

对数变换主成分分析的图像识别

宋昱^{1,2,3}, 孙文贇^{1,2,3}, 陈昌盛^{1,2,3}

(1. 深圳大学电子与信息工程学院, 518060, 广东深圳; 2. 深圳大学深圳市媒体信息安全重点实验室, 518060, 广东深圳; 3. 深圳大学广东省智能信号处理重点实验室, 518060, 广东深圳)

摘要: 为了进一步提升现有鲁棒主成分分析(PCA)算法处理含有异常样本数据的性能,提出了对数变换的 PCA 算法。为降低异常样本对目标函数值的影响,根据对数函数的性质建立了对数变换 PCA 算法的目标函数,证明了所提目标函数值小于标准 PCA 算法的。之后,给出了求解所提目标函数的一种优化算法,即为所提算法。通过迭代计算对角矩阵和进行特征值分解进行优化,证明了所提算法可以近似收敛于所提目标函数的最优解。分析证明了所提算法具有旋转不变性。为了更好地比较各算法处理异常样本数据的能力,使用 AR 数据库中的原样本作为异常样本,在其他数据库中人为添加了异常样本。与标准 PCA 算法、鲁棒 PCA 算法、包括贪婪求解的基于 l_1 范数的 PCA 算法、非贪婪求解的基于 l_1 范数的 PCA 算法、基于 $l_{2,p}$ 范数的 PCA 算法和基于最大相关熵的 PCA 算法在 AR、Extended Yale B、CMU PIE 这 3 个人脸数据库和 MNIST 这 1 个手写字符数据库进行了实验对比,结果表明:所提算法均得到了最低的重构误差和最高的识别精度;所提算法具有良好的收敛性能,一般迭代 5 到 6 次即可收敛。

关键词: 图像识别;主成分分析;异常样本;对数变换

中图分类号: TP391.4 **文献标志码:** A

DOI: 10.7652/xjtuxb202101005 **文章编号:** 0253-987X(2021)01-0033-10



OSID 码

Logarithm Transformation Based Principal Component Analysis
for Image Recognition

SONG Yu^{1,2,3}, SUN Wenyun^{1,2,3}, CHEN Changsheng^{1,2,3}

(1. College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China;

2. Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060, China;

3. Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China)

Abstract: To further improve the capability of the existing robust principal component analysis (PCA) in dealing with data containing outliers, a principal component analysis method based on logarithm transformation is proposed. For decreasing the interference of outliers to the value of the objective function and based on the property of the logarithm function, the objective function of the proposed logarithm transformation based PCA is established. It is theoretically proved that the objective value of the proposed algorithm is smaller than that of the standard PCA algorithm. An optimization algorithm for solving the proposed objective function is given. The proposed objective function is optimized by computing a diagonal matrix and eigenvalue decomposition iteratively. The optimization algorithm can approximately converge to the optimal solution of the proposed objective function. The rotation invariance property of the proposed algorithm is also

analyzed. We use 3 face datasets (AR, Extended Yale B, CMU PIE) and 1 handwritten digit dataset (MNIST) for experimentation. For better comparing the capability of existing algorithms in dealing with outliers, the original data samples in the AR datasets are taken as outliers. Some data samples are artificially added to the other datasets as outliers. Compared with standard PCA algorithm and the existing robust PCA algorithms, including greedy PCA algorithms based on l_1 norm, nongreedy PCA algorithms based on l_1 norm, PCA algorithms based on $l_{2,p}$ norm and PCA algorithms based on maximum correntropy, the proposed algorithm is endowed with the lowest reconstruction error and highest recognition accuracy. Analyzing the relationship between the iteration number and the value of the objective function, the proposed algorithm has good convergence, it usually converges after 5 or 6 iterations.

Keywords: image recognition; principal component analysis; outliers; logarithm transformation

主成分分析(PCA)^[1]是一种线性变换算法,在图像处理和机器学习中有许多重要的应用。PCA将高维数据投影至低维空间,可以有效地表示高维数据。然而,PCA也有一些缺点。因为PCA采用 l_2 范数的平方衡量原样本与重构样本的误差,所以当数据中有些样本明显远离其他样本,即数据中含有异常样本时,这些异常样本会对PCA的目标函数造成显著的影响^[2-5],从而导致标准PCA不能很好地处理含有异常样本的数据。

为了减小异常样本带来的负面效果,研究者提出了很多鲁棒PCA算法。文献[6]和文献[7]通过对样本进行最大似然估计,得到了基于 l_1 范数的PCA算法。文献[6]采用启发式的估计算法检测异常样本,文献[7]采用凸优化算法检测异常样本。文献[8]采用非凸 M 估计算子作为目标函数,学习得到彩色图像表示。尽管这些算法具有对异常样本的鲁棒性,但是有一个共同的缺点,即不具有旋转不变性,而旋转不变性是学习算法应具有的一个重要性质^[9]。为了保留标准PCA算法的旋转不变性,研究者提出了一些鲁棒的旋转不变PCA算法^[3-4,10-13]。文献[10]提出了PCA- L_1 ,采用贪婪算法最大化投影后样本的 l_1 范数。文献[11]提出了一种非贪婪算法,以此求解PCA- L_1 的目标函数,可以同时求解所有的投影方向。为了扩展PCA- L_1 算法,文献[12]提出了PCA- L_p 算法,该算法最大化投影后样本的 l_p 范数。PCA- L_1 或者PCA- L_p 并没有最小化重构误差,而是最大化投影后样本的 l_1 或者 l_p 范数,其目标函数与标准PCA的目标函数有本质不同。文献[3]和文献[4]提出了 R_1 -PCA,其目标函数中使用了2种范数,在样本空间维度上使用了 l_2 范数,在样本间使用了 l_1 范数。文献[13]提出了 Φ -PCA,其目标函数是一个二阶可导的凸函数,可以采

用牛顿迭代法求解。然而, R_1 -PCA和 Φ -PCA假设数据是均值的,该假设在实际中有一定的限制^[14]。为了估计样本均值,文献[15]提出了基于最大相关熵的PCA。为了进一步扩展 R_1 -PCA,文献[16]提出了基于 $l_{2,p}$ 范数的PCA。该算法采用 $l_{2,p}$ 范数衡量原样本与重构样本的误差,当 $p=1$ 时,该算法与 R_1 -PCA算法等价。

尽管研究者已经提出了众多的鲁棒PCA算法,但是这些算法在处理含有异常样本的数据时依然有一些不足。以基于 $l_{2,p}$ 范数的PCA算法为例,不能在理论上保证该算法的目标函数值小于标准PCA算法的,从而其处理异常样本的性能不能得到保证。为了改进现有的鲁棒PCA算法,本文提出了一种基于对数变换的鲁棒PCA算法,可以在理论上证明所提算法的目标函数值小于标准PCA算法的,从而提高现有鲁棒PCA算法处理含有异常样本数据的能力。所提算法保留了标准PCA算法的性质,即具有旋转不变性,且投影矩阵与数据协方差矩阵相关。实际数据集上的实验结果证明了所提算法的有效性。

