

低秩矩阵恢复算法综述*

史加荣, 郑秀云, 魏宗田, 杨威

(西安建筑科技大学理学院, 西安 710055)

摘要: 将鲁棒主成分分析、矩阵补全和低秩表示统称为低秩矩阵恢复, 并对近年来出现的低秩矩阵恢复算法进行了简要的综述。讨论了鲁棒主成分分析的各种优化模型及相应的迭代算法, 分析了矩阵补全问题及求解它的不精确增广拉格朗日乘子算法, 介绍了低秩表示的优化模型及求解算法。最后指出了有待进一步研究的问题。

关键词: 低秩矩阵恢复; 鲁棒主成分分析; 矩阵补全; 低秩表示; 增广拉格朗日乘子算法

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2013)06-1601-05

doi:10.3969/j.issn.1001-3695.2013.06.001

Survey on algorithms of low-rank matrix recovery

SHI Jia-rong, ZHENG Xiu-yun, WEI Zong-tian, YANG Wei

(School of Science, Xi'an University of Architecture & Technology, Xi'an 710055, China)

Abstract: This paper collectively referred robust principal component analysis, matrix completion and low-rank representation to as low-rank matrix recovery, and made a brief survey on the existing algorithms of low-rank matrix recovery. Firstly, it discussed various optimization models and the corresponding iterative algorithms for robust principal component analysis. Next, it analyzed the matrix completion problem and proposed the inexact augmented Lagrange multipliers algorithm to solve the problem. In addition, it introduced the optimization models for the low-rank representation problem and presented the iterative algorithm. Finally, this paper discussed several problems which need further research.

Key words: low-rank matrix recovery; robust principal component analysis; matrix completion; low-rank representation; augmented Lagrange multipliers

在计算机视觉、模式识别、数据挖掘和机器学习等领域中,常用的模型是假设相关数据存在(或近似存在)于一个低维线性子空间中。主成分分析(principal component analysis, PCA)、线性判别分析(linear discriminant analysis, LDA)和独立成分分析(independent component analysis, ICA)等子空间学习模型^[1,2]就是利用这种低维性质来进行维数约简、特征提取和噪声移除。传统的线性子空间方法对含小高斯噪声的数据非常有效,但对含野点和稀疏噪声大的数据却非常敏感。

在过去的几年里,基于稀疏表示的压缩感知(compressed sensing/compressed sampling, CS)^[3,4]在理论上取得了重要的进展,这些进展促使稀疏表示成为一种更加有效和流行的数据表示方式。与传统的子空间学习模型相比,稀疏表示对含野点和稀疏噪声大的数据更加鲁棒^[5,6]。近年来,低秩矩阵恢复(low-rank matrix recovery, LRMR)将向量样例的稀疏表示推广到矩阵的低秩情形,它已成为继CS之后又一种重要的数据获取和表示方式。LRMR先将数据矩阵表示为低秩矩阵与稀疏噪声矩阵之和,再通过求解核范数优化问题来恢复低秩矩阵。目前,LRMR主要由鲁棒主成分分析(robust PCA, RPCA)^[7-9]、矩阵补全(matrix completion, MC)^[10,11]和低秩表示(low-rank representation, LRR)^[12-14]等三类模型组成。

1 预备知识

定义1 向量 p 范数。向量 $\mathbf{a} = (a_1, a_2, \dots, a_n)^T \in \mathbb{R}^{n \times 1}$ 的 p 范数为 $\|\mathbf{a}\|_p = (\sum_{i=1}^n |a_i|^p)^{1/p}$, 其中 $p > 0$ 。

特殊地,当 $p=1$ 时, $\|\mathbf{a}\|_1 = \sum_{i=1}^n |a_i|$; 当 $p=2$ 时, $\|\mathbf{a}\|_2 = \sqrt{\sum_{i=1}^n a_i^2}$ 。此外,规定 $\|\mathbf{a}\|_0$ 为向量 $\mathbf{a}$ 的非零元素的个数, $\|\mathbf{a}\|_\infty = \max_{1 \leq i \leq n} |a_i|$ 。

定义2 矩阵的内积。对于两个同型的 $m \times n$ 维实矩阵 $\mathbf{A}$ 和 $\mathbf{B}$,它们的内积为 $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}$ 。

定义3 矩阵的范数。矩阵 $\mathbf{A} = (a_{ij})_{m \times n} \in \mathbb{R}^{m \times n}$ 的Frobenius范数为 $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$, 0 范数为 $\|\mathbf{A}\|_0$ 为矩阵 $\mathbf{A}$ 的非零元素的个数, ∞ 范数为 $\|\mathbf{A}\|_\infty = \max_{i,j} |a_{ij}|$, $(1,1)$ 范数为 $\|\mathbf{A}\|_{1,1} = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$, $(2,1)$ 范数为 $\|\mathbf{A}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m a_{ij}^2}$ 。

定理1 矩阵奇异值分解。矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n}$, 它的奇异值分解(singular value decomposition, SVD)为

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{U} \begin{pmatrix} \mathbf{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^T \quad (1)$$

