矩阵补全算法研究进展

史加荣[1] 郑秀云[1] 周水生[2]
(西安建筑科技大学理学院 西安 710055)[1] (西安电子科技大学数学与统计学院 西安 710071)[2]

摘 要 作为压缩感知理论的重要发展,矩阵补全与恢复已成为信号与图像处理的一种新的强有力的工具。综述了矩阵补全算法的最新研究进展。首先分析了核范数最小化模型的几种主要的矩阵补全算法,并对这些算法的迭代过程及原理进行了详细的阐述。其次讨论了矩阵补全的低秩矩阵分解模型,并列出了近年来出现的求解此模型的新算法。然后补充了上述两种模型的衍生版本,指出了相应的求解方法。在数值实验中,对文中所讨论的主要矩阵补全算法的性能进行了比较。最后给出了矩阵补全算法的未来研究方向及重点。

关键词 矩阵补全,低秩,核范数最小化,低秩矩阵分解,压缩感知,低秩矩阵恢复
中图法分类号 TP391 **文献标识码** A

Research Progress in Matrix Completion Algorithms

SHI Jia-rong[1] ZHENG Xiu-yun[1] ZHOU Shui-sheng[2]
(School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China)[1]
(School of Mathematics and Statistics, Xidian University, Xi'an 710071, China)[2]

Abstract As an important development of compressed sensing theory, matrix completion and recovery has been a new and remarkable technique for signal and image processing. This paper made a survey on the latest research progress in matrix completion algorithms. Firstly, it analyzed several main algorithms to nuclear norm minimization model, and elaborated their iterative procedure and principle. Secondly, it discussed low-rank matrix factorization model of matrix completion and listed the corresponding new algorithms emerged in recent years. Then it complemented other versions derived from the above two models and pointed out the solving methods. In numerical experiments, performance comparisons were made on the main algorithms to matrix completion. Finally, it gave future research direction and focus for matrix completion algorithms.

Keywords Matrix completion, Low-rank, Nuclear norm minimization, Low-rank matrix factorization, Compressed sensing, Low-rank matrix recovery

1 引言

一般来说,不能根据信号的部分采样元素来恢复所有元素。但信号在一组基下是稀疏的且满足一定条件时,压缩感知理论(Compressed Sensing/Compressed Sampling, CS)证实了可以通过求解 l_1 最小化问题来精确地恢复所有元素[1,2]。当信号用矩阵形式表示时,同样一般不能根据它的部分元素来恢复所有丢失元素。Candès 和陶哲轩等人证明了当矩阵的奇异值具有稀疏性(即矩阵是低秩的)且采样数目满足一定条件时,大多数矩阵可以通过求解核范数最小化问题来精确地恢复所有元素[3]。由矩阵的部分元素来恢复所有元素这一问题也称为矩阵补全(Matrix Completion, MC)。若将部分采样元素这一约束推广到一般的线性约束函数,则矩阵补全就称为低秩矩阵恢复(Low-Rank Matrix Recovery, LRMR)[4,5]。

矩阵补全广泛地应用在计算机视觉[6-9]、机器学习[10,11]、推荐系统[12]和系统辨识[13,14]等诸多科学与工程领域中。作为信号与图像处理技术的一个强大的新兴分支,矩阵补全已成为继压缩感知之后的又一种重要的信号获取工具。在数学形式上,矩阵补全可描述为一个仿射秩最小化问题,但此问题是 NP 难的。近年来涌现出了许多求解秩最小化问题的启发式方法,这些方法主要分为两类:一类是将秩函数凸松弛到矩阵核范数,建立核范数优化模型;另一类是事先给定矩阵的秩,建立低秩分解模型。

与向量的 l_1 范数类似,矩阵的核范数也是一个连续的、不可微的凸函数。梯度或次梯度下降法是求解核范数优化模型的主要方法。矩阵核范数最小化模型等价于半定规划问题[15,16],求解半定规划常用的方法是基于二阶梯度信息的内点算法[17],但此方法计算复杂度非常高。因此,求解压缩感知的众多基于一阶梯度的方法被纷纷应用到矩阵核范数最小

到稿日期:2013-06-02 返修日期:2013-09-17 本文受国家自然科学基金(61179040),陕西省教育厅专项科研计划项目(2013JK0587, 2013JK0588, 2010JK642)资助。

史加荣(1979—),男,博士,副教授,CCF 会员,主要研究方向为机器学习与模式识别,E-mail:shijiarong@xauat.edu.cn;**郑秀云**(1982—),女,博士,讲师,主要研究方向为最优化方法与应用;**周水生**(1972—),男,博士,教授,博士生导师,主要研究方向为机器学习与模式识别。

化问题中。低秩矩阵分解是矩阵补全的另一种重要的模型，它先将数据矩阵近似分解为两个低秩矩阵之积，再通过求解非凸的低秩逼近问题来恢复丢失元素。求解这类模型的经典方法有 Wiberg 算法、阻尼牛顿算法、Levenberg-Marquardt 算法和幂迭代法等[6,7]。由于低秩矩阵分解模型是 Grassmann 或 Riemann 流形上的优化问题，因此也可以使用几何子空间方法求解。

本文对矩阵补全算法的最新研究进展进行了综述。在第 2 节，引入了矩阵补全算法所需的矩阵算子和核范数等方面的基础知识；第 3 节分析了求解核范数最小化问题的主要算法；第 4 节讨论了求解低秩矩阵分解模型的最新方法；第 5 节补充了前两种模型的推广版本；在第 6 节，通过数值实验来比较矩阵补全的主要算法的性能；最后对矩阵补全算法进行总结，并给出了进一步的研究方向和重点。

2 基本知识

先给出矩阵分解的一种重要且常用的形式：奇异值分解(Singular Value Decomposition，SVD)。

定理 1 对于秩为 r 的矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$，它的奇异值分解为：$\boldsymbol{X}=\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{U}\in\mathbb{R}^{m\times r}$ 和 $\boldsymbol{V}\in\mathbb{R}^{n\times r}$ 满足 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{I}_r$（$\boldsymbol{I}_r$ 为 r 阶单位矩阵），$\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)$ 且其对角线元素满足 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$。

称 σ_i 为 $\boldsymbol{X}$ 的第 i 大奇异值，$i=1,2,\cdots,r$。矩阵 $\boldsymbol{U}$ 有 mr 个元素，又由于 $\boldsymbol{U}$ 的各个列向量具有规范正交性，因此它的自由度为 $mr-(1+2+\cdots+r)$。类似地，$\boldsymbol{V}$ 的自由度为 $nr-(1+2+\cdots+r)$。故 $\boldsymbol{X}$ 的自由度为

$$dr(X)=(mr-(1+2+\cdots+r))+(nr-(1+2+\cdots+r))+r=(m+n-r)r \tag{1}$$

定义 1 矩阵 $\boldsymbol{X},\boldsymbol{Y}\in\mathbb{R}^{m\times n}$，它们的内积为 $\langle\boldsymbol{X},\boldsymbol{Y}\rangle=\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}Y_{ij}$，其中 X_{ij} 表示矩阵 $\boldsymbol{X}$ 的第 i 行第 j 列元素。

定义 2 设 $\sigma_1,\sigma_2,\cdots,\sigma_r$ 为矩阵 $\boldsymbol{X}$ 的所有非零奇异值，则它的 Schatten p-范数($p>0$)为 $\|\boldsymbol{X}\|_{Sp}=(\sum_{i=1}^{r}\sigma_i^p)^{1/p}$。

特殊地，当 $p=1$ 时，$\|\boldsymbol{X}\|_{S1}$ 为 $\boldsymbol{X}$ 的核范数(也称为迹范数、Ky Fan 范数)，可用 $\|\boldsymbol{X}\|_*$ 表示；当 $p=2$ 时，$\|\boldsymbol{X}\|_{S2}$ 为 X 的 Frobenious 范数，可用 $\|\boldsymbol{X}\|_F$ 表示；当 $p\to+\infty$ 时，$\|\boldsymbol{X}\|_{S\infty}=\sigma_1$ 为 $\boldsymbol{X}$ 的算子范数(或谱范数)，可用 $\|\boldsymbol{X}\|$ 表示；规定 $\|\boldsymbol{X}\|_{S0}=r$，可用 $\mathrm{rank}(\boldsymbol{X})$ 表示。矩阵 $\boldsymbol{X}$ 的 Frobenious 范数满足：$\|\boldsymbol{X}\|_F=\sqrt{\langle\boldsymbol{X},\boldsymbol{X}\rangle}=\sqrt{\mathrm{trace}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})}=\sqrt{\sum_{i=1}^{r}\sigma_i^2}$，这里 $\mathrm{trace}(\cdot)$ 表示矩阵的迹算子。矩阵 $\boldsymbol{X}$ 的核范数、Frobenious范数和算子范数满足下列不等式[15]：

$$\|\boldsymbol{X}\|\leqslant\|\boldsymbol{X}\|_F\leqslant\|\boldsymbol{X}\|_*\leqslant\sqrt{r}\|\boldsymbol{X}\|_F\leqslant r\|\boldsymbol{X}\| \tag{2}$$

定义 3 在内积空间中，X 的 Schatten p-范数的对偶范数为

$$\|\boldsymbol{X}\|_{dSp}=\max_{\boldsymbol{Y}}\{\langle\boldsymbol{X},\boldsymbol{Y}\rangle:\|\boldsymbol{Y}\|_{Sp}\leqslant1\} \tag{3}$$

