



智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

自步稀疏最优均值主成分分析

许子微, 陈秀宏

引用本文:

许子微, 陈秀宏. 自步稀疏最优均值主成分分析[J]. 智能系统学报, 2021, 16(3): 416–424.

XU Ziwei, CHEN Xiuhong. Sparse optimal mean principal component analysis based on self-paced learning[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(3): 416–424.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201911028>

您可能感兴趣的其他文章

结构化加权稀疏低秩恢复算法在人脸识别中的应用

A low rank recovery algorithm for face recognition with structured and weighted sparse constraint
智能系统学报. 2019, 14(3): 455–463 <https://dx.doi.org/10.11992/tis.201711026>

SUCE:基于聚类集成的半监督二分类方法

SUCE: semi-supervised binary classification based on clustering ensemble
智能系统学报. 2018, 13(6): 974–980 <https://dx.doi.org/10.11992/tis.201711027>

基于卷积特征和贝叶斯分类器的人脸识别

Face recognition based on convolution feature and Bayes classifier
智能系统学报. 2018, 13(5): 769–775 <https://dx.doi.org/10.11992/tis.201706052>

多标记学习自编码网络无监督维数约简

Unsupervised dimensionality reduction of multi-label learning via autoencoder networks
智能系统学报. 2018, 13(5): 808–817 <https://dx.doi.org/10.11992/tis.201804051>

多层递阶融合模糊特征映射的模糊C均值聚类算法

Fuzzy C-means clustering algorithm for multilayered hierarchical fusion fuzzy feature mapping
智能系统学报. 2018, 13(4): 594–601 <https://dx.doi.org/10.11992/tis.201703047>

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>

 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.201911028

自步稀疏最优均值主成分分析

许子微, 陈秀宏

(江南大学 数字媒体学院, 江苏 无锡 214122)

摘 要: 主成分分析 (PCA) 是一种无监督降维方法。然而现有的方法没有考虑样本的差异性, 且不能联合地提取样本的重要信息, 从而影响了方法的性能。针对以上问题, 提出自步稀疏最优均值主成分分析方法。模型以 $L_{2,1}$ 范数定义损失函数, 同时用 $L_{2,1}$ 范数约束投影矩阵作为正则化项, 且将均值作为在迭代中优化的变量, 这样可一致地选择重要特征, 提高方法对异常值的鲁棒性; 考虑到训练样本的差异性, 利用自步学习机制实现训练样本由“简单”到“复杂”的学习过程, 有效地降低异常值的影响。理论分析和实验结果表明, 以上方法能更有效地降低异常值对分类精度的影响, 提高分类精度。

关键词: 图像处理; 主成分分析; 无监督学习; 数据降维; 稀疏; 最优均值; 自步学习; 人脸识别

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2021)03-0416-09

中文引用格式: 许子微, 陈秀宏. 自步稀疏最优均值主成分分析 [J]. 智能系统学报, 2021, 16(3): 416-424.

英文引用格式: XU Ziwei, CHEN Xiuhong. Sparse optimal mean principal component analysis based on self-paced learning[J]. CAAI transactions on intelligent systems, 2021, 16(3): 416-424.

Sparse optimal mean principal component analysis based on self-paced learning

XU Ziwei, CHEN Xiuhong

(School of Digital Media, Jiangnan University, Wuxi 214122, China)

Abstract: Principal component analysis (PCA) can be referred to as an unsupervised dimensionality reduction approach. However, the existing methods do not consider the difference of samples and cannot jointly extract important information of samples, thus affecting the performance of some methods. For the above problems, based on self-paced learning, we proposed a sparse optimal mean PCA algorithm. In our model, loss of function is defined by $L_{2,1}$ norm, the projection matrix is regularized by $L_{2,1}$ norm, and the mean value is taken as a variable to be optimized in the iteration. In this way, important features can be consistently selected, and the robustness of the method to outliers can be improved. Considering the difference in training samples, we utilized self-paced learning mechanism to complete the learning process of training samples from “simple” to “complex” so as to effectively reduce the influence of outliers. Theoretical analysis and the empirical study revealed that the proposed method could effectively reduce the influence of noise or outliers on the classification progress, thus improving the effect of the classification.

Keywords: image processing; principal component analysis; unsupervised learning; data dimension deduction; sparse; optimal mean; self-paced learning; face recognition

降维^[1]是数据分类中的一个重要问题, 其目的是学习一个变换矩阵, 将高维数据投影到低维子空间中, 使数据在低维子空间中得到有效地分

类。经典的降维方法包括主成分分析 (principal component analysis, PCA)^[2-3] 和线性判别分析 (linear discriminant analysis, LDA)^[4-5], 其中 PCA 是一种将数据信息投影到正交线性空间的无监督方法, 原始 PCA 在对数据降维时, 以平方欧氏距离来表示损失, 而欧氏距离对噪声及异常值敏感。

收稿日期: 2019-11-19.

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: 许子微. E-mail: 18256515269@163.com.

于是提出许多 PCA 变体来降低异常值的影响,例如 Nie 等^[6-7]提出非贪婪 L_1 范数最大化的鲁棒主成分分析 (robust principal component analysis, RPCA), 和基于 $L_{2,1}$ 范数的最优均值主成分分析 (optimal mean robust principal component analysis, OMRPCA) 来同时学习最优变换矩阵和最优均值。但通过 RPCA 和 OMRPCA 等模型得到的低维子空间的每个新特征都是高维空间中所有原始特征的线性组合, 由于存在冗余特征, 通常不适合分类。Zou 等^[8]提出了稀疏主成分分析 (sparse principal component analysis, SPCA)。然而, 因为在每个变换向量上施加 L_1 范数, 故 SPCA 不能联合地选择重要特征; Yi 等^[9]提出了联合稀疏主成分分析 (joint sparse principal component analysis, JSPCA), 利用 $L_{2,1}$ 范数来表示损失项和正则项, 在避免异常值影响的同时能联合地选择重要特征, 从而增强算法的分类精度。

