

## 鲁棒自适应概率加权主成分分析

高云龙<sup>1</sup> 罗斯哲<sup>1</sup> 潘金艳<sup>2</sup> 陈柏华<sup>1</sup> 张逸松<sup>1</sup>

**摘要** 主成分分析 (Principal component analysis, PCA) 是处理高维数据的重要方法. 近年来, 基于各种范数的 PCA 模型得到广泛研究, 用以提高 PCA 对噪声的鲁棒性. 但是这些算法一方面没有考虑重建误差和投影数据描述方差之间的关系; 另一方面也缺少确定样本点可靠性 (不确定性) 的度量机制. 针对这些问题, 本文提出一种新的鲁棒 PCA 模型. 首先采用  $L_{2,p}$  模来度量重建误差和投影数据的描述方差. 基于重建误差和描述方差之间的关系建立自适应概率误差极小化模型, 据此计算主成分对于数据描述的不确定性, 进而提出了鲁棒自适应概率加权 PCA 模型 (RPCA-PW). 此外, 本文还设计了对应的求解优化方案. 对人工数据集、UCI 数据集和人脸数据库的实验结果表明, RPCA-PW 在整体上优于其他 PCA 算法.

**关键词** 主成分分析, 加权主成分分析, 维数约简, 鲁棒性

**引用格式** 高云龙, 罗斯哲, 潘金艳, 陈柏华, 张逸松. 鲁棒自适应概率加权主成分分析. 自动化学报, 2021, 47(4): 825–838

**DOI** 10.16383/j.aas.c180743

## Robust PCA Using Adaptive Probability Weighting

GAO Yun-Long<sup>1</sup> LUO Si-Zhe<sup>1</sup> PAN Jin-Yan<sup>2</sup> CHEN Bai-Hua<sup>1</sup> ZHANG Yi-Song<sup>1</sup>

**Abstract** Principal component analysis (PCA) is an important method for processing high-dimensional data. In recent years, PCA models based on various norms have been extensively studied to improve the robustness. However, on the one hand, these algorithms do not consider the relationship between reconstruction error and covariance; on the other hand, they lack the uncertainty of considering the principal component to the data description. Aiming at these problems, this paper proposes a new robust PCA algorithm. Firstly, the  $L_{2,p}$ -norm is used to measure the reconstruction error and the description variance of the projection data. Based on the reconstruction error and the description variance, the adaptive probability error minimization model is established to calculate the uncertainty of the principal component's description of the data. Based on the uncertainty, the adaptive probability weighting PCA is established. The corresponding optimization method is designed. The experimental results of artificial data sets, UCI data sets and face databases show that RPCA-PW is superior than other PCA algorithms.

**Key words** Principle component analysis (PCA), weighted principal component analysis (WPCA), dimensionality reduction, robustness

**Citation** Gao Yun-Long, Luo Si-Zhe, Pan Jin-Yan, Chen Bai-Hua, Zhang Yi-Song. Robust PCA using adaptive probability weighting. *Acta Automatica Sinica*, 2021, 47(4): 825–838

随着技术的进步, 数据采集的效率逐渐提高, 使得数据的规模越来越大、复杂性越来越高. 在大多数情况下, 这些高维数据都存在着能够保留大部分有效信息的低维子空间, 如何移除高维空间中的噪声和无关信息, 提高后续学习算法的性能和效率一直是模式识别和机器学习领域的研究热点. 在过去的几十年中涌现出了许多优秀的算法, PCA<sup>[1]</sup> 是其中

最经典的方法之一, 它通过线性变换把数据投影到一个新的坐标空间中, 希望用较少的变量来表示原数据所提供的大部分信息. PCA 逐渐发展为多种应用的预处理技术方法, 如图像识别、生物信息和数据挖掘<sup>[2–4]</sup>. 由于其用途广泛且原理简单, 研究者们陆续提出了各种改进的 PCA 算法. Koren 等<sup>[5]</sup> 提出的 WPCA 使用了加权距离来减轻离群点对投影方向的影响, 突出了与主成分相关的特征; Schölkopf 等<sup>[6]</sup> 通过非线性映射将原始数据映射到高维特征空间, 再执行 kernel-PCA 以提取特征; 李春娜等<sup>[7]</sup> 极大化带有稀疏正则项的  $L_p$  模样本方差, 同时赋予算法鲁棒性和稀疏性.

衡量算法的优劣, 一个重要的指标就是鲁棒性, 尽管基于  $L_2$  模的 PCA 能够解决许多问题, 但不能有效地处理小样本问题中的离群点<sup>[8]</sup>, 因为  $L_2$  模的非线性变化特征会放大离群点所带来的影响, 使算法倾向于保留外围结构. 为了减轻异常点的负面

收稿日期 2018-11-07 录用日期 2019-04-26  
Manuscript received November 7, 2018; accepted April 26, 2019  
国家自然科学基金 (61203176), 福建省自然科学基金 (2013J05098, 2016J01756) 资助  
Supported by National Natural Science Foundation of China (61203176), Fujian Provincial Natural Science Foundation (2013J05098, 2016J01756)

本文责任编辑 白翔  
Recommended by Associate Editor BAI Xiang  
1. 厦门大学航空航天学院 厦门 361102 2. 集美大学信息工程学院 厦门 361021

1. School of Aeronautics and Astronautics, Xiamen University, Xiamen 361102 2. School of Information Engineering, Jimei University, Xiamen 361021

影响, 目前已经提出了各种增强鲁棒性的解决方案.  $L_1$  模被认为是增强算法鲁棒性的有效手段之一. Ke 等<sup>[9]</sup> 提出了  $L_1$ -PCA 算法, 通过极小化基于  $L_1$  模的重建误差来提取主成分; Kwak<sup>[10]</sup> 则在特征空间中极大化对应的  $L_1$  模并利用贪婪算法求解模型; 在此基础上, Nie 等<sup>[11]</sup> 提出了一种非贪婪迭代算法能够得到比贪婪算法更好的结果.

尽管基于  $L_1$  模的 PCA 鲁棒性较强, 但是由于计算代价大, 而且不具有旋转不变性<sup>[12]</sup>. 因此, 大量具有旋转不变性的鲁棒 PCA 算法相继出现, 这些方法通过采用不同的准则函数或者优化算法来降低异常点对损失函数的影响, 以提高主成分分析过程中对于异常点的鲁棒性. He 等在文献 [13] 中将 PCA 的均方误差 (MSE) 准则修改为最大熵 (MaxEnt) 准则来尽可能地保留数据的不确定性; 进而在文献 [14] 中提出 HQ-PCA, 使用最大相对熵准则 (MCC) 代替 MSE, 并采用半二次 (Half-Quadratic) 优化将原问题转换为一系列二次规划问题进行求解. HQ-PCA 提高了算法对于噪声的鲁棒性, 同时保留了平移与旋转不变性; He 等在文献 [15] 中基于数据的子空间属性, 分析了鲁棒低秩矩阵恢复方法和基于  $M$  估计的鲁棒主成分分析方法之间的联系, 提高主成分提取过程中对任意噪声的处理能力; Ding 等<sup>[16]</sup> 使用旋转不变的 R1 模构造重建误差, 在一定程度上抑制了离群点的影响, 但是该方法依赖于投影空间中的维数; Nie 等<sup>[17]</sup> 在此基础上提出了 RPCA-OM, 计算了在 R1 范数下的最优均值并能够自动删除最优的数据均值; 受此启发, 许多鲁棒 PCA 采用  $L_{21}$  模作为鲁棒降维的有效手段. Nie 等<sup>[18]</sup> 基于  $L_{21}$  范数最大化在理论上与重构误差最小化的关联性提出了 PCA- $L_{21}$ , 并设计了一种有效的非贪婪优化算法来求解相关的最大化问题; Wang 等<sup>[19]</sup> 将  $L_{21}$  模的距离度量扩展为  $L_{2,p}$ , 可针对不同的数据选择适当的  $p$  以达到更好的效果; 但以上鲁棒 PCA 算法缺乏考虑重建误差和投影数据描述方差之间的关系, 在主成分提取的过程中容易造成判别信息的丢失. 对此, Wang 等<sup>[20]</sup> 提出的 Angle PCA 方法通过最大化每个样本点的描述方差和重建误差之间的比率来确定主成分空间, 通过每个数据点与主成分方向的偏移角度进行加权, 但其权值的变化呈余切函数的快速非线性变化特征, 导致其过度强调局部特征, 所提取的主成分泛化性能弱.

基于此, 本文提出了鲁棒自适应概率加权主成分分析 (RPCA-PW). RPCA-PW 基于样本点的重建误差与描述方差在  $L_{2,p}$  模下的变化关系确定每个样本点的可靠性程度. 其核心是选择主成分空间及其补空间作为参考, 通过分析样本点与这两个描述空间的相似度来确定主成分空间及其补空间对数据

描述的不确定性, 结合交替迭代的优化算法, 从而能够在降维过程中自适应地降低噪声和异常样本点的影响. 本文提出的方法不仅对离群点具有鲁棒性, 并可针对不同数据集选择合适的  $p$  以达到更好的效果, 本文将在人工数据集、UCI 数据集和人脸图像数据库上进行实验, 进而证明本文所提出算法的有效性.

