基于目标扰动的本地化差分隐私矩阵分解推荐算法

王 永[1),2)] 罗陈红[2)] 邓江洲[1)] 高明星[2)]

[1)](重庆邮电大学经济管理学院 重庆 400065)

[2)](重庆邮电大学计算机科学与技术学院 重庆 400065)

摘 要 推荐系统作为有效应对信息过载的工具被广泛应用在电子商务、社交媒体和新闻资讯等领域中。矩阵分解具有泛化能力强和计算效率高的优点,是构建推荐系统的主流算法之一。为提高推荐质量,推荐服务器需要收集大量用户数据用于推荐模型的训练。由于推荐服务器不是完全可信的,向服务器共享用户数据会对用户隐私构成极大的威胁。如何构建一个在保护用户隐私的同时,还能确保推荐质量和准确性的系统,成为了一个热门的研究话题。本地化差分隐私是一种分布式的隐私保护机制,它从中心化差分隐私中发展而来,旨在解决服务器不可信场景下的数据的安全收集和分析。这种框架通过精确的数学证明来确保隐私保护的强度。目前,已经有研究工作将本地化差分隐私引入推荐系统,目的是在推荐效果可接受的情况下,确保用户隐私数据的安全。然而,这些研究还面临一些挑战。首先,隐私保护的范围有限。目前的方法大多只关注显式数据的具体数值,认为这是用户的隐私信息。事实上,攻击者可以通过检查数据是否包含在数据集中,来推测用户的隐私信息。其次,推荐质量较低。本地化差分隐私通过引入扰动来保护用户隐私,但这种方法会导致扰动幅度过大和误差累积,进而影响推荐质量。在推荐服务器不可信场景下,本文提出一种基于本地化差分隐私的矩阵分解推荐算法。首先,该算法将评分数值和评分存在性同时作为隐私保护的对象,为用户提供全面的隐私保护。其次,本算法采用目标扰动方法,添加的噪声量不会随着迭代次数增加而增加,有效避免模型训练过程中噪声累积的问题,保证模型训练的有效性。最后,针对分布式场景下多轮迭代导致的中间参数泄露问题,以无放回方式将采样的模型梯度元素发送给推荐服务器,用于模型训练。本文从理论上证明了所提算法满足本地化差分隐私。对所提算法的效用分析证明本文算法在保证有效的推荐质量的同时,能够实现对用户隐私数据的保护。实验结果表明本文算法极大地提高了隐私保护推荐算法的性能,本文算法在公开数据集上的误差下降幅度平均可达18%,在推荐领域数据隐私保护中展现出良好的应用价值。

关键词 矩阵分解;本地化差分隐私;目标扰动;推荐算法;隐私保护

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2025.00451

Matrix Factorization Recommendation Algorithm Based on Local Differential Privacy with Objective Perturbation

WANG Yong[1),2)] LUO Chen-Hong[2)] Deng Jiang-Zhou[1)] GAO Ming-Xing[2)]

[1)](College of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing 400065)

[2)](College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065)

Abstract As an effective tool to deal with information overload, recommendation system has been widely used in E-commerce, social media, news and other fields. Matrix factorization has the advantages of strong generalization ability and high computational efficiency and become one of the main algorithms for constructing recommendation systems. To improve the quality of

收稿日期:2023-07-09;在线发布日期:2024-09-29。本课题得到国家自然科学基金(62272077, 72301050)、重庆市教委科技重大项目(KJZD-M202400604)资助。**王 永**(通信作者),博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为隐私保护、推荐系统、信息管理. E-mail: wangyong1@cqupt.edu.cn。**罗陈红**,硕士研究生,主要研究领域为隐私保护、推荐系统. **邓江洲**,博士,讲师,主要研究领域为推荐系统、决策优化、信息安全. **高明星**,硕士,主要研究领域为隐私保护、推荐系统.

recommendation, the recommendation server needs to collect a large amount of user data for training recommendation models. Sharing user data with recommendation servers poses a significant threat to user privacy due to the inherent lack of complete trustworthiness in these servers. The construction of a recommendation system that ensures user privacy protection while maintaining recommendation efficiency and accuracy has become a hot topic of research. Local Differential Privacy is an extension of Central Differential Privacy and has been developed as a distributed privacy protection mechanism. It is designed to address the secure collection and analysis of data in scenarios where the server's trustworthiness is questionable. This framework ensures the strength of privacy protection through rigorous mathematical proofs. Recently, Local Differential Privacy has been integrated into recommendation systems with the objective of safeguarding user privacy data while maintaining an acceptable level of recommendation efficacy. However, these studies encounter several challenges. Firstly, the scope of privacy protection is somewhat limited. Current methodologies predominantly focus on the specific numerical values of explicit data, considering these as the users' private information. In fact, adversaries can infer users' private information by examining the presence or absence of data within the dataset. Secondly, the quality of recommendations is often compromised. Local Differential Privacy introduces perturbations to protect user privacy, but this approach can lead to excessive perturbation magnitudes and the accumulation of errors, which in turn can adversely affect the quality of recommendations. In this paper, a matrix factorization recommendation algorithm based on Local Differential Privacy is proposed for the scenario where the recommendation server is not trusted. Firstly, the proposed algorithm considers both rating value and the existence of rating as objects of privacy protection, providing users with comprehensive privacy protection. Secondly, the algorithm employs anobjectiveperturbation, where the amount of noise added does not escalate with the number of iterations. This effectively circumvents the issue of noise accumulation during the model training process, ensuring the efficacy of the model training. Finally, to solve the problem of intermediate parameter leakage intheprocessofmultiple iterations of distributed scenario, an element of the model gradient is selected by sampling without replacement and sent to the recommendation server for updatingmodel. This paper theoretically proves that the proposed algorithm satisfies local differential privacy. The utility analysis of the proposed algorithm proves that the proposed algorithm can achieve the protection of user privacy data while ensuring effective recommendation quality. Experimental results show that the proposed algorithm greatly improved the recommendation performance andthe error reduction of the proposed algorithm on the public datasets can reach 18% on average. Therefore, the proposed algorithm shows good application value in data privacy protection inthe recommendation field.

Keywords matrix factorization; local differential privacy; objective perturbation; recommendation algorithm; privacy protection

1 引 言

互联网电子商务应用中品类繁多的海量商品信息让用户目不暇接,难以做出高效且合理的选择,推荐系统应运而生。它利用用户的历史行为数据,分析其偏好,产生推荐列表,辅助用户决策。推荐系统在协助用户处理信息过载问题中发挥了重要作用[1]。与此同时,由于推荐系统在进行推荐计算的过程中需要使用大量的用户个性化数据,由此带来了人们对隐私数据泄露的担心。已有研究表明,在推荐系统中通过深入分析用户的个人数据,确实可以挖掘用户无意透露的敏感信息[2-3]。此外,推荐系统还可能从看似不敏感的数据中推断出用户的敏感

信息，如从用户对电影的评分数据中获取用户的政治倾向、健康状况等敏感属性[4-5]。因此，考虑用户数据隐私保护的推荐算法成为了近年来的研究热点。

矩阵分解[6]推荐模型是推荐系统中的一类基础模型。该模型通过机器学习的方式将高维稀疏的用户-项目评分矩阵分解为低维稠密的用户隐因子矩阵和项目隐因子矩阵。矩阵分解模型具有预测准确性高和计算效率高的特点，因此在推荐系统中得到了广泛的应用。为了满足隐私保护的需求，一些研究者将差分隐私[7]保护技术引入到矩阵分解中，提出了满足差分隐私的矩阵分解推荐模型[8-13]。