易知 Frobenious 范数的对偶范数仍为 Frobenious 范数，算子范数的对偶范数为核范数。

定义 4 对于多元实函数 $f(\boldsymbol{x})$，$\boldsymbol{x}\in\mathscr{C}$，它的凸包络为满足

$$g(\boldsymbol{x})\leqslant f(\boldsymbol{x}),\ \forall\,\boldsymbol{x}\in\mathscr{C} \tag{4}$$

的最大凸函数 g。

定理 2[16] 秩函数 $\mathrm{rank}(\boldsymbol{X})$ 在集合 $\mathscr{C}=\{\boldsymbol{X}\in\mathbb{R}^{m\times n}:\|\boldsymbol{X}\|\leqslant1\}$ 上的凸包络函数为 $\phi(\boldsymbol{X})=\|\boldsymbol{X}\|_*$。

凸包络函数 $\phi(\boldsymbol{X})$ 是连续的凸函数，但不光滑。若非零矩阵 $\boldsymbol{X}$ 的算子范数 $\|\boldsymbol{X}\|=k$，则由定理 2 可知 $\mathrm{rank}(\boldsymbol{X})$ 的凸包络函数为 $\|\boldsymbol{X}\|_*/k$。

定理 3[18] 函数 $\phi(\boldsymbol{X})=\|\boldsymbol{X}\|_*$ 的次梯度为

$$\partial\|\boldsymbol{X}\|_*=\{\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}+\boldsymbol{M}:\|\boldsymbol{M}\|\leqslant1,\boldsymbol{U}^{\mathrm{T}}\boldsymbol{M}=0,\boldsymbol{M}\boldsymbol{V}=0\} \tag{5}$$

其中，$\boldsymbol{U}\in\mathbb{R}^{m\times r}$、$\boldsymbol{V}\in\mathbb{R}^{n\times r}$ 分别为 $\boldsymbol{X}$ 的左奇异矩阵和右奇异矩阵。

根据定理 3，可以证明下列结论。

定理 4[18] 设秩为 r 的 $m\times n$ 维实矩阵 $\boldsymbol{Q}$ 的奇异值分解为：$\boldsymbol{Q}=\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}=\boldsymbol{U}\,\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_r)\boldsymbol{V}^{\mathrm{T}}$，则核范数最小化问题 $\min_{\boldsymbol{X}}\|\boldsymbol{X}-\boldsymbol{Q}\|_F^2/2+\varepsilon\|\boldsymbol{X}\|_*$ 有闭形式解 $\boldsymbol{X}^*=\mathscr{D}_\varepsilon(\boldsymbol{Q})=\boldsymbol{U}\mathscr{S}_\varepsilon(\Sigma)\boldsymbol{V}^{\mathrm{T}}$，其中 $\varepsilon\geqslant0$，$\mathscr{S}_\varepsilon(\Sigma)=\mathrm{diag}(\max(\sigma_1-\varepsilon,0),\cdots,\max(\sigma_r-\varepsilon,0))$。

3 核范数优化模型的矩阵补全算法

3.1 问题描述

对于低秩矩阵 $\boldsymbol{M}\in\mathbb{R}^{m\times n}$，希望根据它的部分观测元素来恢复所有元素，此恢复过程可用下列仿射秩最小化问题来描述：

$$\min_{\boldsymbol{X}}\mathrm{rank}(\boldsymbol{X}),\ \mathrm{s.t.}\ X_{ij}=M_{ij},(i,j)\in\Omega \tag{6}$$

其中，$\boldsymbol{X}\in\mathbb{R}^{m\times n}$，指标集 $\Omega\subset\{1,2,\cdots,m\}\times\{1,2,\cdots,n\}$。记观测或采样元素数目为 p，则 $p=|\Omega|$，这里 $|\cdot|$ 表示集合的势。在上述优化问题中，约束条件也可以表示为 $\mathscr{P}_\Omega(\boldsymbol{X})=\mathscr{P}_\Omega(\boldsymbol{M})$ 或 $\mathscr{A}(\boldsymbol{X})=\boldsymbol{b}$，其中 $\mathscr{P}_\Omega$ 为投影算子，其定义为

$$\mathscr{P}_\Omega(X_{ij})=\begin{cases}X_{ij}, & \text{若}(i,j)\in\Omega\\ 0, & \text{否则}\end{cases} \tag{7}$$

$\mathscr{A}:\mathbb{R}^{m\times n}\to\mathbb{R}^p$ 为线性算子，$\boldsymbol{b}\in\mathbb{R}^p$。

矩阵的秩函数是非连续、非凸的，故直接求解秩最小化问题比较困难。根据定理 2，可将秩函数凸松弛到核范数，于是得到下列凸优化问题

$$\min_{\boldsymbol{X}}\|\boldsymbol{X}\|_*,\ \mathrm{s.t.}\ X_{ij}=M_{ij},(i,j)\in\Omega \tag{8}$$

下面给出求解上述核范数最小化问题的主要算法。

3.2 半定规划法

当矩阵 $\boldsymbol{X}$ 给定时，由定义 3 知 $\boldsymbol{X}$ 的核范数等于下列优化问题的最优值

$$\max_{\boldsymbol{Y}}\langle\boldsymbol{X},\boldsymbol{Y}\rangle,\ \mathrm{s.t.}\ \|\boldsymbol{Y}\|\leqslant1 \tag{9}$$

此优化问题也等价于

$$\max_{\boldsymbol{Y}}\mathrm{trace}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})\quad\mathrm{s.t.}\begin{pmatrix}\boldsymbol{I}_m & \boldsymbol{Y}\\ \boldsymbol{Y}^{\mathrm{T}} & \boldsymbol{I}_n\end{pmatrix}\geq0 \tag{10}$$

其中，“$\boldsymbol{A}\geq0$”表示 $\boldsymbol{A}$ 为半正定矩阵。上述半定规划问题的对偶规划为

$$\min_{\boldsymbol{W}_1,\boldsymbol{W}_2}\frac{1}{2}(\mathrm{trace}(\boldsymbol{W}_1)+\mathrm{trace}(\boldsymbol{W}_2))\quad\mathrm{s.t.}\begin{pmatrix}\boldsymbol{W}_1 & \boldsymbol{X}\\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{W}_2\end{pmatrix}\geq0 \tag{11}$$

因此，核范数优化问题(8)等价于下列半定规划

$$\min_{\boldsymbol{X},\boldsymbol{W}_1,\boldsymbol{W}_2}\frac{1}{2}(\mathrm{trace}(\boldsymbol{W}_1)+\mathrm{trace}(\boldsymbol{W}_2))\quad\mathrm{s.t.}\begin{pmatrix}\boldsymbol{W}_1 & \boldsymbol{X}\\ \boldsymbol{X}^{\mathrm{T}} & \boldsymbol{W}_1\end{pmatrix}\geq0,X_{ij}=M_{ij},(i,j)\in\Omega \tag{12}$$

求解半定规划常用的方法是内点算法，使用此方法求解上述最优化问题的计算复杂度为 $O(p(m+n)^3+p^2(m+n)^2+p^3)$。现有的求解半定规划的 MATLAB 软件包主要有 CVX、SDPT3 和 SeDuMi。下面以 CVX 为例，给出求解半定规划(12)的 MATLAB 代码。

```
%输入低秩矩阵 M、采样行指标向量 I 和列指标向量 J
[m,n]=size(M);J=J+m;
cvx_setup
cvx_begin sdp
variables WX(m+n,m+n)
minimize(trace(WX(1:m,1:m))+trace(WX(m+1:m+n,m+1:m+n)))
for i=1:length(I),WX(I(i),J(i))==M(I(i),J(i)-m);end
WX>=0
cvx_end
MR=full(WX(1:m,m+1:end));%MR 为 M 的恢复矩阵
```

由于求解半定规划的内点算法具有非常高的复杂度，因此当 m 或 n 或 p 比较大时，半定规划法需要较长的运行时间，甚至不可行。例如：取 $p=0.3mn$，当 $m=n=100$ 时，运行时间大约 1 分钟；当 $m=n=120$ 时，运行时间大约 5 分钟；当 $m=n\geqslant 200$ 时，MATLAB 会溢出。为此，人们提出了许多基于一阶梯度的算法。

3.3 奇异值阈值算法

Cai Jian-feng 和 Candès 等人提出了求解核范数最小化问题的奇异值阈值算法(Singular Value Thresholding, SVT)[18]。此算法先将最优化问题(8)正则化，即有

$$\min_{\boldsymbol{X}} \tau\|\boldsymbol{X}\|_*+\frac{1}{2}\|\boldsymbol{X}\|_F^2,\ \text{s.t.}\ \mathcal{P}_\Omega(\boldsymbol{X})=\mathcal{P}_\Omega(\boldsymbol{M}) \tag{13}$$

其中，$\tau>0$。当参数 $\tau\to+\infty$ 时，上述最优化问题的最优解收敛到(8)的最优解。构造最优化问题(13)的拉格朗日函数

$$L(\boldsymbol{X},\boldsymbol{Y})=\tau\|\boldsymbol{X}\|_*+\frac{1}{2}\|\boldsymbol{X}\|_F^2+\langle\boldsymbol{Y},\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{X})\rangle \tag{14}$$

其中，拉格朗日乘子 $\boldsymbol{Y}\in\mathbb{R}^{m\times n}$。如果 $(\boldsymbol{X}^*,\boldsymbol{Y}^*)$ 为优化问题(13)的原-对偶问题的最优解，则有

$$\sup_{\boldsymbol{Y}}\inf_{\boldsymbol{X}} L(\boldsymbol{X},\boldsymbol{Y})=\inf_{\boldsymbol{X}}\sup_{\boldsymbol{Y}} L(\boldsymbol{X},\boldsymbol{Y})=L(\boldsymbol{X}^*,\boldsymbol{Y}^*) \tag{15}$$

