

doi: 10.12052/gdutxb.190048

联合低秩表示与图嵌入的无监督特征选择

滕少华¹, 冯镇业¹, 滕璐瑶², 房小兆¹

(1. 广东工业大学 计算机学院, 广东 广州 510006; 2. 维多利亚大学 应用信息中心, 维多利亚州 墨尔本 VIC 3011)

摘要: 大数据应用带来高维数据急剧增加,数据降维已成为重要问题. 特征选择降维方法已广泛应用于模式识别领域,近年来提出了许多基于流形学习的特征选择方法,然而这类方法往往容易受到各种噪声影响. 对此,本文提出一种联合低秩表示和图嵌入的高效无监督特征选择方法(JLRRGE). 通过低秩表示寻找数据在低秩子空间下的表示,降低噪声的影响从而提高算法的鲁棒性,并通过自适应图嵌入方法,使选择特征保持原有的局部关系. 实验结果表明,本文提出算法的分类准确率优于其他对比算法.

关键词: 无监督学习; 低秩表示; 图嵌入; 特征选择

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1007-7162(2019)05-0007-07

Joint Low-Rank Representation and Graph Embedding for Unsupervised Feature Selection

Teng Shao-hua¹, Feng Zhen-ye¹, Teng Lu-yao², Fang Xiao-zhao¹

(1. School of Computers, Guangdong University of Technology, Guangzhou 510006, China; 2. Centre for Applied Informatics, Victoria University, Melbourne VIC 3011, Australia)

Abstract: Dimensionality reduction becomes a significant problem due to the proliferation of high dimensional data. For dimensionality reduction, feature selection is more analytical than feature extraction. Therefore feature selection plays an important role in pattern recognition. In recent years, many feature selection methods based on manifold learning have been proposed. However, such methods are susceptible to noise data. Therefore, an efficient unsupervised feature selection method is proposed—joint low-rank representation and graph embedding for unsupervised feature selection (JLRRGE). This method not only finds the low rank structure of the data after selecting feature, which makes the algorithm more robust, but also preserves the local manifold structure of the data through the adaptive graph embedding. The experimental results show that the proposed algorithm is superior to other compared methods.

Key words: unsupervised learning; low rank representation; graph embedding; feature selection

在许多现实应用场景中,数据样本的维度总是很高. 比如人脸识别^[1]、手写字符识别^[2]、生物信息学分析^[3]、视觉跟踪^[4]等^[5-7]. 然而高维数据数据量大,且通常含有大量冗余、噪声,因此会降低学习算法的性能^[8]. 为了解决这些问题,人们通常采用降维方法对数据进行预处理.

近些年提出了许多降维方法,可分为特征选择^[9]和特征提取^[10]两大类. 特征提取通过将数据投影到低维子空间来减少数据维度. 由于低维子空间与原

数据样本空间没有直接联系,提取特征难以对应数据原有特征. 特征选择通过直接选择原数据特征的子集作为降维后的特征. 这能保持数据原有特征,降维后的数据具有很好的解析性. 这在许多应用中是很重要的,例如基因分类^[11]、文本分类^[12]等.

从利用标签信息的角度来看,特征选择算法可以分为监督学习^[13]、半监督学习^[14]和无监督学习^[15]. 监督学习和半监督学习在某种程度上依赖于标签信息,通过标签信息评估结果来选择特征. 然而,许多

实际应用是在没有标签的情况下收集大规模数据. 标记这些无标签数据非常昂贵且耗时^[16]. 因此, 无监督特征选择对于许多实际应用更具普遍性和挑战性.

Cai等^[17]提出了多簇特征选择(MCFS), 通过谱分析捕获局部流形结构, 然后选择能够最好地保留聚类结构的特征. Nie等^[18]提出了灵活流形嵌入(FME), FME作为降维的一般框架, 被许多特征选择方法所采用. Hou等^[19]基于FME和 $l_{2,1}$ 范数, 提出了联合嵌入学习和稀疏回归(JELSR), 给出了特征选择的一般框架. Li等^[20]联合使用FME、非负约束和 $l_{2,1}$ 范数来实现非负判别特征选择(NDFS). Qian等^[21]提出了鲁棒的无监督特征选择(RUFS), RUFS使用FME、非负矩阵分解(NMF)和 $l_{2,1}$ 范数同时实现鲁棒的聚类和特征选择. Shi等^[22]提出联合FME和 l_1 范数实现的鲁棒谱特征选择(RSFS). Nie等^[23]提出了联合结构最优图和局部结构的结构最优图特征选择(SOGFS). 上述方法取得了显著的效果, 但在选择特征中并没有考虑到选择特征后数据的低秩子空间结构, 可能会使算法过分依赖局部结构, 导致对噪声数据敏感. 本文通过引入低秩表示不仅保留了转换空间后样本的全局结构, 而且学习其低秩子空间, 极大地降低了噪声的影响. 并且自适应学习样本间的局部结构, 将全局结构学习、局部结构学习和特征选择融合到一个框架中同时进行.