1 标准PCA算法和鲁棒PCA算法

假设数据 $\mathbf{X}=[\mathbf{x}_1, \dots, \mathbf{x}_N]$,其中 $\mathbf{x}_i \in \mathbf{R}^d$ 是数据中的第 i 个样本, d 是维度。不失一般性地,假设数据是中心化的,即 $\sum_{i=1}^N \mathbf{x}_i = \mathbf{0}$ 。PCA算法^[1]希望找到一个投影矩阵 $\mathbf{W} \in \mathbf{R}^{d \times k}$,使得重构误差最小,其中 k 是投影向量数。最优的投影矩阵 $\mathbf{W}^*$ 可以表示为

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2, \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (1)$$

式中 $\mathbf{I}_k$ 是 $k \times k$ 的单位矩阵。通过代数运算,可以得出目标函数式(1)等价于

$$\mathbf{W}^* = \arg \max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{S}_i \mathbf{W}), \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (2)$$

式中 $\mathbf{S}_i = \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T$ 是数据的协方差矩阵。目标函数式(1)或式(2)的最优解是对应于矩阵 $\mathbf{S}_i$ 的 k 个最大特征值的特征向量。

从目标函数式(1)可以看出,PCA 在平方的意义上最小化重构误差,使得目标函数对于异常样本非常敏感^[3-5]。异常样本重构误差的平方非常大,对于目标函数有显著影响,会影响得到的投影矩阵。为此,研究者提出了很多鲁棒 PCA 算法,例如基于 l_1 范数的 PCA 算法^[11-12],该算法的目标函数为

$$\mathbf{W}_{l_1} = \arg \max_{\mathbf{W}} \sum_{i=1}^N \|\mathbf{W}^T \mathbf{x}_i\|_1, \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (3)$$

基于 l_1 范数的 PCA 算法通过求解目标函数式(3)得到投影矩阵。目标函数式(3)的解与协方差矩阵的关系并不清楚,而且因为 $\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_1 + \|\mathbf{W}^T \mathbf{x}_i\|_1 \neq \|\mathbf{x}_i\|_1$,目标函数式(3)并没有最小化重构误差。PCA 算法的目标函数最小化了重构误差,这说明目标函数式(3)与 PCA 算法的目标函数有本质不同。目标函数式(3)的解不等价于

$$\min_{\mathbf{W}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_1, \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (4)$$

目标函数式(4)直接考虑了数据的重构误差,并且对异常样本具有一定的鲁棒性。然而,目标函数式(4)不容易求解,其求解算法的计算量很大^[3],而且目标函数式(4)不具有旋转不变性。为了改进基于 l_1 范数的 PCA 算法,文献[16]中提出了基于 $l_{2,p}$ 范数的 PCA 算法,其目标函数为

$$\min_{\mathbf{W}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^p, \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (5)$$

式中 $0 < p < 2$ 。与基于 l_1 范数的 PCA 算法不同,基于 $l_{2,p}$ 范数的 PCA 算法最小化重构误差,而且 $0 < p < 2$ 降低了异常样本对于重构误差的影响,从而使求得投影矩阵更加准确。然而,当 $0 < p < 2$ 时,并不能从理论上保证目标函数式(5)的值一定小于目标函数式(1)的。

2 对数变换的 PCA 算法

2.1 目标函数建立

为了从理论上保证所提算法的目标函数值小于目标函数式(1)的,采用对数形式。当 $x \geq 0$ 时,有

$$x \geq \ln(x+1) \quad (6)$$

不等式(6)证明如下。令 $f(x) = x, g(x) = \ln(x+1)$,当 $x=0$ 时, $f(x) = g(x)$ 。考虑 $x > 0$ 的情形,对

不等式两边求导数,得到 $f'(x) = 1, g'(x) = 1/(x+1)$,当 $x > 0$ 时,有 $f'(x) > g'(x)$,即不等式左侧的导数大于不等式右侧的,而不等式两侧在 $x=0$ 时相等,所以当 $x > 0$ 时,有 $f(x) > g(x)$ 。从而证得不等式(6)。根据不等式(6)可知, $\ln(x+1)$ 是始终小于等于 x 的,这与 $x^{p'}$ 和 x 的关系不同,当 p' 取值在 $0 \sim 1$ 之间时(这里 x 表示误差的平方,故 p' 取值在 $0 \sim 1$ 之间), $x^{p'}$ 不总是小于 x 的。根据不等式(6),可以将 x 替换为某样本重构误差的平方,即 $\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2$,从而得到

$$\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 \geq \ln(\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1) \quad (7)$$

进一步得到

$$\sum_{i=1}^N \|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 \geq \sum_{i=1}^N \ln(\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1) \quad (8)$$

这样就从理论上保证了重构误差的减小,而不等式(8)的右侧即为本文采用的目标函数,即

$$\min_{\mathbf{W}} \sum_{i=1}^N \ln(\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1), \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (9)$$

2.2 目标函数求解

通过代数运算,可以得到

$$\begin{aligned} & \sum_{i=1}^N \ln(\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1) = \\ & \sum_{i=1}^N (\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1) (\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1)^{-1} \ln(\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1) = \\ & \sum_{i=1}^N (\text{tr}((\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i)^T (\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i)) + 1) d_i = \\ & \sum_{i=1}^N (\text{tr}(\mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i + \mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i) + 1) d_i = \\ & \sum_{i=1}^N (\text{tr}(\mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i) + 1) d_i \quad (10) \end{aligned}$$

式中 $d_i = (\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1)^{-1} \ln(\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i\|_2^2 + 1)$ 。将式(10)代入目标函数式(9),通过代数运算,可得

$$\begin{aligned} & \min_{\mathbf{W}} \sum_{i=1}^N \text{tr}(\mathbf{x}_i^T \mathbf{x}_i) d_i - \sum_{i=1}^N \text{tr}(\mathbf{x}_i^T \mathbf{W}\mathbf{W}^T \mathbf{x}_i) d_i + \\ & \sum_{i=1}^N d_i, \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (11) \end{aligned}$$

在目标函数式(11)中,存在 2 个未知变量 $\mathbf{W}$ 和 d_i ,且 d_i 依赖于 $\mathbf{W}$ 。目标函数式(11)没有闭式解,无法直接求解。可以通过固定 d_i 更新 $\mathbf{W}$,然后固定 $\mathbf{W}$ 更新 d_i 的方式进行求解。在第 $t+1$ 次迭代时,

若 $d_i^{(t)}$ 已知, 可以通过最小化目标函数式(11)更新 $\mathbf{W}$ 。在这种情况下, 目标函数的第1项和第3项都是常数。从而, 目标函数式(11)变为

$$\mathbf{W}^* = \arg \max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W}), \text{ s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}_k \quad (12)$$

