

Truncated Robust Principle Component Analysis With A General Optimization Framework

Feiping Nie, Danyang Wu, Rong Wang, and Xuelong Li

摘要

最近, 为了提高主成分分析 (PCA) 的鲁棒性, 已经有好几种鲁棒主成分分析 (RPCA) 的模型被提出。但是, 对异常值的鲁棒性会影响正确样本的辨别这个重要的问题仍未得到解决。为了解决这个问题, 作者提出一种截断的鲁棒性主成分分析 (T-RPCA) 模型^[1], 这个模型能够很好区分处理正常样本和异常值, 同时该模型执行了隐式截断加权学习的方案, 该方案对鲁棒性学习更加合理。此外, 本文提出了一个重加权 (RW) 优化框架来解决一般问题并在其上面推广两个子框架。其中就包含对 T-RPCA 目标问题的一般截断损失优化问题并提供了严格的理论保证。实验研究证明, 本文提出的 T-RPCA 模型在分类和重建任务上优于先前的 RPCA 模型。

关键词: 鲁棒主成分分析 (RPCA); 无监督降维; 截断损失; 非凸优化

1 引言

主成分分析 (PCA)^[2] 是一种经典无监督降维模型。由于其高性能和跟可解释性, PCA 已经广泛应用在各种不同领域当中, 如人脸识别, 子空间学习, 图像去噪等。PCA 的主要目标是找到最小重构误差的投影矩阵。与严格求解约束秩的 PCA 不同, 一些相关的工作提出在核范数松弛时执行秩约束。本文主要研究 PCA 模型和他的扩展。

由于 PCA 的重构误差时基于平方的 l_2 范数, PCA 不能很好处理数据的异常值, 例如音频信号的噪声带和图像的遮挡。为了减轻异常值的影响, 一些鲁棒的主成分分析 (RPCA) 模型已经被提出。首先被提出的时基于 l_1 范数的模型, 包括直接最小化 l_1 范数重构误差的 L1-PCA^[3] 和最大化通过 l_1 范数度量计算的方差 PCA-L1^[4]。然而基于 l_1 范数的模型对旋转是不变的并且它们不能直接最小化每个样本的欧氏距离。为了解决这些问题, 最小化 $l_{2,1}$ 范数的 R1-PCA^[5] 模型被提出了, 它通过在数据空间施加 l_1 范数和在特征空间上施加 l_2 范数来重构误差。基于 R1-PCA, RPCA-OM^[6] 被提出用于寻找 R1-PCA 的最佳数据均值。最近, 由于 $l_{2,1}$ 范数不能有效地解决严重异常值, 提出了 2p-RPCA^[7] 以最小化 $l_{2,p}$ 范数的重构误差。

在先前的工作中, L1-PCA, PCA-L1, R1-PCA 和 RPCA-OM 的鲁棒性是不足以处理比较严重的异常值。尽管 2p-RPCA 对异常值表现良好, 但在处理异常值的时候, 它会影响正常样本的辨别。例如, 当 p 被设为一个很小的值的时候, 模型可以有效地解决异常值, 但 $l_{2,p}$ 范数距离度量对正确的分辨力很低。为了缓解这个问题, 本文提出了一种截断鲁棒的主成分分析 (T-RPCA) 模型用 $l_{2,1}$ 范数处理正常样本并且忽略异常值。此外, 本文提出了一个重加权 (RW) 优化框架来解决一般问题并在其上面推广两个子框架。其中就包含对 T-RPCA 目标问题的一般截断损失优化问题并提供了严格的理论保证。实验研究证明, 本文提出的 T-RPCA 模型在分类和重建任务上优于先前的 RPCA 模型。本文主要贡献如下:

- (1) 提出一个截断鲁棒主成分分析 (T-RPCA) 模型来分别处理正常样本和异常值。
- (2) 提出一个重加权 (RW) 优化框架来解决一般性问题, 并在此基础上提供两个有用的子框架。
- (3) 为 T-RPCA 模型, RW 框架和两个子框架提供严格的理论保证, 实验结果证明了 T-RPCA 的优越性。

2 相关工作

2.1 PCA^[2]

给定数据矩阵 $\mathbf{X} \in \mathbb{R}^{d \times n}$, $\mathbf{X}_0 \in \mathbb{R}^{d \times n}$ 表示数据均值矩阵, $\mathbf{X}_0 = \mathbf{X} - \frac{1}{n} \mathbf{X} \mathbf{1}_n \mathbf{1}_n^T$, 其中 $\mathbf{1}_n \in \mathbb{R}^{n \times 1}$ 是元素全为 1 的向量, PCA 目标是解决以下问题:

$$\min_{\mathbf{Z}} \|\mathbf{X}_0 - \mathbf{Z}\|_F^2 \quad s.t. \quad rank(\mathbf{Z}) = k. \quad (1)$$

