融合显/隐式信任协同过滤算法的差分隐私保护

鲜征征[1,2],李启良[2],黄晓宇[3],陆寄远[1],李 磊[2]
(1. 广东金融学院互联网金融与信息工程学院,广东广州 510521;2. 中山大学数据科学与计算机学院,广东广州 510006;
3. 华南理工大学经济与贸易学院,广东广州 510006)

摘 要: 融合显/隐式信任关系的社会化协同过滤算法 TrustSVD 在推荐系统中有广泛的应用,但该算法存在用户隐私泄漏的风险. 基于背景知识的用户个人隐私信息推断是当前 Internet 用户隐私信息泄漏的巨大隐患之一,差分隐私作为一种能为保护对象提供严格的理论保证的隐私保护机制而备受关注. 本文把差分隐私保护技术引入 TrustSVD 中,提出了具有隐私保护能力的新模型 DPTrustSVD. 理论分析和实验结果显示,DPTrustSVD 不仅为用户的隐私信息提供了严格的理论保证,而且仍然保持了较高的预测准确率.

关键词: 社会化协同过滤; 个人隐私保护; 差分隐私; 矩阵分解; 信任关系; 隐式信任
中图分类号: TP311 **文献标识码:** A **文章编号:** 0372-2112 (2018)12-3050-10
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2018.12.032

Differential Privacy Protection for Collaborative Filtering Algorithms with Explicit and Implicit Trust

XIAN Zheng-zheng[1,2], LI Qi-liang[2], HUANG Xiao-yu[3], LU Ji-yuan[1], LI Lei[2]
(1. *School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou, Guangdong* 510521, *China*;
2. *School of Data and Computer Science, Sun Yat-sen University, Guangzhou, Guangdong* 510006, *China*;
3. *School of Economics and Commerce, South China University of Technology, Guangzhou, Guangdong* 510006, *China*)

Abstract: TrustSVD, a popular social collaborative filtering algorithm that incorporates both of the explicit and implicit trust information, has been widely used in recommender systems. However, there is a risk of disclosure of user privacy in TrustSVD. Privacy information inference based on background knowledge is one of the great hidden dangers of user's privacy disclosure. Differential privacy has attracted much attentiaon as a privacy protection mechanism that can provide a strict theoretical guarantee for protection objects. In this article, we propose DPTrustSVD, a novel collaborative filtering algorithm that applies Differential privacy to TrustSVD and has the ability of privacy preserving. Theoretical analysis and experimental results show that DPTrustSVD not only provides a strict theoretical guarantee for users' privacy information, but also maintains a high prediction accuracy.

Key words: social collaborative filtering; personal privacy preservation; differential privacy; matrix factorization; trust relationship; implicit trust

1 引言

协同过滤技术经过二十来年的沉淀,已在推荐系统领域得到迅速发展并取得巨大成就. 经典的协同过滤技术是以用户对项目的评分作为唯一的计算依据,学习用户的兴趣偏好,最终将其潜在感兴趣的项目推荐给用户[1]. 现实世界里,人们更乐于从熟人(朋友、同事等)获取信息. 这意味着潜藏在用户间的社交网络之下的信任关系可以用于改善推荐准确率[2,3]. Jamali 等[4]提出的 SocialMF 模型,Yang 等[2]提出的 TrustMF 模型,Yao 等[5]提出的 RoRec 模型以及 Guo 等[6,7]提出的 TrustSVD 模型,都是融入了社会关系的协同过滤技术的典型模型. 其中,TrustSVD 模型不仅在预测准确率上有着明显的优势,而且在解决数据稀疏性和冷启动

收稿日期:2017-06-13;修回日期:2018-04-22;责任编辑:孙瑶
基金项目:广东省自然科学基金(No. 2017A030313391);广东省科技计划(No. 2017A050501042, No. 2016ZC0039, No. 2017ZC0117);广东省哲学社科(No. GD15CGL05)

方面也有突出表现.

协同过滤技术的发展离不开海量数据的大规模采集.然而,近年来随着大数据的相关研究和应用,用户个人隐私信息泄露的风险也随之增大.用户数据的安全与隐私保护无疑成为一个亟待解决的重要问题[8].在推荐系统研究中,用户评分和社会关系都存在隐私泄露的风险.譬如:某电影网站有 100 个用户参与了《变形金刚》的评分,已知所有用户的平均评分为 4.2 分(1 ~ 5 分评分制),当某新用户也对该电影做了评分,且平均评分被更新为 4.17,即使该新用户并未公布他的评分,但可根据 $\lfloor 101 \times 4.17 - 100 \times 4.2 \rfloor = 1$("$\lfloor \bullet \rfloor$"表示向下取整)得出他对该电影评分为 1 分,进而可猜测该用户可能不喜欢科幻片.反之,若用户对某些敏感题材电影给出高评分,则可能反映出用户对这类题材有特殊喜好,而他本人却不希望这些喜好为人所知.从社会化信任关系的角度,通常某用户与其信任列表中的朋友有着类似的兴趣爱好.那么当数据集中包含了用户评分以及用户信任关系时,若用户 A 的个人敏感信息被泄露,且不对其社会关系进行保护,那么信任 A 的其他用户的类似偏好也会被间接泄露.因此,除了用户显式评分,用户间的社会关系同样应该得到保护,特别是利用背景知识间接泄漏隐私的问题,更成为业界研究的热点.

差分隐私保护[9](Differential Privacy,DP)模型是由 Dwork 等人于 2006 年提出的一种定义极为严格、与背景知识无关、可证明的强隐私保护模型.该模型力图实现在"所有可能被公开的数据都已知"的情况下,通过对数据或算法加扰的方式来保护原始数据中潜在的用户敏感信息,旨在达到即便攻击者已掌握除某一条信息以外的其他信息,也无法推测出这条信息的目标[10].换句话说,对某随机算法做差分隐私保护处理后,在可以得到近似结果的前提下,也难以利用任何背景知识推出某条信息是否存在于输入的数据集中.

近年来,将差分隐私引入协同过滤技术中的研究已取得一些成果.相关工作包括 McSherry 等[11]提出的对 item-to-item 协方差矩阵注入噪音的差分隐私保护策略;Hardt 等[12]提出的对评分矩阵作低秩近似的差分隐私保护策略;Friedman 等[13]提出将差分隐私应用于矩阵分解的各个步骤之中,并对原始评分矩阵做相关预处理等;Hua 等[14]提出首先阻止不受信任的推荐者使用用户的评分信息,同时允许用户在矩阵分解过程中离开和加入,通过对矩阵分解的目标函数加扰实现差分隐私保护;Yan 等[15]提出利用差分隐私和用户信任关系来自适应地修改用户评分历史的协同过滤新策略;为追踪用户的兴趣变化,弥补兴趣漂移对推荐质量带来的影响,范等[16]提出一种基于差分隐私和时序的推荐系统新模型,即用时间衰减函数来作为原始评分的时间权重.

现有研究中,针对社会化信任关系的协同过滤技术的隐私保护工作尚不多见.因此,从考虑隐私保护和预测准确率两者间的折中以及协同过滤技术中的数据稀疏性和冷启动问题,本文将差分隐私保护技术引入融合显/隐式信任关系的 SVD + + 协同过滤技术中,提出目标函数加扰的 TrustSVD 差分隐私保护新策略.关于新策略,文中在理论上分析了其隐私保护的性能,实验上验证了其在协同过滤应用中的预测表现.结果表明:所提新策略与无隐私保护的 TrustSVD 具有相近的预测准确率,与做类似差分隐私保护的 SVD + + 相比获得了更优的预测结果,此外还给出了核心参数的调节实验.

2　知识背景

2.1　信任关系

信任关系是最常见的社会关系之一.推荐系统中的信任关系可分为显式信任和隐式信任.前者是某用户直接表明对其他用户的信任[7].譬如在 Epinions① 数据集中,用户 A 可以把那些与自己兴趣相投的用户添加到信任列表中,那么用户 A 对某些物品的喜好程度将受到信任列表中的用户所影响.同样,用户 A 的贡献也将影响那些信任他的用户.这样可为用户 A 建立一个包含信任和被信任关系的信任网络.隐式信任则是没有直接表明信任关系,但可通过用户历史评分或用户间的相互关系等信息推导出来,以弥补显式信任关系无法获取或稀疏的情况.本文将凭借用户历史评分来推导出用户间的隐式信任.