以上方法的每一个样本对应一个损失值, 损失值的大小决定了样本对模型的贡献, 这就表明了样本之间存在差异性, 而这些方法均同等地对待所有训练样本, 没有考虑样本之间的差异性。受人类/动物学习过程的启发, 文献 [10-11] 提出了自定步长 (或课程) (self-paced learning, SPL) 学习理论, 其主要思想是以自定步长方式来实现从“简单”到“复杂”样本的训练, 最终学习得到一个成熟的模型。此后, Lu 等^[12]提出自步课程学习 (self-paced curriculum learning, SPCL) 同时考虑训练前已知的先验知识和训练过程中的学习进度。Yu 等^[13]提出了自步学习 K 均值聚类 (self-paced learning for k-means clustering algorithm, SPKC), 将自步学习机制用于聚类方法。Hu 等^[14]提出自步核联合稀疏表示高光谱图像分类 (kernel joint sparse representation based on self-paced learning for hyperspectral image classification, SPKJSR), 将自步学习机制用于高光谱图像分类。

本文为充分考虑样本之间的差异性, 利用 SPL 思想提出自步稀疏最优均值主成分分析 (sparse optimal mean principal component analysis based on self-paced learning, SPL-OMSPCA), 该模型放宽了对投影矩阵的正交约束, 通过对训练样本由“简单”到“复杂”的渐进学习过程, 得到最优投影矩阵和最优样本均值。该方法能更好地获得样本中重要的特征信息, 避免异常值的影响, 提高了算法的鲁棒性。在 6 个不同数据集上的实验表明, SPL-OMSPCA 算法优于目前最先进的主成分分析方法。

1 相关工作

给定训练样本为 $X = [x_1 \ x_2 \ \cdots \ x_n] \in \mathbf{R}^{m \times n}$, 其中 m 为原始图像的空间维数, n 为训练样本的个数。

1.1 PCA 模型

假设训练样本 X 已经中心化, 即均值为零, 则通常的 PCA 可以表示为

$$\min_W \|X - WW^T X\|_F^2, \text{ s.t., } W^T W = I \quad (1)$$

式中 W 为投影矩阵。该问题的最优解 W 的列对应于矩阵的前 m 个最大特征值对应的特征向量。

基于 L_2 范数的均值的合理性只针对传统的 PCA 方法, 泛化能力较差, 于是考虑以下最优均值 PCA 模型 (OMPCA):

$$\min_{b, W^T W} \|X - b\mathbf{1}^T - WW^T(X - b\mathbf{1}^T)\|_F^2$$

式中: $b \in \mathbf{R}^m$ 是一个待优化的变量; $\mathbf{1}$ 为 n 维全 1 列向量。于是, 投影矩阵 W 的列为矩阵 XXH^T 的前 d 个最大特征值对应的特征向量, 其中 $H = I - \frac{1}{n}(\mathbf{1}\mathbf{1}^T)$; 最优均值为 $b = \frac{1}{n}(X\mathbf{1} + W\alpha)$, $\alpha \in \mathbf{R}^d$ 为任意 d 维列向量。

若式 (1) 中的投影矩阵与恢复矩阵不相同, 并以 $L_{2,1}$ 范数来表示损失函数以及稀疏正则化项, 则有以下联合稀疏主成分分析 (JSPCA) 模型:

$$\min_{Q, P} \|X - PQ^T X\|_{2,1} + \lambda \|Q\|_{2,1}$$

式中: $Q \in \mathbf{R}^{m \times d}$ 是投影矩阵; $P \in \mathbf{R}^{m \times d}$ 为恢复矩阵。

1.2 自步学习 (SPL) 机制

受人类/动物学习过程的启发, SPL 以自定步长的方式实现从“简单样本”到“复杂样本”迭代地训练模型。其优化模型为

$$\min \sum_{i=1}^n v_i L(y_i, g(x_i, w)) + \alpha \phi(w) + \sum_{i=1}^n f(v_i, k)$$

$$\text{s.t. } v_i \in [0, 1], i = 1, 2, \dots, n$$

式中: α 为稀疏正则化控制参数; k 是自步学习参数; 变量 v_i 用于决定选择哪些样本参与训练模型; $L(y_i, g(x_i, w))$ 是样本的损失函数, 损失值的大小决定样本的难易程度; $f(v_i, k)$ 是自步正则化函数。在算法迭代过程中, 模型可以根据样本难易程度, 从简单的样本开始, 学习到复杂的样本。自步学习机制的有效性, 特别是对高度损坏的数据的鲁棒性, 已经在各种计算机视觉任务中得到了验证。

2 自步稀疏最优均值主成分分析模型 (SPL-OMSPCA)

2.1 模型建立

SPL-OMSPCA 的基本思想是: 以 $L_{2,1}$ 范数表示损失函数和稀疏约束, 并利用自步学习方案对

参与训练的样本进行选择。其中,损失函数是训练样本从低维变换空间中恢复原始数据产生的误差,且包含需优化的均值向量。于是有以下优化模型:

$$\begin{aligned} \min_{v,b,P,Q} \sum_{i=1}^n v_i \|x_i - b - PQ^T(x_i - b)\|_2 + \alpha \|Q\|_{2,1} + \\ \sum_{i=1}^n f(v_i, k) \end{aligned} \quad (2)$$

$$\text{s.t. } P^T P = I, v_i \in [0, 1], i = 1, 2, \dots, n$$

式中: x_i 为第 i 个样本; α 为稀疏正则化控制参数; k 为自步控制参数; 变量 v_i 用于决定选择哪些样本参与训练; $b \in \mathbf{R}^{m \times 1}$ 是样本均值; $Q \in \mathbf{R}^{m \times d}$ 是投影矩阵; $P \in \mathbf{R}^{m \times d}$ 是恢复矩阵且服从正交约束 $P^T P = I$ 。式(2)用损失值的大小决定样本的难易程度。

