

基于 L1-范数和弹性网约束的鲁棒稀疏块 PCA

唐肝翌¹, 卢桂馥¹, 王 勇^{1,2}, 范莉莉¹, 杜扬帆¹

(1.安徽工程大学计算机与信息学院,安徽 芜湖 241000)

(2.计算机软件国家重点实验室(南京大学),江苏 南京 210023)

[摘要] 块主成份分析(block principal component analysis, BPCA)是一种重要的子空间学习方法,能充分利用图像矩阵的部分关联. 基于 L1-范数的 BPCA 是近年来发展起来的鲁棒降维的有效方法. 本研究提出了一种新的鲁棒稀疏 BPCA 方法,称之为 BPCAL1-S. 该方法相对于传统的基于 L2-范数的 PCA 对噪声更加鲁棒. 为了建立稀疏模型,优化过程中引入弹性网,联合使用 Lasso 与 Ridge 惩罚因子进行约束. 提出了一种贪心算法逐个提取特征向量,对迭代过程的收敛性做了理论证明. 将 BPCAL1-S 应用于图像分类与图像重构,实验结果验证了该方法的有效性.

[关键词] 块主成份分析, L1-范数, 弹性网, 稀疏建模, 子空间学习

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2022)04-0102-08

Robust and Sparse BPCA with the Constraints of L1-norm and the Elastic Net

Tang Ganyì¹, Lu Guifu¹, Wang Yong^{1,2}, Fan Lili¹, Du Yangfan¹

(1.School of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China)

(2.State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China)

Abstract: Block principal component analysis (BPCA), which can utilize part correlation of image matrix sufficiently, is an important subspace learning approach. L1-norm based BPCA is an effective technique for robust learning in dimensionality reduction developed recently. We propose a novel robust and sparse BPCA method referred to as BPCAL1-S. The approach is more robust to outliers than the traditional L2-norm based PCA. To develop a model with sparsity, the elastic net constraint which combining ridge and lasso penalty, is integrated into the optimization procedure. We present a greedy algorithm to extract basic feature vectors one by one, and proposed theoretical analysis to guarantee the convergence of the iterative process. The proposed BPCAL1-S is applied to the analysis of image classification and image reconstruction, and the experimental results verify its effectiveness.

Key words: BPCA, L1-norm, elastic net, sparse modelling, subspace learning

主成份分析(PCA)^[1]是降维和特征提取的经典方法,广泛应用于机器学习与计算机视觉领域. 传统的 PCA 方法需要把二维矩阵转化为一维向量,这会破坏原始图像的内在结构. PCA 方法需要构造图像库协方差矩阵并求取特征向量和特征值,当样本数量不断增加时,容易出现协方差矩阵过大的维度灾难问题. 文献[2]提出了二维主成份分析方法,可以直接利用图像矩阵,保留了原始图像的空间结构,并避免协方差矩阵过大的问题.

传统的 PCA/2DPCA 并未考虑鲁棒性的问题,在目标函数中采用 L2-范数,对噪声比较敏感. 近年来,作为鲁棒的替代方案,基于 L1-范数的方法得到了快速发展. L1-PCA^[3]和 R1-PCA^[4]通过线性的或二次规划的方法来求得优化的特征基,缺点是计算复杂度高. 文献[5]提出一种简单直观的基于 L1-范数的 PCA

收稿日期:2022-04-16.

基金项目:国家自然科学基金项目(61976005)、安徽省自然科学基金项目(1908085MF183)、安徽高校自然科学基金项目重点项目(KJ2020A0363)、安徽工程大学“中青年拔尖人才培养计划”(201812)、计算机软件新技术国家重点实验室(南京大学)开放基金项目(KFKT2019B23)、安徽省高等教育提升计划项目(TSKJ2016B01)、安徽省高等学校省级质量工程项目(2019jyxm1183).

通讯作者:唐肝翌,硕士,副教授,研究方向:机器学习与图像处理. E-mail:tom365@ahpu.edu.cn

(PCA-L1). 文献[6]提出一种基于 L1-范数的鲁棒 2DPCA(2DPCA-L1). 文献[7]提出了基于 L1-范数的非贪心 PCA(PCAL1-nongreedy), 文献[8]提出了基于 L1-范数的非贪心 2DPCA(2DPCAL1-nongreedy), 非贪心算法可以同时提取各个方向上的特征. 因为 L2-范数和 L1-范数都是 L_p -范数($0 < p \leq 2$)的特例, 文献[9]很自然地将 PCA 方法推广到 L_p -范数(PCA- L_p).

