

多线性鲁棒主成分分析

史加荣¹,周水生²,郑秀云¹

(1. 西安建筑科技大学理学院, 陕西西安 710055; 2. 西安电子科技大学数学与统计学院, 陕西西安 710071)

摘要: 鲁棒主成分分析(RPCA)是恢复低秩与稀疏成分的一种非常有效的方法. 本文将RPCA推广到张量情形, 提出了多线性鲁棒主成分分析(MRPCA)框架. 首先建立了MRPCA模型, 即最小化张量核范数与 l_1 范数的加权组合. 然后使用增广拉格朗日乘子法求解上述张量核范数优化问题. 实验结果证实: 对于具有多线性结构的数据, MRPCA比RPCA更加鲁棒.

关键词: 多线性鲁棒主成分分析; 鲁棒主成分分析; 低秩; 核范数最小化; 增广拉格朗日乘子法

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2014)08-1480-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.08.004

Multilinear Robust Principal Component Analysis

SHI Jia-rong¹, ZHOU Shui-sheng², ZHENG Xiu-yun¹

(1. School of Science, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China;

2. School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: Robust principal component analysis(RPCA) is a very effective method to recover both the low-rank and sparse components. This paper extends RPCA to the case of tensor and proposes a framework of multilinear robust principal component analysis(MRPCA). First, it establishes the model of MRPCA which minimizes a weighted combination of the tensor nuclear norm and l_1 norm. Then, it employs the augmented Lagrange multipliers algorithm to solve the above nuclear norm optimization problem. Experimental results demonstrate that MRPCA is more robust than RPCA for the data with multilinear structure.

Key words: multilinear robust principal component analysis; robust principal component analysis; low-rank; nuclear norm minimization; augmented Lagrange multipliers

1 引言

主成分分析(PCA, Principal Component Analysis)是分析与处理高维数据的一种最常用的统计工具, 它广泛地应用在信号与图像处理、模式识别、计算机视觉和数据挖掘等科学与工程领域中. 待研究的数据样例通常用量来表示, 样例集合则表示成矩阵形式. 在PCA中, 假设所给定的高维数据嵌在某个低维线性子空间中, 即数据矩阵是低秩的. PCA也可以解释为将给定的样例集合分解为一个低秩矩阵与一个高斯噪声矩阵之和.

为了增强PCA对稀疏噪声的鲁棒性, Wright等学者提出了鲁棒主成分分析(RPCA, Robust PCA)^[1], 即先将数据矩阵表示为低秩矩阵与稀疏矩阵之和, 再通过求解核范数优化问题来恢复低秩结构与稀疏成分. Candès等学者从理论上证明了在某些条件下可以通过求解上述凸优化问题来精确地恢复低秩成分和稀疏噪声^[2]. 求解核范数优化问题的方法主要包括半定规划法^[3]、迭代阈

值法^[4]、加速近端梯度法^[5]、增广拉格朗日乘子法^[6]和交替方向法^[7]等. 文献[8,9]从理论上分析了RPCA的鲁棒性和稳定性, 文献[10]将RPCA的思想应用到子空间聚类上, 提出了低秩表示模型, 文献[11]将RPCA与稀疏表示分类方法结合, 并将其应用到人脸识别.

在实际应用中, 图像和视频等数据具有多线性结构. 对于这类数据, 传统的处理方法是每个数据样例表示成向量形式, 但这种向量化破坏了原始样例的空时结构, 还可能导致小样本问题和维数灾难问题的产生^[12,13]. 近年来, 低秩张量分解已成为处理这类多线性数据的新兴方法^[14], 同时也涌现出多线性判别分析^[12]、多线性主成分分析^[15]和增量张量子空间^[16]等诸多多线性子空间方法. 此外, 一些学者根据张量的低秩结构来求解张量补全问题(TC, Tensor Completion)^[17,18]. 本文将鲁棒主成分分析推广到张量情形, 提出了多线性鲁棒主成分分析(MRPCA, Multilinear RPCA)模型, 并设计具有可扩展性的算法.

2 预备知识

本节简要介绍矩阵与张量代数的基本知识. 先引入矩阵的一种非常重要的分解模式, 即奇异值分解 (SVD, Singular Value Decomposition).

定理 1 矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 它的奇异值分解为 $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, 其中 $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m)$ 和 $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ 分别为 m 阶、 n 阶正交矩阵, \mathbf{S} 为 $m \times n$ 维对角矩阵且对角线元素 σ_i 满足 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_{\min(m,n)} = 0$.

称 σ_i 为矩阵 \mathbf{A} 的奇异值, \mathbf{u}_i 和 \mathbf{v}_i 分别为 σ_i 的左、右奇异向量. 易知 \mathbf{A} 的秩 $\text{rank}(\mathbf{A}) = r$, 它的 SVD 可重新表示为 $\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. 若 \mathbf{A} 的列向量经过中心化, 则 SVD 就是 PCA. 用 $\|\mathbf{A}\| = \sigma_1$ 表示 \mathbf{A} 的算子范数 (谱范数). 根据 SVD, 还可定义矩阵核范数^[3].

定义 1 (矩阵核范数) 矩阵 \mathbf{A} 的非零奇异值为 $\sigma_1, \sigma_2, \dots, \sigma_r$, 则它的核范数为 $\|\mathbf{A}\|_* = \sum_{i=1}^r \sigma_i$.

定义 2 (矩阵伸缩算子)^[2,4] 对于任意 $\varepsilon > 0$, 矩阵 \mathbf{A} 的伸缩绝对值算子为 $\mathcal{S}_\varepsilon(\mathbf{A}) = \max(\mathbf{A} - \varepsilon \mathbf{E}, 0) + \min(\mathbf{A} + \varepsilon \mathbf{E}, 0)$, 伸缩奇异值算子为 $\mathcal{D}_\varepsilon(\mathbf{A}) = \sum_{i=1}^r \max(\sigma_i - \varepsilon, 0) \mathbf{u}_i \mathbf{v}_i^T$, 其中 $m \times n$ 维矩阵 \mathbf{E} 的所有元素均为 1.