式中 $\mathbf{D}$ 是一个对角矩阵, 其对角线上的元素是 d_i 。根据矩阵理论, 目标函数式(12)中 $\mathbf{W}^*$ 的列向量应是对应于矩阵 $\mathbf{X} \mathbf{D} \mathbf{X}^T$ 最大的 k 个特征值的特征向量。求解出矩阵 $\mathbf{W}$ 后, 再更新 d_i 。重复上述迭代过程直至收敛, 收敛条件为

$$\left| \frac{O_t - O_{t-1}}{O_{t-1}} \right| < \epsilon \quad (13)$$

式中: O_t 和 O_{t-1} 分别表示在第 t 次和第 $t+1$ 次目标函数式(9)的取值; ϵ 是一个小的常数, 取为 10^{-3} 。

求解目标函数式(9)的优化算法为对数变换的PCA, 即算法1, 伪代码如下。

输入: $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbf{R}^{d \times N}$, k

初始化: $\mathbf{W}_1 \in \mathbf{R}^{d \times k}$, 并且满足 $\mathbf{W}_1^T \mathbf{W}_1 = \mathbf{I}_k$, $t=1$

while 不收敛 do

1. 计算对角矩阵 $\mathbf{D}$, 其对角线上的元素为

$$d_i = (\|\mathbf{x}_i - \mathbf{W}_{t-1} \mathbf{W}_{t-1}^T \mathbf{x}_i\|_2^2 + 1)^{-1} \leq \ln(\|\mathbf{x}_i - \mathbf{W}_{t-1} \mathbf{W}_{t-1}^T \mathbf{x}_i\|_2^2 + 1)$$

2. 计算加权协方差矩阵 $\mathbf{X} \mathbf{D} \mathbf{X}^T$

3. 计算 $\mathbf{W}_t = \arg \max_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W})$ 。最优投影矩阵 $\mathbf{W}_t$ 的列是对应于矩阵 $\mathbf{X} \mathbf{D} \mathbf{X}^T$ 最大 k 个特征值的特征向量。

4. $t=t+1$

end while

输出: $\mathbf{W}_t \in \mathbf{R}^{d \times k}$

2.3 收敛性分析

定理1 在算法1的每一次迭代中, 有

$$\sum_{i=1}^N [\ln(\|\mathbf{x}_i - \mathbf{W}_{t+1} \mathbf{W}_{t+1}^T \mathbf{x}_i\|_2^2 + 1) \cdot \ln(\|\mathbf{x}_i - \mathbf{W}_t \mathbf{W}_t^T \mathbf{x}_i\|_2^2 + 1)] \leq \sum_{i=1}^N [\ln(\|\mathbf{x}_i - \mathbf{W}_t \mathbf{W}_t^T \mathbf{x}_i\|_2^2 + 1)]^2 \quad (14)$$

不等式(14)表明, 算法1在 $\ln(\|\mathbf{x}_i - \mathbf{W}_t \mathbf{W}_t^T \mathbf{x}_i\|_2^2 + 1)$ 加权的意义上减少了目标函数值。

证明 根据算法1, 在第 $t+1$ 次迭代时, 有

$$\text{tr}(\sum_{i=1}^N \mathbf{W}_{t+1}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{W}_{t+1} d_i^{(t)}) \geq \text{tr}(\sum_{i=1}^N \mathbf{W}_t^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{W}_t d_i^{(t)}) \quad (15)$$

不等式(14)两边乘以 -1 , 然后在两边加上

$$\sum_{i=1}^N \text{tr}(\mathbf{x}_i \mathbf{x}_i^T d_i^{(t)}) + d_i^{(t)}, \text{ 可得}$$

$$\sum_{i=1}^N \text{tr}(\mathbf{x}_i \mathbf{x}_i^T d_i^{(t)}) - \text{tr}(\mathbf{W}_{t+1}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{W}_{t+1} d_i^{(t)}) + d_i^{(t)} \leq \sum_{i=1}^N \text{tr}(\mathbf{x}_i \mathbf{x}_i^T d_i^{(t)}) - \text{tr}(\mathbf{W}_t^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{W}_t d_i^{(t)}) + d_i^{(t)} \quad (16)$$

对于任意的 i , 关系式 $\|\mathbf{x}_i - \mathbf{W} \mathbf{W}^T \mathbf{x}_i\|_2^2 = \text{tr}(\mathbf{x}_i \mathbf{x}_i^T) - \text{tr}(\mathbf{W}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{W})$ 成立, 所以不等式(15)变为

$$\sum_{i=1}^N (\|\mathbf{x}_i - \mathbf{W}_{t+1} \mathbf{W}_{t+1}^T \mathbf{x}_i\|_2^2 + 1) d_i^{(t)} \leq \sum_{i=1}^N (\|\mathbf{x}_i - \mathbf{W}_t \mathbf{W}_t^T \mathbf{x}_i\|_2^2 + 1) d_i^{(t)} \quad (17)$$

令 $\mathbf{e}_i^{(t+1)} = \mathbf{x}_i - \mathbf{W}_{t+1} \mathbf{W}_{t+1}^T \mathbf{x}_i$, $\mathbf{e}_i^{(t)} = \mathbf{x}_i - \mathbf{W}_t \mathbf{W}_t^T \mathbf{x}_i$,

并且根据 d_i 的定义, 不等式(17)变为

$$\sum_{i=1}^N (\|\mathbf{e}_i^{(t+1)}\|_2^2 + 1) \frac{\ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1)}{\|\mathbf{e}_i^{(t)}\|_2^2 + 1} \leq \sum_{i=1}^N \ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1) \quad (18)$$

在不等式(18)的两边同除以 $\sum_{i=1}^N \ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1)$, 可得

$$\sum_{i=1}^N \frac{\|\mathbf{e}_i^{(t+1)}\|_2^2 + 1}{\|\mathbf{e}_i^{(t)}\|_2^2 + 1} \frac{\ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1)}{\sum_{j=1}^N \ln(\|\mathbf{e}_j^{(t)}\|_2^2 + 1)} \leq 1 \quad (19)$$

假设 f 为凹函数, 当系数 λ_i 满足 $\sum_{i=1}^N \lambda_i = 1$ 时, 有

$$\sum_{i=1}^N \lambda_i f(x_i) \leq f(\sum_{i=1}^N \lambda_i x_i) \quad (20)$$

对数函数 $\ln(\cdot)$ 是凹函数, 满足式(20), 即

$$\sum_{i=1}^N \lambda_i \ln(x_i) \leq \ln(\sum_{i=1}^N \lambda_i x_i) \quad (21)$$

令 $\lambda_i = \ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1) / \sum_{j=1}^N \ln(\|\mathbf{e}_j^{(t)}\|_2^2 + 1)$,