假设 $\mathbf{F} \in \mathbb{R}^{d \times k}$ 满足正交约束, 即 $\mathbf{F}^T \mathbf{F} = \mathbf{I}_k$, 其中 $\mathbf{I}_k \in \mathbb{R}^{k \times k}$ 是一个单位矩阵, 通过满秩分解 $\mathbf{Z} = \mathbf{F} \mathbf{R}^T$, 其中 $\mathbf{R} \in \mathbb{R}^{n \times k}$ 。所以, 问题 (1) 能够转化为以下问题

$$\min_{\mathbf{F}, \mathbf{R}} \|\mathbf{X}_0 - \mathbf{F} \mathbf{R}^T\|_F^2 \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (2)$$

对 (2) 式中 \mathbf{R} 取导数并赋值为 0, 得到 $\mathbf{R} = \mathbf{X}_0^T \mathbf{F}$ 。故问题 (2) 可以写成

$$\min_{\mathbf{F}} \|\mathbf{X}_0 - \mathbf{F} \mathbf{F}^T \mathbf{X}_0\|_F^2 \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k, \quad (3)$$

这是大多数相关研究中通用 PCA 公式, 根据 (3), 我们能知道 PCA 主要目标式找到最小重构误差的投影矩阵 \mathbf{F} 。

2.2 RPCA

L1-PCA^[3]: 为了提高 PCA 的鲁棒性, L1-PCA 模型使用 l_1 范数去解决以下问题:

$$\min_{\mathbf{F}} \|\mathbf{X}_0 - \mathbf{F} \mathbf{F}^T \mathbf{X}_0\|_1, \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (4)$$

PCA-L1^[4]: 因为问题 (4) 的计算成本昂贵, 把 PCA 转化成以下问题:

$$\begin{aligned} \min_{\mathbf{F}} \|\mathbf{X}_0 - \mathbf{F} \mathbf{F}^T \mathbf{X}_0\|_F^2 &\Rightarrow \max_{\mathbf{F}} Tr(\mathbf{F}^T \mathbf{X}_0 \mathbf{X}_0^T \mathbf{F}) \\ &\Rightarrow \max_{\mathbf{F}} \|\mathbf{F}^T \mathbf{X}_0\|_F^2, \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \end{aligned} \quad (5)$$

该作者提出可以通过解决以下问题使用 PCA-L1 来简化问题 (4)

$$\max_{\mathbf{F}} \|\mathbf{F}^T \mathbf{X}_0\|_1, \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (6)$$

R1-PCA^[5]: 由于问题 (6) 并不能最好地最小化数据的重构误差, 而且 PCA-L1 和 L1-PCA 对旋转都是不变的。为了解决这些问题, 提出了 R1-PCA 模型来优化以下问题:

$$\min_{\mathbf{F}} \|\mathbf{X}_0 - \mathbf{F} \mathbf{F}^T \mathbf{X}_0\|_{2,1}, \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (7)$$

2p-RPCA^[7]: 由于 $l_{2,1}$ 范数对严重异常值的鲁棒性不够, 而且 $l_{2,1}$ 范数是 $l_{2,p}$ 范数的特例, 所以将问题 (7) 转化为 $l_{2,p}$ 范数的模型。

$$\min_{\mathbf{F}} \|\mathbf{X}_0 - \mathbf{F} \mathbf{F}^T \mathbf{X}_0\|_{2,p}^p, \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (8)$$

RPCA-OM^[6]: 不同于先前 RPCA 的工作, RPCA-OM 关注于 RPCA 模型的最有均值问题。因为由 l_2 范数计算均值可能不是数据的最优均值, 所以把最优均值的计算集成到模型优化中, 以得到最有均

值。 $\mathbf{X}_0 = \mathbf{X} - \mathbf{b}\mathbf{1}_n^T$ ，其中 $\mathbf{b} \in \mathbb{R}^{d \times 1}$ 是未知的最优均值。

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{b}} \quad & \|\mathbf{X} - \mathbf{b}\mathbf{1}_n^T - \mathbf{F}\mathbf{F}^T(\mathbf{X} - \mathbf{b}\mathbf{1}_n^T)\|_F^2 \\ \text{s.t.} \quad & \mathbf{F}^T\mathbf{F} = \mathbf{I}_k. \end{aligned} \quad (9)$$

基于 R1-PCA 模型，RPCA-OM 模型可扩展为以下问题：

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{b}} \quad & \|\mathbf{X} - \mathbf{b}\mathbf{1}_n^T - \mathbf{F}\mathbf{F}^T(\mathbf{X} - \mathbf{b}\mathbf{1}_n^T)\|_{2,1} \\ \text{s.t.} \quad & \mathbf{F}^T\mathbf{F} = \mathbf{I}_k. \end{aligned} \quad (10)$$

3 截断鲁棒的主成分分析 (T-RPCA)