作为向量和矩阵的高阶推广, 张量通过多个指标来表示数据, 也称为多阶阵列. 本文遵循文献[14]的符号描述, 并且仅考虑实张量. 在以下定义中, 规定张量 $\mathcal{A} = (a_{i_1 i_2 \dots i_N})_{I_1 \times I_2 \times \dots \times I_N}$, $\mathcal{B} = (b_{i_1 i_2 \dots i_N})_{I_1 \times I_2 \times \dots \times I_N}$.

定义 3 (张量内积) \mathcal{A} 与 \mathcal{B} 的内积为

$$[\mathcal{A}, \mathcal{B}] = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \dots i_N} b_{i_1 i_2 \dots i_N}.$$

定义 4 (张量范数) \mathcal{A} 的 Frobenious 范数为 $\|\mathcal{A}\|_F = \sqrt{[\mathcal{A}, \mathcal{A}]}$, l_0 范数 $\|\mathcal{A}\|_0$ 为它的非零元素数目,

l_1 范数为 $\|\mathcal{A}\|_1 = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} |a_{i_1 i_2 \dots i_N}|$.

定义 5 (张量 n -模式矩阵化) \mathcal{A} 的 n -模式矩阵化是将它的元素重新排列成 $I_n \times \prod_{j \neq n} I_j$ 维矩阵 $\mathbf{A}_{(n)}$, 其第 (i_n, k) 元素为 $a_{i_1 i_2 \dots i_N}$, $k = 1 + \sum_{m \neq n} (i_m - 1) \prod_{m' \neq m, n} I_{m'}$.

定义 6 (张量核范数)^[18] \mathcal{A} 的核范数为 $\|\mathcal{A}\|_* = \sum_{n=1}^N \alpha_n \|\mathbf{A}_{(n)}\|_*$, 其中权重系数 α_n 非负且满足 $\sum_{n=1}^N \alpha_n = 1$.

定义 7 (张量 n -模式积) \mathcal{A} 与矩阵 $\mathbf{U} = (u_{jn})_{J_n \times I_n}$ 的 n -模式积为 $\mathcal{A} \times_n \mathbf{U} = (\sum_{i_n=1}^{I_n} a_{i_1 \dots i_n \dots i_N} u_{jn})_{I_1 \times \dots \times I_n \times \dots \times I_N}$.

定义 8 (张量的秩) \mathcal{A} 的 n -模式秩为 $\text{rank}(\mathbf{A}_{(n)})$,

多线性秩为 N 维向量 $\text{rank}(\mathcal{A}) = (\text{rank}(\mathbf{A}_{(1)}), \dots, \text{rank}(\mathbf{A}_{(N)}))$.

3 多线性鲁棒主成分分析

3.1 鲁棒主成分分析

在实际应用中, 所给定的数据集合可按列排列成一个大的矩阵 $\mathbf{D} \in \mathbb{R}^{m \times n}$, 并假设它们近似存在于一个低维线性子空间中. 基于上述假设, 可将矩阵 \mathbf{D} 分解为低秩矩阵 \mathbf{A} 与噪声矩阵 \mathbf{E} 之和. 经典的 PCA 对小的稠密高斯噪声非常有效, 但对大的稀疏噪声或野点却非常敏感. 而鲁棒主成分分析 (RPCA) 能够有效地恢复矩阵 \mathbf{D} 的低秩结构 \mathbf{A} 和稀疏噪声 \mathbf{E} , 其求解模型如下:

$$\min_{\mathbf{A}, \mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1, \text{ s.t. } \mathbf{A} + \mathbf{E} = \mathbf{D} \quad (1)$$

其中 $\lambda > 0$ 为矩阵低秩与稀疏的折中因子, 文献[2]建议按如下方式选取 $\lambda: \lambda = 1/\sqrt{\max(m, n)}$.

求解最优化问题(1)的各种优化算法可参见文献[6]和[19]. 下面给出求解此优化问题的较为有效的算法: 增广拉格朗日乘子法. 先构造增广拉格朗日函数

$$L(\mathbf{A}, \mathbf{E}, \mathbf{Y}, \mu) = \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 + [\mathbf{Y}, \mathbf{D} - \mathbf{A} - \mathbf{E}] + \mu \|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2/2 \quad (2)$$

当 $\mathbf{E} = \mathbf{E}_k, \mathbf{Y} = \mathbf{Y}_k, \mu = \mu_k$ 时, 矩阵 \mathbf{A} 的更新公式为 $\mathbf{A}_{k+1} = \arg \min_{\mathbf{A}} L(\mathbf{A}, \mathbf{E}_k, \mathbf{Y}_k, \mu_k)$

$$= \arg \min_{\mathbf{A}} \|\mathbf{A}\|_* + \mu_k \|\mathbf{A} - (\mathbf{D} - \mathbf{E}_k + \mathbf{Y}_k/\mu_k)\|_F^2/2 = \mathcal{D}_{1/\mu_k}(\mathbf{D} - \mathbf{E}_k + \mathbf{Y}_k/\mu_k) \quad (3)$$

当 $\mathbf{A} = \mathbf{A}_{k+1}, \mathbf{Y} = \mathbf{Y}_k, \mu = \mu_k$ 时, 矩阵 \mathbf{E} 的更新公式为 $\mathbf{E}_{k+1} = \arg \min_{\mathbf{E}} L(\mathbf{A}_{k+1}, \mathbf{E}, \mathbf{Y}_k, \mu_k)$

$$= \arg \min_{\mathbf{E}} \lambda \|\mathbf{E}\|_1 + \mu_k \|\mathbf{E} - (\mathbf{D} - \mathbf{A}_{k+1} + \mathbf{Y}_k/\mu_k)\|_F^2/2 = \mathcal{S}_{\lambda/\mu_k}(\mathbf{D} - \mathbf{A}_{k+1} + \mathbf{Y}_k/\mu_k) \quad (4)$$

当 $\mathbf{A} = \mathbf{A}_{k+1}, \mathbf{E} = \mathbf{E}_{k+1}, \mu = \mu_k$ 时, 矩阵 \mathbf{Y} 的更新公式为 $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k (\mathbf{D} - \mathbf{A}_{k+1} - \mathbf{E}_{k+1})$ (5)

最后更新参数 μ ,

$$\mu_{k+1} = \begin{cases} \rho \mu_k, & \text{若 } \mu_k \|\mathbf{E}_{k+1} - \mathbf{E}_k\|_F / \|\mathbf{D}\|_F < \varepsilon \\ \mu_k, & \text{否则} \end{cases} \quad (6)$$

其中 $\rho > 1$ 为常数, $\varepsilon > 0$ 为比较小的正数.

