差分隐私保护在推荐系统中的应用研究*

鲜征征[1,2]，李启良[2]

（1. 广东金融学院 计算机科学与技术系，广州 510521；2. 中山大学 信息科学与技术学院，广州 510006）

摘　要：推荐系统已经成为 Internet 商家给用户提供个性化服务的高级商务智能平台之一。然而，用于研究推荐系统的数据信息里往往存在能够被攻击者直接或者间接获取的个人隐私。近年来受到极大关注的差分隐私保护是一种非常严格的、可证明的隐私保护模型。针对目前流行的协同过滤算法之一的矩阵分解进行了研究，提出了采用差分隐私保护技术对原始输入数据进行预处理和扰动处理的新方法。最后通过在真实数据集上进行相关实验验证，结果表明提出的带差分隐私保护的矩阵分解算法达到了预期：既能保护用于做推荐研究的原始数据集的隐私，又没有严重影响推荐的准确率。

关键词：推荐系统；个人隐私保护；差分隐私；矩阵分解

中图分类号：TP309　　**文献标志码**：A　　**文章编号**：1001-3695(2016)05-0001-05

doi:10.3969/j.issn.1001-3695.2016.05.059

Research on application of differential privacy in recommender system

Xian Zhengzheng[1,2], Li Qiliang[2]

(1. *Dept. of Computer Science & Technology, Guangdong University of Finance, Guangzhou* 510521, *China*; 2. *School of Information Science & Technology, Sun Yat-sen University, Guangzhou* 510006, *China*)

Abstract: Recommender system has become one of the most advanced business intelligence platform that Internet businesses provide personalized service to the user. However, the massive data information used to recommender system research often exist some personal privacy information that can be got by an attacker directly or indirectly. As a new privacy notion, differential privacy has shown in popularity recently due to its strict and provable guarantee. Matrix factorization is one of the popular collaborative filtering methods. This paper addressesd the problem of privacy preserving matrix factorization by utilizing differential privacy. The new approach proposed preprocesses and obfuscates the raw input data before using matrix factorization. Finally, some experiment results show that the new proposed approach had achieves the expected goals. As a result, it not only can protect the privacy of the original input data set for recommender systemd, but also don't seriously affect the recommendation accuracy.

Key words: recommender system; personal privacy preserving; differential privacy; matrix factorization

0　引言

当今，互联网和电子商务已经成为人们日常生活、工作不可缺少的重要部分。特别是近十年来，大量的研究和商业活动使得推荐系统得到了快速的发展，众多互联网商家可以提供给用户个性化的在线推荐服务，这些服务甚至注入到了用户的移动终端设备，如电影、音乐、书籍、酒店等推荐服务，从而让用户更加热衷并依赖于互联网。

推荐系统的基本原理是：首先对用户的历史行为数据，如浏览历史、购买记录、商品评论和商品打分等情况进行搜集；然后根据用户这些使用记录挖掘出用户偏好信息，进而分析用户的喜好并构建用户兴趣模型；最后当用户浏览信息时，它可面向不同用户主动推荐其可能感兴趣的信息，从而提供个性化的服务，以满足不同用户的不同需求。由于推荐系统是依赖于用户的个人信息，所以使用这样的系统也就存在一个隐私泄露的风险。例如，某电影推荐系统，根据大量注册用户对已看电影进行的评分，分析出用户的偏好并构建兴趣模型，最终能为新用户或者其他用户推荐其可能喜欢的电影。有人提出，仅仅一个评分能有什么隐私可泄？在此，假设用户 A 和 B 是朋友关系，B 平时较了解 A 的兴趣爱好，那么 B 可以模仿 A 的各种可能的特征去获得推荐的电影，由此 B 就可以得知 A 曾经或者喜欢看的影片，而 A 并不希望其他人知道他看过的某些影片或者影片类型，这样该推荐系统就潜在地向用户 B 暴露了 A 的个人隐私。

近年来人们也越来越重视隐私保护问题，使得更多的数据拥有者不愿意为数据分析者提供自己的数据，或者要么从自己的数据中去除掉一些信息，要么提供一些虚假信息，导致不能为例如推荐系统等数据挖掘研究提供真实可靠的数据源，这样必然严重影响挖掘的结果。因此，如何能够确保个人隐私安全，消除用户疑虑，鼓励用户提供真实可靠的数据源，从而确保挖掘出有效的知识和规则，已经成为数据挖掘领域一个亟待解决的问题。同时，推荐系统中的隐私保护问题也成了目前数据挖掘领域研究的热点之一。协同过滤是典型的推荐方法之一，它依赖于用户的个性化信息。关于个人隐私，仅仅简单的匿名处理并不能得到真正意义上的保护，因为攻击者可以通过相关