现有基于差分隐私矩阵分解推荐系统从结构上可分为两类：服务器可信的集中式结构和服务器不可信的分布式结构。这两类结构通常分别与中心化差分隐私(CDP)和本地化差分隐私(LDP)结合实现隐私保护的目的。相对于服务器可信的集中式结构，服务器不可信的分布式结构的限制条件更少，更契合实际应用场景，因此成为当前的研究热点。从隐私保护的对象来看，不少方案将用户对项目的评分值作为隐私保护的对象[14-17]，也有一些方案将用户与项目之间是否存在交互行为作为隐私保护的对象[18-19]。另外还有一些研究工作专门针对隐式反馈的差分隐私保护算法，也是将交互行为的存在性看作隐私保护对象[20-21]。然而，只保护评分或者只保护用户与项目之间交互的存在性都是有局限性的，它们在一定条件下都存在隐私泄露的可能[12]。因此，近年来的研究转向了同时对两者进行保护的强隐私保护推荐模型。强隐私保护策略的使用会给推荐模型的设计带来新的技术挑战，其中最突出的一点是由于需要保护的内容增多，引入的隐私噪声会增加，从而影响推荐的质量。其次，当前的推荐模型常采用梯度扰动实现强隐私保护，但是随着模型训练迭代轮次的增加，扰动程度会呈指数增长，导致严重的误差累积问题，从而影响推荐模型的准确性[9]。所以，为推荐系统中的核心数据提供强隐私保护的同时，如何有效保留数据效用，提高推荐质量，仍然是当前值得深入研究的问题。

针对上述问题，本文提出了一种本地化差分隐私矩阵分解推荐算法。该算法采用目标扰动来实现隐私保护，可以有效避免梯度扰动导致的误差累积问题，有效提高模型的训练质量。但是，采用目标扰动的方式需要解决噪声抹除的技术性挑战。因为，在服务器不可信的分布式场景下，用户和服务器之间交互的梯度是可以被截获的，并且目标扰动方法对模型施加的是重复噪声，因此攻击者在迭代过程中有可能抹除噪声，进而对用户数据构成极大的隐私威胁。为此，本算法采用目标扰动，同时对梯度采取不放回的随机选择策略以防止噪声抹除问题，从而能够有效防止不可信服务器和攻击者根据中间参数推断出用户的隐私信息。

本工作的主要贡献如下：

(1)文中提出了一种满足本地化差分隐私的矩阵分解推荐算法。该算法能够同时对评分数值和用户-项目交互的存在性进行保护，具有良好的强隐私保护性。

(2)在该推荐算法中，设计了一种基于随机选择的目标扰动机制。它不仅避免了模型迭代中的误差累积问题，而且可以有效解决分布式目标扰动机制所固有的噪声抹除问题，兼顾了推荐结果的有效性和隐私保护的安全性。

(3)理论证明本文算法能够满足ϵ-LDP的强隐私保护，并从理论上说明了本文算法能够有效发挥数据效用，保证算法具有良好的推荐结果。同时，通过实验证明了本文算法在保证强隐私保护力度的基础上确实具有优良的推荐性能。

2　相关工作

推荐系统主要利用用户与项目的基本信息及其交互信息预测用户偏好，为用户提供个性化推荐服务，在电子商务、社交媒体等诸多领域有着广泛的应用。推荐算法总体上可划分为基于内容的推荐、基于邻居的推荐和基于模型的推荐[22]。其中，基于内容和基于邻居的推荐主要是利用相似性度量获取类似的内容或者用户的邻居，然后直接将相似的内容推荐给目标用户，或者是结合邻居的意见完成物品推荐[23-24]。基于模型的推荐则是利用已有用户和项目数据，通过学习发现其内部规律，然后构建相应的模型并根据模型的计算结果完成推荐。当前，基于模型的推荐，特别是其中基于机器学习模型的推荐成为了研究的主流。与此同时，推荐系统中大量的用户数据因其极高的应用价值和隐私性逐渐成为大众和隐私保护领域关注的焦点。差分隐私[25]作为一种可证明、可量化的隐私保护技术，受到了高度重视和广泛应用。当前，差分隐私已被引入到推荐系统，成为了该领域中一项重要的数据隐私保护技术[26]。矩阵分解[6]是一种应用广泛的推荐模型，它

具有计算效率高和可控性强的特点，因而常被选为构造隐私保护推荐系统的基础模型[11,16-17]。从推荐服务器是否可信的角度，可将差分隐私推荐系统划分为两类，即：服务器可信的中心化差分隐私保护推荐系统和服务器不可信的本地化差分隐私保护推荐系统。

在服务器可信的场景下，已有一些基于矩阵分解的中心化差分隐私保护推荐方案被提出，展现出良好的应用价值。Fridman等人[10]提出了中心化差分隐私保护下的矩阵分解框架，给出了输入扰动、梯度扰动和输出扰动三种不同的方式，为设计具有隐私保护性的矩阵分解推荐算法奠定了良好基础。Hua等人[11]通过对矩阵分解的目标函数添加噪声，实现了对输出的项目隐因子矩阵进行差分隐私保护的目的，并提出了相应的隐私保护推荐算法。考虑到不同用户具有不同的隐私需求，Zhang等人[14]将个性化差分隐私保护[15]和目标扰动方法相结合，提出了个性化差分隐私概率矩阵分解模型。该方案在实现隐私保护的同时，从个性化的角度提升了数据效用，为保证推荐算法计算结果的准确性奠定了良好的基础。为了弥补由目标扰动所引起的参数独立更新的不足，Zhang等人[16]提出联合优化用户隐因子和项目隐因子的方案，有效提升了数据可用性和推荐质量。在服务器不可信场景下，出于隐私保护的考虑，与用户相关的隐私数据，如评分矩阵、用户隐因子矩阵等，不能存在服务器中。矩阵分解模型的训练与预测计算需要通过服务器与用户端的交互才能完成。为此，一些基于本地化差分隐私[8-9]的矩阵分解推荐方案被提出。以中心化隐私保护算法为基础，Hua等人[11]针对用户评分的保护，提出引入可信第三方及相应的加密机制，设计了DPMF方案。该方案能够在服务器不可信的条件下对评分数据进行隐私保护同时完成相应的预测与推荐。然而，可信第三方的引入在一定程度上限制了DPMF的应用，加密机制的使用则降低了模型训练的效率。在不引入可信第三方的条件下，Ermiş等人[17]依据本地化差分隐私框架，提出在客户端对矩阵分解模型的梯度进行扰动，在完成推荐计算的同时实现对用户评分的隐私保护。然而，上述方案只保护了推荐系统中的用户评分值，却未考虑用户与项目之间是否存在交互这类隐私信息的保护。已有研究表明，用户与项目之间的交互信息，即用户所评价的项目集合，在特定的情况下也会暴露用户的隐私[5]。因此同时对用户评分值和用户所评价的项目集合实施保护的强隐私保护推荐方案日益受到研究者的重视。Jiang等人[12]提出了满足差分隐私的两阶段随机响应机制，不仅能对用户评分进行隐私保护，还能有效防止不可信服务器推断出用户评过分的项目集，扩大了数据隐私保护的范围。Shin等人[13]将本地化差分隐私引入矩阵分解，实现了同时对用户评分值和用户与项目交互信息进行保护的强隐私保护方案。该算法还对高维稀疏的梯度矩阵进行了压缩和降维，以减小梯度扰动引起的误差，提高数据的效用。

现有基于矩阵分解的隐私保护推荐算法存在的不足主要是数据隐私保护范围的局限。一些研究方案[10-11,14-17]只强调对用户评分的隐私保护，忽略了用户评分的存在性，未对用户与项目之间的交互进行保护。虽然，有一些方案考虑了对用户评分的存在性进行保护[18-19]，但保护的范围仅仅只是用户单个评分的存在性，而非对用户的整个评分集的存在性进行保护。其次，扩大数据的隐私保护范围往往需要推荐模型添加过多的噪声，从而导致推荐误差过大的问题。已有方案通过梯度扰动同时实现对用户评分和评分集的强隐私保护[13]。这些方案需要在模型的迭代过程对梯度进行隐私预算分配，随着迭代轮数的增加，对梯度的扰动程度会指数增长，从而产生较为严重的误差累积问题。如何增强隐私保护的数据内容，协调隐私保护强度与数据效用之间的关系，保证推荐结果具有良好的质量，仍然是当前研究中需要继续深入的问题。

