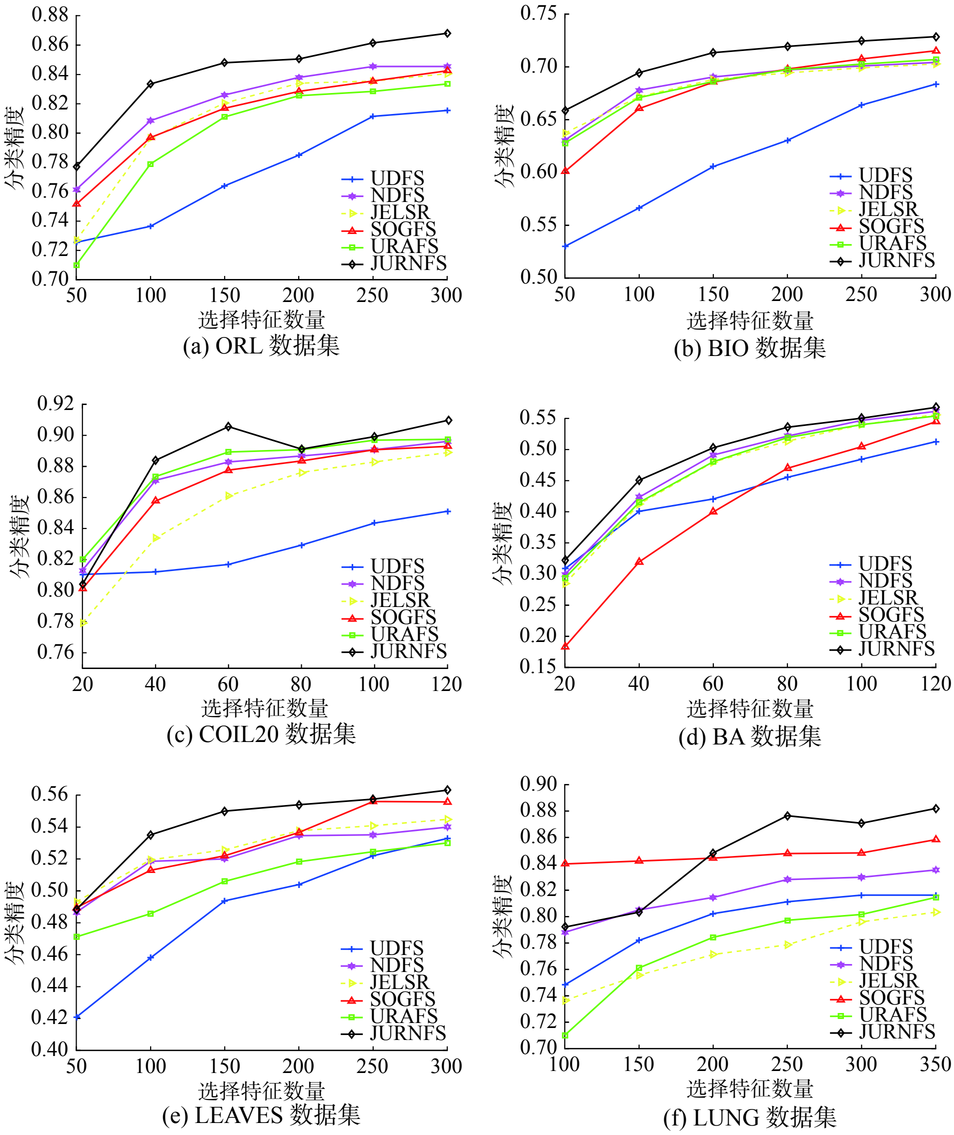


图 3　6 个算法在 6 个数据集上的分类精度

Fig. 3　Classification accuracies of six methods on six different datasets

此外，图 4 通过重构图展示了部分算法在 ORL 数据集上选择的特征，不难看出：1) 随着选择特征数的变化，JURNFS 可以选择信息量较大的特征 (如眼睛、鼻子、嘴巴)，并且 JURNFS 在选择特征数较少的情况下会优先选择具有判别性的特征；2) 对于 SOGFS，尽管它在聚类任务中表现良好，但它选择的特征主要分布在皮肤区域，而不是人脸的区分器官特征。这表明边缘和轮廓上的特征有助于提高这些方法的聚类性能。

