一种面向推荐系统的隐私保护图神经网络

王 锟[1),3)] 王 永[1),2)] 张智强[2)] 刘金源[1)] 邓江洲[2)]

[1)](重庆邮电大学计算机科学与技术学院 重庆 400065)

[2)](重庆邮电大学经济管理学院 重庆 400065)

[3)](西南政法大学教育信息技术中心 重庆 401120)

摘 要 近年来,图神经网络凭借强大的图数据表示学习能力,在推荐领域得到广泛应用。现有基于图神经网络的推荐系统大多依赖中心服务器集中存储用户数据并训练模型。然而,用户数据中通常蕴含着大量的敏感信息,不可信的中心服务器可能通过隐私攻击窃取用户数据,从而严重威胁用户的隐私权益。虽然目前针对图神经推荐系统的隐私保护研究已取得一些进展,但大多都是建立在可信第三方的假设基础之上,因此在实际应用中具有较大的局限性。此外,由于图数据的结构复杂且关联性强,传统针对关系型数据的保护方法已无法满足其隐私保护需求。针对上述问题,本文提出了一种面向图神经推荐系统的隐私保护框架PF-GNR。该框架主要由3个部分组成:本地隐私编码模块、图神经推荐模块以及隐私保护更新模块。具体而言,首先,每个用户利用本地化差分隐私技术对原始数据进行扰动和编码,以保证数据采集过程的隐私安全;其次,在模型学习阶段,服务器使用图神经网络对用户和项目之间的偏好关系进行建模;最后,服务器借助同态加密技术为训练过程提供保护并完成模型的训练任务。与现有方法相比,PF-GNR将本地化差分隐私与同态加密技术相结合,能够在无可信第三方的情况下,为用户数据提供严格的隐私保护,同时保证模型可用性。此外,PF-GNR还是一种通用的解决方案,适用于各种图神经网络模型。在3个公开数据集上进行了大量的实验。实验结果表明,与目前最优方法相比,PF-GNR的整体表现平均提升了9.2%且对隐私预算的敏感度更低,验证了该方法在隐私性与可用性之间能够实现有效平衡。

关键词 推荐系统;图神经网络;隐私保护;本地化差分隐私;同态加密

中图法分类号 TP311 **DOI号** 10.11897/SP.J.1016.2025.01116

A Privacy-Preserving Graph Neural Network Framework for Recommendation

WANG Kun[1),3)] WANG Yong[1),2)] ZHANG Zhi-Qiang[2)] LIU Jin-Yuan[1)] DENG Jiang-Zhou[2)]

[1)](School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065)

[2)](School of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing 400065)

[3)](Education Information Technology Center of Southwest University of Political Science and Law, Chongqing 401120)

Abstract In recent years, graph neural networks (GNNs) have been widely used in the field of recommendation by virtue of their powerful graph representation learning capability. At present, most GNN-based recommender systems rely on central servers to centrally store users' data and training models. However, users' data contains sensitive information, and untrusted central servers or other adversaries may steal users' data through privacy attacks, which seriously threatens users' privacy rights. Although there have been some studies on privacy protection for GNN-based recommender systems, most of them are based on the security assumption that the central server is trusted, which has limitations in practical applications. Furthermore, due to the complex structure and strong correlation of graph data, the traditional privacy protection methods

收稿日期:2024-07-24;在线发布日期:2025-03-11。本课题得到国家自然科学基金面上项目(No.62272077)、国家自然科学基金青年基金(No.72301050)、重庆市教委科学技术研究重大项目(KJZD-M202400604)资助。**王 锟**,博士研究生,高级工程师,主要研究方向为图数据处理、推荐系统、数据隐私保护。E-mail:wangkun@swupl.edu.cn。**王 永(通信作者)**,博士,教授,博士生导师,主要研究领域为数据挖掘、数据隐私安全保护。E-mail:wangyong1@cqupt.edu.cn。**张智强**,博士,讲师,主要研究领域为机器学习、推荐系统。**刘金源**,博士,讲师,主要研究领域为数据挖掘、图像隐私安全保护。**邓江洲**,博士,讲师,主要研究领域为机器学习、推荐系统。