收稿日期：2015-01-15；**修回日期**：2015-03-05　**基金项目**：广东省高校创新强校工程自主创新能力提升类培育项目

作者简介：鲜征征(1977-)，女，四川阆中人，讲师，博士研究生，主要研究方向为数据挖掘中的隐私保护、机器学习(xianzhengzheng@126.com)；李启良(1990-)，男，硕士研究生，主要研究方向为数据挖掘、机器学习.

的背景知识进行攻击。2006 年,由 Dwork 等人提出了定义极为严格的、与背景知识无关的新型隐私保护模型—差分隐私保护(differential privacy,DP)。

本文旨在将差分隐私保护应用于推荐系统中(以协同过滤的矩阵分解为例),并通过在真实数据集上进行实验验证,提出的带差分隐私保护的推荐算法不但可以有效地保护用户的个人信息,同时也能获得良好的推荐准确率。

1　相关工作

数据挖掘领域始终是以数据为核心,自 2006 年由 Dwork 等人提出了差分隐私保护模型后,数据挖掘与机器学习领域出现了许多基于该保护模型的数据分析研究工作。例如,文献[1,2]将差分隐私引入到频繁模式挖掘,使得自顶向下的树划分过程中不但保护了原始数据的敏感信息,而且仍然能够支持有效的 top-k 频繁模式挖掘。分类技术在数据预测分析中起着关键作用,结合差分隐私保护技术与决策树分类的代表是 SuLQ-based ID3 方法[3]、DiffP-C4.5 方法[4] 以及 DiffGen 方法[5]。这三种方法在生成分类器时类似于 ID3 算法[6],同时考虑了决策树各个节点上分割属性的选择问题,且三种方法均采用了信息增益(information gain) 来选择分割属性,并递归地构建决策树。文献[7,8]都结合差分隐私保护技术分别提出了支持向量机分类方法 PrivateSVM 和 ObjectiveSVM,前者是对输出的分类向量进行差分隐私保护处理,后者是对目标函数进行差分隐私保护处理。实验结果表明后者的分类准确率高于前者。文献[9]结合采样与聚类技术提出了一种满足差分隐私的 K-均值聚簇中心发布方法 PK-Means,该方法给出了聚类敏感性的度量方法以及聚类误差的下界。

差分隐私应用于推荐系统方面的研究和成果有限。2009 年 McSherry 等人[10] 将差分隐私用在了协同过滤推荐系统中,它们是对 item-to-item 协方差矩阵进行差分隐私处理,且将推荐系统分为学习阶段和预测阶段(即推荐),阐述了在没有严重损失推荐的准确率情况下实施差分隐私保护是可行的,但它们没有考虑到潜在因素模型(latent factor model)。因此本文从考虑潜在因素角度,将差分隐私保护技术引入到矩阵分解模型中,最终达到既能对用户个人隐私数据进行差分隐私保护,又能在一定程度上保证推荐的准确率。

2　理论基础

2.1　差分隐私保护模型

差分隐私与传统的隐私保护有着本质上的区别,它定义了一个极为严格的攻击模型,并对隐私泄露风险给出了严谨、定量化的表示和证明。差分隐私保护在大大降低隐私泄露风险的同时,也极大地保证了数据的可用性。该方法的最大优点是:虽然基于数据失真技术,但所加入的噪声量与数据集大小无关,因此对于大型数据集,仅通过添加极少量的噪声,就能得到高级别的隐私保护[11]。此外,差分隐私保护模型是不关心攻击者拥有多少背景知识,通过向查询或者分析结果中添加噪声以达到隐私保护效果。迄今,虽然出现了多种基于 k-匿名和划分隐私保护框架的保护方法,但是差分隐私保护模型被公认为最严格和强健的保护模型。接下来详细给出差分隐私的定义及其相关属性。

2.1.1　差分隐私的定义

差分隐私保护模型的核心思想:一方面,可以确保在某一数据集中插入或者删除一条记录的操作不会影响任何计算(如计数查询)的输出结果;另一方面,该模型不关心攻击者所具有的背景知识,即使攻击者已经掌握除某一条记录之外的所有记录的敏感信息,该记录的敏感信息也无法被披露[12]。其形式化定义如下:

定义 1[13,14]　差分隐私。给定两个至多相差一条记录的数据集 D 和 D'(即 $|D\Delta D'|\leqslant 1$),对于一个设定的随机算法 A,其取值范围为 Range(A),若算法 A 在数据集 D 和 D' 上的任意输出结果 $S(S\in \text{Range}(A))$ 满足式(1),则称算法 A 满足 ε-差分隐私。

$$\Pr[A(D)\in S]\leqslant e^{\varepsilon}\times \Pr[A(D')\in S] \tag{1}$$

其中,Pr[.]表示隐私被披露的概率,且由算法 A 的随机性所控制(算法 A 的随机性与攻击者具有的背景知识无关);ε 是隐私保护参数,用来表示隐私保护的程度,ε 越小意味着隐私保护程度越高。

2.1.2　实现差分隐私的方法

实现差分隐私保护的关键技术是添加噪声,常用的噪声添加机制有拉普拉斯机制和指数机制。通过添加噪声来满足差分隐私的算法是与函数的敏感度和隐私保护参数 ε 有着直接的关联。

1)函数的敏感度

函数的敏感度分为全局敏感度和局部敏感度。全局敏感度是指函数从两个只有一条记录不同的数据集中得到的输出的最大差别。其形式化定义如下。

定义 1[13,14]　函数的全局敏感度。对于任意一个函数 f:$D\rightarrow R^d$,d 表示函数 f 的维度,则函数 f 的 L_k-全局敏感度 $S_k(f)$ 为

$$S_k(f)=\max_{D,D'}\|f(D)-f(D')\|_k \tag{2}$$

其中,数据集 D 和 D' 至多相差一条记录;$\|.\|_k$ 表示 L_k-范数。

2)拉普拉斯机制

提出差分隐私保护模型的 Dwork 等人在文献[15]中给出了拉普拉斯机制可以取得差分隐私保护效果,即通过添加拉普拉斯随机噪声来实现差分隐私保护。拉普拉斯分布的概率密度函数为

$$f(x|\mu,b)=\frac{1}{2b}\exp(-|x-\mu|) \tag{3}$$

其中,μ 和 b 分别为变量 x 的期望和尺度参数。为了方便获取噪声,设 $\mu=0$,则拉普拉斯分布可以看做是标准差为 $\sqrt{2}b$ 的对称指数分布。在实现差分隐私时,拉普拉斯随机噪声的计算方法为

$$\text{laplapce}(\Delta f/\varepsilon) \tag{4}$$

其中,Δf 是函数 f 的全局敏感度(由式(2)求得),ε 为差分隐私保护参数。从式(4)可见,加入的噪声与 Δf 成正比,与 ε 成反比。

3)差分隐私的特性

通常,一个复杂的隐私保护问题通常需要多次应用差分隐私保护技术,在这种情况下,为了保证整个过程的隐私保护水平是控制在给定的隐私保护预算 ε 之内,则须要用到差分隐私保护技术本身具有的两种重要的组合特性,即列组合特性或者并行组合特性[16]。

性质 1　序列组合性。设多个随机算法 $A_1,A_2,\cdots,A_n$,每

个随机算法对应的隐私保护预算分别为 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$，且各自满足 ε_i-差分隐私，则这些算法构成的组合算法 A 在同一数据集 D 上满足 $\sum_{i=1}^{n}\varepsilon_i$-差分隐私。

性质 2　并行组合性。设多个随机算法 $A_1,A_2,\cdots,A_n$，每个随机算法对应的隐私保护预算分别为 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$，且各自满足 ε_i-差分隐私，则这些算法构成的组合算法 A 对于不相交的数据集 $D_1,D_2,\cdots,D_n$ 满足 $\max\ \varepsilon_i$-差分隐私。

2.2　矩阵分解

要做推荐系统，最基本的一个数据就是用户—项目评分矩阵。通常，这个评分矩阵还存在着一些缺失的评分，本文的目标是尽可能准确地预测出这些缺失评分。矩阵分解（decomposition factorization，PF）是解决这类问题的有效方法之一，其原理是将矩阵拆解为数个矩阵的乘积。