3 预备知识

3.1 矩阵分解

在包含n个用户，m个项目的分布式推荐系统场景中，令用户-项目评分矩阵为$\boldsymbol{R}=[r_{ij}]_{n\times m}$，其中$r_{ij}$为用户$i$对项目$j$的评分。推荐系统根据$\boldsymbol{R}$中现有的值，预测其中的空缺值，即用户对未交互的项目的评分值。矩阵分解是推荐系统常用的一种基本的机器学习模型，其核心思想是将高维稀疏的矩阵$\boldsymbol{R}$分解成两个低维稠密的隐因子矩阵$\boldsymbol{U}$和$\boldsymbol{V}$，即

$$\boldsymbol{R}=\boldsymbol{U}\boldsymbol{V}^{T} \tag{1}$$

其中，$\boldsymbol{U}=\{\boldsymbol{u}_i^T\}_{1\leqslant i\leqslant n}$表示用户的隐因子矩阵，用于刻画用户的偏好特性；$\boldsymbol{V}=\{\boldsymbol{v}_j^T\}_{1\leqslant j\leqslant m}$表示项目的隐因子矩阵，用于刻画项目属性的特性。其中$\boldsymbol{u}_i\in\mathbb{R}^d$表示用户$i$的隐因子向量，$\boldsymbol{v}_j\in\mathbb{R}^d$表示项目$j$的隐因

子向量，d为隐因子向量维度，且$d \ll min(m,n)$。

矩阵分解模型通常根据正则化最小平方误差求解$\boldsymbol{U}$和$\boldsymbol{V}$，其目标函数如式(2)所示：

$$\min_{U,V} \frac{1}{2N}\sum\nolimits_{(i,j)\in O}(r_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j)^2+\frac{1}{2}\lambda\left(\|\boldsymbol{u}_i\|^2+\|\boldsymbol{v}_j\|^2\right) \tag{2}$$

其中，O为已有用户-项目评分对集合；N为集合O内元素的个数，即$N=|O|$；$\lambda>0$为正则化系数。采用随机梯度下降法求解(2)式时，用户隐因子向量和项目隐因子向量的更新规则如式(3)和式(4)所示：

$$\boldsymbol{u}_i \leftarrow \boldsymbol{u}_i-\gamma\nabla_{u_i} \tag{3}$$

$$\boldsymbol{v}_j \leftarrow \boldsymbol{v}_j-\gamma\nabla_{v_j} \tag{4}$$

其中，γ表示学习率；∇_{u_i}和∇_{v_j}分别表示$\boldsymbol{u}_i$和$\boldsymbol{v}_j$的梯度。在分布式矩阵分解模型中，每个用户均在本地完成梯度计算。用户i计算∇_{u_i}和∇_{v_j}的公式为

$$\nabla_{u_i}=-2\sum\nolimits_{j:(i,j)\in O}\boldsymbol{v}_j(r_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j)+\lambda\boldsymbol{u}_i \tag{5}$$

$$\nabla_{v_j}^{i}=-2\sum\nolimits_{i:(i,j)\in O}\boldsymbol{u}_i(r_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j)+\lambda\boldsymbol{v}_j \tag{6}$$

分布式矩阵分解的工作过程为：首先，用户i在本地完成梯度计算之后，将∇_{v_j}发送给服务器。然后，服务器端对所有用户发送的梯度$\nabla_{v_j}(i=1,2,\cdots,n)$进行聚合，得到聚合后的梯度$\nabla_{v_j}$，如公式(7)所示：

$$\nabla_{v_j}=\frac{1}{n}\sum\nolimits_{i=1}^{n}\nabla_{v_j}^{i} \tag{7}$$

最后，根据公式(4)更新项目j的隐因子向量$\boldsymbol{v}_j$，进而更新项目隐因子矩阵$\boldsymbol{V}$。最后，将$\boldsymbol{V}$分发给各个用户用于计算下一次迭代的梯度。重复上述过程，直至模型收敛。

3.2　本地化差分隐私

本地化差分隐私是差分隐私的一种变体。在本地化差分隐私场景中，服务器对于用户而言是不可信的，用户的数据必须保存在用户本地。它考虑了用户数据在模型训练中的隐私泄露问题，在分布式的机器学习场景展现出良好的应用价值。本地化差分隐私的定义如下：

定义1. 本地化差分隐私(LDP)[9]假设存在n个用户，每个用户对应一条数据，对于随机算法A，其定义域为$Domain(A)$，值域为$Range(A)$，若随机算法A在任意两条不同的记录x和$x'(x,x'\in Domain(A))$上得到相同输出结果$x^*(x^*\in Range(A))$，且满足

$$\Pr[A(x)=x^*]\leqslant e^{\epsilon}\Pr[A(x')=x^*],$$

则称随机算法A满足ϵ-本地化差分隐私，其中$\epsilon>0$表示隐私预算。

并行组合性是本地化差分隐私中的一项重要性质，它在证明分布式模型满足差分隐私保护的过程中常发挥重要的作用，其定义如下：

定义2. 并行组合性[9]给定数据集D，将其划分为m个互不相交的子集，即$D=\{D_1,D_2,\cdots,D_m\}$，设A为任一满足本地化差分隐私的随机算法，则算法A在$\{D_1,D_2,\cdots,D_m\}$上满足ϵ-本地化差分隐私。

4　基于目标扰动的差分隐私

4.1　应用场景分析

本文探讨的本地化差分隐私矩阵分解模型的应用场景如图1所示。在该场景中，假设中心服务器S不可信，用户将评分数据和用户隐因子向量$\boldsymbol{u}_i$存储在本地，项目隐因子矩阵$\boldsymbol{V}$存储在服务器S上。在每一轮训练中，服务器向所有用户发送当前的项目隐因子矩阵$\boldsymbol{V}$；用户在本地计算关于$\boldsymbol{V}$的梯度，然后将梯度上传给服务器用于更新$\boldsymbol{V}$。有研究表明，根据训练过程中的梯度可以推断出用户的评分数据。同时对于用户没有评过分的项目，梯度是0，根据这个信息可以知道用户对哪些项目评过分。

为了同时对用户评分和与用户有交互的项目集进行隐私保护，现有研究主要是以式(2)～(7)所描述的分布式矩阵分解为基础，通过在式(6)中添加噪声对梯度进行扰动来实现[13,17]。然而，基于梯度扰动的差分隐私保护方法在模型训练过程中需要为每一次迭代分配隐私预算并添加噪声。随着迭代次数的增加，每次迭代分配的隐私预算就会减小，添加的噪声会急剧增大，同时还会出现噪声的累积，影响数据效用甚至导致数据完全失去可用性。因此，有必要探寻新的隐私保护方案。

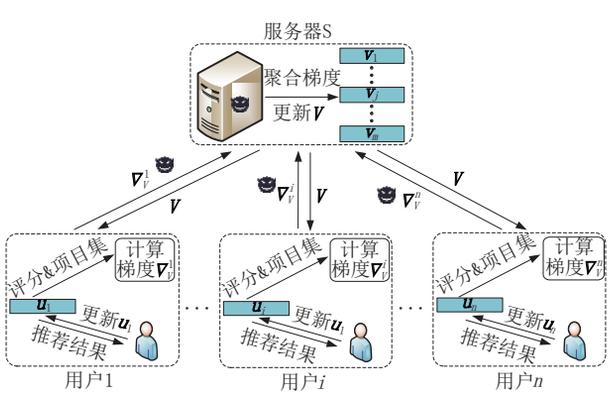


图1　本地化差分隐私矩阵分解的应用场景

4.2 基于目标扰动的本地化差分隐私矩阵分解基础方案

与梯度扰动不同，目标扰动方法不具有迭代累加性，即噪声的影响不会随着迭代次数增多而增多，从而有效保证数据的效用。为此，在上一节所描述的场景下，提出矩阵分解的噪声扰动目标函数为

$$\min_{V}\tilde{L}(V)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_{ij}\times\left(r_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j\right)^2+\lambda\sum_{j=1}^{m}\left\|\boldsymbol{v}_j\right\|_2^2+\sum_{j=1}^{m}\boldsymbol{\eta}_j^T\boldsymbol{v}_j \tag{8}$$