因此本文通过结合低秩子空间学习和自适应图嵌入方法, 提出了一个联合低秩表示与图嵌入的无监督特征选择方法. 本文主要贡献在于: (1)提出了一种保留选择特征局部结构方法, 同时考虑到选择特征的低秩子空间, 使其选择特征后数据具有低秩性质, 从而提高方法的鲁棒性; (2)选择特征的局部结构自适应学习, 无需预先计算, 从而选择最合适模型的局部结构.

1 相关工作

1.1 低秩子空间学习

低秩表示假设原始数据可分解成多个独立子空间来近似表示, 从而发现数据本质结构中, 其子空间最低秩的表示^[24]. 给定数据集 $X = [x_1, x_2, \dots, x_n] \in R^{d \times n}$, 其中 $x_i \in R^d$ 为数据集的某个样本. 本文构建一个字典 $D = [d_1, d_2, \dots, d_m] \in R^{d \times m}$ 来重构数据集 X , 使得 $X = DZ$, 其中 Z 为重构系数矩阵, $Z = [z_1, z_2, \dots, z_n] \in R^{m \times n}$. 通过对 Z 添加低秩约束, 其目的

是为了寻找其最低秩解, 因此上述模型如下:

$$\min_Z \text{rank}(Z) \quad \text{s.t.} \quad X = DZ. \quad (1)$$

上述问题是NP难问题, 难以直接求解. Candès等^[25]通过使用核范数来代替低秩约束, 可转换为如下凸优化问题:

$$\min_Z \|Z\|_* \quad \text{s.t.} \quad X = DZ. \quad (2)$$

其中 $\| \cdot \|_*$ 表示矩阵的核范数, 即矩阵奇异值之和. 使用原数据 X 作为字典, 式(2)可转化为

$$\min_Z \sum_i^n \|x_i - Xz_i\|^2 + \alpha \|Z\|_*, \quad (3)$$

其中 α 为正则项参数.

1.2 图嵌入方法

随着谱分析和流形学习的发展, 许多无监督特征选择方法通过保留样本间的局部流形结构均取得较好的结果. 这表明局部结构在特征选择过程中具有重要作用. 本文利用自适应过程来确定保持局部关系的相似矩阵. 给定数据集 $X = [x_1, x_2, \dots, x_n] \in R^{d \times n}$. 其中 $x_i \in R^d$ 可被其他样本 x_j 通过概率 p_{ij} 表示, 其中 p_{ij} 为相似矩阵 $P \in R^{n \times n}$ 的元素. 任意两个样本作为邻居的概率可以看作它们之间的相似性. 因此较近的样本应具有较大的概率, 即 p_{ij} 与 x_i 和 x_j 之间的距离成反比. 因此相似矩阵 P 可由式(4)求解:

$$\min_{P^T I_n = 1, 0 \leq p_{ij} \leq 1} \sum_{ij} (\|x_i - x_j\|^2 p_{ij} + \mu p_{ij}^2), \quad (4)$$

其中 μ 为正则化参数, 正则项 μp_{ij}^2 可避免退化解. I_n 为长度为 n , 元素全为1的向量.

2 联合低秩表示与图嵌入的无监督特征选择

此处, 给出本文使用的符号. 矩阵 $M \in R^{m \times n}$, m_{ij} 为矩阵 M 的第 i 行 j 列的元素. m_i 为矩阵 M 的第 i 行向量. $\text{Tr}(M)$ 表示矩阵 M 的迹, $\langle M, M \rangle = \text{Tr}(M^T M)$. 矩阵 M 的F范数表示为 $\|M\|_F$. 矩阵 M 的 $l_{2,1}$ 范数表示

$$\text{为} \|M\|_{2,1}, \text{其中} \|M\|_{2,1} = \sum_{i=1}^m \sqrt{\sum_{j=1}^n m_{ij}^2}.$$

2.1 JLRRGE模型

根据上述子空间学习方法和自适应局部结构学习方法, 由式(3)和式(4)可得:

$$\begin{cases} \min_{\mathbf{Z}, \mathbf{P}} \sum_i^n \|\mathbf{x}_i - \mathbf{X}\mathbf{z}_i\|^2 + \alpha \|\mathbf{Z}\|_* + \\ \beta \sum_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2 P_{ij} + \mu P_{ij}^2. \\ \text{s.t.} \quad \mathbf{P}\mathbf{I}_n = \mathbf{I}_n, \mathbf{P} > 0. \end{cases} \quad (5)$$

式(5)中,第一部分为低秩子空间学习, $\mathbf{X}\mathbf{z}_i$ 为样本 $\mathbf{x}_i$ 在低秩子空间下重构样本. 对于整个重构后的数据 $\mathbf{X}\mathbf{Z}$, 其低秩结构能更好地还原数据的本质结构, 降低噪声对数据影响. 第二部分为自适应局部结构学习, 自适应学习数据的局部相似性矩阵 $\mathbf{P}$, 使其保留原有的局部流形结构.

