

基于加权核范数最小化的矩阵填充模型

张玮奇¹ 张宏志¹ 左旺孟¹ 崔梦天^{2,3}

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)¹

(西南民族大学计算机科学与技术学院 成都 610041)²

(电子科技大学计算机科学与工程学院 成都 610054)³

摘要 协同过滤是目前推荐系统最常用的技术之一,相比于传统的推荐技术具有一定优势,但其缺点是受用户对商品评价的稀疏性制约,现阶段一般利用矩阵填充技术来解决这一问题。主要研究了基于低秩的矩阵填充模型,针对原有模型解对所有奇异值用同一值收缩的问题,提出了一种加权核范数最小化模型以提高核范数灵活度,给出了该模型用收缩算子可得到全局最优解的相关定理及证明,同时对模型的另一种形式在求解过程中的迭代收敛性进行了证明。用凸优化主流算法在两种真实数据集上进行的实验表明,改进后的模型一定程度上提高了计算速度与准确性。

关键词 协同过滤,矩阵填充,低秩,凸优化算法

中图法分类号 TP399

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2015.7.054

Weighted Nuclear Norm Minimization Model for Matrix Completion

ZHANG Wei-qi¹ ZHANG Hong-zhi¹ ZUO Wang-meng¹ CUI Meng-tian^{2,3}

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)¹

(School of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610041, China)²

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)³

Abstract Collaborative filtering is one of the popular techniques used in recommendation system. It has some advantages over traditional recommendation technologies. But the limitation is that it is constrained by the data sparsity. Matrix completion technology can be used to solve this problem. This paper proposed an weighted nuclear norm minimization (WNNM) model for matrix completion to improve the flexibility of nuclear norm. Under certain condition, it can be proved to get global optimal solution. Meanwhile, convergence for another form of the proposed model was confirmed. With two real data sets, convex optimization algorithm of nuclear norm minimization was achieved to verify the proposed model. The result proves that to some extent it improves the computational speed and accuracy.

Keywords Collaborative filtering, Matrix completion, Low rank, Convex optimization algorithm

1 绪论

Internet 的快速发展将人类带入了信息社会和网络经济时代,在如今电子商务的虚拟环境下,商家所提供的商品种类和数量非常多,消费者很难快速有效地挑选出需要的商品。在这种背景下,推荐系统(Recommender Systems)应运而生,它根据用户的特征,推荐满足用户要求的对象。协同过滤方法^[1]是目前推荐系统中比较常用的技术,它通过分析用户兴趣,在用户群中找到与指定用户相似(有共同兴趣)的用户,综合这些相似用户对某一信息的评价,形成系统对该指定用户关于此信息的喜好程度预测。协同过滤方法通常分为基于用户的和基于项目的两种,与传统文本过滤相比,协同过滤有下列优点:能够过滤难以进行机器自动基于内容分析的信息,如

艺术品、音乐;能够基于一些复杂的、难以表达的概念(比如品位)进行过滤;推荐具有新颖性。但该方法也有一定的局限性:

- 1)用户对商品的评价非常稀疏,基于用户评价得到的用户间的相似性可能不准确(即稀疏性问题);
- 2)随着用户和商品的增多,系统的性能会越来越低;
- 3)如果从来没有用户对某一商品加以评价,则这个商品就不可能被推荐(冷启动问题)。

为了解决上述用户评分的稀疏问题,我们尝试用一些方法猜测用户对商品的评价。用户-商品的关系一般用矩阵表示,矩阵内的元素即某用户对某商品的评分,这样一来,预测评分其实就是一种矩阵填充^[2](Matrix Completion)。矩阵填充考虑:对于某个矩阵,只能采样得到一部分元素,其它一部

到稿日期:2014-07-01 返修日期:2014-08-02 本文受国家自然科学基金面上项目(61271093,61379019),四川省学术和技术带头人培养资金资助。