式中 $y_{ij}=\begin{cases}1, & \text{如果} r_{ij}>0\\ 0, & \text{其他}\end{cases}$，$\boldsymbol{\eta}=\{\boldsymbol{\eta}_j\}_{1\leqslant j\leqslant m}$ 为噪声矩阵，$\boldsymbol{\eta}_j$ 是用于扰动项目隐因子向量 $\boldsymbol{v}_j$ 的噪声向量，它是 n 个用户分别对 $\boldsymbol{v}_j$ 添加的噪声的聚合结果，即 $\boldsymbol{\eta}_j=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\eta}_j^i$。参考集中式基于目标扰动的中心化差分隐私矩阵分解模型，分布式基于目标扰动的本地化差分隐私矩阵分解推荐算法的基础方案如算法1所示。在初始化阶段，服务器随机初始化项目隐因子矩阵 $\boldsymbol{V}$；每个用户 i 在本地使用SGD方法最小化目标函数(2)，求得用户隐因子向量 $\boldsymbol{u}_i$；用户 i 对(8)式中的损失函数计算梯度 $\nabla_V^{i,*}=\nabla_V^i+\boldsymbol{\eta}^i$；服务器聚合每个用户的梯度并更新 $\boldsymbol{V}$。

算法1. 基于目标扰动的本地化差分隐私矩阵分解基础方案.

输入：评分矩阵 $\boldsymbol{R}\in\mathbb{R}^{n\times m}$，隐私预算 ϵ，迭代次数 $iter$，隐因子向量维度 d

输出：项目隐因子矩阵 $\boldsymbol{V}$

1. 服务器随机初始化项目隐因子 $\boldsymbol{V}\in\mathbb{R}^{m\times d}$
2. 用户 $i(i=1,2,\cdots,n)$ 使用SGD方法最小化目标函数(2)，求得用户隐因子向量 $\boldsymbol{u}_i$
3. FOR $t=0$ TO $iter-1$ DO
4. 用户 $i(i=1,2,\cdots,n)$，固定 u_i 并根据(8)式计算 V 的梯度 $\nabla_V^{i,*}=\nabla_V^i+\boldsymbol{\eta}^i$
5. 服务器接收所有用户的 $\nabla_V^{i,*}(i=1,2,\cdots,n)$，并进行聚合，即 $\nabla_V^*=\frac{1}{n}\sum_{i=1}^{n}\nabla_V^{i,*}$
6. 服务器按公式(4)更新全局的 $\boldsymbol{V}$，即 $\boldsymbol{V}\leftarrow\boldsymbol{V}-\gamma_t\nabla_V^*$，并将 V 分发给各个用户
7. ENDFOR
8. RETURN $\boldsymbol{V}$

在上述基础方案中，模型训练时，服务器不仅能获知最终训练好的 $\boldsymbol{V}$，还能得到每次迭代中各个用户发送的中间参数 $\nabla_V^{i,*}$。在服务器可信情况下，该方案能够提供隐私保护，但是在服务器不可信场景下，该方案存在安全漏洞。因为该方案中的噪声是添加在目标函数中，进而被包含到 $\nabla_V^{i,*}$ 中；又由于任意两次迭代中添加的噪声是相同的，因此在两次迭代中有关系式 $\nabla_V^{i,*}(\boldsymbol{u}_i,V^t)-\nabla_V^{i,*}(\boldsymbol{u}_i,V^{t-1})=\nabla_V^i(\boldsymbol{u}_i,V^t)-\nabla_V^i(\boldsymbol{u}_i,V^{t-1})$ 成立，所以对相邻两次迭代中的梯度做差可以消除隐私保护噪声，从而为获取数据的真实值奠定良好基础。例如，对于用户从未评分的项目，在算法的学习过程中其真实梯度为0，由于连续两次迭代中添加的噪声是相同的，那么对于未评分的项目，这两次梯度之差就是0；相反，对于评过分的项目就不为0。因此攻击者通过检测连续两次迭代过程中的梯度值之差是否为0，可以推断出该用户是否与项目进行交互，从而获取用户的评分项目集。而用户的评分项目集也是强隐私保护方案中需要保护的内容之一，所以上述基础方案不能提供全面的隐私保护效果。

4.3 基于目标扰动的本地化差分隐私矩阵分解推荐算法

从安全角度看，前述基础方案存在的主要缺陷是同一用户在每次迭代中向同一个项目隐因子向量的梯度添加的噪声是相同的。为了避免相同的噪声在服务器端重复出现，本文采用的策略是在模型训练迭代过程中，用户至多只向服务器发送一次含有相同噪声的梯度值。为此，服务器端引入一个位置标记矩阵 $\boldsymbol{M}\in\{0\}^{m\times d}$，每一轮迭代中由服务器根据 M 确定每个用户应该上传哪个位置的梯度值，避免用户将相同位置的梯度值重复发送给服务器。在算法1的基础上，提出本文的改进方案，即基于目标扰动的本地化差分隐私矩阵分解推荐算法，如算法2所示。在算法开始之前，服务器初始化项目隐因子矩阵 $\boldsymbol{V}$ 和位置标记矩阵 $\boldsymbol{M}$，并生成服务器端的噪声矩阵 $\boldsymbol{H}\in\mathbb{R}^{m\times d}$，然后将 $\boldsymbol{V}$ 和 $\boldsymbol{H}$ 发送给用户；用户 i 在本地使用SGD方法最小化目标函数(2)，求得用户隐因子向量 $\boldsymbol{u}_i$；在对项目隐因子矩阵训练的第 t 轮迭代中，服务器根据位置标记矩阵 $\boldsymbol{M}$ 确定用户上传 p^t 这个位置上的梯度，用户在本地根据算法C2计算 p^t 这个位置上的梯度并上传到服务器用于更新 $\boldsymbol{V}$。

算法2. 基于目标扰动的本地化差分隐私矩阵分解推荐算法.

输入：评分矩阵 $\boldsymbol{R}\in\mathbb{R}^{n\times m}$，隐私预算 ϵ，迭代次数 $iter$，隐因子向量维度 d

输出：项目隐因子矩阵 $\boldsymbol{V}$

1. 服务器随机初始化 $\boldsymbol{V}\in\mathbb{R}^{m\times d}$，位置标记矩阵 $\boldsymbol{M}\in\{0\}^{m\times d}$ 和噪声矩阵 $\boldsymbol{H}\in\mathbb{R}^{m\times d}$，且 H 中元素 $H_{jl}\sim$

$E(1)$，并将V, H分发给用户$i(i=1,2,\cdots,n)$

2. 用户$i(i=1,2,\cdots,n)$使用SGD方法最小化目标函数(2)，求得用户隐因子向量$\boldsymbol{u}_i$；

3. FOR $t=0$ TO $iter-1$ DO

4. 　服务器端初始化$\nabla_V^*=\{0\}^{m\times d}$，从$\boldsymbol{M}$中选择一个值为0的元素$\boldsymbol{M}_{p^t}$，其中$p^t=(j,l)$表示其行和列的位置，将$p^t$发送给每个用户并设置$\boldsymbol{M}_{p^t}=1$；

5. 用户i接收服务器发送的p^t，计算梯度$\nabla_V^{i*}=C_2(p^t,\boldsymbol{u}_i,V,t,H)$，将$\nabla_V^{i*}$上传服务器；

6. 服务器接收所有用户的$\nabla_V^{i*}(i=1,2,\cdots,n)$，计算$\nabla_V^*=\frac{1}{n}\sum_{i=1}^n\nabla_V^{i*}$，得到梯度聚合结果；

7. 服务器根据式(4)更新全局的V，即$V\leftarrow V-\gamma_t\nabla_V^*$，并将$V$分发各个用户；

8. END FOR

9. RETURN V

在算法2的第5行中，用户i计算梯度∇_V^{i*}的详细过程如算法C2所示。首先用户i初始化需要上传的噪声梯度矩阵∇_V^{i*}，并生成随机数$C_{p^t}^i$计算噪声$\boldsymbol{\eta}_{p^t}^i$；然后用户$i$计算$p^t$这个位置上的噪声梯度$(\nabla_V^{i*})_{p^t}$并上传到服务器。

算法C2. 用户端的梯度计算.