3.1 本文方法概述

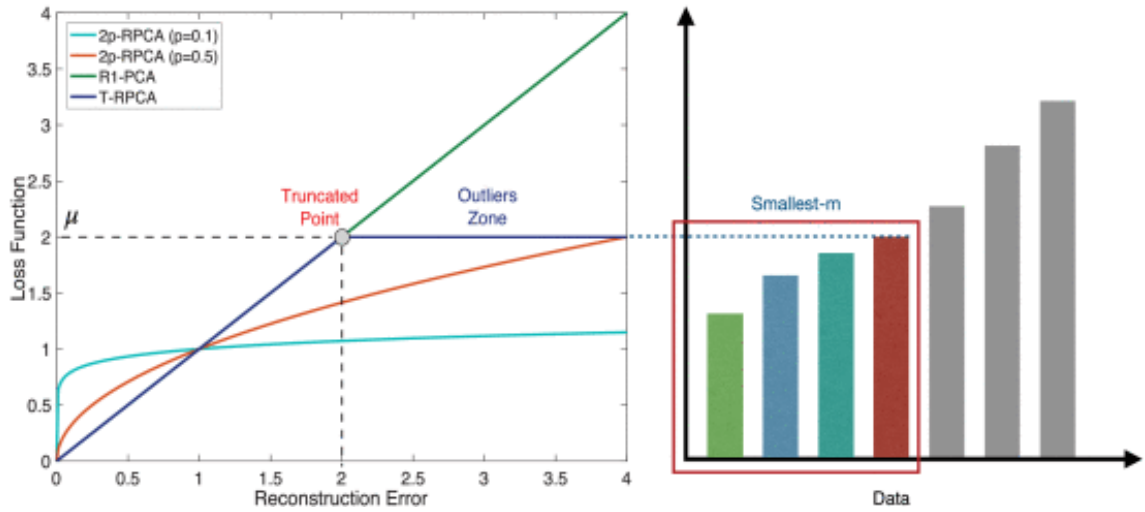


图 1: 截断损失和 m 个最小的损失

在以前的工作中，基于 l_1 范数和 $l_{2,1}$ 范数模型对异常值的鲁棒性是不够的， $l_{2,p}$ 范数当 p 足够小 ($p < 1$) 的时候对异常值有很好的鲁棒性但是会影响对正常样本的辨别力。为了解决这些问题，文中提出了以下两个动机。

(1) 使用 $l_{2,1}$ 范数对正常样本做重构误差，用一个固定的常数来惩罚异常值。

$$\min_{\mathbf{F}, \mathbf{b}} \sum_{i=1}^n \min(\|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2, \mu) \quad \text{s.t.} \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}_k. \quad (11)$$

(2) 使用 $l_{2,1}$ 范数对正常样本做重构误差，忽略异常值的影响。

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{b}} \sum_{i=1}^n \{ \min(\|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2, \mu) \\ - \mu \mathcal{I}(\|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2 \geq \mu) \} \quad \text{s.t.} \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}_k, \end{aligned} \quad (12)$$

其中 μ 是损失上界，是区分正常样本和异常值的界限， $\mathcal{I}(\cdot)$ 是指标函数。问题 (11) 和问题 (12) 有相同的解集，但问题 (12) 是更加合理的。如图 1 所示：

3.2 损失函数定义

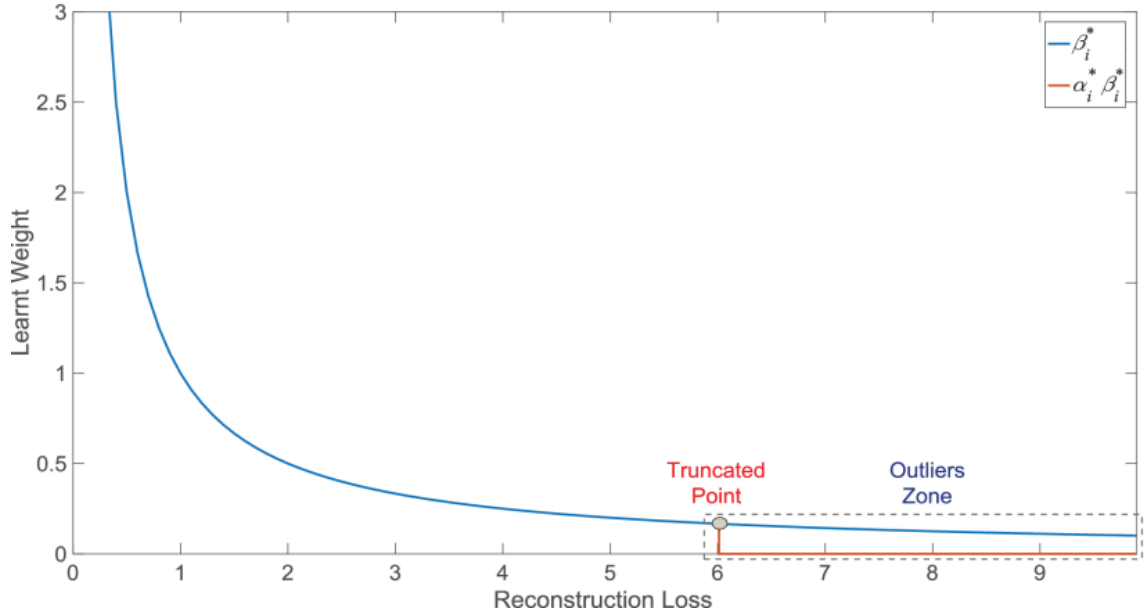


图 2: 截断损失和 m 个最小的损失

对于问题 (12), 存在两个问题: (1) 如何合理地选择 μ 的值是一个困难的事, 因为 $\mu \in (0, \infty)$ 。(2) 如果过多的样本比 μ 大, 模型可能会导致一个平凡解。为了防止这两个问题, 将问题 (12) 转化为问题 (13)。问题 (13) 和问题 (12) 有相同的解