3.2 多线性鲁棒主成分分析模型

考虑具有 N 阶多线性结构的数据样例集合 $X = \{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m\}$, 其中 $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, $i = 1, 2, \dots, m$. 记 $I_{N+1} = m$, 则集合 X 可用一个 $I_1 \times I_2 \times \dots \times I_{N+1}$ 维的 $N+1$ 阶张量 \mathcal{D} 来表示. 通常假设 \mathcal{D} 在各个模式上是低秩的. 当 \mathcal{D} 受噪声腐蚀时, 可将它分解为低秩张量 \mathcal{A} 与噪声张量 \mathcal{E} 之和, 即 $\mathcal{D} = \mathcal{A} + \mathcal{E}$.

当 \mathcal{E} 为小的稠密高斯噪声张量时, 可通过求解下列优化问题来恢复低秩张量 \mathcal{A} :

$$\min_{\mathcal{A}} (\text{rank}(\mathbf{A}_{(1)}), \dots, \text{rank}(\mathbf{A}_{(N+1)})), \text{ s.t. } \|\mathcal{D} - \mathcal{A}\|_F \leq \varepsilon \quad (7)$$

上述多目标优化问题也可转化为如下单目标优化问题

$$\min_{\mathcal{A}} \|\mathcal{D} - \mathcal{A}\|_F, \text{ s.t. rank}(\mathbf{A}_{(n)}) \leq r_n, n = 1, 2, \dots, N+1 \quad (8)$$

这就是多线性主成分分析(MPCA). 通过不同模式的协方差矩阵的特征分解, 可得到相应的特征空间.

当 \mathcal{E} 为大的稀疏噪声张量时, 可通过求解如下优化模型来恢复低秩张量 \mathcal{A} 和稀疏张量 \mathcal{E} :

$$\min_{\mathcal{A}, \mathcal{E}} (\text{rank}(\mathbf{A}_{(1)}), \dots, \text{rank}(\mathbf{A}_{(N+1)}), \|\mathcal{E}\|_0), \text{ s.t. } \mathcal{D} = \mathcal{A} + \mathcal{E} \quad (9)$$

将上述多目标优化问题转化为单目标优化问题

$$\min_{\mathcal{A}, \mathcal{E}} \sum_{n=1}^{N+1} \alpha_n \text{rank}(\mathbf{A}_{(n)}) + \lambda \|\mathcal{E}\|_0, \text{ s.t. } \mathcal{D} = \mathcal{A} + \mathcal{E} \quad (10)$$

其中 $\alpha_n \geq 0$ 为 $\text{rank}(\mathbf{A}_{(n)})$ 的权重且满足 $\sum_{n=1}^{N+1} \alpha_n = 1$. 将最优化问题(10)凸松弛到如下形式

$$\min_{\mathcal{A}, \mathcal{E}} \|\mathcal{A}\|_* + \lambda \|\mathcal{E}\|_1, \text{ s.t. } \mathcal{D} = \mathcal{A} + \mathcal{E} \quad (11)$$

其中 $\|\mathcal{A}\|_* = \sum_{n=1}^{N+1} \alpha_n \|\mathbf{A}_{(n)}\|_*$. 称最优化问题式(11)为多线性鲁棒主成分分析(MRPCA)模型.

3.3 多线性鲁棒主成分分析算法

为求解最优化模型(11), 引入 $N+1$ 个 $I_1 \times I_2 \times \dots \times I_{N+1}$ 维张量 $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_{N+1}$. 令 $\mathbf{M}_{n(n)} = \mathbf{A}_{(n)}$, 其中 $\mathbf{M}_{n(n)}, \mathbf{A}_{(n)}$ 分别表示张量 \mathcal{M}_n 和 \mathcal{A} 按 n -模式的矩阵化, $n = 1, 2, \dots, N+1$. 于是最优化问题式(11)等价于

$$\min_{\mathcal{A}, \mathcal{E}, \mathcal{M}_1, \dots, \mathcal{M}_{N+1}} \sum_{n=1}^{N+1} \alpha_n \|\mathbf{M}_{n(n)}\|_* + \lambda \|\mathcal{E}\|_1, \text{ s.t. } \mathcal{A} = \sum_{n=1}^{N+1} \mathcal{M}_n / (N+1), \mathcal{D} = \mathcal{M}_n + \mathcal{E}, n = 1, 2, \dots, N+1 \quad (12)$$

上述最优化模型也等价于

$$\min_{\mathcal{E}, \mathcal{M}_1, \dots, \mathcal{M}_{N+1}} \sum_{n=1}^{N+1} \alpha_n \|\mathbf{M}_{n(n)}\|_* + \lambda \|\mathcal{E}\|_1, \text{ s.t. } \mathcal{D} = \mathcal{M}_n + \mathcal{E}, n = 1, 2, \dots, N+1 \quad (13)$$

构造最优化问题式(13)的增广拉格朗日函数

$$\begin{aligned} L(\mathcal{E}, \mathcal{M}_1, \dots, \mathcal{M}_{N+1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{N+1}, \mu) &= \sum_{n=1}^{N+1} \alpha_n \|\mathbf{M}_{n(n)}\|_* + \lambda \|\mathcal{E}\|_1 \\ &+ \sum_{n=1}^{N+1} [\mathcal{Y}_n, \mathcal{D} - \mathcal{M}_n - \mathcal{E}] / (N+1) \\ &+ \mu \sum_{n=1}^{N+1} \|\mathcal{D} - \mathcal{M}_n - \mathcal{E}\|_F^2 / (2N+2) \end{aligned} \quad (14)$$