$x_i = (\|\mathbf{e}_i^{(t+1)}\|_2^2 + 1) / (\|\mathbf{e}_i^{(t)}\|_2^2 + 1)$, 并代入不等式(21), 可得

$$\sum_{i=1}^N \ln\left(\frac{\|\mathbf{e}_i^{(t+1)}\|_2^2 + 1}{\|\mathbf{e}_i^{(t)}\|_2^2 + 1}\right) \frac{\ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1)}{\sum_{j=1}^N \ln(\|\mathbf{e}_j^{(t)}\|_2^2 + 1)} \leq \ln\left[\sum_{i=1}^N \frac{\|\mathbf{e}_i^{(t+1)}\|_2^2 + 1}{\|\mathbf{e}_i^{(t)}\|_2^2 + 1} \frac{\ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1)}{\sum_{j=1}^N \ln(\|\mathbf{e}_j^{(t)}\|_2^2 + 1)}\right] \quad (22)$$

结合不等式(19), 可以得到不等式(22)的右侧小于0, 所以可得

$$\sum_{i=1}^N \ln(\|\mathbf{e}_i^{(t+1)}\|_2^2 + 1) \ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1) \leq \sum_{i=1}^N [\ln(\|\mathbf{e}_i^{(t)}\|_2^2 + 1)]^2 \quad (23)$$

由此证明了不等式(14)成立。不等式(14)表

明,在 $\ln(\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1)$ 加权意义上,算法 1 在每次迭代中单调减少目标函数式(9)的值。从而在每次迭代中, $\mathbf{W}$ 近似朝着最优解移动。证毕。

定理 2 算法 1 近似收敛于目标函数式(9)的最优解处。

证明 目标函数式(9)是约束优化问题。采用拉格朗日乘子法,将其转换为非约束优化问题。目标函数式(9)对应的拉格朗日函数为

$$L(\mathbf{W}) = \sum_{i=1}^N \ln(\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1) + \text{tr}(\mathbf{A}^T (\mathbf{W}^T \mathbf{W} - \mathbf{I})) \quad (24)$$

式中 $\mathbf{A}$ 是拉格朗日乘子。为了满足约束条件 $\mathbf{W}^T \mathbf{W} = \mathbf{I}_k$, $\mathbf{A}$ 应设为对角矩阵。根据 KKT 条件,在最优解处有 $dL=0$ 。对式(24)取微分,可得

$$dL = 2 \sum_{i=1}^N \text{tr} \left(\frac{-\mathbf{x}_i \mathbf{x}_i^T \mathbf{W} (d\mathbf{W})^T}{\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1} \right) + \text{tr}(\mathbf{W} \mathbf{A}^T (d\mathbf{W})^T) \quad (25)$$

为了使得 $dL=0$,应满足

$$\sum_{i=1}^N \frac{\mathbf{x}_i \mathbf{x}_i^T \mathbf{W}}{\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1} = \mathbf{W} \mathbf{A}^T \quad (26)$$

目标函数式(12)的最优解在算法 1 的第 3 步得到,所以算法 1 收敛时的解满足目标函数式(12)的 KKT 条件。目标函数式(12)的拉格朗日函数为

$$L_2(\mathbf{W}) = \text{tr}(\mathbf{W}^T \mathbf{D} \mathbf{X} \mathbf{X}^T \mathbf{W}) + \text{tr}(\mathbf{A}^T (\mathbf{W}^T \mathbf{W} - \mathbf{I})) \quad (27)$$

对 L_2 取微分,并令其微分为 0,可得

$$\mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W} + \mathbf{W} \mathbf{A}^T = \mathbf{0} \quad (28)$$

式(28)与式(26)在形式上一致,主要的差别有 2 点:①式(28)中 $\mathbf{D}$ 的对角线元素是 $d_i = (\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1)^{-1} \ln(\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1)$,而式(26)如果写成矩阵相乘的形式,其对角线元素是 $(\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1)^{-1}$,两者之间差了 $\ln(\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1)$;②在算法 1 的迭代过程中, $\mathbf{D}$ 是已知的。可以看出,算法 1 的解近似满足目标函数式(9)的 KKT 条件,即 $\mathbf{W}$ 会朝着一个近似解移动。证毕。

2.4 旋转不变性分析

旋转不变性意味着投影结果,即高维数据的低维表示,在样本空间经过旋转后保持不变。

定理 3 算法 1 的解具有旋转不变性。

证明 任意给定一个旋转矩阵 $\mathbf{\Gamma}(\mathbf{\Gamma}^T \mathbf{\Gamma} = \mathbf{I})$,投影矩阵 $\mathbf{W}$ 和样本 $\mathbf{x}_i$ 的旋转变换表示为

$$\tilde{\mathbf{W}} = \mathbf{\Gamma} \mathbf{W}; \tilde{\mathbf{x}}_i = \mathbf{\Gamma} \mathbf{x}_i \quad (29)$$

因为 $\mathbf{\Gamma}^T \mathbf{\Gamma} = \mathbf{I}$,从而可得

$$\| \mathbf{x}_i \|_2^2 = \| \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{x}_i \|_2^2 = \| \mathbf{\Gamma} \mathbf{x}_i \|_2^2 \quad (30)$$

将式(29)和式(30)进行结合,算法 1 的目标函

数可以表示为

$$\begin{aligned} \sum_{i=1}^N \ln(\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2 + 1) &= \\ \sum_{i=1}^N \ln(\| (\mathbf{\Gamma}^T \mathbf{\Gamma})(\mathbf{x}_i - \mathbf{W}\mathbf{u}_i) \|_2^2 + 1) &= \\ \sum_{i=1}^N \ln(\| \mathbf{\Gamma}^T (\mathbf{\Gamma} \mathbf{x}_i - \mathbf{\Gamma} \mathbf{W} \mathbf{u}_i) \|_2^2 + 1) &= \\ \sum_{i=1}^N \ln(\| \mathbf{\Gamma} \mathbf{x}_i - \mathbf{\Gamma} \mathbf{W} \mathbf{u}_i \|_2^2 + 1) &= \\ \sum_{i=1}^N \ln(\| \mathbf{\Gamma} \mathbf{x}_i - \tilde{\mathbf{W}} \mathbf{u}_i \|_2^2 + 1) \end{aligned} \quad (31)$$

式中 $\mathbf{u}_i = \mathbf{W}^T \mathbf{x}_i$ 是样本 $\mathbf{x}_i$ 的低维表示。式(31)表明,如果 $\mathbf{W}$ 是算法 1 目标函数的解,那么 $\tilde{\mathbf{W}}$ 就是样本经过旋转变换 $\mathbf{\Gamma}$ 下目标函数的解。由此可知

$$\tilde{\mathbf{W}}^T \mathbf{\Gamma} \mathbf{x}_i = \mathbf{W}^T \mathbf{x}_i = \mathbf{u}_i \quad (32)$$

样本 $\mathbf{x}_i$ 的低维表示在旋转变换下并不改变。证毕。

图 1 是某样本误差与目标函数的关系。 $E_i = \| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \|_2^2$ 是某样本误差的平方, O 是目标函数值。基于 $l_{2,p}$ 范数的 PCA 含有参数 p ,图中 p 取典型值 1。可以看出:标准 PCA 的目标函数值在误差较小时大于基于 $l_{2,p}$ 范数的 PCA 的,即基于 $l_{2,p}$ 范数的 PCA 放大了较小的重构误差,对于求得最优投影矩阵有不利的影响;本文算法的目标函数值始终小于标准 PCA 的,对于较大的重构误差有更多的抑制,对于含有异常样本的数据鲁棒性更强。