## 1 相关工作

### 1.1 PCA

考虑如下样本矩阵:  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$ , 其中  $n$  和  $d$  分别为样本数量和维数. 不失一般性, 这里假设  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$  已经去中心化, 即  $\sum_{i=1}^n \mathbf{x}_i = 0$ , 定义投影矩阵  $W \in \mathbf{R}^{d \times m}$  ( $m < d$ ), 其中  $m$  是降维后的维数. 从均方差的角度来看, PCA 通过求解以下优化问题, 使得投影空间中样本点的方差最大化:

$$\max_{W^T W = I} \sum_{i=1}^n \|W^T \mathbf{x}_i\|_2^2 \quad (1)$$

其中,  $I$  是单位矩阵, 因为  $\|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^2 + \|W^T \mathbf{x}_i\|_2^2 = \|\mathbf{x}_i\|_2^2$ , 因此通过简单的变换可以得到式 (1) 的另一种等价形式:

$$\min_{W^T W = I} \sum_{i=1}^n \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^2 \quad (2)$$

由式 (1) 和式 (2) 可知, 由于  $L_2$  模对离群点敏感, 传统 PCA 对噪声的鲁棒性不强, 噪声的存在会使得 PCA 的计算结果会出现很大的误差.

### 1.2 $L_{2,p}$ -PCA

$L_{2,p}$ -PCA 采用  $L_{2,p}$  模作为重建误差的距离度量, 可针对不同的数据选择适当的  $p$  以达到更好的效果.  $L_{2,p}$ -PCA 不仅能在一定程度上削弱噪声点的影响, 而且还保留了 PCA 所需的特性, 如旋转不变性. 此外, 基于  $L_{2,1}$  模的鲁棒 PCA 可作为  $L_{2,p}$ -PCA 的特例.  $L_{2,p}$ -PCA 的目标函数定义为:

$$\min_{W^T W = I} \sum_{i=1}^n \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p \quad (3)$$

$L_{2,p}$  模的非线性函数特征在一定程度上降低了噪声和异常样本点的影响力, 但是仍然不能完全剔除噪声和异常样本点的影响. 究其原因,  $L_{2,p}$ -PCA 的根本特征在于仅考虑了样本点与数据簇整体统计特征的偏差程度 (bias), 但没有考虑噪声点与可靠数据点在不同的子空间属性下潜在的可分性, 从而造成判别信息的丢失.

### 1.3 Angle PCA

Angle PCA 采用  $L_2$  模来构造投影数据的重建误差和描述方差, 通过最大化方差和重构误差之比来确定投影矩阵, 即 Angle PCA 通过求解以下目标函数来确定主成分:

$$\max_{W^T W=I} \sum_{i=1}^n \frac{\|W^T \mathbf{x}_i\|_2^1}{\|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^1} \quad (4)$$

根据  $\|W^T \mathbf{x}_i\|_2^2 + \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^2 = \|\mathbf{x}_i\|_2^2$ , 若将  $\|\mathbf{x}_i\|_2^1$  视为直角三角形的斜边, 则  $\|W^T \mathbf{x}_i\|_2^1$  和  $\|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^1$  分别为直角三角形的两条直角边, 令直角边  $\|W^T \mathbf{x}_i\|_2^1$  与斜边  $\|\mathbf{x}_i\|_2^1$  的夹角为  $\alpha_i$ , 观察模型 (4) 可知, 求和函数中的每一项是第  $i$  个数据的重建误差与协方差之间角度的余切值, 即

$$\frac{\|W^T \mathbf{x}_i\|_2^1}{\|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^1} = \cot \alpha_i \quad (5)$$

因此, 目标函数 (4) 被称为 Angle PCA. 通过对样本点迭代加权的方式来降低噪声和异常样本的影响. 这种建模方式的核心是能够减少重建误差较大样本点产生的损失, 从而提升对噪声的鲁棒性. 但  $\cot \alpha_i$  的非线性快速衰减特征造成了 Angle PCA 对数据的全局结构特征提取能力差. 例如: 当样本点与主成分方向之间的夹角增大时,  $\cot \alpha_i$  迅速减小, 使得主成分对数据的局部结构特征描述能力强, 但是对数据的全局结构特征描述能力差. 这一特征造成 Angle PCA 最优解的稳定性差, 对初始投影矩阵  $W$  的选取依赖性很强. 例如: 当模型的初始  $W$  选择恰当时, 则 Angle PCA 对噪声点具有很强的鲁棒性, 若初始  $W$  确定的投影方向与实际主成分方向垂直的时候, 式 (5) 起到的作用则正好相反, 表现为突出非主成分样本点在模型中所占的比重 (此时非主成分样本点的重建误差很小).

## 2 鲁棒自适应概率加权主成分分析

为了能够充分考虑重建误差和投影数据描述方差之间的联系, 并根据数据主要统计特征及其互补信息确定各样本点的可靠程度, 在提取主成分的过程中, 提高可靠度较高样本点的影响力, 同时削弱可靠度较低样本点的影响程度, 本文建立以下 RPCA-PW 模型:

$$\max_{W^T W=I} \sum_{i=1}^n \left( \|W^T \mathbf{x}_i\|_2^p - \frac{1-a_i}{a_i+\varepsilon} \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p \right) \quad (6)$$

其中,  $0 \leq a_i \leq 1$ ,  $\varepsilon$  为一个较小的正数常数, 目的是防止分母为 0. 令  $\frac{1-a_i}{a_i+\varepsilon} = \delta_i$ , 则式 (6) 变为:

$$\max_{W^T W=I} \sum_{i=1}^n (\|W^T \mathbf{x}_i\|_2^p - \delta_i \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p) \quad (7)$$

模型 (7) 中采用了  $L_{2,p}$  模作为度量标准, 不仅可以降低噪声的影响, 而且具有旋转不变性, 改变  $p$  值的大小可应用于不同类型的数据集, 大大提高了算法的灵活性和鲁棒性.

为了提高 RPCA-PW 对数据全局结构特征的描述能力和对噪声的鲁棒性, 需要模型 (7) 中  $\delta_i$  满足以下要求:

1) 能够反映出样本点的可靠性 (不确定性). 对于可靠样本点,  $\delta_i$  应取较大的值, 对于噪声和异常样本点,  $\delta_i$  应取较小的值, 通过分析样本点的不确定性削弱噪声和异常值的影响.

2) 在主成分空间所确定的一个较大的邻域内,  $\delta_i$  的取值应保持稳定, 以提高主成分对数据全局结构特征的提取能力, 从而提高主成分的泛化性能.

3) 能够根据数据受噪声污染的实际情况动态调整各局部邻域的大小, 即能够动态调整主成分局部邻域和主成分互补空间局部邻域的划分界限.

为了满足上述要求, 本文通过以下模型确定权值:

$$\min_{0 \leq a_i \leq 1} a_i \|W^T \mathbf{x}_i\|_2^p + (1-a_i) \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p + \lambda_i (a_i^2 + (1-a_i)^2) \quad (8)$$

其中,  $\|W^T \mathbf{x}_i\|_2^p$  ( $\|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p$ ) 表示主成分空间 (主成分空间的补空间) 对数据的描述能力,  $\|W^T \mathbf{x}_i\|_2^p$  越大 ( $\|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p$  越小), 代表样本点隶属于主成分空间的程度越大, 该样本点越可靠, 反之亦然. 令式 (8) 中的  $\lambda_i = 0$ , 有  $a_i \in \{0, 1\}$ , 上述模型类似于 1NN 规则, 即对于每个样本点要么选择主成分空间, 要么选择其补空间来描述. 但是这种“非此即彼”结果大大降低了算法的泛化性能, 也使得算法在验证集上的实验结果缺乏鲁棒性. 如何软化模型对于样本点的判定结果使得结果更加契合主成分提取的过程是我们重点考虑的一个问题. 基于以上的分析, 在模型中添加基于  $L_2$ -norm 的正则化项  $\lambda (a_i^2 + (1-a_i)^2)$ . 因此  $a_i$  可以理解把第  $i$  个样本点划分到主成分空间的错误概率, 通过  $L_2$ -norm 的正则化项, 上述模型考虑了每个样本点划分到主成分空间和其补空间的错分概率所带来的误差.

令  $u_{1,i} = \|W^T \mathbf{x}_i\|_2^p, u_{2,i} = \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p$ , 则式 (8) 可写成:

$$\min_{0 \leq a_i \leq 1} a_i u_{1,i} + (1 - a_i) u_{2,i} + \lambda_i (a_i^2 + (1 - a_i)^2) \quad (9)$$

由于  $a_i$  对于不同  $i$  的取值相互独立, 因此通过简单的数学变换:

$$\begin{aligned} \min_{0 \leq a_i \leq 1} a_i u_{1,i} + (1 - a_i) u_{2,i} + \lambda_i (a_i^2 + (1 - a_i)^2) &\Rightarrow \min_{0 \leq a_i \leq 1} 2\lambda_i a_i^2 - 2\lambda_i a_i + \lambda_i + a_i u_{1,i} + (1 - a_i) u_{2,i} \\ &\Rightarrow \min_{0 \leq a_i \leq 1} 2a_i^2 - 2a_i + \frac{a_i u_{1,i} + (1 - a_i) u_{2,i}}{\lambda_i} \end{aligned} \quad (10)$$

求得  $a_i$  最优解的解析形式为:

$$a_i = \begin{cases} \frac{2\lambda_i - u_{1,i} + u_{2,i}}{4\lambda_i}, & 0 \leq \frac{2\lambda_i - u_{1,i} + u_{2,i}}{4\lambda_i} \leq 1 \\ 0, & \frac{2\lambda_i - u_{1,i} + u_{2,i}}{4\lambda_i} < 0 \\ 1, & \frac{2\lambda_i - u_{1,i} + u_{2,i}}{4\lambda_i} > 1 \end{cases} \quad (11)$$

式 (11) 中基于  $a_i$  最优解的稀疏性结构特征使得在主成分周围的邻域内,  $\delta_i$  为某个固定的较大正数, 在重建误差相对大的噪声样本局部邻域内,  $\delta_i$  为某个固定的非常小的正数或等于零. 另外式 (11) 中的参数  $\lambda_i$ , 决定了局部稀疏邻域的大小, 即决定了主成分样本的局部邻域与异常样本局部邻域的划分, 通过调节  $\lambda_i$  可以动态调整各自局部邻域的大小.