输入：梯度位置标记p^t，用户隐因子向量u_i，项目隐因子矩阵V，当前迭代次数t，随机数矩阵$\boldsymbol{H}$

输出：∇_V^{i*}

1. 用户i初始化$\nabla_V^{i*}=\{0\}^{m\times d}$

2. 用户i生成随机数$\boldsymbol{C}_{p^t}^i\sim N\left(0,\frac{1}{n}\right)$

3. 生成噪声$\boldsymbol{\eta}_{p^t}^i=\frac{2\Delta B}{\epsilon}\cdot\sqrt{2H_{p^t}}\,\boldsymbol{C}_{p^t}^i$

4. 计算梯度$\nabla_V^i=\left\{y_{ij}\times\boldsymbol{u}_i\left(r_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j\right)+\lambda\boldsymbol{v}_j\right\}_{1\leqslant j\leqslant m}$

5. 扰动梯度$(\nabla_V^{i*})_{p^t}=(md-t)\left[(\nabla_V^i)_{p^t}+\boldsymbol{\eta}_{p^t}^i\right]$

6. RETURN ∇_V^{i*}

5　算法分析

5.1　安全性分析

引理1[27]. 假设有两个随机数$h\sim Exp(1)$，$c\sim N(0,1)$，并且随机数h和随机数c相互独立，则$X=\mu+b\sqrt{2h}\,c\sim Lap(\mu,b)$。

用户在本地生成随机数矩阵$\boldsymbol{C}^i\in\mathbb{R}^{m\times d}$，$\boldsymbol{C}^i$的每个分量$C_{jl}^i$服从$N(0,\frac{1}{n})$，服务器生成随机数矩阵$\boldsymbol{H}\in\mathbb{R}^{m\times d}$，$\boldsymbol{H}$的每个分量服从$Exp(1)$，并将随机矩阵$\boldsymbol{H}$发送给各个用户。用户在本地计算$\boldsymbol{\eta}_{jl}^i=\frac{2\Delta B}{\epsilon}\cdot\sqrt{2H_{jl}}\,\boldsymbol{C}_{jl}^i$，服务器聚合所有用户的噪声可得到$\eta_{jl}=\sum\eta_{jl}^i=\frac{2\Delta B}{\epsilon}\cdot\sqrt{2H_{jl}}\sum\boldsymbol{C}_{jl}^i$。由于$C_{jl}^i$是满足$N(0,\frac{1}{n})$的随机变量，根据高斯随机变量之和仍然满足高斯分布的性质，$\sum C_{jl}^i$满足$N(0,1)$。根据引理1，η_{jl}满足$Lap\left(0,\frac{2\Delta B}{\epsilon}\right)$。尽管噪声一部分来源于服务器，但由于随机数$C^i$是用户在本地生成的，因此服务器无法推断出噪声的确切值，保证了差分隐私保护的安全性。

定理1. 若噪声矩阵$\boldsymbol{\eta}$中每个分量η_{jl}服从$p(\eta_{jl})\sim lap\left(0,\frac{2\Delta B}{\epsilon}\right)$，其中，$\Delta$为评分数据集的敏感度，$B$为用户隐因子向量的模长的上界，则算法2满足$\epsilon$-LDP。

证明. 假设两个相邻数据集分别为$D=\{r_1,r_2,r_p\cdots r_n\}$和$D'=\{r_1,r_2,r_p'\cdots r_n\}$，其中$r_i=\{r_{i1},r_{i2},\cdots,r_{im}\}$表示用户$i$的所有评分数据。若用户未对项目评分，则对应的评分值用0表示。η和η'分别表示数据集D和D'训练模型时添加的噪声矩阵。以项目为标准，将数据集划分为m个不相交的子集，即$D=\{D_1,D_2,\cdots,D_m\}$，其中$D_j=\{r_{1j},r_{2j},r_{pj},\cdots,r_{nj}\}$表示所有用户对项目$j$的评分集合。按照相同的方式，将数据集$D'$也划分为$m$个不相交的子集，即$D'=\{D'_1,D'_2,\cdots,D'_m\}$，其中$D'_j=\{r_{1j},r_{2j},r'_{pj},\cdots,r_{nj}\}$。

令$\overline{V}=\{\overline{\boldsymbol{v}}_j\}_{1\leqslant j\leqslant m}$为根据算法2生成的项目隐因子矩阵。由于事先已求出$U$且它在算法2中固定不变，因此目标函数(8)式是关于V的凸函数。所以，对于$\forall j\in\{1,2,\cdots,m\}$，有

$$\nabla_{v_j}\tilde{L}(D,\overline{\boldsymbol{v}}_j)=\nabla_{v_j}\tilde{L}(D',\overline{\boldsymbol{v}}_j)=0。$$

将上述等式根据公式(6)展开，得到

$$\boldsymbol{\eta}_j-2\sum_{i=1}^n y_{ij}\times\boldsymbol{u}_i\left(r_{ij}-\boldsymbol{u}_i^T\overline{\boldsymbol{v}}_j\right)=\boldsymbol{\eta}'_j-2\sum_{i=1}^n y_{ij}\times\boldsymbol{u}_i\left(r'_{ij}-\boldsymbol{u}_i^T\overline{\boldsymbol{v}}_j\right),$$

当且仅当$i\neq p$时，有$r_{ij}=r'_{ij}$，因此有

$$\boldsymbol{\eta}_j-2\boldsymbol{u}_p\left(r_{pj}-\boldsymbol{u}_p^T\overline{\boldsymbol{v}}_j\right)=\boldsymbol{\eta}'_j-2\boldsymbol{u}_p\left(r'_{pj}-\boldsymbol{u}_p^T\overline{\boldsymbol{v}}_j\right),$$

又因为$|r_{pj}-r'_{pj}|\leqslant\Delta$，所以可以得到

$$\left\|\boldsymbol{\eta}_j-\boldsymbol{\eta}'_j\right\|\leqslant 2\Delta B。$$

对于相邻数据集D_j和D'_j，有

$$\frac{\Pr\left[\boldsymbol{v}_j=\overline{\boldsymbol{v}}_j|D_j\right]}{\Pr\left[\boldsymbol{v}_j=\overline{\boldsymbol{v}}_j|D_j'\right]}=\frac{\Pr\left[\boldsymbol{v}_j=\overline{\boldsymbol{v}}_j|r_p\right]}{\Pr\left[\boldsymbol{v}_j=\overline{\boldsymbol{v}}_j|r_p'\right]}=$$

$$\prod_{(t,l)\in[mark]}\frac{(md-t)}{(md-t)}\frac{\Pr\left[\eta_{jl}|r_{pj}\right]}{\Pr\left[\eta_{jl}'|r_{pj}'\right]}=\frac{\Pr\left[\eta_j|r_{pj}\right]}{\Pr\left[\eta_j'|r_{pj}'\right]}\leqslant$$

$$e^{\frac{\epsilon\left\|\boldsymbol{\eta}_j-\boldsymbol{\eta}_j'\right\|}{2\Delta B}}\leqslant e^{\epsilon},$$

其中$mark$表示迭代轮数t和当前迭代用户p选择的v_j的第l维度的元组集合。同理可得

$$\frac{\Pr\left[\boldsymbol{v}_j=\overline{\boldsymbol{v}}_j|D_j\right]}{\Pr\left[\boldsymbol{v}_j=\overline{\boldsymbol{v}}_j|D_j'\right]}\leqslant e^{\epsilon},j=1,2,\cdots,m。$$