当 $\mathcal{M}_1, \dots, \mathcal{M}_{N+1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{N+1}, \mu$ 给定时, 张量 \mathcal{E} 的更新公式为

$$\begin{aligned} \mathcal{E} &:= \arg \min_{\mathcal{E}} L(\mathcal{E}, \mathcal{M}_1, \dots, \mathcal{M}_{N+1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{N+1}, \mu) \\ &= \arg \min_{\mathcal{E}} \lambda \|\mathcal{E}\|_1 + \mu \|\mathcal{E} - \mathcal{D}\|_F^2 / 2 = \mathcal{S}_{\lambda/\mu}(\mathcal{D}) \end{aligned} \quad (15)$$

其中, $\mathcal{D} = \mathcal{D} - \sum_{n=1}^{N+1} (\mu \mathcal{M}_n - \mathcal{Y}_n) / (N\mu + \mu)$, $\mathcal{S}_{\lambda/\mu}(\mathcal{D})$ 为矩阵伸缩绝对值算子的高阶推广.

当 $\mathcal{M}_1, \dots, \mathcal{M}_{n-1}, \mathcal{M}_{n+1}, \dots, \mathcal{M}_{N+1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{N+1}, \mu$ 给定时, 矩阵 $\mathbf{M}_{n(n)}$ 的更新公式为

$$\begin{aligned} \mathbf{M}_{n(n)} &:= \arg \min_{\mathbf{M}_{n(n)}} L(\mathcal{E}, \mathcal{M}_1, \dots, \mathcal{M}_{N+1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{N+1}, \mu) \\ &= \arg \min_{\mathbf{M}_{n(n)}} \alpha_n \|\mathbf{M}_{n(n)}\|_* + \mu \|\mathbf{M}_{n(n)} - \mathbf{G}_n\|_F^2 / (2N+2) \\ &= \mathcal{S}_{(N+1)\alpha_n/\mu}(\mathbf{G}_n) \end{aligned} \quad (16)$$

其中 $\mathbf{D}_{(n)}, \mathbf{E}_{(n)}, \mathbf{Y}_{n(n)}$ 分别为 $\mathcal{D}, \mathcal{E}, \mathcal{Y}_n$ 的 n -模式矩阵化, $\mathbf{G}_n = \mathbf{D}_{(n)} - \mathbf{E}_{(n)} + \mathbf{Y}_{n(n)}/\mu$. 将更新后的矩阵 $\mathbf{M}_{n(n)}$ 张量化即得到 \mathcal{M}_n .

当 $\mathcal{M}_1, \dots, \mathcal{M}_{N+1}, \mathcal{Y}_1, \dots, \mathcal{Y}_{n-1}, \mathcal{Y}_{n+1}, \dots, \mathcal{Y}_{N+1}, \mu$ 给定时, 根据 Uzawa 算法更新 \mathcal{Y}_n :

$$\mathcal{Y}_n := \mathcal{Y}_n + \mu(\mathcal{D} - \mathcal{M}_n - \mathcal{E}) \quad (17)$$

参数 μ 的更新公式为:

$$\mu := \min(\rho\mu, \bar{\mu}) \quad (18)$$

其中常量 $\rho > 1, \bar{\mu}$ 为比较大的常数.

迭代公式(15)~(18)的终止条件可设置为 $\|\mathcal{D} - \mathcal{M}_{N+1} - \mathcal{E}\|_F / \|\mathcal{D}\|_F < \epsilon$, 其中 ϵ 表示比较小的正数. 变量的初始化可按如下方式选取: $\mathcal{M}_1 = \dots = \mathcal{M}_{N+1} = \mathcal{O}, \mathcal{E} = \mathcal{O}, \mathcal{Y}_1 = \dots = \mathcal{Y}_{N+1} = (N+1)\mathcal{D} / \|\mathbf{D}_{(N+1)}\|_d, \mu = 1.25(N+1) / \|\mathbf{D}_{(N+1)}\|_d$, 其中 \mathcal{O} 表示 $I_1 \times I_2 \times \dots \times I_{N+1}$ 同维零张量, $\|\mathbf{D}_{(N+1)}\|_d = \max(\|\mathbf{D}_{(N+1)}\|, \|\text{vec}(\mathbf{D}_{(N+1)})\|_\infty / \lambda)$ 表示 $\mathbf{D}_{(N+1)}$ 的对偶范数, $\text{vec}(\cdot)$ 为矩阵向量化算子.

3.4 与鲁棒主成分分析的联系

对于多线性样例集合 X , 若使用 RPCA, 则需要将 \mathcal{X}_i 向量化为 $\mathbf{d}_i \in \mathbb{R}^{I_1 \times \dots \times I_N \times 1}$, 再由 \mathbf{d}_i 构造观测矩阵 $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m)$. 下面讨论 RPCA 与 MRPCA 之间的联系.

在 MRPCA 算法中, 只有 $\mathcal{M}_n (n = 1, \dots, N+1)$ 的迭代公式与各模式的权重向量 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{N+1})$ 有关. 取 $\boldsymbol{\alpha} = (0, 0, \dots, 1)$. 当 $n = 1, 2, \dots, N$ 时, \mathcal{M}_n 的迭代公式为 $\mathcal{M}_n := \mathcal{D} - \mathcal{E} + \mathcal{Y}_n / \mu$; 当 $n = N+1$ 时, $\mathbf{M}_{N+1(N+1)}$ 的迭代公式为 $\mathbf{M}_{N+1(N+1)} := \mathcal{S}_{(N+1)/\mu}(\mathbf{G}_{N+1})$. 在 $\mathcal{Y}_1 = \dots = \mathcal{Y}_{N+1}$ 条件下, 等式 $\mathcal{M}_1 = \dots = \mathcal{M}_N$ 成立. 如果参数 μ 的取值比较大, 则有 $\mathbf{M}_{N+1(N+1)} \approx \mathbf{G}_{N+1}$ 或 $\mathcal{M}_{N+1} \approx \mathcal{D} - \mathcal{E} + \mathcal{Y}_{N+1}/\mu$, 于是有 $\mathcal{M}_1 = \dots = \mathcal{M}_N \approx \mathcal{M}_{N+1}$. 反之, 当 $\mathcal{M}_1 = \dots = \mathcal{M}_N = \mathcal{M}_{N+1}$ 时, 显然有 $\mathcal{Y}_1 = \dots = \mathcal{Y}_{N+1}$. 因此, 当 $\alpha_1 = \alpha_2 = \dots = \alpha_N = 0, \alpha_{N+1} = 1$ 时, MRPCA 算法近似等价于 RPCA 算法. 由上述分析可知, RPCA 可以看作 MRPCA 的特例.