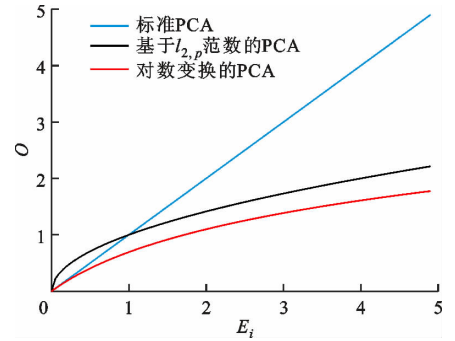


图 1 重构误差与目标函数值的关系

与现有的 PCA- L_1 算法相比,所提的基于对数变换的 PCA 算法最小化重构误差,与 PCA 算法的目标一致。与目标函数式(4)对应的算法相比,所提算法保留了 PCA 算法的旋转不变性。与基于 $l_{2,p}$ 范数的 PCA 算法相比,在理论上所提算法的目标函数值小于 PCA 算法的,避免了 $l_{2,p}$ 范数放大较小重构误差的缺点。所提算法具有良好的收敛性。

3 实验结果与分析

为了验证所提算法的有效性,在 AR、Extended

Yale B 和 CMU PIE 共 3 个人脸数据集及 MNIST 共 1 个手写字符数据集上进行了实验,和标准 PCA 算法^[1]、贪婪求解的基于 l_1 范数的 PCA 算法^[10](记为贪婪 PCA- L_1 算法)、非贪婪求解的基于 l_1 范数的 PCA 算法^[11](记为非贪婪 PCA- L_1 算法)、基于最大相关熵的 PCA 算法^[15](记为 HQPCA 算法)以及基于 $l_{2,p}$ 范数的 PCA 算法^[16](记为 PCA- $l_{2,p}$)进行了比较。设置 PCA- $l_{2,p}$ 算法的 $p=1$ 。使用最近邻算法对降维后的数据进行分类,近邻数取为 1。投影向量数 k 从 10 取到 200。重构误差为

$$E = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \| \mathbf{x}_i^{\text{clean}} - \mathbf{W}\mathbf{W}^T \mathbf{x}_i^{\text{clean}} \|_2 \quad (33)$$

式中: N_{train} 表示训练集中的正常样本数; $\mathbf{x}_i^{\text{clean}}$ 是训练集中的第 i 个正常样本。

3.1 在 AR 数据集上的实验

AR 数据集^[17]包含 126 人的超过 4 000 幅彩色正面人脸图像,实验中转为灰度图像。实验选择其中 119 人(65 名男性和 54 名女性)的 3 094 幅图像作为数据集。这些图像在 2 个时段拍摄(相隔 2 周)。每个时段包含 13 幅图像,6 幅图像有眼镜或围巾遮挡,7 幅图像有不同表情和光照。从每幅图像中选出人脸区域,并将图像大小调整为 50×40 像素^[18]。图 2 是 AR 数据集中的某人图像,图像 a~m 来自第 1 个时段,图像 n~z 来自第 2 个时段。

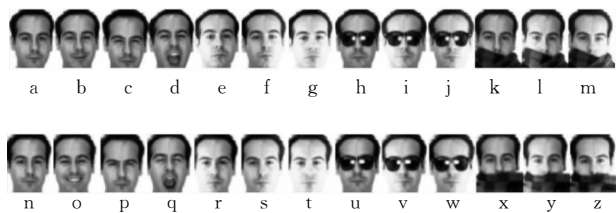
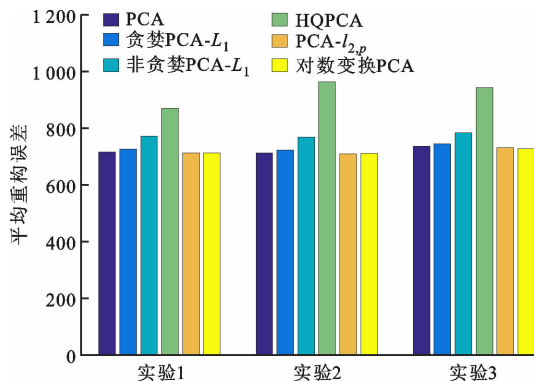


图 2 AR 数据集中的某人图像

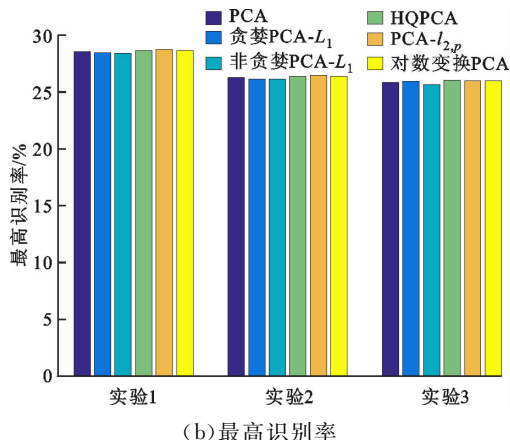
在 AR 数据集上,进行 3 组实验,实验 1 是眼镜遮挡实验,实验 2 是围巾遮挡实验,实验 3 是眼镜和围巾遮挡实验。在实验 1 中,选择每人 10 幅图像,即图像 a~j 作为训练样本,图像 n~w 作为测试样本,其中被眼镜遮挡的图像是异常样本。在实验 2 中,选择每人 10 幅图像,即图像 a~g 和 k~m 作为训练样本,图像 n~t 和 x~z 作为测试样本,其中被围巾遮挡的图像是异常样本。在实验 3 中,选择每人 13 幅图像,包括第 1 个时段的 13 幅图像 a~m 作为训练样本,其他图像作为测试样本,被眼镜或者围巾遮挡的图像是异常样本。

图 3 是 AR 数据集上的平均重构误差和最高识别率。从图 3a 可以看出:在实验 1 和 2 中,所提算

法的重构误差与 PCA- $l_{2,p}$ 算法的接近,2 种算法的重构误差小于其他 4 种算法的;在实验 3 中,所提算法的重构误差最小,优于其他算法的;在所有实验中,HQPCA 算法的重构误差明显高于其他算法的。从图 3b 可以看出,所提算法和 HQPCA 算法以及 PCA- $l_{2,p}$ 算法的识别率接近,高于其他 3 种算法的。