3.4　算法参数与性能分析

本节将分析 JURNFS 中参数的稳定性。在所提出的模型 (9) 中，有 3 个参数：α、β、λ。其中，正

则化参数 α 用于保证在学习相似矩阵 $\boldsymbol{S}$ 时, 每个样本都有机会成为 $\boldsymbol{x}_i$ 的近邻; 正则化参数 β 用于调整投影矩阵 $\boldsymbol{W}$ 的稀疏性; 正则化参数 λ 确保学习到的样本标签足够平滑。因此, 参数 α、β、λ 共同调节回归与特征选择之间的效果。在实验中, 可以根据文献 [15] 中的方法自适应确定参数 α, 因此, 本节将专注于研究参数 β 和 λ 对模型的影响。这里只给出 JURNFS 在数据集 ORL、BIO 和 COIL20 上的不同参数组合变化图。

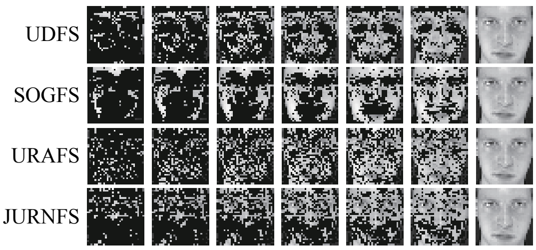
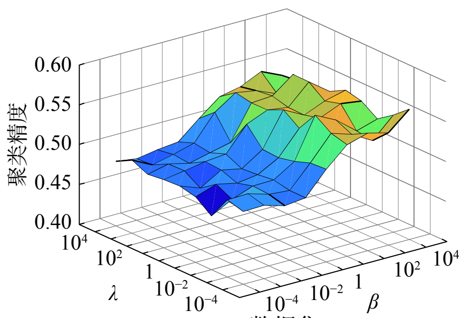


图 4　6 个算法在 ORL 数据集上的重构图, 所选特征个数从左到右依次为 {150, 250, 350, 450, 550, 650, 1 024}

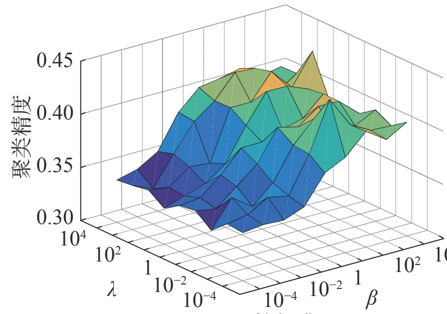
Fig. 4　Reconstruction graph of 6 algorithms on ORL data set, the number of selected features from left to right is {150, 250, 350, 450, 550, 650, 1 024}

观察图 5 可知, 在 ORL 和 BIO 数据集上, JURNFS 受 λ 的影响较小, 而随着 β 的增大, JURNFS 的聚类效果显著提高。这可能是由于在人脸数据集上, JURNFS 经过有限次迭代学习到的标签矩阵接近于真实标签, 故 λ 对目标函数的影响不大, 而随着 β 增大, 使得 $\boldsymbol{W}$ 更加稀疏, 因而选择出的特征更加具有判别性及代表性, 故聚类效果得到提升。在 COIL20 数据集上, JURNFS 的聚类效果随着 β 和 λ 的变化而波动, 这表明此时 β 和 λ 同时影响着 JURNFS 的聚类效果, 并且随着 β 的增大, JURNFS 的聚类精度会呈现出一个先增加后减少的趋势。因此在实验中, 对于不同数据集仍需要找到一组合适的参数。

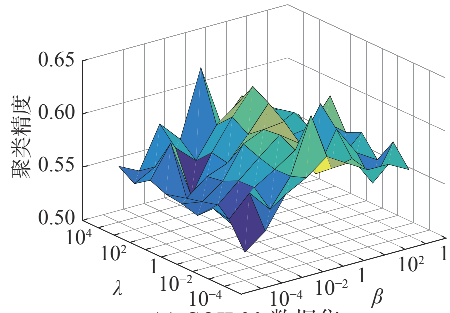
为验证 JURNFS 算法的收敛性, 本文考虑在数据集 ORL 和 BA 上验证式 (9) 中目标函数随迭代次数增加的变化情况。从图 6 可见, 随着迭代次数增加, 目标函数值单调减少, 且在 5 次迭代后即可收敛, 这也验证了 JURNFS 的高效性。



(a) ORL 数据集



(b) BIO 数据集



(c) COIL20 数据集

图 5　JURNFS 在不同数据集上不同参数组合的聚类精度

Fig. 5　Clustering accuracies of JURNFS with different parameter combinations on different dataset, respectively