2.2　TrustSVD 模型

2.2.1　SVD + + 矩阵分解模型

记 $\boldsymbol{R} \in \mathbf{R}^{n \times m}$ 是 n 名用户对 m 个项目的评分矩阵,其中 $r_{u,i}$ 表示用户 u 对项目 i 的评分,$r_{u,i}$ 缺失时令 $r_{u,i} = 0$.协同过滤研究的主要任务是如何根据 $\boldsymbol{R}$ 中非 0 的评分数据,对缺失的数据给出合理的预测.Koren 在文献[17]中提出了如下的 SVD + + 模型

$$L = \sum_u \sum_i (r_{u,i} - \tilde{r}_{u,i})^2 + \lambda(\mu^2 + \sum_u b_u^2 + \sum_i b_i^2 + \sum_u \| \boldsymbol{p}_u \|_F^2 + \sum_i \| \boldsymbol{q}_i \|_F^2 + \sum_u \sum_{j \in \boldsymbol{I}_u} \| \boldsymbol{y}_j \|_F^2) \tag{1}$$

其中,$\tilde{r}_{u,i}$ 表示 u 对项目 i 的预测评分,计算方法为

$$\tilde{r}_{u,i} = \mu + b_u + b_i + \boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u + |\boldsymbol{I}_u|^{-1/2} \sum_{j \in \boldsymbol{I}_u} \boldsymbol{y}_j) \tag{2}$$

$\boldsymbol{P} \in \mathbf{R}^{n \times d}$ 表示用户特征矩阵,第 u 个行向量 $\boldsymbol{p}_u$ 对应用户 u 的一个 d 维特征描述;$\boldsymbol{Q} \in \mathbf{R}^{d \times m}$ 表示项目特征矩阵,第 i 个列向量 $\boldsymbol{q}_i$ 对应项目 i 的一个 d 维特征描述,d 通常取低

① http://trustlet.org/wiki/Epinions_datasets

维. $\boldsymbol{B}_u=[b_1,b_2,\cdots,b_n]'$是用户评分的偏置向量,其分量 b_u 对应用户 u 的评分偏好相对于所有用户对所有项目给出的评分的平均分数 μ 的偏置;$\boldsymbol{B}_i=[b_1,b_2,\cdots,b_m]'$是项目得分的偏置向量,其分量 b_i 对应项目 i 的得分相对于 μ 的偏置. $\boldsymbol{I}_u$ 记录 u 给出了评分的项目的集合,$|\boldsymbol{I}_u|$表示 u 给出评分的项目数. 对每个 $j\in\boldsymbol{I}_u$,d 维向量 $\boldsymbol{y}_j$ 对应了基于 u 对 j 的评分行为而得到的隐式特征反馈,$Y=I_1\cup I_2\cup\cdots\cup I_n$ 是所有隐式反馈向量集合之并.

2.2.2 信任关系建模

有向图 $G=(V,E)$表示一个社会化网络,其中 V 是 n 个用户的集合,E 为用户间显式信任关系的有向边的集合. 将信任网络转换为对应的信任关系邻接矩阵 $\boldsymbol{T}\in\mathbf{R}^{n\times n}$,$t_{u,v}$代表用户 u 对用户 v 的显式信任程度. 与用户评分矩阵类似,显式信任关系矩阵也是稀疏的. Yang 等[2]提出把用户的显式信任关系映射到 Truster 和 Trustee 两个低维空间,前者向量表示用户 u 的信任行为,本质上刻画了 u 受其信任用户的影响,后者向量表示 u 的被信任行为,本质上刻画了 u 对信任他的用户所产生的影响. 对同一数据集,评分矩阵 $\boldsymbol{R}$ 和信任关系矩阵 $\boldsymbol{T}$ 所涉及的用户是同一组,因此可通过共享用户特征空间 $\boldsymbol{P}$ 将 $\boldsymbol{R}$ 和 $\boldsymbol{T}$ 联合到同一个(都是 d 维)矩阵分解过程中[2]. 用$\tilde{t}_{u,v}=\boldsymbol{p}_u\boldsymbol{w}_v^{\mathrm{T}}$来预测用户 u 对用户 v 的信任值,用$\tilde{t}_{k,u}=\boldsymbol{p}_k\boldsymbol{w}_u^{\mathrm{T}}$来预测信任 u 的用户 k 对其的信任值. 由此,建立如下线性关系来描述用户 u 的两种显式信任

$$\sum_u\left(\alpha\sum_{v\in T_u^+}(t_{u,v}-\tilde{t}_{u,v})^2+(1-\alpha)\sum_{k\in T_u^-}(t_{k,u}-\tilde{t}_{k,u})^2\right) \tag{3}$$

其中 T_u^+ 表示 u 直接信任的用户集合,T_u^- 表示信任 u 的用户集合. 信任关系的隐私反馈信息可根据用户评分隐式反馈信息来定义. 类似评分,用$|T_u^+|^{-1/2}\sum_{v\in T_u^+}w_v$表示 u 信任的用户对其评分产生的隐式影响. 用$|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k$表示信任 u 的用户对其评分产生的隐式影响. Guo 等[7]提出通过融合显/隐式信任关系来对用户特征进行更详细地建模,预测评分即为

$$\begin{aligned}\tilde{r}_{u,i}=&\mu+b_u+b_i+\boldsymbol{q}_i^{\mathrm{T}}\Big(\boldsymbol{p}_u+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in I_u}\boldsymbol{y}_j\\&+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k\Big)\end{aligned} \tag{4}$$

其中,$\alpha\in[0,1]$用于控制两种隐式信任对预测评分影响的重要性. 特别地,当 $\alpha=0$ 意味着只考虑信任 u 的用户带来的隐式影响,$\alpha=1$ 意味着只考虑 u 信任的用户带来的隐式影响.

2.2.3 融合显/隐式信任的 TrustSVD 模型

将显/隐式信任融合到 SVD + + 中,Guo 等[7]提出 TrustSVD 模型,其损失函数为

$$\begin{aligned}L=&\frac{1}{2}\sum_u\sum_{i\in\boldsymbol{I}_u}(r_{u,i}-\tilde{r}_{u,i})^2\\&+\frac{\lambda_t}{2}\sum_u\Big(\alpha\sum_{v\in T_u^+}(t_{u,v}-\tilde{t}_{u,v})^2+(1-\alpha)\sum_{k\in T_u^-}(t_{k,u}-\tilde{t}_{k,u})^2\Big)\\&+\frac{\lambda}{2}\Big(\mu^2+\sum_u|\boldsymbol{I}_u|^{-1/2}b_u^2+\sum_i|\boldsymbol{U}_i|^{-1/2}b_i^2\Big)\\&+\sum_u\left(\frac{\lambda}{2}|\boldsymbol{I}_u|^{-1/2}+\frac{\lambda_t}{2}\begin{pmatrix}(\delta(\alpha)|T_u^+|^{-1/2})\\+\delta(1-\alpha)|T_u^-|^{-1/2}\end{pmatrix}\right)\|\boldsymbol{p}_u\|_F^2\\&+\frac{\lambda}{2}\Big(\sum_i|\boldsymbol{U}_i|^{-1/2}\|\boldsymbol{q}_i\|_F^2+\sum_j|\boldsymbol{U}_j|^{-1/2}\|\boldsymbol{y}_j\|_F^2\Big)\\&+\frac{\lambda}{2}\sum_u\sum_{v\in T_u^+}\delta(\alpha)|T_v^+|^{-1/2}\|\boldsymbol{w}_v\|_F^2\\&+\frac{\lambda}{2}\sum_u\sum_{k\in T_u^-}\delta(1-\alpha)|T_k^-|^{-1/2}\|\boldsymbol{p}_k\|_F^2\end{aligned} \tag{5}$$

其中,$|\boldsymbol{U}_i|$、$|\boldsymbol{U}_j|$分别表示参与项目 i,j 评分的用户数;$\delta(x)$是一个指示函数,当 $x>0$ 时 $\delta(x)=1$,否则 $\delta(x)=0$. λ 和 λ_t 是正则化参数,λ_t 用于控制用户的信任程度.