(a) 平均重构误差

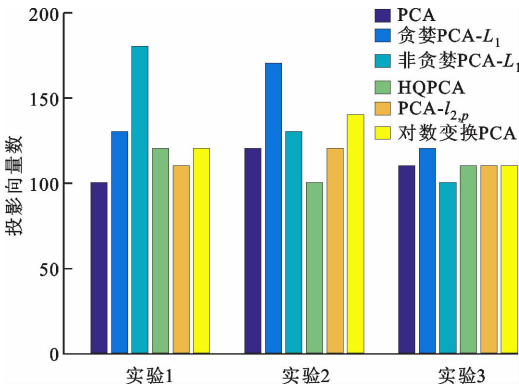


(b) 最高识别率

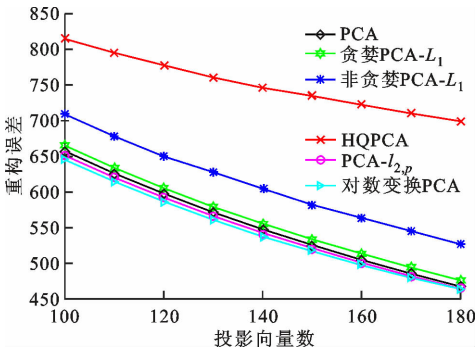
图 3 AR 数据集上的平均重构误差和最高识别率

图 4 是 AR 数据集上的性能比较。从图 4a 可以看出,当投影向量数在 $100 \sim 180$ 时,各算法取得最高的识别率,由此表明投影向量数在 $100 \sim 180$ 变化时,对各算法有重要影响。从图 4b 中可以看出,当投影向量数在 $100 \sim 180$ 时:所提算法取得了最小重构误差;PCA- $l_{2,p}$ 算法取得了较小的重构误差;HQPCA 算法的重构误差较大。对于 AR 数据集上的 3 类实验,在第 1、2 类实验中分别加入了被眼镜遮挡的样本和被围巾遮挡的样本作为异常样本,异常样本数量占的比例较小。实验 1 和实验 2 不能充分体现鲁棒 PCA 算法处理含有异常样本数据集的优越性。实验 3 中加入了被眼镜和围巾遮挡的样本作为异常样本,异常样本占比明显高于实验 1 和实验 2 的。实验 3 上的结果更能体现鲁棒 PCA 算法的优越性。综合可知,所提算法在实验 3 上取得了

最小的重构误差,且取得了较高的识别率。



(a)取得最高识别率时对应的投影向量数



(b)实验 3 的重构误差

图 4 AR 数据集上的性能比较

3.2 在 Extended Yale B 数据集上的实验

Extended Yale B 数据集^[19]包含了来自 38 人的 2 144 幅在不同光照下的图像。在该实验中,将人脸图像的大小调整为 32×32 像素^[18]。从每个人的图像中随机选取 14 幅图像,然后随机选取图像的一部分,将这部分替换为含有椒盐噪声的图像块。该图像块只包含像素值为 0 和 255 的点,大小取为 12×12 像素。图 5 是 Extended Yale B 数据集集中的部分正常图像和含噪图像。对于每一个人,随机选取 25 幅正常图像和 7 幅含噪图像作为训练样本,该人的其他图像作为测试样本。实验重复 10 次。



(a)正常图像



(b)含噪图像

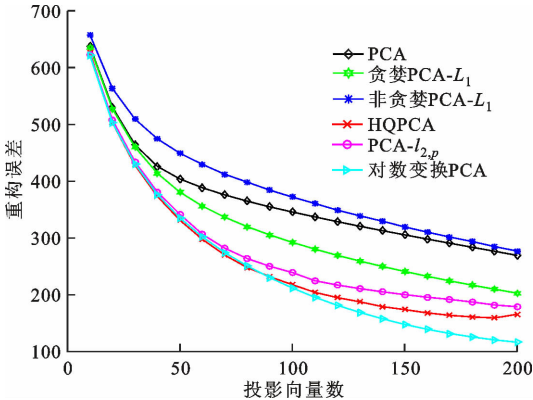
图 5 Extended Yale B 数据集集中的部分图像

表 1 是 Extended Yale B 数据集上各算法的最低平均重构误差和对应的识别率,表中以“量值±标准差”的形式表示。

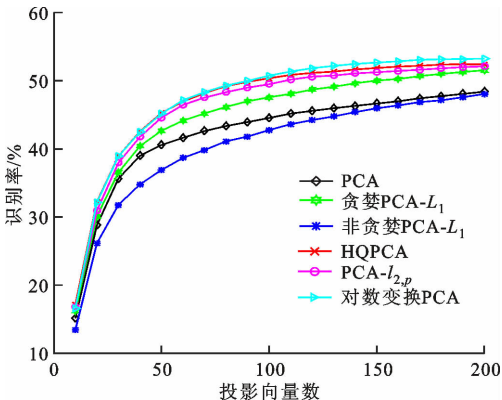
表 1 Extended Yale B 数据集上各算法的最低平均误差和对应的识别率

算法	最低平均误差	识别率/%
PCA	269.47±1.59	48.40±1.33
贪婪 PCA- L_1	203.00±1.29	51.55±1.43
非贪婪 PCA- L_1	277.13±1.77	48.07±1.24
HQPCA	159.38±4.75	52.45±1.47
PCA- $l_{2,p}$	179.12±0.92	52.12±1.52
所提算法	116.55±2.20	53.25±1.49

图 6 是 Extended Yale B 数据集上各算法的性能比较。从图 6 和表 1 中可以看出:所提算法在数据重构和分类上的性能明显优于其他 5 种算法的,性能仅次于所提算法的是 HQPCA 算法,再次是 PCA- $l_{2,p}$ 算法;HQPCA 在投影向量数从 190 变为 200 时,重构误差下降,而其他 5 种算法依然保持上升;PCA 算法性能不佳,这是因为数据集中含有异常样本,而异常样本对 PCA 算法的重构误差有较大



(a)重构误差和投影向量数的关系



(b)识别率和投影向量数的关系

图 6 Extended Yale B 数据集上各算法的性能比较

影响;贪婪和非贪婪 PCA- L_1 算法由于没有直接最小化重构误差,而是最大化投影后向量的 l_1 范数,所以性能也不佳;所提算法优于 PCA- $l_{2,p}$ 算法,这是因为所提算法没有放大小的重构误差,对于大的重构误差也有更大的抑制,所以所提算法取得了最优的性能。在 Extended Yale B 数据集上实验结果与理论分析一致,且与 AR 数据集上的实验结果一致。

3.3 在 CMU PIE 数据集上的实验

CMU PIE 数据集^[20]包含了来自 68 人的 2 856 幅正面人脸图像。每人含有 42 幅在不同光照条件下拍摄的图像。在该实验中,将每幅图像的大小调整为 32×32 像素^[18],并且加入了类似于在 Extended Yale B 数据集的实验中加入的噪声。从每人的图像中随机选择 10 幅图像加入噪声。图 7 是部分 CMU PIE 数据集上的人脸图像和含噪图像。随机选取每个人的 21 幅图像(16 幅正常图像和 5 幅含噪图像)作为训练样本,其他图像作为测试样本。实验重复 10 次。