其中: $\mathbf{U} \in \mathbb{R}^{m \times m}$ 和 $\mathbf{V} \in \mathbb{R}^{n \times n}$ 均为正交矩阵, 对角矩阵 $\mathbf{\Sigma}_r = \text{diag}$

收稿日期: 2012-10-10; **修回日期:** 2012-11-25 **基金项目:** 国家自然科学基金资助项目(61179040); 陕西省教育厅专项科研计划资助项目(2013JK0587, 2013JK0588, 12JK1000)

作者简介: 史加荣(1979-), 男, 山东东阿人, 副教授, 博士, 主要研究方向为机器学习与模式识别(shijiarong@xauat.edu.cn); 郑秀云(1982-), 女, 山东日照人, 讲师, 博士, 主要研究方向为最优化方法与应用; 魏宗田(1966-), 男, 陕西横山人, 副教授, 博士研究生, 主要研究方向为组合优化及其应用; 杨威(1979-), 女, 辽宁抚顺人, 讲师, 博士, 主要研究方向为信息融合及金融优化。

$(\sigma_1, \sigma_2, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$, 且其对角线元素满足 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ 。

显然矩阵 A 的秩为 $\text{rank}(A) = r$ 。记 $U = (u_1, u_2, \dots, u_m)$, $V = (v_1, v_2, \dots, v_n)$, 则式(1)可重新表示为

$$A = \sum_{i=1}^r \sigma_i u_i u_i^T \quad (2)$$

u_i 和 v_i 分别为奇异值 σ_i 对应的左奇异向量和右奇异向量。易知 $\|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}$, 矩阵 A 在 Frobenius 范数意义下的最优秩 r_0 逼近为 $A_{r_0} = \sum_{i=1}^{r_0} \sigma_i u_i u_i^T$, 且 $\|A - A_{r_0}\|_F^2 = \sum_{i=r_0+1}^r \sigma_i^2$, 其中 $r_0 < r$ 。

矩阵 A 的谱范数定义为 $\|A\|_2 = \sigma_1$, 核范数^[15] 定义为

$$\|A\|_* = \max_{U, V} \{ \text{trace}(U^T A V) : U^T U = I_m, V^T V = I_n \} \quad (3)$$

其中: I_m 表示 m 阶单位矩阵; $\text{trace}(\cdot)$ 表示矩阵的求迹算子。容易证明矩阵 A 的核范数可用它的奇异值来表示, 即 $\|A\|_* = \sum_{i=1}^r \sigma_i$, 且当 A 满足 $\|A\|_2 \leq 1$ 时, 此范数是 $\text{rank}(A)$ 的包络。下面引入矩阵的三种算子。

令 $Q = (q_1, q_2, \dots, q_n)$ 为 $m \times n$ 维实矩阵, 则优化问题

$$\min_X \varepsilon \|X\|_{1,1} + \|X - Q\|_F^2/2 \quad (4)$$

的最优解为 $X^* = S_\varepsilon(Q)$, 它的第 (i, j) 元素为 $\max(|q_{ij}| - \varepsilon, 0) \text{sgn}(q_{ij})$, 其中参数 $\varepsilon > 0$ ^[8]。易验证 $S_\varepsilon(Q) = \varepsilon S_1(Q/\varepsilon)$ 。若将优化问题式(4)中的 $(1, 1)$ 范数替换成核范数, 则有如下核范数优化问题:

$$\min_X \varepsilon \|X\|_* + \|X - Q\|_F^2/2 \quad (5)$$

此优化问题有闭形式解 $X^* = D_\varepsilon(Q)$, 其中 $D_\varepsilon(Q) = U S_\varepsilon(\Sigma) V^T$, $U \Sigma V^T$ 为矩阵 Q 的奇异值分解^[16]。显然有 $D_\varepsilon(Q) = \varepsilon D_1(Q/\varepsilon)$ 成立。若将优化问题式(4)中的 $(1, 1)$ 范数替换成 $(2, 1)$ 范数, 则得优化问题:

$$\min_X \varepsilon \|X\|_{2,1} + \|X - Q\|_F^2/2 \quad (6)$$

记上述优化问题的最优解为 $X^* = T_\varepsilon(Q)$ ^[12]。当 $\varepsilon < \|q_i\|_2$ 时, $T_\varepsilon(Q)$ 的第 i 列为 $(1 - \varepsilon/\|q_i\|_2) q_i$; 否则, $T_\varepsilon(Q)$ 的第 i 列为零向量。

2 鲁棒主成分分析

在许多实际应用中, 所给定的数据矩阵 $D \in \mathbb{R}^{m \times n}$ 往往是低秩或近似低秩的。为了恢复矩阵 D 的低秩结构, 可将矩阵 D 分解为两个矩阵之和, 即 $D = A + E$, 其中矩阵 A 和 E 未知, 但 A 是低秩的。当矩阵 E 的元素服从独立同分布的高斯分布时, 可用经典的 PCA 来获得最优的矩阵 A , 即求解下列最优优化问题:

$$\min_{A, E} \|E\|_F \quad \text{s. t. } \text{rank}(A) \leq r, D = A + E \quad (7)$$