SVT 算法使用交替迭代方法求解优化问题(13)或(15)。初始化 $\boldsymbol{Y}^0=0$，当 $\boldsymbol{Y}^{k-1}$ 给定时，

$$\begin{aligned}\boldsymbol{X}^k&=\arg\min_{\boldsymbol{X}} L(\boldsymbol{X},\boldsymbol{Y}^{k-1})\\&=\arg\min_{\boldsymbol{X}}\tau\|\boldsymbol{X}\|_*+\frac{1}{2}\|\boldsymbol{X}\|_F^2-\langle\boldsymbol{Y}^{k-1},\mathcal{P}_\Omega(\boldsymbol{X})\rangle\\&=\arg\min_{\boldsymbol{X}}\tau\|\boldsymbol{X}\|_*+\frac{1}{2}\|\boldsymbol{X}\|_F^2-\langle\boldsymbol{X},\mathcal{P}_\Omega(\boldsymbol{Y}^{k-1})\rangle\\&=\mathcal{D}_\tau(\mathcal{P}_\Omega(\boldsymbol{Y}^{k-1}))\end{aligned} \tag{16}$$

当 $\boldsymbol{X}^k$ 给定时，记 $g(\boldsymbol{Y})=L(\boldsymbol{X}^k,\boldsymbol{Y})$，则 $\partial g(\boldsymbol{Y})=\partial_{\boldsymbol{Y}}L(\boldsymbol{X}^k,\boldsymbol{Y})=\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{X}^k)$。使用梯度下降法(Uzawa 算法)来更新 $\boldsymbol{Y}$：

$$\boldsymbol{Y}^k=\boldsymbol{Y}^{k-1}+\delta_k\partial g(\boldsymbol{Y}^{k-1})=\boldsymbol{Y}^{k-1}+\delta_k\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{X}^k) \tag{17}$$

其中，$\delta_k>0$ 为步长大小，$k\geqslant 1$。由更新公式(17)及 $\boldsymbol{Y}$ 的初始化可知迭代公式(16)还等价于 $\boldsymbol{X}^k=\mathcal{D}_\tau(\boldsymbol{Y}^{k-1})$。

上述迭代算法收敛速度往往比较慢，文献[19,20]分别给出了 SVT 算法的加速版本。

3.4 不动点延续算法

最优化问题(8)的另一种正则化版本为

$$\min_{\boldsymbol{X}}\mu\|\boldsymbol{X}\|_*+\frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{X})-\mathcal{P}_\Omega(\boldsymbol{M})\|_F^2 \tag{18}$$

其中，$\mu>0$。上述优化问题的目标函数关于 $\boldsymbol{X}$ 的次梯度为 $\mu\partial\|\boldsymbol{X}\|_*+\mathcal{P}_\Omega(\boldsymbol{X})-\mathcal{P}_\Omega(\boldsymbol{M})$。若 $\boldsymbol{X}^*$ 为最优化问题(18)的最优解，则它满足

$$\begin{aligned}0&\in\mu\tau\partial\|\boldsymbol{X}^*\|_*+\tau(\mathcal{P}_\Omega(\boldsymbol{X}^*)-\mathcal{P}_\Omega(\boldsymbol{M}))\\&=\mu\tau\partial\|\boldsymbol{X}^*\|_*+\boldsymbol{X}^*-(\boldsymbol{X}^*-\tau(\mathcal{P}_\Omega(\boldsymbol{X}^*)-\mathcal{P}_\Omega(\boldsymbol{M})))\end{aligned} \tag{19}$$

记 $\boldsymbol{Y}^*=\boldsymbol{X}^*-\tau(\mathcal{P}_\Omega(\boldsymbol{X}^*)-\mathcal{P}_\Omega(\boldsymbol{M}))$，则 $0\in\mu\tau\partial\|\boldsymbol{X}^*\|_*+\boldsymbol{X}^*-\boldsymbol{Y}^*$，即 $\boldsymbol{X}^*$ 是核范数最小化问题

$$\min_{\boldsymbol{X}}\mu\tau\|\boldsymbol{X}\|_*+\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{Y}^*\|_F^2 \tag{20}$$

的最优解。综上分析，不动点延续算法(Fixed Point Continuation, FPC)[21]的迭代过程如下

$$\begin{cases}\boldsymbol{Y}^k=\boldsymbol{X}^{k-1}-\tau(\mathcal{P}_\Omega(\boldsymbol{X}^{k-1})-\mathcal{P}_\Omega(\boldsymbol{M}))\\\boldsymbol{X}^k=\mathcal{D}_{\mu_k}(\boldsymbol{Y}^k)\\\mu_{k+1}=\max(\mu_k\eta_\mu,\bar{\mu})\end{cases} \tag{21}$$

其中，$\bar{\mu}>0$ 为参数 μ 取值的下限，$0<\eta_\mu<1$ 为常数。文献[22]还建议将 FPC 和 Bregman 迭代技术结合在一起。

3.5 加速近端梯度算法

记 $f(\boldsymbol{X})=\frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{X})-\mathcal{P}_\Omega(\boldsymbol{M})\|_F^2$，$g(\boldsymbol{X})=\mu\|\boldsymbol{X}\|_*$，则最小化问题(18)的目标函数为 $F(\boldsymbol{X})=f(\boldsymbol{X})+g(\boldsymbol{X})$。显然 $f(\boldsymbol{X})$ 为凸光滑函数，$g(\boldsymbol{X})$ 为连续、不光滑的凸函数。对于任意给定的 $\boldsymbol{Y}\in\mathbb{R}^{m\times n}$，考虑 $F(\boldsymbol{X})$ 在 $\boldsymbol{Y}$ 处的部分二次逼近

$$\begin{aligned}F(\boldsymbol{X})&\approx Q(\boldsymbol{X},\boldsymbol{Y})\\&=f(\boldsymbol{Y})+\langle\boldsymbol{X}-\boldsymbol{Y},\partial f(\boldsymbol{Y})\rangle+\frac{1}{2t}\|\boldsymbol{X}-\boldsymbol{Y}\|_F^2+g(\boldsymbol{X})\end{aligned} \tag{22}$$

根据上述逼近公式，Toh 等人提出了加速近端梯度算法(Accelerated Proximal Gradient, APG)[23]。APG 算法交替更新 $\boldsymbol{X}$、$\boldsymbol{Y}$ 和 t，其迭代过程如下

$$\begin{aligned}\boldsymbol{X}^k&=\arg\min_{\boldsymbol{X}} Q(\boldsymbol{X},\boldsymbol{Y}^{k-1})\\&=\arg\min_{\boldsymbol{X}}\frac{1}{2}\|\boldsymbol{X}-(\boldsymbol{Y}^{k-1}-t_k\partial f(\boldsymbol{Y}^{k-1}))\|_F^2+\mu t_k\|\boldsymbol{X}\|_*\\&=\mathcal{D}_{\mu t_k}(\boldsymbol{Y}^{k-1}-t_k(\mathcal{P}_\Omega(\boldsymbol{Y}^{k-1})-\mathcal{P}_\Omega(\boldsymbol{M})))\end{aligned} \tag{23}$$

$$\boldsymbol{Y}^k=\boldsymbol{X}^k+\frac{t_{k-1}-1}{t_k}(\boldsymbol{X}^k-\boldsymbol{X}^{k-1}) \tag{24}$$

$$t_{k+1}=\frac{1+\sqrt{1+4t_k^2}}{2} \tag{25}$$

在上述迭代过程中，文献[23]还提出并入类似线性搜索的加速算法。

3.6 增广拉格朗日乘子法

通过引入 $m\times n$ 维实矩阵变量 $\boldsymbol{E}$，最小化问题(8)可重新表示为

$$\min_{\boldsymbol{X},\boldsymbol{E}}\|\boldsymbol{X}\|_*,\ \text{s.t.}\ \boldsymbol{X}+\boldsymbol{E}=\boldsymbol{M},\mathcal{P}_\Omega(\boldsymbol{E})=0 \tag{26}$$

其中，$\boldsymbol{M}$ 的丢失元素设置为 0。上述最优化问题的部分增广拉格朗日函数为

$$L(\boldsymbol{X},\boldsymbol{E},\boldsymbol{Y},\mu)=\|\boldsymbol{X}\|_*+\langle\boldsymbol{Y},\boldsymbol{M}-\boldsymbol{X}-\boldsymbol{E}\rangle+\frac{\mu}{2}\|\boldsymbol{M}-\boldsymbol{X}-\boldsymbol{E}\|_F^2 \tag{27}$$