自步正则化函数 $f(v_i, k)$ 有多种形式。若 $f(v_i, k) = -\frac{1}{k} v_i$, k 值较大, 倾向于选择损失值较小的样本, 于是在算法迭代过程中, 利用一个步长参数 μ 来逐渐减小 k 的值, 使得越来越多的样本参与模型学习, 从而实现从简单样本学习到复杂样本的过程。当 k 小于一个预定义的确值时, 算法结束。

上述方法是硬性阈值权值选择方法, 通过给每个样本分配二元权值 ($v_i \in [0, 1]$) 来决定是否选择样本。但异常值在所有样本中分布不均匀, 所以硬性正则函数不能确定是否选择这些样本。相比于硬性阈值权值, 软权值给每个样本分配 0~1(包括 0 和 1) 的权值, 显示了样本的潜在重要性, 更好地实现了从简单样本逐步学习到复杂样本的过程。所提最终优化算法为

$$\begin{aligned} \min_{v,b,P,Q} \sum_{i=1}^n v_i \|x_i - b - PQ^T(x_i - b)\|_2 + \alpha \|Q\|_{2,1} + \\ \sum_{i=1}^n \frac{\beta^2}{v_i + \beta k} \end{aligned} \quad (3)$$

$$\text{s.t. } P^T P = I, v_i \in [0, 1], i = 1, 2, \dots, n$$

式中 β 是间隔控制参数, 控制 0~1 的模糊区间。因此软阈值权值可以通过明确样本间差异, 准确选择样本, 进一步避免异常值的影响。另外, 这里存在一个隐形变量 $\mu (\mu > 1)$, 用来作为自步控制参数 k 减小的步长。目标函数式(3)转化为

$$\begin{aligned} \min_{v,b,P,Q} \|(X - b\mathbf{1}^T - PQ^T(X - b\mathbf{1}^T)) \sqrt{V} \sqrt{D}\|_F^2 + \\ \alpha \|\sqrt{H}Q\|_F^2 + \sum_{i=1}^n \frac{\beta^2}{v_i + \beta k} \end{aligned} \quad (4)$$

$$\text{s.t. } P^T P = I, v_i \in [0, 1], i = 1, 2, \dots, n$$

式中: $V = \text{diag}(v_1, v_2, \dots, v_n)$; $\mathbf{1}$ 是全为 1 的列向量; 对角矩阵 $D = \text{diag}(d_{11}, d_{22}, \dots, d_{nn})$ 的对角元素定义为

$$d_{ii} = \begin{cases} 1/\|A - PQ^T A\|_2, & \|A - PQ^T A\|_2 \neq 0 \\ 1/\delta, & \text{其他} \end{cases} \quad (5)$$

式中: $A = X - b\mathbf{1}^T$; $i(i = 1, 2, \dots, n)$ 表示矩阵的第 i 列; 对角矩阵 $H = \text{diag}(h_{11}, h_{22}, \dots, h_{nn})$ 的对角元素定义为

$$h_{jj} = \begin{cases} 1/\|Q^j\|_2, & \|Q^j\|_2 \neq 0 \\ 1/\delta, & \text{其他} \end{cases} \quad (6)$$

式中: $j(j = 1, 2, \dots, m)$ 表示矩阵的第 j 行; δ 为一个很小的数。

2.2 模型求解

因为式(4)含有 P 、 Q 、 b 及 v 共 4 个变量, 直接求解非常困难, 所以本文使用交替迭代求解的方法, 分别求得自步学习权重 v 、最优投影矩阵 Q 、最优均值 b 以及恢复矩阵 P 。

2.2.1 固定 P 、 Q 、 b 求自步学习权重 v

当 P 、 Q 和 b 固定时, 式(4)转化为

$$\begin{aligned} \min_{v,b,P,Q} \sum_{i=1}^n v_i \|x_i - b - PQ^T(x_i - b)\|_2 + \sum_{i=1}^n \frac{\beta^2}{v_i + \beta k} \\ \text{s.t. } P^T P = I, v_i \in [0, 1], i = 1, 2, \dots, n \end{aligned}$$

令 $L_i = \|x_i - b - PQ^T(x_i - b)\|_2$, 则上式可表示为

$$\begin{aligned} \min_v \sum_{i=1}^n v_i L_i + \sum_{i=1}^n \frac{\beta^2}{v_i + \beta k} \\ \text{s.t. } v_i \in [0, 1], i = 1, 2, \dots, n \end{aligned} \quad (7)$$

式(7)的目标函数关于 v_i 求偏导并令其为 0 得 v_i 的解

$$v_i = \begin{cases} 1, & L_i \leq 1/(k + 1/\beta)^2 \\ 0, & L_i \geq 1/k^2 \\ \beta \left(\frac{1}{\sqrt{L_i}} - k \right), & \text{其他} \end{cases} \quad (8)$$

2.2.2 固定 P 、 Q 、 v 求解最优均值 b

当固定 P 、 Q 和 v 时, 式(4)转化为

$$\min_b \|(X - b\mathbf{1}^T - PQ^T(X - b\mathbf{1}^T)) \sqrt{V} \sqrt{D}\|_F^2 \quad (9)$$

式(9)的目标函数关于 b 求偏导并令其为零, 得到

$$(I - PQ^T - QP^T + QQ^T)(b\mathbf{1}^T - X)VD\mathbf{1} = 0 \quad (10)$$

令 $c = (b\mathbf{1}^T - X)VD\mathbf{1}$, 则式(10)转化为

$$(I - PQ^T - QP^T + QQ^T)c = 0 \quad (11)$$

若 $I - PQ^T - QP^T + QQ^T$ 为非奇异的, 则式(11)的解为 $c = 0$, 从而有

$$b = \frac{XVD\mathbf{1}}{\mathbf{1}^T VD\mathbf{1}} \quad (12)$$

若 $I - PQ^T - QP^T + QQ^T$ 为奇异的, 设其奇异值分解为 $E_1 W_1 U_1^T$, 其中 $W_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_s, 0, \dots, 0)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s > 0$, 则式(11)的解为

$$c = U_1 z$$

其中 $z = [0, \dots, 0, z_{s+1}, \dots, z_m]^T \in \mathbf{R}^m$ 为任意的。特别地, 如令 $z_{s+1} = 1, z_{s+2} = \dots = z_m = 0$, 则式(11)的一个