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{b}, \alpha} \sum_{i=1}^n \alpha_i \|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2 \\ s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k, \quad \alpha^T \mathbf{1}_n = m, \quad 0 \leq \alpha \leq 1. \end{aligned} \quad (13)$$

其中 $m = \sum_{i=1}^n \mathcal{I}(\|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2 \geq \mu)$, 这样一来, 把问题 (12) 转化为问题 (13), 把参数 $\mu \in (0, \infty)$ 转化为 $m \in (1, n)$, 使得参数的范围大大缩减, 减少了参数选择的工作量, 也避免了 μ 值选择过小而导致的平凡解。如图 1 和图 2 所示展示了 T-RPCA 的主要思想。

3.3 损失函数的求解

设 $g_i(\mathbf{F}, \mathbf{b}) = \|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2^2$, $h_i(g_i(\mathbf{F}, \mathbf{b})) = \sqrt{g_i(\mathbf{F}, \mathbf{b})} = \|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2$, $e_i(\mathbf{F}, \mathbf{b}) = \|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2$, 所以通过以下式子更新 α_i 。

$$\alpha_i = \begin{cases} 1, & \text{if } e_i(\mathbf{F}, \mathbf{b}) \leq e_{[m]}(\mathbf{F}, \mathbf{b}) \\ 0, & \text{Otherwise,} \end{cases} \quad (14)$$

其中 $e_{[m]}(\mathbf{F}, \mathbf{b})$ 是所有 $e_i(\mathbf{F}, \mathbf{b})$ 中 m 个最小的值。

通过隐式加权的思想把 $l_{2,1}$ 范数转化为 l_2 范数, 以此定义 d_i 如下。

$$d_i = \begin{cases} 0, & \text{if } \alpha_i = 0 \\ \frac{1}{2\|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2}, & \text{Otherwise.} \end{cases} \quad (15)$$

我们通过解决以下子问题来求解 \mathbf{F}, \mathbf{b}

$$\min_{\mathbf{F}, \mathbf{b}} \sum_{i=1}^n d_i \|(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{x}_i - \mathbf{b})\|_2^2 \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (16)$$

为了更好地解释，我们定义了一个对角矩阵 \mathbf{D} ，其中对角项每一个元素代表 d_i 。通过求 \mathbf{b} 的导数并设置为 0，我们得到：

$$(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{b}\mathbf{1}_n^T - \mathbf{X})\mathbf{D}\mathbf{1}_n = \mathbf{0}. \quad (17)$$

\mathbf{F} 的正交补码为 \mathbf{F}^\perp ，因此有如下式子。其中 z_1 和 z_2 代表向量。

$$(\mathbf{b}\mathbf{1}_n^T - \mathbf{X})\mathbf{D}\mathbf{1}_n = \mathbf{F}z_1 + \mathbf{F}^\perp z_2, \quad (18)$$

联合 (17) 和 (18) 得，

$$(\mathbf{I}_d - \mathbf{F}\mathbf{F}^T)(\mathbf{F}z_1 + \mathbf{F}^\perp z_2) = \mathbf{0} \Rightarrow z_2 = \mathbf{0} \quad (19)$$

把 $z_2 = \mathbf{0}$ 带入 (18) 式得，

$$\mathbf{b} = \frac{1}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} (\mathbf{X} \mathbf{D} \mathbf{1}_n + \mathbf{F} z_1). \quad (20)$$

由于 z_1 可以是任何 k 维列向量，而不损失泛化，因此我们将 z_1 设置为 0，并具有：

$$\mathbf{b} = \frac{1}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} \mathbf{X} \mathbf{D} \mathbf{1}_n. \quad (21)$$

最后，把 (21) 代入 (16)，设 $\left(\mathbf{D} \frac{1}{2} - \frac{\mathbf{D} \mathbf{1}_n \mathbf{1}_n^T \mathbf{D}}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} \right) \mathbf{X}^T \mathbf{H} = \mathbf{D} - \frac{\mathbf{D} \mathbf{1}_n \mathbf{1}_n^T \mathbf{D}}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n}$ ，因此问题 (16) 变成了如下形式。

$$\max_{\mathbf{F}} \text{Tr}(\mathbf{F}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{F}) \quad s.t. \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_k. \quad (22)$$

可以看出，问题 (22) 不受任何 z_1 的影响，这也是选择 z_1 等于 0 的一个重要原因。由于 $\mathbf{H} = \left(\mathbf{D} \frac{1}{2} - \frac{\mathbf{D} \mathbf{1}_n \mathbf{1}_n^T \mathbf{D}}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} \right)^T \left(\mathbf{D} \frac{1}{2} - \frac{\mathbf{D} \mathbf{1}_n \mathbf{1}_n^T \mathbf{D}}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} \right)$ ，所以 \mathbf{F} 能够通过 $\left(\mathbf{D} \frac{1}{2} - \frac{\mathbf{D} \mathbf{1}_n \mathbf{1}_n^T \mathbf{D}}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} \right) \mathbf{X}^T$ 做奇异值分解得到。 \mathbf{F} 是对应 k 个最大奇异值得到 k 个有奇异向量。

4 复现细节

4.1 与已有开源代码对比

本文的算法和实验没有参考其他源代码，均为自己复现。

算法的伪代码如 Procedure 1 所示，在实践过程中，我们通过初始化 \mathbf{D} 为一个单位矩阵比初始化 \mathbf{F}, \mathbf{b} 要更加方便。

Procedure 1 The Algorithm for Problem (13).