对矩阵 D 进行 SVD 便可得到上述优化问题的最优解。但当 E 为稀疏的大噪声矩阵时, PCA 不再适用。此时, 恢复低秩矩阵 A 为一个双目标优化问题:

$$\min_{A, E} (\text{rank}(A) + \|E\|_0) \quad \text{s. t. } D = A + E \quad (8)$$

通过引入折中因子 $\lambda (> 0)$, 将双目标优化问题式(8)转换为如下的单目标优化问题:

$$\min_{A, E} \text{rank}(A) + \lambda \|E\|_0 \quad \text{s. t. } A + E = D \quad (9)$$

优化问题式(9)是 NP 难的, 因此需要对此优化问题的目标函数进行松弛。由于矩阵的核范数是矩阵秩的包络, 矩阵的 $(1, 1)$ 范数是 0 范数的凸包, 故将问题式(9)松弛到如下凸优化问题:

$$\min_{A, E} \|A\|_* + \lambda \|E\|_{1,1} \quad \text{s. t. } A + E = D \quad (10)$$

在实际计算中, 文献[8]建议取 $\lambda = 1/\sqrt{\max(m, n)}$ 。优化问题式(10)也称为鲁棒主成分分析。下面简要介绍求解 RP-CA 的几种主要的算法。

2.1 迭代阈值算法^[8,16,17]

将最优化问题式(10)正则化, 便得到优化问题:

$$\min_{A, E} \|A\|_* + \lambda \|E\|_{1,1} + \mu (\|A\|_F^2 + \|E\|_F^2)/2 \quad (11)$$

s. t. $A + E = D$

其中: 参数 μ 为比较小的正数。优化问题式(11)的拉格朗日 (Lagrange) 函数为

$$L(A, E, Y) = \|A\|_* + \lambda \|E\|_{1,1} + \mu (\|A\|_F^2 + \|E\|_F^2)/2 + \langle Y, D - A - E \rangle \quad (12)$$

使用迭代阈值算法 (iterative thresholding, IT) 交替更新矩阵 A 、 E 和 Y 。当 $E = E_k, Y = Y_k$ 时,

$$A_{k+1} = \arg \min_A L(A, E_k, Y_k) = \arg \min_A \|A\|_* / \mu + \|A - Y_k / \mu\|_F^2/2 = D_{1/\mu}(Y_k / \mu) = \mu^{-1} D_1(Y_k) \quad (13)$$

当 $A = A_{k+1}, Y = Y_k$ 时,

$$E_{k+1} = \arg \min_E L(A_{k+1}, E, Y_k) = \arg \min_E \lambda \|E\|_{1,1} / \mu + \|E - Y_k / \mu\|_F^2/2 = S_{\lambda/\mu}(Y_k / \mu) = \mu^{-1} S_\lambda(Y_k) \quad (14)$$

当 $A = A_{k+1}, E = E_{k+1}$ 时,

$$Y_{k+1} = Y_k + \delta_k (D - A_{k+1} - E_{k+1}) \quad (15)$$

其中: 步长 δ_k 满足 $0 < \delta_k < 1$ 。

IT 算法的迭代形式简单且收敛, 但它的收敛速度比较慢, 且难以选取合适的步长 δ_k ^[18]。因此, IT 算法具有有限的应用范围。

2.2 加速近端梯度算法^[19]

将优化问题式(10)的等式约束松弛到目标函数中, 得到如下的拉格朗日函数:

$$L(A, E, \mu) = \mu (\|A\|_* + \lambda \|E\|_{1,1}) + \|D - A - E\|_F^2/2 \quad (16)$$

基于式(16), 文献[19]提出使用加速近端梯度算法 (accelerated proximal gradient, APG) 求解优化问题式(10)。

记 $g(A, E, \mu) = \mu (\|A\|_* + \lambda \|E\|_{1,1})$, $f(A, E) = \|D - A - E\|_F^2/2$, 于是 $L(A, E, \mu) = g(A, E, \mu) + f(A, E)$ 。函数 $g(A, E, \mu)$ 不可微, 而 $f(A, E)$ 光滑且具有李普希兹 (Lipschitz) 连续梯度, 即存在 $L_f > 0$, 使得

$$\|\nabla f(A, E) - \nabla f(\tilde{A}, \tilde{E})\|_F \leq L_f \|(A - \tilde{A}, E - \tilde{E})\|_F$$

其中: $\nabla f(A, E)$ 表示函数 $f(A, E)$ 关于矩阵变量 A 和 E 的 Fréchet 梯度。此处取 $L_f = 2$ 。

对于给定的与 D 同型的两个矩阵 Y_A 和 Y_E , 作 $L(A, E, \mu)$ 的部分二次逼近:

$$Q(A, E, \mu, Y_A, Y_E) = g(A, E, \mu) + f(Y_A, Y_E) + \langle \nabla f(Y_A, Y_E), (A - Y_A, E - Y_E) \rangle + L_f \|(A - Y_A, E - Y_E)\|_F^2/2 = g(A, E, \mu) + f(Y_A, Y_E) + \langle Y_A - (D - Y_E), A - Y_A \rangle + \langle Y_E - (D - Y_A), E - Y_E \rangle + L_f \|A - Y_A\|_F^2/2 + L_f \|E - Y_E\|_F^2/2 \quad (17)$$

