软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2017,28(6):1547−1564 [doi: 10.13328/j.cnki.jos.005260]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

矩阵补全模型及其算法研究综述*

陈 蕾[1,2,3], 陈松灿[1]

[1](南京航空航天大学 计算机科学与技术学院,江苏 南京 210016)

[2](江苏省无线传感网高技术研究重点实验室(南京邮电大学),江苏 南京 210003)

[3](南京邮电大学 计算机学院,江苏 南京 210003)

通讯作者: 陈蕾, E-mail: chenlei@njupt.edu.cn

摘 要: 近年来,随着压缩感知技术在信号处理领域的巨大成功,由其衍生而来的矩阵补全技术也日益成为机器学习领域的研究热点,诸多研究者针对矩阵补全问题展开了大量卓有成效的研究.为了更好地把握矩阵补全技术的发展规律,促进矩阵补全理论与工程应用相结合,针对矩阵补全模型及其算法进行了综述.首先,对矩阵补全技术进行溯源,介绍了从压缩感知到矩阵补全的自然演化历程,指出压缩感知理论的发展为矩阵补全理论的形成奠定了基础;其次,从非凸非光滑秩函数松弛的角度将现有矩阵补全模型进行分类,旨在为面向具体应用的矩阵补全问题建模提供思路;然后综述了适用于矩阵补全模型求解的代表性优化算法,其目的在于从本质上理解各种矩阵补全模型优化技巧,从而有利于面向应用问题的矩阵补全新模型求解;最后分析了矩阵补全模型及其算法目前存在的问题,提出了可能的解决思路,并对未来的研究方向进行了展望.

关键词: 稀疏学习;矩阵补全;压缩感知;矩阵分解;随机优化

中图法分类号: TP181

中文引用格式: 陈蕾,陈松灿.矩阵补全模型及其算法研究综述.软件学报,2017,28(6):1547−1564. http://www.jos.org.cn/1000-9825/5260.htm

英文引用格式: Chen L, Chen SC. Survey on matrix completion models and algorithms. Ruan Jian Xue Bao/Journal of Software, 2017,28(6):1547−1564 (in Chinese). http://www.jos.org.cn/1000-9825/5260.htm

Survey on Matrix Completion Models and Algorithms

CHEN Lei[1,2,3], CHEN Song-Can[1]

[1](College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

[2](Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks (Nanjing University of Posts and Telecommunications), Nanjing 210003, China)

[3](School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: In recent years, with the great success of compressed sensing (CS) in the field of signal processing, matrix completion (MC), derived from CS, has increasingly become a hot research topic in the field of machine learning. Many researchers have done a lot of fruitful studies on matrix completion problem modeling and their optimization, and constructed relatively complete matrix completion

* 基金项目: 国家自然科学基金(61472186, 61572263, 61403208); 江苏省自然科学基金(BK20161516, BK20151511); 中国博士后科学基金(2015M581794); 江苏省高校自然科学研究面上项目(15KJB520027); 江苏省博士后科研资助计划(1501023C); 南京邮电大学校级科研基金(NY214127, NY215097)

Foundation item: National Natural Science Foundation of China (61472186, 61572263, 61403208); National Science Foundation of Jiangsu Province (BK20161516, BK20151511); Postdoctoral Science Foundation of China (2015M581794); Project of Natural Science Research of Jiangsu University (15KJB520027); Postdoctoral Science Foundation of Jiangsu Province (1501023C); NUPTSF (NY214127, NY215097)

收稿时间: 2016-05-26; 修改时间: 2016-11-03; 采用时间: 2017-01-19; jos 在线出版时间: 2017-02-20

CNKI 网络优先出版: 2017-02-20 14:05:55, http://www.cnki.net/kcms/detail/11.2560.TP.20170220.1405.018.html

theory. In order to better grasp the development process of matrix completion, and facilitate the combination of matrix completion theory and engineering applications, this article reviews the existing matrix completion models and their algorithms. First, it introduces the natural evolution process from CS to MC, and illustrates that the development of CS theory has laid the foundation for the formation of MC theory. Second, the article summarizes the existing matrix completion models into the four classes from the perspective of the relaxation of non-convex and non-smooth rank function, aiming to provide reasonable solutions for specific matrix completion applications; Third, in order to understand the inherent optimization techniques and facilitate solving new problem-dependent matrix completion model, the article studies the representative optimization algorithms suitable for various matrix completion models. Finally, article analyzes the existing problems in current matrix completion technology, proposes possible solutions for these problems, and discusses the future work.

Key words: sparse learning; matrix completion; compressed sensing; matrix decomposition; stochastic optimization

近年来,随着压缩感知[1,2]理论的流行,矩阵补全技术(matrix completion)[3]也越来越受到研究者的广泛关注.众所周知,压缩感知研究的是如何根据低维的测量值向量恢复出高维稀疏的原始信号向量.而在很多实际问题中,例如图像和视频处理、推荐系统、文本分析等等,待恢复的数据通常是用矩阵表示的,这使得对问题的理解、建模、处理和分析更为方便.然而,这些数据经常面临缺失、损毁和噪声污染等问题,如何在这些情形下得到准确的数据,就是研究者们所要解决的问题.因此,压缩感知理论便自然地从向量空间被拓展至矩阵空间,从而利用矩阵的二阶稀疏性(即矩阵的低秩性),通过采样矩阵的部分元素来恢复目标矩阵.在机器学习领域,这类问题通常被刻画为矩阵补全问题.迄今为止,矩阵补全理论已在图像修复[4]、多标记图像分类[5]、网络异常流量监测[6]、无线传感器网络数据收集[7]、核磁共振图像分割[8]、声源定位与阵列校准[9]、社交网络链接关系预测[10]、运动侦测与估计[11]、半监督聚类[12]、推荐系统[13]、相位检索[14]、地震数据重构[15]、问答社区专家发现[16]、无线传感器网络节点定位[17]、Web 服务 QoS 属性预测[18]以及 Web 服务标签修正[19]等领域得到了重要应用,已逐渐成为机器学习、模式识别以及计算机视觉领域中的主要研究热点之一.

就矩阵补全理论的学术影响而言,在著名的学术论文数字图书馆 ACM,IEEE,Elsevier,Springer 以及 Google Scholar 中用“Matrix Completion”作为关键字进行检索发现:与矩阵补全有关的学术论文发表数量,近年来一直呈增长趋势.国内外众多研究机构如卡内基梅隆大学[20]、斯坦福大学[21]、加州理工学院[22]、明尼斯达大学[23]、亚利桑那州立大学[24]、哥伦比亚大学[25]、莱斯大学[26]、华盛顿大学[27]、加州大学伯克利分校[28]、加州大学洛杉矶分校[29]、IBM [30]、德克萨斯州大学奥斯汀分校[31]、新加坡国立大学[32]、密歇根州立大学[33]、鲁汶大学[34]、伊朗沙力夫理工大学[35]、滑铁卢大学[36]、香港中文大学[37]、北京大学[38]、清华大学[39]、南京大学[40]、上海交通大学[41]、西安电子科技大学[42]、浙江大学[43]、南京邮电大学[44]等研究机构都开展了矩阵补全研究工作.他们的研究成果不仅推动了矩阵补全理论的发展,而且展现了矩阵补全理论诱人的应用前景.

本文首先对标准的矩阵补全模型进行溯源,探讨压缩感知和矩阵补全的关系,指出压缩感知的相关理论和技术可以自然地推广到矩阵补全;然后综述近年来涌现的各种矩阵补全模型,从非凸非光滑秩函数松弛的角度将其分为基于核范数松弛的矩阵补全模型、基于矩阵分解的矩阵补全模型、基于非凸函数松弛的矩阵补全模型和其他类型的矩阵补全模型;接着,从独立于问题模型的角度综述适用于矩阵补全模型求解的代表性优化方法,其目的在于从本质上理解矩阵补全模型优化技巧,从而有利于问题依赖的新模型求解;最后,本文分析了现有矩阵补全研究存在的问题,指出现有矩阵补全模型及其算法仍然存在噪声容错性差、归纳推广性不足和先验信息融合性不强等问题,同时,所设计的算法大多受制于单机内存和计算效率的约束,难以适用于大规模问题求解;针对这些问题,作者提出了可能的解决思路,并对未来研究方向进行了展望.

本文第 1 节对矩阵补全技术进行溯源,指出压缩感知理论的发展为矩阵补全理论的形成奠定了基础.第 2 节综述现有矩阵补全模型.第 3 节综述适用于矩阵补全模型求解的代表性优化算法.第 4 节对矩阵补全模型及其算法研究目前存在的问题进行分析,指出可能的解决方案,并对未来研究工作进行展望.最后是本文的总结.

1 矩阵补全溯源

由 Donoho[1]提出的压缩感知技术是信号处理领域的研究热点,其核心思想是:基于信号的可压缩性或稀疏

性,通过低分辨率、欠 Nyquist 采样数据的非相关观测来实现高维信号的感知.压缩感知理论突破了香农定理对信号采样频率的限制,能够以较少的采样资源、较高的采样速度和较低的软硬件复杂度获得原始型号的测量值,已被广泛应用于数字相机、医学成像、遥感成像、地震勘探、多媒体混合编码、通讯等领域,在理论和应用方面均取得了极大成功[39].从数学上来看,压缩感知主要研究对于以向量 $x\in\mathbb{R}^n$ 表示的稀疏信号,如何在 $m<<n$ 的情形下仅通过测量较少的采样 $y\in\mathbb{R}^m$ 并能从中恢复出原始信号 x.标准的压缩感知问题可建模为

$$\min_{x\in\mathbb{R}^n}\|x\|_0 \ \text{ s.t. } Ax=y \tag{1}$$

其中,矩阵 $A\in\mathbb{R}^{m\times n}$ 称为测量矩阵.可以证明只要测量矩阵 A 中任意 $2k$ 列向量都是线性无关的,那么至少存在一个 k-稀疏的解向量 x_0 满足 $y=Ax_0$.但不幸的是,求解该问题是 NP 难的.为此,Candes 和 Tao 等人[2]合作证明了在约束等距性(restricted isometry property,简称 RIP)条件下,向量空间 L_0 范数优化问题(1)与如下 L_1 范数优化问题等价:

$$\min_{x\in\mathbb{R}^n}\|x\|_1 \ \text{ s.t. } Ax=y \tag{2}$$

由于 L_1 范数优化问题(2)是凸优化问题,所以其具有唯一解,且该唯一解即为 L_0 范数优化问题(1)的唯一解.这是一种非常重要的松弛技巧,已经在机器学习领域得到了广泛引用.