根据式(5), 本文提出一种联合低秩表示与图嵌入的无监督特征选择学习, 如式(6)所示:

$$\begin{cases} \min_{\mathbf{W}, \mathbf{Z}, \mathbf{P}} \sum_i^n \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{X}\mathbf{z}_i\|^2 + \alpha \|\mathbf{Z}\|_* + \\ \beta \sum_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2 P_{ij} + \mu P_{ij}^2 + \gamma \|\mathbf{W}\|_{2,1}. \\ \text{s.t.} \quad \mathbf{P}\mathbf{I}_n = \mathbf{I}_n, \mathbf{P} > 0, \mathbf{W}^T \mathbf{X}\mathbf{X}^T \mathbf{W} = \mathbf{I}. \end{cases} \quad (6)$$

其中 $\mathbf{W} \in R^{d \times c}$ 为特征选择矩阵, c 为选择特征的维持, β, γ 为正则化参数. 上述模型的第一和第二项通过低秩表示学习样本在转换空间后的低秩子空间, 并保留了样本的全局结构. 其中在重构系数矩阵 $\mathbf{Z}$ 的低秩约束下, 重构后的样本能极大降低噪声样本的影响. 并且 $\mathbf{W}$ 具有行稀疏, 能极大降低噪声或冗余特征的影响. 通过引入低秩表示重构数据, 同时降低了噪声样本、噪声特征以及冗余特征对特征选择的影响. 第三和第四项自适应学习样本间的局部流形结构, 保留了样本间的局部结构.

2.2 模型优化

式(6)中 $\mathbf{W}, \mathbf{Z}, \mathbf{P}$ 均未知, 很难直接求出目标函数的最优解. 因此本文使用增广拉格朗日乘子法 (ALM) 来尝试求解目标函数的最小值, 如式(7)所示:

$$\begin{cases} \Gamma(\mathbf{W}, \mathbf{S}, \mathbf{P}, \mathbf{Q}, \mathbf{T}, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \mu_1) = \|\mathbf{W}^T \mathbf{X} - \mathbf{W}^T \mathbf{X}\mathbf{Z}\|_F^2 + \\ \alpha \|\mathbf{T}\|_* + \beta \sum_{ij} \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2 P_{ij} + \mu P_{ij}^2 + \\ \gamma \|\mathbf{W}\|_{2,1} + \langle \mathbf{Y}_1, \mathbf{P}\mathbf{I}_n - \mathbf{I}_n \rangle + \langle \mathbf{Y}_2, \mathbf{P} - \mathbf{Q} \rangle + \\ \langle \mathbf{Y}_3, \mathbf{Z} - \mathbf{T} \rangle + \frac{\mu_1}{2} (\|\mathbf{P}\mathbf{I}_n - \mathbf{I}_n\|_F^2 + \|\mathbf{P} - \mathbf{Q}\|_F^2 + \|\mathbf{Z} - \mathbf{T}\|_F^2). \\ \text{s.t.} \quad \mathbf{W}^T \mathbf{X}\mathbf{X}^T \mathbf{W} = \mathbf{I}, \mathbf{Q} > 0. \end{cases} \quad (7)$$

其中 $\mathbf{Y}_1, \mathbf{Y}_2$ 和 $\mathbf{Y}_3$ 为拉格朗日乘子, μ_1 为惩罚参数. 式(7)可通过不精确拉格朗日乘子法 (IALM) 求解.

IALM 在求解最小值过程中, 在每一步中更新一个变量, 同时固定其他变量不变. 并依次迭代更新所

有变量直到收敛. 主要的求解步骤如下:

1) 更新 $\mathbf{W}$. 在式(6)中对 $\mathbf{W}$ 求偏导数可得

$$\frac{(\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{G}^T + \mathbf{G})}{2} \mathbf{W} = \mathbf{A}\mathbf{W}, \quad (8)$$

其中 $\mathbf{G} = \mathbf{X}\mathbf{L}\mathbf{X}^T + \gamma \mathbf{D}_W$, $\mathbf{D}_W$ 为对角矩阵且第 i 个对角元素为

$$D_{W_{ii}} = \frac{1}{2\|\mathbf{W}_i\|_2}, \quad (9)$$

$\mathbf{L} = \mathbf{I} + \mathbf{Z}\mathbf{Z}^T - \mathbf{Z}^T - \mathbf{Z} + \beta(\mathbf{D}_P + \mathbf{D}_{P^T} - 2\mathbf{P})$, 其中 $\mathbf{D}_P$ 和 $\mathbf{D}_{P^T}$ 为对角矩阵, 其中对角元素分别为 $D_{P_{ii}} = \sum_j P_{ij}$ 和 $D_{P_{ii}^T} = \sum_i P_{ij}$. 显而易见, $\mathbf{W}$ 的最优解对应于特征值分解问题的 c 个最小特征值的特征向量.