can no longer meet the privacy protection requirements of GNN-based recommender systems. To address the above problems, we propose a privacy-preserving framework for GNN-based recommender systems named PF-GNR. The framework consists of three main parts: a local encoding module, a recommendation module based on graph neural networks, and a privacy-preserving model update module. Specifically, each user first randomly scrambles and encodes the raw data using local differential privacy techniques to ensure the privacy of users' data during transmission and use. Secondly, in the learning phase, the central server uses GNNs to model the preference relationship between the users and the items. Lastly, the central server leverages homomorphic encryption technology to safeguard the training process and complete the model training task. Compared to existing approaches, PF-GNR is able to provide strict privacy protection for users' data without trusted third parties while guaranteeing model availability by combining local differential privacy with homomorphic encryption technique. In addition, PF-GNR is a generalized protection scheme suitable for various GNN-based recommendation methods. Extensive experiments on three real-life datasets demonstrate that PF-GNR improves the overall performance by an average of 9.2% and is less sensitive to the privacy budget than the state-of-the-art approaches, verifying that the proposed method can effectively balance between privacy and usability.

Keywords　recommender system; graph neural network; privacy protection; local differential privacy; homomorphic encryption

1　引　言

在大数据时代,随着移动互联网技术的快速发展和普及,导致网络信息量呈爆炸式增长,信息过载问题也日益严重。为有效缓解这一问题,个性化推荐系统被广泛应用于电子商务、社交媒体、搜索引擎等各种在线服务平台。这些推荐系统通过分析用户的数据和行为模式,来提供高效的个性化信息服务,从而帮助改善用户体验和促进商业盈利。在现实世界中,推荐系统的大部分信息可以自然地表示成图数据结构。以电商平台为例,用户的历史交互信息可转化为用户－项目交互图,其中,每个节点代表一个实体(即用户或项目),每条边代表两个实体之间的关系。如此,推荐问题便转化为图的链接(边)预测问题。近年来,涌现出了许多基于图神经网络(Graph Neural Networks,GNN)的推荐算法[1-7]。这些方法凭借强大的图数据表示学习能力,在各类推荐任务中都取得了优异的表现,并逐渐成为解决推荐问题最有效的工具之一。

虽然基于 GNN 的推荐系统能够为用户提供更加准确、高效的个性化信息服务,但在实际应用中也面临着严重的隐私风险。一方面,随着云计算的发展,大多数企业选择在云端部署推荐系统,并利用云服务器来完成用户数据的存储、分析和处理等一系列工作。然而,用户数据中通常包含了大量的敏感信息,如点赞、评分、购买等。如果用户数据缺乏必要的隐私保护措施,不可信的云服务提供商或其他攻击者可能会恶意窃取用户的敏感信息,从而导致隐私泄露[8-10];另一方面,图神经推荐系统采用消息传播机制[11]来训练模型,这种特有的训练方式使其更容易受到各种隐私攻击的威胁,如模型逆向攻击[12]、成员推理攻击[13-14]、链接窃取攻击[15-17]等。已有研究表明[18],在图数据的应用场景下,由于攻击者掌握了更加丰富的背景知识,例如结构信息、节点属性以及链接关系等,使得攻击模型的能力更强且造成的后果也更加严重。此外,随着人们隐私意识的增强和一系列数据安全法规[18-19]的相继实施,也在很大程度上影响了数据的分享和流通,这使得推荐系统的发展受到了严重限制。因此,如何在保证用户数据隐私安全的同时保持模型的可用性,已成为当前图神经推荐系统隐私保护研究所面临的关键问题之一。

近年来,国内外学者对图神经推荐系统中存在的隐私问题进行了深入研究,并提出了许多解决方案。这些方案根据所采用的隐私技术手段总体可分为两类:基于信息模糊的方法[20-22]和基于数据加密[23-24]的方法。(1)基于信息模糊的方法主要是利