如果把矩阵行向量看成一个计算单元, 2DPCA 与 PCA 本质上是等价的^[10]. 更一般地, 块 PCA(BPCA)^[11-12]将每一个图像矩阵分成若干个具有相同数量像素的块, 把每一个块向量化并作为一个计算单元应用于 PCA 框架. 基于这一视角, 2DPCA 和 PCA 都是 BPCA 的特例. 众所周知, L1-范数比 L2-范数更加鲁棒, 基于 L1-范数的鲁棒 BPCA(BPCA-L1)^[13]及其非贪心版本(BPCA-L1 non-greedy)^[14]也相继被提出.

以上方法获取的特征向量还存在大量非零元, 这样可能存在冗余信息并很难解释. 稀疏性通常有助于算法的高效执行^[15], 稀疏表示也往往对应着数据的局部线性结构^[16], 从大量特征中寻找最相关或者最突出的元素是非常有意义的. 文献[17]提出了一种基于 L1-范数的鲁棒稀疏的 2DPCA 模型(2DPCAL1-S). 随后, 文献[18]提出一种基于 L_p -范数的通用 2DPCA 稀疏模型(G2DPCA). 文献[19]和文献[20]分别提出了基于 $L_{2,p}$ -范数和 R1-范数的 $L_{2,p}$ -PCA 和 R1-2-DPCA 方法.

受以上文献启发, 考虑到 BPCA 能充分利用近邻像素间的逻辑关联, 本文提出一种基于 L1-范数和弹性网约束的鲁棒稀疏的块 PCA 方法, 将其命名为 BPCAL1-S(BPCA-L1 with sparsity). BPCAL1-S 有以下优点:

- (1) 计算单元为更加一般的块, 块中包含行和列的像素, 该方法能充分利用图像的部分关联;
- (2) 用 L1-范数代替 L2-范数, 所提取的特征更加鲁棒;
- (3) 所提取的特征是稀疏的, 有更好的语义解释性, 也有助于提升性能.

1 相关方法

1.1 PCA-L1

将图像集表示为 $X = \{x_1, \dots, x_n\} \in R^{d \times n}$, 其中 n 为样本数, d 为图像矩阵向量化后的维数. 假设样本 $x_i (i=1, 2, \dots, n)$ 已做去均值化处理. PCA-L1 寻求最优的投影向量 u^* , 使得基于 L1-范数的目标函数最大化, 即:

$$u^* = \arg \max_u \|X^T u\|_1 = \arg \max_u \sum_{i=1}^n |u^T x_i|. \quad (1)$$

式中, $\|u\|_2 = 1$, $\|\cdot\|_1$ 表示 L1-范数.

PCA-L1 采用迭代过程来解决基于 L1-范数的非线性优化问题, 由此求出投影向量 u^* . 投影向量迭代公式如(2)所示:

$$u(t+1) = \frac{\sum_{i=1}^n p_i(t) x_i}{\left\| \sum_{i=1}^n p_i(t) x_i \right\|_2}. \quad (2)$$

式中, 极性函数 $p_i(t)$ 定义如下:

$$p_i(t) = \begin{cases} 1, & u^T(t) x_i \geq 0 \\ -1, & u^T(t) x_i < 0 \end{cases} \quad (3)$$

式中, $u(t)$ 为第 t 轮迭代的投影向量.

1.2 2DPCA-L1

令 $X = \{X_1, \dots, X_n\}$ 表示 n 个图像矩阵. $X_i (i=1, 2, \dots, n)$ 的行列数为 $h \times v$. 不失一般性, X_i 已做去均值化处理. 2DPCA-L1 寻求最优的投影向量 u^* , 使得基于 L1-范数的目标函数值取最大, 即:

$$u^* = \arg \max_u \sum_{i=1}^n \|X_i u\|_1 = \arg \max_u \sum_{i=1}^n \sum_{j=1}^h |u^T x_{ij}| \quad (4)$$

式中, $\|u\|_2 = 1$, x_{ij} 为列向量, 由第 i 个图像矩阵的第 j 行转置而成.

2DPCA-L1 也采用类似于 PCA-L1 的迭代过程,迭代公式如(5)所示:

$$\mathbf{u}(t+1) = \frac{\sum_{i=1}^n \sum_{j=1}^h p_{ij}(t) \mathbf{x}_{ij}}{\left\| \sum_{i=1}^n \sum_{j=1}^h p_{ij}(t) \mathbf{x}_{ij} \right\|_2}. \quad (5)$$

式中, $p_{ij}(t)$ 为极性函数.

1.3 BPCA-L1

式(2)和式(5)表明,2DPCA-L1 相当于把 PCA-L1 的过程应用到图像的行. 如果在 2DPCA-L1 的迭代公式中用更加一般的像素块(包含行和列)代替图像的行,则把 2DPCA-L1 推广到了 BPCA-L1.