2.3 差分隐私保护

差分隐私保护技术能在数据的隐私保护和可用性之间做很好的权衡. 本文对于差分隐私的相关概念大多源于文献[18~22],结合论文研究内容,略有改动.

定义 1 相邻评分矩阵. 记 $\boldsymbol{R}$ 和 $\boldsymbol{R}'$是由相同 n 名用户对 m 个项目做的评分矩阵(可能是不完全的),$\boldsymbol{R}$ 和 $\boldsymbol{R}'$是“相邻”的,当且仅当它们两者间至多相差一个元素.

例 假设有 u_1、u_2 两名用户对 $i_1\sim i_3$ 三部电影做出评分的矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{R}'$,其中 $\boldsymbol{R}$ 包含了 u_1 对 i_1、u_2 对 i_2 的评分,$\boldsymbol{R}'$除了这两个评分外,还包含了 u_1 对 i_3 的评分,则 $\boldsymbol{R}$ 和 $\boldsymbol{R}'$是相邻的.

定义 2 全局敏感度. 给定两个相邻评分矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{R}'$,对于任意预测函数 $f:\mathbf{R}^{n\times m}\to\mathbf{R}^{n\times m}$,它的 L_n 全局敏感度 GS_f为

$$GS_f=\max_{\boldsymbol{R},\boldsymbol{R}'}\|f(\boldsymbol{R})-f(\boldsymbol{R}')\|_n \tag{6}$$

其中 $\|.\|$ 表示 n 范数.

定义 3 ε-差分隐私. 给定值域为 Range(A)的随机算法 $A:\mathbf{R}^{n\times m}\to\mathbf{R}^{n\times m}$,当输入两个相邻的评分矩阵 $\boldsymbol{R}$、$\boldsymbol{R}'$和非负实数 ε,若对于任意 $S\subset\mathrm{Range}(A)$,都有

$$\Pr[A(\boldsymbol{R})\in S]\le e^{\varepsilon}\times\Pr[A(\boldsymbol{R}')\in S] \tag{7}$$

则称算法 A 满足 ε-差分隐私. 其中,$\Pr[.]$表示隐私被披露的概率,它是由算法 A 的随机性所控制(与攻击者的背景知识无关);ε 是隐私保护参数,表示隐私保护的力度,ε 越小意味着隐私保护力度越强. 定义 3 本质上刻画了基于随机算法 A 输出的两个相邻矩阵的不可分程度.

拉普拉斯机制(Laplace Mechanism)是最常见的差

分隐私实现机制之一,其基本策略是通过向聚合数据中注入服从 Laplace 分布的噪音变量,从而实现对个体数据的保护.

定义 4　Laplace 机制. 给定两个相邻的评分矩阵 $\boldsymbol{R}$、$\boldsymbol{R}'$和非负实数 ε,对于任意预测函数 $f:\mathbf{R}^{n\times m}\rightarrow\mathbf{R}^{n\times m}$ 的全局敏感度为 GS_f,如果随机噪声 $Y\sim\mathrm{Lap}(GS_f/\varepsilon)$,则随机算法 $A(\boldsymbol{R})=f(\boldsymbol{R})+Y$ 满足 ε-差分隐私.

从定义 4 可知,加入的随机噪声与 GS_f 正比,与 ε 成反比.

拉普拉斯分布的概率密度函数为

$$f(x|\mu,b)=\frac{1}{2b}\mathrm{e}^{-|x-\mu|} \tag{8}$$

其中 μ 和 b 分别为变量 x 的位置和尺度参数,为方便获取噪声,设 $\mu=0$,则尺度参数为 b 的 Laplace 分布记为 $\mathrm{lap}(b)$.

3　TrustSVD 的差分隐私保护新策略

针对社会化协同过滤算法 TrustSVD,本文通过引入目标函数加扰策略[23]来实现对其的差分隐私保护,提出新算法 DPTrustSVD.

为求解 TrustSVD 模型,文献[6,7]采用梯度下降,考虑到梯度下降容易陷入局部最优,本文采用最小二乘法(Alternating Least Squares,ALS)优化策略. 其基本思路是:对一阶可导的具有 c 维自变量的凸函数 $f:\mathbf{R}^c\rightarrow\mathbf{R}$,为求得 $\min\limits_{x_1,x_2,\cdots,x_c} f(x_1,x_2,\cdots,x_c)$,可对 $l=1,2,\cdots,c$ 顺次令 $\dfrac{\partial f}{\partial x_l}=0$ 进行求解. 具体地,根据 ALS 优化策略求解 TrustSVD 目标函数(式(5))中的 $\mu,b_u,b_i,q_i,p_u,y_j,w_v,p_k$,将分别产生相应的凸优化函数. 篇幅有限,仅以 b_u 和 p_u 的凸优化函数为例,如式(9)和式(10)所示. 为简化式(5),设 $e_{u,i}=r_{u,i}-\tilde{r}_{u,i}$ 为评分误差.

$$J(b_u,\boldsymbol{R},\boldsymbol{T})=\frac{1}{2}\sum_u\sum_i(e_{u,i})^2+\frac{\lambda}{2}\sum_u|\boldsymbol{I}_u|^{-1/2}b_u^2 \tag{9}$$

$$\begin{aligned}J(\boldsymbol{p}_u,\boldsymbol{R},\boldsymbol{T})=&\frac{1}{2}\sum_u\sum_i(e_{u,i})^2+\sum_u\Big(\frac{\lambda}{2}|\boldsymbol{I}_u|^{-1/2}\\&+\frac{\lambda_t}{2}\big((\delta(\alpha)|T_u^+|^{-1/2}\\&+\delta(1-\alpha)|T_u^-|^{-1/2})\big)\Big)\|\boldsymbol{p}_u\|_F^2\end{aligned} \tag{10}$$

针对各项参数的凸优化函数,再将相应因子的偏导设为 0,如:$\dfrac{\partial J(b_u,R)}{\partial b_u}=0$. 则得到各项因子的解,分别为式(11)~(18).

$$\begin{aligned}\mu=&(\lambda+1)^{-1}+(r_{u,i}-b_u-b_i-\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u\\&+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k))\end{aligned} \tag{11}$$

$$\begin{aligned}b_i=&(\lambda|\boldsymbol{U}_i|^{-1/2}+1)^{-1}\sum_{i\in\boldsymbol{I}_u}(r_{u,i}-b_u-\mu\\&-\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k))\end{aligned} \tag{12}$$

$$\begin{aligned}b_u=&(\lambda|\boldsymbol{I}_u|^{-1/2}+1)^{-1}\sum_{i\in\boldsymbol{I}_u}(r_{u,i}-b_i-\mu\\&-\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k))\end{aligned} \tag{13}$$

$$\begin{aligned}\boldsymbol{q}_i=&\Big(\sum_{u\in\boldsymbol{U}_i}(\boldsymbol{p}_u+|\boldsymbol{I}_u|^{-1/2}\sum_{i\in\boldsymbol{I}_u}\boldsymbol{y}_i\\&+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k)(\boldsymbol{p}_u\\&+|\boldsymbol{I}_u|^{-1/2}\sum_{i\in\boldsymbol{I}_u}\boldsymbol{y}_i+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k)^{\mathrm{T}}+\lambda|\boldsymbol{U}_i|^{-1/2}\boldsymbol{Z}\Big)^{-1}\\&\cdot\Big(\sum_{u\in\boldsymbol{U}_i}((r_{u,i}-\mu-b_i-b_u)(\boldsymbol{p}_u+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j\\&+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k))\Big)\end{aligned} \tag{14}$$

$$\begin{aligned}\boldsymbol{p}_u=&\Big(\sum_{i\in\boldsymbol{I}_u}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}+\lambda_t\alpha\sum_{v\in T_u^+}\boldsymbol{w}_v\boldsymbol{w}_v^{\mathrm{T}}+(\lambda|\boldsymbol{I}_u|^{-1/2}\\&+\lambda_t(\delta(\alpha)|T_u^+|^{-1/2}+\delta(1-\alpha)|T_u^-|^{-1/2}))\boldsymbol{Z}\Big)^{-1}\\&\cdot\Big(\sum_{i\in\boldsymbol{I}_u}(r_{u,i}-\mu-b_i-b_u)\boldsymbol{q}_i\\&-\sum_{i\in\boldsymbol{I}_u}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}(|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k)+\lambda_t\alpha\sum_{v\in T_u^+}t_{u,v}\boldsymbol{w}_v\Big)\end{aligned} \tag{15}$$