2.2.1　目标函数

设定一个有 n 个用户 m 个项目的评分矩阵 $\boldsymbol{R}_{n\times m}$，矩阵中每一个元素 r_{ui} 表示用户 u 对项目 i 的评分。经过矩阵分解处理后，原始评分矩阵 $\boldsymbol{R}_{n\times m}$ 被分解成两个隐含矩阵：用户因素矩阵 $\boldsymbol{P}_{n\times d}$ 和项目因素矩阵 $\boldsymbol{Q}_{d\times m}$，其中 d 通常取低维。设 $\tilde{\boldsymbol{R}}_{n\times m}=\boldsymbol{P}_{n\times d}\times\boldsymbol{Q}_{d\times m}$，新矩阵 $\tilde{\boldsymbol{R}}_{n\times m}$ 即为原始矩阵 $\bar{\boldsymbol{R}}_{n\times m}$ 的近似矩阵，且原始矩阵中的每一个评分 r_{ui} 也有着相应的近似值，即 $\tilde{r}_{ui}=p_u\cdot q_i^T$，在本文也称其为预测值。为了获得最优的 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 矩阵，本文采用求最小化正则化的平方误差作为目标函数，即式(5)：

$$\min_{P,Q}\sum_{r_{ui}\in R}\left[e_{ui}^2+\lambda(\|p_u\|_2+\|q_i\|_2)\right] \tag{5}$$

其中，e_{ui} 表示真实值与预测值之间的误差，即 $e_{ui}=r_{ui}-\tilde{r}_{ui}$；$\|\cdot\|_2$ 表示取二范数；λ 为正则化参数，用来防止过拟合。

针对式(5)，本文采用两种优化算法即梯度下降法（stochastic gradient descent，SGD）、交替最小二乘法（alternating least square，ALS）。下面分别阐述这两种优化算法的核心思想。

2.2.2　优化算法

1）梯度下降法（SGD）

在 SGD 中，是按照式(6)中的方式不断迭代地更新隐含矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 中的每一个元素来求解式(5)的。

$$\begin{aligned}p_u&\leftarrow p_u+\gamma(e_{ui}q_i-\lambda p_u)\\ q_i&\leftarrow q_i+\gamma(e_{ui}p_u-\lambda q_i)\end{aligned} \tag{6}$$

其中，p_u 为矩阵 $\boldsymbol{P}$ 中的任一元素；q_i 为矩阵 $\boldsymbol{Q}$ 中的任一元素；λ 为式(5)中的正则化参数；γ 是一个常量，用来决定最小化误差梯度下降的速率，通常称为学习率。

2）交替最小二乘法（ALS）

在 ALS 中，优化问题是通过不断迭代来实现的。在每一迭代中，首先固定一个隐含矩阵，例如矩阵 $\boldsymbol{P}$，这样就变成一个凸优化问题来求解另一矩阵 $\boldsymbol{Q}$；然后固定另一个隐含矩阵，如矩阵 $\boldsymbol{Q}$，同理可以求出另一矩阵 $\boldsymbol{P}$。多次迭代后直到收敛即求得式(5)中最优的 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$。

3　基于差分隐私的矩阵分解

在采用矩阵分解的方法来实现推荐的问题，使用的数据集是原始评分矩阵，为了保护原始数据中个人隐私，本文提出将差分隐私保护技术引入到矩阵分解中。根据矩阵分解的原理，可以看出矩阵分解过程可以分为四个步骤：a）数据输入（原始评分矩阵）；b）具体的分解处理（SGD 或 ALS）；c）输出分解后用户矩阵和项目矩阵；d）评分预测（即作推荐）。本章将详细阐述采用对输入数据进行差分隐私保护的算法。

算法原理简单地说，就是对原始输入数据（原始评分矩阵）合理添加拉普拉斯噪声。但是为了增强隐私保护力度又不失数据的有效性，本文提出新的预处理算法，并对原始评分矩阵作相关预处理，然后将处理后的数据作为矩阵分解的输入数据。

3.1　带差分隐私保护的预处理

为了得到更好的推荐准确率，通常会对输入数据采取一些有效的预处理。本文采用的预处理思想源于文献[10]，但与之不同的是，本文考虑到了用户的平均分，并将其并入到了评分预测中。该预处理包含以下步骤。