4 实验分析

在人工数据集和图像数据集上进行实验. 先将给定的数据集表示成高阶张量, 再分别使用 RPCA 和 MRPCA 两种方法来恢复低秩与稀疏成分, 最后对这两种方法的实验结果进行对比分析. 在实验中, 参数选取如下: $\rho = 1.5, \epsilon = 10^{-8}, \bar{\mu} = 10^7 \|\mathbf{D}_{(N+1)}\|_d$.

4.1 人工数据集

在本节实验中,按如下方式生成 $I_1 \times I_2 \times \dots \times I_{N+1}$ 维 $N+1$ 阶低秩张量: $\mathcal{A} = \mathcal{L} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \dots \times_{N+1} \mathbf{U}^{(N+1)}$, 其中 $\mathcal{L} \in \mathbb{R}^{J_1 \times J_2 \times \dots \times J_{N+1}}$ 和 $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times J_n}$ 的元素服从相互独立的标准正态分布, $J_n < I_n, n = 1, \dots, N+1$. 由上述生成方式可知 $\text{rank}(\mathcal{A}) = (J_1, J_2, \dots, J_{N+1})$, 对 \mathcal{A} 添加稀疏噪声 \mathcal{E} , 则得到含噪张量 $\mathcal{D} = \mathcal{A} + \mathcal{E}$, 其中 \mathcal{E} 的非零元素服从区间 $(-500, 500)$ 上的均匀分布且相互独立.

根据 MRPCA 算法得到 \mathcal{D} 的低秩成分 $\hat{\mathcal{A}}$ 和稀疏成分 $\hat{\mathcal{E}}$. 用 $\|\mathcal{E}\|_0 / \prod_{n=1}^{N+1} I_n$ 表示噪声 \mathcal{E} 的稀疏度, 逆信噪比 $\|\mathcal{E}\|_F / \|\mathcal{A}\|_F$ 表示噪声的大小, $\|\hat{\mathcal{A}} - \mathcal{A}\|_F / \|\mathcal{A}\|_F$ 表示低秩逼近误差. 将 \mathcal{D} 按 $(N+1)$ -模式矩阵化为 $\mathbf{D}_{(N+1)}$, 根据 RPCA 算法得到它的低秩与稀疏成分, 类似定义逆信噪比和低秩逼近误差. 当低秩逼近误差小于 10^{-4} 时, 称低秩张量或矩阵得到精确恢复. 在 MRPCA 中, 取 $\alpha = (1, 1, \dots, 1) / (N+1)$. 将张量 \mathcal{D} 随机生成 20 次, 分别执行 RPCA 和 MRPCA, 最终报告逆信噪比和低秩逼近误差的平均值.

先考虑 3 阶张量情形, 取 $(I_1, I_2, I_3) = (50, 50, 50)$, $\lambda \in [0.002, 0.08]$. 图 1 绘出了 $(J_1, J_2, J_3) = (3, 3, 3)$ 、噪声稀疏度为 0.05 时不同 λ 取值下的低秩逼近误差, 可以看出: 对于相同的 λ , MRPCA 的逼近误差均优于 RPCA; 精确恢复时, RPCA 的 λ 取值区间是 MRPCA 区间的子集. 对于 0.05, 0.1, 0.15 三种稀疏度和 $(3, 3, 3)$ 、 $(5, 5, 5)$ 两组不同低秩取值, 6 组实验结果如表 1 所示.

在表 1 中, 低秩逼近误差与最优的 λ 值相对应. 在 6 组实验中, 对于比较大的逆信噪比, MRPCA 均精确地恢复出了低秩成分, 而 RPCA 仅有 2 组精确恢复, 且其精确恢复时 λ 的取值区间均劣于 MRPCA 对应的区间. 这说明 MRPCA 对 λ 的选取比 RPCA 更加灵活稳定. 此外, 随噪声稀疏度或秩增加, MRPCA 精确恢复时 λ 的取值区间的长度变小, 即大的噪声稀疏度和秩对精确恢复产生不利影响.