$\forall j\in\boldsymbol{I}_u$,

$$\begin{aligned}\boldsymbol{y}_j=&\Big(\sum_{i\in\boldsymbol{I}_u}|\boldsymbol{I}_u|^{-1}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}+\lambda|\boldsymbol{U}_j|^{-1/2}\boldsymbol{Z}\Big)^{-1}\\&+\Big(\sum_{i\in\boldsymbol{I}_u}|\boldsymbol{I}_u|^{-1/2}(r_{u,i}-\mu-b_i-b_u)\boldsymbol{q}_i\\&-\sum_{i\in\boldsymbol{I}_u}|\boldsymbol{I}_u|^{-1/2}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k)\\&-\Big(\sum_{i\in\boldsymbol{I}_u}|\boldsymbol{I}_u|^{-1}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}+\lambda|\boldsymbol{U}_j|^{-1/2}\boldsymbol{Z}\Big)\sum_{k\in\boldsymbol{I}_u,k\neq i}\boldsymbol{y}_k\Big)\end{aligned} \tag{16}$$

$\forall v\in T_u^+$,

$$\begin{aligned}\boldsymbol{w}_v=&\Big(\sum_{i\in\boldsymbol{I}_u}\alpha|T_u^+|^{-1}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}+\lambda_t\alpha\boldsymbol{p}_u\boldsymbol{p}_u^{\mathrm{T}}\\&+\lambda\delta(\alpha)|T_u^+|^{-1/2}\boldsymbol{Z}\Big)^{-1}\Big(\sum_{i\in\boldsymbol{I}_u}\alpha|T_u^+|^{-1/2}(r_{u,i}\\&-\mu-b_i-b_u)\boldsymbol{q}_i-\sum_{j\in\boldsymbol{I}_u}(\alpha|T_u^+|^{-1/2}\boldsymbol{q}_j\boldsymbol{q}_j^{\mathrm{T}}(\boldsymbol{p}_u\\&+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k))\\&+\lambda_t\alpha t_{u,u}\boldsymbol{p}_u\Big)\end{aligned} \tag{17}$$

$\forall k \in T_u^-$,

$$\boldsymbol{p}_k = ((1-\alpha)\sum_{i\in \boldsymbol{I}_u} |T_u^-|^{-1/2}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}} + \lambda_t(1-\alpha)\boldsymbol{w}_u\boldsymbol{w}_u^{\mathrm{T}} + \lambda\delta(1-\alpha)|T_k^-|^{-1/2}\boldsymbol{Z})^{-1}(\sum_{j\in \boldsymbol{I}_u}(1-\alpha)|T_u^-|^{-1/2}(r_{u,i} - \mu - b_j - b_u)\boldsymbol{q}_i - \sum_{j\in \boldsymbol{I}_u}(1-\alpha)|T_u^-|^{-1/2}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u + |\boldsymbol{I}_u|^{-1/2}\sum_{j\in \boldsymbol{I}_u}\boldsymbol{y}_j + \alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v) + \lambda_t(1-\alpha)t_{k,u}\boldsymbol{w}_u) \tag{18}$$

上述式中,$\boldsymbol{I}_u = \{i \mid r_{u,i} > 0\}$,$\boldsymbol{U}_i = \{u \mid r_{u,i} > 0\}$,$\boldsymbol{Z}$ 为 $d \times d$ 的单位矩阵.

新算法 DPTrustSVD(算法 1)的核心思想是通过对求解 TrustSVD 目标函数的 ALS 优化策略实施差分隐私保护. 下面给出新算法的详细描述、分析、相关定理及关键证明.

算法 1 DPTrustSVD 算法

输入: $\boldsymbol{R} \in \mathbf{R}^{n\times m}$——$n$ 名用户对 m 个项目的评分矩阵(若 $R_{u,i}$ 值缺失,则令 $R_{u,i}=0$)
$\boldsymbol{T} \in \mathbf{R}^{n\times n}$——$n$ 名用户之间的信任关系邻接矩阵
d——矩阵分解隐含特征矩阵的特征个数
λ、λ_t——目标函数中的正则化参数
ε——差分隐私保护预算参数
c——计算松弛项的参数
h——ALS 求解迭代次数
输出: $\mu, B_u, B_i, \boldsymbol{P}, \boldsymbol{Q}, Y$ 和 W

```
1   初始化:μ 为所有有效评分的均值,户偏置因子 B_u 和 B_i 为所有项目的平均评分,用户矩阵 P,项目矩阵 Q,隐式反馈项 Y 和 W 为随机值.
2   FOR l = 1 TO h - 1 DO
3       CALL ALSforTrustSVD (false)
4   END FOR
5   CALL ALSforTrustSVD(true)
6   FOR u = 1 TO n DO
7       ε' = ε - log(1 + 2c/(NΛ) + c^2/(N^2Λ^2))
8       IF ε' > 0 THEN
9           t = 0
10      ELSE
11          t = c/(N(e^{ε/4} - 1)) - Λ
12      END IF
13      ε' = ε/2
14      由 p(o) ∝ e^{(-ε'‖o‖/(2Δ))} 生成 Lap 噪声向量 o
15      计算加扰后的 p_u^priv(式(20))
16  END FOR
17  返回 μ, B_u, B_i, P, Q, Y 和 W
```

函数 1 ALSforTrustSVD(bool isPrivate)

```
/* 功能:交替最小二乘法求解 TrustSVD
说明:若 isPrivate 为 false,则为 ALS 求解的常规版本,即对所有参数都作更新;若 isPrivate 为 true,则为 ALS 的差分隐私保护版本,只对除 p_u 外的其他参数作更新. */
1   更新 μ(式(11))
2   FOR i = 1 TO m DO
3       更新 b_i(式(12))
4       更新 q_i(式(14))
5   END FOR
6   FOR u = 1 TO n DO
7       更新 b_u(式(13))
8       ∀j ∈ I_u,更新 y_j(式(16))
9       ∀v ∈ T_u^+,更新 w_v(式(17))
10      ∀k ∈ T_u^-,更新 p_k(式(18))
11      IF !isPrivate THEN
12      更新 p_u(式(15))
13      END IF
14      END FOR
```

针对算法 1,给出以下几点说明:

(1)算法 1 总体设计策略:前 $h-1$ 次迭代先执行常规的 ALS 求解来拟合 TrustSVD 模型中的各参数,然后在最后一次迭代中对拟合的结果加以扰动来实现差分隐私保护. 其中,算法 1 的 2~4 行是调用函数 ALSforTrustSVD 来更新全部参数,第 5 行仅更新除 $\boldsymbol{p}_u$ 外的其他参数.

(2)算法 1 的 6~16 行是差分隐私保护的实现. 对每一个用户 u,生成一个 Laplace 噪声变量 $\boldsymbol{o}$(第 14 行),其中 Δ 为最高评分与最低评分之差. ALS 求解用户特征矩阵 $\boldsymbol{P}$ 的式(10)变为

$$J^{\mathrm{priv}}(\boldsymbol{p}_u, \boldsymbol{R}, \boldsymbol{T}) = J(\boldsymbol{p}_u, \boldsymbol{R}, \boldsymbol{T}) + \frac{1}{n}\boldsymbol{o}^{\mathrm{T}}\boldsymbol{p}_u \tag{19}$$

针对式(19),采用经验风险最小化(Empirical Risk Minimization, ERM)[23] 来获得差分隐私处理的最优 $\boldsymbol{p}_u$,即

$$\boldsymbol{p}_u^{\mathrm{priv}} = \arg\min_{\boldsymbol{p}_u} J^{\mathrm{priv}}(\boldsymbol{p}_u, \boldsymbol{R}, \boldsymbol{T}) + \frac{1}{2}t\|\boldsymbol{p}_u\|^2 \tag{20}$$