张玮奇(1992—),女,硕士生,主要研究方向为机器学习、计算机视觉, E-mail: weiqizhanghit@gmail.com; 张宏志(1976—),男,博士,副教授,主要研究方向为计算机视觉、机器学习; 左旺孟(1977—),男,博士,副教授,主要研究方向为计算机视觉、机器学习; 崔梦天(1972—),女,博士后,教授,主要研究方向为可信软件、优化理论和算法, E-mail: happyzg3@163.com(通信作者)。

分或大部分元素由于各种原因丢失了或无法得到,如何将这
些空缺的元素合理准确地填充。比如我们熟知的 Netflix 问
题,就是典型的矩阵填充问题。

矩阵填充问题其实是压缩传感的推广,但它利用的是矩
阵奇异值的稀疏性,即矩阵的低秩性。根据不同的情境,低秩
矩阵填充问题也可以被称为低秩矩阵分解、鲁棒主成分分析
(RPCA)、低秩稀疏非相干分解(Rank-sparsity incoherence)
等。

机器学习是人工智能的重要分支之一,是一种用于创建
数据集分析程序的方法,已经广泛用于生物特征识别、医学诊
断、搜索引擎、语音和手写识别及计算机视觉等方面,其现有
的一些关键技术包括人工神经网络、决策树、线性判别分析、
支持向量机等也发挥了重要的作用。而近两年,基于凸优化
的矩阵分析技术作为一种有效的数据分析工具受到了更多的
关注。

本文将通过研究矩阵填充问题与凸优化方法,比较现阶
段较为有效的几种矩阵填充算法,得出求解核范数最小化的
一般性方法;并针对原核范数最小化模型的不足之处做进一
步的改进,使得在矩阵中只有很少部分元素明确的情况下,更
快速、精确地填充剩余的部分,以期在协同过滤技术的应用中
能得到更好的推荐效果。

本文第2节介绍传统的矩阵填充方法;第3节给出一般
性矩阵填充模型、矩阵填充需满足的条件及解决此类问题的
主流凸优化算法;第4节介绍针对矩阵填充的加权核范数最
小化模型,给出了该模型能得到全局最优解的相关理论,证明
了模型的另一种形式在求解过程中的迭代收敛性;第5节给
出实验结果与分析;最后总结本文工作。

2 传统矩阵填充方法

最简单的矩阵填充办法就是将用户对未评分项目的评分
设为一个固定的缺省值,或者设为其他用户对该项目的平均
评分。然而用户对未评分项目的评分不可能都相同,这种办
法不能从根本上解决稀疏性问题。能够产生理想推荐效果的
矩阵填充技术主要有以下几类。

1) BP 神经网络^[3]

BP 神经网络对复杂的输入输出关系有比较强大的学习
和建模能力,能够有效地处理非完整信息。BP 神经网络是一
个3层网络,分别为输入层、隐含层和输出层。BP 神经网络
把用户对各个项目的评分看作训练样本,分别输入到输入层
的各个单元中;这些单元经过加权,输出到隐含层的各个单
元;隐含层的加权输出再经过一次加权作为输出层的单元输
入;最后由输出层产生给定样本的预测值。这种矩阵填充技
术对噪声数据有较强的承受能力,可以有效降低用户-项目矩
阵的稀疏性,达到提高推荐精度的目的。然而,BP 算法的缺
点是随着训练时间的增加,收敛速度有变慢的趋势,以致会延
长最近邻的查找时间。

2) Naive Bayes 分类方法

Naive Bayes 分类方法^[4]基于概率模型进行分类,可以使
用该方法估算一个实例属于某一类的概率,在得到某一个项
目所属的分类之后,可以利用此分类中其他项目的评分情况
来预测未评分项目的评分,从而可以填充用户-项目矩阵,降
低稀疏性。