下面通过与 PCA、Angle PCA 的比较实例来说明 RPCA-PW 对离群点的鲁棒性:

如图 1 所示, 图 1(a) 中是一组由 500 个数据点组成的服从高斯分布的数据簇, 在图 1(b) 中将随机插入 10 个与原分布差异很大的离群点. 结果显示, PCA 对样本点的分布非常敏感, 主方向计算结果误差非常大; 因为  $\cot(\alpha_i)$  函数非线性快速衰减的加权方式, 使 Angle PCA 在离群点的影响之下, 投影方向也发生了一定程度的偏离, 图 1(c) 显示了  $\cot(\alpha_i)$  的变化过程; 而 RPCA-PW 的投影方向不受离群点的影响, 主成分空间仍然很好地保留了原数据的全局结构特征, 有效排除少数离群点造成的影响. 这是因为 RPCA-PW 充分考虑了数据在描述

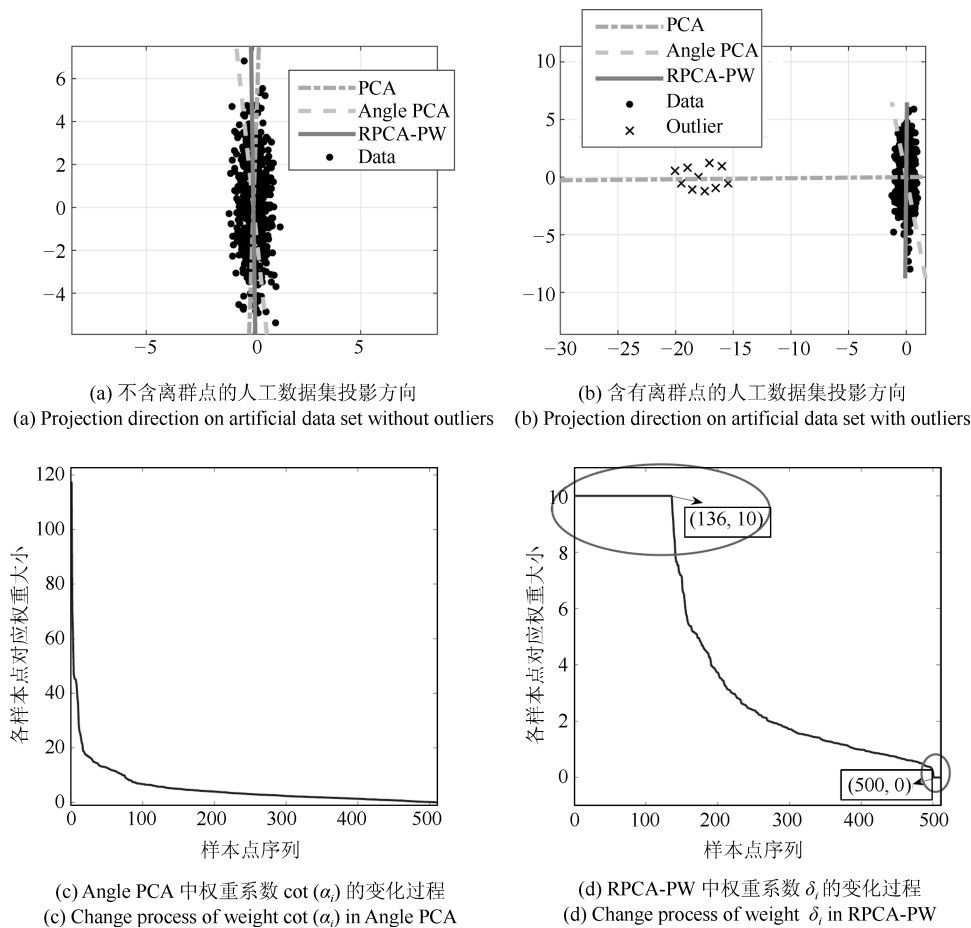


图 1 人工数据集上的鲁棒性实验

Fig. 1 Robustness experiment on artificial data set



空间中的分布特征, 可以将脱离数据集的异常样本造成的影响降到最低. 图 1(d) 中显示了权重系数  $\delta_i$  的变化过程, 根据  $\delta_i$  的变化趋势, 可以分为三个阶段, 分别代表着 RPCA-PW 为主成分点、过渡点和离群点所属的三个局部区域所分配的不同权重 (对于离群点, 有  $\delta_i = 0$ ), 通过调节参数  $\lambda_i$  可以控制两个跳变点的位置, 即 (138, 10) 与 (500, 0) 的位置, 进而调整三类区域的范围.

### 3 优化算法

本节中利用优化理论中的交替优化算法求解模型 (7), 通过简单的数学代换, 可得:

$$\begin{aligned} & \sum_{i=1}^n (\|W^T \mathbf{x}_i\|_2^p - \delta_i \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p) = \\ & \sum_{i=1}^n \|W^T \mathbf{x}_i\|_2^2 \|W^T \mathbf{x}_i\|_2^{p-2} - \\ & \sum_{i=1}^n \delta_i \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^2 \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^{p-2} \end{aligned} \quad (12)$$

令  $d_{1,i} = \|W^T \mathbf{x}_i\|_2^{p-2}$ ,  $d_{2,i} = \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^{p-2}$ , 并代入式 (12) 可得

$$\begin{aligned} & \sum_{i=1}^n \|W^T \mathbf{x}_i\|_2^2 \|W^T \mathbf{x}_i\|_2^{p-2} - \\ & \sum_{i=1}^n \delta_i \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^2 \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^{p-2} = \\ & \sum_{i=1}^n \left( d_{1,i} \text{Tr}(\mathbf{x}_i^T WW^T \mathbf{x}_i) - \right. \\ & \left. \delta_i d_{2,i} \text{Tr}(\mathbf{x}_i^T WW^T \mathbf{x}_i) \right) = \\ & \sum_{i=1}^n \text{Tr}(W^T (d_{1,i} + \delta_i d_{2,i}) \mathbf{x}_i \mathbf{x}_i^T W) \end{aligned} \quad (13)$$

将式 (13) 代入模型 (7) 中, 模型 (7) 最后变为:

$$\max_{W^T W = I} \sum_{i=1}^n \text{Tr}(W^T (d_{1,i} + \delta_i d_{2,i}) \mathbf{x}_i \mathbf{x}_i^T W) \quad (14)$$

RPCA-PW 的目标是找到一个投影矩阵  $W$ , 以最大化目标函数 (14) 的值, 其中有三个与  $W$  相关的未知变量分别是  $W$ 、 $d_{1,i}$  和  $d_{2,i}$ . 因此, 目标函数没有闭式解, 难以直接求解目标函数 (14). 本文采用文献 [20] 中的算法来交替地更新  $W$ 、 $d_{1,i}$  和  $d_{2,i}$ , 具体来说, 是在第  $i$  次迭代中, 当  $d_{1,i}$  和  $d_{2,i}$  已知时, 通过最小化目标函数 (14) 更新投影矩阵  $W$ . 在这种

情况下, 将目标函数 (14) 简化为:

$$\max_{W^T W = I} \text{Tr}(W^T X D X^T W) \quad (15)$$

其中,  $D$  是对角矩阵, 对角线上的元素为  $D_{ii} = d_{1,i} + \delta_i d_{2,i}$ ,  $a_i$  可由式 (11) 获得. 根据矩阵理论, 目标函数 (15) 中投影矩阵  $W$  的列向量由矩阵  $X D X^T$  的前  $m$  个最大特征值所对应的特征向量组成, 随后再使用得到的  $W$  来更新  $d_{1,i}$  和  $d_{2,i}$ , 重复该迭代过程直至算法收敛. 算法 1 中给出了求解 RPCA-PW 的具体算法.

#### 算法 1. RPCA-PW 算法

输入:  $X \in \mathbf{R}^{d \times n}$ ,  $k = 1, m, p, \varepsilon$ , 其中  $\sum_{i=1}^n \mathbf{x}_i = 0$

初始化:  $W^0 \in \mathbf{R}^{d \times m}$  且  $W^{0T} W^0 = I$ , 目标函数大小  $J(W^0)$

#### 1: While not converge

2: 对于所有的样本点, 计算  $d_{1,i} = \|W^T \mathbf{x}_i\|_2^{p-2}$ ,  $d_{2,i} = \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^{p-2}$ ;

3: 对于每个  $i$ , 通过式 (11) 来更新权值  $a_i$ ;

4: 计算对角矩阵  $D$ , 其中对角线上的元素为  $D_{ii} = d_{1,i} + \delta_i d_{2,i}$ ;

5: 更新投影矩阵  $W$ , 其中  $W$  的列向量由矩阵  $X D X^T$  的前  $m$  个最大特征值所对应的特征向量组成;

6: 若  $J(W^k) \geq J(W^{k-1})$ , 转到第 8 步, 否则转到第 7 步;

7: 通过线搜索 Armijo 算法确定步长的子梯度下降法<sup>[21]</sup> 找到满足  $J(W^k) \geq J(W^{k-1})$  的  $W^k$ , 如果没有解决方案, 转到步骤 9, 否则, 转到步骤 8;

8:  $k \leftarrow k + 1$ ;

#### 9: End while

输出:  $W \in \mathbf{R}^{d \times m}$

## 4 讨论

本节中讨论了 RPCA-PW 算法的收敛性以及  $L_{2,p}$  范数下不同  $p$  值所对应的损失函数变化情况, 并给出本文算法与其他相关鲁棒 PCA 算法的关系.