特解为 U_1 的第 $s+1$ 列, 即 $c = (U_1)_{s+1}$, 从而有

$$b = \frac{(U_1)_{s+1} + XVD\mathbf{1}}{\mathbf{1}^T VD\mathbf{1}} \quad (13)$$

2.2.3 固定 P 、 v 、 b 求解投影矩阵 Q

固定 P 、 v 、 b 后, 式 (4) 转化为

$$\min_Q \left\| (X - b\mathbf{1}^T - PQ^T(X - b\mathbf{1}^T)) \sqrt{V} \sqrt{D} \right\|_F^2 + \alpha \| \sqrt{H} Q \|_F^2 \quad (14)$$

令 $(X - b\mathbf{1}^T) \sqrt{V} \sqrt{D} = G$, 则式 (14) 转化为

$$\min_Q \| G - PQ^T G \|_F^2 + \alpha \| \sqrt{H} Q \|_F^2 \quad (15)$$

式 (15) 的目标函数关于 Q 求偏导, 并令其为 0 得

$$(GG^T + \alpha H)Q = GG^T P$$

从而有

$$Q = (GG^T + \alpha H)^{-1} GG^T P \quad (16)$$

2.2.4 固定 Q 、 v 、 b 求解恢复矩阵 P

固定 Q 、 v 、 b 后, 式 (4) 转化为

$$\min_P \left\| ((X - b\mathbf{1}^T) - PQ^T(X - b\mathbf{1}^T)) \sqrt{V} \sqrt{D} \right\|_F^2$$

$$\text{s.t. } P^T P = I$$

或等价形式为

$$\min_P \left\| \sqrt{D} \sqrt{V} (X - b\mathbf{1}^T)^T - \sqrt{D} \sqrt{V} (X - b\mathbf{1}^T)^T Q P^T \right\|_F^2$$

$$\text{s.t. } P^T P = I \quad (17)$$

约束条件 $P^T P = I$ 意味着 P 是列正交矩阵。设 $(X - b\mathbf{1}^T)VD(X - b\mathbf{1}^T)^T Q$ 的奇异值分解为 $E_2 W_2 U_2^T$, 则由文献 [9] 中定理知, 式 (17) 的最优解为

$$P = E_2 U_2^T \quad (18)$$

以上过程不断迭代, 直到收敛条件满足为止, 具体算法如算法 1。

算法 1 SPL-OMSPCA

输入 样本 X , 维度 d , 参数 k 、 α 、 β 和步长 μ ;

输出 投影矩阵 Q 。

迭代执行以下操作, 直到模型收敛:

- 1) 通过式 (8) 计算 v ;
- 2) 通过式 (12)、(13) 计算 b ;
- 3) 通过式 (16) 计算 Q ;
- 4) 通过式 (18) 计算 P ;
- 5) 通过式 (5) 计算 D ;
- 6) 通过式 (6) 计算 H ;
- 7) 更新 k : $k = k/\mu$ 。

得到投影矩阵 Q 之后, 运用最近邻分类器 (nearest neighbor classifier, NN) 进行分类:

$$c = \operatorname{argmin}_{c_i} \left(\sum_{j=1}^m (X_i^j - y_j)^2 \right)^{1/2} \quad (19)$$

式中: c 为测试样本 y 的预测标签; c_i 为第 i 个样

本的类标签。

在算法 1 中, 主要的计算复杂度为奇异值分解和矩阵求逆, 故本算法时间复杂度最高为 $O(m^3)$, 假设算法迭代 T 次, 则该部分的时间复杂度为 $O(Tm^3)$, 从而整个算法的时间复杂度为 $O(Tm^3)$ 。

3 实验

为评估 SPL-OMSPCA 算法的有效性, 本文将在 UMIST^[16]、JAFPE^[17]、AR^[18]、BIO^[19]、COIL20^[20] 及 MNIST^[21] 数据集上进行测试, 每次实验独立随机, 重复 20 次取平均识别率和标准差作为最后实验结果, 并与 PCA、SPCA、OMRPCA 和 JSPCA 算法作对比。

3.1 数据集

为测试 SPL-OMSPCA 对异常值的鲁棒性, 将 UMIST 数据集的每个图像随机添加 13×13 的块遮挡; 随机抽取 JAFPE 数据集 50% 的图像添加 13×13 块遮挡, 而对 AR、BIO、COIL20 和 MNIST 数据集随机添加 10% 的椒盐噪声。各数据集的具体特征如表 1 所示, 而各数据集的部分图像如图 1。

表 1 不同数据集的属性及特点

Table 1 Attributes and characteristics of different datasets

数据集	样本数	类别数	规模	特征
UMIST	575	20	28×32	有较大姿态变化
JAFPE	200	10	32×32	有面部表情变化
AR	3 120	120	32×32	有光照表情变化
BIO	1 460	22	32×32	样本较相似
COIL20	1 440	20	32×32	姿态间隔拍摄
MNIST	1 000	10	28×28	样本较相似

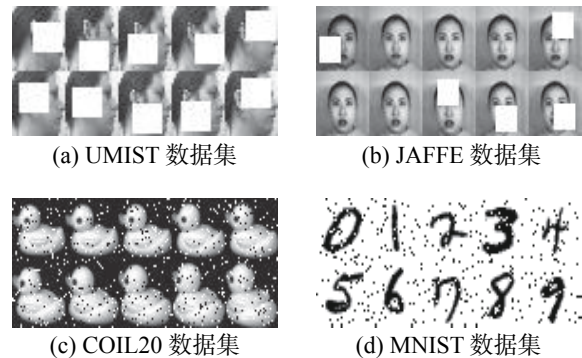


图 1 部分含噪数据集的可视化

Fig. 1 Visualization of some datasets with noise

3.2 实验结果及分析

在各数据集的每类中随机选取 L 张图像作为

训练样本,其余作为测试样本。图2给出了每个算法对不同数据集,在不同子空间维度上的分类精度。对UMIST、JAFFE和COIL20数据集,分别取9、10和5,则由图2(a)、(b)和(e)可看出, SPL-OMSPCA 整体优于其他对比算法;对AR数据集,取13,由图2(c)可知, SPL-OMSPCA 明显优于 OMRPCA 和 SPCA,而略优于其他算法;对BIO和MNIST数据集,分别选取20和10,由图2(d)与2(f)可以看出,由于此类数据集中的样本较为相似,当子