3) 基于内容的预测

基于内容的预测又称基于属性的预测或基于语义的预
测,该方法根据项目的属性联系以及项目所处的地位、相互关
系和项目元信息等内容计算项目之间的内容相似性,而不依
赖于用户对项目的评分。得到项目之间的内容相似性后,选
择与目标项目相似性最大的若干个项目进行评分预测,用预
测评分填充用户-项目矩阵中的空项,降低其稀疏性。由于不
同类别的项目在属性描述上有较大差别,因此基于语义的方
法无法计算跨类别的项目之间的相似性,也就无法进行跨类
别的评分预测。另外,基于语义的相似性计算需要提取项目
的属性特征,涉及到领域知识,应用面较窄。

3 矩阵填充问题研究

本节给出了矩阵填充的一般模型及实现矩阵填充的必须
条件,同时介绍了几种求解矩阵填充模型的主流算法。

3.1 矩阵填充模型及条件

为解决矩阵填充问题,通常假设矩阵是有信息冗余的,比
如低秩,于是可以通过如下优化问题来实现:

$$\begin{aligned} \min \text{rank}(X) \\ \text{s. t. } X_{ij} = M_{ij}, (i, j) \in \Omega \end{aligned} \quad (1)$$

其中, Ω 是已知元素下标的集合, X 是填充所得矩阵,这个模
型的意义是将空缺的元素填充之后使得矩阵的秩尽可能低。
其已经被证明是一个 NP-Hard 的问题,在理论和实践中,均
只存在指数复杂度(相对于矩阵维数 n)的算法。所以现阶段
方法大多用矩阵的奇异值的和,即核范数,来近似地替代矩阵
的秩,将问题转化为如下形式:

$$\begin{aligned} \min \|X\|_* \\ \text{s. t. } X_{ij} = M_{ij}, (i, j) \in \Omega \end{aligned} \quad (2)$$

对于满足一定条件的矩阵,通过求解问题(2)可以实现空
缺值的填充。文献[5]给出了实现矩阵填充需要满足的条件,
首先,如果一个秩为 r 的矩阵的奇异值分解为:

$$M = \sum_{k=1}^r \sigma_k u_k v_k^* \quad (3)$$

第一,假定 k 个左奇异向量和 k 个右奇异向量都是均匀
分布在所有由 k 个正交归一化向量的集合中,对于这样的矩
阵,有很大的概率可以正确填充;

第二,一般都假定矩阵的采样方式也是均匀采样;

第三,采样的矩阵元素数目必须大于一定范围。在上述
两个假定条件下,设 n 为矩阵维数, r 为矩阵的秩,则存在常
数 C 和 c ,使得采样数目 m 满足

$$m \geq C n^{5/4} r \log n \quad (4)$$

时,式(2)的最优解正好为 M 的概率满足:

$$p \geq 1 - c n^{-3} \quad (5)$$

3.2 奇异值收缩算子

一个秩为 r 的矩阵 X ,其奇异值分解为:

$$X = U \Sigma V^T, \Sigma = \text{diag}(\{\sigma_i\}, 1 \leq i \leq r) \quad (6)$$

对于 $\forall \tau > 0$,软阈值算子可以表示为:

$$S_\tau(\Sigma) = \text{diag}(\{\sigma_i - \tau\}_+) \quad (7)$$

其中

$$\{\sigma_i - \tau\}_+ = \max(\sigma_i - \tau, 0) \quad (8)$$

文献[9]中定义了奇异值收缩算子 D_τ :

$$D_\tau(X) = U S_\tau(\Sigma) V^T$$

从上式可以看出, D_τ 是对 X 进行了软阈值收缩。奇异

值收缩算子和核范数的求解有着很重要的关系,下面的章节中将会对其说明。

3.3 矩阵填充算法

3.3.1 APG 算法

APG (Accelerated Proximal Gradient) 算法是一种利用 Nesterov 方法的一阶算法,收敛速度相比于 SVT、FPC 更有竞争力。文献[6]中将矩阵填充问题转化为一个与原问题近似的无约束优化问题:

$$\min F(X) = \frac{1}{2} \|P_{\Omega}(X-M)\|_F^2 + \mu \|X\|, \quad (10)$$

APG 算法每次最小化上述函数在某个点 Y 处的一个二阶近似:

$$Q(X, Y) = \frac{1}{2} \|P_{\Omega}(X-M)\|_F^2 + \langle P_{\Omega}(Y), X-Y \rangle + \frac{L_f}{2} \|X-Y\|_F^2 + \mu \|X\|. \quad (11)$$