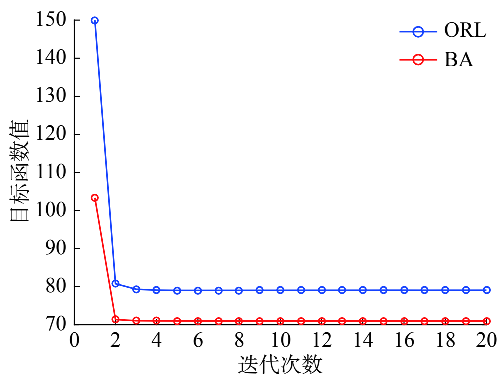


图 6　JURNFS 在不同数据集上的收敛性

Fig. 6　Convergence behaviors of JURNFS on different datasets

此外, 本文还通过算法的运行时间来评估 JURNFS 的性能。本文将所有算法在 ORL、BIO、COIL20 和 BA 数据集上的聚类时间进行比较。不同算法的运行时间见表 3。可以看到, JURNFS 的运行时间与其他算法相比仍具有很好的竞争力。

表 3　不同方法的运行时间

Table 3　Computational time of different methods　　s

算法	ORL	BIO	COIL20	BA
UDFS	8.752	21.190	20.191	5.715
NDFS	3.919	14.339	15.209	7.213
JELSR	3.167	20.665	22.618	18.578
SOGFS	12.275	16.779	17.968	12.589
URAFS	5.436	13.217	14.344	9.357
JURNFS	2.443	8.616	7.693	3.851

4 结束语

本文提出了一种新颖的无监督特征选择方法，将广义不相关回归、结构化最优图和非负谱分析结合，通过广义不相关回归模型选择不相关和判别性的特征，利用结构化最优图自适应学习数据间的局部结构信息，同时结合非负谱聚类来学习局部流形中的标签矩阵，这使得所学习到的标签矩阵更真实可靠。文中还提出了一个有效的迭代算法来求解提出的模型，并在真实数据集上进行了大量实验证明了该方法的有效性和优越性。后续的工作主要集中在以下 2 个方面：1) 提高算法在不同噪声(如光照、遮挡等)水平下的鲁棒性；2) 提高算法处理高维数据的速度。

参考文献：

[1] BLUM A L, LANGLEY P. Selection of relevant features and examples in machine learning[J]. Artificial Intelligence, 1997, 97(1-2): 245–271.

[2] LIU H, MOTODA H. Feature selection for knowledge discovery and data mining[M]. Boston: Kluwer Academic Publishers, 1998: 1−214.

[3] SAEYS Y, INZA I, LARRANAGA P. A review of feature selection techniques in bioinformatics[J]. Bioinformatics (Oxford, England), 2007, 23(19): 2507–2517.

[4] ZHANG R, NIE F P, LI X L. Self-weighted supervised discriminative feature selection[J]. IEEE transactions on neural networks and learning systems, 2018, 29(8): 3913–3918.

[5] NIE F P, HUANG H, CAI X, et al. Efficient and robust feature selection via joint $L_{2,1}$ -norms minimization [C]//International Conference on Neural Information Processing Systems. Vancouver, Canada, 2010: 1813−1821.

[6] 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. Barcelona, Spain, 2011: 1589–1594

[7] NIE F P, XU D, TSANG I W, et al. Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction[J]. IEEE transactions on image processing, 2010, 19(7): 1921–1932.

[8] LI Z C, YANG Y, LIU J, et al. Unsupervised feature selection using nonnegative spectral analysis[C]// Proceedings of the National Conference on Artificial Intelligence. Toronto, Canada, 2012: 1026−1032.

[9] HOU C P, NIE F P, Li X L, et al. Joint embedding learning and sparse regression: A framework for unsupervised feature selection[J]. IEEE transactions on systems, man, and cybernetics, 2014, 44(6): 793–804.

[10] NIE F P, ZHU W, LI X L. Unsupervised feature selection with structured graph optimization[C]//Proceedings of the National Conference on Artificial Intelligence. Phoenix, USA, 2016: 1302–1308.

[11] 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.

[12] NIE F P, WANG X Q, HUANG H. Clustering and projected clustering with adaptive neighbors[C]//knowledge discovery and data mining. New York, USA, 2014: 977−986.

[13] NIE F P, WU D Y, WANG R, et al. Self-weighted clustering with adaptive neighbors[J]. IEEE transactions on neural networks and learning systems, 2020, 31(9): 3428–3441.

[14] JERIBI A. Spectral graph theory[M]. Providence: Published for the Conference Board of the mathematical sciences by the American Mathematical Society, 1997.

[15] FAN K. On a theorem of weyl concerning eigenvalues of linear transformations ii[J]. Proceedings of the National Academy of Sciences, 1949, 35(11): 652–655.

[16] NIE F P, ZHANG R, LI X L, et al. A generalized power iteration method for solving quadratic problem on the stiefel manifold[J]. Science China information sciences, 2017, 60(11): 142–151.