然而在实际应用中,由于硬件设备或者环境影响等原因,信号测量过程通常容易引入误差,当采样值存在误差时,标准的压缩感知问题可修正为

$$\min_{x\in\mathbb{R}^n}\|x\|_0 \ \text{ s.t. } \|Ax-y\|_2\leqslant\varepsilon \tag{3}$$

该问题既可以松弛为经典的规一化 LASSO(least absolute shrinkage and selection operator)形式[45]:

$$\min_{x\in\mathbb{R}^n}\|Ax-y\|_2^2 \ \text{ s.t. } \|x\|_1\leqslant\varepsilon \tag{4}$$

也可以松弛为如下的拉格朗日乘子形式:

$$\min_{x\in\mathbb{R}^n}\|x\|_1+\frac{\lambda}{2}\|Ax-y\|_2^2 \tag{5}$$

压缩感知理论得以成功应用的一个重要前提是信号向量的稀疏性,但在很多实际问题中,我们面临的数据并不是向量空间中的一维数据,而是矩阵空间中的二维数据,因此,如何选择矩阵稀疏性的度量准则就成为将压缩感知理论由向量空间推广到矩阵空间的一个关键问题.如果将矩阵的低秩性视为矩阵稀疏性,那么向量空间的压缩感知理论就自然拓展为矩阵空间的矩阵补全理论.矩阵补全理论主要研究如何在数据不完整的情况下将矩阵空间的缺失数据进行填补,即:对于某个矩阵,假设只能采样到矩阵的一部分元素,其他一部分或者大部分元素由于各种原因丢失了或无法得到,如何将这些缺失的元素合理准确地填补.标准矩阵补全问题可建模为如下形式的秩最小化约束优化模型[3]:

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} rank(X) \text{ s.t. } P_\Omega(M)=P_\Omega(X) \tag{6}$$

其中,$\Omega\subseteq[n_1]\times[n_2]$($[n_1]=\{1,\ldots,n_1\}$,$n_2=\{1,\ldots,n_2\}$)为采样元素的索引集合;$P_\Omega(\cdot)$为正交投影算子,表示当$(i,j)\in\Omega$时,$M_{ij}$ 为采样元素,即:

$$[P_\Omega(M)]_{ij}=\begin{cases}M_{ij}, & \text{if } (i,j)\in\Omega\\ 0, & \text{otherwise}\end{cases} \tag{7}$$

当采样数据存在误差时,上述模型可进一步修正为

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} rank(X) \text{ s.t. } \|P_\Omega(M-X)\|_F\leqslant\delta \tag{8}$$

理论上,Candes 和 Recht[3]证明了在不相关假设条件下(即,目标矩阵的 k 个左奇异向量和 k 个右奇异向量都均匀分布在所有由 k 个正交归一化向量组成的集合中),从某个秩为 r 的矩阵 $M\in\mathbb{R}^{n\times n}$ 中均匀采样 m 个元素,且满足 $m\geqslant Cn^{5/4}r\log n$,则矩阵精确恢复的概率为 $p\geqslant 1-cn^{-3}$.进一步,如果 $r\leqslant n^{1/5}$,则采样个数 m 满足条件 $m\geqslant Cn^{6/5}r\log n$ 时能够以 $1-cn^{-3}$ 的概率精确地恢复出原矩阵.

实际上,如果我们将公式(6)中的正交投影算子理解为线性算子:$[P_\Omega(M)]_{ij}=\langle e_ie_j^T,M\rangle$,其中,$e_i$ 和 e_j 分别表示

欧氏空间中的标准正交基,$\langle\cdot,\cdot\rangle$表示矩阵空间的标准内积运算.由此可以看出:矩阵补全问题中的元素采样是通过特殊矩阵 $e_ie_j^T$ 与目标矩阵的内积运算获取的,而压缩感知问题中的元素采样则是通过测量矩阵中的行向量与原始信号向量的内积运算获取的.如果将特殊矩阵 $e_ie_j^T$ 组合成一个大矩阵集合,则该矩阵集合与压缩感知问题中的测量矩阵获得了形式上的统一.进一步地,我们不难发现:压缩感知和矩阵补全不仅在采样约束的形式上是一致的,而且都是通过稀疏约束缩小问题求解的解空间,所不同的是压缩感知理论是基于原始信号的向量稀疏性对数据进行重构,而矩阵补全理论则是基于目标矩阵的低秩性对缺失数据进行补全.更加值得注意的是:由于矩阵的秩在数值上等于矩阵奇异值向量中非零元素的个数,目标矩阵的低秩性因此等同于其奇异值向量的稀疏性,这又与压缩感知理论中原始信号的向量稀疏性是一致的.正是由于矩阵补全和压缩感知在采样约束和稀疏先验上的一致性,才使得压缩感知领域的诸多理论和方法得以直接推广到矩阵补全领域,从而为矩阵补全理论的形成奠定了良好的基础.

2 矩阵补全模型

由于秩函数 $rank(\cdot)$的非凸性和非光滑性,标准矩阵补全问题求解成为了一个NP难问题,在所有已知的求解算法中,其求解复杂度均随着矩阵维数的增加呈平方倍指数增长.为此,众多研究者对此展开了广泛研究,取得了大量成果[20−56].从秩函数松弛的角度分析,这些研究成果大致可分为以下4类,即:基于核范数松弛的矩阵补全模型、基于矩阵分解的矩阵补全模型、基于非凸函数松弛的矩阵补全模型及其他类型的矩阵补全模型.

2.1 基于核范数松弛的矩阵补全模型

Fazel 在文献[46]中证明了矩阵核范数是秩函数在矩阵谱范数意义下单位球上的最佳凸逼近.因此,类似于压缩感知理论中常用的将向量 L_0 范数松弛为向量 L_1 范数的技巧,为了使标准矩阵补全问题易于求解,一个自然想法就是利用凸核范数代替非凸秩函数,将标准的矩阵补全问题松弛为如下形式的凸约束优化模型:

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \|X\|_* \ \text{ s.t. } P_\Omega(M)=P_\Omega(X) \tag{9}$$

随后,一系列基于核范数松弛的矩阵补全模型被陆续提出.例如,Ma[25]和 Toh[47]等人同时将标准矩阵补全问题松弛为如下形式的矩阵 LASSO 模型:

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \lambda\|X\|_*+\frac{1}{2}\|P_\Omega(M)-P_\Omega(X)\|_F^2 \tag{10}$$

并且分别采用近邻梯度算法和加速近邻梯度算法求解.然而,矩阵LASSO模型对于相关性非常强的数据通常会出现解不稳定的情况.Cai等人[48]则引入弹性网(elastic-net)正则化项来增加矩阵补全问题求解的稳定性,将标准矩阵补全问题建模为如下形式:

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \lambda\|X\|_*+\frac{1}{2}\|X\|_F^2 \ \text{ s.t. } P_\Omega(M)=P_\Omega(X) \tag{11}$$

该模型采用线性 Bregman 迭代算法求解,然而由于模型构建时没有显式考虑噪声信息,因此该模型对采样矩阵的容错性不是很好.林宙辰等人[49]则将矩阵补全问题视为 Robust PCA 问题的特例,将矩阵补全问题建模为如下形式:

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \|X\|_* \ \text{ s.t. } X+E=M, P_\Omega(E)=0 \tag{12}$$

该模型采用了和交替方向乘子法等价的非精确增广拉格朗日乘子法求解.同样是采用核范数逼近矩阵秩函数,但 Hu 等人[43]认为,这些核范数最小化方法有一个共同的缺陷在于,它们都是通过同时最小化所有奇异值之和而获得结果矩阵的低秩性,这将导致所求解的结果矩阵虽然低秩但却不能很好地逼近真实解.因此,他们将最小的那些奇异值之和定义为截断核范数$\|X\|_r$,并采用截断核范数来代替标准核范数$\|X\|_*$,从而将矩阵补全问题建模为如下形式:

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \|X\|_r \ \text{ s.t. } P_\Omega(M)=P_\Omega(X) \tag{13}$$

该模型分别采用了交替方向乘子法和加速近邻梯度算法求解,文献[43]的实验结果表明,该模型较好地解

决了所求解的结果矩阵不能精确地逼近目标矩阵真实解的问题,矩阵补全精度得到了很大程度的提高.但该模型的不足之处在于,问题求解过程中要预先估计矩阵的秩信息.