令 $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ 表示 n 个去均值的图像矩阵. BPCA-L1 把每一个矩阵分成 m 个具有相同行列数的小块,即 $\mathbf{X}_i = \{\mathbf{X}_1^{(i)}, \dots, \mathbf{X}_m^{(i)}\}$, $\mathbf{X}_j^{(i)}$ 为第 i 个图像中的第 j 个小块. 容易理解,每个小块中的像素之间因距离较近而有很强信息关联. \mathbf{X}_i 划分为 2×3 的小块并完成向量化,过程如图 1 所示, $\mathbf{x}_j^{(i)}$ 为块 $\mathbf{X}_j^{(i)}$ 向量化再转置而成的列向量.

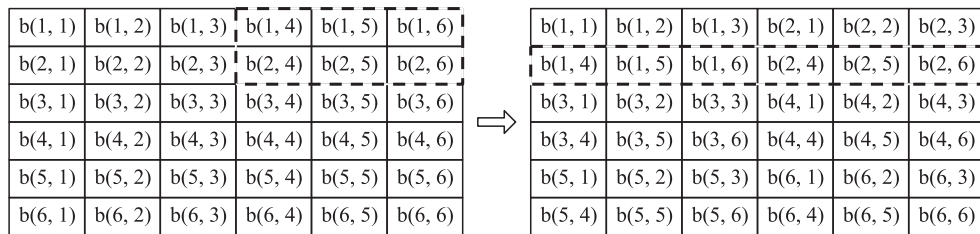


图 1 图像矩阵划分为 3×3 的小块并完成向量化

Fig. 1 Transforming original image matrix to vectored representation of 3×3 block

BPCA-L1 寻求最优的投影向量 \mathbf{u}^* , 以最大化基于 L1-范数的目标函数, 即:

$$\mathbf{u}^* = \arg \max_{\mathbf{u}} \sum_{i=1}^n \sum_{j=1}^m |\mathbf{u}^T \mathbf{x}_j^{(i)}|. \quad (6)$$

为解决 L1-范数优化问题, 特征向量 $\mathbf{u}(t+1)$ 的迭代公式如(7):

$$\mathbf{u}(t+1) = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \mathbf{x}_j^{(i)}}{\left\| \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \mathbf{x}_j^{(i)} \right\|_2}. \quad (7)$$

式中, $p_{ij}(t)$ 为极性函数.

2 BPCAL1-S

2.1 基本思路

图像通常拥有很多特征, 其中有些是相关或者冗余的. 抽取稀疏特征有重要的意义, 稀疏特征中的非零元可能对应了诸如人脸图像中的嘴巴或眼睛等的显著区域. 稀疏特征更加典型, 解释性更好, 对于分类也更为有效. 弹性网(Elastic Net)联合使用 Lasso 和 Ridge 惩罚因子, 可以规避单纯使用 Lasso 惩罚因子的潜在限制^[17,21]. 目标函数中引入弹性网约束可用于稀疏特征提取, 如式(8):

$$H(\mathbf{u}) = \sum_{i=1}^n \sum_{j=1}^m |\mathbf{u}^T \mathbf{x}_j^{(i)}| - \frac{\eta}{2} \|\mathbf{u}\|_2 - \lambda \|\mathbf{u}\|_1. \quad (8)$$

提出的 BPCAL1-S 通过最大化 $H(\mathbf{u})$ 求得稀疏特征 \mathbf{u}^* , 如式(9):

$$\mathbf{u}^* = \arg \max_{\mathbf{u}} H(\mathbf{u}) = \arg \max_{\mathbf{u}} \sum_{i=1}^n \sum_{j=1}^m |\mathbf{u}^T \mathbf{x}_j^{(i)}| - \frac{\eta}{2} \|\mathbf{u}\|_2 - \lambda \|\mathbf{u}\|_1. \quad (9)$$

2.2 特征提取

众所周知, 同时提取多个稀疏特征是非常困难的. 本研究采用贪心策略, 先获取一个特征, 然后逐个提取其余特征. 第 t 轮迭代目标函数如式(10):

$$H(u(t)) = \sum_{i=1}^n \sum_{j=1}^m |u^T(t) \mathbf{x}_j^{(i)}| - \frac{\eta}{2} \|u(t)\|_2 - \lambda \|u(t)\|_1 = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) u^T(t) \mathbf{x}_j^{(i)} - \frac{\eta}{2} \|u(t)\|_2 - \lambda \|u(t)\|_1. \quad (10)$$

式中, $u^T u = 1$, p_{ij} 为极性函数.