所以，根据本地化差分隐私的并行组合性，算法2满足ϵ-LDP。

根据定理1可知，对于任意的相邻数据集，本文改进算法生成相同的$\boldsymbol{V}$矩阵的概率非常接近，因此攻击者无法通过差分攻击推断出用户的整条评分向量。同时，在改进算法中，无论用户是否对某个项目有过评分，在传递梯度矩阵时都会对未评分项目隐因子的梯度加噪声，从而服务器无法直接获取或推断出与用户有交互的项目集，所以实现了对用户数据的强隐私保护。另外，在训练过程中，对于梯度矩阵$\nabla_V^{i,*}$中每一个位置上的值，用户i只上传一次。所以攻击者没有办法得到连续两次迭代的关系式$\nabla_V^{i,*}(\boldsymbol{u}_i,\boldsymbol{V}^t)-\nabla_V^{i,*}(\boldsymbol{u}_i,\boldsymbol{V}^{t-1})=\nabla_V^i(\boldsymbol{u}_i,\boldsymbol{V}^t)-\nabla_V^i(\boldsymbol{u}_i,\boldsymbol{V}^{t-1})$，从而避免了噪声消除问题。

5.2 效用性分析

定理2. 算法2中，每次迭代时，服务器聚合的梯度∇_V^*是不做目标扰动时梯度∇_V的无偏估计。

证明：不失一般性，令当前为第t次迭代。由算法2知聚合梯度为$\nabla_V^*=1/n\sum_{i=1}^n\nabla_V^{i,*}$。根据算法C2的计算过程可知，梯度$\nabla_V^{i,*}(i=1,2,\cdots,n)$中的任一元素$\left(\nabla_V^{i,*}\right)_{jl}$以$\frac{1}{md-t}$的概率等于$(md-t)\left[\left(\nabla_V^i\right)_{jl}+\eta_{jl}^i\right]$，以$1-\frac{1}{md-t}$的概率等于0。因此，容易得到$\left(\nabla_V^{i,*}\right)_{jl}$的期望满足：

$$E\left(\left(\nabla_V^{i,*}\right)_{jl}\right)=\frac{1}{md-t}(md-t)\left[\left(\nabla_V^i\right)_{jl}+\eta_{jl}^i\right]+\left(1-\frac{1}{md-t}\right)\times 0=\left(\nabla_V^i\right)_{jl}+\eta_{jl}^i。$$

类似地，由单一元素扩展至整个矩阵，可得$E(\nabla_V^{i,*})=\nabla_V^i+\boldsymbol{\eta}^i$。所以，服务器端聚合后的梯度矩阵的期望为

$$E(\nabla_V^*)=E(1/n\sum_{i=1}^n\nabla_V^{i,*})=\nabla_V+E(\boldsymbol{\eta})=\nabla_V,$$

即服务器聚合的扰动梯度∇_V^*是真实梯度∇_V的无偏估计，证毕。

定理2说明在算法2的执行过程中，尽管用户在每次迭代中只发送单个带有噪声的梯度元素给服务器，但是当服务器收集到所有用户上传的梯度后，能比较准确地估计出不做目标扰动的梯度∇_V，从而有效保证了算法的推荐性能。此外，相比基于梯度扰动的本地化差分隐私矩阵分解算法，本文算法是对目标函数进行扰动，添加的噪声量不会随着迭代次数增加而增加，有效避免噪声累积的问题，更好地保留了数据的效用，为提高推荐结果的准确性奠定了良好的基础。

6 实 验

6.1 实验设置

本文使用两个公开数据集（MovieLens100K和MovieLens1M）来评估模型的性能，各数据集的特征信息如表1所示。同时，以8:2的比例划分训练集和测试集。实验环境：Windows10 64b操作系统，Intel CoreTM i7-9700@ 3.00 GHz处理器，16 GB RAM。

表1 数据集参数

数据集	用户数	项目数	评分数	稀疏度
MovieLens100K	943	1682	100 000	93.70%
MovieLens1M	6040	3706	1 000 208	95.63%

采用RMSE和MAE为度量算法性能的指标，其计算公式为

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{(i,j)\in O}\left(r_{ij}-\hat{r}_{ij}\right)^2} \tag{9}$$

$$\mathrm{MAE}=\frac{\sum_{(i,j)\in O}|r_{ij}-\hat{r}_{ij}|}{N} \tag{10}$$

其中，N为评分个数，r_{ij}表示真实评分，$\hat{r}_{ij}=\boldsymbol{u}_i^T\boldsymbol{v}_j$表示预测评分。RMSE和MAE值越小，则表示评分预测越准确，推荐质量越高。

为了保证实验的可靠性，取10次实验的平均值作为最终结果。实验中设置迭代次数$iter=1000$，正则化系数$\lambda=10^{-8}$。在movieLens100K数据集和movieLens1M数据集中，设置初始学习率$\gamma=0.015$。同时，为了避免梯度下降越过最优点，设定学习率随迭代次数增加而逐渐减少，衰减率为

0.95。同时，设置隐私预算范围为[0.1,1.0]。

为了进一步说明本文算法的性能，选取如下算法进行对比：

(1) DPMF[11]是不可信服务器场景下的分布式矩阵分解算法。它利用可信第三方暂存用户经过目标扰动后的梯度信息，并基于差分隐私框架产生项目隐因子矩阵，实现了对用户评分值的隐私保护。

(2)LTPS[28]是一种轻量级两阶段隐私保护算法。它采用分阶段的方式对用户评分和用户评分存在性分别进行隐私保护。对用户评分项目集使用随机响应机制来实现本地化差分隐私保护；针对用户评分则利用二值扰动对用户发送的梯度矩阵进行扰动，同时实现了对用户评分值和评分存在性的隐私保护。

(3) Private-avgSVD[29]是一种针对用户评分值进行隐私保护的方法。它将用户数据分为敏感和不敏感两种类型，仅对用户的敏感评分数据添加拉普拉斯噪声，较好地实现了隐私保护和数据效用的平衡。

(4)DPLRMF[20]是一种针对隐式反馈数据(即0/1数值)的中心化差分隐私矩阵分解推荐方法，该方法假设存在一个可信的中心服务器收集用户数据。它首先在没有任何隐私保护机制的情况下训练用户隐因子矩阵，然后将训练后的用户矩阵代入目标函数中，以获取关于项目隐因子矩阵的目标函数，最后向目标函数添加随机噪声以提供差分隐私保护。本文对DPLRMF进行修改使其适于本文的对比实验。由于DPLRMF的输出结果在区间(0,1)中，本文将DPLRMF的输出重新映射到评分区间，即将输出结果乘以评分数据的最大值来作为评分的预测结果。

(5) MF[6]是经典的矩阵分解算法。它不包含任何隐私保护的操作，此处将其用作对比实验的基线。

(6) RSMF是本文为了便于算法的性能比较而设计的针对分布式场景的MF算法。它以MF算法为基础，其学习过程中的梯度信息收集方式与本文算法相同，即客户端采用随机不重复采样的方式向服务器发送梯度更新的信息，但是发送的梯度信息中不包含隐私保护的噪声。

6.2　结果分析

6.2.1　隐私保护范围

在推荐系统的数据隐私保护中，根据应用场景的不同，对数据的保护方式和保护范围均有所不同。整体上看，推荐系统中的数据隐私保护可以分为对用户评分值的保护和对用户评分的项目集(评分存在性)的保护。表2中列出了本文算法和对比算法所采用的扰动方法以及数据隐私保护的范围。DPLRMF本身是针对隐式反馈数据，保护的就是用户对项目交互行为的存在性。由于本文是将DPLRMF修改后作为对比实验，并且DPLRMF是基于中心化差分隐私的，本文在此处不讨论DPLRMF的隐私保护范围与方式。