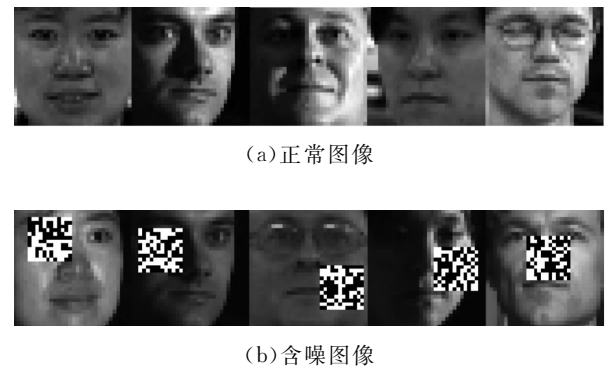


图 7 CMU PIE 数据集的部分图像

表 2 是 CMU PIE 数据集上各算法的平均最小重构误差和对应的识别率。

表 2 CMU PIE 数据集上各算法的平均最小重构误差和对应的识别率

算法	平均最小重构误差	识别率/%
PCA	290.84 ± 1.18	50.97 ± 1.95
贪婪 PCA- L_1	228.96 ± 1.47	52.93 ± 1.81
非贪婪 PCA- L_1	287.95 ± 2.18	50.91 ± 1.95
HQPCA	166.23 ± 1.49	54.25 ± 1.71
PCA- $l_{2,p}$	212.57 ± 1.65	53.23 ± 1.68
所提算法	145.85 ± 1.54	54.89 ± 1.56

图 8 是 CMU PIE 数据集上各算法的性能比较。从表 2 和图 8 可以看出:所提算法取得了最优的性能,性能仅次于所提算法的是 HQPCA 算法;

对于 HQPCA 算法,当投影向量数较小时,其性能与所提算法接近,随着投影向量数的增加,其性能达到最优后开始变差,而所提算法的性能随着投影向量数的增加一直增加,并且优于 HQPCA 算法的最优性能。在 CMU PIE 数据集上的实验结果与理论分析一致,并且与在 AR 数据集和 Extended Yale B 数据集上的实验结果一致。

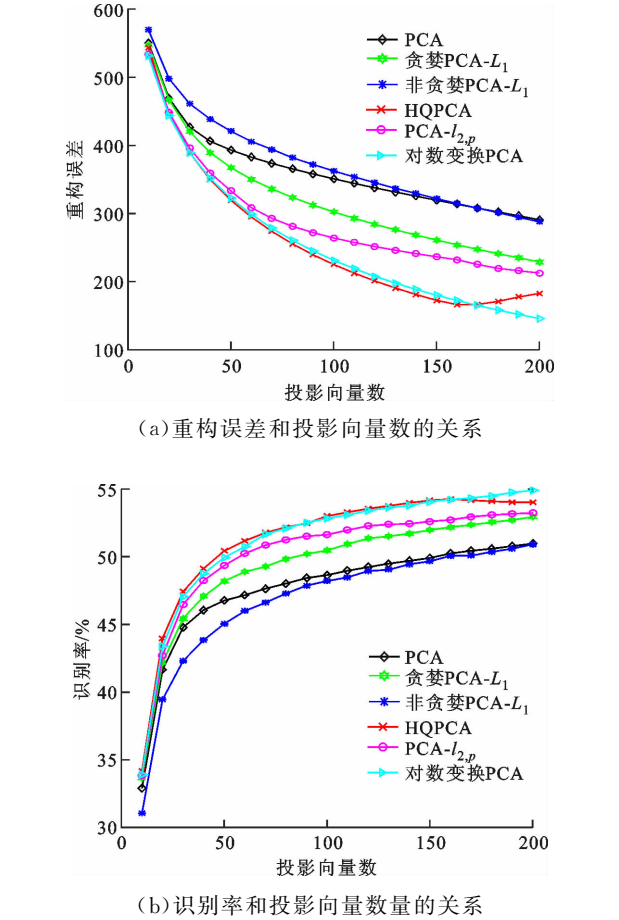


图 8 CMU PIE 数据集上各算法的性能比较

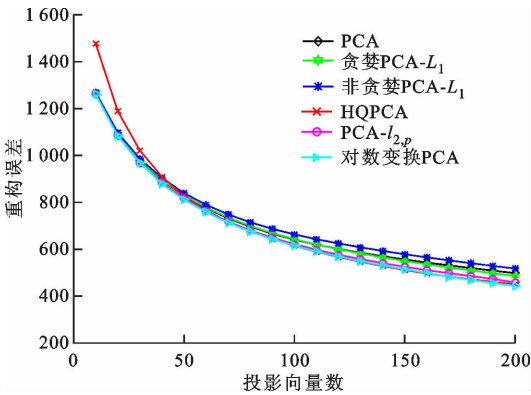
3.4 在 MNIST 数据集上的实验

MNIST 数据集^[21]共包含了 70 000 幅手写数字图像,其中训练集有 60 000 幅图像,测试集有 10 000 幅图像。每幅图像的大小是 28×28 像素,手写数字位于图像中间。在本实验中,从测试集中选取 5 000 幅图像,每个字符(即 0~9)包含 500 幅图像。从每个字符的图像中随机选取 120 幅图像,然后随机选取图像的一部分,将这部分替换为含有椒盐噪声的图像块,图像块的大小取为 10×10 像素。对于每一个字符,随机选取 190 幅正常图像和 60 幅含噪图像作为训练样本,该字符的其他图像为测试样本。实验重复 10 次。表 3 是 MNIST 数据集上各算法的平均最小重构误差和对应的识别率。

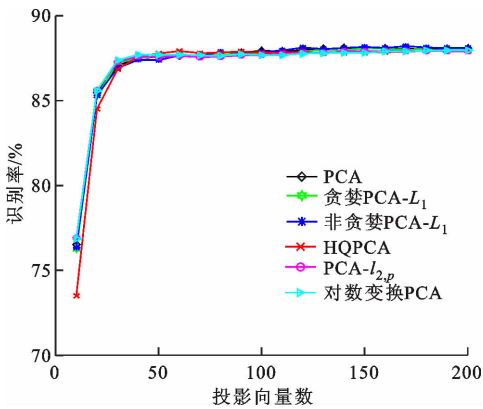
表 3 MNIST 数据集上各算法的平均最小重构误差和对应的识别率

算法	平均最小重构误差	识别率/%
PCA	361.93±1.68	88.08±0.54
贪婪 PCA- L_1	347.41±1.87	88.05±0.61
非贪婪 PCA- L_1	385.94±1.97	88.10±0.69
HQPCA	304.45±2.53	87.92±0.63
PCA- $l_{2,p}$	313.34±2.07	87.94±0.59
所提算法	296.32±2.12	87.97±0.64