试图寻找 $\mathbf{u}(t+1)$ 单调递增的迭代公式, 即 $\mathbf{u}(t+1) \geq \mathbf{u}(t)$. 考虑到 $u(t)$ 的稀疏性, $u(t)$ 中的一些元素为零. 将 $u(t)$ 中的零元素移除并表示为 $\bar{\mathbf{u}}(t)$, 把 $\mathbf{x}_j^{(i)}$ 中与 $u(t)$ 位置对应的零元素移除并表示为 $\bar{\mathbf{x}}_j^{(i)}$. 例如: $u(t) = (7, 0, 4, 0, 2)^T$, $\mathbf{x}_j^{(i)} = (21, 22, 23, 24, 25)^T$, 那么 $\bar{\mathbf{u}}(t) = (7, 4, 2)^T$, $\bar{\mathbf{x}}_j^{(i)} = (21, 23, 25)^T$. 因此式 (10) 转换为:

$$H(u(t)) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{\mathbf{u}}^T(t) \bar{\mathbf{x}}_j^{(i)} - \frac{\eta}{2} \|\bar{\mathbf{u}}(t)\|_2 - \lambda \|\bar{\mathbf{u}}(t)\|_1. \quad (11)$$

式 (11) 可以写成:

$$H(u(t)) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{\mathbf{u}}^T(t) \bar{\mathbf{x}}_j^{(i)} - \frac{\eta}{2} \|\bar{\mathbf{u}}(t)\|_2 - \frac{\lambda}{2} (\bar{\mathbf{u}}^T(t) \bar{\mathbf{R}}(t) \bar{\mathbf{u}}(t) + \|\bar{\mathbf{u}}(t)\|_1). \quad (12)$$

式中, 对角矩阵 $\bar{\mathbf{R}}(t) = \text{diag}(|\bar{u}_1(t)|^{-1}, \dots, |\bar{u}_h(t)|^{-1})$, 其中 $\bar{u}_1(t), \bar{u}_2(t), \dots, \bar{u}_h(t)$ 为向量 $\bar{\mathbf{u}}(t)$ 的分量.

优化分析中引入替代函数是一种传统的方法, 将替代函数 $Q(u|\bar{\mathbf{u}}(t))$ 定义如下:

$$Q(u|\bar{\mathbf{u}}(t)) = u^T \left(\sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{\mathbf{x}}_j^{(i)} \right) - \frac{\eta}{2} \|u\|_2 - \frac{\lambda}{2} (u^T \bar{\mathbf{R}}(t) u + \|\bar{\mathbf{u}}(t)\|_1). \quad (13)$$

$Q(u|\bar{\mathbf{u}}(t))$ 为 u 的函数, 而 $\bar{\mathbf{u}}(t)$ 为固定参数. 对 u 求偏导数并置为零以求得函数 Q 的最大值:

$$\frac{\partial Q(u|\bar{\mathbf{u}}(t))}{\partial u} = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{\mathbf{x}}_j^{(i)} - \eta u - \lambda \bar{\mathbf{R}}(t) u = 0. \quad (14)$$

由此得到,

$$\bar{\mathbf{u}}^* = \arg \max_u Q(u|\bar{\mathbf{u}}(t)) = (\eta \mathbf{I}_h + \lambda \bar{\mathbf{R}}(t))^{-1} \left(\sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{\mathbf{x}}_j^{(i)} \right), \quad (15)$$

式中, \mathbf{I}_h 为 h 维单位矩阵.

令 $\bar{\mathbf{u}}(t+1) = \bar{\mathbf{u}}^*$, 因此得到:

$$Q(\bar{\mathbf{u}}(t+1)|\bar{\mathbf{u}}(t)) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{\mathbf{u}}^T(t+1) \bar{\mathbf{x}}_j^{(i)} - \frac{\eta}{2} \|\bar{\mathbf{u}}(t+1)\|_2 - \frac{\lambda}{2} (\bar{\mathbf{u}}^T(t+1) \bar{\mathbf{R}}(t) \bar{\mathbf{u}}(t+1) + \|\bar{\mathbf{u}}(t)\|_1) \geq Q(\bar{\mathbf{u}}(t)|\bar{\mathbf{u}}(t)) = H(u(t)), \quad (16)$$

将零元素放回 $\bar{\mathbf{u}}(t+1)$, 得到 $\mathbf{u}(t+1)$ 为:

$$\mathbf{u}(t+1) = (\eta \mathbf{I} + \lambda \mathbf{R}(t))^{-1} \left(\sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \mathbf{x}_j^{(i)} \right), \quad (17)$$

显然, $(\eta \mathbf{I} + \lambda \mathbf{R}(t))^{-1}$ 为对角矩阵, 式 (17) 转换为

$$\mathbf{u}(t+1) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \mathbf{x}_j^{(i)} \cdot \left(\frac{|u_1(t)|}{\lambda + \eta |u_1(t)|}, \frac{|u_2(t)|}{\lambda + \eta |u_2(t)|}, \dots, \frac{|u_h(t)|}{\lambda + \eta |u_h(t)|} \right)^T. \quad (18)$$

式中, 运算符 \cdot 表示向量点乘.