表2　各算法的隐私保护范围与方式比较

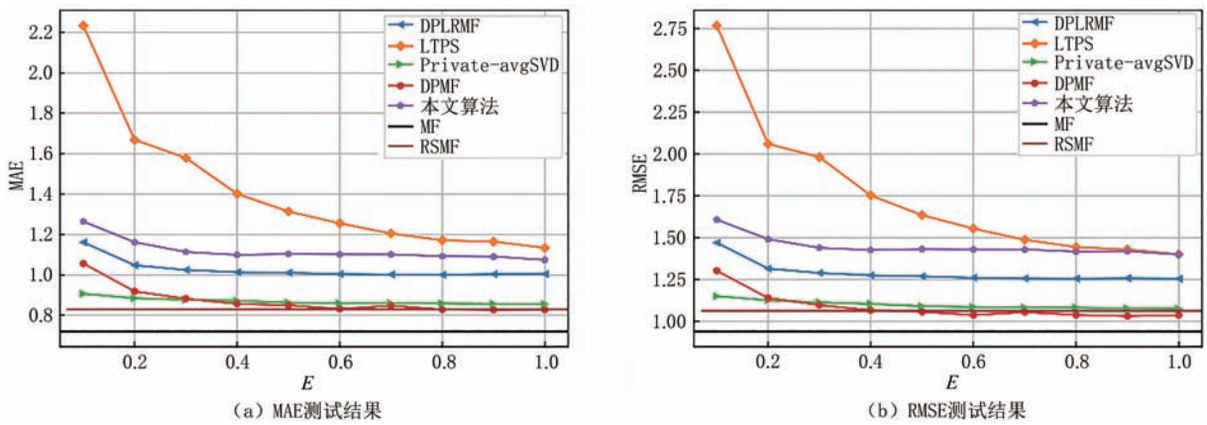
算法	隐私保护范围	扰动方法
本文算法	用户评分值、用户评分存在性	目标扰动
LTPS	用户评分值、用户评分存在性	输入扰动、梯度扰动
DPMF	用户评分值	目标扰动
Private-avgSVD	用户评分值	输入扰动

由表中可以看出，本文算法和LTPS提供了更全面的隐私保护，它们不仅对用户评分值进行了保护，而且也对用户评分的存在性进行保护。另外，从表2中还可以发现，尽管这两个算法提供的隐私保护范围是相同的，但是各自采用的扰动方法却不相同。关于扰动方法之间的优劣，将在下一小节中做进一步的分析。DPMF和Private-avgSVD仅对用户评分值进行了隐私保护，而未对用户评分的存在性进行隐私保护。从隐私保护范围的角度看，本文算法和LTPS的隐私保护强度优于DPMF和Private-avgSVD，能够为数据提供综合性的隐私保护。

6.2.2　性能分析

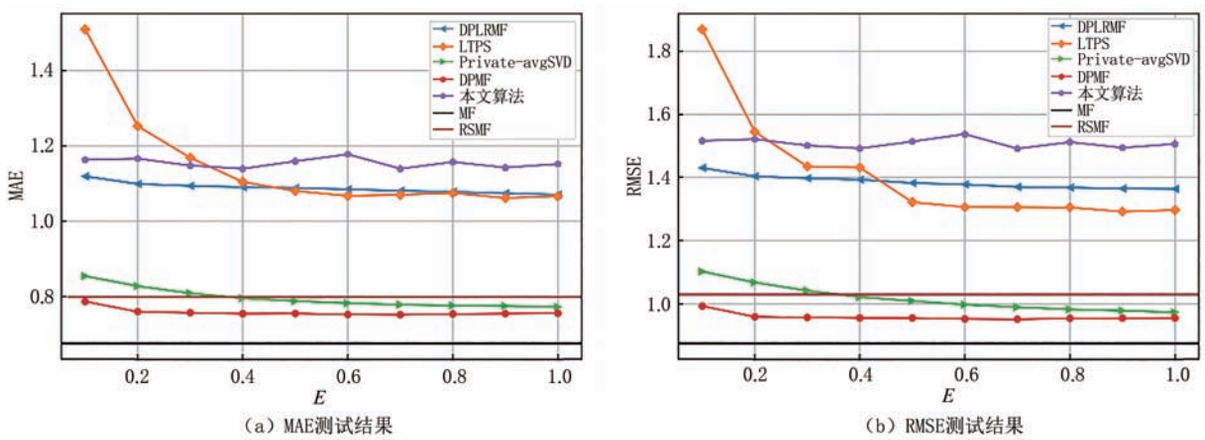
在不同隐私预算条件下，计算本文算法和各对比算法在MovieLens100K和MovieLens1M数据集上的MAE值和RMSE值，结果如图2和图3所示。总体而言，在图2和图3中，随着隐私预算增大，各算法在不同数据集上的RMSE与MAE值均逐渐减小。这符合差分隐私的一般规律，即随着隐私预算的逐步增加，扰动噪声减少，数据效用增加，推荐系统的性能随之提升。

首先在提供相同隐私保护范围的条件下，测试本文算法的性能。本文算法和LTPS都对用户评分值和评分存在性提供隐私保护，它们提供的隐私保护强度是相同的。比较这两个算法的性能，由图2和图3可知，在MovieLens100K数据集上，本文算法



（a）MAE测试结果 （b）RMSE测试结果

图2 MovieLens100K数据集中的测试结果



（a）MAE测试结果 （b）RMSE测试结果

图3 MovieLens1M数据集中的测试结果

的MAE值和RMSE值明显小于LTPS的测试值。在MovieLens1M数据集上，隐私预算较小时，本文算法的MAE值和RMSE值明显小于LTPS的测试值。相对于LTPS，本文算法具有性能优势的主要原因是本文算法采用的是目标扰动，而LTPS采用了梯度扰动的方式。梯度扰动会在每次迭代中添加扰动噪声。随着迭代次数的增加，会引起严重的误差累积问题，从而影响算法的性能。而目标扰动可以避免误差累积问题，因此具有更好的性能。另外，图2和图3中DPLRMF的性能虽然优于本文算法，但是DPLRMF是基于中心化差分隐私的，其隐私假设性更强。虽然本文算法是基于LDP的，但本文算法性能与DPLRMF差异不大。此外，从图2和图3还可以看出，本文算法的MAE和RMSE值随隐私预算增加，其波动的幅度远小于LTPS。这表明即使在隐私预算较小的情况下，本文算法仍然具有良好的预测准确性，能够更好地适用于对隐私保护要求较为严格的应用场景。

其次，在不同隐私保护范围的条件下，测试本文算法的性能。DPMF和Private-avgSVD仅对用户评分值进行隐私保护，比较本文算法与这两个算法的性能。由图2和图3可知，DPMF在整体上具有比Private-avgSVD和本文算法更优的MAE和RMSE。DPMF和Private-avgSVD在Movielens100K和Movielens1M数据集上具有优于本文算法的MAE和RMSE。产生上述实验结果主要的原因是本文算法除了对用户评分进行隐私保护之外，还会对用户评分的存在性进行隐私保护。由于本文算法提供的数据隐私保护范围更广，故添加的噪声更大，对数据效用的影响更大，因此其MAE和RMSE值大于DPMF和Private-avgSVD。此外，本文算法还能保证服务器端聚合的梯度是对不考虑隐私保护情况下的梯度的无偏估计，进一步保证了本文算法的性能。对比图2和图3中LTPS与DPMF和Private-avgSVD的性能差距，这进一步验证了本文算法能更好地控制噪声的影响，在提供更全面隐私保护的同时还能有效保证推荐算法的性能。

最后，通过与基线算法比较，进一步测试本文算法的性能。此处的实验中选用了两个基线算法：MF和RSMF。其中，MF采用的是集中式的数据处

理方式，RSMF采用了分布式的数据处理方式。从图2和图3中可以看出MF展现出明显性能优势。这是因为MF以集中方式处理数据，相对于分布式的数据处理方式，它能够更有效地获取数据信息。RSMF采用了与本文算法相同的数据处理方式，只是在模型训练过程中未添加隐私保护的噪声。

7 结　论

本文提出了一个基于目标扰动的本地化差分隐私矩阵分解推荐算法。该算法能够为用户提供较为全面的隐私保护，同时实现了对用户评分和评分存在性的保护。本文算法采用目标扰动来实现本地化差分隐私矩阵分解，有效避免了由梯度扰动引起的噪声累积问题。此外，本文算法在模型迭代学习过程中采用了梯度不放回抽样的方式，解决了分布式环境中模型中间参数存在泄露风险的问题，并从理论上证明了本文算法满足本地化差分隐私。同时还证明了依据本文方案在服务器端聚合的含有噪声的梯度是不考虑隐私保护情况下不含有噪声的梯度的无偏估计。从理论上说明了本文算法能够有效控制隐私保护噪声的影响，很好地保留了数据效用。最后，三个公开数据集上的实验结果验证本文算法不仅能够满足用户强隐私保护的需求，而且具有良好的推荐性能，展现出很好的应用价值。