其中，正则化参数 $\mu>0$。初始化 $\boldsymbol{E}^0=\boldsymbol{Y}^0=0$，$\mu_0>0$，$\rho>1$，$\varepsilon>0$，则不精确增广拉格朗日乘子法(Inexact Augmented Lagrange Multipliers, IALM)[24]交替更新 $\boldsymbol{X},\boldsymbol{E},\boldsymbol{Y},\mu$，其迭代公式为

$$\begin{aligned}\boldsymbol{X}^k &= \arg\min_{\boldsymbol{X}} L(\boldsymbol{X},\boldsymbol{E}^{k-1},\boldsymbol{Y}^{k-1},\mu_{k-1})\\ &= \arg\min_{\boldsymbol{X}} \frac{1}{\mu_{k-1}}\|\boldsymbol{X}\|_* + \frac{1}{2}\|\boldsymbol{X}-(\boldsymbol{M}-\boldsymbol{E}^{k-1}+\boldsymbol{Y}^{k-1}/\mu_{k-1})\|_F^2\\ &= \mathcal{D}_{1/\mu_{k-1}}(\boldsymbol{M}-\boldsymbol{E}^{k-1}+\boldsymbol{Y}^{k-1}/\mu_{k-1}) \qquad (28)\\ \boldsymbol{E}^k &= \arg\min_{\mathcal{P}_\Omega(\boldsymbol{E})=0} L(\boldsymbol{X}^k,\boldsymbol{E},\boldsymbol{Y}^{k-1},\mu_{k-1})\\ &= \arg\min_{\mathcal{P}_\Omega(\boldsymbol{E})=0}\|\boldsymbol{E}-(\boldsymbol{M}-\boldsymbol{X}^k+\boldsymbol{Y}^{k-1}/\mu_{k-1})\|_F^2\\ &= \mathcal{P}_{\bar{\Omega}}(\boldsymbol{M}-\boldsymbol{X}^k+\boldsymbol{Y}^{k-1}/\mu_{k-1}) \qquad (29)\end{aligned}$$

$$\boldsymbol{Y}^k = \boldsymbol{Y}^{k-1}+\mu_{k-1}(\boldsymbol{M}-\boldsymbol{X}^k-\boldsymbol{E}^k) \qquad (30)$$

$$\mu_k = \begin{cases}\rho\mu_{k-1}, & \text{若 } \mu_{k-1}\|\boldsymbol{E}^k-\boldsymbol{E}^{k-1}\|_F/\|\boldsymbol{M}\|_F<\varepsilon\\ \mu_{k-1}, & \text{否则}\end{cases} \qquad (31)$$

其中，$\bar{\Omega}$表示 Ω 的补集。在更新 $\boldsymbol{X}$ 和 $\boldsymbol{E}$ 的过程中，若将式(28)和式(29)循环迭代，直到满足终止条件再执行式(30)和式(31)，称这种方法为精确增广拉格朗日乘子法(Exact Augmented Lagrange Multipliers，EALM)。EALM 方法较好地求解了$\min_{\boldsymbol{X},\boldsymbol{E}} L(\boldsymbol{X},\boldsymbol{E},\boldsymbol{Y}^{k-1},\mu_{k-1})$的近似解，但对于大的 μ_{k-1}，其收敛速度往往非常慢。实验结果证实 IALM 比 EALM 运行速度快，也称 IALM 为交替方向方法(Alternating Direction Method，ADM)[25,26]。

4 低秩矩阵分解模型的矩阵补全算法

4.1 问题描述

对于核范数最小化模型，用矩阵核范数来度量秩的大小。在实际问题中，若事先知道秩的取值，则可以用低秩矩阵分解模型来恢复丢失元素。这种方法还可以避免矩阵的奇异值分解。对于给定的含丢失元素的低秩或近似低秩矩阵 $\boldsymbol{M}$，对它作如下分解：$\boldsymbol{M}\approx\boldsymbol{XY}$，其中 $\boldsymbol{X}\in\mathbb{R}^{m\times r}$，$\boldsymbol{Y}\in\mathbb{R}^{r\times n}$，$r$ 为 $\boldsymbol{M}$ 的事先规定或根据某种方法确定的秩。

为了恢复丢失元素，建立下列最小化模型

$$\min_{\boldsymbol{X},\boldsymbol{Y}}\|\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{XY})\|_F^2,\ \text{s.t. } \mathcal{P}_\Omega(\boldsymbol{M})=\mathcal{P}_\Omega(\boldsymbol{XY}) \qquad (32)$$

下面给出近年来求解上述低秩矩阵分解模型的几种最新的方法。

4.2 最优空间算法

维数为 r 的$\mathbb{R}^n$的所有子空间的集合为 Grassmann 流形。Grassmann 流形也是紧的 Riemannn 流形，其测地线可以计算。Grassmann 矩阵流形中的一个点为 $m\times r$ 维正交矩阵 $\boldsymbol{A}$ 的等价类：$[\boldsymbol{A}]=\{\boldsymbol{AQ}:\boldsymbol{Q}$ 为任意的 r 阶正交方阵$\}$。基于 Grassmann 流形，文献[27]提出了求解优化问题(32)的最优空间算法(OptSpace)，此算法由以下 4 步组成。

第 1 步　剪切。记 $\boldsymbol{M}^\Omega=\mathcal{P}_\Omega(\boldsymbol{M})$，若 $\boldsymbol{M}^\Omega$ 的某个列向量的非零元素的个数大于 $2p/n$，则在此列向量中随机保留 $2p/n$ 个非零元素，其余非零元素值变为 0；若某行向量的非零元素的个数大于 $2p/m$，则在此列向量中随机保留 $2p/m$ 个非零元素，其余非零元素值改为 0。剪切后的矩阵记为 $\widetilde{\boldsymbol{M}}^\Omega$。

第 2 步　估计 $\boldsymbol{M}$ 的秩。对 $\widetilde{\boldsymbol{M}}^\Omega$ 进行奇异值分解：$\widetilde{\boldsymbol{M}}^\Omega=\sum_{i=1}^{k}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}$，其中 k 为 $\widetilde{\boldsymbol{M}}^\Omega$ 的秩，奇异值满足 $\sigma_1\geqslant\cdots\geqslant\sigma_k>0$。根据 $\widetilde{\boldsymbol{M}}^\Omega$ 的奇异值估计 $\boldsymbol{M}$ 的秩：

$$r=\arg\min_i\{(\sigma_{i+1}+\sigma_1\sqrt[4]{mn}\sqrt{i/p})/\sigma_i;i=1,2,\cdots,k-1\}$$

第 3 步　秩 r 投影。$\widetilde{\boldsymbol{M}}^\Omega$ 的秩 r 投影为 $\mathcal{P}_r(\widetilde{\boldsymbol{M}}^\Omega)=\boldsymbol{X}^0\boldsymbol{S}^0(\boldsymbol{Y}^0)^{\mathrm{T}}$，其中 $\boldsymbol{X}^0=\sqrt{m}(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r)$，$\boldsymbol{Y}^0=\sqrt{n}(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r)$，$\boldsymbol{S}^0=(\sqrt{mn}/p)\mathrm{diag}(\sigma_1,\cdots,\sigma_r)$。

第 4 步　Grassman 流形上的梯度下降法。定义差异性函数：$F(\boldsymbol{X},\boldsymbol{Y})=\min_{\boldsymbol{S}\in\mathbb{R}^{r\times r}}\mathcal{F}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S})$，其中 $\mathcal{F}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S})=\frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{M}^\Omega-\boldsymbol{XSY}^{\mathrm{T}})\|_F^2+\frac{\lambda}{2}\|\mathcal{P}_\Omega(\boldsymbol{XSY}^{\mathrm{T}})\|_F^2$，$\boldsymbol{X},\boldsymbol{Y}$ 满足 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=m\boldsymbol{I}_r$，$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}=n\boldsymbol{I}_r$。为简便起见，考虑 $\lambda=0$。使用交替和梯度下降法求解优化问题$\min_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S}}\mathcal{F}(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{S})$，迭代过程如下

$$\begin{cases}\boldsymbol{S}^k=\arg\min_{\boldsymbol{S}}\mathcal{F}(\boldsymbol{X}^{k-1},\boldsymbol{Y}^{k-1},\boldsymbol{S})\\ t_k=\arg\min_{t\geqslant 0}F(\boldsymbol{X}^{k-1}-t\partial_{\boldsymbol{X}}F(\boldsymbol{X}^{k-1},\boldsymbol{Y}^{k-1}),\boldsymbol{Y}^{k-1}-\\ \qquad t\partial_{\boldsymbol{Y}}F(\boldsymbol{X}^{k-1},\boldsymbol{Y}^{k-1}))\\ \boldsymbol{X}^k=\sqrt{m}\ \mathrm{orth}(\boldsymbol{X}^{k-1}-t_k\partial_{\boldsymbol{X}}F(\boldsymbol{X}^{k-1},\boldsymbol{Y}^{k-1}))\\ \boldsymbol{Y}^k=\sqrt{n}\ \mathrm{orth}(\boldsymbol{Y}^{k-1}-t_k\partial_{\boldsymbol{Y}}F(\boldsymbol{X}^{k-1},\boldsymbol{Y}^{k-1}))\end{cases} \qquad (33)$$

其中，$\partial_{\boldsymbol{X}}F(\boldsymbol{X},\boldsymbol{Y})=\mathcal{P}_\Omega(\boldsymbol{XSY}^{\mathrm{T}}-\boldsymbol{M}^\Omega)\boldsymbol{YS}^{\mathrm{T}}$，$\partial_{\boldsymbol{Y}}F(\boldsymbol{X},\boldsymbol{Y})=\mathcal{P}_\Omega(\boldsymbol{XSY}^{\mathrm{T}}-\boldsymbol{M}^\Omega)\boldsymbol{XS}$，$\mathrm{orth}(\cdot)$表示矩阵列正交化算子。当迭代公式(33)达到终止条件时，$\boldsymbol{M}$ 补全后的结果为 $\boldsymbol{X}^k\boldsymbol{S}^k(\boldsymbol{Y}^k)^{\mathrm{T}}$。

4.3 Grassmann 秩 1 更新子空间估计算法

由于噪声的存在，最优化模型(32)中的约束 $\mathcal{P}_\Omega(\boldsymbol{M})=\mathcal{P}_\Omega(\boldsymbol{XY})$不一定成立。为此，考虑下列优化模型

$$\min_{\boldsymbol{X},\boldsymbol{Y}}\|\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{XY})\|_F^2 \qquad (34)$$