Input: Initialize \mathbf{F} and \mathbf{b}

Output: \mathbf{F} , \mathbf{b}

while not converge **do**

- 1 Update α by Eq. (35);
- 2 Update \mathbf{D} by Eq. (36);
- 3 Update \mathbf{F} the k right singular vectors corresponding to the

k -largest singular values of $\left(\mathbf{D} \frac{1}{2} - \frac{\mathbf{D} \mathbf{1}_n \mathbf{1}_n^T \mathbf{D}}{\mathbf{1}_n^T \mathbf{D} \mathbf{1}_n} \right) \mathbf{X}^T$;

- 4 Update \mathbf{b} by Eq. (42);
-

4.2 数据集和比较算法

在此部分实验中，选取了 3 个真实的图像数据集，包括 AR,Yale 和 ORL 这三个数据集，选择了经典的 PCA 算法作为对比算法。数据集的描述信息如下：

Dateset	Sample	Dimensions	Classes
AR	2400	2000	120
ORL	400	1024	40
YALE	165	1024	15

4.3 重建实验的可视化

为了展示 T-RPCA 的性能，在 AR 和 ORL 数据集上进行了可视化实验。随机选择 1/5 的图像，并在 AR 数据集上放置 1/4 大小的黑色遮挡，在 ORL 数据集上放置 1/4 的随机遮挡。并将维度缩减到 20，在实验中，我们主要观察遮挡图像的重建性能。实验结果如图 3和 4所示。其中图片的第一行为原始图像，第二行为加入遮挡后的图像，第三行为经 PCA 处理重建后的图像，第四行为经 T-RPCA 处理重建后的图像。我们可以看到，PCA 并不能有效地重建遮挡图像,而对于本文所提的 T-RPCA 模型能够重建出比较干净的图像。这证明了 T-RPCA 能够通过截断损失分别处理干净图像和遮挡图像，这表明了 T-RPCA 模型对干净图像的具有很好的辨别力，对遮挡图像有很强的鲁棒性。而且，T-RPCA 能有效重建遮挡图像，这表明了它对遮挡图像的去噪能力。



图 3: 在 AR 数据集上具有遮挡的重建图像



图 4: 在 ORL 数据集上具有遮挡的重建图像

4.4 重建误差评估

在本部分中，在 AR, ORL 和 YALE 真实数据集上使用 PCA 和 T-RPCA 进行了重建实验。降维尺寸分别为 10, 20, 30, 40, 50。并分别计算了干净图像和遮挡图像的重建误差。因为对干净图像的评估侧重于投影 F 的鲁棒性，而对遮挡图像的评估不仅关注投影 F 的健壮性，还关注遮挡图像的去噪能力。图 5-7 显示了干净样本的重建误差，可以看出，当降维维度增加时，重建误差减少，这表明当降维增大时，投影矩阵 F 的能力提高。此外，T-RPCA 在三个数据集上，所有维度中以最低的重建误差重建干净的图像，这证明了它对遮挡的鲁棒性和有效性。