2.2 基于矩阵分解的矩阵补全模型

基于核范数松弛的建模方法由于涉及复杂的矩阵奇异值分解,从而导致模型的求解效率和可扩放性(scalability,也称为可扩展性)受限.基于矩阵分解的建模方法是一类可替代的矩阵补全模型构建方法,其基本思路是:将目标矩阵分解为两个低秩矩阵的乘积,从而避免了复杂的矩阵奇异值分解,加速了算法的执行效率.例如,Wen 等人[50]将矩阵补全问题建模为如下形式:

$$\min_{U\in\mathbb{R}^{n_1\times k},V\in\mathbb{R}^{k\times n_2},Z\in\mathbb{R}^{n_1\times n_2}} \|UV-Z\|_F^2 \ \text{ s.t. } P_\Omega(Z)=P_\Omega(M) \tag{14}$$

其中,k 是预测的矩阵秩界.该模型采用分块坐标下降算法(俗称交替最小化算法)求解,如果能够预先获取合适的 k 值,该模型可以在较小的时间复杂度内获得相当精度的解.焦李成教授等人[51]也基于矩阵分解的思想提出了一种矩阵双分解模型:

$$\min_{L\in\mathbb{R}^{n_1\times k},Q\in\mathbb{R}^{k\times n_2}} \left\{\lambda\|Q\|_*+\frac{1}{2}\|P_\Omega(LQ-W)\|_F^2,\ \text{s.t. } L^TL=I\right\} \tag{15}$$

该模型采用交替方向乘子法求解,合成数据和真实数据集上的实验结果均表明,该模型无论是补全精度还是收敛速度都明显优于已有的核范数松弛模型.此外,Srebro 等人在文献[52]中已经证明:对任意秩为 r 的矩阵 $X\in\mathbb{R}^{n_1\times n_2}$,若 $k>r$,有如下等式成立:

$$\|X\|_*=\min_{L\in\mathbb{R}^{n_1\times k},Q\in\mathbb{R}^{n_2\times k}} \frac{1}{2}\{\|L\|_F^2+\|Q\|_F^2 \ \text{ s.t. } X=LQ^T\} \tag{16}$$

基于此,标准矩阵补全问题也可以直接建模为

$$\min_{L\in\mathbb{R}^{n_1\times k},Q\in\mathbb{R}^{k\times n_2}} \frac{1}{2}\|P_\Omega(M-LQ)\|_F^2+\frac{\lambda}{2}(\|L\|_F^2+\|Q\|_F^2) \tag{17}$$

从而,采用分块坐标下降算法即可容易求解.然而,由于问题(15)非凸,该问题可能存在非全局最优的驻点解.幸运的是,Mardani 等人在文献[53]中已经证明,只要满足如下较为宽松的条件:

$$\|P_\Omega(M-LQ)\|_F\leqslant\lambda \tag{18}$$

则问题(16)的解 $\hat{X}=\hat{L}\hat{Q}$ 即为全局最优解.

2.3 基于非凸函数松弛的矩阵补全模型

基于非凸函数松弛的建模方法是基于核范数松弛建模方法的另一种可替代方法.研究者认为:虽然核范数是秩函数在单位球内的最佳凸逼近,但二者之间仍存在较大差异.因此,他们提出可以引入一些更加紧致的非凸函数来松弛秩函数.例如,Nie 等人在文献[31]中将矩阵 Schatten p-范数引入模型中代替秩函数,将矩阵补全问题建模为

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \lambda\|X\|_{S_p}^p+\|P_\Omega(M-X)\|_P^p \tag{19}$$

其中,$\|X\|_{S_p}^p=\sum_{i=1}^{\min\{n_1,n_2\}}\sigma_i^p(X)$,$\|X\|_p^p=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}|X_{ij}|^p$,$0<p\leqslant1$,$\sigma_i(X)$表示矩阵 X 的第 i 个奇异值.该模型采用增广拉格朗日乘子法求解,数值实验结果表明,用 Schatten p-范数松弛秩函数比用核范数松弛秩函数能够得到更加精确的补全性能.Ghasemi 等人在文献[35]中则用高斯函数:

$$f_\delta(s_i)=\exp(-s_i^2/2\delta^2) \tag{20}$$

替代秩函数,将标准矩阵补全问题建模为

$$\min_{X\in\mathbb{R}^{n_1\times n_2}} \sum_{i=1}^{\min\{n_1,n_2\}} f_\delta(\sigma_i(X)) \text{ s.t. } P_\Omega(M)=P_\Omega(X) \tag{21}$$

Ghasemi 等人充分利用高斯函数可微的性质,采用梯度投影算法进行求解.受 Ghasemi 等人的启发,Geng 等人[54]提出采用光滑双曲正切函数:

$$f_\delta(s_i)=\frac{\exp(s_i^2/2\delta^2)-\exp(-s_i^2/2\delta^2)}{\exp(s_i^2/2\delta^2)+\exp(-s_i^2/2\delta^2)} \tag{22}$$

代替秩函数建立光滑非凸优化模型.同样采用梯度投影算法进行求解,对随机生成的低秩矩阵补全问题和图像恢复问题的测试结果表明,其所提出的算法能够得到更好的补全精度.此外,Kang 等人在文献[55]中还提出,可以采用非凸光滑的 logdet($I+X^TX$)函数作为秩函数 $rank(X)$的包络来对矩阵补全问题建模,该模型求解可以采用交替方向乘子法.

2.4 其他类型的矩阵补全模型

基于核范数松弛的矩阵补全模型涉及复杂的矩阵奇异值分解,尽管算法实现时可以调用 PROPACK 软件包(http://sun.stanford.edu/~rmunk/PROPACK/)进行部分奇异值分解,从而有效降低奇异值分解的时间复杂度,一定程度上加速算法收敛速度,但仍然无法适用于大规模矩阵补全问题.Wang 等人[24]提出将向量空间的正交匹配追踪技术拓展到矩阵空间,构建了一种有效且可扩放的基于 R1MP(rank-one matrix pursuit)的矩阵补全模型,该模型的主要技巧是,将待补全的目标矩阵看作是若干基矩阵的线性组合,即:

$$X=R(\theta)=\sum\theta_iR_i \tag{23}$$

其中,R_i 为秩为 1 且具有单位 Frobenius 范数的基矩阵.在这个意义上,标准矩阵补全问题可建模为

$$\min_{\theta,R}\|\theta\|_0 \ \text{ s.t. } P_\Omega(M)=P_\Omega(R(\theta)) \tag{24}$$

Wang 等人很自然地将该问题进一步转换为噪声模型:

$$\min_{\theta,R}\|P_\Omega(R(\theta))-P_\Omega(M)\|_F^2 \ \text{ s.t. } \|\theta\|_0\leqslant r \tag{25}$$

该模型采用正交匹配追踪算法求解,实验结果表明:即使对于高达 69878×10677 阶的 Movielens 10M 数据集,其运算时间也仅达 13.79s,比采用奇异值分解的矩阵补全算法运算时间提高了 10 倍~100 倍.

2.5 各类矩阵补全模型优缺点分析

上述 4 类矩阵补全模型关注的都是如何基于目标矩阵的先验低秩性从少量采样观察中补全缺失元素,主要区别在于秩函数的松弛方式不同,从而导致模型的凸性各异,模型求解效率和可扩放性也因此不同.表 1 分析比较了这 4 类矩阵补全模型的适应场景及各自的优缺点.

Table 1 Application scenerios, advantages and disadvantages analysis of four kinds of matrix completion models

表 1 4 类矩阵补全模型适用场景及优缺点分析

矩阵补全模型	优点	缺点	适用场景
基于核范数松弛的矩阵补全模型	(1) 凸优化模型,存在全局最优解; (2) 核范数近邻算子存在闭式解	(1) 模型求解涉及复杂的奇异值分解,求解效率和可扩放性受限; (2) 核范数不能紧致地逼近目标矩阵的真实秩	小规模问题
基于矩阵分解的矩阵补全模型	(1) 避免了复杂的矩阵奇异值分解; (2) 能够分布式并行实现	(1) 非凸优化模型,可能存在非全局最优的驻点解; (2) 需要预先估计目标矩阵的秩信息	大规模问题
基于非凸函数松弛的矩阵补全模型	非凸函数逼近秩函数的紧致性增强,补全性能好	(1) 非凸优化模型,可能存在非全局最优的驻点解; (2) 模型求解效率和可扩放性受限	小规模问题
其他类型的矩阵补全模型	模型求解效率和可扩放性好	(1) 非凸优化模型,可能存在非全局最优的驻点解; (2) 需要预先估计目标矩阵的秩信息	大规模问题

2.6 正则化技术在矩阵补全模型中的应用

正则化技术诞生于 20 世纪 60 年代,最初在数学领域提出,用于解决不适定的反问题,基本思想是,通过引入含有解的先验知识的非负辅助泛函正则化项来增加解的稳定性[56].随着 20 世纪 80 年代机器学习的兴起,正则化技术被广泛应用于分类器设计领域,并衍生出许多著名的算法[57].这些算法大部分都能归纳为经验风险最小化正则化算法的框架,不同之处在于所用的正则化项,不同的正则化项能产生不同性质的解,而如何选取正则化项,通常由问题的具体要求决定.例如,为了增加矩阵补全模型解的稳定性,通常将 Tikhonov 正则化技术(Frobenius 范数正则化)引入目标函数[48],其作用一是能将目标函数由凸变为强凸,从而使得目标问题最优解唯

一;二是能够组合目标矩阵中的高度相关变量,有利于克服数据强相关所导致的解不稳定性;三是由于目标矩阵的 Frobenius 范数等同于其奇异值向量的 L_2 范数,因此有利于目标矩阵奇异值收缩,从而更好地逼近原始矩阵的真实秩.正则化技术在矩阵补全模型中的另一个典型应用是各类噪声的拟合,例如,研究者们通常将实际应用中遇到的高斯噪声类型拟合为高斯分布模型(Frobenius 范数),将野值噪声类型拟合为拉普拉斯分布模型(矩阵 L_1 范数),将结构化带状噪声类型拟合为矩阵 $L_{2,1}$ 范数模型.关于噪声拟合的详细阐述参见本文第 4.1 节.虽然正则化技术在矩阵补全模型稳定性和噪声解析方面应用较为广泛,但是正则项权重参数的设定仍然是一个开放性问题,目前广泛采用的策略依然是传统的网格搜索和交叉验证方法.