2) 更新 $\mathbf{Z}$. 对 $\mathbf{Z}$ 求偏导数可得

$$\mathbf{Z}^* = (\mathbf{F} + \mu_1 \mathbf{I})^{-1} \mathbf{F} - \mathbf{T}\mu_1 - \mathbf{Y}_3, \quad (10)$$

其中 $\mathbf{F} = 2\mathbf{X}^T \mathbf{W} \mathbf{W}^T \mathbf{X}$.

3) 更新 $\mathbf{P}$. 对 $\mathbf{P}$ 求偏导数可得

$$\mathbf{P}^* = \frac{\mathbf{E}}{2\mu + \mu_1} \left(\mathbf{I} + \frac{\mu_1}{2\mu + \mu_1} \mathbf{I}_n \mathbf{I}_n^T \right)^{-1}, \quad (11)$$

其中

$$\begin{aligned} \mathbf{E} &= (\mu_1 \mathbf{I}_n \mathbf{I}_n^T + \mu_1 \mathbf{Q} - \beta \mathbf{B} - \mathbf{Y}_1 \mathbf{I}_n^T - \mathbf{Y}_2), \\ B_{ij} &= \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|^2. \end{aligned}$$

4) 更新 $\mathbf{T}$. 对 $\mathbf{T}$ 求偏导数可得

$$\mathbf{T}^* = \nu_{1/\mu_1} \left(\mathbf{Z} + \frac{\mathbf{Y}_3}{\mu_1} \right), \quad (12)$$

其中 $\nu_{1/\mu_1}(\mathbf{X}) = \mathbf{U}\mathbf{S}_\tau(\mathbf{\Sigma})\mathbf{V}$ 是关于奇异值 τ 的阈值算子, $\mathbf{S}_\tau(\mathbf{\Sigma}_{ij}) = \text{sign}(\mathbf{\Sigma}_{ij}) \max(0, |\mathbf{\Sigma}_{ij} - \tau|)$ 为软阈值算子, $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$ 为 $\mathbf{X}$ 的奇异值分解.

5) 更新 $\mathbf{Q}$. 对 $\mathbf{Q}$ 求偏导数可得

$$\mathbf{Q}^* = \max \left(\mathbf{P} + \frac{\mathbf{Y}_2}{\mu_1}, 0 \right). \quad (13)$$

6) 最后更新拉格朗日乘子和惩罚参数:

$$\mathbf{Y}_1 = \mathbf{Y}_1 + \mu_1(\mathbf{P}\mathbf{I}_n - \mathbf{I}_n). \quad (14)$$

$$\mathbf{Y}_2 = \mathbf{Y}_2 + \mu_1(\mathbf{P} - \mathbf{Q}). \quad (15)$$

$$\mathbf{Y}_3 = \mathbf{Y}_3 + \mu_1(\mathbf{Z} - \mathbf{T}). \quad (16)$$

$$\mu_1 = \min(\mu_{\max}, \rho \mu_1). \quad (17)$$

其中 $\rho > 0$ 为迭代步长, $\mu_{\max}$ 为常数.

2.3 时间复杂度

上述优化过程中时间复杂度分析如下: 第1步为特征值分解问题, 其运算矩阵的大小 $d \times d$, 所以时间复杂度为 $o(d^3)$; 第2步中矩阵求逆需要 $o(n^3)$; 第4步中的时间复杂度为 $o(n^3)$. 因此整个优化算法的时间复

杂度为 $o(\tau(d^3 + 2n^3))$,其中 τ 为算法优化过程中的迭代次数.

2.4 收敛性分析

在本节中, 将JLRRGE运行1-最近邻分类器(NN). 在2个数据集上运行本文所提方法, 以显示其收敛性质, 包括JAFPE 和MFEA数据集. 当所有变量和分类准确率都稳定, 则该算法趋于收敛. 在上述优化过程中, 使用变量 $\boldsymbol{Q}$ 和 $\boldsymbol{T}$ 来分别逼近变量 $\boldsymbol{P}$ 和 $\boldsymbol{Z}$. 本文认为当 $\boldsymbol{Q}$ 接近于 $\boldsymbol{P}$ 且 $\boldsymbol{T}$ 接近于 $\boldsymbol{Z}$ 时, 所有变量趋于稳定. 因此定义了如下目标函数:

$$\text{obj} = \|\boldsymbol{P} - \boldsymbol{Q}\|_F + \|\boldsymbol{Z} - \boldsymbol{T}\|_F.$$

(18)

在上述两个数据集上运行本文所提方法并进行100次迭代, 并在图1中显示了目标函数值、分类精度和迭代之间的关系.