当 $E = E_k, Y_A = Y_A^k, Y_E = Y_E^k, \mu = \mu_k$ 时,

$$A_{k+1} = \arg \min_A Q(A, E_k, \mu_k, Y_A^k, Y_E^k) = \arg \min_A \mu_k \|A\|_* + \langle Y_A^k - (D - Y_E^k), A \rangle + L_f \|A - Y_A^k\|_F^2/2 = \arg \min_A \mu_k \|A\|_* + L_f \|A - (Y_A^k + (D - Y_A^k - Y_E^k)/L_f)\|_F^2/2 = D_{\mu_k/L_f}(Y_A^k + (D - Y_A^k - Y_E^k)/L_f) \quad (18)$$

类似地, 当 $A = A_{k+1}, Y_A = Y_A^k, Y_E = Y_E^k, \mu = \mu_k$ 时,

$$E_{k+1} = \arg \min_E Q(A_{k+1}, E, \mu_k, Y_A^k, Y_E^k) =$$

$$D_{\mu_k/L_f}(Y_k^i + (D - Y_A^i - Y_E^i)/L_f) \quad (19)$$

为了得到更新 Y_A 和 Y_E 时的步长,需先确定参数 t_{k+1} :

$$t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2 \quad (20)$$

于是 Y_A 和 Y_E 的迭代更新公式为

$$Y_A^{k+1} = A_k + (t_k - 1)(A_k - A_{k+1})/t_{k+1} \quad (21)$$

$$Y_E^{k+1} = E_k + (t_k - 1)(E_k - E_{k+1})/t_{k+1} \quad (22)$$

参数 μ 的迭代公式为

$$\mu_{k+1} = \max(\eta\mu_k, \tilde{\mu}) \quad (23)$$

其中: $\tilde{\mu}$ 为事先给定的正数; $0 < \eta < 1$ 。

尽管 APG 与 IT 算法类似,但它却大大降低了迭代次数。

2.3 对偶方法^[19]

由于核范数的对偶范数为谱范数,所以优化问题式(10)的对偶问题为

$$\max_Y \langle D, Y \rangle \quad \text{s. t. } J(Y) \leq 1 \quad (24)$$

其中: $J(Y) = \max(\|Y\|_2, \|Y\|_\infty)$; $\|\cdot\|_\infty$ 表示矩阵元素绝对值最大的值。当优化问题式(24)取得最优值 Y^* 时,必定满足 $J(Y^*) = 1$,即此优化问题等价于

$$\max_Y \langle D, Y \rangle \quad \text{s. t. } J(Y) = 1 \quad (25)$$

上述优化问题是非线性、非光滑的,可以使用最速上升法求解。当 $Y = Y_k$ 时,定义正规锥 $N(Y_k) = \{aX; a \geq 0, X \in \partial J(Y_k)\}$,其中 $\partial J(\cdot)$ 表示函数 $J(\cdot)$ 的次梯度。此时,优化问题式(25)的最速上升方向为 $W_k = D - D_k$,其中 D_k 为 D 在 $N(Y_k)$ 上的投影。使用线性搜索方法确定步长大小:

$$\delta_k = \arg \max_{\delta \geq 0} \langle D, (Y_k + \delta W_k) / J(Y_k + \delta W_k) \rangle \quad (26)$$

于是 Y_k 的更新过程为

$$Y_{k+1} = (Y_k + \delta_k W_k) / J(Y_k + \delta_k W_k) \quad (27)$$

上述方法称为对偶方法(DUAL),它比 APG 算法具有更好的可扩展性,这是因为在每次迭代过程中对偶方法不需要矩阵的完全奇异值分解。

2.4 增广拉格朗日乘子法

文献[18]使用增广拉格朗日乘子法(augmented Lagrange multipliers, ALM)求解优化问题式(10)。构造增广拉格朗日函数:

$$L(A, E, Y, \mu) = \|A\|_* + \lambda \|E\|_{1,1} + \langle Y, D - A - E \rangle + \mu \|D - A - E\|_F^2/2 \quad (28)$$

当 $Y = Y_k, \mu = \mu_k$,使用交替式方法求解优化问题 $\min_{A, E} L(A, E, Y_k, \mu_k)$ 。

使用精确拉格朗日乘子法(exact ALM, EALM)交替迭代矩阵 A 和 E ,直到满足终止条件为止。若 $E = E_{k+1}^j$,则

$$\begin{aligned} A_{k+1}^{j+1} &= \arg \min_A L(A, E_{k+1}^j, Y_k, \mu_k) = \\ &= \arg \min_A \|A\|_* + \mu_k \|A - (D - E_{k+1}^j + Y_k/\mu_k)\|_F^2/2 = \\ &= D_{1/\mu_k}(D - E_{k+1}^j + Y_k/\mu_k) \end{aligned} \quad (29)$$

再根据得到的 A_{k+1}^{j+1} 更新矩阵 E :

$$\begin{aligned} E_{k+1}^{j+1} &= \arg \min_E L(A_{k+1}^{j+1}, E, Y_k, \mu_k) = \\ &= \arg \min_E \lambda \|E\|_{1,1} + \mu_k \|E - (D - A_{k+1}^{j+1} + Y_k/\mu_k)\|_F^2/2 = \\ &= S_{\lambda/\mu_k}(D - A_{k+1}^{j+1} + Y_k/\mu_k) \end{aligned} \quad (30)$$