3 矩阵补全模型优化算法

本节将综述适用于求解矩阵补全模型的优化算法.为便于描述,我们首先将约束优化类矩阵补全模型统一为如下形式的约束优化一般框架(F1):

$$\text{(F1)}: \min_{X_1,\dots,X_m\in\mathbb{R}^{n_1\times n_2}} J(X_1,\dots,X_m) \text{ s.t. } C(X_1,\dots,X_m)=0 \tag{26}$$

将无约束优化类矩阵补全模型统一为如下形式的无约束优化一般框架(F2):

$$\text{(F2)}: \min_{X_1,\dots,X_m\in\mathbb{R}^{n_1\times n_2}} F(X_1,\dots,X_m)=J(X_1,\dots,X_m)+H(X_1,\dots,X_m) \tag{27}$$

事实上,如果令 $H(X)=\mu\|C(X)\|_F^2$,则上述无约束优化一般框架(F2)可视为约束优化一般框架(F1)的噪声版本,更具一般性.进一步,如果令 $X=(X_1,\dots,X_m)$,则无约束优化一般框架(F2)可简化为

$$\text{(F3)}: \min_{X=(X_1,\dots,X_m)\in\mathbb{R}^{m\times n}} F(X)=J(X)+H(X) \tag{28}$$

结合这两类矩阵补全模型一般框架的特点,综合分析现有矩阵补全模型所采用的优化求解算法不难发现,适用于矩阵补全问题模型求解的代表性算法包括近邻梯度下降算法(proximal gradient descend,简称 PGD)、分块坐标下降算法(block coordinate descent,简称 BCD)、Bregman 迭代算法(bregman iterative,简称 BI)、交替方向乘子法(alternating direction method of multipliers,简称 ADMM)、随机优化算法(stochastic optimization,简称 SO).

3.1 近邻梯度下降算法

近邻梯度下降算法[58](也称近邻前向后向分裂算法(proximal forward backward splitting algorithm,简称 PFBS)适用于求解无约束优化一般框架(F3):

$$\min_{X\in\mathbb{R}^{m\times n}} J(X)+H(X) \tag{29}$$

假定 $J(X)$是$\mathbb{R}^{m\times n}$ 上具有下半连续性质的凸函数,$H(X)$是$\mathbb{R}^{m\times n}$ 上凸光滑下半连续函数且具有 L_J-Lipschitz 连续导数,即:

$$\|\nabla J(X)-\nabla J(Y)\|_F\leqslant L_J\|X-Y\|_F,\forall X,Y\in\mathbb{R}^{m\times n} \tag{30}$$

则对任意的初始值 X^0 及 $0<\delta<1/L_J$,用如下方法生成的迭代序列 X^{k+1} 收敛到无约束优化一般框架(F3)的唯一解:

$$X^{k+1}=prox_{\delta J}(X^k-\delta\nabla H(X^k))=\arg\min_{X\in\mathbb{R}^{m\times n2}} \delta J(X)+\frac{1}{2}\|X-(X^k-\delta\nabla H(X^k))\|_F^2 \tag{31}$$

算法 1 描述了近邻梯度下降算法的详细步骤.

算法 1. 近邻梯度下降算法(proximal gradient descent,简称 PGD).

Input: maximum iteration time N;

Output: X^{opt}.

Initialize $X^0=0$;

for $k=0,\dots,N$ **do**

$$X^{k+1}=prox_{\delta J}(X^k-\delta\nabla g(X^k))=\arg\min_{X\in\mathbb{R}^{m\times n}}\left\{\delta J(X)+\frac{1}{2}\|X-(X^k-\delta\nabla H(X^k))\|_F^2\right\};$$

if stopping criterion is satisfied **then**

return X^{k+1}

end if

end for

虽然上述近邻梯度下降算法的收敛率可达到 $O(N^{-1})$,然而 Nemirovski 等人[59]的研究结果早已表明,采用一阶优化算法求解无约束凸优化问题的收敛率理论上可以达到 $O(N^{-2})$.受此激励和启发,Beck 和 Teboulle[60]把加速一阶凸光滑优化问题的 Nesterov 技巧[61]进行推广,提出了收敛率为 $O(N^{-2})$的加速近邻梯度算法如下.

算法 2. 加速近邻梯度算法(accelerated proximal gradient,简称 APG).

Input: maximum iteration time N;

Output: X^{opt}.

Initialize $X^0=Y^0=0$, $t_1=1$;

for $k=0,\ldots,N$ **do**

$$X^{k+1} = prox_{\delta J}(Y^k - \delta\nabla g(Y^k)) = \arg\min_{X\in\mathbb{R}^{m\times n}} \left\{\delta J(X) + \frac{1}{2}\| X - (Y^k - \delta\nabla H(Y^k))\|_F^2\right\};$$

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

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

if stopping criterion is satisfied **then**

return X^{k+1}

end if

end for

需要指出的是:(i) 采用 Nesterov 技巧的加速近邻梯度算法并不是下降算法,其收敛过程类似涟漪(ripples)的形状,并不呈单调下降;(ii) 近邻梯度下降算法及加速近邻梯度算法只有在近邻算子易于求解的情况下才有效,所幸,在本文所述的矩阵补全问题中,条件(ii)通常可以满足,因为函数 $J(X)$一般取为矩阵范数,此时,该子问题有闭式解;(iii) 对于矩阵补全问题中常见的线性约束优化模型:

$$\min_{X\in\mathbb{R}^{m\times n}} J(X) \text{ s.t. } AX = B \tag{32}$$

可以把约束项的平方以惩罚函数的方式转化为无约束优化问题:

$$\min_{X\in\mathbb{R}^{m\times n}} J(X) + \frac{\lambda}{2}\| AX - B\|_F^2 \tag{33}$$

再采用加速近邻梯度算法求得近似解.但是需要注意的是:如果一开始让惩罚参数λ很大,则收敛速度将会极其地慢.所以一般采用 Continuation 技巧,即:一开始λ取较小的值,随着迭代进行再逐渐增大到一个较大的值,这样简单处理后,会显著地加快收敛速度.另外,加速近邻梯度算法需要估计 Lipschitz 系数 L_J,如果估计值太大,则会影响收敛速度,所以 Beck 和 Teboulle 还在文献[60]中进一步提出了使用动态估计的 L_J 以加速收敛.

3.2 分块坐标下降算法

在大规模数据处理方面,坐标下降[62]算法以简洁的操作流程和快速的实际收敛效果,成为处理大规模优化问题的首选方法.标准的坐标下降算法是一种非梯度优化算法,在每次迭代中,在当前点处沿一个坐标方向进行一维线搜索(line search),固定其他的坐标方向,以求得目标函数的一个局部极小值.在整个迭代过程中,循环使用不同的坐标方向,一个周期的一维搜索迭代过程相当于一个梯度迭代.对于不可分的目标函数而言,算法可能无法在较小的迭代步数中求得最优解.

由于矩阵补全问题通常表示为形如无约束优化一般框架(F2)的多变量可分离函数优化模型,而标准的坐标下降算法每次只针对一个挑选的坐标分量进行更新,因此该算法不能直接应用于矩阵补全问题模型求解.分块坐标下降算法[63](俗称交替最小化(alternating minimization)算法)每次迭代时以分块的方式更新多个坐标分量.算法 3 描述了分块坐标下降算法的详细步骤.

算法 3. 分块坐标下降算法(block coordinate descent,简称 BCD).

Input: maximum iteration time N;

Output: $(X_1,\ldots,X_m)^{opt}$.

Initialization: choose $(X_1^0,\ldots,X_m^0)$;

for k=1,2,…,N **do**

  for i=1,2,…,m **do**

    $X_i^k=\arg\min_{X_i} F(X_{<i}^k,X_i,X_{>i}^{k-1})$;

  end for

  if stopping criterion is satisfied **then**

    return $(X_1^k,\ldots,X_m^k)$

  end if

end for

分块坐标下降算法具有以下 3 个特点:(i) 如果函数 $F(X_1,\ldots,X_m)$连续可微且每个子问题均可解,那么算法能够保证收敛;(ii) 如果函数 $F(X_1,\ldots,X_m)$非光滑,那么算法不一定收敛;(iii) 如果非光滑部分是可分离的,那么算法也能保证收敛.

3.3 矩阵空间Bregman迭代算法

Bregman 迭代算法起源于凸分析中一种寻求等式约束下目标函数极值的优化方法[64],Osher 等人[65]于 2005 年首先将该方法应用于全变分图像复原,随之被拓展应用到压缩感知[66]、图像去噪[67]等领域,已成为求解低秩矩阵恢复问题的有效方法之一.

对于任一凸函数 $J(X)$在点 $\hat{X}$ 处的 Bregman 距离定义为

$$D_J^P(X,\hat{X})=J(X)-J(\hat{X})-\langle P,X-\hat{X}\rangle \tag{34}$$

其中, $P\in\partial J(\hat{X})$ 为 J 在点 $\hat{X}$ 处的一个次梯度.假设 $J(X)$和 $H(X)$为$\mathbb{R}^{m\times n}$ 上两个凸函数,其中,函数 $H(X)$可微,$J(X)$不可微,考虑无约束优化一般框架:

$$\min_{X\in\mathbb{R}^{m\times n}} J(X)+H(X) \tag{35}$$

可以对不可微函数 $J(X)$引入 Bregman 距离,转化为迭代求解下述子问题序列:

$$X^{k+1}=\arg\min_{X\in\mathbb{R}^{m\times n}} D_J^{P^k}(X,X^k)+H(X)=\arg\min_{X\in\mathbb{R}^{m\times n}} J(X)-\langle P^k,X\rangle+H(X) \tag{36}$$

进一步地,由于 X^{k+1} 是上述子问题的最优解,因此,0 矩阵属于目标函数在 X^{k+1} 处的次微分,即:

$$0\in\partial J(X^{k+1})-P^k+\nabla H(X^{k+1}) \tag{37}$$

由于 $P^{k+1}\in\partial J(X^{k+1})$,化简后可得:

$$P^{k+1}=P^k-\nabla H(X^{k+1}) \tag{38}$$

求解上述无约束优化问题(35)的 Bregman 迭代公式如下:

$$\begin{cases} X^{k+1}=\arg\min_{X\in\mathbb{R}^{m\times n}} J(X)-\langle P^k,X\rangle+H(X) \\ P^{k+1}=P^k-\nabla H(X^{k+1}) \end{cases} \tag{39}$$

算法 4 详细描述了求解无约束优化一般框架的矩阵空间 Bregman 迭代算法实现步骤.