### 4.1 算法收敛性

本文的基本思路是对于给定的主成分空间, 通过模型 (8) 学习每个样本点的误差极小化概率描述, 将学习到的概率描述反馈到模型 (7) 中, 从而自适应地修改模型 (7) 中描述误差项的权重, 从而达到提高 PCA 鲁棒性的目的. 因此算法 1 的收敛性应该从以下三个方面考虑:

1) 给定  $W$  条件下, 对模型 (8) 的优化

对于  $W$  已经给定的条件下, 模型 (8) 存在解析解, 其最优解由式 (11) 给出.

2) 给定  $\delta_i$  条件下, 对模型 (7) 的优化模型 (7) 的拉格朗日函数为:

$$L_1(W) = \sum_{i=1}^n (\|W^T \mathbf{x}_i\|_2^p - \delta_i \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p) - \text{Tr}(\Lambda(W^T W - I)) \quad (16)$$

其中, 拉格朗日乘子  $\Lambda$  是用于强制正交约束  $W^T W = I$  的对角矩阵. 在第  $k$  次迭代中, 当  $\delta_i$  已知时, 为了满足 KKT 条件, 式 (16) 的梯度必须等于零, 即

$$\frac{\delta L_1}{\delta W} = XDX^T W - \Lambda W = 0 \quad (17)$$

根据算法 1 中的步骤 5, 能够找到式 (17) 的最优解. 因此, 算法 1 的收敛解决方案满足问题的 KKT 条件. 式 (16) 的拉格朗日方程为:

$$L_2(W) = \text{Tr}(W^T XDX^T W) - \text{Tr}(\Lambda(W^T W - I)) \quad (18)$$

对  $W$  求导可得到关于式 (18) 的 KKT 条件:

$$XDX^T W - \Lambda W = 0 \quad (19)$$

注意到在式 (19) 中, 矩阵  $D$  与  $W_{k-1}$  有关. 假设在第  $k$  次迭代中获得局部最优解  $W^*$ , 即  $W^* = W_k = W_{k-1}$ . 在这种情况下, 式 (17) 与式 (19) 保持一致, 这意味着算法 1 的收敛解决方案满足模型 (7) 的 KKT 条件, 即

$$\left. \frac{\delta L_1}{\delta W} \right|_{W=W^*} = 0 \quad (20)$$

上述分析保证了算法 1 的最优解是模型的一个驻点. 此外, 算法 1 中的步骤 6 和步骤 7 说明 RPCA-PW 的目标函数值在每次迭代中都是非递减的, 这样就保证了算法具有单调收敛的特性.

3) 迭代更新  $W$  和  $\delta_i$ , 算法 1 的收敛性验证

这里通过实验验证迭代更新  $W$  和  $\delta_i$  时算法 1 的收敛性, 具体内容见实验部分第 5.3.5 节.

#### 4.2 不同 $p$ 值下的损失函数

图 2 绘制了  $p$  在不同取值下的目标函数取值变化曲线. 从图中可以看出, 当  $p$  为 2 时, 具有大重建误差的样本点将会显著地支配目标函数, 因此传统的基于 MSE 的 PCA 算法对于噪声敏感. 与  $p = 2$  相比,  $p = 0.5$  或 1 可以在一定程度上减弱大距离样本点的影响, 对于噪声点具有更强的鲁棒性, 此外, 与  $p = 1$  或 2 相比,  $p = 0.5$  可以进一步削弱异常样本点的影响, 同时提高主成分邻域中的样本点在求解最优解时的影响.

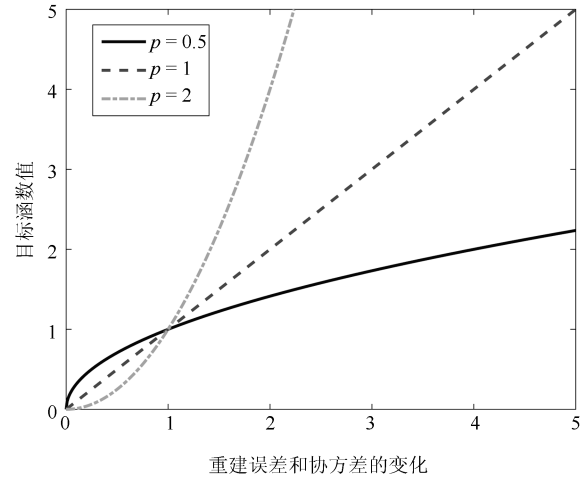


图 2  $p$  在不同取值下的目标函数取值变化曲线

Fig. 2 Objective function values under different  $p$

#### 4.3 与相关算法之间的联系

基于对于  $p$  的不同取值的分析, 下面将讨论 RPCA-PW 与几种鲁棒 PCA 算法之间的相互联系, 从模型的角度出发分析各自算法的特点, 包括 PCA, R1-PCA,  $L_{2,p}$ -PCA 和 Angle PCA.

##### 4.3.1 与 PCA 的联系

经典 PCA 算法建立在均方差或者重建误差意义上, 目的是使得投影空间中样本点协方差最大或者重建误差最小, 本质是在最小均方差意义下寻找最能代表数据特征的投影向量空间, 因此有:

1) 若模型 (6) 中的  $a_i = 1$  ( $i = 1, \dots, n$ ),  $p = 2$ , 得到:

$$\max_{W^T W = I} \sum_{i=1}^n \|W^T \mathbf{x}_i\|_2^2 \quad (21)$$

2) 若模型 (6) 中的  $a_i = 0$  ( $i = 1, \dots, n$ ),  $p = 2$ , 得到:

$$\min_{W^T W = I} \sum_{i=1}^n \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^2 \quad (22)$$

模型 (21) 和 (22) 是 PCA 的两种标准等价形式.

##### 4.3.2 与 R1-PCA 的联系

一个数据矩阵的 R1 模就是每个数据点的  $L_2$  模之和, R1-PCA 将原 PCA 模型中的  $L_2$  模改成 R1 模进行求解, R1 模在降低噪声影响的同时能保持旋转不变性. 若模型 (6) 中的  $a_i = 0$  ( $i = 1, \dots, n$ ),  $p = 1$ , 可得

$$\min_{W^T W = I} \sum_{i=1}^n \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2 \quad (23)$$

模型 (23) 即为 R1-PCA 的目标函数.

### 4.3.3 与 $L_{2,p}$ -PCA 的联系

$L_{2,p}$ -PCA 采用  $L_{2,p}$  模构造每个样本点的重建误差, 与大多数现有的 PCA-L1 方法相比,  $L_{2,p}$ -PCA 直接最小化了样本点的重建误差, 与 L1-PCA 方法相比,  $L_{2,p}$ -PCA 保留了 PCA 的旋转不变性. 若模型 (6) 中的  $a_i = 0$  ( $i = 1, \dots, n$ ), 则它的形式简化为:

$$\min_{W^T W = I} \sum_{i=1}^n \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p \quad (24)$$

模型 (24) 即为  $L_{2,p}$ -PCA 的目标函数.

### 4.3.4 与 Angle PCA 的联系

Angle PCA 采用最大化每个样本点的协方差和重建误差之比的总和所得到的结果作为投影矩阵, 根据文献 [22] 中的定理 1, 模型 (4) 可以等价转换为以下形式:

$$\max_{W^T W = I} \sum_{i=1}^n \left( \|W^T \mathbf{x}_i\|_2^1 - \mu_i \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^1 \right) \quad (25)$$

其中,  $\mu_i^k = \frac{\|(W^{k-1})^T \mathbf{x}_i\|_2^1}{\|\mathbf{x}_i - W(W^{k-1})^T \mathbf{x}_i\|_2^1}$ ,  $k$  表示优化算法中当前的迭代次数, 根据第 2 节中的定义, 可以认为  $\mu_i^k = \cot(\alpha_i^{k-1})$  ( $\alpha_i$  为直角边  $\|W^T \mathbf{x}_i\|_2^1$  与斜边  $\|\mathbf{x}_i\|_2^1$  的夹角). 若令模型 (6) 中

$$\frac{1 - a_i}{a_i + \varepsilon} = \frac{\|(W^{k-1})^T \mathbf{x}_i\|_2^1}{\|\mathbf{x}_i - W(W^{k-1})^T \mathbf{x}_i\|_2^1} \quad (26)$$

则它与模型 (4) 中的加权效果相同.