算法 $r(r>0)$ 个特征抽取的 BPCAL1-S 算法.

输入: 样本矩阵的块向量 $\mathbf{x}_j^{(i)}$: $\mathbf{x}_j^{(i)}$ 表示为第 i 个图像中的第 j 个块转化而来的列向量 ($i=1, 2, 3, \dots, n; j=1, 2, 3, \dots, m$).

输出: 特征向量 \mathbf{u}_τ ($\tau=1, 2, 3, \dots, r$).

1. 令 $\tau=1$.

2. 计算特征向量 \mathbf{u}_τ .

(1) 初始化: $t=0$, 随机生成 $u(t)$, 使得 $u^T(t)u(t)=1$

(2) 计算 $a(t) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \mathbf{x}_j^{(i)}$

$$(3) \text{ 计算 } b(t) = \left(\frac{|u_1(t)|}{\lambda + \eta |u_1(t)|}, \frac{|u_2(t)|}{\lambda + \eta |u_2(t)|}, \dots, \frac{|u_h(t)|}{\lambda + \eta |u_h(t)|} \right)^T.$$

$$(4) \mathbf{u}(t+1) = a(t) \circ b(t)$$

$$(5) \mathbf{u}(t+1) = \mathbf{u}(t+1) / \|\mathbf{u}(t+1)\|_2$$

(6) 收敛条件: 如果 $\|\mathbf{u}(t+1) - \mathbf{u}(t)\|_2 > \varepsilon$ 则转到(2)

$$3. \text{ 使用贪心策略更新样本: } x^{(i)} = x^{(i)} - \sum_{k=1}^r u_k (u_k^T x^{(i)})$$

4. 如果 $\tau < r$, 则 $\tau = \tau + 1$ 并转到 2

2.3 收敛性证明

定理 在 BPCAL1-S 过程中, 每一轮迭代, 目标函数是单调递增的. 即:

$$H(\mathbf{u}(t+1)) \geq H(\mathbf{u}(t)) \quad (19)$$

证明:

$$H(\mathbf{u}(t+1)) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t+1) u^T(t+1) x_j^{(i)} - \frac{\eta}{2} \|\mathbf{u}(t+1)\|_2 - \lambda \|\mathbf{u}(t+1)\|_1 \quad (20)$$

上式中, 第 1 项为

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t+1) u^T(t+1) x_j^{(i)} &= \sum_{i=1}^n \sum_{j=1}^m |u^T(t+1) x_j^{(i)}| \geq \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) u^T(t+1) x_j^{(i)} = \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}(t) \bar{u}^T(t+1) \bar{x}_j^{(i)}, \end{aligned} \quad (21)$$

第 2 项为

$$-\frac{\eta}{2} \|\mathbf{u}(t+1)\|_2 = -\frac{\eta}{2} \|\bar{\mathbf{u}}(t+1)\|_2, \quad (22)$$

我们注意到式(16) $Q(\bar{\mathbf{u}}(t+1) | \bar{\mathbf{u}}(t))$ 的第 3 项中:

$$\begin{aligned} \bar{\mathbf{u}}^T(t+1) \bar{R}(t) \bar{\mathbf{u}}(t+1) + \|\bar{\mathbf{u}}(t)\|_1 &= \sum_{i=1}^{\bar{h}} \frac{\bar{u}_i^2(t+1)}{|\bar{u}_i(t)|} + \|\bar{\mathbf{u}}(t)\|_1 \geq \sum_{i=1}^{\bar{h}} \frac{\bar{u}_i^2(t+1)}{|\bar{u}_i(t+1)|} + \|\bar{\mathbf{u}}(t+1)\|_1 = \\ &= 2 \|\bar{\mathbf{u}}(t+1)\|_1 = 2 \|\bar{\mathbf{u}}(t+1)\|_1 \end{aligned} \quad (23)$$

即

$$-\lambda \|\mathbf{u}(t+1)\|_1 = -\lambda \|\bar{\mathbf{u}}(t+1)\|_1 \geq -\frac{\lambda}{2} (\bar{\mathbf{u}}^T(t+1) \bar{R}(t) \bar{\mathbf{u}}(t+1) + \|\bar{\mathbf{u}}(t)\|_1) \quad (24)$$

式(23)中的不等性, 是由于如下引理^[22]: 对于任意向量 $\mathbf{v} \in \mathbf{R}^p$, 有

$$\|\mathbf{v}\|_1 = \min_{z \in \mathbf{R}_+^p} \frac{1}{2} \sum_{j=1}^p \frac{v_j^2}{z_j} + \frac{1}{2} \|z\|_1 \quad (25)$$

当 $z_j = |v_j|$ ($j=1, 2, \dots, p$) 时, $\frac{1}{2} \sum_{j=1}^p \frac{v_j^2}{z_j} + \frac{1}{2} \|z\|_1$ 取得最小值 $\|\mathbf{v}\|_1$.