算法 4. 矩阵空间 Bregman 迭代算法.

Input: maximum iteration time N;

Output: X^{opt}.

Initialization: choose $X^0=P^0=0$;

for k=0,1,…,N **do**

$X^{k+1} = \arg\min_{X\in\mathbb{R}^{n_1\times n_2}} J(X) - \langle P^k, X\rangle + H(X)$

$P^{k+1}=P^k-\nabla H(X^{k+1})$

if stopping criterion is satisfied **then**

return X^{k+1}

end if

end for

需要指出的是,矩阵空间 Bregman 迭代算法通常假定子问题 $X^{k+1} = \arg\min_{X\in\mathbb{R}^{n_1\times n_2}} J(X) - \langle P^k, X\rangle + H(X)$ 有解析解.当该子问题没有解析解的时候,可以采用第 3.1 节所述的近邻梯度下降算法或者加速近邻梯度算法迭代求解.考虑到实际上不需要求出子问题的精确解就足以使算法最终收敛到原问题的最优解,因此只需更新 X^{k+1} 一次得到子问题的一个近似解即可,据此可以得到一个更为简洁且实际收敛速度较快的矩阵空间线性 Bregman 迭代算法.此外,与传统的求解约束优化问题的惩罚函数与连续策略相比,Bregman 迭代算法主要有两个优点:(i) Bregman 迭代收敛速度快,特别是当目标函数中包含 L_1 范数正则化项时,只需少许几次迭代求解无约束问题即可获得很好的结果;(ii) 连续策略需要在迭代过程中不断减小步长的取值,而步长的取值在 Bregman 迭代过程中可保持不变,因此可以选择一个合适的步长值最小化子问题的条件数,同时可避免参数调整程中连续策略的数值不稳定性.

3.4 交替方向乘子法

交替方向乘子法[68](alternating direction method of multipliers,简称 ADMM)是一种求解凸优化问题的算法框架,通常也称为 Douglas-Rachford 分裂算法[69],适用于求解分布式凸优化问题.ADMM 算法最早由 Gabay 和 Mercier[70]于 1976 年提出,并被 Boyd 等人于 2011 年重新综述并证明其适用于大规模分布式优化问题.ADMM 算法融合了对偶分解和增广 Lagrangian 的优点,在解决大规模分布式计算系统和面向大规模问题的优化求解中起着非常重要的作用,近年来受到了广泛关注,尤其在机器学习领域得到了很好的应用.

标准 ADMM 算法旨在求解约束优化一般框架(F1)的特例,即,如下带线性约束的可分离目标函数优化问题:

$$\min_{X\in\mathbb{R}^{m\times n}} F_1(X) + F_2(Z) \text{ s.t. } AX + BZ = C \tag{40}$$

该算法先构造相应的增广拉格朗日函数:

$$L_\rho(X,Z,Y) = F_1(X) + F_2(Z) + trace(Y^T(AX+BZ-C)) + (\beta/2)\|AX+BZ-C\|_F^2 \tag{41}$$

然后,通过交替最小化增广拉格朗日函数来更新变量 X,Y,Z.算法 5 描述了标准 ADMM 算法的详细步骤.

算法 5. 交替方向乘子法(alternating direction method of multipliers,简称 ADMM).

Input: maximum iteration time N, parameter β;

Output: X^{opt}, Z^{opt}.

Initialize $Y^0=Z^0=0$;

for k=0,…,N **do**

$X^{k+1} = \arg\min_x L_\rho(X, Z^k, Y^k)$

$Z^{k+1} = \arg\min_z L_\rho(X^{k+1}, Z, Y^k)$;

$Y^{k+1} = Y^k + \rho(AX^{k+1} + BZ^{k+1} - C)$

if stopping criterion is satisfied **then**

return X^{k+1}, Z^{k+1};

end if

end for

为了更好地适应大规模问题求解,众多学者对标准 ADMM 算法进行了改进,分别提出了同步分布式

ADMM[68]、异步分布式 ADMM[71]及快速随机 ADMM[72]等.特别地,对于只涉及两个变量的标准 ADMM 算法,南京大学何炳生等人在文献[73]中已经证明,其收敛率为 $O(N^{-1})$.然而,对于 3 个或 3 个以上变量的凸约束优化问题,如果沿用上述标准的 ADMM 求解步骤,则其收敛性难以保证.为此,何炳生教授等人提出了一种有收敛性保证的基于高斯回代方法的多变量 ADMM 算法[74],北京大学林宙辰教授等人则提出了一种更为简单的有收敛性保证的基于自适应惩罚的并行分裂线性 ADMM 算法[75].

3.5 随机优化算法

随着信息技术和计算技术的迅猛发展,行业应用和科学研究中的数据正以惊人的速度增长,产生了大规模海量数据,机器学习也正面临着数据规模日益扩大的严峻挑战,传统的机器学习优化方法受制于单机计算效率和内存空间的限制,每次迭代都要遍历所有样本,已经不能满足实际应用中大规模问题快速求解的需求.由于随机优化算法通常假设样本是独立同分布的,从而随机抽取单个样本的梯度是目标函数梯度的无偏估计[76],进而可以通过每次迭代时仅随机选用单个或部分样本的方法来代替批处理方法.因此,随机优化算法可视为求解大规模机器学习问题的高效方法之一.常用的随机优化算法包括随机梯度下降算法[77]、随机坐标下降算法[78]、随机交替方向乘子算法[79]和随机近邻梯度下降算法[80]等.鉴于随机梯度下降算法、随机坐标下降算法和随机交替方向乘子算法分别是其各自确定性算法版本的自然推广,本节将主要介绍随机近邻梯度下降算法.

如第 2.1 节所述,基于核范数松弛的矩阵补全模型在求解过程中均不可避免地需要求解如下子问题:

$$\min_{X\in\mathbb{R}^{m\times n}} F(X)=J(X)+\lambda\|X\|_*,\text{其中},J(X)=\frac{1}{2}\|X-W\|_F^2 \tag{42}$$

该子问题固然可以采用奇异值阈值算子求得闭式解,但由于涉及到复杂的奇异值分解,其时间复杂度和空间复杂度分别达到了 $O(mn\min(m,n))$和 $O(mn)$,即使采用改进的部分奇异值分解[81]或者随机奇异值分解[81]算法,依然无法适用于大规模问题.众所周知,随机梯度下降算法通过采用目标函数随机样本梯度的无偏估计代替目标函数所有样本的实际梯度,从而使得每次迭代的计算量非常小,计算复杂度因而不会随着样本量的增加而急剧增加.受此启发,南京大学张利军等人针对现有的近邻梯度下降算法,提出了求解上述子问题的随机近邻梯度下降算法[80],该算法通过计算函数 $J(X)$在 X_t 处的低秩随机无偏梯度 $\widehat{G}_t=G_tY_tY_t^T$ 并引入增量奇异值分解[81],使得算法的时间和空间复杂度分别减少到 $O((m+n)r^2)$和 $O((m+n)r)$,其中,r 为目标矩阵的秩.算法 6 给出了随机近邻梯度下降算法的详细步骤.

算法 6. 随机近邻梯度下降算法(stochastic proximal gradient descent,简称 SPGD).

Input: maximum iteration time N, the regularization parameter λ;

Output: X^{opt}.

Initialize X_1=0;

for t=1,2,…,N **do**

$Y=Z/\sqrt{k}$, where each columns of Z is drawn uniformly at random and independently of each other from $\{\sqrt{n}e_1,...,\sqrt{n}e_n\}$;

$\widehat{G}_t=G_tY_tY_t^T$, where G_t is a subgradient of $J(\cdot)$ at X_t;

$X_{t+1}=\arg\min_{X\in\mathbb{R}^{m\times n}}\frac{1}{2}\|X-(X_t-\eta_t\widehat{G}_t)\|_F^2+\eta_t\lambda\|X\|_*$;

if stopping criterion is satisfied **then**

return X_{t+1};

end if

end for

3.6 各类优化算法优缺点分析

上述 5 类优化算法是求解矩阵补全模型的代表性算法.表 2 分析了这些算法各自的收敛率及其求解优势、

存在缺点,其中:随机优化算法涉及到多种具体算法,限于篇幅,表 2 仅给出了随机近邻梯度下降算法的收敛率;其他随机优化算法的收敛率可查阅相应的参考文献.

Table 2 Convergence rate, advantages and disadvantages analysis of different representative optimizing algorithms

表 2 各类代表性优化算法收敛率及其优缺点分析

优化算法	优点	缺点	收敛率
近邻梯度下降算法	面向小规模问题求解时补全性能好,收敛速度快	求解矩阵补全模型时涉及到复杂的矩阵奇异值分解,因此无法求解中大规模问题	近邻梯度下降算法的收敛率为 $O(N^{-1})$;加速近邻梯度下降算法的收敛率为 $O(N^{-2})$
分块坐标下降算法	应用范围广,简单易实现,是矩阵补全模型求解的基本算法	当目标函数非光滑时算法无收敛性保证	目标函数凸光滑或者非光滑部分可分离时算法有收敛性保证,其收敛率为 $O(N^{-1})$
矩阵空间 Bregman 迭代算法	实际应用中无需求解子问题的精确解,算法实际收敛速度快	不可分布式并行实现,无法求解大规模问题	$O(N^{-1})$
交替方向乘子法	融合了对偶分解和增广 Lagrangian 方法的优点,可分布式并行实现,适用于大规模问题求解	在高精度要求下算法收敛很慢;但在中等精度要求下,ADMM 的收敛速度可以接受(一般迭代数十次即可)	$O(N^{-1})$
随机优化算法	可分布式并行实现,适用于大规模问题求解	仅适用于样本独立同分布情形,实际应用中该假设并不总是成立	随机近邻梯度下降算法的收敛率为 $O(\log N/\sqrt{N})$

4 存在的问题及未来研究方向

4.1 存在的问题

然而,尽管现有矩阵补全研究在模型及算法方面取得了众多进展,但仍然存在如下几点不足.