其中, L_f 为 ∇f 的 Lipschitz 常数,并且:

$$\arg \min_X Q(X, Y) = \arg \min_X \frac{L_f}{2} \|X-Y + \frac{1}{L_f} P_{\Omega}(Y)\|_F^2 + \mu \|X\|. \quad (12)$$

于是得到更新规则:

$$\text{Step1 } Y_k = X_k + \frac{t_{k-1}-1}{t_k} (X_k - X_{k-1})$$

$$\text{Step2 } X_{k+1} = D_{m/L_f} (Y_k - \frac{1}{L_f} P_{\Omega}(Y_k))$$

$$\text{Step3 } t_{k+1} = \frac{1 + \sqrt{t_k^2 + 1}}{2}$$

为了加速 APG 方法,文献[6]使用了线性搜索和截断等技术,文中后续部分将改进后的算法称为 APGL 算法。APGL 算法中 L_f 被变量 τ_k 代替,每次迭代过程中减小 τ_k 并用线性搜索的方法确定其值。同时,将对角阵中的奇异值分为两类,当两类比值大于一定阈值时舍去后一类,主要是考虑到奇异值较大的一类能更好地估计出低秩优化结果矩阵的秩,而另外一类则更多源于噪声。

3.3.2 ALM 算法

$$\min_A \|A\|, \quad (13)$$

$$\text{s.t. } A+E=D, \pi_{\Omega}(E)=0$$

A 为所求矩阵, D 为已知矩阵, π_{Ω} 算子保持 Ω 中的元素不变,而 Ω 外的元素为 0。 E 是 D 中未知元素的补偿,在这里初始设为 0,矩阵恢复问题的增广拉格朗日函数可以表示为:

$$L(A, E, Y, \mu) = \|A\|_* + \lambda \|E\|_1 + \langle Y, D-A-E \rangle + \frac{\mu}{2} \|D-A-E\|_F^2 \quad (14)$$

每一步最小化时可以利用交替更新的方式,即先固定 E 和 Y ,求一个使 L 最小化的 A ,然后固定 A 和 Y ,求一个使 L 最小的 E ,从而迭代数次就可以收敛到这个子问题的最优解。更新 A, E :

$$\arg \min_A \|A\|_* + \frac{\mu}{2} \|D-A-E+\mu^{-1}Y\|_F^2 = D_{\mu^{-1}}(D-A-E+\mu^{-1}Y) \quad (15)$$

$$\arg \min_E \frac{\mu}{2} \|D-A-E+\mu^{-1}Y\|_F^2 = P_{\Omega}(D-A-E+\mu^{-1}Y) \quad (16)$$

交替更新直到对子问题的求解收敛,称此算法为矩阵恢

复的精确增广拉格朗日乘子法 (EALM)^[7]。实际上,只需要更新 A 与 E 各一次,得到子问题的一个近似解,已足以使算法最终收敛到原问题的最优解,于是得到一个更简洁且收敛更快的算法,称之为非精确增广拉格朗日乘子法 (IALM)^[7],也称为交替方向法 (ADM)^[8,9]。更新规则为:

$$\text{Step1 } (U, S, V) = \text{svd}(D-E_k + \mu_k^{-1}Y_k)$$

$$\text{Step2 } A_{k+1} = US_{\mu_k}^{-1}V^T$$

$$\text{Step3 } E_{k+1} = \pi_{\Omega}(D-E_k + \mu_k^{-1}Y_k)$$

$$\text{Step4 } Y_{k+1} = Y_k + \mu_k(D-A_{k+1}-E_{k+1})$$

为了解决子问题没有闭合解的问题,文献[10,11]还提出了线性化的 ALM 和 ADM 方法。

4 一种加权的核范数最小化模型

在本节中,提出了一种加权的核范数最小化模型。首先,4.1 节给出了该模型的具体形式,然后在 4.2 节中给出了该模型有可行解的证明。

4.1 WNNM 模型

基于核范数最小化的矩阵填充问题可以很容易地用软阈值的方法解决,具体证明参见文献[12],模型如式(17):