将矩阵 $\boldsymbol{M}$ 和 $\boldsymbol{Y}$ 按列向量表示：$\boldsymbol{M}=(\boldsymbol{M}_{(1)},\boldsymbol{M}_{(2)},\cdots,\boldsymbol{M}_{(n)})$，$\boldsymbol{Y}=(\boldsymbol{Y}_{(1)},\boldsymbol{Y}_{(2)},\cdots,\boldsymbol{Y}_{(n)})$。记 $\mathcal{H}_{\Omega i}$ 为向量、矩阵选择算子，即 $\mathcal{H}_{\Omega i}(\boldsymbol{M}_{(i)})$按照 Ω 来选取 $\boldsymbol{M}_{(i)}$ 的非丢失元素，$\mathcal{H}_{\Omega i}(\boldsymbol{X})$按照 $\boldsymbol{M}_{(i)}$ 的非丢失元素来选取 $\boldsymbol{X}$ 的相应的行向量，$i=1,2,\cdots,n$。于是最优化问题(34)等价于

$$\min_{\boldsymbol{X},\boldsymbol{Y}}\sum_{i=1}^{n}\|\mathcal{H}_{\Omega i}(\boldsymbol{M}_{(i)})-\mathcal{H}_{\Omega i}(\boldsymbol{X})\boldsymbol{Y}_{(i)}\|^2 \qquad (35)$$

其中，$\|\cdot\|$表示向量的 l_2 范数。文献[13]提出了求解上述最优化问题的 Grassmann 秩 1 更新子空间估计算法(Grassmannian Rank-One Update Subspace Estimation，GROUSE)。在 GROUSE 中，更新矩阵 $\boldsymbol{X}$ 的算法如下。

For i=1:n

$\mathbf{Y}_{(i)}^*=\arg\min_{\mathbf{Y}_{(i)}}\|\mathcal{H}_{\Omega i}(\mathbf{M}_{(i)})-\mathcal{H}_{\Omega i}(\mathbf{X})\mathbf{Y}_{(i)}\|^2=(\mathcal{H}_{\Omega i}(\mathbf{X}))^\dagger\mathcal{H}_{\Omega i}(\mathbf{M}_{(i)})$；

预测 $\mathbf{M}_{(i)}$ 的补全向量 $\mathbf{p}_i=\mathbf{X}\mathbf{Y}_{(i)}^*$；

计算残差向量 $\mathbf{r}_{\Omega i}=\mathcal{H}_{\Omega i}(\mathbf{M}_{(i)})-\mathcal{H}_{\Omega i}(\mathbf{X})\mathbf{Y}_{(i)}^*$；

根据 $\mathbf{r}_{\Omega i}$ 和 $\mathcal{H}_{\Omega i}(\cdot)$得到 m 维向量 $\mathbf{r}_i$，即 $\mathbf{r}_i$ 对应的丢失元素取值为 0；

更新矩阵 $\mathbf{X}$：$\mathbf{X}\leftarrow\mathbf{X}+[(\cos(\|\mathbf{r}_i\|\|\mathbf{p}_i\|\eta_i)-1)\mathbf{p}_i/\|\mathbf{p}_i\|+\sin(\|\mathbf{r}_i\|\|\mathbf{p}_i\|\eta_i)\mathbf{r}_i/\|\mathbf{r}_i\|](\mathbf{Y}_{(i)}^*)^{\mathrm{T}}/\|\mathbf{Y}_{(i)}^*\|$。

End

在上述算法中，“†”表示矩阵的伪逆，η_i 为步长大小。为了得到更优的解，可将上述算法重复多遍。最终矩阵 Y 的求解公式为

$$\boldsymbol{Y}^*=((\mathcal{H}_{\Omega 1}(\boldsymbol{X}))^\dagger\mathcal{H}_{\Omega 1}(\boldsymbol{M}_{(1)}),\cdots,(\mathcal{H}_{\Omega n}(\boldsymbol{X}))^\dagger\mathcal{H}_{\Omega n}(\boldsymbol{M}_{(n)})) \qquad (36)$$

矩阵补全的其它的 Riemann 或 Grassmann 流形子空间方法可参见文献[28,29]。

4.4 奇异值投影算法

记 $\boldsymbol{Z}=\boldsymbol{XY}$，则最小化问题(32)的鲁棒形式为

$$\begin{aligned}&\min_{\boldsymbol{Z}}\psi(\boldsymbol{Z})=\frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{Z})-\mathcal{P}_\Omega(\boldsymbol{M})\|_F^2\\ &\text{s.t. } \boldsymbol{Z}\in\mathscr{C}(r)=\{\boldsymbol{U}\in\mathbb{R}^{m\times n}:\mathrm{rank}(\boldsymbol{U})\leqslant r\}\end{aligned} \qquad (37)$$

由于集合 $\mathscr{C}(r)$是非凸、非连续的，因此直接求解上述最优化问题是非常困难的。

利用启发式方法，文献[30]提出了求解最小化问题(37)的奇异值投影算法(Singular Value Projection，SVP)。在SVP中，先不考虑低秩约束，直接使用梯度下降法更新矩阵 $\boldsymbol{Z}$。$\boldsymbol{Z}$ 的更新公式为

$$\widehat{\boldsymbol{Z}}^k=\boldsymbol{Z}^{k-1}-\eta_k\partial\psi(\boldsymbol{Z}^{k-1})=\boldsymbol{Z}^{k-1}-\eta_k(\mathcal{P}_\Omega(\boldsymbol{Z}^{k-1})-\mathcal{P}_\Omega(\boldsymbol{M})) \tag{38}$$

步长大小 η_k 可按如下方式选取：$\eta_k=mn/(p+\delta p)$，其中 $0<\delta<1/3$ 且取值依赖于 p。再将 $\widehat{\boldsymbol{Z}}^k$ 投影到 $\mathscr{C}(r)$ 上，即

$$\boldsymbol{Z}^k=\arg\min_{\boldsymbol{Z}}\{\|\boldsymbol{Z}-\widehat{\boldsymbol{Z}}^k\|_F:\boldsymbol{Z}\in\mathscr{C}(r)\} \tag{39}$$

对 $\widehat{\boldsymbol{Z}}^k$ 进行奇异值分解，由其前 r 个最大的奇异值及对应的奇异向量可构成矩阵 $\boldsymbol{Z}^k$。

4.5 低秩矩阵拟合算法

通过引入辅助变量矩阵 $\boldsymbol{Z}$，最小化问题(32)等价于

$$\min_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}}\frac{1}{2}\|\boldsymbol{XY}-\boldsymbol{Z}\|_F^2,\ \text{s.t. }\mathcal{P}_\Omega(\boldsymbol{Z})=\mathcal{P}_\Omega(\boldsymbol{M}) \tag{40}$$

此优化问题的拉格朗日函数为

$$L(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\Lambda)=\frac{1}{2}\|\boldsymbol{XY}-\boldsymbol{Z}\|_F^2-\langle\Lambda,\mathcal{P}_\Omega(\boldsymbol{Z}-\boldsymbol{M})\rangle \tag{41}$$

对 $L(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\Lambda)$ 关于各个块变量求梯度，得到一阶最优性条件

$$\begin{cases}(\boldsymbol{XY}-\boldsymbol{Z})\boldsymbol{Y}^{\mathrm{T}}=0\\ \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{XY}-\boldsymbol{Z})=0\\ \mathcal{P}_\Omega(\boldsymbol{Z}-\boldsymbol{XY})=0\\ \mathcal{P}_\Omega(\boldsymbol{Z}-\boldsymbol{M})=0\\ \mathcal{P}_\Omega(\boldsymbol{Z}-\boldsymbol{XY})=\Lambda\end{cases} \tag{42}$$

使用非线性逐次超松弛迭代法求解上述方程系统，交替迭代公式如下

$$\begin{cases}\boldsymbol{X}_+\leftarrow\boldsymbol{Z}\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}})^{-1}\\ \boldsymbol{X}_+(\omega)\leftarrow\omega\boldsymbol{X}_++(1-\omega)\boldsymbol{X}\\ \boldsymbol{Y}_+\leftarrow(\boldsymbol{X}_+(\omega)^{\mathrm{T}}\boldsymbol{X}_+(\omega))^{-1}(\boldsymbol{X}_+(\omega)^{\mathrm{T}}\boldsymbol{Z})\\ \boldsymbol{Y}_+(\omega)\leftarrow\omega\boldsymbol{Y}_++(1-\omega)\boldsymbol{Y}\\ \boldsymbol{Z}_+(\omega)\leftarrow\boldsymbol{X}_+(\omega)\boldsymbol{Y}_+(\omega)+\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{X}_+(\omega)\boldsymbol{Y}_+(\omega))\end{cases} \tag{43}$$

其中，参数 $\omega\geqslant1$。当 $\omega=1$ 时，上述迭代方法即为 Gauss-Seidel 方法。迭代公式(43)称为求解矩阵补全的低秩矩阵拟合算法(Low-rank matrix Fitting，LmaFit)[31]。

5 其它优化模型

由核范数最小化模型和低秩矩阵分解模型衍生出许多求解矩阵补全的其它模型，本节对部分模型及求解方法作简要介绍。

5.1 剪切核范数正则化模型

在核范数最小化模型中，同时最小化所有奇异值，这在实际问题中可能不能更好地逼近矩阵的秩。文献[32]提出了剪切核范数正则化模型(Truncated Nuclear Norm Regularization，TNNR)，它在一定程度上避免了核范数优化模型的上述缺陷。