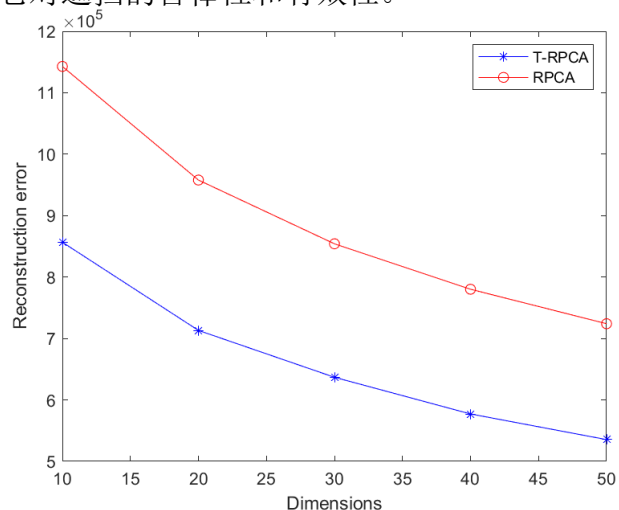


图 5: 在 AR 数据集上干净图像的重建误差

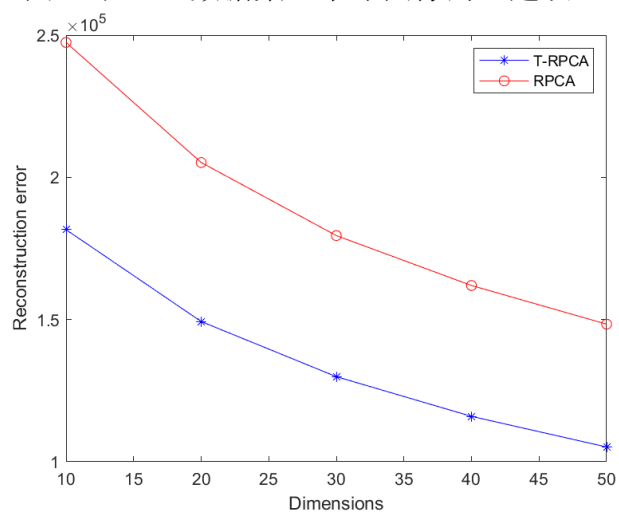


图 6: 在 ORL 数据集上干净图像的重建误差

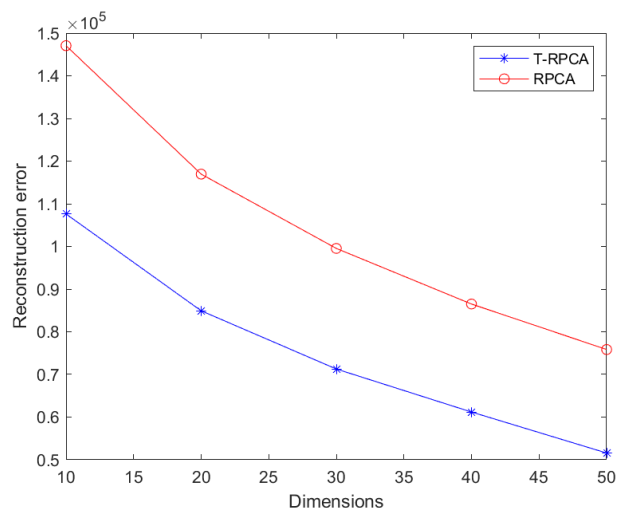


图 7: 在 YALE 数据集上干净图像的重建误差

在图 8-11 中展示了 T-RPCA 对遮挡图像的重建误差，可以明显看出，它在三个数据集上以最低的重建误差重建了所有维度的遮挡图像。这证明了 T-RPCA 的鲁棒性和有效性。说明 TRPCA 对遮挡图像有更好的去噪效果。