用泛化、混淆、噪声扰动等方式对真实数据进行隐私化处理,从而抵抗敌手获取特定数据的隐私信息。典型的信息模糊技术包括中心化差分隐私[25](Centralized Differential Privacy,CDP)、本地化差分隐私[26](Local Differential Privacy,LDP)等。这类基于信息模糊的方法因具有计算效率高、通信开销低等优点[27]而受到大量的关注。然而,由于图数据的结构复杂且关联性强,噪声带来的负面影响较大,始终存在隐私与效用难以平衡的问题。(2)基于数据加密的方法是利用密码学工具对数据进行加密处理,确保数据在存储以及运算过程中的机密性和完整性。典型的加密技术包括同态加密[28](Homomorphic Encryption,HE)、秘密共享等。这类基于数据加密方法的最大优点是能够同时保证数据的隐私性和可用性。然而,由于密码技术普遍存在计算效率低、存储开销大等问题[18],目前还难以应用于复杂的机器学习模型。综上所述,尽管当前对图神经推荐系统的隐私保护研究已取得一定进展,但总体仍处于早期探索阶段,面临着以下问题与挑战。挑战 1:不可信的第三方。现有研究大多以中心服务器完全可信为前提,无法抵御来自不可信第三方的隐私攻击,因此在实际中具有较大的局限性;挑战 2:新的隐私保护对象。与关系型数据不同,图结构数据中的隐私保护对象不仅包括节点隐私,还包括节点之间的链接关系(边隐私),因此传统针对关系型数据的保护方法[29-30]已无法满足其隐私保护需求;挑战 3:隐私性与可用性难以平衡。图神经推荐系统采用消息传播机制进行建模,这使得图数据的节点之间并不是独立存在而是相互关联的,如果直接应用传统的机器学习隐私保护方法,将会破坏节点之间的关系,从而极大地损害模型的可用性。

为应对上述问题和挑战,本文提出了一种面向图神经网络推荐系统的隐私保护框架 PF-GNR (Privacy-Preserving Framework for GNN-based Recommender Systems)。该框架主要由三个部分组成:本地隐私编码模块、图神经推荐模块以及隐私保护更新模块。首先,针对挑战 1 和挑战 2,本文设计了本地隐私编码模块,用于为数据收集阶段提供严格的隐私保证。该模块部署在客户端,每个用户利用 LDP 技术对即将上传的数据进行随机扰动和编码,可避免服务器直接收集或接触到本地原始数据,从而有效保证了数据在传输和存储中的隐私性。其次,针对挑战 3,本文设计了隐私保护更新模块。在训练阶段,该模块协调多个参与方与中心服务器共同训练模型,并利用同态加密技术为各参与方的数据交换和运算提供隐私保护,防止中心服务器在训练过程中窃取参与方的隐私信息,以保证模型更新的安全性和隐私性。同时,通过采用这种协作训练模型的方式,还能有效减少差分噪声带来的影响,实现隐私性与可用性之间的有效平衡。总而言之,本文基于图神经推荐系统的隐私保护需求,将 LDP 和同态加密技术相结合,设计了一种两阶段的混合隐私保护机制(Hybrid Privacy-Preserving Mechanism,HPM)。该机制能够在不依赖可信第三方的情况下,为本地数据提供严格的隐私保护,同时保持模型的可用性。此外,HPM 机制已实现模块化,具有良好的通用性和可扩展性,能直接应用于其他 GNN 模型,例如 GraphSage[3]、FastGCN、ClusterGCN 等,以满足不同应用场景的需求。

本文工作的主要贡献总结如下:

(1)本文研究了图神经推荐系统的隐私保护问题,并提出一种新的隐私保护框架 PF-GNR。该框架不仅能够为用户提供从数据收集到模型训练的全周期隐私保护,同时还可保持模型的可用性。此外,PF-GNR 框架已实现模块化,具备良好的通用性,可以满足不同应用场景的需求。

(2)本文为 PF-GNR 框架设计了一种两阶段的混合隐私保护机制。该机制通过结合本地化差分隐私和同态加密技术的优势,能够在保证用户隐私安全的同时,有效降低差分噪声带来的负面影响,从而实现隐私保护与推荐性能的兼顾。

(3)本文对 PF-GNR 的隐私性进行了详细的理论分析,并在三个真实数据集上开展多组实验。实验结果表明,与现有最优方法相比,PF-GNR 的整体表现平均提升了约 9.2%,并且对隐私预算的敏感度更低,这验证了所提方法能够在隐私性与可用性之间实现有效的平衡。

本文第 2 节介绍国内外相关工作的研究进展;第 3 节介绍相关的背景知识以及对本文的研究问题进行定义;第 4 节详细阐述了面向图神经网络推荐系统的隐私保护框架 PF-GNR;第 5 节在真实数据集上对本文所提方法进行实验,并对实验结果进行了详细分析;第 6 节主要总结全文以及对未来进行展望。