空间维数较低时, SPL-OMSPCA 与其他对比算法几乎保持一致的精度,但随着子空间维数的增加,本文算法精度持续保持平稳状态,且明显优于其他算法。表2列出了不同训练样本数下,5种算法在不同数据集上的分类精度与标准差,表中数据为子空间维数从5~100变化时,每个算法取得的最高精度,其中黑色加粗字体表示训练样本相同时,同一设置下几种算法中的最高分类精度。

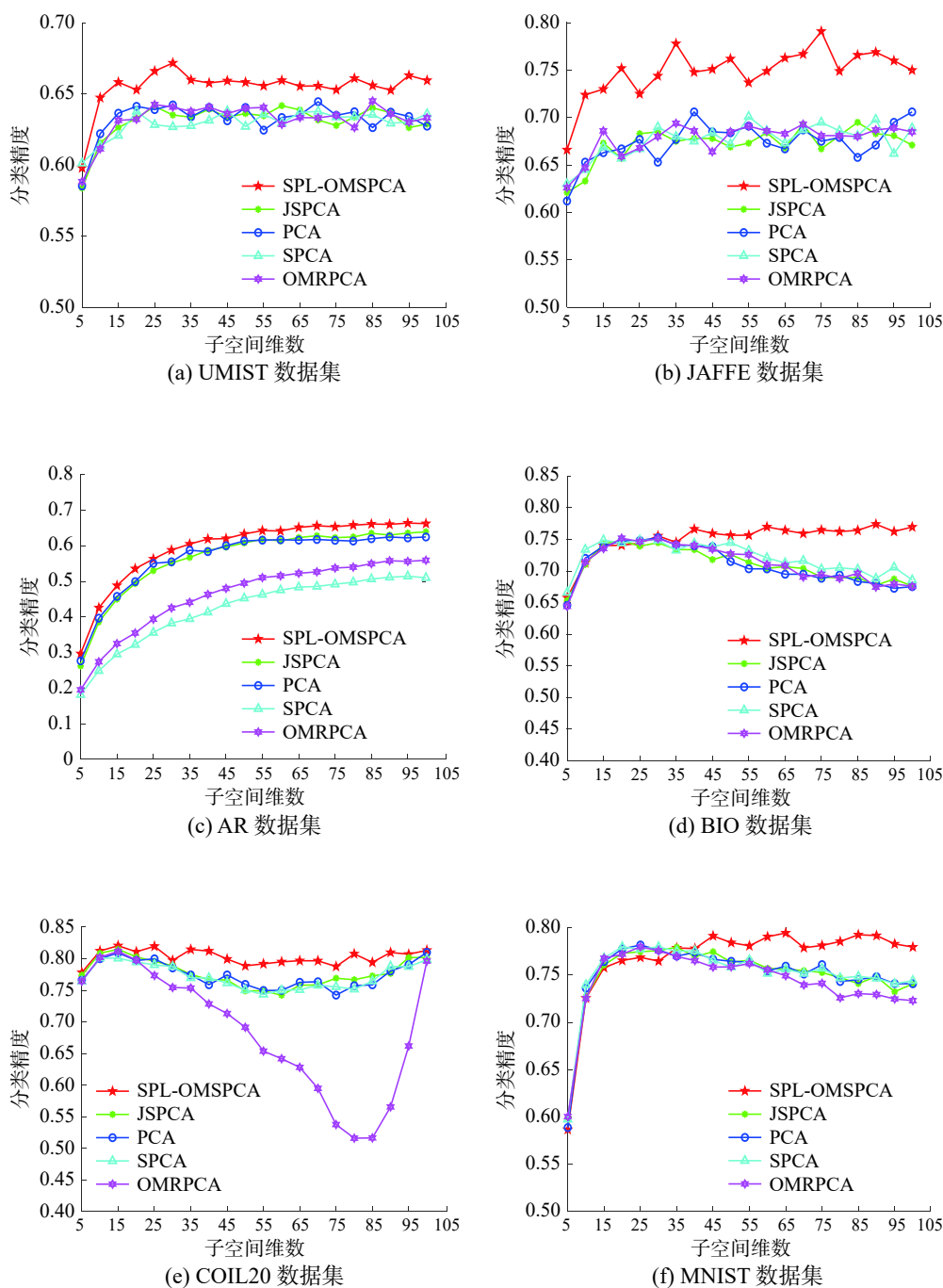


图2 不同算法在6个数据集上不同子空间维度的分类精度

Fig. 2 Classification accuracies of different algorithm in different subspace dimensions on six datasets

表2 在6个不同噪声数据集上各种PCA算法的分类精度及其标准差

Table 2 Average classification accuracy and standard deviation of various PCA algorithm on six data sets with different noise training samples