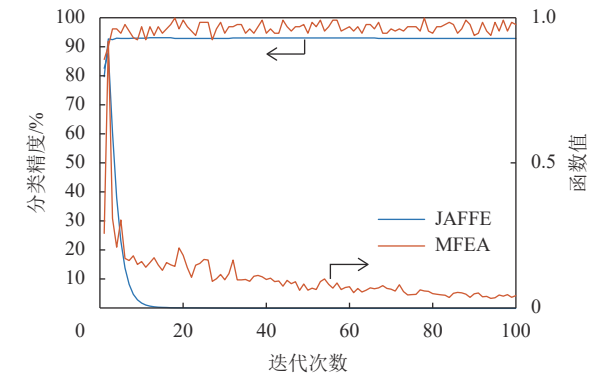


图1 数据集JAFPE和MFEA收敛性实验
Fig.1 Convergence experiments on JAFPE and MFEA datasets

由图1可知, 随着迭代次数的增加, 目标函数整体上不断减少且分类准确率整体上不断升高, 并在一定迭代次数后趋于稳定. 在JAFPE数据集中目标函数值先上升后下降, 可能是由于在优化过程中, 随机初值所造成的影响. 总体而言, 由目标函数值和分类准确率曲线可知本文提出的方法具有较强的鲁棒性, 能够快速逼近局部最优解, 并具有良好的收敛性.

3 实验

本文所提出的JLRRGE将与4种具有代表性的特征选择算法如拉普拉斯评分的特征选择(Laplacian Score, LS)、多类簇特征选择(MCFS)、无监督判别特征选择(UDFS)和最优结构图的特征选择(SOGFS)进行详细对比.

3.1 实验数据集

本文使用以下5个常用数据集, 如表1所示.
COIL数据集包含20个对象, 每个对象有72个样本, 共1 440个样本. 样本以5度为间隔从同方向进行

表 1 数据集信息

Tab.1 Summary of the benchmark data sets

No.	Dataset	Sample	Feature	Label
1	COIL20	1 440	1 024	20
2	ORL32	400	1 024	40
3	JAFPE	213	676	10
4	UMIST	575	644	20
5	MFEA	2 000	240	10

捕获, 所有图像被调整为1 024个像素.
ORL数据集包含40个对象, 每个对象拍摄10幅图片, 共400张图. 每个对象在不同的时间、光照、面部表情(睁/闭眼、微笑/不微笑)和面部细节(是否戴眼镜)下拍摄, 所有图像被调整为 32×32 个像素.
JAFPE数据集包含213张7种面部表情的图像, 包括6种基本面部表情和1种中性表情.
UMIST数据集包含575张20人的图像. 每个人从侧面到正面都摆出一系列姿势, 而且他们来自不同的种族. 这些文件都是PGM格式, 灰度为256个阴影.
MFEA数据集包含一系列的手写数字(0~9), 这些数字是从一组荷兰实用地图中提取的. 一幅图像的大小为 15×16 像素.

3.2 参数设置与评价指标

本文实验对每个数据集做4组对比实验. 每组实验在每一类中分别随机抽取10%、20%、30%和40%作为训练样本, 其余作为测试样本. 在训练过程中, 每组实验选取特征数目分别为 $c = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$. 最终结果运行在1-最近邻分类器(NN)中验证本文算法的性能.

对于LS和MCFS算法生成的近邻矩阵时, 参数 k 统一设置为5. LS热核计算权重矩阵 $\boldsymbol{W}$ 时, σ 取值1. UDFS和SOGFS中参数 γ 设置为 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$. 本文提出的算法JLRRGE中参数 α, β, γ 均设置为 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$.

此外, 算法的鲁棒性也是需要考虑的范畴. 由算法目标函数可知, 参数 α 和 β 对优化过程起重要作用. 本文在各数据集抽取10%作为训练样本的情景下, 随机固定参数 λ 和选择特征个数, 在不同的参数 α 和 β 下, 进行实验并记录其对应的准确率, 结果如图2所示.

3.3 实验结果与分析

3.3.1 COIL20数据集

由表2可知, JLRRGE在随机抽取20%的训练样本的实验中, 比MCFS的分类效果略差. 其余情况均取得较好的分类结果, 且整体分类效果比其他对比算法好.