2 相关工作

2.1 基于图神经网络的推荐系统

推荐系统中大多数信息可以自然表示成图数

据。近年来,GNN 在图数据表示学习方面展现出了卓越性能,越来越多的学者开始关注基于 GNN 的推荐系统。文献[2]提出了一种基于图神经网络的协同过滤推荐模型 NGCF,通过利用 GNN 对用户—项目交互图中的高阶连接信息进行建模,以增强用户和项目嵌入表示的准确性。在 NGCF 模型的基础上,文献[1]提出了轻量级图卷积网络模型 LightGCN,该模型去除了传统图卷积网络中非线性激活函数和特征变换等组件,使得模型的参数和计算复杂度大大减少,不仅可以实现更高的计算效率,还能显著提升模型的推荐性能。为支持在大规模图数据上进行应用,文献[3]借鉴 GAT[4] 方法的思想,提出一种基于图卷积网络的推荐算法 PinSage,该算法结合随机游走策略和图卷积操作生成节点的嵌入表示,能够在大规模图数据上实现高效、准确的图神经网络训练。文献[5]提出了一种用于社交推荐的增强型异构图神经网络模型 GL-HGNN,该模型旨在学习一个异构全局图,从中捕捉到高阶复杂的语义信息,以帮助提升社交推荐的准确性。以上研究表明,基于 GNN 的推荐方法能够充分发挥对图数据的建模优势,有效提升推荐性能。

2.2　基于图神经网络的隐私保护推荐系统

近年来,许多学者对图神经推荐系统中的隐私问题展开了研究,并提出了一些解决方案。这些方案依据所采用的隐私保护技术,可大致分为两类:基于信息模糊的方法和基于数据加密的方法。

基于信息模糊的方法是通过泛化、混淆、添加差分噪声等方式对真实数据进行隐私化处理,防止敌手从中窃取特定用户的敏感信息。典型的信息模糊技术包括 CDP、LDP 等。文献[12]提出了一种基于 CDP 的图神经推荐模型 GERAI,能够同时为用户的敏感特征和模型训练过程提供双重隐私保护。为保证训练过程中数据的隐私安全,文献[20]设计了一种隐私保护图神经网络算法 GAP,该算法利用高斯机制对参数的聚合过程进行随机扰动,使得训练好的 GNN 模型满足中心化差分隐私。基于 CDP 的方法具有计算效率高、部署简单等优点,但是仅适用于中心服务器完全可信的特定场景,因此在实际应用中具有较大的局限性。鉴于此,LDP 技术应运而生。基于 LDP 的方法将隐私保护过程从数据收集方转移到客户端,使得每个用户能够独立地对个人敏感信息进行隐私处理,以避免不可信第三方带来的隐私攻击。目前,LDP 技术主要应用于图神经推荐模型的本地保护策略。文献[21]提出了一种基于 GNN 的联邦推荐框架 FedGNN,在每轮迭代中参与方采用 LDP 技术对上传至服务器的梯度参数进行随机扰动,以防止服务器推断出用户的隐私数据。文献[17]针对链接推理攻击,提出了一种基于边本地化差分隐私的保护算法 LapGraph。该算法利用拉普拉斯机制在原始图数据上添加差分隐私噪声,以避免图中边的信息被攻击者恶意窃取。虽然基于 LDP 的方法能够提供严格且可量化的隐私保证,但是由于图数据的结构复杂且关联性强,直接对其添加噪声会极大破坏数据可用性,进而严重影响推荐系统的准确性。

基于数据加密的方法是利用密码学工具将数据转换为密文进行分析和处理,确保数据在传输、存储以及运算过程中的机密性和完整性。典型的加密技术包括同态加密、秘密共享等。文献[23]提出了一种去中心化联邦图神经网络 D-FedGNN,允许多个参与者在没有中心服务器的情况下训练图神经网络模型,并通过引入 Diffie-Hellman 密钥交换方法来实现客户端之间的安全模型聚合。针对云计算环境中的隐私泄露问题,文献[24]设计了一种轻量级隐私保护模型 SecGNN。该模型利用秘密共享技术构建一个安全多方计算协议,使得云服务器能够在不访问原始数据的情况下训练 GNN 模型,并提供安全的推理服务。虽然基于数据加密的方法能够在不损害模型准确性的前提下提供严格的隐私保护,但是由于引入了密文运算,普遍存在计算效率低、存储开销大等性能问题,目前还难以应用于复杂的机器学习模型。