(1) 噪声容错性差

早期的矩阵补全模型大多只是简单地假设采样矩阵受到单一的高斯噪声污染,并不能解决实际问题中遇到的野值(outlier)噪声和结构化噪声污染问题.为此,针对野值噪声情形下的矩阵补全问题,南京大学陶敏和何炳生等人在文献[82]中采用 Laplacian 分布拟合野值噪声,提出了一种基于核范数松弛的 L_1 范数正则化矩阵补全模型,并设计了相应的交替分裂增广拉格朗日求解算法.Yan 等人则在文献[83]中提出一种基于矩阵分解的 L_1 范数正则化矩阵补全模型,在求解该模型时,他们采用了自适应匹配追踪算法.更进一步,Chen 等人在文献[84]中对何等人的核范数松弛模型进行了严格的理论分析,从理论上证实了基于核范数松弛的 L_1 范数正则化矩阵补全模型的有效性.另一方面,针对结构化噪声污染问题,Chen 等人在文献[85]中率先提出一种基于增广拉格朗日乘子法的核范数松弛矩阵补全模型.类似地,我们也在文献[44]中提出了一类基于线性 Bregman 迭代算法的结构化噪声矩阵补全模型.与已有矩阵补全模型相比,我们的模型不仅能够更好地恢复结构化噪声矩阵的缺失元素,还能精确地辨识出采样矩阵中被污染元素所在行的位置信息.此外,我们也对复合噪声情形下的矩阵补全应用展开了一些探索,在文献[17]中提出了一类适用于高斯、野值复合噪声情形的矩阵补全算法,并将其成功应用于无线传感网节点定位.然而实际问题中,采样矩阵遇到的噪声类型常常是不可预知的,并且通常表现为多种噪声类型的复合.因此在构建矩阵补全问题模型时,如何自适应地拟合不可预知的复杂噪声信息从而提高矩阵补全模型的噪声容错性能,仍然有待进一步的深入研究.

(2) 归纳推广性不足

现有的矩阵补全模型仅能处理观测样本的数据补全问题,而对于未见样本则无法处理,在实际应用中表现出了明显的归纳推广性缺陷.例如:Gogn 等人在文献[13]中将矩阵补全技术应用到推荐系统时,无法克服冷启动问题;Yi 等人在文献[12]中将矩阵补全应用到半监督聚类时,所设计的成对相似性矩阵补全算法不能进行联机半监督聚类,对新加入样本无法进行类别预测;南京大学 Xu 等人在文献[86]中提出了一种融合辅助信息的加速矩阵补全算法应用于多标记学习,然而他们同样未能解决未见样本的分类问题.

(3) 先验信息融合性不强

现有的矩阵补全理论往往忽略了数据固有的先验信息,这类先验信息通常包括目标矩阵的先验稀疏信息、应用问题的先验特征信息和结构信息等.例如:Cabral 等人在文献[5]中利用矩阵补全的方法解决多标记图像分类问题时,忽略了图像与图像、标记与标记之间的先验相似性关联信息;Mardani 等人在文献[6]利用矩阵补全方法监测网络异常流量时,忽略了网络流量的先验特征和结构信息;Natarajan 等人在文献[87]中将矩阵补全算法应用于基因-疾病预测时,忽略了基因-疾病关联矩阵的先验稀疏信息.

(4) 算法扩放性及计算效率低下

现有矩阵补全算法的可扩放性和计算效率也已成为其大规模应用的瓶颈.针对这个问题,研究者已经进行过一些探索.针对核范数松弛类矩阵补全算法所涉及的奇异值分解问题,普遍采用的方法是调用 PROPACK 软件包进行部分奇异值分解,从而有效降低奇异值分解的时间复杂度,一定程度上加速了算法收敛速度,但仍然无法适用于大规模矩阵补全问题.为此,Teflioudi 等人在文献[88]中基于 MapReduce 并行编程模型设计了一类适用于集群平台的并行交替最小均方算法,Recht 等人在文献[75]中基于随机投影增量梯度方法设计了一类矩阵补全并行算法,Mosabbeb 等人在文献[89]中基于交替方向乘子法设计了一类分布式矩阵补全并行算法并实际应用于大规模多标记分类,而 Makari 等人则在文献[30]中基于随机梯度下降方法设计了一类既能运行在共享存储又能运行在分布式存储平台上的并行算法.然而,所有这些算法一方面局限于解决直传型矩阵补全问题,缺乏归纳推广性能;另一方面,基本都是基于 MapReduce 并行编程模型设计,仅适合运行于 Hadoop 平台,而且由于矩阵补全算法需要大量的迭代计算,使用 MapReduce 编程模型效率显得极为低下,仍然难以满足很多复杂的大数据处理需求.

4.2 未来研究方向

矩阵补全技术研究方兴未艾,在模型、算法及其应用方面仍需要做很多工作.下面分别从这几个方面对其未来研究方向进行探讨.

(1) 模型方面

现有的矩阵补全模型往往局限于数据表示信息的重建,重建的目的是获得最佳描述特征,未能有效利用数据的判别信息.例如:在多标记学习任务中经常遇到基于不完全样本特征信息的多标记分类问题,对于此类问题,我们实际上并不关心补全后的样本数据和真实的样本数据是否在数据表示上完全一致,我们真正关心的是这些补全后的样本数据能否被正确分类,因此,如果在对这些样本特征信息进行预测和补全时能够融入监督或半监督的类别信息,那么将有可能提高后续分类任务的分类性能.因此,研究如何在问题建模时融入问题的判别信息将具有很好的理论与现实意义,极有可能成为矩阵补全技术研究的新热点.此外,矩阵补全是压缩感知从一维向量空间向二维矩阵空间的拓展,而张量补全[93]则是矩阵补全从二维矩阵空间向多维张量空间的拓展.与压缩感知和矩阵补全技术相比,张量补全技术研究起步较晚,张量补全模型多样性的缺乏阻碍了其在实际应用中的推广,因此可以预见,张量补全技术的研究将是一个值得持续关注的热点.

(2) 算法方面

现有的矩阵补全算法大多属于集中式串行处理算法,受制于单机计算效率和内存空间限制,算法的扩放性和计算效率低下,难以应用于大规模问题.对现有算法进行分布式并行扩展,应该是一种可行的解决思路.典型的分布式并行编程模型诸如 MPI,Hadoop 和 Spark 各有其优缺点,其中,源于加州伯克利大学的 Spark 是近年来大数据处理平台的新锐代表.Spark 已经在批处理、流计算、机器学习、图计算、SQL 查询等一系列领域得到广泛应用,并随着愈发活跃的开发者社区以及 Twitter,Adobe,Intel,Amazon,Redhat 等公司的加入而渐成气候.相比于 Hadoop,Spark 是内存计算框架,拥有 Hadoop MapReduce 所具有的优点;但不同于 Hadoop,Job 中间输出和结果可以保存在内存中,从而不再需要读写 HDFS,因此尤其适用于需要多次迭代计算的矩阵补全模型求解.而且从数值优化上看,Spark MLlib 已成功实现了梯度下降、牛顿法、ADMM 等经典优化算法.因此,Spark 有望成为矩阵补全分布式并行扩展的首选编程模型.

(3) 应用方面

现有的矩阵补全技术虽然在诸多领域得到了广泛应用,但是大多基于一些严格的假设,实际上这些假设并不一定符合实际,因此,如何放宽矩阵补全应用中的理想假设使其更加契合问题的本源,将是值得关注的研究方向.例如在多标记图像分类中,往往假设图像标记和图像特征之间满足线性映射关系,而实际情况是更可能满足非线性映射关系.再如,在推荐系统的用户-评分预测中,通常假设用户偏好在某一个时间段内是一成不变的,只在跨时间段间发生变化,这种假设显然也是过于严苛的.针对这两种理想假设的不足,最近已有学者分别提出了非线性矩阵补全[91]和动态矩阵补全模型[92],模型性能有了显著提高.此外,矩阵补全模型的噪声容错性能也是一个影响矩阵补全应用推广的普适性问题,现有的矩阵补全模型在噪声类型先验假设的前提下取得了较好的研究成果,但是不可预知的复杂噪声背景下的矩阵补全技术研究才刚刚起步.其中,西安交通大学徐宗本院士和孟德宇博士团队在低维子空间学习模型噪声容错性领域的研究成果为此提供了一个很好的思路,他们在 ICCV 2015 会议上撰文指出[93]:如果采用混合指数幂分布(mixture of exponential power)函数来拟合不可预知的复合噪声类型,将取得令人鼓舞的效果.其理论依据是,混合指数幂分布函数理论上能逼近任意分布.虽然该方法目前尚存在算法扩放性和初值敏感性问题,但其无疑为机器学习领域众多学习模型的噪声容错性问题提供了一种很好的解决思路.

5 本文总结

作为稀疏学习理论的重要组成部分,衍生于压缩感知的矩阵补全技术近年来在机器学习领域获得了广泛关注,短短几年,从模型、算法到应用方面均得到了快速发展,取得了诸多研究成果.本文首先从秩函数松弛的角度综述了 4 类不同的矩阵补全模型,旨在为相关应用领域的矩阵补全问题建模提供参考;然后从独立于问题模型的角度综述了适用于矩阵补全模型统一框架的的常用优化算法,其目的在于从本质上加深对矩阵补全模型优化技巧的理解,从而有利于面向应用问题的矩阵补全新模型的优化求解;最后指出了现有矩阵补全模型在噪声容错性、归纳推广性和先验信息融合性方面存在的不足;同时,本文也关注到现有的矩阵补全算法大多受制于单机计算效率和内存空间限制,导致算法扩放性和计算效率低下,难以应用于大规模问题求解的问题.可以想见,如果上述这些问题能够得到较好的解决,矩阵补全技术必将迎来更加广阔的应用前景.