记 $A_{k+1}^{j+1}, E_{k+1}^{j+1}$ 分别收敛于 A_{k+1}^*, E_{k+1}^* ,则矩阵 Y 的更新公式为

$$Y_{k+1} = Y_k + \mu_k (D - A_{k+1}^* - E_{k+1}^*) \quad (31)$$

最后更新参数 μ :

$$\mu_{k+1} = \begin{cases} \rho\mu_k & \text{若 } \mu_k \|E_{k+1}^* - E_k^*\|_F / \|D\|_F < \varepsilon \\ \mu_k & \text{否则} \end{cases} \quad (32)$$

其中: $\rho > 1$ 为常数; $\varepsilon > 0$ 为比较小的正数。

不精确拉格朗日乘子法(inexact ALM, IALM)改善了 EALM,它不需要 $\min_{A, E} L(A, E, Y_k, \mu_k)$ 的精确解,即矩阵 A 和 E 的迭代更新公式为

$$A_{k+1} = \arg \min_A L(A, E_{k+1}, Y_k, \mu_k) = D_{1/\mu_k}(D - E_{k+1} + Y_k/\mu_k) \quad (33)$$

$$E_{k+1} = \arg \min_E L(A_{k+1}, E, Y_k, \mu_k) = S_{\lambda/\mu_k}(D - A_{k+1} + Y_k/\mu_k) \quad (34)$$

IALM 也称为交替方向方法(alternating direction methods, ADM)^[20]。

数值实验证实 ALM 算法比 APG 算法快很多,而且它可以达到较高的精度,需要较低的存储空间^[18]。下面考虑将 RPCA 应用到 Lobby 视频(http://perception.izir.a-star.edu.sg/bk_model/bk_index.html)的背景建模中^[21]。在 Lobby 视频中,有几次剧烈的照明变化,这里仅考虑其中前 200 帧灰度图像序列,而每帧图像的分辨率为 128×160 。这些图像组成的数据矩阵 D 的维数为 20480×200 。使用 IALM 算法对矩阵 D 进行低秩恢复,部分实验结果如图 1 所示。

图 1 中,(a)给出了 Lobby 视频中的两帧原始图像,(b)为相应的低秩图像,(c)为稀疏的背景图像。从图 1 可以看出,低秩图像正确识别出了主要的照明背景,同时稀疏成分对应场景中人的移动。

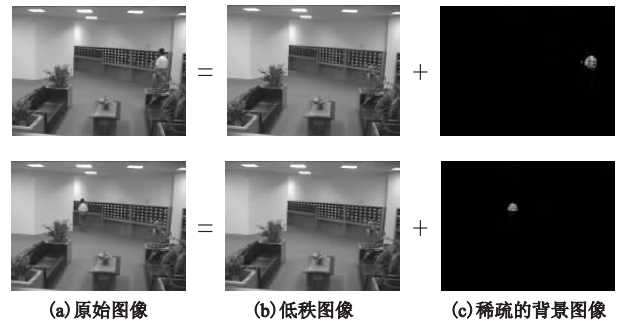


图1 Lobby视频背景建模

3 矩阵补全

当数据矩阵 D 含丢失元素时,可根据矩阵的低秩结构来恢复矩阵的所有元素,称此恢复过程为矩阵补全(MC)。Candès 等人^[10,11]证明了当矩阵的奇异值及采样数目满足一定的条件时,大多数低秩矩阵可以通过求解简单的凸优化问题(即半定规划)而以较大的概率近乎完美地恢复。文献[22~24]对 MC 的理论与算法作了进一步探讨。

记 Ω 为集合 $[m] \times [n]$ 的子集,这里 $[m]$ 表示集合 $\{1, 2, \dots, m\}$ 。MC 的原始模型可描述为如下的优化问题:

$$\min_A \text{rank}(A) \quad \text{s. t. } P_\Omega(A) = P_\Omega(D) \quad (35)$$

其中: $P_\Omega: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ 为一线性投影算子,即

$$P_\Omega(D_{ij}) = \begin{cases} D_{ij} & (i, j) \in \Omega \\ 0 & (i, j) \notin \Omega \end{cases} \quad (36)$$

为了便于求解优化问题式(35),将矩阵的秩松弛到矩阵的核范数,即得到凸优化问题:

$$\min_A \|A\|_* \quad \text{s. t. } P_\Omega(A) = P_\Omega(D) \quad (37)$$

上述优化问题等价于半定规划问题:

$$\begin{aligned} & \min_A (\langle W_1, I_m \rangle + \langle W_2, I_n \rangle) / 2 \\ \text{s. t. } & P_\Omega(A) = P_\Omega(D), \begin{pmatrix} W_1 & A \\ A^T & W_2 \end{pmatrix} \succeq 0 \end{aligned} \quad (38)$$