综上,虽然目前针对图神经推荐系统隐私保护的研究已有不少,但仍有许多问题和挑战亟待解决,如严重依赖第三方可信度的假设、隐私保护方式单一、隐私性—可用性不平衡等,远未达到成熟的水平。

3　预备知识及问题描述

3.1　本地差分隐私

差分隐私[25] 是机器学习中常用的一种隐私保护技术,因其严格的数学定义和可量化的隐私保证,受到了学者们的广泛关注。其核心思想是:通过设计随机化算法,保证任意个体的数据无论是否在数据集中,对算法的输出结果几乎没有影响。差分隐私最初是为集中式学习而设计的,对于敏感信息的保护始终建立在可信第三方的基础之上,因此在实际应用中具有很大的局限性。为了解决这个问题,

LDP 技术[26]应运而生。在 LDP 模型中,每个用户独立地对上传服务器的数据进行随机化处理。由于隐私保护过程不再依赖第三方的介入,从而避免了服务器直接访问或接触原始数据,可以有效抵御不可信第三方带来的隐私攻击。

定义 1. ε-本地化差分隐私[27]。给定一个隐私算法 M 及其定义域 $Dom(M)$ 和值域 $Ran(M)$,若算法 M 在任意个体数据记录 t 和 $t'(t,t' \in Dom(M))$ 上得到相同的输出结果 $t^*(t^* \subseteq Ran(M))$ 满足下列不等式,则说明 M 满足 ε -本地化差分隐私。

$$\Pr[\mathcal{M}(t)=t^*] \leqslant e^{\varepsilon} \times \Pr[\mathcal{M}(t')=t^*] \quad (1)$$

其中,$\Pr(\cdot)$ 表示算法 M 的随机性,ε 为隐私预算,值越小表示算法 M 的隐私保护程度越高,而数据效用越低。

在图数据应用场景下使用 LDP 技术时,根据隐私保护对象的不同,主要分为:边本地化差分隐私[27]和节点本地化差分隐私[31]。边本地化差分隐私定义在只相差一条边的相邻图上,而节点本地化差分隐私则定义在相差一个节点及与该节点相连的所有边的相邻图上。本文主要是对用户-项目交互图中边的隐私保护问题开展研究,下面重点介绍与边本地化差分隐私的相关概念。

定义 2. 边-相邻图[17]。给定图 $G_1=(V_1,E_1)$,$G_2=(V_2,E_2)$,且 $V_1=V_2$。如果 G_1 可以通过在 G_2 中添加或删除一条边而得到,即满足 $|E_1 \oplus E_2|=1$,则称 G_1 和 G_2 边-相邻图,记为 $G_1 \sim G_2$。

定义 3. L_1-边敏感度[20]。给定一个函数 $f: G \to \mathcal{G}$,G_1 和 G_2 为任意两个边-相邻图,则函数 f 的 L_1-边敏感度可以表示如下:

$$\Delta f = \max_{G_1,G_2} \| f(G_1)-f(G_2) \|_1 \quad (2)$$

其中,$f(G_1)$ 和 $f(G_2)$ 分别代表函数 f 在图 G_1、G_2 上的输出,$\|\cdot\|_1$ 代表 L_1 范式。

定义 4. ε -边本地差分隐私[27]。给定一种隐私算法 M 及其定义域 $Dom(M)$ 和值域 $Ran(M)$。对于任意两个边-相邻图 G_1 和 G_2,若经过算法 M 处理后得到的输出结果 $S(S \subseteq Ran(M))$ 满足下列不等式,则隐私算法 M 满足 ε -本地化差分隐私。

$$\Pr[M(G_1)=S] \leqslant e^{\varepsilon} \Pr[M(G_2)=S] \quad (3)$$

定义 5. Laplace 机制[17]。设函数 $f: G \to G'$,其敏感度 Δf,若算法 M 的输出结果满足下列等式,则称算法 M 满足 ε -差分隐私。

$$M(G)=f(G)+Lap(\Delta f/\varepsilon) \quad (4)$$

其中,$Lap(\Delta f/\varepsilon)$ 是添加的随机噪声,它服从期望值为 0,尺度参数为 $\Delta f/\varepsilon$ 的拉普拉斯分布。