式(16)与式(20)均可看成三项之和. 联合式(20)、(21)、(22)、(24)以及(16), 可以看到, $H(\mathbf{u}(t+1))$ 的每一项均大于等于 $Q(\bar{\mathbf{u}}(t+1) | \bar{\mathbf{u}}(t))$ 的对应项, 可以得到:

$$H(\mathbf{u}(t+1)) \geq Q(\bar{\mathbf{u}}(t+1) | \bar{\mathbf{u}}(t)) \geq H(\mathbf{u}(t)) \quad (26)$$

3 实验与结果

为了评价 BPCAL1-S 的性能表现, 设计了图像分类与重构实验, 使用了 3 个基准图像集: ORL、Yale 和 Feret 库. 在 BPCAL1-S 中, 有两个可调参数 η 和 λ . 式(18)表明, 单位特征向量 \mathbf{u} 的提取实际上依赖于 η/λ . 定义 $\rho = \log_{10} \eta/\lambda$, 实验中从 $\rho = -2, -1, \dots, 1, 2$ 中找出最优的 ρ .

3.1 图像分类

在分类实验中, 将 BPCAL1-S 应用于 ORL 和 Yale 库. 基于不同的 ρ 值, 比较 BPCAL1-S 与 2DPCA、2DPCA-L1、BPCA 及 BPCA-L1 的分类性能.

图 2 左边部分的结果显示, ρ 的不同取值对 BPCAL1-S 算法性能影响很大, 在 ORL 库实验中, $\rho = -1$ 是较好的选择. 图 3 左边部分也显示了 Yale 实验中存在类似的情况, 但 ρ 的较优取值为 -2 .

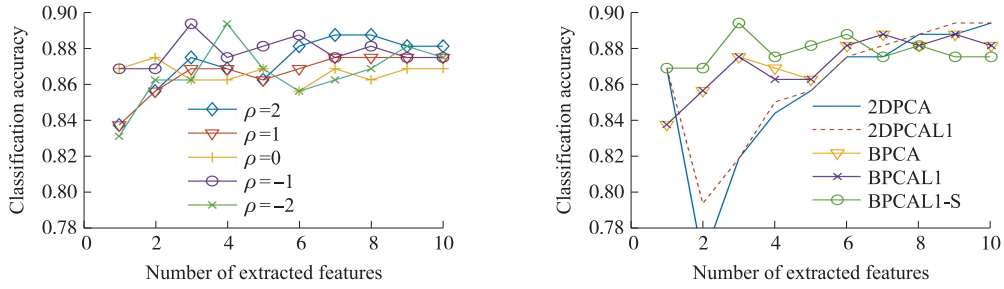


图 2 基于 ORL 库, BPCAL1-S 的分类准确率及其与 2DPCA, 2DPCA-L1, BPCA, BPCA-L1 的对比

Fig. 2 Classification accuracies of BPCAL1-S upon ORL and compare with 2DPCA, 2DPCA-L1, BPCA, BPCA-L1

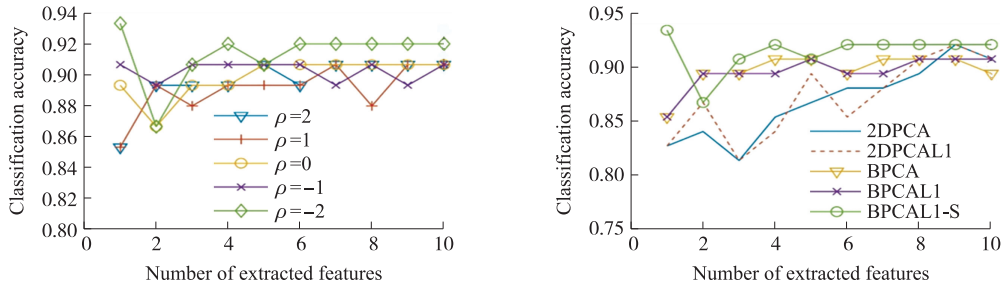


图 3 基于 Yale 库, BPCAL1-S 的分类准确率及其与 2DPCA, 2DPCA-L1, BPCA, BPCA-L1 的对比

Fig. 3 Classification accuracies of BPCAL1-S upon Yale and compare with 2DPCA, 2DPCA-L1, BPCA, BPCA-L1

图 2 和图 3 还显示, 分类准确率基于最初的几个特征就达到了最大值, 这说明最初的几个特征足以完成分类. 当特征数量继续增加时, 准确率逐步开始下降, 这说明过多的特征会相互影响, 反而可能造成算法性能下降. 在分类性能比较中, 对于 ORL 库和 Yale 库, 分别选择 $\rho = -1$ 和 $\rho = -2$, 图 2、图 3 的右边部分显示 BPCAL1-S 有较好的性能.