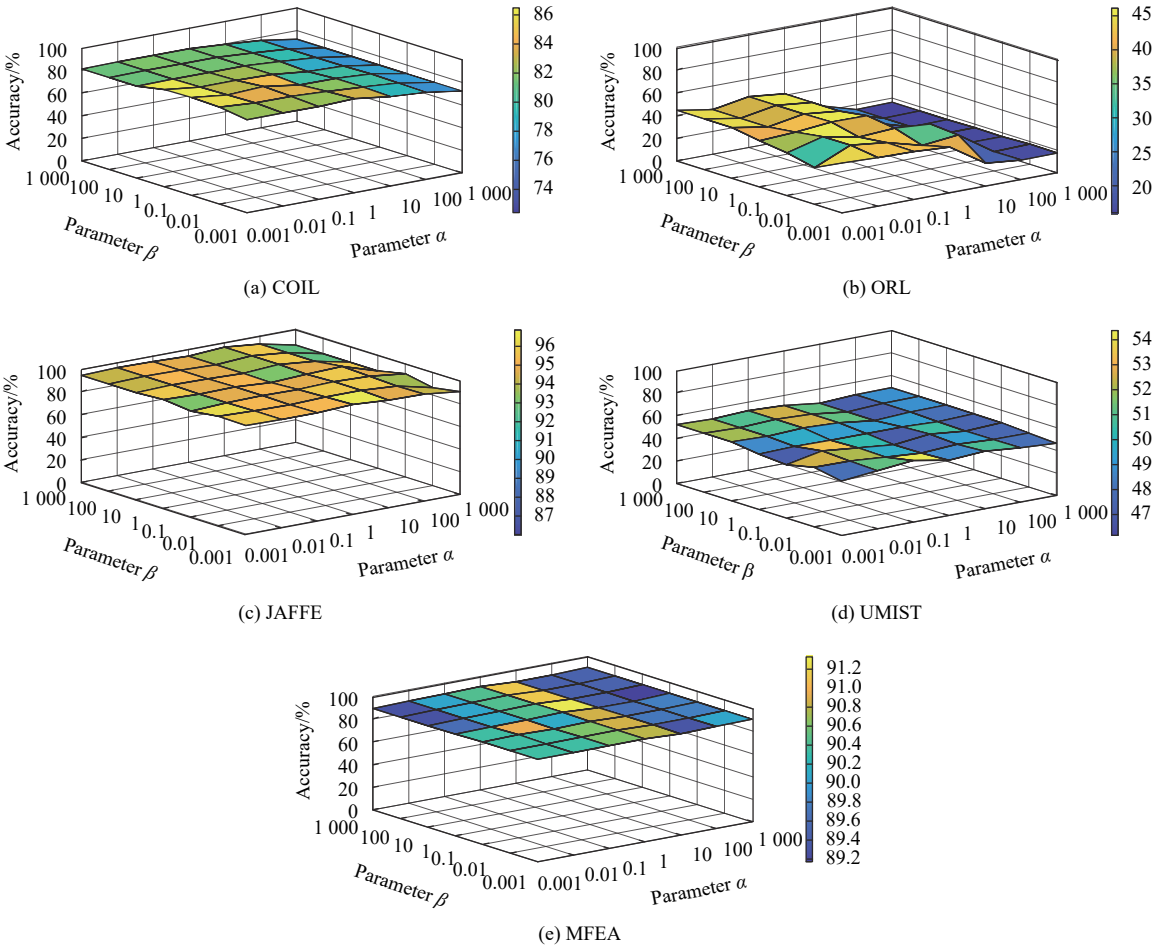


图2 各数据集的参数敏感实验

Fig.2 Parameter sensitivity on different datasets

表2 不同算法在COIL数据集上的分类准确率

Tab.2 Aggregated classification results measured by accuracy of the compared methods on COIL

抽取数比例/%	LS	MCFS	UDFS	SOGFS	JLRRGE
10	78.46	83.20	85.38	86	87.69
20	86.38	93.97	93.10	93.36	93.36
30	86.57	94.90	96.37	95.20	96.08
40	89.88	97.16	97.27	94.55	98.75
Average	85.32	92.31	93.03	92.28	93.97

由图2(a)可知, 当 $\alpha \in [0.001, 10]$ 时, JLRRGE取得较好的结果. 当 $\alpha \in [10, 1000]$ 时, JLRRGE分类性能略微下降. 整体上, JLRRGE在数据集COIL中对参数 α 和 β 的变化并不敏感, 即JLRRGE对于数据集COIL具有较好的鲁棒性.

3.3.2 ORL32数据集

由表3可知, 除了在抽取20%作为训练样本的情况, JLRRGE均取得最好的分类结果.

由图2(b)可知, 当 $\alpha \in [0.001, 1]$ 时, 取得较好的结果. 当 $\alpha \in [10, 1000]$ 时, 分类准确率骤降. 即过于注重

ORL数据的全局结构, 而导致其分类准确率下降. 由此表明参数 α 的变化对JLRRGE在数据集ORL上的分类效果具有较大的影响.

表3 不同算法在ORL数据集上的分类准确率

Tab.3 Aggregated classification results measured by accuracy of the compared methods on ORL

抽取数比例/%	LS	MCFS	UDFS	SOGFS	JLRRGE
10	31.39	46.39	49.72	51.39	52.22
20	40.63	63.13	65	70.63	68.75
30	61.79	76.79	71.07	74.29	78.21
40	60.83	80.42	77.08	80	82.08
Average	48.66	66.68	65.72	69.08	70.32

3.3.3 JAFFE数据集

由表4可得, JLRRGE在JAFFE数据集上, 对于不同百分比的训练样本, 均取得较好的效果.

由图2(c)可得, 当 $\alpha \in [0.001, 100]$ 时, JLRRGE取得较好的结果. 当 $\alpha \in [100, 1000]$ 时, JLRRGE分类性能略微下降. 整体上, JLRRGE在数据集JAFFE上具有较强的鲁棒性, 对参数 α 和 β 并不敏感.