定义 5[32] 矩阵 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$ 的奇异值按大小排列为 $\sigma_1,\sigma_2,\cdots,\sigma_{\min(m,n)}$，则它的 r-剪切核范数为

$$\|\boldsymbol{X}\|_{-r}=\sum_{i=r+1}^{\min(m,n)}\sigma_i \tag{44}$$

在核范数最小化模型(8)中，将矩阵的核范数替换成矩阵的剪切核范数，便得如下的 TNNR 模型

$$\min_{\boldsymbol{X}}\|\boldsymbol{X}\|_{-r},\ \text{s.t. }\mathcal{P}_\Omega(\boldsymbol{X})=\mathcal{P}_\Omega(\boldsymbol{M}) \tag{45}$$

上述最优化问题也等价于

$$\min_{\boldsymbol{X}}(\|\boldsymbol{X}\|_*-\max_{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{I}_r}\mathrm{trace}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}^{\mathrm{T}})),\ \text{s.t. }\mathcal{P}_\Omega(\boldsymbol{X})=\mathcal{P}_\Omega(\boldsymbol{M}) \tag{46}$$

其中，$\boldsymbol{A}\in\mathbb{R}^{r\times m},\boldsymbol{B}\in\mathbb{R}^{r\times n}$。初始化 $\boldsymbol{X}^0=\mathcal{P}_\Omega(\boldsymbol{M})$，上述最优化问题可转化为交替求解两个子优化问题

$$\max_{\boldsymbol{A},\boldsymbol{B}}\mathrm{trace}(\boldsymbol{A}\boldsymbol{X}^{k-1}\boldsymbol{B}^{\mathrm{T}}),\ \text{s.t. }\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{I}_r,\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}=\boldsymbol{I}_r \tag{47}$$

$$\min_{\boldsymbol{X}}\|\boldsymbol{X}\|_*-\langle\boldsymbol{X},\boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{A}_k\rangle,\ \text{s.t. }\mathcal{P}_\Omega(\boldsymbol{X})=\mathcal{P}_\Omega(\boldsymbol{M}) \tag{48}$$

最优化问题(47)的最优解记为 $\boldsymbol{A}^k=(\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r)^{\mathrm{T}}$，$\boldsymbol{B}^k=(\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_r)^{\mathrm{T}}$，其中单位向量 $\boldsymbol{u}_i$、$\boldsymbol{v}_i$ 分别为 $\boldsymbol{X}^{k-1}$ 的第 i 大奇异值对应的左、右奇异向量。对于最小化问题(48)，文献[32]提出了两种求解方法，即 ADM 和基于线性搜索的 APG。

5.2 Schatten p-范数优化模型

在核范数优化模型(8)中，将矩阵的 Schatten 1-范数推广到 Schatten p-范数($0<p\leqslant2$)，则有最小化模型

$$\min_{\boldsymbol{X}}\|\boldsymbol{X}\|_{Sp},\ \text{s.t. }\mathcal{P}_\Omega(\boldsymbol{X})=\mathcal{P}_\Omega(\boldsymbol{M}) \tag{49}$$

上述最优化问题的目标函数也可用迹函数来表示，即 $\|\boldsymbol{X}\|_{Sp}=(\mathrm{trace}((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{p/2}))^{1/p}$。文献[33]提出了求解上述最小化模型的拉格朗日乘子法，并证明了所提算法的收敛性。

5.3 最小秩逼近原子分解模型

令 Γ 为 $\mathbb{C}^{m\times n}$ 中所有秩为 1、Frobenius 范数也为 1 的矩阵构成的集合。对 Γ 的元素进行精炼，即任意两个相异的矩阵不共线，精炼后的集合称为原子集，用 O 表示。$\forall\boldsymbol{Z}\in\mathbb{C}^{m\times n}$，它的原子分解形式为 $\boldsymbol{Z}=\sum_i\alpha_i\phi_i,\phi_i\in O$。易知 $rank(\boldsymbol{Z})=\min_\Phi\{|\Phi|:\Phi\subset O,\boldsymbol{Z}\in span(\Phi)\}$。

秩最小化问题(6)对应的鲁棒优化问题为

$$\min_{\boldsymbol{X}}\|\mathscr{A}(\boldsymbol{X})-\boldsymbol{b}\|,\ \text{s.t. }\mathrm{rank}(\boldsymbol{X})\leqslant r \tag{50}$$

文献[34]提出了使用最小秩逼近原子分解模型(Atomic Decomposition for Minimum Rank Approximation，ADMiRA)求解上述最优化问题。对于初始化的 $\boldsymbol{X}^0$ 和 Φ^0，ADMiRA 方法交替更新 Φ 和 $\boldsymbol{X}$：

$$\begin{cases}\widetilde{\Phi}^k=\arg\max_{\Phi\subset O}\{\|\mathcal{P}_\Phi\mathscr{A}^*(\mathscr{A}(\boldsymbol{X}^{k-1})-\boldsymbol{b})\|_F:|\Phi|\leqslant2r\}\cup\Phi^{k-1}\\ \widetilde{\boldsymbol{X}}^k=\arg\min_{\boldsymbol{X}}\{\|\mathscr{A}(\boldsymbol{X})-\boldsymbol{b}\|:\boldsymbol{X}\in\mathrm{span}(\widetilde{\Phi}^k)\}\\ \Phi^k=\arg\min_{\Phi\subset O}\{\|\mathcal{P}_\Phi\widetilde{\boldsymbol{X}}^k\|_F:|\Phi|\leqslant r\}\\ \boldsymbol{X}^k=\mathcal{P}_{\Phi^k}\widetilde{\boldsymbol{X}}^k\end{cases} \tag{51}$$

其中，$\mathscr{A}^*$ 为 $\mathscr{A}$ 的伴随算子，$\mathcal{P}_\Phi$ 表示 Φ 张成子空间上的正交投影。

5.4 最大范数约束的优化模型

对于秩为 r 的矩阵 $\boldsymbol{Z}\in\mathbb{R}^{m\times n}$，补充它的几种范数：$\|\boldsymbol{Z}\|_\infty=\max_{i,j}|Z_{ij}|$，$\|\boldsymbol{Z}\|_{2,\infty}=\max_i\|\boldsymbol{Z}_{(i)}\|$，$\|\boldsymbol{Z}\|_{\max}=\min\{\|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty}:\boldsymbol{Z}=\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}},\boldsymbol{U}\in\mathbb{R}^{m\times r},\boldsymbol{V}\in\mathbb{R}^{n\times r}\}$，其中 $\boldsymbol{Z}_{(i)}$ 为 $\boldsymbol{Z}$ 的第 i 列构成的向量。对于给定的 $R\geqslant\alpha>0$，文献[35]提出的最大范数约束优化模型为

$$\min_{\boldsymbol{Z}}\|\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{Z})\|_F^2,\ \text{s.t. }\|\boldsymbol{Z}\|_\infty\leqslant\alpha,\|\boldsymbol{Z}\|_{\max}\leqslant R \tag{52}$$

上述优化问题可以转化为半定规划模型

$$\min_{\boldsymbol{X},\boldsymbol{W}_1,\boldsymbol{W}}\|\mathcal{P}_\Omega(\boldsymbol{M}-\boldsymbol{Z})\|_F^2$$
$$\text{s.t. }\begin{pmatrix}\boldsymbol{W}_1&\boldsymbol{Z}\\ \boldsymbol{Z}^{\mathrm{T}}&\boldsymbol{W}_2\end{pmatrix}\geqslant0 \tag{53}$$
$$\mathrm{diag}(\boldsymbol{W}_1)\leqslant R,\mathrm{diag}(\boldsymbol{W}_2)\leqslant R,\|\boldsymbol{Z}\|_\infty\leqslant\alpha$$

或分解形式模型

$$\min_{\boldsymbol{X}} \| \mathcal{P}_{\Omega}(\boldsymbol{M}-\boldsymbol{XY}) \|_F^2$$
$$\text{s. t. } \max\{ \| \boldsymbol{X} \|_{2,\infty}^2, \| \boldsymbol{Y} \|_{2,\infty}^2 \} \leqslant R \tag{54}$$
$$\max_{i,j} |(\boldsymbol{XY})_{ij}| \leqslant \alpha$$

可使用一阶梯度下降法求解最小化问题(54)。

5.5 低秩逼近模型

若矩阵 $\boldsymbol{X}$ 的秩为 r，则它可以分解为：$\boldsymbol{X}=\boldsymbol{LR}^{\mathrm{T}}$，其中 $\boldsymbol{L}\in\mathbb{R}^{m\times r}$，$\boldsymbol{R}\in\mathbb{R}^{n\times r}$。此时，$\boldsymbol{X}$ 的核范数满足下列性质[10]

$$\| \boldsymbol{X} \|_* = \min_{\boldsymbol{X}=\boldsymbol{LR}^{\mathrm{T}}} \| \boldsymbol{L} \|_F \| \boldsymbol{R} \|_F = \frac{1}{2} \min_{\boldsymbol{X}=\boldsymbol{LR}^{\mathrm{T}}} (\| \boldsymbol{L} \|_F^2 + \| \boldsymbol{R} \|_F^2) \tag{55}$$

最优化问题(8)对应的低秩逼近模型为

$$\min_{\boldsymbol{L},\boldsymbol{R}} \| \boldsymbol{L} \|_F^2 + \| \boldsymbol{R} \|_F^2, \text{ s. t. } \mathcal{A}(\boldsymbol{LR}^{\mathrm{T}}) = \boldsymbol{b} \tag{56}$$

如果 r 比式(8)的最优解矩阵的秩充分大，则最优化问题(8)与式(56)等价。文献[15]提出了求解式(56)的增广拉格朗日乘子法，文献[36]对式(56)进行正则化，并使用梯度下降法求解。