其中: $\succeq$ 为半正定符号。内点算法是求解半定规划的主要算法,但其计算复杂度高达 $O(p(m+n)^3 + p^2(m+n)^2 + p^3)$, 这里 p 表示矩阵 D 的观测元素的数目。现有的求解半定规划的 SDPT3、SeDuMi 和 CVX 等软件包能求解的矩阵维数一般不超过 100×100 维,因此半定规划模型不适用于较大维数的矩阵。由于 MC 是 LRMR 的特例,可将 LRMR 的 IALM 算法应用到 MC 问题上。

将优化问题式(37)重新表示为

$$\min_A \|A\|_* \text{ s. t. } A + E = D, P_\Omega(E) = 0 \quad (39)$$

于是构造上述问题的部分增广拉格朗日函数为

$$L(A, E, Y, \mu) = \|A\|_* + \langle Y, D - A - E \rangle + \mu \|D - A - E\|_F^2 / 2 \quad (40)$$

当 $E = E_k, Y = Y_k, \mu = \mu_k$ 时,

$$A_{k+1} = \arg \min_A L(A, E_k, Y_k, \mu_k) = D_{1/\mu_k}(D - E_k + Y_k / \mu_k) \quad (41)$$

当 $A = A_{k+1}, Y = Y_k, \mu = \mu_k$ 时,

$$\begin{aligned} E_{k+1} &= \arg \min_{P_\Omega(E)=0} L(A_{k+1}, E, Y_k, \mu_k) = \\ & P_\Omega(D - A_{k+1} - Y_k / \mu_k) \end{aligned} \quad (42)$$

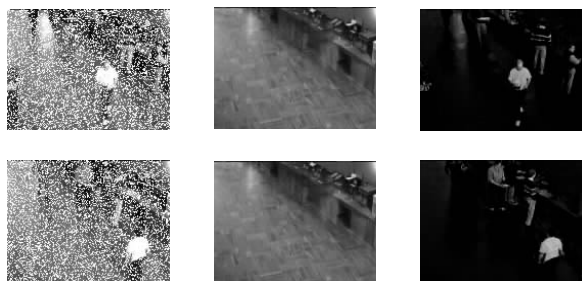
根据 A_{k+1} 和 E_{k+1} 更新 Y :

$$Y_{k+1} = Y_k + \mu_k (D - A_{k+1} - E_{k+1}) \quad (43)$$

最后更新 μ 为 $\mu_{k+1} = \max(\eta \mu_k, \tilde{\mu})$ 。

在 Bootstrap 视频 (<http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm>) 中进行矩阵补全实验。考虑图像序列的前 200 帧,每帧图像的分辨率为 120×160 ,则图像集合组成的数据矩阵 D 的维数为 19200×200 。对于矩阵 D ,每个元素以 0.3 的概率随机丢失。使用 IALM 算法对不完全矩阵进行低秩恢复,部分实验结果如图 2 所示。

图 2 中,(a) 为含丢失元素的原始图像,其中丢失元素用白色像素表示;(b) 和 (c) 分别表示使用 IALM 算法得到的低秩图像和稀疏成分。从图 2(b)(c) 可以看出,丢失元素并不影响矩阵的低秩恢复,低秩成分正确地识别出了餐馆的背景,而稀疏成分则反映出了人们的活动。



(a) 含丢失元素的原始图像 (b) 低秩图像 (c) 稀疏成分
图2 基于矩阵补全的Bootstrap视频背景建模

4 低秩表示

低秩矩阵表示(LRR)是将数据集矩阵 D 表示成字典矩阵 B (也称为基矩阵)下的线性组合,即 $D = BZ$,并希望线性组合系数矩阵 Z 是低秩的^[12]。为此,需要求解下列优化问题:

$$\min_Z \text{rank}(Z) \text{ s. t. } D = BZ \quad (44)$$

将优化问题式(44)进行凸松弛,得到

$$\min_Z \|Z\|_* \text{ s. t. } D = BZ \quad (45)$$

若选取数据集 D 本身作为字典,则有

$$\min_Z \|Z\|_* \text{ s. t. } D = DZ \quad (46)$$

为了对噪声和野点更加鲁棒,一个更合理的模型为

$$\min_{Z, E} \|Z\|_* + \lambda \|E\|_{2,1} \text{ s. t. } D = DZ + E \quad (47)$$

在求解优化问题式(47)时,引入新的矩阵 X ,即此优化问题等价于

$$\min_{Z, E, X} \|X\|_* + \lambda \|E\|_{2,1} \text{ s. t. } D = DZ + E, Z = X \quad (48)$$

构造上述优化问题的增广拉格朗日乘子函数为

$$\begin{aligned} L(Z, E, X, Y_1, Y_2, \mu) &= \|X\|_* + \lambda \|E\|_{2,1} + \langle Y_1, D - DZ - E \rangle + \\ & \langle Y_2, Z - X \rangle + \mu (\|D - DZ - E\|_F^2 + \|Z - X\|_F^2) / 2 \end{aligned} \quad (49)$$

当 $Z = Z_k, E = E_k, Y_1 = Y_1^k, Y_2 = Y_2^k, \mu = \mu_k$ 时, X 的更新公式为

$$\begin{aligned} X_{k+1} &= \arg \min_X \|X\|_* / \mu_k + \|X - (Z_k + Y_2^k / \mu_k)\|_F^2 / 2 = \\ & D_{1/\mu_k}(Z_k + Y_2^k / \mu_k) \end{aligned} \quad (50)$$