## 5 实验结果与分析

本节将文中所提算法在人工数据集、UCI 数据集和人脸数据库上进行实验, 并将其性能与传统 PCA 和相关鲁棒 PCA 算法进行比较, 例如: PCA-L21、RPCA-OM、Angle PCA、MaxEnt-PCA、HQ-PCA 和  $L_{2,p}$ -PCA. 其中式 (6) 中的参数  $\varepsilon$  是为了防止除零, 本文在实验中统一设置为 0.05, 另一参数  $\lambda_i$  的取值决定主成分及非主成分的局部邻域大小, 实验中根据不同的数据集动态地进行调整, 为了方便超参数设置, 本文在实验中统一置  $\lambda_i$  为:

$$\lambda_i = \lambda = \frac{1}{2n} \sum_{i=1}^n \left( \|\mathbf{x}_i - WW^T \mathbf{x}_i\|_2^p - \|W^T \mathbf{x}_i\|_2^p \right) \quad (27)$$

在实验中使用最近邻域 (1-NN) 分类器<sup>[23]</sup> 进行分类以计算识别准确率, 在人脸重构实验中, 平均重建误差由原始未被遮挡的图像与重构图像之间的平均距离定义, 如下所示:

$$error = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i^{org} - WW^T \mathbf{x}_i\|_2 \quad (28)$$

其中,  $n$  为图片总数,  $\mathbf{x}_i^{org}$  为未加入噪声块的原始图像,  $\mathbf{x}_i$  为加入噪声块后的图像.

### 5.1 人工数据集实验

为了验证本文所提算法在进行特征提取时能够有效挖掘数据集中的底层结构, 实验选用的人工数据集是一组随机生成的双高斯数据集. 在该数据集中, 存在两个服从高斯分布的数据簇, 目标是找到一个合适的低维投影空间, 使得两组数据簇在低维子空间中能够明显分开. 然后将 RPCA-PW 与 PCA 和 Angle PCA 进行比较, 结果如图 3 所示.

从图 3 中可以看出, 当这两组数据簇相距很远时, 三种 PCA 方法都能够找到一个合适的投影方向使得数据集在一维空间中保留绝大多数的信息. 随着两个数据簇之间的距离逐渐减少, PCA 变得不再有效, 具体表现为在一维投影空间中, 两组数据簇互相重叠, 完全丢失判别信息. 随着两个集群变得更加接近, PCA 与 Angle PCA 都无法得到最佳的投影方向. 而 RPCA-PW 所找到的一维投影在各种情况下都能将两组数据簇明显分开. 实验结果说明 PCA 由于倾向于保留数据集整体的外围结构, 受到大距离样本的干扰明显; Angle PCA 更加注重保留局部结构, 但是它采用的快速衰减加权方式将过多的样本点当成劣点处理, 丢失了过多的全局结构判别信息; RPCA-PW 在提取主成分的过程中, 在突出样本点主成分与互补子空间的区域性差异的同时, 能够有效揭示数据集中的底层结构.

### 5.2 UCI 数据集实验

为了验证本文提出算法的特征提取能力, 本实验选用了 UCI 数据集中常用的几个数据集, 如: Australian、Cars、Cleve 和 Solar 等, 这些数据集已被用于许多研究, 具体信息在表 1 中列出. 为了消除随机影响, 实验时采用了十折交叉验证法, 将每个数据集降至  $c - 1$  维, 其中  $c$  为类别数, 然后采用最近邻分类器对每一数据集求出识别正确率并计算出对应的重建误差. 其中对于采用  $L_{2,p}$ -norm 的作为距离度量算法, 实验中分别将  $p$  设为 0.5, 1, 1.5, 2 四个值进行实验. 最终得到 8 种算法在 10 个数据集上的平均识别正确率和重建误差大小. 实验结果如表 2 和表 3 所示.

表 1 实验中使用的 UCI 数据集

Table 1 UCI data sets used in the experiment

数据集	维数	类别数	样本数
Australian	14	2	690
Cars	8	3	392
Cleve	13	8	303
Solar	12	6	323
Zoo	19	7	101
Control	60	6	600
Crx	15	2	690
Glass	9	6	214
Iris	4	3	150
Wine	13	3	178

通过观察发现, 在完全相同且随机的实验条件下 RPCA-PW 能够在 8 组 UCI 数据集上取得更高的平均识别正确率, 同时在 5 组 UCI 数据集上有着最小的重建误差. 此外, HQ-PCA 在 4 组 UCI 数据集上取得最小的重建误差, 也展现出了优越的性能. 在  $p$  取不同值的情况下, 对应  $L_{2,p}$ -PCA 和 RPCA-PW 算法的识别正确率也有所区别, 其中当  $p = 0.5$  的时候,  $L_{2,p}$ -PCA 和 RPCA-PW 在总体上拥有更好的降维性能, 这是因为与  $p = 1$  或 2 相比,  $p = 0.5$  可以进一步削弱异常样本点的影响, 同时提高主成分邻域中的样本点在求解最优解时的影响. 因此在 UCI 真实数据集上的实验表明 RPCA-PW 在拥有更高精度的基础上, 还通常具有更小的重建误差.

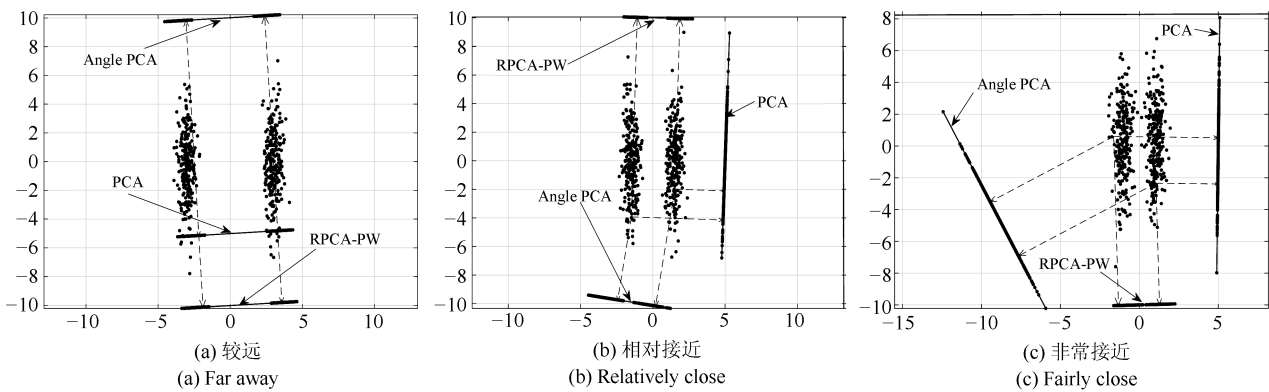


图 3 三种算法在双高斯人工数据集上的投影结果

Fig. 3 Projection result of three algorithms on two-Gaussian artificial dataset

表 2 UCI 数据集上各算法的平均分类正确率 (%)

Table 2 Average classification accuracy of each algorithm on UCI data sets (%)

数据集	Australian	Cars	Cleve	Solar	Zoo	Control	Crx	Glass	Iris	Wine	
PCA	57.32	69.27	52.16	64.11	<b>96.00</b>	91.83	54.93	74.35	96.00	73.04	
PCA- $L_{21}$	62.41	69.17	52.74	64.44	95.60	92.50	53.39	75.33	<b>97.33</b>	74.13	
RPCA-OM	63.04	69.93	52.16	65.27	95.00	92.83	56.96	74.35	96.00	74.15	
Angle PCA	60.39	69.43	52.16	65.60	94.00	84.28	59.59	73.94	95.53	74.15	
MaxEnt-PCA	60.14	69.17	52.45	64.76	95.10	92.15	59.32	74.42	95.33	73.59	
HQ-PCA	60.77	69.27	52.46	65.46	95.80	91.50	60.87	76.16	92.00	<b>75.55</b>	
$L_{2,p}$ -PCA	$p = 0.5$	60.68	69.60	52.84	65.42	93.40	92.10	60.29	73.56	96.47	73.59
	$p = 1$	62.39	69.58	52.16	64.51	94.90	92.50	59.43	73.73	96.67	73.59
	$p = 1.5$	62.81	69.43	52.16	65.28	95.00	92.50	58.20	75.57	96.33	73.04
	$p = 2$	62.32	69.43	52.16	64.11	<b>96.00</b>	91.83	54.93	75.82	96.00	73.59
RPCA-PW	$p = 0.5$	<b>63.77</b>	<b>71.21</b>	<b>53.76</b>	<b>66.85</b>	95.00	92.50	58.84	74.35	95.33	74.15
	$p = 1$	63.04	69.43	52.49	65.94	95.00	91.67	60.14	73.94	96.00	74.15
	$p = 1.5$	62.17	69.43	52.16	64.72	95.00	<b>97.83</b>	<b>61.16</b>	<b>76.28</b>	96.67	74.15
	$p = 2$	62.32	69.43	52.16	64.11	<b>96.00</b>	92.33	54.93	75.82	96.00	73.59

表 3 UCI 数据集上各算法的重建误差  
Table 3 Reconstruction error of each algorithm on UCI data sets

数据集	Australian	Cars	Cleve	Solar	Zoo	Control	Crx	Glass	Iris	Wine	
PCA	197.53	1 979.82	5.47	2.97	1.43	115.06	6 265.42	37.91	0.67	24.49	
PCA-L21	197.21	1 979.16	6.90	2.90	1.56	132.84	6 545.72	38.59	1.56	34.44	
RPCA-OM	193.23	1 977.98	5.61	2.65	1.51	108.13	<b>5 684.38</b>	38.64	0.64	24.85	
Angle PCA	197.96	1 981.34	13.27	2.95	2.04	123.62	6 583.82	38.42	0.59	25.85	
MaxEnt-PCA	203.82	1 976.70	6.24	2.61	1.44	113.01	6 805.98	38.12	0.58	25.01	
HQ-PCA	195.79	<b>1 929.26</b>	5.92	<b>2.37</b>	<b>1.39</b>	131.35	6 987.68	<b>36.38</b>	4.31	25.85	
$L_{2,p}$ -PCA	$p = 0.5$	193.29	1 978.76	9.22	2.93	2.05	106.69	6 601.74	38.33	1.04	26.14
	$p = 1$	193.71	1 978.33	5.60	2.75	1.72	108.06	6 588.17	38.37	0.84	25.44
	$p = 1.5$	193.50	1 977.93	5.50	2.69	1.44	109.69	6 807.38	37.82	0.72	25.05
	$p = 2$	193.53	1 977.82	5.47	2.97	1.43	115.06	5 865.42	37.91	0.67	24.49
RPCA-PW	$p = 0.5$	<b>192.93</b>	1 976.94	5.75	2.95	1.57	<b>106.11</b>	6 467.83	38.43	<b>0.52</b>	<b>24.12</b>
	$p = 1$	193.27	1 979.96	6.84	2.84	1.83	110.46	6 491.19	38.17	0.70	24.36
	$p = 1.5$	193.36	1 977.82	<b>5.06</b>	2.83	1.46	110.45	6 073.11	37.91	0.71	25.26
	$p = 2$	193.53	1 977.82	5.47	2.97	1.43	115.06	5 865.42	37.91	0.67	24.49