为防止加扰后的结果过拟合,添加正则项 $\frac{1}{2}t\|\boldsymbol{p}_u\|^2$,$t$ 通过算法 1 的第 8~12 行求得. 式(20)右端的和式对 $\boldsymbol{p}_u$ 为凸,令右端和式对 $\boldsymbol{P}$ 的偏导为 0,可求得 $\boldsymbol{p}_u^{\mathrm{priv}}$ 的解为

$$\boldsymbol{p}_u^{\mathrm{priv}} = (\sum_{i\in \boldsymbol{I}_u}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}} + \lambda_t\alpha\sum_{v\in T_u^+}\boldsymbol{w}_v\boldsymbol{w}_v^{\mathrm{T}} + (\lambda|\boldsymbol{I}_u|^{-1/2} + \lambda_t(\delta(\alpha)|T_u^+|^{-1/2} + \delta(1-\alpha)|T_u^-|^{-1/2}) + \frac{1}{2}\Delta)\boldsymbol{Z})^{-1}(\sum_{i\in \boldsymbol{I}_u}(r_{u,i} - \mu - b_i - b_u)\boldsymbol{q}_j - \sum_{i\in \boldsymbol{I}_u}\boldsymbol{q}_i\boldsymbol{q}_i^{\mathrm{T}}(|\boldsymbol{I}_u|^{-1/2}\sum_{j\in \boldsymbol{I}_u}\boldsymbol{y}_j + \alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v + (1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k) + \lambda_t\alpha\sum_{v\in T_u^+}t_{u,v}\boldsymbol{w}_v - \frac{1}{N}\boldsymbol{o}) \tag{21}$$

类似可分别求得 n 个用户对应的 $\boldsymbol{p}_1^{\mathrm{priv}}, \boldsymbol{p}_2^{\mathrm{priv}}, \cdots, \boldsymbol{p}_n^{\mathrm{priv}}$.

(3)算法第 7、11 行中,N 表示评分总数,Λ 为 ALS 求解 $\boldsymbol{p}_u$ 的正则项系数,即

$$\Lambda=\frac{\lambda}{2}|\boldsymbol{I}_u|^{-1/2}+\frac{\lambda_t}{2}(\delta(\alpha)|T_u^+|^{-1/2}+\delta(1-\alpha)|T_u^-|^{-1/2}) \tag{22}$$

(4)算法 1 的第 7 和 11 行中参数 c 的确定. 根据文献[23]的推论 5 和推论 6,对式(10)中的损失函数 $\ell(e_{u,i})=(e_{u,i})^2$ 求二阶导数,可得

$$\ell'(e_{u,i})=\frac{\partial\ell(e_{u,i})}{\partial e_{u,i}}=2e_{u,i},\ell''(e_{u,i})=\frac{\partial\ell'(e_{u,i})}{\partial e_{u,i}}=2 \tag{23}$$

则有$|\ell''(e_{u,i})|\leq c$,此时 $c=2$.

关于算法 1 的隐私保护性能,给出如下定理.

定理 1　算法 1 满足 ε-差分隐私.

证明

(1)对每一个 $u\in[n]$,由式(5)得到经验误差因子为

$$\begin{aligned}e_{u,i}&=r_{u,i}-\tilde{r}_{u,i}\\&=r_{u,i}-\mu-b_u-b_i-\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&\quad+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k)\end{aligned} \tag{24}$$

且损失函数 $\ell(e_{u,i})$是二阶可导的凸函数.

(2)假设 $\boldsymbol{p}_u^{\mathrm{priv}}$ 是算法 1 中对 $\boldsymbol{p}_u$ 实施目标函数加扰后的输出结果. 如果给定任何一个固定的 $\boldsymbol{p}_u^{\mathrm{priv}}$ 和固定的数据集 $\boldsymbol{R}$,总是存在一个 $\boldsymbol{o}$(Laplace 噪声)使得算法 1 在数据集 $\boldsymbol{R}$ 输出 $\boldsymbol{p}_u^{\mathrm{priv}}$. 因为损失函数 $\ell(e_{u,i})=(e_{u,i})^2$ 是二阶可导的凸函数,当给定两个相邻评分矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{R}'$,那么根据式(20),再由式(25)

$$\nabla_{p_u}J^{\mathrm{priv}}(\boldsymbol{p}_u,\boldsymbol{R},\boldsymbol{T})=\nabla_{p_u}J^{\mathrm{priv}}(\boldsymbol{p}_u,\boldsymbol{R}',\boldsymbol{T})=0 \tag{25}$$

可得

$$\boldsymbol{o}_u-\sum_u\sum_i\boldsymbol{q}_ie_{u,i}=\boldsymbol{o}'_u-\sum_u\sum_i\boldsymbol{q}_ie'_{u,i} \tag{26}$$

由式(26)可知,如设相邻评分矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{R}'$的最后一个元素$(r_{n,m})$不同,则有

$$\begin{aligned}\frac{1}{2}\boldsymbol{o}_n-\boldsymbol{q}_me_{n,m}&=\frac{1}{2}\boldsymbol{o}'_n-\boldsymbol{q}_me'_{n,m}\\\boldsymbol{o}_n-\boldsymbol{o}'_n&=2\boldsymbol{q}_m(e_{n,m}-e'_{n,m})\end{aligned} \tag{27}$$

按照式(24)展开 $e_{n,m}$,可得

$$\boldsymbol{o}_n-\boldsymbol{o}'_n=2\boldsymbol{q}_m(r_{n,m}-r'_{n,m}) \tag{28}$$

又因为 $\|\boldsymbol{q}_m\|\leq 1$ 且 $|r_{n,m}-r'_{n,m}|\leq\Delta$,则有

$$\|\boldsymbol{o}_n-\boldsymbol{o}'_n\|\leq 2\Delta \tag{29}$$

因此,当差分隐私保护参数为 ε'(由算法的 7 ~ 13 行计算而得)的相邻评分矩阵上的 Laplace 噪声之比为

$$\frac{\boldsymbol{p}(\boldsymbol{o}|\boldsymbol{R})}{\boldsymbol{p}(\boldsymbol{o}'|\boldsymbol{R}')}=\mathrm{e}^{\left(-\frac{\varepsilon'(\|\boldsymbol{o}_n\|+\|\boldsymbol{o}'_n\|)}{2\Delta}\right)}\leq\mathrm{e}^{\left(\frac{\varepsilon'2\Delta}{2\Delta}\right)}\leq\mathrm{e}^{\varepsilon'} \tag{30}$$

再设 $J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}|\boldsymbol{R})$和 $J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}'|\boldsymbol{R}')$是从 $\boldsymbol{p}_u^{\mathrm{priv}}$ 映射到 $\boldsymbol{o}$ 的 Jacobian 矩阵,可以获得

$$\frac{\Pr[\boldsymbol{p}_u|\boldsymbol{R}]}{\Pr[\boldsymbol{p}_u|\boldsymbol{R}']}=\frac{\boldsymbol{p}(\boldsymbol{o}|\boldsymbol{R})}{\boldsymbol{p}(\boldsymbol{o}'|\boldsymbol{R}')}\cdot\frac{|\det(J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}|\boldsymbol{R}))|}{|\det(J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}'|\boldsymbol{R}'))|} \tag{31}$$

再根据文献[23]的定理 2,可知

$$\frac{|\det(J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}|\boldsymbol{R}))|}{|\det(J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}'|\boldsymbol{R}'))|}=\mathrm{e}^{(\varepsilon-\varepsilon')} \tag{32}$$

综合式(30)和式(32),得到

$$\begin{aligned}\frac{\Pr[\boldsymbol{p}_u|\boldsymbol{R}]}{\Pr[\boldsymbol{p}_u|\boldsymbol{R}']}&=\frac{\Pr(\boldsymbol{o}|\boldsymbol{R})}{\Pr(\boldsymbol{o}'|\boldsymbol{R}')}\cdot\frac{|\det(J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}|\boldsymbol{R}))|}{|\det(J(\boldsymbol{p}_u^{\mathrm{priv}}\rightarrow\boldsymbol{o}'|\boldsymbol{R}'))|}\\&\leq\mathrm{e}^{\varepsilon'}\mathrm{e}^{(\varepsilon-\varepsilon')}\leq\mathrm{e}^{\varepsilon}\end{aligned} \tag{33}$$

从式(33)的结论以及定义 1 可知,$\boldsymbol{p}_u^{\mathrm{priv}}$ 满足 ε-差分隐私.