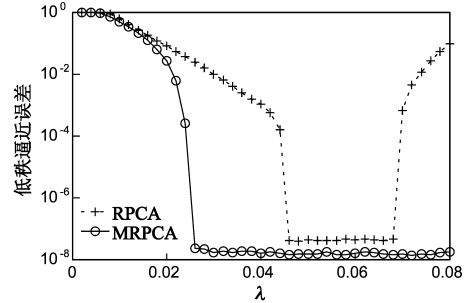


图1 3阶张量不同 λ 取值下的低秩逼近误差

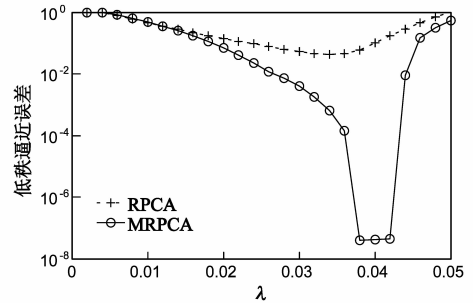


图2 4阶张量不同 λ 取值下的低秩逼近误差

表 1 RPCA 和 MRPCA 在 3 阶张量上的实验结果

(J_1, J_2, J_3)	噪声稀疏度	逆信噪比	RPCA			MRPCA		
			精确恢复时 λ 取值区间	最优 λ 值	低秩逼近误差	精确恢复时 λ 取值区间	最优 λ 值	低秩逼近误差
(3, 3, 3)	0.05	12.53	[0.046, 0.068]	0.048	3.81e-8	[0.024, 0.08]	0.072	1.32e-8
(3, 3, 3)	0.10	17.69	∅	0.046	1.54e-3	[0.028, 0.056]	0.042	2.96e-8
(3, 3, 3)	0.15	22.18	∅	0.038	6.66e-3	[0.028, 0.046]	0.044	3.92e-8
(5, 5, 5)	0.05	5.69	[0.042, 0.052]	0.052	3.72e-8	[0.032, 0.076]	0.072	9.59e-9
(5, 5, 5)	0.10	8.55	∅	0.044	3.80e-3	[0.03, 0.054]	0.05	1.90e-8
(5, 5, 5)	0.15	10.40	∅	0.036	1.70e-2	[0.03, 0.044]	0.038	3.36e-8

再考虑 4 阶张量情形, 取 $(I_1, I_2, I_3, I_4) = (20, 20, 20, 20)$, $\lambda \in [0.002, 0.05]$. 图 2 比较了 $(J_1, J_2, J_3, J_4) = (2, 2, 2, 2)$ 、噪声稀疏度为 0.05 时 MRPCA 和 RPCA 的低秩逼近误差, 可以看出: MRPCA 的低秩逼近误差明显优于 RPCA; 通过选取适当的 λ , MRPCA 能精确恢复低秩

成分, 而 RPCA 却不能. 设计噪声稀疏度为 0.05、0.1、0.15 和低秩为 $(2, 2, 2, 2)$ 、 $(4, 4, 4, 4)$ 的 6 组实验, 其中 λ 依据 MRPCA 的最优值选取, 精确恢复百分比表示在 20 次实验中能精确恢复低秩成分的比例, 实验结果如表 2 所示.

表 2 RPCA 和 MRPCA 在 4 阶张量上的实验结果

(J_1, J_2, J_3, J_4)	噪声稀疏度	逆信噪比	λ	RPCA		MRPCA	
				低秩逼近误差	精确恢复百分比(%)	低秩逼近误差	精确恢复百分比(%)
(2,2,2,2)	0.05	17.22	0.038	0.0605	0	3.93e-8	100
(2,2,2,2)	0.10	28.23	0.03	0.2440	0	5.99e-3	70
(2,2,2,2)	0.15	30.52	0.024	0.2580	0	0.0458	15
(4,4,4,4)	0.05	4.48	0.034	0.0728	0	6.18e-3	70
(4,4,4,4)	0.10	5.63	0.024	0.1193	0	0.0582	0
(4,4,4,4)	0.15	7.58	0.02	0.2072	0	0.1550	0

由表 2 可知稀疏噪声也比较大. 通过比较 RPCA 和 MRPCA 的实验结果, 可得出如下结论: MRPCA 的低秩逼近性能均优于 RPCA, 在前 4 组实验中, 前者的低秩逼近误差比后者低 1~6 个数量级; RPCA 的精确恢复百分比均为 0, 而 MRPCA 在前 4 组实验中的精确恢复百分比介于 15%~100%. 对于 MRPCA, 后 2 组实验未能精确恢复, 这可能是由大的逆信噪比、噪声稀疏度、张量阶数和秩共同导致. 通过对比表 1 和表 2 的实验结果, 可知 MRPCA 在低秩逼近误差和 λ 的选取上有非常大的优势, 这是因为它充分考虑了数据集本身的多线性结构.

4.2 图像数据集

在图像数据集或视频中, 由于各幅图像之间的关联性, 可将其视为低秩模型. 考虑 $I_1 \times I_2$ 维的灰度图像情形, I_3 幅图像构成的集合可表示 $I_1 \times I_2 \times I_3$ 维的 3 阶张量. 在 RPCA 和 MRPCA 模型中, 选取相同的折中因子: $\lambda = 1/\sqrt{\max(I_1, I_2, I_3)}$.

先考虑由低秩图像与稀疏图像生成的混合图像, 如图 3 所示. 混合图像的分辨率为 256×256 , 将它复制 50 次, 并分别对每次复制的图像添加密度为 0.1 的椒盐噪声, 于是得到 $256 \times 256 \times 50$ 的含稀疏噪声的 3 阶

张量 Δ . 为了恢复 Δ 的低秩成分和稀疏成分, 对 Δ 分别执行 RPCA 和 MRPCA 算法, 部分实验结果如图 4 所示, 其中在 MRPCA 中取 $\alpha = (1, 1, 1)/3$.

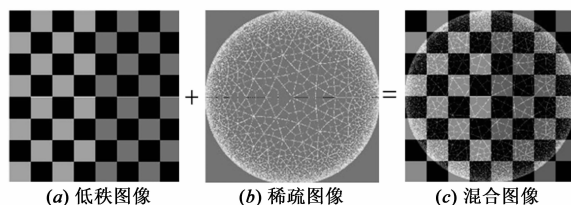


图 3 混合图像的生成

通过观察图 4 可以看出: RPCA 只恢复了部分噪声, 未能正确地恢复低秩成分, 而 MRPCA 在近乎完美地恢复低秩成分的同时, 也成功分离出稀疏噪声. 为了恢复不同类型的稀疏噪声, 可以执行两次 MRPCA. 在第一次 MRPCA 中, 按模式 3 恢复低秩成分和稀疏噪声, 即取 $\alpha = (0, 0, 1)$; 在第二次 MRPCA 中, 对第一次所得的低秩成分再按模式 1 进行低秩恢复, 即取 $\alpha = (1, 0, 0)$, 部分实验结果如图 5 所示. 从图 5 的结果可以看出所提方法成功地恢复了低秩成分和在模式 3、模式 1 上的稀疏噪声.