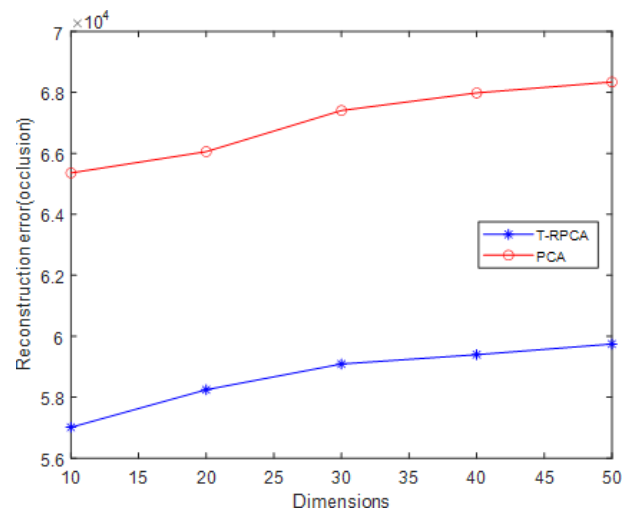


图 8: 在 AR 数据集上遮挡图像的重建误差

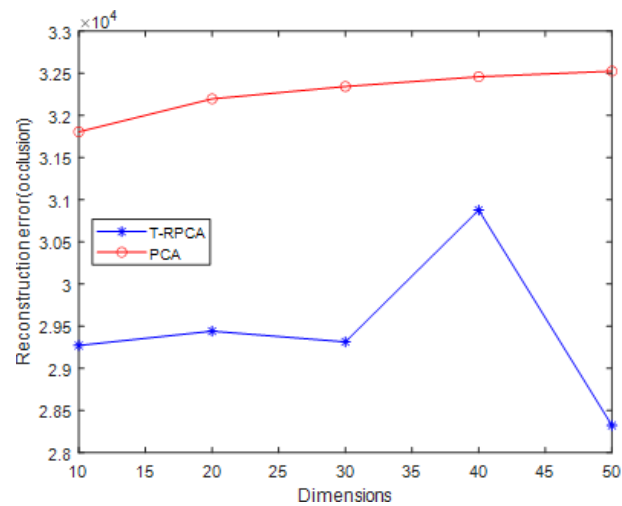


图 9: 在 ORL 数据集上遮挡图像的重建误差

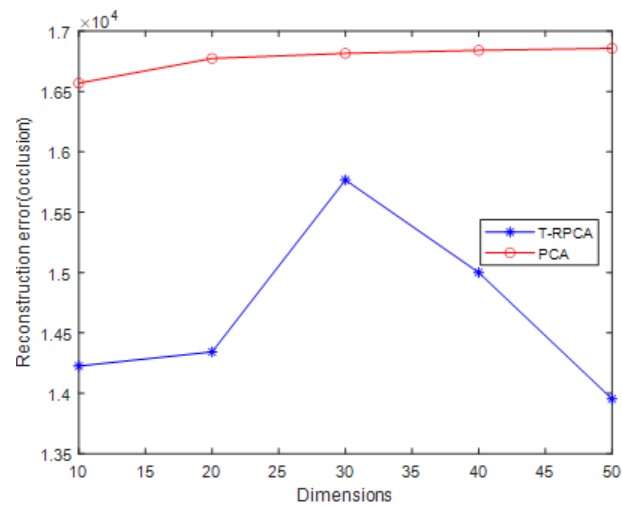


图 10: 在 YALE 数据集上遮挡图像的重建误差

4.5 聚类效果评估

在这个实验中，在 AR,YALE 和 ORL 数据集上使用 PCA 和 T-RPCA 进行降维，然后对降维后的数据用 k-means 进行聚类实验。将缩减的维度设置为 10, 20, 30, 40, 50。并记录最后的聚类精度。聚类实验的结果在图 12-14 中展示。我们可以发现，在绝大多数的情况下，精度和减少的维数成正比。

对于 T-RPCA 模型，它绝大多数情况下都优于比较模型。

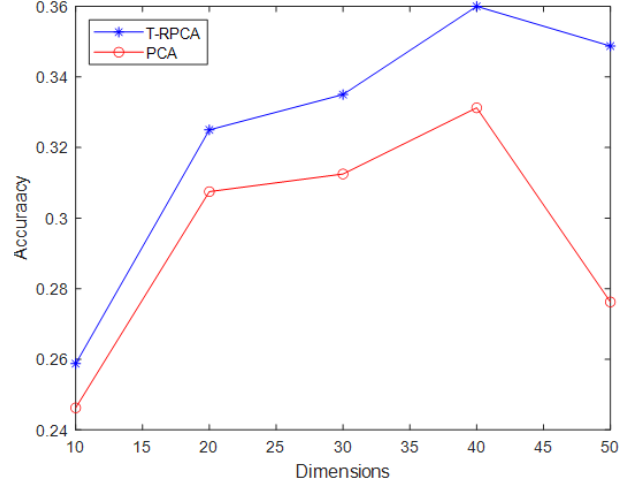


图 11: 在 AR 数据集上降维 +kmeans 的聚类性能

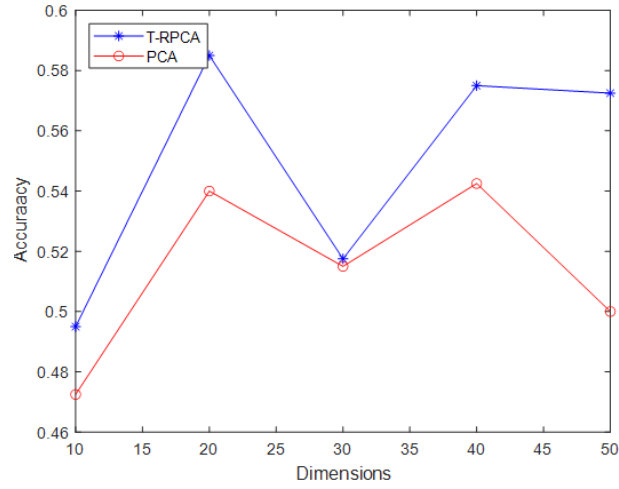


图 12: 在 ORL 数据集上降维 +kmeans 的聚类性能

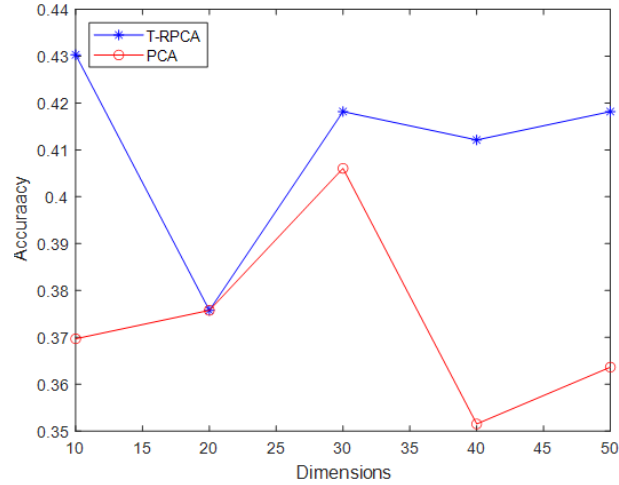


图 13: 在 YALE 数据集上降维 +kmeans 的聚类性能

4.6 创新点

1. 在 NPE 中引入截断的思想

Neighborhood Preserving Embedding(NPE) 领域保持嵌入，这是一个经典的降维算法，主要从数据的表征来反映数据的局部结构，NPE 的目标函数如下：

$$\min \sum_i \left\| \mathbf{y}_i - \sum_j W_{ij} \mathbf{y}_j \right\|^2 \quad \text{where } \mathbf{y}^T = \mathbf{a}^T \mathbf{X} \quad (23)$$