数据集	L	PCA	SPCA	JSPCA	OMRPCA	SPL-OMSPCA
UMIST (13×13遮挡)	6	0.5980±0.0222	0.5853±0.0449	0.5936±0.0200	0.5897±0.0292	0.6198±0.0220
	9	0.6430±0.0253	0.6377±0.0225	0.6418±0.0316	0.6425±0.0324	0.6718±0.0377
	12	0.6872±0.0203	0.6761±0.0373	0.6764±0.0412	0.6830±0.0382	0.7185±0.0128
	Mean	0.6427±0.0226	0.6330±0.0349	0.6373±0.0309	0.6384±0.0332	0.6700±0.0242
JAFPE (13×13遮挡)	5	0.6520±0.0587	0.6447±0.0553	0.6447±0.0580	0.6413±0.0920	0.7100±0.0367
	10	0.7060±0.0660	0.7010±0.0960	0.6950±0.0450	0.6940±0.0560	0.7910±0.0590
	15	0.7260±0.0660	0.7280±0.0680	0.7240±0.0640	0.7340±0.0660	0.8120±0.0640
	Mean	0.6947±0.0636	0.6912±0.0731	0.6879±0.0730	0.6898±0.0713	0.7710±0.0532
AR (10% 椒盐噪声)	8	0.5456±0.0214	0.5669±0.0174	0.5444±0.0263	0.5681±0.0205	0.5744±0.0255
	13	0.6450±0.0135	0.6509±0.0280	0.6495±0.0105	0.6517±0.0279	0.6659±0.0197
	18	0.7368±0.0091	0.7292±0.0292	0.7211±0.0211	0.7397±0.0207	0.7440±0.0240
	Mean	0.6424±0.0147	0.6490±0.0249	0.6383±0.0193	0.6532±0.0230	0.6614±0.0231
BIO (10% 椒盐噪声)	18	0.7391±0.0222	0.7424±0.0263	0.7438±0.0365	0.7362±0.0172	0.7621±0.0131
	20	0.7483±0.0258	0.7589±0.0286	0.7480±0.0206	0.7520±0.0196	0.7741±0.0206
	22	0.7661±0.0116	0.7769±0.0168	0.7681±0.0364	0.7682±0.0223	0.7860±0.0247
	Mean	0.7512±0.0199	0.7594±0.0239	0.7533±0.0312	0.7521±0.0197	0.7741±0.0195
COIL20 (10% 椒盐噪声)	4	0.7864±0.0156	0.7735±0.0306	0.7836±0.0215	0.7840±0.0296	0.7895±0.0392
	5	0.8100±0.0245	0.8011±0.0202	0.8149±0.0239	0.8116±0.0350	0.8205±0.0265
	6	0.8357±0.0190	0.8394±0.0227	0.7939±0.0251	0.8476±0.0365	0.8522±0.0211
	Mean	0.8107±0.0197	0.8047±0.0245	0.8201±0.0215	0.8144±0.0337	0.8207±0.0283
MINIST (10% 椒盐噪声)	15	0.7515±0.0209	0.7449±0.0273	0.7528±0.0284	0.7439±0.0329	0.7718±0.0271
	20	0.7819±0.0156	0.7780±0.0228	0.7795±0.0249	0.7794±0.0194	0.7928±0.0165
	25	0.8016±0.0211	0.8021±0.0208	0.8019±0.0275	0.8017±0.0231	0.8168±0.0301
	Mean	0.7783±0.0192	0.7750±0.0236	0.7781±0.0269	0.7750±0.0251	0.7938±0.0246

为进一步证明实验的合理性,选取 JAFPE 和 COIL20 数据集, L 分别取 10 和 9, 子空间维度取 10。SPL-OMSPCA 模型中,每一次迭代中更新均值,图 3 展示了 2 个数据集从 1024 个特征中随机抽取的 10 个特征的均值变化,由图 3 可看出均值随迭代次数增加逐渐趋于平稳。模型考虑样本自身差异性,引入自步学习机制,利用损失值判断样本的难易程度,赋予样本软权重 v_i ,并且放宽对投影矩阵的正交约束,以 $L_{2,1}$ 范数来表示损失函数和稀疏约束,由此即使对于有噪声的样本,也能极大程度地进行重构,图 4 展示了对于随机添加 50% 块噪声的 JAFPE 数据集,5 种算法得到的

部分重构图像,明显 SPL-OMSPCA 模型对含噪声块图像的重构更好。表 3 展示了迭代过程中 JAFPE 数据集部分样本的 v_i 的变化情况,其中一行表示一个样本的变化。从表 3 可看出,随迭代次数的增加,对损失值大小的限制逐渐放宽,实现从“简单样本”到“复杂样本”的学习。此外,图 5 给出了算法在 UMIST 和 JAFPE 上的收敛情况,其中,训练样本数 L 分别为 9、10,子空间维数为 100,最大迭代次数为 30。由图 5 可见,本文所提出的算法具有收敛性。

综上所述, SPL-OMSPCA 模型对噪声样本具有更高的鲁棒性,且能取得比其他算法更高的识别率。

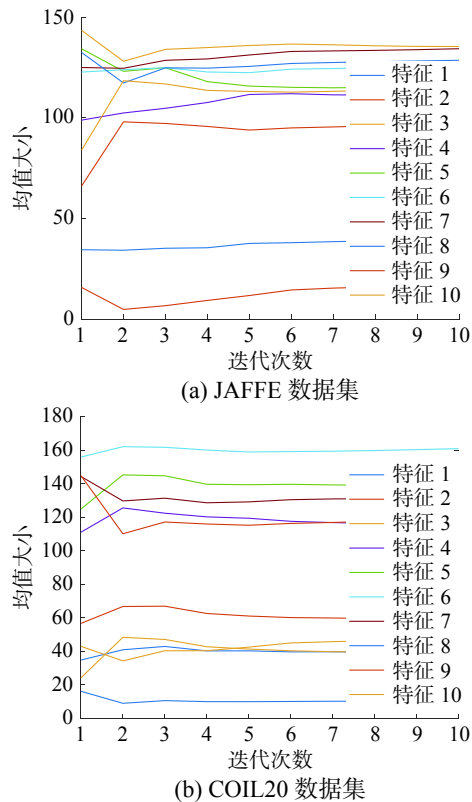


图3 均值变化图

Fig. 3 Graph of the mean value

表3 自步软权重 v_i 在 JAFFE 数据集上的变化Table 3 Change of self-stepping soft weight v_i on JAFFE dataset