References:

[1] Donoho DL. Compressed sensing. IEEE Trans. on Information Theory, 2006,52(4):1289−1306. [doi: 10.1109/TIT.2006.871582]

[2] Candμes EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. on Information Theory, 2006,52(2):489−509. [doi: 10.1109/TIT.2005.862083]

[3] Candes EJ, Recht B. Exact matrix completion via convex optimization. Foundations of Computational Mathematics, 2009,9(6): 717−772. [doi: 10.1007/s10208-009-9045-5]

[4] Li W, Zhao L, Lin ZJ, Xu DQ, Lu DM. Non-Local image inpainting using low-rank matrix completion. Computer Graphics Forum, 2015,34(6):111−122. [doi: 10.1111/cgf.12521]

[5] Cabral R, De la Torre F, Costeira JP, Bernardino A. Matrix completion for weakly-supervised multi-label image classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(1):121−135. [doi: 10.1109/TPAMI.2014.2343234]

[6] Mardani M, Giannakis GB. Estimating traffic and anomaly maps via network tomography. IEEE/ACM Trans. on Networking, 2016, 24(5):1533−1547. [doi: 10.1109/TNET.2015.2417809]

[7] Yi KF, Wan JW, Yao L, Bao TY. Partial matrix completion algorithm for efficient data gathering in wireless sensor networks. IEEE Communications Letters, 2015,19(1):54−57. [doi: 10.1109/LCOMM.2014.2371998]

[8] Otazo R, Candès E, Sodickson DK. Low-Rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. Magnetic Resonance in Medicine, 2015,73(3):1125−1136. [doi: 10.1002/mrm.25240]

[9] Taghizadeh MJ, Parhizkar R, Garner PN, Bourlard H, Asaei A. Ad hoc microphone array calibration: Euclidean distance matrix completion algorithm and theoretical guarantees. Signal Processing, 2015,107:123−140. [doi: 10.1016/j.sigpro.2014.07.016]

[10] Hsieh CJ, Natarajan N, Dhilon IS. PU learning for matrix completion. In: Proc. of the Int'l Conf. on Machine Learning. 2015.

[11] Adeli-Mosabbeb E, Fathy M. Non-Negative matrix completion for action detection. Image and Vision Computing, 2015,39(1): 38−51. [doi: 10.1016/j.imavis.2015.04.006]

[12] Yi J, Zhang L, Jin R, Qian Q, Jain AK. Semi-Supervised clustering by input pattern assisted pairwise similarity matrix completion. In: Proc. of the Int'l Conf. on Machine Learning. 2013.

[13] Gogn A, Majumdar A. Matrix completion incorporating auxiliary information for recommender system design. Expert System with Applications, 2015,42(12):5789−5799. [doi: 10.1016/j.eswa.2015.04.012]

[14] Candes EJ, Eldar YC, Strohmer T, Vladislav H. Phase retrieval via matrix completion. SIAM Review, 2015,57(2):225−251. [doi: 10.1137/151005099]

[15] Kumar R, Silva CD, Akalin O, Aravkin AY, Mansour H; Recht B; Herrmann FJ. Efficient matrix completion for seismic data reconstruction. Geophysics, 2015, 80(5):97−114. [doi: 10.1190/geo2014-0369.1]

[16] Zhao Z, Zhang LJ, He XF, Ng W. Expert finding for question answering via graph regularized matrix completion, IEEE Trans. on Knowledge and Data Engineering, 2015,27(4):993−1004. [doi: 10.1109/TKDE.2014.2356461]

[17] Xiao F, Sha CH, Chen L, Sun LJ, Wang RC. Noise-Tolerant localization from incomplete range measurements for wireless sensor networks. In: Proc. of the IEEE Int'l Conf. on Computer Communications. 2015. 2794−2802. [doi: 10.1109/INFOCOM.2015.7218672]

[18] Chen L, Yang G, Chen ZY, Xiao F, Xu J. Web services QoS prediction via matrix completion with structural noise. Journal on Communications, 2015,36(6):49−59 (in Chinese with English abstract).

[19] Chen L, Yang G, Chen ZY, Xiao F, Shi JY. Correlation consistency constrained matrix completion for Web service tag refinement. Neural Computing and Applications, 2015,26(1):101−110. [doi: 10.1007/s00521-014-1704-z]

[20] Bishop WE, Yu BM. Deterministic symmetric positive semidefinite matrix completion. In: Proc. of the Advances in Neural Information Processing Systems. 2014.

[21] Li XD. Compressed sensing and matrix completion with constant proportion of corruptions. Constructive Approximation, 2013,37: 73−99. [doi: 10.1007/s00365-012-9176-9]

[22] Candes EJ, Plan Y. Matrix completion with noise. Proc. of the IEEE, 2010,98(6):925−936. [doi: 10.1109/JPROC.2009.2035722]

[23] Mardani M, Mateos G, Giannakis GB. Subspace learning and imputation for streaming big data matrices and tensors. IEEE Trans. on Signal Processing, 2015,63(10):2663−2677. [doi: 10.1109/TSP.2015.2417491]

[24] Wang Z, Lai J, Lu Z S, Ye JP. Orthogonal rank-one matrix pursuit for low rank matrix completion SIAM Journal on Scientific Computing, 2015,37(1):488−514. [doi: 10.1137/130934271]

[25] Ma S, Goldfarb D, Chen L. Fixed point and Bregman iterative methods for matrix rank minimization. Mathematics Programming, 2011,128(1):321−353. [doi: 10.1007/s10107-009-0306-5]

[26] Xu YY, Yin WT, Wen ZW, Zhang Y. An alternating direction algorithm for matrix completion with nonnegative factors. Frontiers of Mathematics in China, 2012,7(2):365−384. [doi: 10.1007/s11464-012-0194-5]

[27] Mohan K, Fazel M. Iterative reweighted algorithms for matrix rank minimization. Journal of Machine Learning Research, 2013,13: 3253−3285.

[28] Mackey L, Talwalkar A, Jordan M. Distributed matrix completion and robust factorization. Journal of Machine Learning Research, 2015,16(1):913−960.

[29] Candes EJ, Tao T. The power of convex relaxation: Near-optimal matrix completion. IEEE Trans. on Information Theory, 2009, 56(5):2053−2080. [doi: 10.1109/TIT.2010.2044061]

[30] Makari F, Teflioudi C, Gemulla R, Haas P, Sismanis Y. Shared-Memory and shared-nothing stochastic gradient descent algorithms for matrix completion. Knowledge and Information Systems, 2015,42(3):493−523. [doi: 10.1007/s10115-013-0718-7]

[31] Nie FP, Wang H, Huang H, Ding C. Joint schatten *p*-norm and *Lp*-norm roubst matrix completion for missing value recovery. Knowledge and Information Systems, 2015,42(3):525−544. [doi: 10.1007/s10115-013-0713-z]

[32] Lu CY, Tang JH, Yan SC, Lin Z. Generalized nonconvex non-smooth low-rank minimization. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 4130−4137. [doi: 10.1109/CVPR.2014.526]

[33] Zhang LJ, Yang T, Jin R, Zhou ZH. Stochastic proximal gradient descent for nuclear norm regularization. 2015. http://arxiv.org/abs/1511.01664 CoRR abs/1511.01664

[34] Boumal N, Absil P. RTRMC: A riemannian trust region method for matrix completion. In: Proc. of the Annual Conf. on Neural Information Processing Systems. 2011.

[35] Ghasemi H, Malek-Mohammadi M, Babaie-Zadeh M. SRF: Matrix completion based on smoothed rank function. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. 2011. 3672−3675. [doi: 10.1109/ICASSP.2011.5947147]

[36] Winlaw M, Hynes M, Caterini A, Sterck H. Algorithmic acceleration of parallel ALS for collaborative filtering; speeding up distributed big data recommendation in Spark. In: Proc. of the IEEE Int'l Conf. on Parallel and Distributed Systems. 2015. 682−691. [doi: 10.1109/ICPADS.2015.91]

[37] Sun RY, Luo ZQ. Guaranteed matrix completion via non-convex factorization. In: Proc. of the Annual Symp. on Foundations of Computer Science. 2015. [doi: 10.1109/FOCS.2015.25]

[38] Lin ZC. Rank minimization: Theory, algorithms and applications. In: Zhang CS, Zhou ZH, eds. Proc. of the Machine Learning and Their Applications. Beijing: Tsinghua University Press, 2013. 143−164 (in Chinese with English abstract).

[39] Peng YG, Suo JL, Dai QH, Xu LW. From compressed sensing to low-rank matrix recovery: Theory and aplications. Acta Actomatica Sinica, 2013,39(7):981−994 (in Chinese with English abstract). [doi: 10.1016/S1874-1029(13)60063-4]

[40] Chen CH, He BS, Yuan XM. Matrix completion via an alternating direction method. IMA Journal of Numerical Analysis, 2012, 32(1):227−245. [doi: 10.1093/imanum/drq039]

[41] Wen SW, Xu FF, Wen ZW, Chen L. Robust linear optimization under matrix completion. Science in China: Mathematics, 2014,57: 699−710. [doi: 10.1007/s11425-013-4697-7]

[42] Jiao LC, Zhao J, Yang SY, Liu F, Xie W. Research advances on sparse cognitive learning, computing and recognition. Chinese Journal of Computers, 2016,39(4):835−852 (in Chinese with English abstract).

[43] Hu Y, Zhang DB, Ye JP, Li X, He XF. Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013,35(9):2117−2130. [doi: 10.1109/TPAMI.2012.271]

[44] Chen L, Yang G, Chen ZY, Xiao F, Chen SC. Linearized bregman iteration algorithm for matrix completion with structural noise. Chinese Journal of Computers, 2015,38(7):1357−1371 (in Chinese with English abstract).

[45] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, 1996,58(1):267−288.

[46] Fazel M. Matrix rank minimization with applications [Ph.D. Thesis]. Standford University, 2002.