3.2 图像重构

图 4 和图 5 展示了图像重构在 Yale 库和 Feret 库上运行的实验结果. 实验中随机选择 20% 的图像样本添加随机矩形噪声, 作为原始带噪声样本. 结果显示, 当特征数量从 1 开始增加时, 各种方法的重构误差

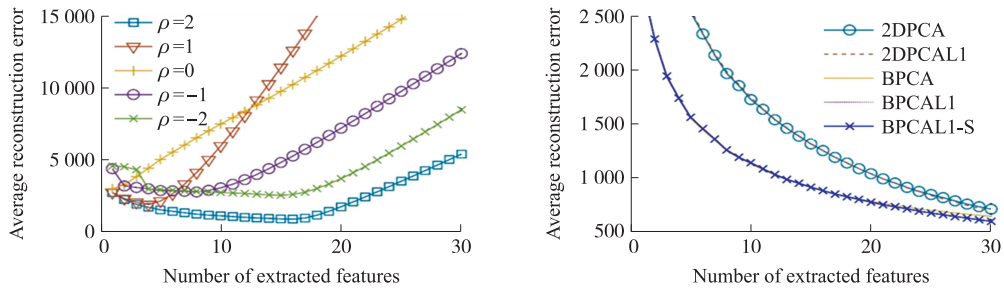


图 4 Yale 库平均重构误差

Fig. 4 Average reconstruction errors for Yale database

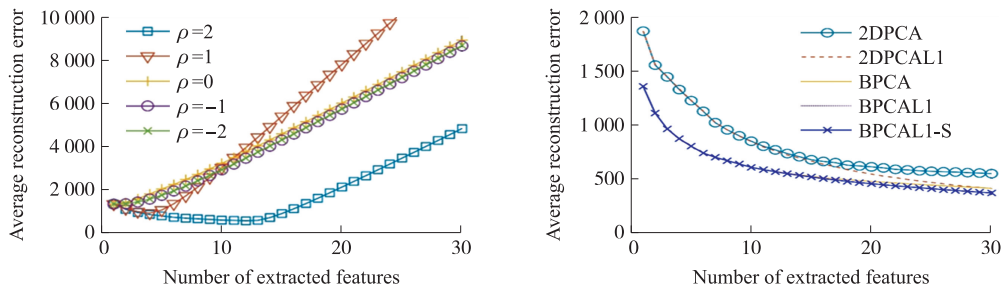


图 5 Feret 库平均重构误差

Fig. 5 Average reconstruction errors for Feret database

都在下降,但随着特征数量不断增加,重构误差反而开始上升,这与分类性能表现类似. 综合来说,相比所列举的其他方法,BPCAL1-S 表现出了更好的性能.

3.3 带噪声的图像重构

实验基于 Feret 库,图 6 为带噪声人脸图像用 5 个投影向量做图像重构的效果. 直观上看,与 2DPCA、2DPCA-L1 相比,块 PCA(BPCA,BPCA-L1,BPCAL1-S)重构图像更加接近于原始带噪声的样本,这与图 4、图 5 的平均重构误差曲线图显示的情况一致.