性质 1. 序列组合性[20]。给定数据集 D 和 n 个隐私算法 $\{M_1,M_2,\cdots,M_n\}$,每个算法 $M_i(1 \leqslant i \leqslant n)$ 分别作用于数据集 D 且满足 ε_i-本地化差分隐私,则这 n 个隐私算法在数据集 D 上构成的序列组合 M,满足 $\sum_{i=1}^{n} \varepsilon_i$ -本地化差分隐私。

性质 2. 并行组合性[32]。将数据集 D 分成 n 个不相交的子集合 $\{D_i\}(i \in \{1,2,\cdots,n\})$,每个子集 D_i 分别作用于一个随机算法 M_i。如果 M_i 满足 ε_i-本地化差分隐私,则 $\{M_1,M_2,\cdots,M_n\}$ 在 D 上的构成的并行组合 M,满足 ε-本地化差分隐私,其中,$\varepsilon=\max(\varepsilon_i)$。

3.2 同态加密

同态加密是一种非对称加密系统,允许第三方对密文进行特定代数运算并得到加密形式的计算结果,且解密后与其在明文上进行运算得到的结果是一致的。同态加密技术使得在密文执行计算成为可能,从而可以保证数据在存储以及运算过程中的隐私性和完整性。

定义 6. 同态性[33]。给定加密算法 Enc 和明文域 P_m 上的运算。,若对于 $\forall m_1,m_2 \in P_m$ 都满足下列不等式,则称加密算法 Enc 具有同态性。

$$Dec(k_s,Enc(k_p,m_1) \diamondsuit Enc(k_p,m_2))=m_1 \circ m_2 \quad (5)$$

其中,k_s、k_p 分别代表私钥与公钥,$Dec(\cdot)$ 是解密算法,$\diamondsuit$ 为密文域上的运算。

定义 7. 部分同态加密[33]。对于加密算法 Enc_e 和明文域 P_e 上的运算 $(+,\times)$,若 $\forall p_1,p_2,\cdots,p_n \in P_e$ 仅满足加法或者乘法运算在式(5)中成立,则称加密算法 Enc_ε 为满足部分同态的加密算法。

3.3 链接窃取攻击

链接窃取攻击是指攻击者试图获取图数据中目标节点对之间敏感链接的恶意活动。以图神经推荐系统为例,用户-项目交互图中的链接通常对应于节点(即用户和项目)之间的关联信息,例如评分、点击、购买等。如果这些敏感的链接关系被恶意攻击者推断出来,可能会泄露用户的兴趣、行为等隐私信息,使用户面临巨大的隐私风险。

在链接窃取攻击中,攻击者通常利用 GNN 模型的某些特性或漏洞来实施攻击,主要的攻击方法包括:利用后验概率进行攻击、结合文本特征进行攻击以及跨数据集进行攻击等。例如,Wu 等人提出的链接攻击模型 LinkTeller,无需任何背景知识,仅

通过访问 GNN 模型获取节点的后验概率信息,然后比较不同节点之间后验概率的相似性,则可推断出目标攻击对象之间的链接关系。

3.4　问题描述

本文提出的方法采用客户－服务器架构,主要由一个中央服务器和大量客户端组成。各客户端保存用户的本地数据且独立地参与模型训练。中央服务器则负责收集用户数据并训练 GNN 模型,进而预测用户潜在的兴趣或偏好行为。假设推荐系统的用户集合为 $U=\{u_1,\cdots,u_m\}(|U|=m)$,项目集合为 $V=\{v_1,v_2,\cdots,v_n\}(|V|=n)$。用户 u_i 的历史交互信息以图数据 $G_i=(N_{u_i},E_{u_i})$ 的形式保存在本地客户端。其中,N_{u_i} 表示与用户 u_i 相关联的项目节点集合,E_{u_i} 是边集合。本地子图 G_i 的结构信息采用邻接表 $A_i=[a_{i,1},\cdots,a_{i,n}]$ 来表示,如果用户 u_i 和项目 v_j 之间存在交互行为,则 $a_{i,j}=1$,否则 $a_{i,j}=0$。