[47] Toh KC, Yun SW. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. Pacific Journal of Optimization, 2010,6(3):615−640.

[48] Cai JF, Candes EJ, Shen Z. A singular value thresholding algorithm for matrix completion, SIAM Journal of Optimization, 2010, 20(4):1956−1982. [doi: 10.1137/080738970]

[49] Lin ZC, Chen MM, Wu LQ, Ma Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical Report, UILU-ENG-09-2214, UIUC University, 2010.

[50] Wen ZW, Yin WT, Zhang Y. Solving a low rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. Mathematic Programming Computation. 2012,4:333−361. [doi: 10.1007/s12532-012-0044-1]

[51] Liu YY, Jiao LC, Shang FH. A fast tri-factorization method for low-rank matrix recovery and completion. Pattern Recognition, 2013,46(1):163−173. [doi: 10.1016/j.patcog.2012.07.003]

[52] Srebro N, Shraibman A. Rank, trace-norm and max-norm. In: Proc. of the Annual Conf. on Learning Theory. 2005. 545−560. [doi: 10.1007/11503415_37]

[53] Mardani M, Mateos G, Giannakis GB. Rank minimization for subspace tracking from incomplete data. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. 2013. 5681−5685. [doi: 10.1109/ICASSP.2013.6638752]

[54] Geng J, Wang LS, Wang YF. A non-convex algorithm framework based on DC programming and DCA for matrix completion. Numerical Algorithms, 2015,68:903−921. [doi: 10.1007/s11075-014-9876-2]

[55] Kang Z, Peng C, Cheng J, Cheng Q. LogDet rank minimization with application to subspace clustering. Computational Intelligence and Neuroscience, 2015, Article ID 824289. [doi: 10.1155/2015/824289]

[56] Honerkamp J, Weese J. Tikhonovs regularization method for ill-posed problems. Continuum Mechanics and Thermodynamics, 1990,2(1):17−30. [doi: 10.1007/BF01170953]

[57] Chen Z, Haykin S. On different facets of regularization theory. Neural Computation, 2002,14(12):2791−2846. [doi: 10.1162/089976602760805296]

[58] Combettes PL, Wajs VR. Signal recovery by proximal forward-backward splitting. Multiscale Modeling and Simulation, 2005,4(4): 1168−1200. [doi: 10.1137/050626090]

[59] Nemirovski A. Efficient methods in convex programming. Lecture Notes, 1995,23(3):24−39.

[60] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2009,2(1):183−202. [doi: 10.1137/080716542]

[61] Nesterov Y. A method of solving a convex programming problem with convergence rate $O(k^{-2})$. Soviet Mathematics Doklady, 1983, 27(2):372−376.

[62] Nesterov Y. Efficiency of coordinate descent methods on huge-scale optimization problems. SIAM Journal on Optimization, 2012, 22:341−362. [doi: 10.1137/100802001]

[63] Beck A, Tetruashvili L. On the convergence of block coordinate descent methods. SIAM Journal on Optimization, 2013,23(4): 2037−2060. [doi: 10.1137/120887679]

[64] Bregman L. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. Computational Mathematics and Mathematical Physics, 1967,7(3):200−217. [doi: 10.1016/0041-5553(67)90040-7]

[65] Osher S, Burger M, Goldfarb M. An iterative regularization method for total variation-based image restoration. Multiscale Modeling and Simulation, 2005,4(2):460−489. [doi: 10.1137/040605412]

[66] Yin WT, Osher S, Goldfarb D. Bregman iterative algorithms for L_1-minimization with applications to compressed sensing. SIAM Journal on Imaging Sciences, 2008,1(1):143−168. [doi: 10.1137/070703983]

[67] Cai JF, Osher S, Shen Z. Linearized Bregman iteration for frame-based image deblurring. SIAM Journal on Imaging Sciences, 2009, 2(2):226−252. [doi: 10.1137/080733371]

[68] Boyd S, Parikh N, Chu E. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning, 2011,3(1):1−122. [doi: 10.1561/2200000016]

[69] Douglas J, Rachford HH. On the numerical solution of heat conduction problems in two and three space variables. Trans. of the American Mathematical Society, 1956,82:421−439. [doi: 10.1090/S0002-9947-1956-0084194-4]

[70] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Computers and Mathematics with Applications, 1976,2:17−40. [doi: 10.1016/0898-1221(76)90003-1]

[71] Zhang RL, Kwok J. Asynchronous distributed ADMM for consensus optimization. In: Proc. of the Int'l Conf. on Machine Learning. 2014.

[72] Zhang WL, Kwok J. Fast stochastic alternating direction method of multipliers. In: Proc. of the Int'l Conf. on Machine Learning. 2014.

[73] He BS, Yuan XM. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM Journal of Numerical Analysis. 2012,50(2):700−709. [doi: 10.1137/110836936]

[74] He BS, Tao M, Yuan XM. Alternating direction method with Gaussian back substitution for separable convex programming. SIAM Journal on Optimization, 2012,22(2):313−340. [doi: 10.1137/110822347]

[75] Lin ZC, Liu RS, Li H. Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. Machine Learning, 2015,99(2):287−325. [doi: 10.1007/s10994-014-5469-5]

[76] Nemirovski A, Juditsky A, Lan G, Shapiro A. Robust stochastic approximation approach to stochastic programming. SIAM Journal on Optimization, 2009,19(4):1574−1609. [doi: 10.1137/070704277]

[77] Recht B, Re C. Parallel stochastic gradient algorithms for large-scale matrix completion. Mathematical Programming Computation, 2013,5(2):201−226. [doi: 10.1007/s12532-013-0053-8]

[78] Liu J, Wright SJ, Re C, Bittorf V, Sridhar S. An asynchronous parallel stochastic coordinate descent algorithm. Journal of Machine Learning Research, 2015,16(1):285−322.

[79] Azadi S, Sra S. Towards an optimal stochastic alternating direction method of multipliers. In: Proc. of the Int'l Conf. on Machine Learning. 2014.

[80] Zhang LJ, Yang T, Yi J, Jin R, Zhou ZH. Stochastic optimization for kernel PCA. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2016. 2316 –2322.

[81] Zhang ZH. The singular value decomposition, applications and beyond. arXiv: 1510.08532v1, http://arxiv.org/abs/1510.08532

[82] Tao M, He BS, Yuan XM. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM Journal on Optimization, 2011,21(1):57–81. [doi: 10.1137/100781894]

[83] Yan M, Yang Y, Osher S. Exact low-rank matrix completion from sparsely corrupted entries via adaptive outlier pursuit. Journal of Scientific Computing, 2013,56(3):433–449. [doi: 10.1007/s10915-013-9682-3]

[84] Chen YD, Jalali A, Sanghavi S, Caramanis C. Low-Rank matrix recovery from errors and erasures. IEEE Trans. on Information Theory, 2013,59(7):4324–4337. [doi: 10.1109/TIT.2013.2249572]

[85] Chen YD, Xu H, Caramanis C, Sanghavi S. Robust matrix completion and corrupted columns. In: Proc. of the Int'l Conf. on Machine Learning. 2011.

[86] Xu M, Jin R, Zhou ZH. Speedup matrix completion with side information: Application to multi-label learning. In: Proc. of the Advances in Neural Information Processing Systems. 2013.

[87] Natarajan N, Dhillon IS. Matrix completion for predicting gene-disease associations. Bioinformatics, 2014,30(12):60–68. [doi: 10.1093/bioinformatics/btu269]

[88] Teflioudi C, Makari F, Gemulla R. Distributed matrix completion. In: Proc. of the IEEE Int'l Conf. on Data Mining. 2012. 655–664. [doi: 10.1109/ICDM.2012.120]

[89] Mosabbeb EA, Fathy M. Distributed matrix completion for large-scale multi-label classification. Intelligent Data Analysis, 2014,18: 1137–1151.

[90] Liu J, Musialski P, Wonka P, Ye JP. Tensor completion for estimating missing values in visual data. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013,35(1):208–220. [doi: 10.1109/TPAMI.2012.39]

[91] Alameda-Pineda X, Yan Y, Ricci E, Sebe N. Recognizing emotions from abstract paintings using non-linear matrix completion. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 5240–5248. [doi: 10.1109/CVPR.2016.566]

[92] Xu LB, Davenport MA. Dynamic matrix recovery from incomplete observations under an exact low-rank constraint. In: Proc. of the Advances in Neural Information Processing Systems. 2016.

[93] Cao XR, Chen Y, Zhao Q, Meng DY, Wang Y, Wang D, Xu ZB. Low-Rank matrix factorization under general mixture noise distributions. In: Proc. of the Int'l Conf. on Computer Vision. 2015. [doi: 10.1109/ICCV.2015.175]

附中文参考文献:

[18] 陈蕾,杨庚,陈正宇,肖甫,许建.基于结构化噪声矩阵补全的 Web 服务 QoS 预测.通信学报,2015,36(6):49–59.

[38] 林宙辰.秩极小化:理论、算法与应用.见:张长水,周志华,编.机器学习及其应用会议论文集.北京:清华大学出版社,2013.143–164.

[39] 彭义刚,索津莉,戴琼海,徐文立.从压缩传感到低秩矩阵恢复:理论与应用.自动化学报,2013,39(7):981–994. [doi: 10.1016/S1874-1029(13)60063-4]

[42] 焦李成,赵进,杨淑媛,刘芳,谢雯.稀疏认知学习、计算与识别的研究进展.计算机学报,2016,39(4):835–852.

[44] 陈蕾,杨庚,陈正宇,肖甫,陈松灿.基于线性 Bregman 迭代的结构化噪声矩阵补全算法.计算机学报,2015,38(7):1357–1371.

陈蕾(1975－),男,江西宜春人,博士,副教授,CCF 专业会员,主要研究领域为大规模机器学习,计算机视觉,基于医学影像的脑疾病诊断.

陈松灿(1962－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为模式识别,机器学习.