1）计算带差分隐私保护的项目平均评分

无须做任何隐私保护，对于原始评分矩阵，可以由式(7)求得每一个项目的平均分。

$$\mathrm{IAvg}(j)=\frac{\sum_{R_j}r_{ui}}{|R_j|} \tag{7}$$

其中，j 表示第 j 个项目，$|R_j|$ 表示参与第 j 个项目评分的用户数。

在这一步骤中，本文将用户平均评分做了差分隐私保护处理，具体计算方法如算法 1 描述。在算法 1 中，实现差分隐私保护是通过添加拉普拉斯噪声，用户评分是属于 L_1-敏感度，因此评分的敏感度 $\Delta r=r_{\max}-r_{\min}$；$\beta_i$ 是稳定参数，目的为了限制添加的噪声对于只获得少量评分项目的影响。

算法 1　计算带差分隐私保护的项目平均分

输入：$\boldsymbol{R}_{n\times m}=\{r_{ui}\}$— 原始评分矩阵，$\beta_i$ 为稳定参数，ε_1 为全局平均分隐私预算参数，ε_2 为项目平均分隐私预算参数。

输出：每个项目的平均分 $\mathrm{IAvg}(j)$。

a）计算做了差分隐私保护处理的全局平均分 $GAvg$：

$$\mathrm{GAvg}=\frac{(\sum_{R}r_{ui})+\mathrm{Laplace}(\Delta r/\varepsilon_1)}{|R|}$$

b）for $j=1$ to m do

设 $\boldsymbol{P}_j=\{r_{ui}\in\boldsymbol{R}\mid i=j\}$

计算 $\mathrm{IAvg}(j)$：$\mathrm{IAvg}(j)=\dfrac{(\sum_{R_j}r_{ui})+\beta_i\times\mathrm{GAvg}+\mathrm{Laplace}(\Delta r/\varepsilon_2)}{|R_j|+\beta_i}$

将 $\mathrm{UAvg}(v)$ 控制在 $[r_{\min},r_{\max}]$ 中。

end for

2）计算带差分隐私保护的用户平均评分

无须做任何隐私保护，对于原始评分矩阵，可以由式(8)求得每一个用户对参与的项目的平均评分。

$$\mathrm{UAvg}(v)=\frac{\sum_{R_v}r_{ui}}{|R_v|} \tag{8}$$

其中，v 表示第 v 个用户，$|Rv|$ 表示用户 v 参与评分的项目数。

在这一步骤中，本文采用类似算法 1 的方法来计算带差分隐私保护的用户平均分。具体计算方法如算法 2 所描述。算法 2 中，首先利用算法 1 计算出来的项目平均分来对原始评分矩阵 $\boldsymbol{R}$ 作初始化，这样做的目的是为了减少用户对个别项目过高或过低的评分给实验结果带来的影响。

算法 2　计算带差分隐私保护的用户平均评分

输入：$\boldsymbol{P}_{n\times m}=\{r_{ui}\}$ 为原始评分矩阵，β_u 为稳定参数，ε_3 为算法 1 处理后的全局平均隐私预算参数，ε_4 为用户平均分隐

私预算参数。

输出:每个用户的平均评分 UAvg(v)。

a)设 $R'\{r_{u_i} - \mathrm{IAvg}(i) \mid r_{u_j} \in R\}$ 即得 $R'\{r'_{u_i}\}$

b)针对 a)中的新评分矩阵 R'来计算做了差分隐私保护处理的全局平均分 $GAvg'$:

$$\mathrm{GAvg}' = 00\frac{(\sum_{R'} r'_{ui}) + \mathrm{Laplace}(\Delta r/\varepsilon_3)}{|R'|}$$

c)for $v = 1$ to n do

设 $R_v = \{r'_{ui} \in R' \mid u = v\}$

计算 UAvg(v): $\mathrm{UAvg}(v) = \dfrac{(\sum_{R'} r'_{ui}) + \beta_i \times \mathrm{GAvg}' + \mathrm{Laplace}(\Delta r/\varepsilon_4)}{|R_v| + \beta_u}$

将 UAvg(v)控制在范围[-2,2]中。

end for

3)预测评分的最终处理

由于对输入的原始评分矩阵(作为训练数据集)做了预处理,所以为了保证最终的预测评分准确率更高,预测评分应该做一些恢复处理,具体由式(9)来获得。

$$\tilde{R} = \{\tilde{r}_{ui}\}, \text{其中 } \tilde{r}_{ui} = \mathrm{IAvg}(i) + \mathrm{UAvg}(u) + p_u q_i^T \tag{9}$$