### 5.3 人脸识别与重建实验

#### 5.3.1 人脸数据库

本节中运用两个人脸数据库的人脸图像进行实验, 其中 Extended Yale B 人脸数据库由 2 414 个正面人脸图像组成, 这些正面人脸图像是从 38 个具有不同光照的个体中采样的, 其中每人有 65 张图像. AR 数据库包含超过 4 000 张 126 人 (70 名男性和 56 名女性) 的面部图像, 包括不同的面部表情、光照强度与遮挡范围.

#### 5.3.2 人脸识别实验

实验中, 将每张图像的大小调整为  $32 \times 32$  像素, 这样每幅图像就被转化为一个 1 024 维的向量, 并且让所有数据被归一化至均值为 0. 为消除随机影响, 实验方法选择十折交叉验证法, 并在训练集图像中添加遮挡噪声块、高斯噪声和椒盐噪声三种不同类型的噪声以测试各算法的鲁棒性, 其中噪声块的位置是随机的, 遮挡噪声像素与图像像素数的比率为 0.05~0.10. 图 4(a)~(b) 分别列出了两个数据库的原始图像以及添加噪声后对应的图像.

比较结果如图 5 所示, 不同的噪声类型添加到同一数据库上表现出各不相同的实验结果. 从图中不难发现本文算法对不同维度降维后的分类性能总体上要优于其他的鲁棒 PCA 算法. 此外, 从图中可以得到两个结论: 1) 特征空间中的分类结果要优于原始空间中的分类结果, 特别是当特征空间的维数增加时. 这是因为噪声可能发生在原始空间中, 而降维方法可以找到具有大部分判别信息的主成分空间

并抛弃补空间中的冗余信息, 从而更准确和快速地进行分类. 2) 在不同的情况下, RPCA-PW 在总体上要优于其他方法. 因为 RPCA-PW 能够有效地排除少数异常样本的不良影响, 最终收敛到正确结果并且保持很高的识别精度, 所以当图像保留足够的空间信息时, 它具有更好的性能并且取得更高的平均识别率.

#### 5.3.3 人脸重构实验

为了直观观察使用特征空间对遮盖噪声图进行重构之后的效果图, 本节在 Extended Yale B 和 AR 两个人脸数据库中进行人脸重构实验. 为了方便观察, 将每张图片的大小调整为  $64 \times 64$  像素, 并在训练集图像中添加遮挡噪声块以测试各算法的鲁棒性, 遮挡噪声块的位置是随机的, 噪声像素与图像像素数的比率为 0.05~0.10. 实验中首先使用各算法提取出前 30 个特征值所对应的特征向量组成的特征空间, 再使用这些特征对遮盖噪声图像进行重构. 图 6 是各算法的人脸重构效果图.

从图中可以看出, 经典 PCA 算法对原始图片的重构效果最差, 因为基于均方误差的 PCA 算法, 其重构性能易受噪声点影响, 导致质量差. PCA-L21、Angle PCA 和 RPCA-OM 三种算法能够较为清晰地还原人脸的轮廓, 但是对于噪声块遮挡部位的还原效果并不理想. 而 HQ-PCA 与基于  $L_{2,p}$ -norm 的 RPCA-PW 和  $L_{2,p}$ -PCA 对于遮盖部分的还原更加清晰, 表现出对噪声有更好的鲁棒性. 而本文算法在降维过程中对于图片全局结构特征的提取

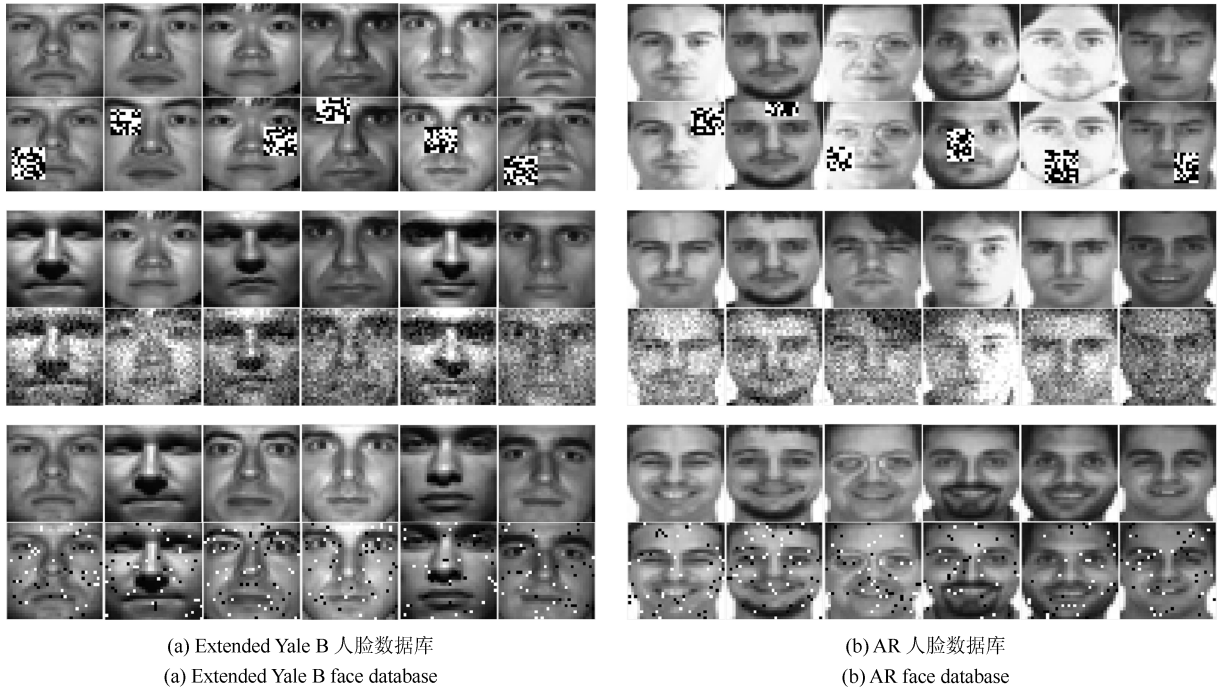


图 4 Extended Yale B 和 AR 人脸数据库原始图像与加入三种不同噪声后对应的图像 (从上至下分别为黑白噪声块、高斯噪声和椒盐噪声)

Fig. 4 Original images in Extended Yale B and AR face database and the corresponding images with three different noise types (top to bottom are noise block, Gaussian noise and salt-and-pepper noise respectively)