(3)经过算法 1 处理后获得的预测评分为

$$\begin{aligned}\tilde{r}_{u,i}&=\mu+b_u+b_i+\boldsymbol{q}_i^{\mathrm{T}}(\boldsymbol{p}_u^{\mathrm{priv}}+|\boldsymbol{I}_u|^{-1/2}\sum_{j\in\boldsymbol{I}_u}\boldsymbol{y}_j\\&\quad+\alpha|T_u^+|^{-1/2}\sum_{v\in T_u^+}\boldsymbol{w}_v\\&\quad+(1-\alpha)|T_u^-|^{-1/2}\sum_{k\in T_u^-}\boldsymbol{p}_k)\end{aligned} \tag{34}$$

式(34)中,$\boldsymbol{p}_u^{\mathrm{priv}}$ 是算法 1 的隐私保护步骤,其余参数均按照常规更新. 针对差分隐私保护预算 ε 的分配问题,McSherry 等[19]提出了两个差分隐私的组合特性:序列组合性质和并行组合性质. 前者说明了一系列串行的差分隐私保护算法构成的组合算法,它所提供的差分隐私保护水平为这些串行算法的全部预算的总和. 因为算法中除 $\boldsymbol{p}_u$ 外,未对其余参数实施差分隐私保护,也就是说,可以视为其余参数分配的隐私保护参数为 0. 因此,根据序列组合特性可得:算法 1 满足 ε-差分隐私.

关于算法 1,接下来针对其性能进行量化分析. 由于算法 1 中是对 $\boldsymbol{p}_u$ 做目标函数加扰的差分隐私保护处理,因此是对 $\boldsymbol{p}_u$ 的性能进行分析.

推论 1　设 $\boldsymbol{p}_0$ 是不做差分隐私保护处理时由期望的损失 $L(\boldsymbol{p}_0)$ 求得的一个最优解,损失函数 $\ell(e_{u,i})=(e_{u,i})^2$ 是可凸可导且 $|\ell''(e_{u,i})|\leq c$. 当给定一个泛化误差 $\varepsilon_g(\varepsilon_g<1)$和整数 $d(d<n)$,那么存在一个常量 C,如果能使

$$n>C\max\left(\frac{\|\boldsymbol{p}_0\|^2\log(1/\alpha)}{\varepsilon_g^2},\frac{c\|\boldsymbol{p}_0\|^2}{\varepsilon_g\varepsilon},\frac{d\log(d/\alpha)\|\boldsymbol{p}_0\|}{\varepsilon_g\varepsilon}\right) \tag{35}$$

成立,那么算法 1 中经过目标函数加扰求解的 $\boldsymbol{p}_u^{\mathrm{priv}}$ 将满足

$$\Pr[L(\boldsymbol{p}_u^{\mathrm{priv}})\leq L(\boldsymbol{p}_0)+\varepsilon_g]\geq 1-\alpha \tag{36}$$

证明　因为损失函数 $\ell(e_{u,i})$是可凸可导,那么对

任意两个误差有

$$|\ell'(e_{u,i}) - \ell'(e_{u,i})| \leq c(e_{u,i} - e'_{u,i}) \tag{37}$$

其中 $c=2$(算法 1 的第 4 点说明). 因此由文献[23]的定义 3 可知,$\ell'(e_{u,i})$ 是 2-Lipschitz[24],再由其中的定理 4 可得结论. 该推论说明了当样本个数大于某一取值时,经目标函数加扰策略求解的结果与无差分隐私处理的结果相差的量化程度.

4 实验结果及分析

4.1 实验数据和评估指标

4.1.1 实验数据

本文实验选用了两个公开数据集:FilmTrus① 和 Epinions②. 它们均含有用户-项目评分和用户间的信任关系,统计属性分别如表 1 所示.

表 1 两个数据集的统计属性

	特征	FilmTrust	Epinions
评分信息	用户数	1508	40163
	项目数	2071	139738
	评分数	35497	664824
	稀疏度	1.14%	0.051%
信任信息	信任者数	609	33960
	被信任者数	732	49288
	信任数	1853	487183
	密度	0.42%	0.029%

4.1.2 实验评估指标

本文实验采用 10 - 折交叉验证训练和预测(训练集:验证集 =9:1),结果取 10 次的平均值. 为评估新算法的预测准确率,选用绝对值误差(Mean Absolute Error,MAE)和根均方误差(Root Mean Squared Error,RMSE)来评估预测评分 $\tilde{r}_{u,i}$. MAE 和 RMSE 越小意味着预测的越准确,N 为有效评分的总数目,二者计算方法分别为

$$\mathrm{MAE} = \frac{\sum_{u,i} |r_{u,i} - \tilde{r}_{u,i}|}{N} \tag{38}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{u,i} (r_{u,i} - \tilde{r}_{u,i})^2}{N}} \tag{39}$$

4.2 实验结果及分析

4.2.1 实验参数设定

对于实验中涉及的参数,本文沿用 Guo 等在文献[7]中设定的参数值. 具体为:两个数据集的隐含特征矩阵的特征个数均为 $d=10$;正则化参数 $\lambda=0.6$,$\lambda_t=0.1$;迭代次数 $h=20$. 除此之外,针对 FilmTrust 数据集,$\alpha=0.6$,Epinions 数据集,$\alpha=0.4$. 关于权重 α 取值对预测结果的影响将在 4.2.3 小节中给出相应的调节实验.

4.2.2 实验结果及分析

实验分别从所有(ALL)用户和仅考虑冷启动(Cold-Start)用户两个角度来进行对比. 图 1 ~ 图 4 给出了本文算法 DPTrustSVD 与两个相关算法从 All 角度得到的 MAE 和 RMSE 结果比较. 其中,DPSVD 代表对无社会化关系的 SVD + + 做目标函数加扰的算法. 图 1 和图 2 是 FilmTrust 数据集上的结果,图 3 和图 4 是 Epinions 数据集上的结果.

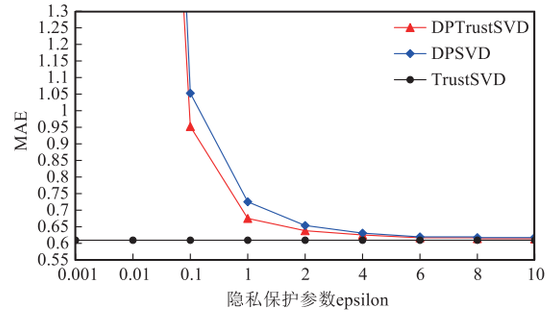

图1 FilmTrust数据集上MAE比较(ALL角度)

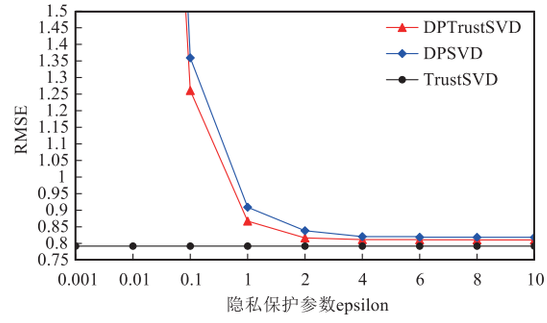

图2 FilmTrust数据集上RMSE比较(ALL角度)

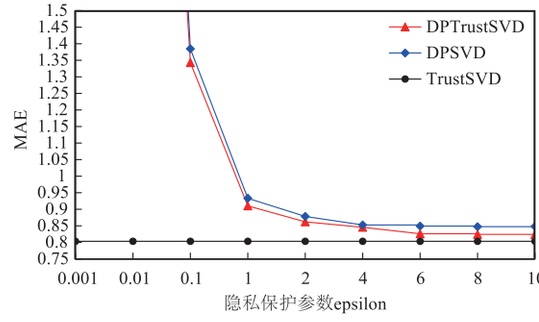

图3 Epinions数据集上MAE比较(ALL角度)

从图 1 ~ 图 4 可以看出,在两个数据集上,本文提出的差分隐私保护新策略 DPTrustSVD 的 MAE 和 RMSE 结果均可接受. 特别是当隐私保护参数 $\varepsilon \in [1, 10]$ 时,DPTrustSVD 更接近 TrustSVD 的结果,这是因为 ε 越大,添加的噪声会越小,那么与不做扰动的结果就越接近. 也可观察到,无论在哪个数据集,$\varepsilon < 0.1$ 时,