基于本文中截断的思想，可以把它迁移到 NPE 中，NPE 用数据表征误差的平方来进行降维，故若数据表征损失较大，能力较弱时，此样本很可能是数据中的一个异常值。所以，为了提高模型的鲁棒性，提出截断的 NPE, 若样本的表征能力损失高于一个一定值时，把它表征损失置为一个常数。目标函数如下：

$$\min \sum_i \min \left(\left\| \mathbf{y}_i - \sum_j W_{ij} \mathbf{y}_j \right\|^2, \mu \right) \text{ where } \mathbf{y}^T = \mathbf{a}^T \mathbf{X} \quad (24)$$

2. 在 TRPCA 中入权值的思想。

在上文我们可以看到，基于 PCA 的改进都是基于样本属性的重要度是平等的这个思想来做降维的。但是基于这个思想做数据的降维是不太合适的。由于数据存在各种特征。噪声的特征和样本原有的特征的重要度是不一样的，对数据本身的特征的刻画的重要度是要远大于噪声特征的重要性的。所以，映入权值对原数据做一个权值的分配，同时把权值的更新集成到模型自更新中。故提出以下的模型。

$$\min_{\mathbf{F}, \mathbf{b}, \alpha} \sum_{i=1}^n \alpha_i \left\| (\mathbf{I}_d - \mathbf{F}\mathbf{F}^T) (\Theta \mathbf{x}_i - \mathbf{b}) \right\|_2 \text{ s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}_k, \alpha^T \mathbf{1}_n = m, 0 \leq \alpha \leq 1, \Theta = \text{diag}(\boldsymbol{\theta}), \boldsymbol{\theta} \geq 0, \boldsymbol{\theta}^T \mathbf{1} = 1. \quad (25)$$

其中 Θ 是一个对角权值矩阵， θ 是每个特征的权值向量。

3. 对 RSFCM 进行优化

RSFCM 是一个鲁棒和稀疏的模糊 C 均值聚类算法，它是在模糊 C 均值聚类 FCM 的基础上做的一个改进，RSFCM 的目标函数也是用截断的思想做的一个优化，但是损失上界是不好确定的，故利用本文提出的一个优化方法，我们将其目标函数做了一个该改进，从而能够进一步减少参数的选择范围。原始和改进后的目标函数如下所示：原始的 RSFCM 目标函数：

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^N \sum_{k=1}^c u_{ik} \min (\|\mathbf{x}_i - \mathbf{v}_k\|_2, \epsilon) + \gamma \|\mathbf{U}\|_F^2 \text{ s.t. } \mathbf{U} \geq 0, \mathbf{U}\mathbf{1} = \mathbf{1} \quad (26)$$

提出改进的 RSFCM 目标函数：

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^N \sum_{k=1}^C u_{ik} \alpha_i \|\mathbf{x}_i - \mathbf{v}_k\|_2 + \gamma \|\mathbf{U}\|_F^2 \text{ s.t. } \mathbf{U} \geq 0, \mathbf{U}\mathbf{1} = \mathbf{1}, \alpha^T \mathbf{1}_n = m, 0 \leq \alpha \leq 1 \quad (27)$$

5 总结与展望

本文着重在不影响正确样本辨别力的情况下提高 PCA 的鲁棒性，为此，提出了截断鲁棒的主成分分析 (TRPCA) 模型。通过一些简练的定理，所提出的模型可以控制通过截断损失忽略样本的百分比。减少了参数的选择量，同时提高了对正常样本辨别和异常样本的鲁棒性。基于本文的思想，提出了另外的三个想法。一是对 NPE 这个降维算法引入截断的思想来提高对噪声的鲁棒性。二是在 TRFCM 中对样本特征引入权值以进一步提高 PCA 的鲁棒性。三是对 RSFCM 模型的目标函数做了一个优化。后续将会对这三个模型进行实验以验证模型的正确性和可行性。

参考文献

- [1] NIE F, WU D, WANG R, et al. Truncated Robust Principle Component Analysis With A General Optimization Framework[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(2): 1081-1097.
- [2] JOLLIFFE I. Principal Component Analysis.[J]. Berlin, Germany: Springer, 2011.
- [3] KE Q, KANADE T. Robust $l_{1/2}$ norm factorization in the presence of outliers and missing data by alternative convex programming[J]. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2005: 739-746.
- [4] KWAK N. Principal component analysis based on L1-norm maximization[J]. IEEE Trans. Pattern Anal. Mach. Intell., 2008, 30(9): 1672-1680.
- [5] C.DING X, D. Zhou, ZHA H. R1-PCA: rotational invariant l1-norm principal component analysis for robust subspace factorization[J]. Proc. 23rd Int. Conf.Mach. Learn., 2006: 281-288.
- [6] F. NIE J Y, HUANG H. Optimal mean robust principal component analysis[J]. Proc. Int. Conf. Mach. Learn., 2014: 1062-1070.
- [7] Q.WANG X G, Q. Gao, NIE F. “l2p norm based PCA for image recognition[J]. IEEE Trans. Image Process., 2018, 27(3): 1336-1346.