循环次数	1	2	3	4	5	6	7	8	9	10
v_1	0.2211	0.4198	0.5108	0.6372	0.7597	0.8402	0.8871	0.9123	0.9247	0.9303
v_2	0.3248	0.5901	0.8300	1	1	1	1	1	1	1
v_3	1	1	1	1	1	1	1	1	1	1
v_4	0.0077	0.0435	0.1781	0.1431	0.1934	0.2112	0.2139	0.2162	0.2175	0.2181
v_5	0	0	0	0	0	0	0	0	0	0
v_6	0.6918	1	1	1	1	1	1	1	1	1
v_7	0	0	0	0	0	0	0	0	0	0
v_8	0.4027	0.7193	1	1	1	1	1	1	1	1
v_9	0	0	0.1253	0.4601	0.5323	0.5571	0.5819	0.5902	0.5921	0.5999
v_{10}	1	1	1	1	1	1	1	1	1	1

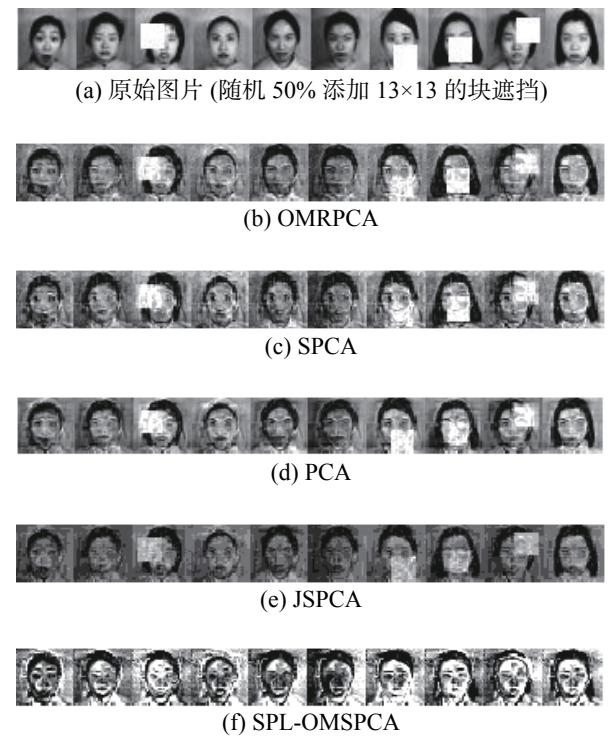
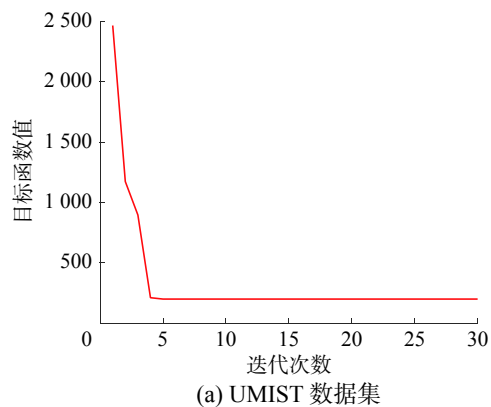


图4 不同算法对部分 JAFFE 数据集的重构图像

Fig. 4 Reconstructed images of partial JAFFE datasets by different algorithms

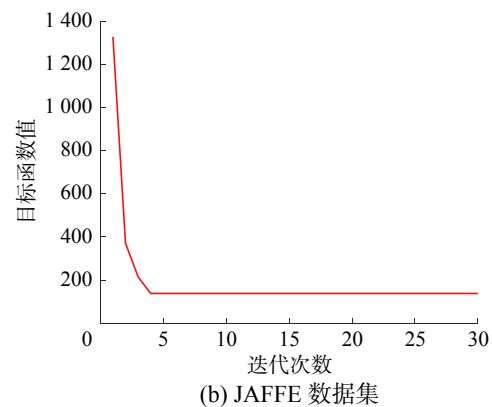


图5 目标函数收敛图

Fig. 5 Convergence graph of objective value

3.3 参数设置

由式(3)提出的问题可知,参数 k 和 β 的值决定了学习过程中样本的选取,故选择合适的参数可以提升算法的分类效果。假设初始训练时最大损失为 L_m ,由式(3)取适当的 k 和 β 满足:

$$L_m = 1/(k+1/\beta)^2 \quad (20)$$

为简化计算,令 $k=1/\beta$,则有

$$k = 1/2\sqrt{L_m}, \quad \beta = 2\sqrt{L_m} \quad (21)$$

对步长参数 μ 和稀疏正则化控制参数 α 将通过实验确定,由图6可见,SPL-OMSPCA算法在不同参数组合下得到不同分类效果。具体地,UMIST数据集上 $L=9$ 时,在 $(\alpha, \mu) = (0.001, 1.25)$ 处可取得最高精度;JAFPE数据集上 $L=10$ 时,在 $(\alpha, \mu) = (10, 1.15)$ 处取得最高精度;AR数据集上 $L=13$ 时,在 $(\alpha, \mu) = (0.01, 1.15)$ 处取得最高精度;BIO数据集上 $L=20$ 时,在 $(\alpha, \mu) = (1\ 000, 1.1)$ 时取得最高精度;COIL20数据集上 $L=5$ 时,在 $(\alpha, \mu) = (1\ 000, 1.15)$ 时取得最高精度;MNIST数据集上 $L=20$ 时,在 $(\alpha, \mu) = (1\ 000, 1.15)$ 取到最高精度。