① www. librec. net/datasets. html

② trustlet. org/wiki/Epinions_datasets

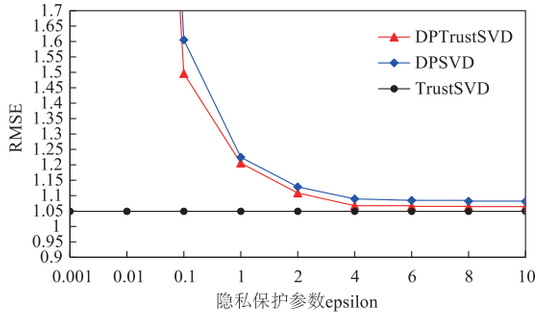

图4　Epinions数据集上RMSE比较（ALL角度）

MAE 和 RMSE 结果开始变差,由于添加的噪声过多,虽然隐私保护能力增强,但算法的预测准确率大大降低. 与未考虑社会化关系的 SVD + + 的差分隐私保护算法 DPSVD 相比,本文的 DPTrustSVD 更有优势,特别是在 Epinions 数据集上优势更明显. 其主要原因是 TrustSVD 算法自身优于 SVD + + ,那么在做了相关差分隐私保护后 TrustSVD 的隐私保护版本仍占优势. 总体来说,三个算法在 FilmTrust 数据集上结果均比 Epinions 数据集上的结果优秀,这是因为后者的评分数据和信任关系数据较前者稀疏,那么必然影响预测准确率,即使做了差分隐私保护后,该特性也没有改变.

TrustSVD 协同过滤算法对于冷启动问题做了较大改善[6,7],本文的 DPTrustSVD 算法也不会因为引入了差分隐私保护而明显破坏算法对冷启动问题的改善程度. 图 5 ~ 图 8 给出了从冷启动角度,本文的 DPTrustSVD 与无隐私保护的 TrustSVD 和无社会化关系的 DPSVD 在两个数据集上的预测准确率的比较情况.

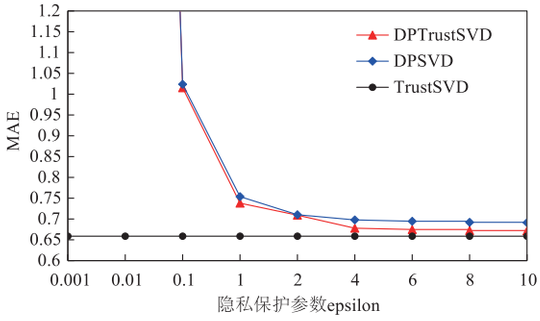

图5　FilmTrust数据集上MAE比较（Cold-Start角度）

从图 5 ~ 图 8 可以看出,本文提出的差分隐私保护算法 DPTrustSVD 仍可以较好地缓解冷启动问题,特别是当 $\varepsilon \in [2,10]$ 时,隐私保护力度和预测准确率达到了比较合理的折中,即一方面实现了对原始数据的差分隐私保护,另一方面算法的预测准确率更接近未做隐私保护的 TrustSVD 算法. 对于 SVD + + 算法来说,由于自身未考虑信任关系而不擅长改善冷启动问题,当做了差分隐私保护后,预测准确率在两个数据集上的表现均弱于 DPTrustSVD.

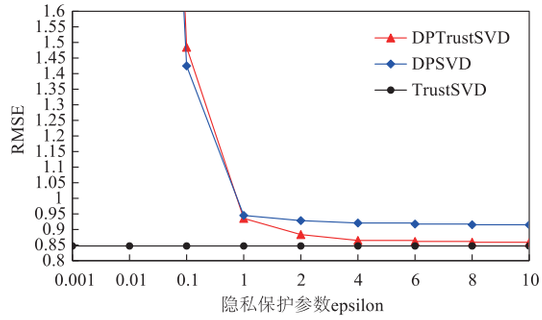

图6　FilmTrust数据集上RMSE比较（Cold-Start角度）

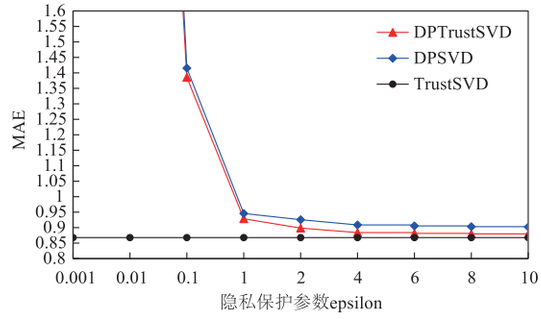

图7　Epinions数据集上MAE比较（Cold-Start角度）

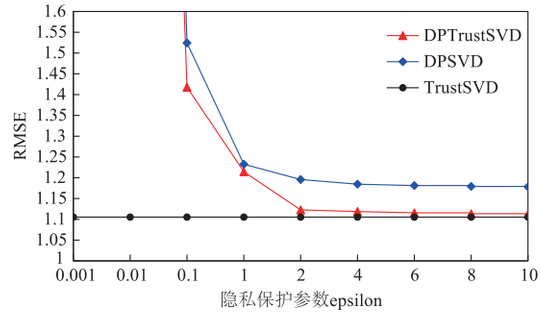

图8　Epinions数据集上RMSE比较（Cold-Start角度）

4.2.3　权重 α 对预测准确率的影响

TrustSVD 是通过权重 α 来控制显、隐式信任对预测评分的贡献程度. 本文新算法 DPTrustSVD 仍然保留 α 的作用,但对于预测准确率,还需考虑差分隐私保护参数 ε 的影响. 表 2 和表 3 分别给出了在两个数据集上,当固定差分隐私保护参数 ε,观察 α 在区间 $[0,1]$ 并按照 0.1 的步长取值对预测准确率的影响力. 考虑到 MAE 比 RMSE 更容易看出预测结果的变化,在此仅给出 MAE 的结果. 表 2 和表 3 中其他参数设置情况:$\lambda = 0.6$,$\lambda_t = 0.1$,$d = 10$. 两表中粗体行表示最优的 MAE. 得到结论:FilmTrust 数据集上,针对不同的差分隐私保护参数 ε,当 $\alpha = 0.6$ 时,DPTrustSVD 获得最优的预测准确率;Epinions 数据集上,则是 $\alpha = 0.4$ 时表现最优.

表 2 不同 ε 下 α 对预测准确率的影响(FilmTrust)

	$\varepsilon=0.1$	$\varepsilon=1$	$\varepsilon=2$	$\varepsilon=4$
$\alpha=0$	0.9586	0.6786	0.6416	0.6288
$\alpha=0.1$	0.9537	0.6765	0.6395	0.6267
$\alpha=0.2$	0.9537	0.6765	0.6396	0.6268
$\alpha=0.3$	0.9537	0.6766	0.6396	0.6268
$\alpha=0.4$	0.953	0.6757	0.6387	0.6259
$\alpha=0.5$	0.9528	0.6756	0.6386	0.6257
$\alpha=0.6$	**0.9523**	**0.6753**	**0.6382**	**0.6253**
$\alpha=0.7$	0.9529	0.6755	0.6385	0.6257
$\alpha=0.8$	0.954	0.6767	0.6398	0.6269
$\alpha=0.9$	0.9541	0.6769	0.6399	0.6272
$\alpha=1$	0.955	0.6778	0.6408	0.6279

表 3 不同 ε 下 α 对预测准确率的影响(Epinions)

	$\varepsilon=0.1$	$\varepsilon=1$	$\varepsilon=2$	$\varepsilon=4$
$\alpha=0$	1.3461	0.9135	0.8652	0.8487
$\alpha=0.1$	1.3447	0.912	0.8638	0.8471
$\alpha=0.2$	1.3443	0.9117	0.8634	0.8469
$\alpha=0.3$	1.3434	0.9106	0.8624	0.8459
$\alpha=0.4$	**1.3431**	**0.9103**	**0.8621**	**0.8456**
$\alpha=0.5$	1.3439	0.911	0.8629	0.8463
$\alpha=0.6$	1.3435	0.9108	0.8625	0.846
$\alpha=0.7$	1.3443	0.9115	0.8632	0.8467
$\alpha=0.8$	1.3437	0.9111	0.8626	0.8463
$\alpha=0.9$	1.3452	0.9126	0.8642	0.8478
$\alpha=1$	1.3513	0.9185	0.8703	0.8539

5 结束语

社会化协同过滤算法是充分利用用户间的信任关系来改善协同过滤面临的数据稀疏和冷启动问题.然而无论是用户的评分数据还是信任关系数据,都存在隐私泄露的风险.加强隐私保护是推荐系统改善服务质量的目标之一[25].针对目前流行的社会化关系的协同过滤算法 TrustSVD,本文提出将定义严格、可证明的差分隐私保护引入其中.针对提出的新策略 DPTrustSVD,文中从理论上分析了其隐私保护的性能,并在真实数据集上检验了其在社会化协同过滤应用中的预测表现,结果表明,本文提出的新策略与无隐私保护的 TrustSVD 具有相近的预测准确率,与做类似保护的 DPSVD 算法相比获得了更优的预测准确率.