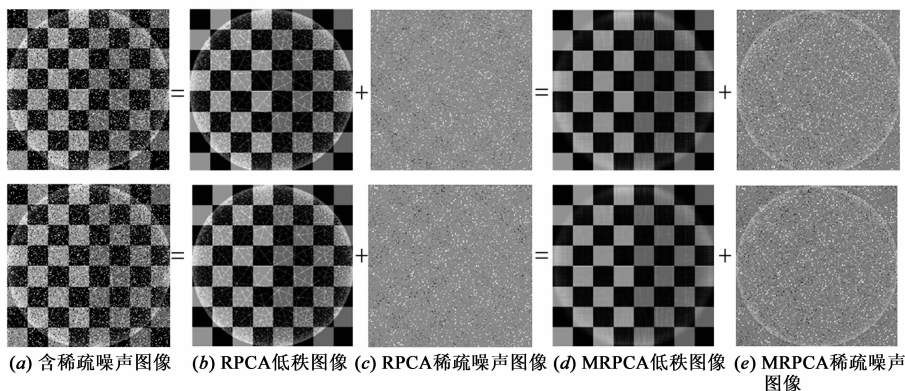


图 4 混合图像集的低秩恢复结果

再在 Lobby 视频数据集上进行实验. 选取该视频的前 200 帧灰度图像, 每帧图像的分辨率为 128×160 , 因此数据集可构成 $128 \times 160 \times 200$ 的 3 阶张量 Δ . 首先分析 MRPCA 的权重因子 α 的选取对视频背景建模结果的影响. 考虑 $\alpha = (1, 0, 0)$ 、 $(0, 1, 0)$ 、 $(0, 0, 1)$ 和 $(1, 1, 1)/3$ 四种

情形, 部分实验结果如图 6 所示, 其中第三种情形实际上对应于 RPCA.

在图 6 中, 第一行表示低秩图像, 第二行表示稀疏图像, 图 6(a)、(b) 和 (c) 分别只考虑了张量数据在单个模式上的低秩恢复, (d) 可视作前三种方式的平均.

由 Lobby 视频建模结果可知:图 6(a)或(b)中低秩图像淡化了原始图像中的亮斑,将部分亮斑作为稀疏成分的一部分;(c)中的低秩图像充分保持了亮斑;(d)是前

三种结果的折中,即在一定程度上抑制亮斑,同时还较好地恢复了稀疏成分.

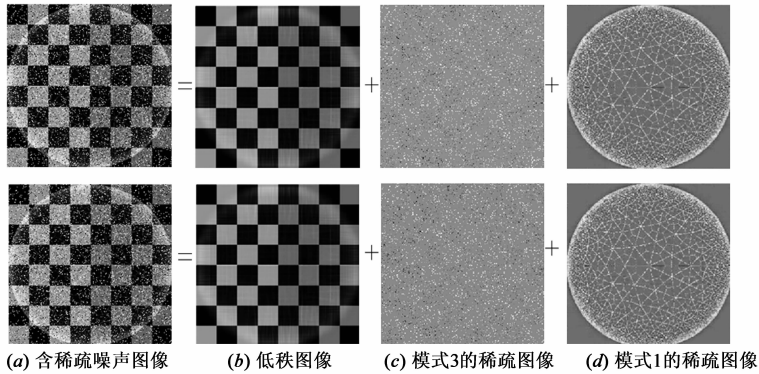


图5 混合图像集执行两次MRPCA的结果

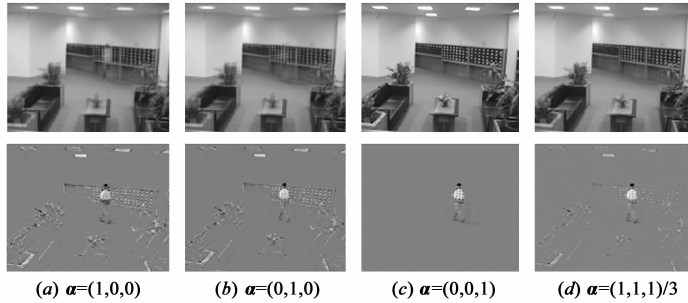


图6 不同权重因子下Lobby视频背景建模

对于前述视频数据集,设计第二组实验,即考虑图像存在丢失数据情形.对于每幅图像,按照相同的模式 φ_Ω 选取图像的部分元素.对于 $I_1 \times I_2$ 维矩阵 X 和指标集 $\Omega \subset \{1, 2, \dots, I_1\} \times \{1, 2, \dots, I_2\}$, 定义投影算子 φ_Ω :

当 $(i, j) \in \Omega$ 时, $\varphi_\Omega(X_{ij}) = X_{ij}$; 否则 $\varphi_\Omega(X_{ij}) = 0$. 记 Ω 的元素个数为 p , 在执行 RPCA 和 MRPCA 算法时,丢失元素用 255 替代.在 MRPCA 中取 $\alpha = (1, 1, 1)/3$, 图 7 列出了 $p/(I_1 I_2) = 0.1$ 时部分 Lobby 视频背景建模结果.

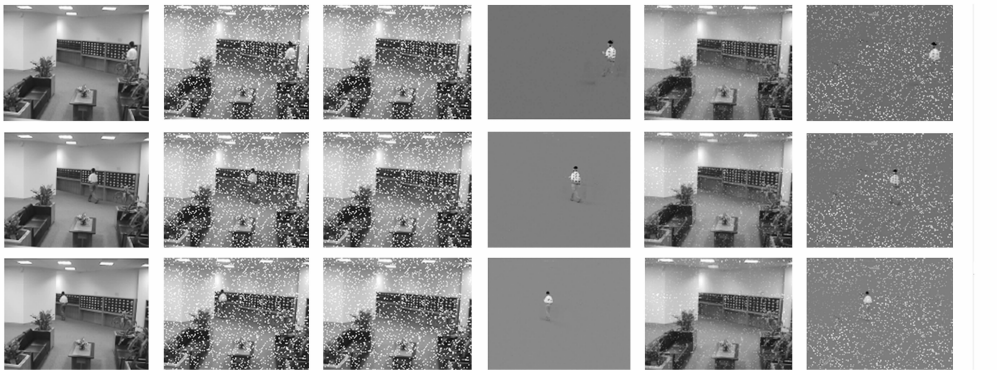


图7 含丢失元素的Lobby视频背景建模

由于丢失元素的值用 255 代替,所以这些元素往往对应着稀疏噪声.在图 7 中,将图 7(c)和(e)分别与(b)作比较,可以看出 MRPCA 较好地恢复了低秩成分,而 RPCA 几乎对丢失元素没有任何恢复;将图 7(d)和(f)作比较,发现(d)没有将丢失元素对应的稀疏噪声分离

出来,而(f)却较好地丢失元素对应位置的噪声进行恢复.若依次在模式 3 和模式 1 上执行 MRPCA,则可将(f)中的稀疏图像进一步分解为两稀疏图像之和.综上所述,MRPCA 在恢复低秩成分与稀疏噪声方面性能优于 RPCA.