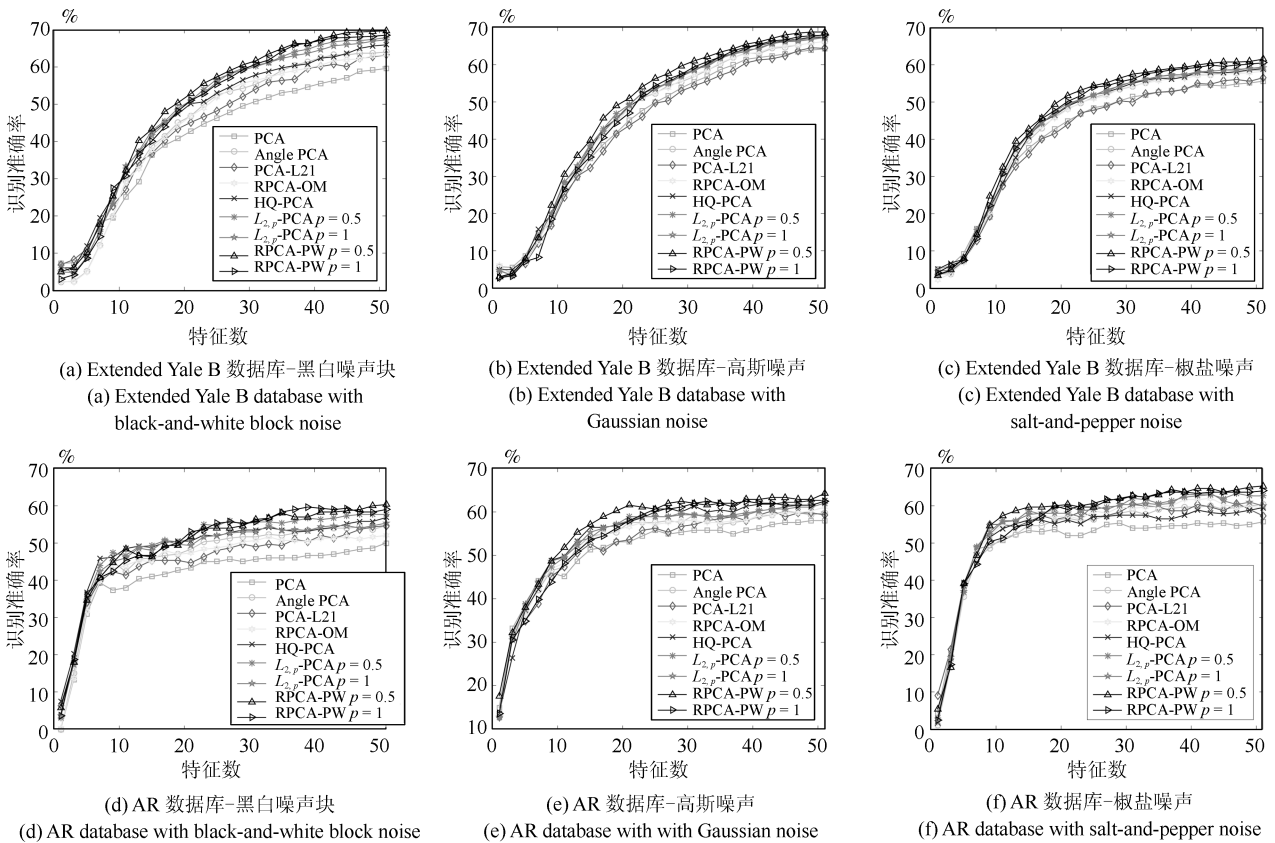
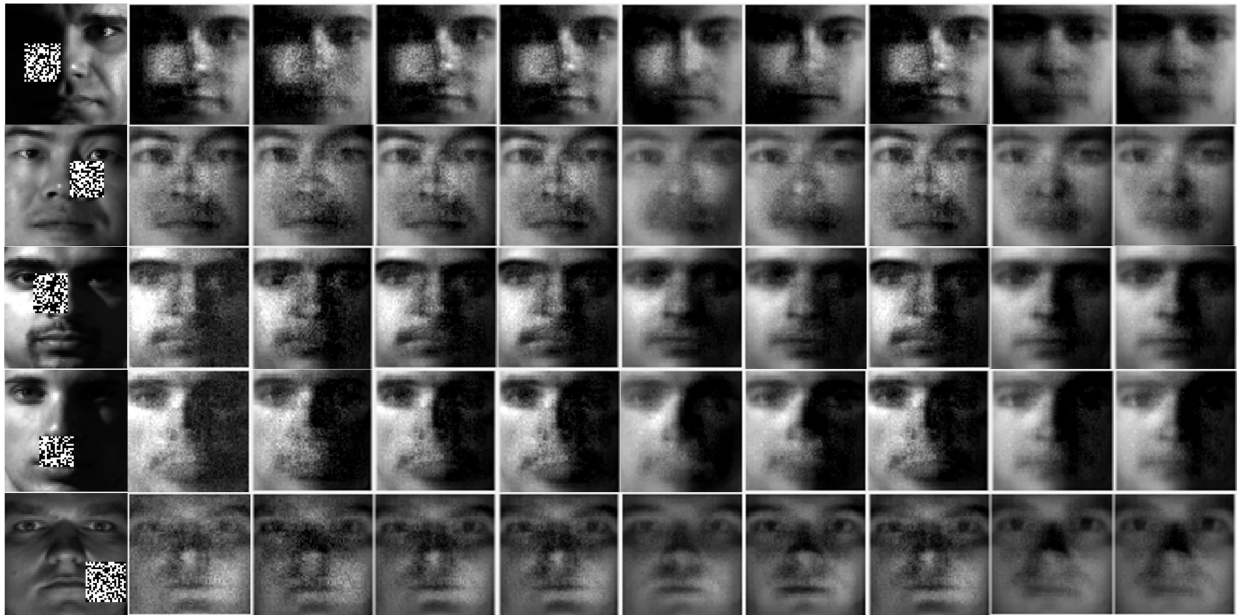


图 5 不同维度下的各算法平均识别准确率

Fig. 5 Recognition accuracy with different reduced dimensions



(a) Extended Yale B 数据库重构结果  
 (a) Reconstruction results of Extended Yale B



(b) AR 数据库重构结果  
 (b) Reconstruction results of AR

图 6 人脸重构效果图, 每一列从左到右依次是原图、PCA、PCA-L21、RPCA-OM、Angle PCA、HQ-PCA、 $L_{2,p}$ -PCA  $p = 0.5$ 、 $L_{2,p}$ -PCA  $p = 1$ 、RPCA-PW  $p = 0.5$ 、RPCA-PW  $p = 1$

Fig. 6 Face reconstruction pictures, each column represents original image, PCA, PCA-L21, RPCA-OM, Angle PCA, HQ-PCA,  $L_{2,p}$ -PCA  $p = 0.5$ ,  $L_{2,p}$ -PCA  $p = 1$ , RPCA-PW  $p = 0.5$  and RPCA-PW  $p = 1$  from left to right

能力强, 因此在图像重构中能够获得更为清晰的人脸轮廓, 尤其是对于遮盖部分的还原几乎不受影响。

### 5.3.4 算法时间比较结果

为了比较本文提出的 RPCA-PW 算法和其他算法的运行效率, 这里统计了各算法在不同数据集

上的平均运行时间, 其中包括 10 个 UCI 数据集和 2 个人脸图像数据集. 实验结果如图 7 所示。

从实验结果可以看出, PCA 由于只需要进行一次特征分解, 因此运行时间远远低于其他的 PCA 算法; PCA-L21 采用的是非贪婪优化算法, 整体上优化了迭代过程从而减少了运行时间; 相比之下,

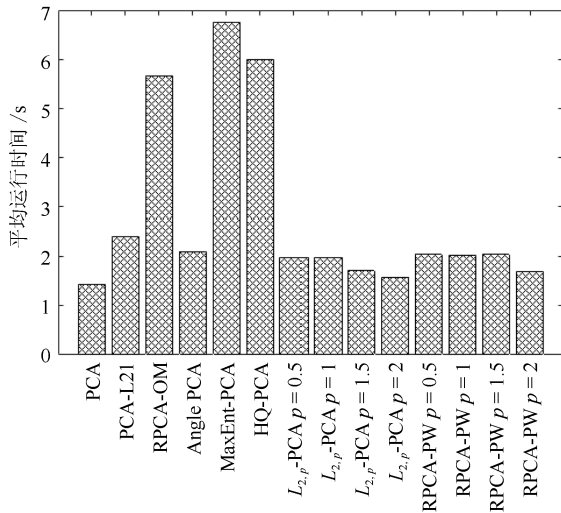
RPCA-PW 虽然需要更长的平均迭代时间,但是算法拥有更好的分类性能和鲁棒性,而且从总体而言,平均运行时间也要低于各算法的平均值.

### 5.3.5 算法收敛性实验

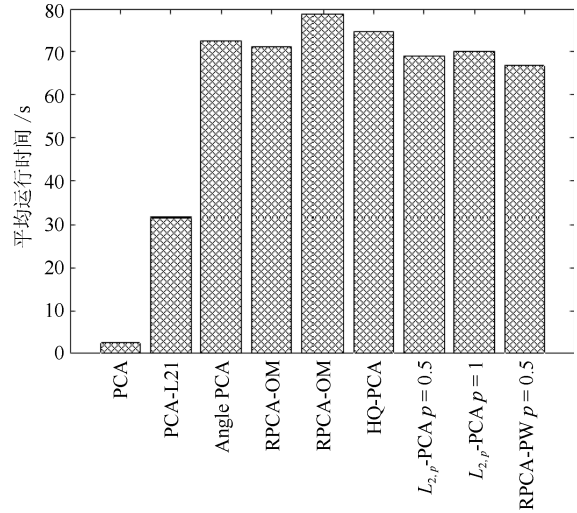
为了测试 RPCA-PW 算法的收敛性能,本文进行如下的实验以验证算法在不同类型标准数据集上

的收敛性能.这里选取了 6 个标准数据集进行相关的收敛性实验(包括 4 个不同类型的 UCI 数据集与 2 个人脸图像数据集).所提算法在标准数据集上的收敛曲线如图 8 所示.