5.6 非负矩阵分解模型

非负矩阵分解是处理非负低秩矩阵的一种流行的方法，它将非负数据矩阵 $\boldsymbol{M}$ 分解为非负低秩矩阵 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之积。基于非负矩阵分解的矩阵补全描述如下

$$\min_{\boldsymbol{X},\boldsymbol{Y}} \| \mathcal{P}_{\Omega}(\boldsymbol{M}-\boldsymbol{XY}) \|_F^2$$
$$\text{s. t. } X_{ik} \geqslant 0, Y_{kj} \geqslant 0, i=1,\cdots,m, \tag{57}$$
$$k=1,\cdots,r, j=1,\cdots,n$$

其中，$\boldsymbol{X}\in\mathbb{R}^{m\times r}$，$\boldsymbol{Y}\in\mathbb{R}^{r\times n}$，$r\ll\min(m,n)$。对于上述最优化问题，文献[37]先将它转化为交替求解两个非负最小二乘问题，再对每个非负最小二乘问题使用内点梯度法求解；文献[38]提出使用 ADM 方法求解。

此外，矩阵补全的基于其它模型的算法还有：鲁棒低秩矩阵分解框架的交替方向法[39]、正则化算法[40]、迭代重新加权算法[41]和原-对偶算法[42]等。

6 实验分析

先在人工数据集上比较矩阵补全的主要算法的性能，再将部分算法应用到图像恢复中。

6.1 人工数据集

随机生成秩为 r 的 $m\times n$ 维矩阵 $\boldsymbol{M}$，生成方式如下：$\boldsymbol{M}=\boldsymbol{M}_L\boldsymbol{M}_R$，其中 $\boldsymbol{M}_L\in\mathbb{R}^{m\times r}$，$\boldsymbol{M}_R\in\mathbb{R}^{r\times n}$，$r<\min(m,n)$。矩阵 $\boldsymbol{M}_L$、$\boldsymbol{M}_R$ 的各元素之间相互独立且服从均值为 0、标准差为 1 的正态分布。按照均匀分布对 $\boldsymbol{M}$ 进行随机采样，采样指标集合为 Ω，采样数目记为 p。根据核范数优化模型和低秩矩阵分解模型来补全 $\boldsymbol{M}$ 的未采样元素，补全后的矩阵记为 $\widetilde{\boldsymbol{M}}$。

对于各种矩阵补全算法(GROUSE 除外)，设终止条件为 $\| \mathcal{P}_{\Omega}(\boldsymbol{M}^k)-\mathcal{P}_{\Omega}(\boldsymbol{M}) \|_F / \| \mathcal{P}_{\Omega}(\boldsymbol{M}) \|_F \leqslant \varepsilon$，其中 $\boldsymbol{M}^k$ 为 $\boldsymbol{M}$ 的第 k 次迭代时的补全结果，ε 为容许误差。使用相对误差 $\| \widetilde{\boldsymbol{M}}-\boldsymbol{M} \|_F / \| \boldsymbol{M} \|_F$ 来度量 $\widetilde{\boldsymbol{M}}$ 的补全性能。在每种算法中，对于给定的秩 r 和采样数目 p，将矩阵补全算法重复 10 次，最后报告运行时间、迭代次数和相对误差的平均值。在实验中，取 $m=n=1000$，$\varepsilon=10^{-4}$，则 $\boldsymbol{M}$ 的自由度为 $dr=r(2000-r)$。考虑 $r=10$、50、100 这 3 种情形，实验结果如表 1 所列。

表 1 各种矩阵补全算法的性能比较

算法	秩	p/dr	p /n²	运行时间(s)	迭代次数	相对误差
SVT	10	6	0.12	85.0	117	1.66e−4
	50	4	0.39	559.3	113.1	1.62e−4
	100	3	0.57	1359.1	128.7	1.70e−4
FPC	10	6	0.12	18.9	91.6	4.69e−4
	50	4	0.39	83.7	70.1	3.05e−5
	100	3	0.57	132.3	79	2.25e−5
APG	10	6	0.12	10.2	37.7	2.71e−4
	50	4	0.39	28.2	51	3.77e−2
	100	3	0.57	360.7	49	9.60e−3
IALM	10	6	0.12	8.1	45	1.40e−4
	50	4	0.39	30.6	25.6	1.44e−4
	100	3	0.57	89.9	28	1.53e−4
OptSpace	10	6	0.12	66.2	24.6	4.94e−5
	50	4	0.39	6804.5	34	2.00e−5
GROUSE	10	6	0.12	4.2	—	2.05e−6
	50	4	0.39	39.7	—	1.57e−8
	100	3	0.57	137.1	—	4.56e−5
SVP	10	6	0.12	6.4	26.1	1.56e−4
	50	4	0.39	28.4	29	7.07e−5
	100	3	0.57	75.8	33	6.25e−5
LmaFit	10	6	0.12	1.8	28	1.54e−4
	50	4	0.39	4.6	21	1.43e−4
	100	3	0.57	8.6	21	1.58e−4

对于低秩矩阵分解模型的 OptSpace 方法，当 $r=50$，$p/dr=4$ 时，由于运行时间非常长，因此只报告了一次实验的结果；当 $r=100$，$p/dr=3$ 时，运行时间更长，没有报告实验结果。从表 1 中可以看出：对于核范数优化模型的 4 种算法，APG 的相对误差最大，其它 3 种方法都达到了 10^{-4} 或更低的数量级，而 IALM 运行时间最短，且迭代次数较少。对于低秩矩阵分解模型，4 种算法都达到了较好的相对误差，而 LmaFit 的运行时间最短，迭代次数也较低。IALM 与 LmaFit 相比，后者的运算速度更快，但需要事先确定矩阵的秩，而如何确定不完全矩阵的秩是一个公开问题。

6.2 图像恢复

通过 6.1 节的分析可知，IALM 和 LmaFit 分别代表两类模型的最有效的算法。本节将这两种算法应用到图像恢复中。“熊猫”灰度图像的分辨率为 194×259，采样数目记为 p，则采样概率为 $sp=p/(194\times259)$。考虑 4 种采样概率，即 $sp=0.1,0.3,0.5,0.7$，分别使用 IALM 和 LamFit 来补全丢失元素。用峰值信噪比(psnr)来度量算法的恢复性能，其中在 LamFit 算法中，考虑 psnr 最大时秩 r 的取值情形。实验结果如图 1 所示。

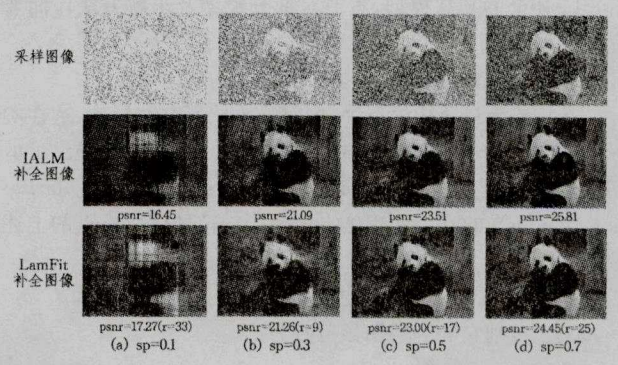

图 1 不同采样概率下 IALM 和 LamFit 的恢复性能比较

在图 1 的采样图像中，未采样元素或丢失元素用白像素值(即 255)来表示。从图 1 可以看出：当 $sp=0.1$ 和 0.3 时，LamFit 的恢复性能优于 IALM；当 $sp=0.5$ 和 0.7 时，IALM

的恢复性能优于 LamFit。对于 $sp=0.1$，两种方法均未能较好地恢复熊猫的轮廓，这可能是由于图像的秩取值较大且未充分采样所致。此外，在 LamFit 方法中需要事先给定矩阵的秩，而秩的大小直接影响着恢复性能。

结束语 本文报告了矩阵补全算法的最新研究进展。将求解矩阵补全的主要算法按模型归纳为核范数优化模型和低秩矩阵分解模型两类。对于这两类模型，总结了各种主流的算法，并通过数值实验比较了它们的性能。本文也补充了这两类模型的衍生版本及相应的求解方法。

对于大规模问题，设计可扩展的快速算法是当前矩阵补全算法的研究核心，这涉及到部分奇异值分解的快速算法和稀疏矩阵的存储与运算。此外，在今后的矩阵补全算法研究中，下面几个方向值得关注：(a)建立适合不同问题背景的矩阵补全优化模型；(b)将矩阵补全推广到高阶张量上，即研究张量补全和非负张量补全算法[43-45]；(c)Grassmann 流形子空间方法求解矩阵补全问题；(d)如何根据矩阵的部分元素来确定它的秩[46]。

参 考 文 献

[1] Donoho D. Compressed sensing [J]. IEEE Transactions on Information Theory, 2006, 52(4): 1289-1306

[2] Candès E J, Michael W. An introduction to compressive sampling [J]. IEEE Signal Processing Magazine, 2008, 25(2): 21-30

[3] Candès E J, Tao T. The power of convex relaxation: Near-optimal matrix completion [J]. IEEE Transactions on Information Theory, 2010, 56(5): 2053-2080

[4] Candès E J, Plan Y. Matrix completion with noise [J]. Proceedings of the IEEE, 2010, 98(6): 925-936

[5] 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

[6] Chen P. Optimization algorithms on subspaces: revisiting missing data problem in low-rank matrix [J]. International Journal of Computer Vision, 2008, 80(1): 125-142

[7] Vidal R, Tron R, Hartley R. Multiframe motion segmentation with missing data using PowerFactorization and GPCA [J]. International Journal of Computer Vision, 2008, 79(1): 85-105

[8] Ji Hui, Liu Chao-qiang, Shen Zuo-wei, et al. Robust video denoising using low rank matrix completion [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2010: 1791-1798