其中,$\tilde{R}$表示预测评分矩阵,$\tilde{r}_{ui}$则表示 $\tilde{R}$中的用户 u 对项目 i 的预测评分。

3.2　输入扰动差分隐私保护

输入扰动差分隐私保护的基本原理就是直接对每一个评分添加相应的拉普拉斯噪声。具体过程见算法 3 的描述。

算法 3　输入扰动差分隐私保护

输入:$R_{n\times m} = \{r_{ui}\}$为采用算法 1 和算法 2 做了预处理的用户评分矩阵,k 为矩阵分解隐含特征矩阵的特征个数,λ 为正则化参数,ε_5 为隐私保护预算参数。输出:矩阵分解后的隐含因素矩阵 $P_{n\times k}$和 $Q_{k\times m}$。

a)对输入的用户评分矩阵 R 做差分隐私保护处理,得到新的评分矩阵 R'。

$$R' = \{r_{ui} + \mathrm{Laplace}(\Delta r/\varepsilon_5)\} \quad (r_{ui} \in R_{n\times m})$$

b)将新的评分矩阵 R'中的评分控制在范围[1,5](有效的评分)中。

c)进行矩阵分解,分别采用 SGD 和 ALS 来求解最优值:

d)Return $P_{n\times k}$和 $Q_{k\times m}$

3.3　新算法满足 ε-差分隐私保护

迄今,差分隐私保护跟其他匿名化技术相比是更为严格的隐私保护模型,这种严格体现在需要通过利用其各项属性来证明某算法满足差分隐私保护的定义(节定义 1)。实现差分隐私保护的关键技术是添加噪声,本文提出的算法 1 ~ 3 中均采用的是添加拉普拉斯噪声。本节给出算法 3 满足 ε-差分隐私的关键证明步骤。

a)算法 1 中的 a)对全局平均分做了拉普拉斯噪声处理(严格按照差分隐私实现机制添加),然后将该处理后的全局平均分用于 c)中求得每一个项目的平均分,而在求每一个项目的平均分时又再次做了拉普拉斯噪声处理。可见,两次加噪声的过程是串行的,根据 2.1.2 节的差分隐私保护序列组合特性,算法 1 满足($\varepsilon_1 + \varepsilon_2$)-差分隐私。

b)同理,可以证得算法 2 满足($\varepsilon_3 + \varepsilon_4$)-差分隐私。由于算法 2 中的数据来源于算法 1,因此,两个算法仍然满足序列组合特性。

c)算法 3 的 a)中,输入数据是严格按照差分隐私实现机制添加拉普拉斯噪声,所以算法 3 满足 ε_5-差分隐私。由于这一步中被加噪的数据来自算法 1、2 预处理后的数据,那么算法 3 实际上是满足($\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4 + \varepsilon_5$)-差分隐私。为了统一和简化隐私保护参数,本文从整体上给出一个隐私保护参数 ε,然后将这个总的 ε 合理地分割为五份(具体分割请见实验结果部分),因此,算法 3 就满足 ε-差分隐私。

4　实验结果及分析

针对本问题提出的新算法,本章将给出在真实数据集上进行的实验结果,以此验证新算法不但可以对原始数据集进行差分隐私保护,同时也能达到一定程度的推荐准确率。

4.1　实验数据和评估指标

4.1.1　实验数据

本文实验数据选用的 Movielens 数据集来自 http://grouplens.org/datasets/movielens/,Movielens 数据集包括 100 KB、1 MB 和 10 MB 数据集,本文选用了 1 MB 数据集。该数据集的统计属性如表 1 所示。

表 1　1 MB Movielens 数据集的统计属性

属性名	统计值
users(用户数)	6 040
movies(电影数)	3 952
density(密度)	4.19%
average rating(平均评分)	3.581 6
variance rating(评分的方差)	1.247 9
avg. rating per user(每个用户参与的平均评分)	165.6
avg. rating per item(每部电影的平均评分)	253

4.1.2　实验评估指标

为了更有效地评估推荐系统,本文实验采用 10-折交叉验证来训练和预测。在评估推荐准确率方面本文采用根均方误差(RMSE)指标来评估预测评分 $\tilde{r}_{ui}$,RMSE 越小意味着预测越准确。RMSE 的计算方法如公式(10)。

$$RMSE = \frac{\sqrt{\sum_{R}(r_{ui} - \tilde{r}_{ui})^2}}{|R|} \tag{10}$$

其中,$|R|$表示评分的个数。注意,这里评分是指有效的评分,不包括那些缺失的评分。

4.2　数据的预处理

为了能够达到本文研究的最终目标:在对原始数据集进行差分隐私保护的同时又不影响推荐效果,本文采用了算法 1 和 2 对原始评分数据集进行了一些预处理。

本文 3.3 节中提到了差分隐私保护参数 ε 的分配问题,经过多次实验,下面给出实验结果相对理想的分配方案,总体来说有 0.3ε 用于数据的预处理。最终对 ε 的分配如表 2 所示。