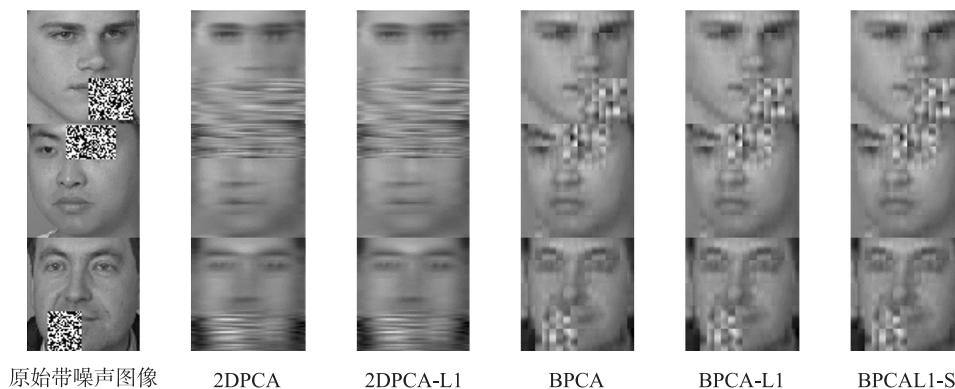


图 6 原始带噪声人脸图像和重构图像

Fig. 6 Face images with occlusion and the reconstructed faces

4 结论

本文提出的 BPCAL1-S 基于块计算单元,能更好地利用原始数据的部分关联. 目标函数中使用 L1-范数并整合了稀疏约束的弹性网,提出了 BPCAL1-S 特征抽取算法,所抽取的特征对噪声更加鲁棒且具有稀疏性,有更好的解释性与性能表现. 对算法过程做了理论分析,并证明了迭代过程的收敛性. 通过 ORL、Yale 和 Feret 库的图像分类、图像重构实验,验证了该方法的有效性.

[参考文献]

- [1] JOLLIFFE I. Principal component analysis[M]. New York,XY,USA:Springer,2004.
- [2] YANG J,ZHANG D,FRANGI A F,et al. Two-dimensional PCA;a new approach to appearance-based face representation and recognition[J]. IEEE transaction on pattern analysis and machine intelligence,2004,28(1):131-137.
- [3] KE Q,KANADE T. Robust L1 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. San Diego, USA:IEEE,2005.
- [4] DING C,ZHOU D,HE X,et al. R1-PCA:Rotational invariant L1-Norm principal component analysis for robust subspace factorization[C]//Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh,USA;the International Conference on Machine Learning,2006.
- [5] KWAK N. Principal component analysis based on L1-Norm maximization[J]. IEEE transaction on pattern analysis and machine intelligence,2008,30(9):1672-1680.
- [6] LI X,PANG Y,YUAN Y. L1-Norm-based 2DPCA[J]. IEEE transactions on systems man and cybernetics,2009,40(4):1170-1175.
- [7] NIE F,HUANG H,DING C,et al. Robust principal component analysis with non-greedy L1-norm maximization[C]// Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona,Spain;the International Joint Conferences on Artificial Intelligence,2011.
- [8] WANG R,NIE F,YANG X,et al. Robust 2DPCA with non-greedy L1-Norm maximization for image analysis[J]. IEEE transactions on cybernetics,2015,45(5):1108-1112.
- [9] KWAK N. Principal component analysis by Lp-Norm maximization[J]. IEEE transactions on cybernetics,2014,44(5):594-609.

- [10] GAO Q X. Is two-dimensional PCA equivalent to a special case of modular PCA [J]. Pattern recognition letters, 2007, 28(10):1250–1251.
- [11] GOTTUMUKKAL R, ASARI V K. An improved face recognition technique based on modular PCA approach [J]. Pattern recognition letters, 2004, 25(4):429–436.
- [12] KIM C, CHOI C H. Image covariance-based subspace method for face recognition [J]. Pattern recognition, 2007, 40(5):1592–1604.
- [13] WANG H. Block principal component analysis with L1-Norm for image analysis [J]. Pattern recognition letters, 2012, 33(5):537–542.
- [14] LI BN, YU Q, WANG R, et al. Block principal component analysis with nongreedy L1-Norm maximization [J]. IEEE transactions on cybernetics, 2015, 46(11):2543–2547.
- [15] 孙茹君, 张鲁飞. 基于动态指导的深度学习模型稀疏化执行方法 [J]. 南京师大学报(自然科学版), 2019, 42(3):11–19.
- [16] 张明华, 罗红玲, 宋巍, 等. 基于稀疏表示和学习图正则的高光谱图像特征提取 [J]. 光子学报, 2021, 50(4):0410002.
- [17] WANG H, WANG J. 2DPCA with L1-Norm for simultaneously robust and sparse modelling [J]. Neural networks, 2013, 46:190–198.
- [18] WANG J. Generalized 2-D principal component analysis by Lp-Norm for image analysis [J]. IEEE transactions on cybernetics, 2016, 46(3):792–803.
- [19] WANG Q, GAO Q, GAO X. L2, p-Norm based PCA for image recognition [J]. IEEE transactions on image processing, 2018, 27(3):1336–1346.
- [20] GAO Q, XU S, CHEN F, et al. R1-2-DPCA and face recognition [J]. IEEE transactions on cybernetics, 2019, 49(4):1212–1223.
- [21] ZOU H, HASTIE T, TIBSHIRANI R. Sparse principal component analysis [J]. Journal of computational and graphical statistics, 2019, 15(2):265–286.
- [22] JENATTON R, OBOZINSKI G, BACH F. Structured sparse principal component analysis [C]//Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. Chia Laguna Resort, Italy: JMLR Workshop and Conference Proceedings, 2010.

[责任编辑:黄 敏]