$$\hat{X} = \arg \min_X \|Y-X\|_F^2 + \lambda \|X\|, \quad (17)$$

Y 是已知矩阵, X 是要得到的近似 Y 的低秩矩阵, λ 是正常数,问题(17)的解 $\hat{X}$ 可以由下式求得:

$$\hat{X} = US_{\lambda}(\Sigma)V^T \quad (18)$$

其中,

$$S_{\lambda}(\Sigma)_{ii} = \max(\Sigma_{ii} - \lambda, 0) \quad (19)$$

为了保证凸性,软阈值算子对所有奇异值都用 λ 进行收缩,这样就忽略了先验知识。通常较大的奇异值会起一定的作用,对不同重要性的奇异值做同样的收缩必然不合理。最好用小值收缩大的奇异值,而用大值收缩小的奇异值,以保护数据中的主要部分,忽略不重要的或是产生噪声的部分。所以为了提高核范数的灵活度,本文提出了用于矩阵填充的加权核范数模型^[13]:

$$\min_X \|Y-X\|_F^2 + \lambda \|X\|_{w,*}, \quad (20)$$

加权核范数作为 X 的正则化项,定义为:

$$\|X\|_{w,*} = \sum_i |w_i \sigma_i(X)|_1 \quad (21)$$

其中, $w = [w_1, w_n, \dots, w_N]$ 。WNNM 问题是核范数最小化问题的扩展,但由于其非凸性,次梯度的条件不再满足,不能用传统的方法解决,于是求解变得困难。

4.2 最优解相关证明

由于 σ_i 是 Y 的奇异值,且在 Σ 中是降序排列的,为了保证越大的奇异值用越小值收缩, w 需要满足 $w_n \geq \dots \geq w_2 \geq w_1 \geq 0$ 。

文献[13]中提出,当 w 满足 $w_n \geq \dots \geq w_2 \geq w_1 \geq 0$ 时,即 w 为升序排列时,可得到全局最优解 $\hat{X}$:

$$\hat{X} = US_w(\Sigma)V^T \quad (22)$$

其中

$$S_w(\Sigma)_{ii} = \max(\Sigma_{ii} - w_i, 0) \quad (23)$$

引理 1 $\forall A \in R^{n \times n}$ 和非负对角矩阵 $W \in R^{n \times n}$,其中对角元素是升序排列的,令 $A = XQY^T$ 是 A 的 SVD 分解,则有

$$\sum_i \sigma_i(A) \sigma_i(W) = \max_{U^T U = I, V^T V = I} \text{tr}[WU^T A V] \quad (14)$$

其中, I 是单位阵, $\sigma_i(A)$ 、 $\sigma_i(W)$ 是 A 和 W 的奇异值。当 $U=X$ 、 $V=Y$ 时, $\text{tr}[WU^TAV]$ 达到最大值。

定理 1 问题(20)的解 X^* 可以表示为 $X^*=UDV^T$, 则 D 中对角元是如下优化问题的解:

$$\begin{aligned} \min_{d_1, d_2, \dots, d_n} \sum_{i=1}^n (d_i - \sigma_i)^2 + w_i d_i \\ \text{s. t. } d_1 \geq d_2 \geq \dots \geq d_n \geq 0 \end{aligned} \quad (25)$$

定理 2 当 $0 \leq w_1 \leq w_2 \leq \dots \leq w_n$ 时, 问题(20)的全局最优解为 $D = (\text{diag}(\bar{d}_1, \bar{d}_2, \dots, \bar{d}_n))$, 其中 $\bar{d}_i = \max(\sigma_i - \frac{w_i}{2}, 0)$ 。