5 结论

本文提出了多线性鲁棒主成分分析模型,基于增广拉格朗日乘子法给出了迭代算法.与鲁棒主成分分析相比,多线性鲁棒主成分分析充分考虑了数据集的空时结构,能更有效地恢复低秩成分和稀疏成分.对多线性鲁棒主成分分析算法及应用的研究仍需值得进一步研究.此外,在多线性鲁棒主成分分析中考虑含丢失元素情形(即张量补全问题)也将是今后研究的一个方向.

参考文献

- [1] Wright J, Ganesh A, Rao S, et al. Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization [A]. Proc Neural Information Processing Systems [C]. British Columbia, Canada, 2009. 2080 – 2088.
- [2] Candès E J, Li X, Ma Y, et al. Robust principal component analysis? [J]. Journal of the ACM, 2011, 58(3): 1 – 37.
- [3] Recht B, Fazel M, Parrilo P A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization [J]. SIAM Review, 2010, 52(3): 471 – 501.
- [4] Cai J F, Candès E J, Shen Z W. A singular value thresholding algorithm for matrix completion [J]. SIAM Journal on Optimization, 2010, 20(4): 1956 – 1982.
- [5] Toh K C, Yun S. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems [J]. Pacific Journal of Optimization, 2010, 6(3): 615 – 640.
- [6] Lin Z C, Chen M, Wu L, et al. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices [R]. Technical Report UILU-ENG-09-2215, UIUC, October, 2009.
- [7] Yuan X, Yang J. Sparse and low-rank matrix decomposition via alternating direction methods [R]. Dept of Mathematics, Hong Kong Baptist University, 2009.
- [8] Xu H, Caramanis C, Sanghavi S. Robust PCA via outlier pursuit [J]. IEEE Transactions on Information Theory, 2012, 58(5): 3047 – 3064.
- [9] Tang G G, Nehorai A. The stability of low-rank matrix reconstruction: a constrained singular value view [J]. IEEE Transactions on Information Theory, 2012, 58(9): 6079 – 6092.
- [10] Liu G C, Lin Z C, Yan S C, et al. Robust recovery of subspace structures by low-rank representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171 – 184.
- [11] 胡正平, 李静. 基于低秩子空间恢复的联合稀疏表示人脸识别算法 [J]. 电子学报, 2013, 41(5): 987 – 991.
Hu Zhengping, Li Jing. Face recognition of joint sparse representation based on low-rank subspace recovery [J]. Acta Elec-

tronica Sinica, 2013, 41(5): 987 – 991. (in Chinese)

- [12] Yan S C, Xu D, Yang Q, et al. Multilinear discriminant analysis for face recognition [J]. IEEE Transactions on Image Processing, 2007, 16(1): 212 – 220.
- [13] Shi J R, Jiao L C, Shang F H. Metric learning for high-dimensional tensor data [J]. Chinese Journal of Electronics, 2011, 20(3): 495 – 498.
- [14] Kolda T G, Bader B W. Tensor decompositions and applications [J]. SIAM Review, 2009, 51(3): 455 – 500.
- [15] Lu H, Plataniotis K N, Venetsanopoulos A N. MPCA: multilinear principal component analysis of tensor objects [J]. IEEE Transactions on Neural Networks, 2008, 19(1): 18 – 39.
- [16] 温静, 李洁, 高新波. 基于增量张量空间学习的自适应目标跟踪 [J]. 电子学报, 2009, 37(7): 1618 – 1623.
Wen Jing, Li Jie, Gao Xinbo. Adaptive object tracking with incremental tensor subspace learning [J]. Acta Electronica Sinica, 2009, 37(7): 1618 – 1623. (in Chinese)
- [17] 史加荣, 焦李成, 尚凡华. 张量补全算法及其在人脸识别中的应用 [J]. 模式识别与人工智能, 2011, 24(2): 255 – 261.
Shi Jiarong, Jiao Licheng, Shang Fanhua. Tensor completion algorithm and its applications in face recognition [J]. Pattern Recognition and Artificial Intelligence, 2011, 24(2): 255 – 261. (in Chinese)
- [18] Liu J, Musialski P, Wonka P, et al. Tensor completion for estimating missing values in visual data [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 208 – 220.
- [19] 史加荣, 郑秀云, 魏宗田, 等. 低秩矩阵恢复算法综述 [J]. 计算机应用研究, 2013, 30(6): 1601 – 1605.
Shi Jiarong, Zheng Xiuyun, Wei Zongtian, et al. Survey on algorithms of low-rank matrix recovery [J]. Application Research of Computers, 2013, 30(6): 1601 – 1605. (in Chinese)

作者简介



史加荣 男, 1979 年生于山东东阿, 2012 年于西安电子科技大学获得智能信息处理专业博士学位, 现为西安建筑科技大学理学院副教授, 主要研究方向为机器学习与模式识别。
E-mail: shijiarong@xauat.edu.cn

周水生 男, 1972 年 3 月生于陕西洛南县, 2005 年于西安电子科技大学获得计算机科学与技术专业博士学位, 现为西安电子科技大学数学与统计学院教授, 博士导师, 目前研究方向为最优化理论、算法及其在模式识别、机器学习中的应用。
E-mail: sszhou@mail.xidian.edu.cn