表 4 不同算法在JAFFE数据集上的分类准确率

Tab.4 Aggregated classification results measured by accuracy of the compared methods on JAFFE %

抽取数量/%	LS	MCFS	UDFS	SOGFS	JLRRGE
10	78.23	94.82	95.34	96.89	98.45
20	89.60	99.42	98.27	100	100
30	96.73	100	99.35	98.69	100
40	97.74	99.25	99.25	99.25	100
Average	90.56	98.37	98.05	98.71	99.61

3.3.4 UMIST数据集

由表5可知, JLRRGE的整体分类性能略差于MCFS, 仅在随机抽取10%训练样本的情况下取得较好的分类结果.

表 5 不同算法在UMIST数据集上的分类准确率

Tab.5 Aggregated classification results measured by accuracy of the compared methods on UMIST %

抽取数比例/%	LS	MCFS	UDFS	SOGFS	JLRRGE
10	45.61	57.20	57.57	55.70	58.32
20	70.53	87.58	85.68	84.42	87.16
30	76.87	90.84	87.95	86.02	89.40
40	88.45	97.18	95.21	96.34	96.34
Average	70.37	83.20	81.60	80.62	82.81

由图2(d)可得, 当 $\alpha \in [0.001, 1]$ 时, JLRRGE取得较好的结果. 当 $\alpha \in [1, 1\ 000]$ 时, JLRRGE分类性能略微下降. 整体上, JLRRGE在数据集UMIST对参数 α 和 β 并不敏感.

3.3.5 MFEA数据集

由表6可得, JLRRGE在各情况下均取得较好的分类结果.

表 6 不同算法在MFEA数据集上的分类准确率

Tab.6 Aggregated classification results measured by Accuracy of the compared methods on MFEA %

抽取数比例/%	LS	MCFS	UDFS	SOGFS	JLRRGE
10	90.44	93.22	91.89	92.89	93.28
20	93.75	94.56	92.63	94.63	95.37
30	94.86	96.14	92.64	96.29	96.36
40	94.43	95.50	92.42	95.42	96.50
Average	93.37	94.86	92.40	94.81	95.38

由图2(e)可得, 当 $\alpha \in [10, 1\ 000]$ 时, JLRRGE分类性能略微下降. 整体上, JLRRGE在数据集MFEA对参数 α 和 β 并不敏感, 具有较强的鲁棒性.

4 结束语

本文提出了一种新颖的无监督特征选择方法. 通过低秩表示选择具有低秩结构的特征, 使选择的

特征更具鲁棒性. 同时结合图嵌入方法, 自适应学习数据的局部相似性矩阵, 使数据降维后仍保留其局部流形结构. 实验表明, 本文所提的方法比其他对比方法在绝大部分情况下均取得更好的分类结果, 并且该方法具有良好的鲁棒性. 在未来将尝试优化模型的复杂度, 减少模型的训练时间.

参考文献:

[1] XU Y, FANG X Z, LI X L, *et al.* Data uncertainty in face recognition [J]. *IEEE Transactions on Cybernetics*, 2014, 44(10): 1950-1961.

[2] YAO C, CHENG G. Approximative bayes optimality linear discriminant analysis for Chinese handwriting character recognition [J]. *Neuro-computing*, 2016, 207: 346-353.

[3] FEI L K, ZHANG B, ZHANG W, *et al.* Local apparent and latent direction extraction for palm-print recognition [J]. *Information Sciences*, 2019, 473: 59-72.

[4] LAN X, MA A J, YUEN P C, *et al.* Joint sparse representation and robust feature-level fusion for multicue visual tracking [J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5826-5841.

[5] FANG X Z, HAN N, WU J G, *et al.* Approximate low-rank projection learning for feature extraction [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(11): 5228-5241.

[6] 滕少华, 郑明, 刘冬宁. 面向音乐推荐的全变差图非负矩阵分解方法[J]. *计算机应用研究*, 2018, 35(4): 1010-1013.

TENG S H, ZHENG M, LIU D L. Facing music recommended total variation non-negative matrix decomposition method [J]. *Application Research of Computers*, 2018, 35(4): 1010-1013.

[7] 滕少华, 宋欢, 霍颖翔, 等. 一种增量式学习的语音字典构造方法[J]. *广东工业大学学报*, 2018, 35(3): 29-36.

TENG S H, SONG H, HUO Y X, *et al.* An incremental learning approach in voice compression via sparse dictionary learning [J]. *Journal of Guangdong University of Technology*, 2018, 35(3): 29-36.

[8] JA IN, DU IN, MAO J. Statistical pattern recognition: a review [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, 27(11): 1502-1502.

[9] FANG X Z, XU Y, LI X L, *et al.* Locality and similarity preserving embedding for feature selection [J]. *Neurocomputing*, 2014, 128: 304-315.

[10] 滕少华, 卢东略, 霍颖翔, 等. 基于正交投影的降维分类方法研究[J]. *广东工业大学学报*, 2017, 34(3): 1-7.

TENG S H, LU D L, HUO Y X, *et al.* Classification method based on dimension reduction [J]. *Journal of Guangdong University of Technology*, 2017, 34(3): 1-7.