具体证明参见文献[13, 14]。综上所述, 当 w 满足条件时, WNNM 模型是凸的, 用软阈值收缩的方法依旧可以得到问题的最优解。

求解 WNNM 模型的算法描述如下。

算法 1 APGL 算法求解 WNNM 模型

输入: 矩阵 Y

输出: 填充后矩阵 X

S1: 初始化 $\mu_0, X_0 = X_{-1}, t_0 = t_{-1} = 1, \sigma_0 = L_f$

S2: //更新 Y

$$Y_k = X_k = \frac{t_{k-1} - 1}{t_k} (X_k - X_{k-1})$$

S3: 用 line search 技术确定 τ_k

S4: $\lambda^i = \mu_k / \tau_k$

S5: //计算权值 w

$$(U, \Sigma, V) = \text{svd}(Y_k - \frac{1}{t_k} P_{\Omega}(Y_k)_k)$$

$$w_i = \lambda \Sigma_i^{-1}$$

S6: //更新 X

$$X_{k+1} = US_{\lambda w}(\Sigma) V^T$$

$$S7: t_{k+1} = \frac{1 + \sqrt{t_k^2 + 1}}{2}$$

S8: 更新 μ_k

4.3 收敛性证明

$$f(X) = \min_X \frac{1}{2} \|AX - b\|^2 + \lambda \|X\|_{w,*} \quad (26)$$

此模型是模型(20)的另外一种形式, A 是从 $R^{m \times n}$ 到 R^p 的线性映射, 在矩阵填充问题中 $AX = X_{\Omega}$, X_{Ω} 为 X 在 Ω 中的元素组成的向量。

定理 3 模型(26)可通过以下方法迭代求解:

$$Y^k = X^k - A^T(AX^k - b) \quad (27)$$

$$X^{k+1} = \arg \min_X \|X - Y^k\|_F^2 + \lambda \|X\|_{w,*} \quad (28)$$

证明: 定义辅助函数

$$Q(X, Z) = \frac{1}{2} \|AZ - b\|^2 + \frac{1}{2} \langle (I - A^T A)(Z - X),$$

$$Z - X \rangle + \lambda \|Z\|_{w,*} \quad (29)$$

假设 $\|A\|_2 \leq 1$, 则 $I - A^T A$ 为半正定矩阵。所以有

$$Q(X, Z) \geq Q(X, X) = f(X) \quad (30)$$

对于给定的 X^k , 将其代入式(29), 消去常数项得

$$\begin{aligned} Q(X^k, Z) = \frac{1}{2} \|Z - (I - A^T A)X^k - A^T b\|_F^2 + \\ \lambda \|Z\|_{w,*} + C \end{aligned} \quad (31)$$

也可以写为

$$\begin{aligned} Q(X^k, Z) = \frac{1}{2} \|Z - [X^k - A^T(AX^k - b)]\|_F^2 + \\ \lambda \|Z\|_{w,*} + C \end{aligned} \quad (32)$$

C 是与 X 有关的常数。式(32)中, $X^k - A^T(AX^k - b)$ 其实就是迭代过程 Y^k 的更新, 其实质就是梯度下降。

根据上节所述定理, 可知式(31)存在最优解 Z^* :

$$Z^* = \arg \min_X Q(X^k, Z)$$

令 $X^{k+1} = Z^*$, 那么有

$$Q(X^k, X_{k+1}) \leq Q(X^k, X^k) = f(X^k)$$

对于给定的 $Z = X^{k+1}$, 将其代入式(29)得

$$Q(X, X^{k+1}) = \frac{1}{2} (X^{k+1} - X)^T (I - A^T A) (X^{k+1} - X) \quad (33)$$