[9] Jing Guo-dong, Shi Yun-hui, Yin Bao-cai. Image super-resolution reconstruction based on sparse representation and low-rank matrix completion [J]. Journal of Information & Computational Science, 2012, 9(13): 3859-3866

[10] Rennie D M, Srebro N. Fast maximum margin matrix factorization for collaborative prediction [C]// Proceedings of International Conference on Machine Learning (ICML). 2005: 713-719

[11] Cabral R S, Torre F D, Costeira J P, et al. Matrix completion for multi-label image classification [C]// Proceedings of Neural Information Processing Systems (NIPS). 2011: 190-198

[12] Keshavan R H, Oh S. OptSpace: A gradient descent algorithm on the Grassman manifold for matrix completion [EB/OL]. http://arxiv.org/pdf/0910.5260v2.pdf, 2009

[13] Balzano L, Nowak R, Recht B. Online identification and tracking of subspaces from highly incomplete information [C]// Proceedings of Annual Allerton Conference on Communication, Control, and Computing. 2010: 704-711

[14] Liu Zhang, Hansson A, Vandenberghe L. Nuclear norm system identification with missing inputs and outputs [EB/OL]. http://www.ee.ucla.edu/~vandenbe/publications/subspace.pdf, 2012

[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] Fazel M. Matrix rank minimization with applications [D]. Stanford University, 2002

[17] Liu Zhang, Vandenberghe L. Interior-point method for nuclear norm approximation with application to system identification [J]. SIAM Journal on Matrix Analysis and Applications, 2009, 31(3): 1235-1256

[18] Cai Jian-feng, Candès E J, Shen Zuo-wei. A singular value thresholding algorithm for matrix completion [J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982

[19] Hu Yao, Zhang De-bing, Liu Jun, et al. Accelerated singular value thresholding for matrix completion [C]// Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2012: 298-306

[20] Cai Jian-feng, Osher S. Fast singular value thresholding without singular value decomposition [J]. Methods and Applications of Analysis, 2013, 20(4): 335-352

[21] Donald Goldfarb D, Ma Shi-qian. Convergence of fixed-point continuation algorithms for matrix rank minimization [J]. Foundations of Computational Mathematics, 2011, 11(2): 183-210

[22] Ma Shi-qian, Goldfarb D, Chen Li-feng. Fixed point and Bregman iterative methods for matrix rank minimization [J]. Mathematical Programming, 2011, 128(1/2): 321-353

[23] Toh K C, Yun S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems [J]. Pacific Journal of Optimization, 2010, 6: 615-640

[24] 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]. http:// arxiv.org/pdf/1009.5055, 2010

[25] Chen Cai-hua, He Bing-sheng, Yuan Xiao-ming. Matrix completion via an alternating direction method [J]. IMA Journal of Numerical Analysis, 2012, 32(1): 227-245

[26] Yang Jun-feng, Yuan Xiao-ming. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization [J]. Mathematics of Computation, 2013, 82(281): 301-329

[27] Keshavan R H, Montanari A, Sewoong Oh. Matrix completion from a few entries [J]. IEEE Transactions on Information Theory, 2010, 56(6): 2980-2998

[28] Dai Wei, Kerman E, Milenkovic O. A geometric approach to low-rank matrix completion [J]. IEEE Transactions on Information Theory, 2012, 58(1): 237-247

[29] Boumal N, Absil P A. RTRMC: A Riemannian trust-region method for low-rank matrix completion [C]// Proceedings of Neural Information Processing Systems (NIPS). 2011

[30] Jain P, Meka R, Dhillon I. Guaranteed rank minimization via singular value projection [EB/OL]. http://arxiv.org/abs/0909.5457, 2009

[31] Wen Zai-wen, Yin Wo-tao, Zhang Yin. Solving a low-rank factorization model for matrix completion by a nonlinear successive

over-relaxation algorithm[J]. Mathematical Programming Computation,2012,4(4):333-361

[32] Zhang De-bing, Hu Yao, Ye Jie-ping, et al. Matrix completion by truncated nuclear norm regularization[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2012:2192-2199

[33] Nie Fei-ping, Huang H, Ding C. Low-rank matrix recovery via efficient schatten p-norm minimization [C] // Proceedings of AAAI Conference on Artificial Intelligence. 2012:655-661

[34] Lee K, Bresler Y. ADMiRA: Atomic decomposition for minimum rank approximation[J]. IEEE Transactions on Information Theory,2010,56(9):4402-4416

[35] Cai T T, Zhou Wen-xin. Matrix completion via max-norm constrained optimization [EB/OL]. http:// arxiv. org/pdf/1303.0341v1. pdf,2013

[36] Recht B, Christopher Re C. Parallel stochastic gradient algorithms for large-scale matrix completion [EB/OL]. http://pages. cs. wisc. edu/~brecht/papers/11. Rec. Re. IPGM. pdf,2013

[37] 史加荣,焦李成,尚凡华. 不完全非负矩阵分解的加速算法[J]. 电子学报,2011,39(2):291-295

[38] Xu Yang-yang, Yin Wo-tao, Wen Zai-wen, et al. An alternating direction algorithm for matrix completion with nonnegative factors [J]. Frontiers of Mathematics in China,2012,7(2):365-384

[39] 尚凡华. 基于低秩结构学习数据表示[D]. 西安:西安电子科技大学,2012

[40] Mazumder R, Hastie T, Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices[J]. Journal of Machine Learning Research,2010,11:2287-2322

[41] Mohan K, Fazel M. Iterative reweighted algorithms for matrix rank minimization [J]. Journal of Machine Learning Research, 2012,13:3441-3473

[42] Xin Yu, Tommi Jaakkola T. Primal-Dual methods for sparse constrained matrix completion[J]. Journal of Machine Learning Research-Proceedings Track,2012,22:1323-1331

[43] Liu Ji, 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

[44] Signoretto M, Van de Plas, R, De Moor B, et al. Tensor versus matrix completion:a comparison with application to spectral data[J]. IEEE Signal Processing Letters,2011,18(7):403-406

[45] 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

[46] Julià C, Sappa A D, Lumbreras F, et al. Rank estimation in missing data matrix problems[J]. Journal of Mathematical Imaging and Vision,2011,39:140-160

(上接第 12 页)

[25] Bilenko M, Mooney R, Cohen W, et al. Adaptive name matching in information integration[J]. IEEE Intelligent Systems, 2003, 18(5):16-23

[26] Gravano L, Ipeirotis P G, Koudas N, et al. Text joins in an RDBMS for web data integration[C]// Proceedings of the 12th international conference on World Wide Web. ACM, 2003:90-101

[27] Gill L. OX-LINK: The Oxford Medical Record Linkage System [C]// Proc. Int'l Record Linkage Workshop and Exposition. 1997:15-33

[28] 刁兴春,谭明超,曹建军. 一种融合多种编辑距离的字符串相似度计算方法[J]. 计算机应用研究,2010,27(12):4523-4525

[29] Wang J, Li G, Yu J X, et al. Entity matching: how similar is similar[J]. Proceedings of the VLDB Endowment, 2011, 4(10):622-633

[30] Herzog T N, Scheuren F J, Winkler W E. Data quality and record linkage techniques[M]. Springer, 2007

[31] Verykios V S, Moustakides G V, Elfeky M G. A Bayesian decision model for cost optimal record matching [J]. The VLDB Journal, 2003, 12(1):28-40

[32] Naumann F, Herschel M. An introduction to duplicate detection [J]. Synthesis Lectures on Data Management, 2010, 2(1):1-87

[33] Fan W, Jia X, Li J, et al. Reasoning about record matching rules [J]. Proceedings of the VLDB Endowment, 2009, 2(1):407-418

[34] Cochinwala M, Kurien V, Lalk G, et al. Efficient data reconciliation[J]. Information Sciences, 2001, 137(1):1-15

[35] Christen P. Automatic record linkage using seeded nearest neighbour and support vector machine classification[C]// Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008:151-159

[36] Christen P. Automatic training example selection for scalable unsupervised record linkage[M]// Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2008:511-518

[37] Arasu A, Götz M, Kaushik R. On active learning of record matching packages[C]// Proceedings of the 2010 International Conference on Management of Data. ACM, 2010:783-794

[38] Whang S E, Garcia-Molina H. Entity resolution with evolving rules[J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 1326-1337

[39] Monge A E. Matching algorithms within a duplicate detection system[J]. IEEE Data Engineering Bulletin, 2000, 23(4):14-20

[40] Hassanzadeh O, Miller R J. Creating probabilistic databases from duplicated data[J]. The VLDB Journal—The International Journal on Very Large Data Bases, 2009, 18(5):1141-1166

[41] Hernández M A, Stolfo S J. The merge/purge problem for large databases[C]// ACM SIGMOD Record. ACM, 1995, 24(2):127-138

[42] Chaudhuri S, Ganti V, Motwani R. Robust identification of fuzzy duplicates[C]// Proceedings of 21st International Conference on Data Engineering 2005. IEEE, 2005:865-876

[43] Kalashnikov D V, Mehrotra S. Domain-independent data cleaning via analysis of entity-relationship graph[J]. ACM Transactions on Database Systems (TODS), 2006, 31(2):716-767

[44] Dong X, Halevy A, Madhavan J. Reference reconciliation in complex information spaces [C] // Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. ACM, 2005:85-96

[45] Bhattacharya I, Getoor L. Collective entity resolution in relational data[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1):5

[46] Rastogi V, Dalvi N, Garofalakis M. Large-scale collective entity matching[J]. Proceedings of the VLDB Endowment, 2011, 4 (4):208-218

[47] Fan W. Dependencies revisited for improving data quality[C]// Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. ACM, 2008:159-170