[17] LEE D D, SEUNG H S. Algorithms for non-negative matrix factorization[C]//International Conference on Neural Information Processing Systems. Cambridge, USA, 2000: 556−562.

[18] HUANG J, NIE F P, HUANG H. A new simplex sparse learning model to measure data similarity for clustering [C]//International Conference on Artificial Intelligence. Buenos Aires, Argentina, 2015: 3569–3575.

[19] SAMARIA F S, HARTER A. 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.

[20] JESORSKY O, KIRCHBERG K J, FRISCHHOLZ R, et al. Robust face detection using the hausdorff distance[J]. Lecture notes in computer science, 2001, 2091: 90–95.

[21] NENE S A, NAYAR S K, MURASE H. Columbia Object Image Library (COIL-20)[R]. CUCS-005-96, Department of Computer Science, Columbia University, 1996(CUCS-005-96).

[22] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J, et al. Eigenfaces vs. fisher faces: recognition using class specific linear projection[J]. IEEE transactions on

pattern analysis and machine intelligence, 1997, 19(7): 711–720.
[23] MALLAH C D, ORWELL J. Probabilistic classification from a K-Nearest-Neighbour classifier[J]. Computational research, 2013, 1(1): 1–9.
[24] SINGH D, FEBBO P G, ROSS K N, et al. Gene expression correlates of clinical prostate cancer behavior[J]. Cancer cell, 2002, 1(2): 203–209.
[25] PAPADIMITRIOU C H, STEIGLITZ K. Combinatorial Optimization: algorithms and complexity[J]. IEEE transactions on acoustics speech and signal processing, 1982, 32(6): 1258–1259.
[26] HE X F, CAI D, NIYOGI P. Laplacian score for feature selection[C]//Advances in Neural Information Processing Systems. Vancouver, Canada, 2005: 507-514.

作者简介:

朱星宇, 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, 教授, 博士后, 主要研究方向为数字图像处理和模式识别。发表学术论文 120 余篇。

第十一届中国智能产业高峰论坛
The 11th China Intelligence Industry Summit (2022)

4 月 29—30 日, 2022 第十一届中国智能产业高峰论坛(简称 CIIS 2022)将在厦门举办, 以多位中外院士为代表的人工智能及相关领域的科学家、工程师、企业家及传统行业代表将共聚一堂, 探讨智能产业融合发展的技术创新动态与商业应用趋势, 建言区域智能生态建设和数字经济崛起。

智“汇”产学融合

本次论坛由戴琼海院士、王恩东院士、赵春江院士三位专家担任主席, 依旧延续阵容强大、成果新颖、议题丰富的办会特色, 注重产学融合、生态协同的共创效应, 将集中分享最新的学术研究结晶、技术实践路径、行业解决方案, 以及多个热门产业发展方向的思考和观点。

备受瞩目的主题报告环节, 戴琼海院士、赵春江院士、何友院士、桂卫华院士、张勤院士、周伯文博士、周斌博士等多位横跨产学两界的专家学者将带来重磅报告。

尖峰对话环节, 产学研用各界同台发声, 将打造一场以院士为主要阵容的多元碰撞的观点盛宴, 围绕智能产业发展的热点命题迸发思路, 输出见解。

同时, 为期两天的活动还将举办十五个专题论坛和两场同期活动, 涉及情感智能、数基生命、脑科学、模式识别、知识工程、智能媒体、人工智能服务平台、智慧法律、智能交通、以及人工智能伦理、青年科学家成长等议题, 各领域资深专家将进行交流探讨。

智“荟”浦江之滨

厦门这座东南沿海重要的中心城市, 是旅游者眼中的风景胜地, 是创新者心中的软件名城。立足于信息技术优势, 厦门正在疾步迈向数字时代, 加快建设“人工智能+”新经济产业体系, 全力打造集创新政策、应用示范、制度供给、人才培训于一体的全球顶尖人工智能标杆城市。

随着 CIIS 2021 的到来, 智能产业的标杆人物、标杆创新、标杆应用、标杆案例等一系列优质资源将“智荟”厦门, 为地方带来实实在在的战略价值。

智“绘”数字未来

回看过去十年, 人工智能经历了从概念到落地、从实验室走向生产线的实践过程, 与实体经济的融合日益深化, 推动数字产业蓬勃发展。

我国“十四五”规划和 2035 年远景目标纲要提出“打造数字经济新优势”, 提出到 2025 年, 数字经济核心产业增加值占 GDP 比重将达到 10%。

智能产业是数字经济的重要部分, 未来十年, 哪些新技术、新模式、新场景将成为传统行业数字化、智能化升级的关键变量, CIIS 2022 将以技术引领、应用驱动的角度“智绘”数字未来, 为各行各业给出具有实践性的答案。