Z 的更新公式为

$$\begin{aligned} Z_{k+1} &= (I_n + D^T D)^{-1} (D^T D - D^T E_k + X_{k+1} + \\ & (D^T Y_1^k - Y_2^k) / \mu_k) \end{aligned} \quad (51)$$

E 的更新公式为

$$\begin{aligned} E_{k+1} &= \arg \min_E \lambda \|E\|_{2,1} / \mu_k + \|E - (D - DZ_{k+1} + Y_1^k / \mu_k)\|_F^2 / 2 = \\ & T_{\lambda/\mu_k}(D - DZ_{k+1} + Y_1^k / \mu_k) \end{aligned} \quad (52)$$

拉格朗日乘子的迭代公式为

$$Y_1^{k+1} = Y_1^k + \mu_k (D - DZ_{k+1} - E_{k+1}) \quad (53)$$

$$Y_2^{k+1} = Y_2^k + \mu_k (Z_{k+1} - X_{k+1}) \quad (54)$$

参数 μ 的更新公式为 $\mu_{k+1} = \max(\eta \mu_k, \tilde{\mu})$ 。

下面在人工数据集上进行实验。随机生成正交矩阵 $U_i \in \mathbb{R}^{100 \times 4}$ 和旋转矩阵 $T \in \mathbb{R}^{100 \times 100}$ 。再由 U_i 和 T 可构造四个正交矩阵: $U_{i+1} = TU_i, i = 1, 2, 3, 4$ 。以 U_i 的列向量为基,可构造出 20 个数据向量 $X_{i+1} = U_i Q_i, i = 1, \dots, 5$, 其中 $Q_i \in \mathbb{R}^{4 \times 20}$ 的各元素是由相互独立的标准正态分布随机生成。记 $X = [X_1, \dots, X_5]$, 随机选取 X 的 20 个列向量,对所选取的列向量 x 的每个元素添加均值为 0、标准差为 $0.1 \|x\|_2$ 的高斯噪声。使用 LRR 方法得到低秩表示矩阵 Z , 根据 Z 可构造相似度矩阵 $S = (S_{ij})_{100 \times 100}$, 其中 $S_{ij} = (|Z_{ij}| + |Z_{ji}|) / 2$ 。文献[14]提出了构造相似度矩阵的另外一种方式:先对矩阵 Z 进行 SVD, 即 $Z = U \Sigma V^T$;再构造矩阵 $\tilde{U} = U(\Sigma)^{1/2}$;最后得到相似度矩阵 $S = \tilde{U} \tilde{U}^T$ 。本文选取第一种相似度构造方式,实验结果如图 3 所示。

图 3 依次绘出了相似度矩阵、原始噪声矩阵和恢复噪声矩阵。从图 3(a) 可以看出,相似度矩阵的块效应非常显著,这与给定的五个子空间相吻合。通过比较(b)和(c)可以看出, LRR 方法近乎完美地恢复了噪声。此实验说明 LRR 在具有良好的去噪性能的同时,还对子空间分割具有鲁棒性。

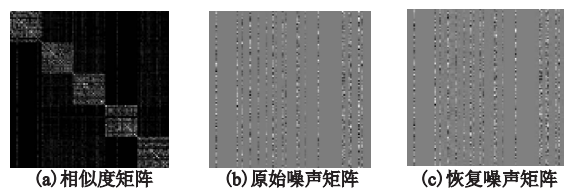


图3 人工数据集的LRR

5 结束语

本文简要综述了低秩矩阵恢复的各种优化模型及相应的算法。对于鲁棒主成分分析,系统地总结了迭代阈值算法、加速近端梯度算法、对偶算法和增广拉格朗日乘子法,并评价了这些算

法的优缺点;对于矩阵补全问题,探讨了它的优化模型,分析了求解此模型的增广拉格朗日乘子法;最后讨论了矩阵的低秩表示模型,并列出了相应的迭代算法。本文使用的部分 MATLAB 代码来自 <http://perception.csl.illinois.edu/matrix-rank/home.html>。低秩矩阵恢复在理论上还不完善,在实际应用中还不成熟。因此,对低秩矩阵恢复的研究仍将是未来的一个研究热点,下面的几个方向在今后的研究中值得关注。

a)对可扩展快速算法的研究。随着计算机与传感器技术的快速发展,人们面对的数据往往具有规模大、维数高的特点,设计快速、高效的算法是当前迫切需要的。为了提高算法的性能,需要建立更优的模型。此外在现有的算法中,矩阵奇异值分解是最耗时的,因此设计部分奇异值分解算法将大大提高算法的速度。

b)将低秩矩阵恢复推广到高阶张量情形^[25~28]。在实际应用中,图像和视频等数据具有多线性性,在使用传统的数据处理方法时需先将它们表示为向量,这种向量化破坏了原始数据固有的空时结构。而低秩张量恢复能充分利用数据的空时结构,它比低秩矩阵恢复具有更好的性能。对于低秩张量恢复,模型建立与算法设计仍值得进一步研究。

c)对低秩矩阵恢复模型的研究。根据实际问题的需要,构造相应的低秩恢复模型。如在鲁棒主成分分析模型中,可将给定的数据矩阵表示为低秩矩阵、稀疏矩阵与小的稠密噪声矩阵之和^[29];在提取 3D 场景的低秩纹理中,考虑变换不变性^[30]。