从图 8 的实验结果可以看出,相比较于只需要进行一次特征分解即可得到结果的传统 PCA 算法, RPCA-PW 虽然需要更多的迭代次数,但是通常能



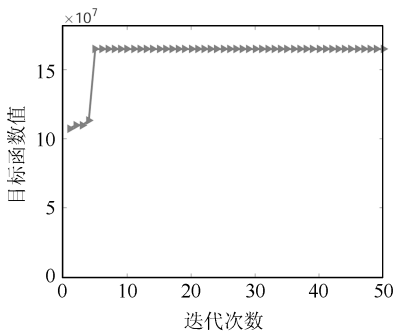
(a) 10 个 UCI 数据集上的平均运行时间  
(a) Average run time on 10 UCI data sets



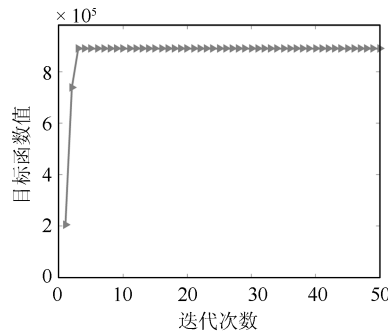
(b) 2 个人脸数据库上的平均运行时间  
(b) Average run time on two face image databases

图 7 各算法对不同数据集的平均时间比较结果

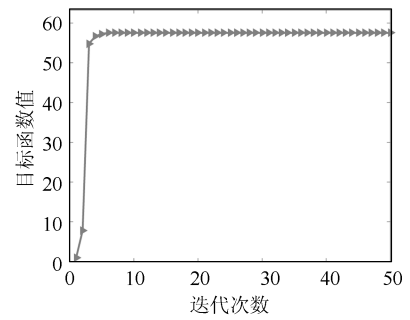
Fig. 7 Comparison of average iteration time for different data sets by different algorithms



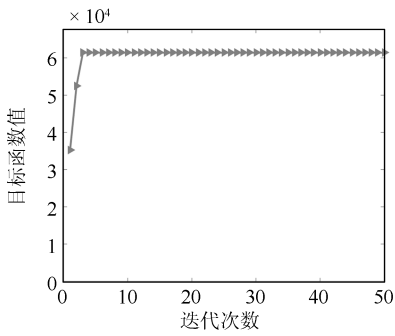
(a) Cleve



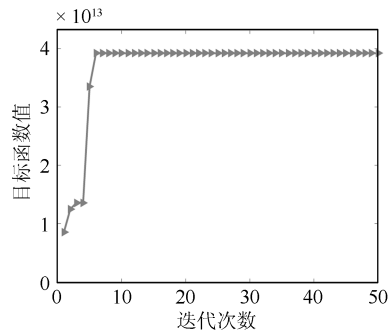
(b) Zoo



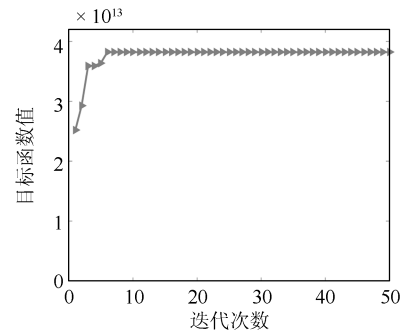
(c) Control



(d) Iris



(e) Extended Yale B



(f) AR

图 8 RPCA-PW 收敛曲线

Fig. 8 Convergence curves of RPCA-PW



够保证在迭代十次之内收敛, 而且在迭代的过程中, 稳定性也没有受到破坏. 因此所提算法能够在仅仅增加少量计算量的前提下, 大大提高算法的鲁棒性和泛化性能.

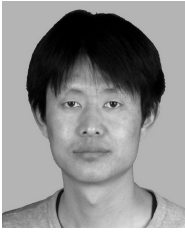
## 6 结论

本文提出了一种新的 PCA 模型, 称为鲁棒自适应概率加权 PCA. RPCA-PW 较经典 PCA 算法在鲁棒性上有了明显的改进, 具体表现在采用  $L_{2,p}$  模降低离群点对于模型的影响, 并且基于投影空间中数据的结构信息与重建误差之间的关系, 在提取主成分的过程中加强对识别关键的样本点的影响, 削弱那些与识别过程关系不大或者冗余大的样本点来提高精度. 从而在计算过程中能够自动识别异常点样本, 有效地降低了样本中离群点的干扰, 这一点在实际应用中也有着一定的意义. 此外, 所提出的模型可作为广义公式, 几种现有的算法都能作为其特例. 在人工数据集、UCI 数据集和人脸数据库上与其他 PCA 方法进行了对比实验, 结果表明本文提出的模型拥有更好的识别精度, 并且对噪声有显著高的鲁棒性.

## References

- Jolliffe I. *Principal Component Analysis*. New York: Springer-Verlag, 1986. 115–128
- Gao Quan-Xue, Pan Quan, Liang Yan, Zhang Hong-Cai, Cheng Yong-Mei. Face recognition based on expressive features. *Acta Automatica Sinica*, 2006, **32**(3): 386–392  
(高全学, 潘泉, 梁彦, 张洪才, 程咏梅. 基于描述特征的人脸识别研究. *自动化学报*, 2006, **32**(3): 386–392)
- Zhang Xian-Peng, Chen Fan, He Hong-Jie. Shadow detection in high resolution remote sensing images using multiple features. *Acta Automatica Sinica*, 2016, **42**(2): 290–298  
(张先鹏, 陈帆, 红杰. 结合多种特征的高分辨率遥感影像阴影检测. *自动化学报*, 2016, **42**(2): 290–298)
- Gao Q X, Ma L, Liu Y, Gao X B, Nie F P. Angle 2DPCA: A new formulation for 2DPCA. *IEEE Transactions on Cybernetics*, 2018, **48**(5): 1672–1678
- Koren Y, Carmel L. Robust linear dimensionality reduction. *IEEE Transactions on Visualization & Computer Graphics*, 2004, **10**(4): 459–470
- Schölkopf B, Smola A, Müller K R. *Nonlinear component analysis as a kernel eigenvalue problem*. MIT Press, 1998, **10**(5): 1299–1319
- Li Chun-Na, Chen Wei-Jie, Shao Yuan-Hai. Robust sparse  $L_p$ -norm principal component analysis. *Acta Automatica Sinica*, 2017, **43**(1): 142–151  
(李春娜, 陈伟杰, 邵元海. 鲁棒的稀疏  $L_p$  模主成分分析. *自动化学报*, 2017, **43**(1): 142–151)
- Xu L, Yuille A L. Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, 1995, **6**(1): 131–143
- Ke Q, Kanade T. Robust  $L_1$ -norm factorization in the presence of outliers and missing data by alternative convex Programming. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA: IEEE, 2005, **1**: 739–746
- Kwak N. Principal component analysis based on L1-norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(9): 1672–1680
- Nie F, Huang H, Ding C, Luo D, Wang H. Robust principal component analysis with non-greedy L1-norm maximization. In: *Proceedings of IJCAI International Joint Conference on Artificial Intelligence*. Barcelona, Catalonia, Spain: AAAI, 2011, **22**(1): 1433
- Ng A Y. Feature selection, L1 vs. L2 regularization, and rotational invariance. In: *Proceedings of the 21st international conference on Machine learning*. Banff, Canada: ACM, 2004, 78
- He R, Hu B, et al. Principal component analysis based on non-parametric maximum entropy. *Neurocomputing*, 2010, **73**(10): 1840–1852
- He R, Hu B, et al. Robust principal component analysis based on maximum correntropy criterion. *IEEE Transactions on Image Processing*, 2011, **20**(6): 1485–1494
- He R, Tan T, Wang L. Robust recovery of corrupted low-rank matrix by implicit regularizers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(4): 770–783
- Ding C, Zhou D, He X F, Zha H Y. R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. In: *Proceedings of the 23rd International Conference on Machine Learning*. Pittsburgh, USA: ACM, 2006. 281–288
- Nie F, Yuan J, Huang H. Optimal mean robust principal component analysis. In: *Proceedings of 31st International Conference on Machine Learning*. Beijing, China: ACM 2014. 2755–2763
- Nie, F, Huang, H. Non-greedy L21-norm maximization for principal component analysis. arXiv 2016, arXiv: 1603.08293
- Wang Q, Gao Q, Gao X, Nie F.  $L_{2,p}$ -norm based PCA for image recognition. *IEEE Transactions on Image Processing*, 2018, **27**(3): 1336–1346
- Wang Q, Gao Q, Gao X, Nie F. Angle principal component analysis. In: *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia: AAAI, 2017. 2936–2942
- Liu Y, Gao Q, Miao S, et al. A non-greedy algorithm for L1-norm LDA. *IEEE Transactions on Image Processing*, 2016, 1–1
- Guo Y F, Li S J, Yang J Y, Shu T T, Wu L D. A generalized Foley-Sammon transform based on generalized fisher discriminant criterion and its application to face recognition. *Pattern Recognition Letters*, 2003, **24**(1-3): 147–158

23 Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 1967, **13**(1): 21–27



**高云龙** 厦门大学副教授. 2005 年获得兰州大学计算机专业硕士学位. 2011 年获得西安交通大学控制科学与工程专业博士学位. 主要研究方向为模式识别, 时间序列分析. 本文通信作者.

E-mail: gaoyl@xmu.edu.cn  
(**GAO Yun-Long** Associate professor at Xiamen University. He received

the master degree in computer science from Lanzhou University in 2005, and the Ph. D. degree in control science and engineering from Xian Jiaotong University in 2011. His research interest covers pattern recognition and time series analysis. Corresponding author of this paper.)



**罗斯哲** 厦门大学硕士研究生. 主要研究方向为机器学习与维数约简.

E-mail: sizheluo@foxmail.com  
(**LUO Si-Zhe** Master student at Xiamen University. His research interest covers machine learning and dimensionality reduction.)



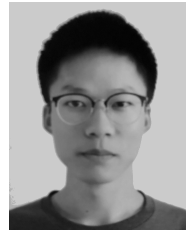
**潘金艳** 集美大学副教授. 主要研究方向为最优化方法和数据挖掘.

E-mail: jypan@jmu.edu.cn  
(**PAN Jin-Yan** Associate professor at Jimei University. Her research interest covers optimization method and data mining.)



**陈柏华** 厦门大学博士研究生. 主要研究方向为机器学习和数据降维.

E-mail: chenbaihua001@163.com  
(**CHEN Bai-Hua** Ph. D. candidate at Xiamen University. His research interest covers machine learning and data dimensionality reduction.)



**张逸松** 厦门大学硕士研究生. 主要研究方向为机器学习和数据聚类.

E-mail: yisongzhang@foxmail.com  
(**ZHANG Yi-Song** Master student at Xiamen University. His research interest covers machine learning and data cluster analysis.)