令 X^* 为式(33)的最优解。考虑该式的非负性, 则

$$X^* = \arg \min_X Q(X, X^{k+1}) = X^{k+1}$$

那么有

$$f(X^{k+1}) = Q(X^{k+1}, X^{k+1}) \leq Q(X^k, X^{k+1})$$

可得迭代过程满足

$$f(X^{k+1}) \leq f(X^k)$$

证毕。

5 实验

为了验证本文算法的效率和正确性, 进行相关实验。实验测试了两种主流算法在随机数据及真实数据集上的性能表现, 以验证算法效率。同时, 测试了新模型在相同数据上的性能。所有的实验都采用 Matlab(2013a)编程, 实验环境为 PC 机, 3.3GHz 的 8 核处理器, 16G 内存, Windows 7 操作系统。实验采用了两种真实数据集: Movielens-100k 数据集和 1M 数据集, 包含 1000 名用户对 1700 部电影的 100000 个评分, 评分区间为 1~5 分。

Jester 数据集包含 2 万余名用户对 100 个笑话的评分, 分数分布在 -10 到 +10 区间。Jester-1、Jester-2 数据集中每名用户评价了 36 个以上的项目, jester-3 数据集中每名用户评价了 15~36 个项目, 评分相对稀疏。

算法参数设置上, 将 w 定义为 σ_i 的函数, 满足 w 升序性质, 令:

$$w_i = \lambda \sigma_i \quad (34)$$

采用 APGL 算法在 Jester 和 Movielens 上进行的实验结果分别如表 1、表 2 所列, m/n 表示矩阵行列值, sv 为奇异值个数, $iter$ 表示迭代次数, 精度衡量采用平均绝对偏差(MAE)和均方根误差(RMSE):

$$\delta_{MAX} = \frac{\sum_{(i,j) \in \Omega} |X_{ij} - M_{ij}|}{N} \quad (35)$$

$$\delta_{RMSE} = \sqrt{\frac{\sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2}{N}} \quad (36)$$

表 1 Jester 数据集结果

模型	数据集	m/n	sv	iter	Time(s)	MAE
原模型	Jester-1	24983/100	90	25	17.2	3.4
	Jester-2	23500/100	90	25	16.5	3.415
	Jester-3	24938/100	90	25	12.6	2.546
WNNM	Jester-1	24983/100	88	26	16.8	3.303
	Jester-2	23500/100	88	26	16.6	3.302
	Jester-3	24938/100	83	24	13.5	2.430

表 2 Movielens 数据集结果

模型	数据集	m/n	sv	iter	Time(s)	MAE	RMSE
原模型	100k	943/1682	100	34	12.2	0.51	0.91
	1M	6040/3706	100	50	73.8	5.67	9.07
WNNM	100k	943/1682	100	34	10.2	0.50	0.88
	1M	6040/3706	100	52	73.6	5.68	9.36

(下转第 290 页)

- [18] Liang J Y, Wang F, Dang C Y, et al. A group incremental approach to feature selection applying rough set technique [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(2):294-308
- [19] Kryszkiewicz M. Rough set approach to incomplete information systems [J]. Information Sciences, 1998, 112:39-49
- [20] Slowinski R, Vanderpooten D. A generalized definition of rough approximations based on similarity [J]. IEEE Transactions on Data and Knowledge Engineering, 2000, 12(2):331-336
- [21] Liang J Y, Xu Z B. The algorithm on knowledge reduction in incomplete information systems [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(1):95-103
- [22] 黄兵, 周献中, 张蓉蓉. 基于信息量的不完备信息系统属性约简 [J]. 系统工程理论与实践, 2005, 25(4):55-60
Huang Bing, Zhou Xian-zhong, Zhang Rong-rong. Attribute reduction based on information quantity under incomplete information systems [J]. Chinese Journal of System Engineering Theory and Practice, 2005, 25(4):55-60
- [23] 钱宇华, 梁吉业, 王锋. 面向非完备决策表的正向近似特征选择加速算法 [J]. 计算机学报, 2011, 34(3):435-442
Qian Yu-hua, Liang Ji-ye, Wang Feng. A positive approximation based accelerated algorithm to feature selection from incomplete decision tables [J]. Chinese Journal of Computers, 2011, 34(3):435-442
- [24] Liang J Y, Shi Z Z, Li D Y, et al. The information entropy, rough entropy and knowledge granulation in incomplete information systems [J]. International Journal of General Systems, 2006, 34(1):641-654
- [25] Li T R, Ruan D, Geert W, et al. A rough sets based characteristic relation approach for dynamic attribute generalization in data mining [J]. Knowledge-Based Systems, 2007, 20(5):485-494
- [26] Pawlak Z, Skowron A. Rudiments of rough sets [J]. Information Sciences, 2007, 177(1):3-27