表 2　差分隐私保护参数 ε 的分配方案

隐私保护参数的作用	分配份额
计算全局评分隐私保护参数(算法 1、2 中的 ε_1 和 ε_3)	0.01ε
计算项目平均分隐私保护参数(算法 1 的 ε_2)	0.14ε
计算用户平均分隐私保护参数(算法 2 的 ε_4)	0.14ε
输入扰动差分隐私保护参数(算法 3 的 ε_5)	0.7ε

4.3 实验结果及分析

首先简单介绍本文各个算法中相关参数的选择情况：

a)算法 1 和 2 中的稳定参数的取值源于文献[10]，分别取值为 $\beta_i = 15, \beta_u = 20$；

b)算法 3 中的隐含特征矩阵的特征个数 $k = 5$；

c)算法 3 中的正则化参数 $\lambda = 0.01$；

d)算法 3 中的 SGD 和 ALS 求解迭代次数设为 20，SGD 中学习率取值为 $\gamma = 0.001$。

由于篇幅有限，在此未给出参数选择的具体分析和说明。

实验结果如图 1 和 2 所示。图 1 给出了先对输入数据集作相关预处理，然后为预处理后的数据集添加噪声，最后分别采用 SGD 和 ALS 进行矩阵分解得到的实验结果与它们各自的 baseline 进行比较。具体地，图 1 中的 SGDBaseline 表示未作任何差分隐私保护且用 SGD 来求解的矩阵分解方法，ALSBaseline 类似；DPInSGD 表示对原始数据做了预处理，然后在对预处理后的数据添加噪声，最后采用 SGD 来求解，DPInALS 类似。从图 1 的实验结果可以看出：当差分隐私保护参数 $\varepsilon \geqslant 4$ 时，有差分隐私保护的两种方法均逐渐趋近于它们的 baseline，且 SGD 方法略优于 ALS 方法；当 $\varepsilon < 4$ 特别是 $\varepsilon < 1$ 以后，本文的方法偏离 baseline 就很远了，这是因为 ε 越小则对数据添加的拉普拉斯噪声越大，虽然对数据隐私保护得较好，但是加噪后的数据必定偏移原始数据越多，也就极大地降低了推荐的准确率。

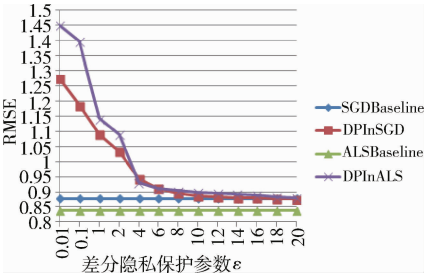
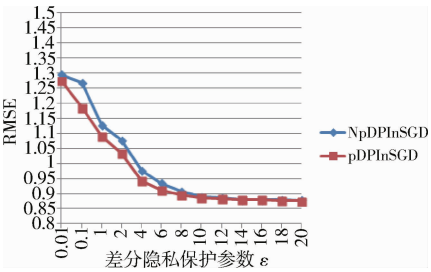


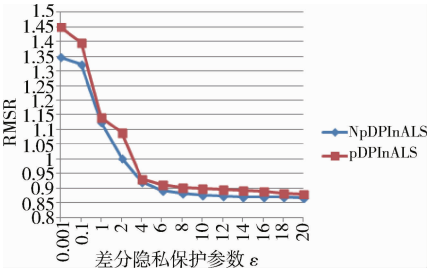
图 1 差分隐私保护的矩阵分解实验结果比较

图 2 给出了不采用数据预处理和采用本文提出的算法 1 和 2 的预处理方法的实验结果比较，表示 SGD 和 ALS 的结果。图 2(a)中 NpDPInSGD 表示未作预处理的 DPInSGD，pDPInSGD 表述做了预处理，图 2(b)类似。从图 2 可以看出，做了预处理后，DPInSGD 的 RMSE 优于未作预处理的结果，而 DPInALS 却相反，这是跟 SGD 和 ALS 自身的计算方法有关，在 2.2.2 节的式(6)可以看出，SGD 的每一次迭代更新与误差有很大关系，而 ALS 的每一次迭代是直接与训练的数据集有关，因此 ALS 对数据的变化表现得更为明显。