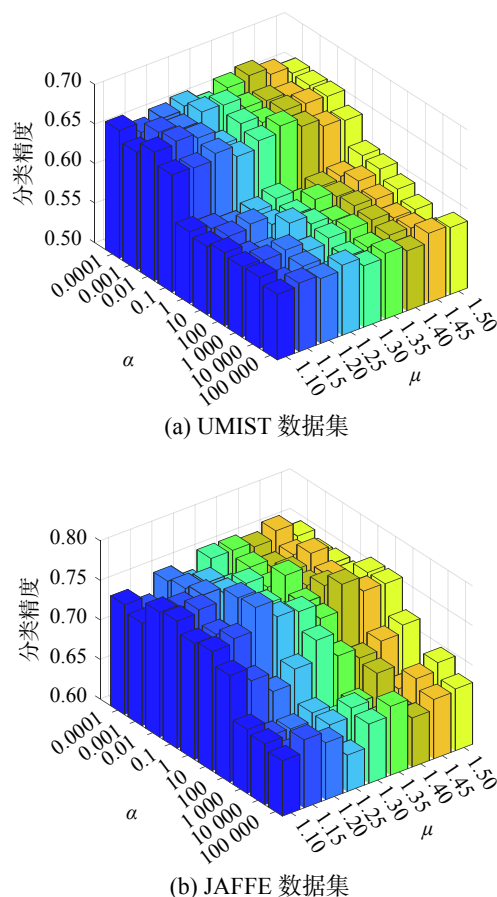


图6 不同参数上的分类精度

Fig. 6 Classification accuracy on different parameters

4 结束语

本文提出了一种新的无监督降维算法——自步稀疏最优均值主成分分析(SPL-OMSPCA)算法。基于对样本自身差异性的考虑,引入自步学习的思想,实现样本从“简单”到“复杂”的训练,且在迭代过程中,自动更新均值,而后通过对投影矩阵添加 $L_{2,1}$ 范数的行稀疏约束,一致选择样本,进一步提高算法的识别精度。理论分析和实验结果都证明了SPL-OMSPCA模型的分类优势。虽然SPL-OMSPCA的性能优于其他PCA方法,但它还是一种无监督方法,不能使用类标签来提取有区别的特征,因此它还不能提取最有效的分类特征。在未来,可将之扩展到半监督分类问题。

参考文献:

- [1] OU Weihua, YOU Xinge, TAO Dacheng, et al. Robust face recognition via occlusion dictionary learning[J]. *Pattern recognition*, 2014, 47(4): 1559–1572.
- [2] WEI Lai, ZHOU Rigui, YIN Jun, et al. Latent graph-regularized inductive robust principal component analysis[J]. *Knowledge-based systems*, 2019, 177: 68–81.
- [3] HE Jinrong, BI Yingzhou, LIU Bin, et al. Graph-dual Laplacian principal component analysis[J]. *Journal of ambient intelligence and humanized computing*, 2019, 10(8): 3249–3262.
- [4] ZHAO Haifeng, WANG Zheng, NIE Feiping. A new formulation of linear discriminant analysis for robust dimensionality reduction[J]. *IEEE transactions on knowledge and data engineering*, 2018, 31(4): 629–640.
- [5] 朱焕荣, 郑智超, 孙怀江. 面向局部线性回归分类器的判别分析方法[J]. *智能系统学报*, 2019, 14(5): 959–965.
- [6] ZHU Huanrong, ZHENG Zhichao, SUN Huaijiang. Locality-regularized linear regression classification-based discriminant analysis[J]. *CAAI transactions on intelligent systems*, 2019, 14(5): 959–965.
- [7] NIE Feiping, HUANG Heng, DING C, et al. Robust principal component analysis with non-greedy ℓ_1 -norm maximization[C]//Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Catalonia, Spain, 2011: 1433–1438.
- [8] NIE Feiping, YUAN Jianjun, HUANG Heng. Optimal mean robust principal component analysis[C]//Proceedings of the 31st International Conference on Machine Learning. Beijing, China, 2014: 1062–1070.
- [9] ZOU Hui, HASTIE T, TIBSHIRANI R. Sparse principal component analysis[J]. *Journal of computational and graphical statistics*, 2006, 15(2): 265–286.

- [9] YI Shuangyan, LAI Zhihui, HE Zhenyu, et al. Joint sparse principal component analysis[J]. *Pattern recognition*, 2017, 61: 524–536.
- [10] BENGIO Y, LOURADOUR J, COLLOBERT R, et al. Curriculum learning[C]//Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Quebec, Canada, 2009: 41–48.
- [11] KUMAR M P, PACKER B, KOLLER D. Self-paced learning for latent variable models[C]//Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada, 2010: 1189–1197.
- [12] JIANG Lu, MENG Deyu, ZHAO Qian, et al. Self-paced curriculum learning[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, Texas, USA, 2015: 2694–2700.
- [13] YU Hao, WEN Guoqiu, GAN Jiangzhang, et al. Self-paced learning for K -means clustering algorithm[J]. *Pattern recognition letters*, 2020, 132: 69–75.
- [14] HU Sixiu, PENG Jiangtao, FU Yingxiong, et al. Kernel joint sparse representation based on self-paced learning for hyperspectral image classification[J]. *Remote sensing*, 2019, 11(9): 1114–1132.
- [15] HAJINEZHAD D, SHI Qingjiang. Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications[J]. *Journal of global optimization*, 2018, 70(1): 261–288.
- [16] HOU Chenping, NIE Feiping, YI Dongyun, et al. Feature selection via joint embedding learning and sparse regression[C]//Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Spain, 2011: 1324–1329.
- [17] LYONS M J, BUDYNEK J, AKAMATSU S. Automatic classification of single facial images[J]. *IEEE transactions on pattern analysis and machine intelligence*, 1999, 21(12): 1357–1362.
- [18] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation.[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2009, 31(2):210-227.
- [19] JESORSKY O, KIRCHBERG K J, FRISCHHOLZ R W. Robust face detection using the hausdorff distance[C]// Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication. Halmstad, Sweden, 2001: 90–95.
- [20] ZHENG Miao, BU Jiajun, CHEN C, et al. Graph regularized sparse coding for image representation[J]. *IEEE transactions on image processing*, 2011, 20(5): 1327–1336.
- [21] KUSSL E, BAIDYK T. Improved method of handwritten digit recognition tested on MNIST database[J]. *Image and vision computing*, 2004, 22(12): 971–981.

作者简介:



许子微, 硕士研究生, 主要研究方向为模式识别、图像处理。



陈秀宏, 教授, 主要研究方向为数字图像处理和模式识别、目标检测与跟踪、优化理论与方法。参与国家自然科学基金项目 2 项, 主持省部级研究项目 3 项, 省博士后基金 1 项。获得省部级奖 1 项、校级教学成果奖 1 项、市厅级奖 6 项。发表学术论文 120 余篇。