[11] IMOTO S, MIVANO S. A top-r feature selection algorithm for microarray gene expression data [J]. *IEEE/ACM Transactions on Computational Bi-ology and Bioinformatics*, 2012, 9(3): 754-764.

[12] SHANG C, LI M, FENG S, *et al.* Feature selection via maximizing global information gain for text classification [J]. *Knowledge-Based Systems*, 2013, 54: 298-309.

[13] NIE F P, HUANG H, CAI X, *et al.* Efficient and robust feature selection via joint ℓ_2 , 1-norms minimization[C]//Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada. Curran Associates Inc. 2010.

[14] FANG X Z, XU Y, LI X, *et al.* Robust semi-supervised subspace clustering via non-negative low-rank representation [J]. *IEEE Transactions on Cybernetics*, 2016, 46(8): 1828-1838.

[15] ZHU X F, LI X L, *et al.* Robust joint graph sparse coding for unsupervised spectral feature selection [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 6(28): 1263-1275.

[16] CHENG Q, ZHOU H, CHENG J. The fisher-markov selector: fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data [J]. *IEEE Transactions on Software Engineering*, 2011, 33(6): 1217-1233.

[17] CAI D, ZHANG C, HE X. Unsupervised feature selection for multi-cluster data[C]//ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. USA, Washington, DC: ACM, 2010.

[18] NIE F P, XU D, TSANG W H, *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.

[19] 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 Cybernetics*, 2013, 44(6): 793-804.

[20] LI Z, YANG Y, LIU J, *et al.* Unsupervised feature selection using nonnegative spectral analysis [C]// Proceedings of 26th AAAI Conference on Artificial Intelligence. Toronto: AAAI Press, 2012: 1026-1032.

[21] QIAN M, ZHAI C. Robust unsupervised feature selection[C]//International Joint Conference on Artificial Intelligence. China, Beijing: ACM, 2013.

[22] SHI L, DU L, SHEN Y D. Robust spectral learning for unsupervised feature selection[C]//IEEE International Conference on Data Mining. China, Shenzhen: IEEE, 2015.

[23] NIE F P, ZHU W, LI X L. Unsupervised feature selection with structured graph optimization[C]//Thirtieth Aaai Conference on Artificial Intelligence. USA, Phoenix: AAAI Press, 2016.

[24] WEN J, HAN N, FANG X Z, *et al.* Low-rank preserving projection Via graph regularized reconstruction [J]. *IEEE Transactions on Cybernetics*, 2018: 1-13.

[25] CANDES E J, RECHT B. Exact matrix completion via convex optimization [J]. *Foundations of Computational Mathematics*, 2009, 9(6): 717.

(上接第6页)

[15] RAPPUOLI R, ADEREM A. A 2020 vision for vaccines against HIV, tuberculosis and malaria [J]. *Nature*, 2011, 473(7348): 463-471.

[16] RINGEL D M, SKIERA B. Visualizing asymmetric competition among more than 1000 products using big search data [J]. *Marketing Science*, 2016, 35(3): 511-534.

[17] 杨建梅, 周恋, 周连强. 中国汽车产业竞争关系与轿车社团企业对抗行动研究[J]. *管理学报*, 2013, 10(1): 49-55.

YANG J M, ZHOU L, ZHOU L Q. Competitive relationships of auto industry and rivalry actions of car community enterprises in China [J]. *Chinese Journal of Management*, 2013, 10(1): 49-55.

[18] 谢逢洁, 崔文田, 武小平. 快递产业竞争关系网络模型构建及结构特性分析[J]. *系统工程*, 2017, 35(7): 101-106.

XIE F J, CUI W T, WU X P. The model of competitive relationship network of express enterprises and its structural properties [J]. *Systems Engineering*, 2017, 35(7): 101-106.

[19] MOE W W. An empirical two-stage choice model with varying decision rules applied to internet clickstream data [J]. *Journal of Marketing Research*, 2006, 43(4): 680-692.

[20] IWANAGA J, NISHIMURA N, SUKEGAWA N, *et al.* Estimating product-choice probabilities from recency and frequency of page views [J]. *Knowledge-Based Systems*, 2016, 99(C): 157-167.

[21] URBAN G L, HULLAND J S, WEINBERG B D. Pre-market forecasting for new consumer durable goods: model [J]. *Journal of Marketing*, 1993, 57(2): 47-63.

[22] DESARBO W S, MANRAI A K, MANRAI L A. Non-spatial tree models for the assessment of competitive market structure: an integrated review of marketing and psychometric literature [J]. *Handbooks in Operations Research & Management Science*, 1993, 5(5): 193-257.

[23] CHO S, CHONG A. Co-query volume as a proxy for brand relatedness [J]. *Industrial Management & Data Systems*, 2018, 118(4): 930-944.

[24] NEWMAN M E J. Fast algorithm for detecting community structure in networks [J]. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2003, 69(6): 066133.

[25] FORTUNATO S, BARTHELEMY M. Resolution limit in community detection [J]. *Proceedings of the National Academy of Sciences*, 2007, 104(1): 36-41.