(上接第 257 页)

在两个数据集上的实验结果表明, WNNM 模型应用在矩阵填充上得到了较好的效果, 在速度和精度上有所提高, 但是在矩阵规模大的情况下可能优势不明显。

结束语 矩阵填充是目前重要的数据分析工具, 在图像处理、文本分析等方面都有重要应用, 是近些年的研究热点。本文总结了矩阵填充问题的研究现状, 针对原有矩阵填充模型, 提出了一种加权核范数最小化模型, 提高了核范数的灵活度。在真实数据集上进行的实验表明, 该模型可以使矩阵填充在速度和精度上得到一定程度的提升。

同矩阵填充问题等价的低秩矩阵分解、鲁棒性主成分分析问题等将应用于越来越多的方向, 所以解决这类问题的算法仍需要不断优化, 更要在并行和分布式上做考虑。未来的工作将进一步完善大规模数据上的矩阵填充算法, 使其在动态变化的矩阵上也能得到较好的应用。

参 考 文 献

- [1] 马宏伟, 张光卫, 李鹏. 协同过滤推荐算法综述 [J]. 小型微型计算机系统, 2009, 30(7):1282-1288
Ma Hong-wei, Zhang Guang-wei, Li Peng. Survey of Collaborative Filtering Algorithms [J]. Journal of Chinese Computer Systems, 2009, 30(7):1282-1288
- [2] 陈敏铭. 矩阵恢复与重建 [D]. 北京: 中国科学院计算技术研究所, 2010
Chen Min-ming. Algorithms and Implementation of Matrix Reconstruction [D]. Beijing: Chinese Academy of Sciences, 2010
- [3] Zhang F, Chang H. A collaborative filtering algorithm embedded BP network to ameliorate sparsity issue [C] // Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005. IEEE, 2005:1839-1844
- [4] Jung K Y, Hwang H J, Kang U G. Constructing full matrix through naive Bayesian for collaborative filtering [M] // Computational Intelligence. Springer Berlin Heidelberg, 2006: 1210-1215
- [5] Candès E J, Recht B. Exact matrix completion via convex optimization [J]. Foundations of Computational mathematics, 2009, 9(6):717-772
- [6] 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(15):615-640
- [7] Lin Zhou-chen, Chen Min-ming, Ma Yi. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices [J]. arXiv preprint arXiv:1009.5055, 2010
- [8] Tao M, Yuan X. Recovering low-rank and sparse components of matrices from incomplete and noisy observations [J]. SIAM Journal on Optimization, 2011, 21(1):57-81
- [9] Yuan X, Yang J. Sparse and low-rank matrix decomposition via alternating direction methods [J]. Preprint, 2009, 9(1):1-16
- [10] Yang J, Yuan X. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization [J]. Mathematics of Computation, 2013, 82(281):301-329
- [11] Lin Z, Liu R, Su Z. Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation [J] // arXiv:1109.0367, 2011
- [12] Cai J F, Candès E J, Shen Z. A singular value thresholding algorithm for matrix completion [J]. SIAM Journal on Optimization, 2010, 20(4):1956-1982
- [13] Gu S, Zhang L, Zuo W, et al. Weighted Nuclear Norm Minimization with Application to Image Denoising [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2862-2869
- [14] Xie Q, Meng D, Gu S, et al. On the Optimal Solution of Weighted Nuclear Norm Minimization [J]. arXiv preprint arXiv:1405.6012, 2014