因此，上述实验结果表明，本文提出的新的具有差分隐私保护能力的矩阵分解方法既能在某种程度上保护原始数据的隐私，又没有大幅影响推荐的准确率，权衡了隐私和推荐两个方面的效能。



(a)有无预处理的 DPInSGD 实验结果比较



(b)有无预处理的 DPInALS 实验结果比较

图 2 有无预处理的矩阵分解实验结果比较

5 结束语

推荐系统日前已经成为众多 Internet 商家提供的必备服务之一，该服务受到了大量在线用户的青睐，但无论从学术价值还是商业价值的角度，绝大多数研究都是考虑如何提高推荐的准确率，而忽略了用来作研究的原始数据的隐私保护问题。本文从考虑保护原始数据的隐私信息出发，继而将当今被证明为最为严格的隐私保护模型——差分隐私应用于解决推荐问题的矩阵分解方法之中，最终通过在真实数据集上的实验验证，提出的新的带隐私保护的矩阵分解方法既能保护用于作推荐的原始数据集的隐私，又没有严重影响推荐的准确率。

根据目前 Internet 的飞速发展和越来越多的用户期望更个性化的推荐服务，隐私保护问题也必将成为用户堪忧的问题，推荐系统乃至数据挖掘领域需要健康的发展，离不开隐私保护问题的深入研究。本文仅从数据的预处理和输入添加噪声的角度进行了差分隐私保护的研究，下一步工作拟对参数的自动调节、推荐算法自身等多个角度来展开更深入的研究。

参考文献：

[1] Chen R, Mohammed N, Fung B C M, *et al*. Publishing set-valued data via differential privacy[C]//Proc of the 37th Conference on Very Large Databases. 2011: 1087-1098.

[2] Chen R, Fung B. C. M, Desai B C, *et al*. Differentially private transit data publication: a case study on the montreal transportation system[C]//Proc of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2012: 493-502.

[3] Blum A, Dwork C, Mcsherry F, *et al*. Practical privacy: the SuLQ framework[C]//Proc of the 24th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. 2005:128-138.

[4] Friedman A, Shuster A. Data mining with differential privacy[C]//Proc of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2010: 493-502.

[5] Mohammed N, Chen R, Fung B C M, *et al*. Differentially private data release for data mining[C]//Proc of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. CA, USA, 2011:493-501.

[6] Quinlan J. Induction of decision tress[J]. Machine Learning, 1989, 1(1):81-106.

[7] Smith A. Privacy-preserving statistical estimation with optimal convergence rate[C]//Proc of the 43th Annual ACM Symposium on Theory of Computing. 2011:813-822.

[8] Chaudhuri K, Monteleoni C, Sarwate A. Differentially private empirical risk minimization[J]. Journal of Machine Learning Research, 2011,12(2):1069-1109.

(下转第 页)

(上接第 1553 页)

[9] Nissim K, Raskhodnikova K, Smith A. Smooth sensitivity and sampling in private data analysis[C]//Proc of the 39th Annual ACM Symposium on Theory of Computing. 2007:75-84.

[10] Mcsherry F, Mironov I. Differential private recommender system: building privacy into Netflix prize conteders[C]//Proc of the 15th ACM SIGKDD International Camference on Knouledge Discovery and Data Mining. New York:ACM Press,2009: 627-636.

[11] 李杨,张新政. 差分隐私保护综述[J]. 计算机应用研究,2012,29(9):3201-3211.

[12] 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.

[13] Dwork C. Differential privacy[C]//Proc of the 33rd International Colloquium on Automata, Languages and Programming, part Ⅱ. 2006:1-12.

[14] Dwork C. Differential privacy: a survey of results[C]//Proc of the 5th International Conference on Theory and Applications of Models of Computation. Berlin:Springer,2008:1-9.

[15] Dwork C, Mcsherry F, Nissim K, *et al.* Calibrating noise to sensitivity in private data analysis[C]//Proc of the 3rd Theory of Cryptography Conference. USA, 2006: 363-385.

[16] Mcsherry F. Privacy integrated queries: an extensible platform for privacy-preserving data analysis[C]//Proc of the ACM SIGMOD International Conference on Management of Data. 2009: 19-30.