推荐系统乃至数据挖掘领域需健康的发展,离不开隐私保护问题的深入研究.下一步工作:

(1)研究算法中其他参数的选择问题,特别是更深入的研究隐私保护参数 ε 的选择,使得隐私保护力度和推荐准确率之间得到更优的折中.

(2)显式信任关系用[0,1]权重来更确切地表示信任的程度,而不仅仅采用二值信任,即将新算法用于其他非二值信任的数据集.

参考文献

[1] ADOMAVICIUS G,TUZHILIN A. Toward the next generation of recommender systems:A survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering,2005,17(6):734-749.

[2] YANG Bo,LEI Yu,LIU Da-you et al. Social collaborative filtering by trust[A]. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI 13)[C]. Beijing,China:IJCAI,2013. 2747-2753.

[3] 俞春花,刘学军,李斌,等. 基于上下文相似度和社会网络的移动服务推荐方法[J]. 电子学报,2017,45(6):1530-1536.
YU Chun-hua,LIU Xue-jun,LI Bin,et al. Mobile service recommendation based on context similarity and social network[J]. Acta Electronica Sinica,2017,45(6):1530-1536. (in Chinese)

[4] JAMALI M,ESTER M. A transitivity aware matrix factorization model for recommendation in social networks[A]. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI'11)[C]. Barcelona,Catalonia,Spain:IJCAI,2011. 2644-2649.

[5] YAO Wei-long,HE Jing,HUANG Guang-yan,et al. Modeling dual role preferences for trust-aware recommendation[A]. Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR'14)[C]. New York,US:ACM,2014. 975-978.

[6] GUO Gui-bing,ZHANG Jie,NEIL Y S. TrustSVD:Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings[A]. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)[C]. Austin,Texas:AAAI,2015. 123-129.

[7] GUO Gui-bing,ZHANG Jie,NEIL Yorke-Smith. A novel recommendation model regularized with user trust and item ratings[J]. IEEE Transactions on Knowledge and Data Engineering,2016,28(7):1607-1620.

[8] 曹珍富,董晓蕾,周俊,等. 大数据安全与隐私保护研究进展[J]. 计算机研究与发展,2016,53(10):2137-2151.
CAO Zhen-fu,DONG Xiao-lei,ZHOU Jun,et al. Research advances on big data security and privacy preserving[J]. Journal of Computer Research and Development,2016,53(10):2137-2151. (in Chinese)

[9] DWORK C. Differential privacy[A]. Proceedings of the 33rd International Colloquium on Automata,Languages and Programming,Venice (ICALP'06)[C]. Italy:ICALP,

2006. 1 – 12.
[10] 张啸剑,孟小峰. 面向数据发布和分析的差分隐私保护研究[J]. 计算机学报,2014,37(4):927 – 949.
ZHANG Xiao-jian, MENG Xiao-feng. Differential privacy in data publication and analysis[J]. Chinese Journal of Computers, 2014, 37(4): 927 – 949. (in Chinese)
[11] MCSHERRY F, MIRONOV I. Differential private recommender system: Building privacy into Netflix prize contenders[A]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)[C]. New York: ACM, 2009. 627 – 636.
[12] HARDT M, ROTH A. Beating randomized response on incoherent matrices[A]. Proceeding of the Annual Acm Symposium on Theory of Computing (STOC'12)[C]. New York: ACM, 2012. 1255 – 1268.
[13] FRIEDMAN A, BERLIOZ A, KAAFAR S, et al. A differential privacy framework for matrix factorization recommender systems[J]. User Modeling and User-Adapted Interaction, 2016, 26(5): 1 – 34.
[14] HUA Jing-yu, XIA Chang, ZHONG Sheng. Differentially private matrix factorization[A]. Proceeding of the 24th International Conference on Artificial Intelligence(IJCAI'15)[C]. New York: ACM, 2015. 1763 – 1770.
[15] YAN Shen, PAN Shi-ran, ZHU Wen-tao, et al. DynaEgo: Privacy-preserving collaborative filtering recommender system based on social-aware differential privacy[A]. International Conference on Information and Communications Security (ICICS'16)[C]. US: ICICS, 2016. 347 – 357.
[16] 范利云,左万利,王英,等 一种基于差分隐私和时序的推荐系统模型研究[J]. 电子学报,2017,45(9):2057 – 2064.
FAN Li-yun, ZUO Wan-li, WANG ying, et al. Research on recommender system model based on differential privacy and time series[J]. Acta Electronica Sinica, 2017, 45(9): 2057 – 2064. (in Chinese)
[17] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30 – 37.
[18] DWORK C. Differential privacy: A survey of results[A]. Proceeding of Theory and Applications of Models of Computation (TAMC'08)[C]. US: TAMC, 2008. 1 – 9.
[19] MCSHERRY F. Privacy integrated queries: An extensible platform for privacy-preserving data analysis[A]. Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'09)[C]. Providence, Rhode Island: ACM, 2009. 19 – 30.
[20] NISSIM K, RASKHODNIKOVA S, SMITH A. Smooth sensitivity and sampling in private data analysis[A]. Proceedings of the 39th Annual ACM Symposium on Theroy of Computing (STOC'07)[C]. US: ACM, 2007. 75 – 84.
[21] DWORK C, MCSHERRY F, NISSIM K, SMITH A. Calibrating noise to sensitivity in private data analysis[A]. Proceedings of the 3th Theory of Cryptography Conference (TCC'06)[C]. US, 2006. 363 – 385.
[22] 熊平,朱天清,王晓峰. 差分隐私保护及其应用[J]. 计算机学报,2014,37(1):101 – 122.
XIONG Ping, ZHU Tian-qing, WANG Xiao-feng. A survey on differential privacy and applications[J]. Chinese Journal of Computer, 2014, 37(1): 101 – 122. (in Chinese)
[23] CHAUDHURI K, MONTELEONI C, SARWATE A. Differentially private empirical risk minimization[J]. Journal of Machine Learning Research, 2011, 12(2): 1069 – 1109.
[24] SHALEV-SHWARTZ S, SREBRO N. SVM optimization: Inverse dependence on training set size[A]. Proceedings of the 25th International Conference on Machine Learning (ICML'08)[C]. US: ICML, 2008. 928 – 935.
[25] SU Kai, MA Liang-li, XIAO Bin, et al. Web service QoS prediction by neighbor information combined non-negative matrix factorization[J]. Journal of Intelligent & Fuzzy Systems, 2016, 30(6): 3593 – 3604.

作者简介

鲜征征　女. 1977 年 8 月出生于四川省阆中市. 博士,现为广东金融学院讲师,CCF 会员. 主要研究方向为数据挖掘、隐私保护等.
E-mail: xianzhengzheng@126.com

李启良　男. 1990 年 5 月出生于广东省云浮市. 硕士,现为华为技术有限公司集成服务部的软件工程师. 主要研究方向为数据挖掘和隐私保护.
E-mail: liqiliang90@163.com

黄晓宇(通讯作者)　男. 1977 年 10 月生. 博士,副教授. 主要研究兴趣为机器学习、隐私保护等.
E-mail: echxy@scut.edu.cn