图 9 是 MNIST 数据集上是各算法的性能比较。从表 4 和图 9a 可以看出:所提算法取得了最小的重构误差,HQPCA 算法和 PCA- $l_{2,p}$ 算法也取得了较好的性能,它们的性能都优于其他 3 类算法的;当投影向量数较小时,HQPCA 算法的重构误差略小于所提算法的,而当投影向量数量增大时,所提算法的性能逐渐优于 HQPCA 算法的。从图 9b 可以看出:各算法的识别率非常接近,最高的识别率(PCA)与最低的识别率(HQPCA)之间的差距为 0.16%,基本可以忽略;当投影向量数较少时,所提



(a) 重构误差和投影向量数的关系



(b) 识别率和投影向量数的关系

图 9 MNIST 数据集上各算法的性能比较

算法取得了最高的识别率。综上可知,所提算法的特征提取性能优于其他对比算法的。

3.5 收敛性实验

AR 数据集上实验 3 的目标函数值与迭代次数的关系曲线如图 10 所示,投影向量数 $k=10$ 。可以看出,所提算法的目标函数值迅速减小至收敛,大概在第 5 次迭代时即可达到收敛。图 10 的结果与 2.3 小节中对于算法收敛性的理论分析一致,所提算法可以得到目标函数的最优解。

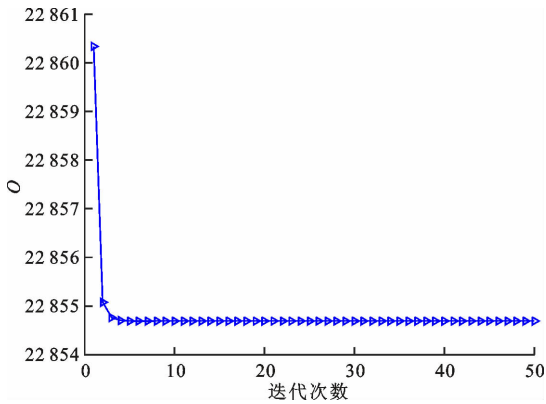


图 10 所提算法目标函数值与迭代次数的关系曲线

4 结 论

本文提出了一种鲁棒 PCA 算法——基于对数变换的 PCA 算法。从对数变换的性质出发,导出了所提算法使用的目标函数。所提算法的目标函数值和标准 PCA 算法的呈对数变换关系,从理论上保证了所提算法的目标函数值小于标准 PCA 算法的。因为对标准 PCA 算法的目标函数采用了对数变换,大幅降低了异常样本对于目标函数值的影响,从而保证了所提算法的鲁棒性。理论分析表明,所提算法的每次迭代可以近似降低目标函数值,并且可以近似收敛于目标函数的最优解。

人脸数据集和手写字符数据集上的实验充分表明了所提算法的优越性。与现有的鲁棒 PCA 算法相比,所提算法对于异常样本的鲁棒性和处理能力有较大提升,取得了最低的重构误差并且取得了最高的识别率。与 PCA- $l_{2,p}$ 算法相比,所提算法的目标函数能够降低较小的重构误差,同时对于较大的重构误差有更强的抑制。与 PCA- L_1 算法相比,所提算法保留了 PCA 算法具有的旋转不变性。实验分析表明,所提算法的收敛性能较好,一般迭代数次即可收敛,与理论分析一致。

未来可将对数变换应用于稀疏 PCA 算法,从而

得到能更好地处理含有异常样本的稀疏 PCA 算法。

参考文献:

- [1] TURK M A, PENTLAND A P. Face recognition using eigenfaces [C] // Proceedings of the 1991 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 1991: 586-591.
- [2] LOCANTORE N, MARRON J S, SIMPSON D G, et al. Robust principal component analysis for functional data [J]. *Test*, 1999, 8(1): 1-73.
- [3] DING C, ZHOU Ding, HE Xiaofeng, et al. R_1 -PCA: Rotational invariant L_1 -norm principal component analysis for robust subspace factorization [C] // Proceedings of the 23rd International Conference on Machine learning. New York USA: ACM, 2006: 281-288.
- [4] NIE F P, YUAN J J, HUANG H. Optimal mean robust principal component analysis [C] // Proceedings of the 31st International Conference on Machine learning. New York, USA: ACM, 2014: 2755-2763.
- [5] MARONNA R. Principal components and orthogonal regression based on robust scales [J]. *Technometrics*, 2005, 47(3): 264-273.
- [6] BACCINI A, BESSE P, FALGUEROLLES A. A L_1 -norm PCA and a heuristic approach [M] // Studies in Classification, Data Analysis, and Knowledge Organization. Cham, Germany: Springer, 1996: 359-368.
- [7] KE Q, KANADE T. Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming [C] // Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2005: 739-746.
- [8] DAHYOT R, CHARBONNIER P, HEITZ F. Robust visual recognition of colour images [C] // Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2000: 685-690.
- [9] NG A Y. Feature selection, L_1 vs. L_2 regularization, and rotational invariance [C] // Proceedings of the 21st International Conference on Machine learning. New York, USA: ACM, 2004: 78-85.
- [10] KWAK N. Principal component analysis based on L_1 -norm maximization [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(9): 1672-1680.
- [11] NIE F, HUANG H, 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. California, USA: IJCAI, 2011: 1433-1438.
- [12] KWAK N. Principal component analysis by L_p -norm maximization [J]. *IEEE Transactions on Cybernetics*, 2014, 44(5): 594-609.
- [13] IGLESIAS J E, DE BRUIJNE M, LOOG M, et al. A family of principal component analyses for dealing with outliers [J]. *Medical Image Computing and Computer-Assisted Intervention*, 2007, 10(2): 178-185.
- [14] SUBBARAO R, MEER P. Subspace estimation using projection based M -estimators over Grassmann manifolds [C] // Proceedings of the 9th European Conference on Computer Vision. Cham, Germany: Springer, 2006: 301-312.
- [15] HE Ran, HU Baogang, ZHENG Weishi, et al. Robust principal component analysis based on maximum correntropy criterion [J]. *IEEE Transactions on Image Processing*, 2011, 20(6): 1485-1494.
- [16] WANG Qianqian, GAO Quanxue, GAO Xinbo, et al. $l_{2,p}$ -norm based PCA for image recognition [J]. *IEEE Transactions on Image Processing*, 2018, 27(3): 1336-1346.
- [17] MARTINEZ A M, BENAVENTE R. The AR face database [EB/OL]. [2020-06-01]. <http://cat.cvc.uab.es/Public/Publications/1998/MaB1998/CVCReport24.pdf>.
- [18] HE X, NIYOGI P. Locality preserving projections [C] // Proceedings of the 2003 Conference on Advances in Neural Information Processing Systems. Vancouver, Canada: NIPS, 2003: 153-160.
- [19] GEORGHIADES A S, BELHUMEUR P N, KRIEGSMAN D J. From few to many: illumination cone models for face recognition under variable lighting and pose [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 643-660.
- [20] SIM T, BAKER S, BSAT M. The CMU pose, illumination, and expression (PIE) database [C] // Proceedings of 5th IEEE International Conference on Automatic Face Gesture Recognition. Piscataway, NJ, USA: IEEE, 2002: 53-58.
- [21] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.

(编辑 陶晴)