参考文献:

- [1] 刘青山,卢汉清,马颂德. 综述人脸识别中的子空间方法[J]. 自动化学报, 2003, 29(6): 900-911.
- [2] JIANG Xu-dong. Linear subspace learning-based dimensionality reduction[J]. IEEE Signal Processing Magazine, 2011, 28(2): 16-26.
- [3] DONOHO D L. Compressed sensing[J]. IEEE Trans on Information Theory, 2006, 52(4): 1289-1306.
- [4] CANDES E J, MICHAEL W. An introduction to compressive sampling [J]. IEEE Signal Processing Magazine, 2008, 25(2): 21-30.
- [5] WRIGHT J, ALLEN Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [6] WRIGHT J, MA Yi, MAIRAL J, et al. Sparse representation for computer vision and pattern recognition [J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044.
- [7] WRIGHT J, GANESH A, RAO S, et al. Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization[C]//Proc of the 24th Annual Conference on Neural Information Processing Systems. 2009: 2080-2088.
- [8] CANDES E J, LI Xiao-dong, MA Yi, et al. Robust principal component analysis? [J]. Journal of the ACM, 2011, 58(3): 1-37.
- [9] XU Huan, CARAMANIS C, SANGHAVI S. Robust PCA via outlier pursuit[J]. IEEE Trans on Information Theory, 2012, 58(5): 3047-3064.
- [10] CANDES E J, RECHT B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6): 717-772.
- [11] CANDES E J, TAO T. The power of convex relaxation: Near-optimal matrix completion[J]. IEEE Trans on Information Theory, 2010, 56(5): 2053-2080.
- [12] LIU Guang-can, LIN Zhou-chen, YU Yong. Robust subspace segmentation by low-rank representation[C]//Proc of International Conference on Machine Learning. 2010: 663-670.
- [13] LIU Guang-can, XU Huan, YAN Shui-cheng. Exact subspace segmentation and outlier detection by low-rank representation[C]//Proc of International Conference on Artificial Intelligence and Statistics. 2012: 703-711.
- [14] LIU Guang-can, LIN Zhou-chen, YAN Shui-cheng, et al. Robust recovery of subspace structures by low-rank representation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171-184.
- [15] 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.
- [16] CAI Jian-feng, CANDES E J, SHEN Zuo-wei. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.
- [17] BECK A, TEBoulLE M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1): 183-202.
- [18] LIN Zhou-chen, CHEN Min-ming, MA Yi, et al. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices[EB/OL]. (2010-09-26). <http://arxiv.org/pdf/1009.5055>.
- [19] LIN Zhou-chen, GANESH A, WRIGHT J, et al. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix [EB/OL]. (2009). http://yima.csl.illinois.edu/psfile/rpca_algorithms.pdf.
- [20] YUAN Xiao-ming, YANG Jun-feng. Sparse and low-rank matrix decomposition via alternating direction methods [EB/OL]. (2009). http://www.optimization-online.org/DB_FILE/2009/11/2447.pdf.
- [21] LI Li-yuan, HUANG Wei-min, GU I Yu-hua, et al. Statistical modeling of complex background for foreground object detection [J]. IEEE Trans on Image Processing, 2004, 13(11): 1459-1472.
- [22] KESHAVAN A H, MONTANARI A, OH S. Matrix completion from noisy entries[J]. Journal of Machine Learning Research, 2010, 11(3): 2057-2078.
- [23] RECHT B. A simpler approach to matrix completion[J]. Journal of Machine Learning Research, 2011, 12(7): 3413-3430.
- [24] CANDES E J, PLAN Y. Matrix completion with noise[J]. Proceedings of the IEEE, 2010, 98(6): 925-936.
- [25] LIU Ji, MUSIALSKI P, WONKA P, et al. Tensor completion for estimating missing values in visual data[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(1): 208-220.
- [26] SIGNORETTO M, De LATHAUWER L, SUYKENS J A K. Nuclear norms for tensors and their use for convex multilinear estimation[EB/OL]. (2010-09-30). ftp://wgs.esat.kuleuven.ac.be/sista/signoretto/Signoretto_nucTensors.pdf.
- [27] 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.
- [28] 史加荣, 杨威, 姜淑艳. 非负张量分解的快速算法[J]. 计算机应用研究, 2011, 28(12): 4475-4477.
- [29] WRIGHT J, MA Yi. The pursuit of low-dimensional structures in high-dimensional (visual) data: formulation and theory[EB/OL]. (2011). <http://www.stanford.edu/group/mmds/slicles2012/s-ma.pdf>.
- [30] ZHANG Zheng-dong, GANESH A, LIANG Xiao, et al. TILT: transform-invariant low-rank textures[J]. International Journal of Computer Vision, 2012, 99(1): 1-24.