参 考 文 献

[1] He X, Liao L, Zhang H, et al. Neural collaborative filtering//Proceedings of the 26th International Conference on World Wide Web. Perth, Australia, 2017: 173-182

[2] Calandrino J, Kilzer A, Narayanan A, et al. "You might also like:" privacy risks of collaborative filtering//Proceedings of the IEEE Symposium on Security and Privacy. Oakland, USA, 2011:231-246

[3] McSherry F, Mironov I. Differentially private recommender systems: Building privacy into the netflix prize contenders//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2009: 627-636

[4] Canny J. Collaborative filtering with privacy//Proceedings of the 2002 IEEE Symposium on Security and Privacy. Berkeley, USA, 2002:45-57

[5] Weinsberg U, Bhagat S, Ioannidis S, et al. Blurme: Inferring and obfuscating user gender based on ratings//Proceedings of the 6th ACM Conference on Recommender Systems. New York, USA, 2012:195-202

[6] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30-37

[7] Dwork C. Differential privacy//Proceedings of the International Colloquium on Automata, Languages, and Programming. Berlin, Germany, 2006:1-12

[8] Duchi J, Jordan M, Wainwright M. Local privacy and statistical minimax rates//Proceedings of the IEEE 54th Annual Symposium on Foundations of Computer Science. Berkeley, USA, 2013: 429-438

[9] Ye Qing-Qing, Meng Xiao-Feng, Zhu Min-Jie, Huo Zheng. Survey on local differential privacy. Journal of Software, 2018, 29(7):1981-2005 (in Chinese)
(叶青青,孟小峰,朱敏杰, 霍峥.本地化差分隐私研究综述.软件学报, 2018, 29(7): 1981-2005)

[10] Friedman A., Berkovsky S, Kaafar M. A differential privacy framework for matrix factorization recommender systems. User Modeling and User-Adapted Interaction, 2016:425-458

[11] Hua J, Chang X, Sheng Z. Differentially private matrix factorization//Proceedings of the 24th International Joint Conference on Artificial Intelligence. Palo Alto, USA, 2015: 1763-1770

[12] Jiang J, Li C, Lin S. Towards a more reliable privacy-preserving recommender system. Information Sciences, 2019, 482: 248-265

[13] Shin H, Kim S, Shin J, et al. Privacy enhanced matrix factorization for recommendation with local differential privacy. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(9): 1770-1782

[14] Zhang S, Liu L, Chen Z, et al. Probabilistic matrix factorization with personalized differential privacy. Knowledge-Based Systems, 2019, 183:104864

[15] Jorgensen Z, Yu T, Cormode G. Conservative or liberal? Personalized differential privacy//Proceedings of the IEEE 31st International Conference on Data Engineering. Seoul, Republic of Korea, 2015:1023-1034

[16] Zhang F, Lee V, Choo K. Jo-dpmf: Differentially private matrix factorization learning through joint optimization. Information Sciences, 2018, 467: 271-281

[17] Ermiş B, Cemgil A. Data sharing via differentially private coupled matrix factorization. ACM Transactions on Knowledge Discovery from Data (TKDD), 2020, 14(3): 1-27

[18] Shen Y, Jin H. Privacy-preserving personalized recommendation: An instance-based approach via differential privacy//Proceedings of the IEEE International Conference on Data Mining. Shenzhen, China, 2014:540-549

[19] Shen Y, Jin H. Epicrec: Towards practical differentially private framework for personalized recommendation//Proceedings of the ACM Conference on Computer and Communications Security. New York, USA, 2016:180-191

[20] Du Mao-Kang, Peng Jun-Jie, Hu Yong-Jin, Xiao Ling. Logistic regression matrix factorization recommendation algorithm for differential privacy.Journal of Beijing University of Posts and Telecommunications, 2023, 46 (3): 115-120 (in Chinese)

(杜茂康,彭俊杰,胡勇进,肖玲.满足差分隐私的逻辑回归矩阵分解推荐算法.北京邮电大学学报,2023,46(3):115-120)

[21] Liu H, Wang Y, Zhang Z, et al. Matrix factorization recommender based on adaptive Gaussian differential privacy for implicit feedback. Information Processing & Management, 2024, 61(4): 103720

[22] Huang Li-Wei, Jiang Bi-Tao, Lv Shou-Ye, et al. Survey on deep learning based recommender systems. Chinese Journal of computers, 2018, 41(07):1619-1647(in Chinese)
(黄立威,江碧涛,吕守业,等.基于深度学习的推荐系统研究综述.计算机学报,2018,41(07):1619-1647)

[23] Balabanović M, Shoham Y. Fab: Content-based, collaborative recommendation. Communications of the ACM, 1997, 40(3): 66-72

[24] Wang D, Yih Y, Ventresca M. Improving neighbor-based collaborative filtering by using a hybrid similarity measurement. Expert Systems with Applications, 2020, 160: 113651

[25] Dwork C, Roth A. The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 2014, 9(3-4): 211-407

[26] Xiong Ping, Zhu Tian-Qing, Wang Xiao-Feng. A survey on differential privacy and applications. Chinese Journal of Computers, 2014, 37(1):101-122 (in Chinese)
(熊平,朱天清,王晓峰.差分隐私保护及其应用.计算机学报,2014,37(1):101-122)

[27] Kotz S, Kozubowski T, Podgórski K. The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance. Berlin, Germany: Springer Science & Business Media, 2001

[28] Zhou H, Yang G, Xiang Y, et al. A lightweight matrix factorization for recommendation with local differential privacy in big data. IEEE Transactions on Big Data, 2021, 9(1): 160-173

[29] Zheng X, Guan M, Jia X, et al. A matrix factorization recommendation system-based local differential privacy for protecting users' sensitive data. IEEE Transactions on Computational Social Systems, 2023, 10(3): 1189-1198

WANG Yong, Ph. D., professor. His research interests include privacy protection, recommendation systems and information management.

LUO Chen-Hong, M. S. candidate. Her research interests include privacy protection and recommendation systems.

DENG Jiang-Zhou, Ph. D., lecturer. His research interests include recommendation systems, decision optimization and information security.

GAO Ming-Xing, M. S.. Her research interests include privacy protection and recommendation systems.

Background

The topic in this paper belongs to the domain of privacy-preserving recommendation algorithms. Although there exist some related works, they mainly focus on the scenario where recommender is trusted. In fact, the scenario of untrusted servers is more common. Currently, most local privacy preserving schemes are designed based on gradient perturbation and provide privacy for the values of ratings. Furthermore, some local privacy preserving schemes begin to explore the protection of the existence. Since more factors are considered, how to improve the comprehensive performance of recommendation system need to be further studied.

In this paper, a matrix factorization recommendation algorithm based on local differential privacy is proposed for the distributed scenario where the recommendation server is not trusted. In our algorithm, the private user data is stored in user's own local device and the privacy of rating values and their existence are both protected. Moreover, our scheme is designed based on an objective perturbation mechanism with random selection. Our scheme not only avoids the error accumulation problem which is commonly existed in the gradient perturbation, but also effectively resist the attack of noise elimination which is commonly existed in the distributed objective perturbation. We the oretically prove that the proposed algorithm satisfies local differential privacy. Experimental results show that the proposed algorithm greatly improved the recommendation performance. Therefore, the proposed algorithm shows high potential to be applied in privacy-preserving recommendation systems.

This work is supported in part by the National Natural Science Foundation of China (No. 62272077), the Natural ScienceFoundation of Chongqing, China (No. cstc2021jcyj msxmX0557) and the MOE Layout Foundation of Humanities and Social Sciences, China (No. 20YJAZH102).