随着云计算技术的发展,许多企业选择将图神经网络推荐系统部署到云平台,并利用云服务器来完成用户数据的收集、存储、分析等一系列的工作。然而,当云平台的可信程度较低时,用户数据可能会遭受不可信的云服务提供商或者其他攻击者恶意窃取,从而导致用户的隐私泄露。本文的目标是设计一种面向图神经推荐系统的隐私保护方案,该方案能够有效抵御不可信服务器的隐私攻击,实现为用户提供严格隐私保护的同时,保证模型的可用性。

4　隐私保护图神经网络推荐方法

4.1　PF-GNR 框架概述

为了解决不可信环境中图神经网络推荐系统面临的隐私问题,本文提出了一种基于 GNN 的隐私保护推荐框架 PF-GNR。该框架将本地化差分隐私和同态加密技术相结合,设计了一种两阶段的混合隐私保护机制 HPM,在无可信第三方的情况下,能够为用户本地数据提供严格的隐私保护,并同时保持模型的可用性。PF-GNR 主要包含 3 个部分:(i)本地隐私编码模块:该模块部署在本地用户端,每个用户可以利用 LDP 技术对原始数据进行随机扰动和编码,以确保数据采集过程中用户数据的隐私安全;(ii)图神经网络推荐模块:该模块部署在中心服务器上,利用 GNN 学习节点的嵌入表示,进而对用户和项目之间的偏好关系进行建模;(iii)隐私保护更新模块:在训练阶段,该模块使用同态加密技术为各参与方的数据交换和运算提供保护,防止中心服务器在训练过程中窃取参与方的数据信息,从而保证模型更新的安全性和隐私性。

PF-GNR 的总体结构如图 1 所示,主要由一个中心服务器和多个客户端组成,每个客户端上保存了用户的交互数据(即本地子图 G_i)。PF-GNR 的工作流程可分为三个阶段:

(1)本地隐私编码模块:在数据收集过程中,每个用户 u_i 首先利用 LDP 技术对其邻接表 A_i 和度信息 D_i(与用户 u_i 相关联的项目数量)进行随机扰动。然后,对 $\widetilde{A}_i$ 进行重新编码,并将其上传至中心服务器 S_c;

(2)图神经推荐模块:中心服务器 S_c 根据收到的扰动数据 $\{\widetilde{A}_1,\cdots,\widetilde{A}_m\}$,构建用户－项目交互图 G,并将其作为图神经推荐模块的输入,用于学习节点的嵌入表示,从而生成用户与项目之间的偏好分数 $\{\hat{R}_1,\cdots,\hat{R}_m\}$;

(3)隐私保护更新模块:模型训练过程中,参与方 u_i 首先计算其本地损失函数 $\mathcal{L}_{u_i}$ 并对其进行加密 $E(\mathcal{L}_{u_i})$,然后发送给第三方服务器 S_a;其次,S_a 利用同态加密机制 Paillier[33] 对接收的密文执行加法同态运算,并将加密运算的结果 $E(\mathcal{L}_{PF-GNR})$ 上传给中心服务器 S_c;最后,中心服务器 S_c 对密文 $E(\mathcal{L}_{PF-GNR})$ 进行解密,得到模型的损失函数 $\mathcal{L}_{PF-GNR}$,并据此更新模型参数。

在当前迭代次数 t 未达到最大迭代次数 T 之前,将依次循环执行阶段 2 和阶段 3,直到模型收敛。接下来,对 PF-GNR 的 3 个模块进行详细介绍。

4.2　本地隐私编码模块

本地隐私编码模块是基于 LDP 技术实现的,用来保证用户数据在收集过程中的隐私安全。由于每个用户能够独立地利用该模块对本地数据进行隐私化处理,不再依赖可信第三方的介入,因此可有效抵御中心服务器或其他敌手发起的隐私攻击。该模块包括以下步骤:

步骤 1:对用户的本地数据进行随机扰动。本文中,每个用户 u_i 的交互信息都是以图 G_i 的形式保存在本地。因此,需要保护的对象包括两个:一是用户 u_i 的邻接表 A_i,二是用户 u_i 的度数信息 D_i(与用户 u_i 相关联的项目数量 $|N_{u_i}|$)。本文采用拉普拉斯机制分别向邻接表 A_i 和度数信息 D_i 添加噪声扰动,使其受到本地差分隐私的保护。假设总隐私预算为 ε,隐私分配参数为 δ。这里使用 δ 将 ε

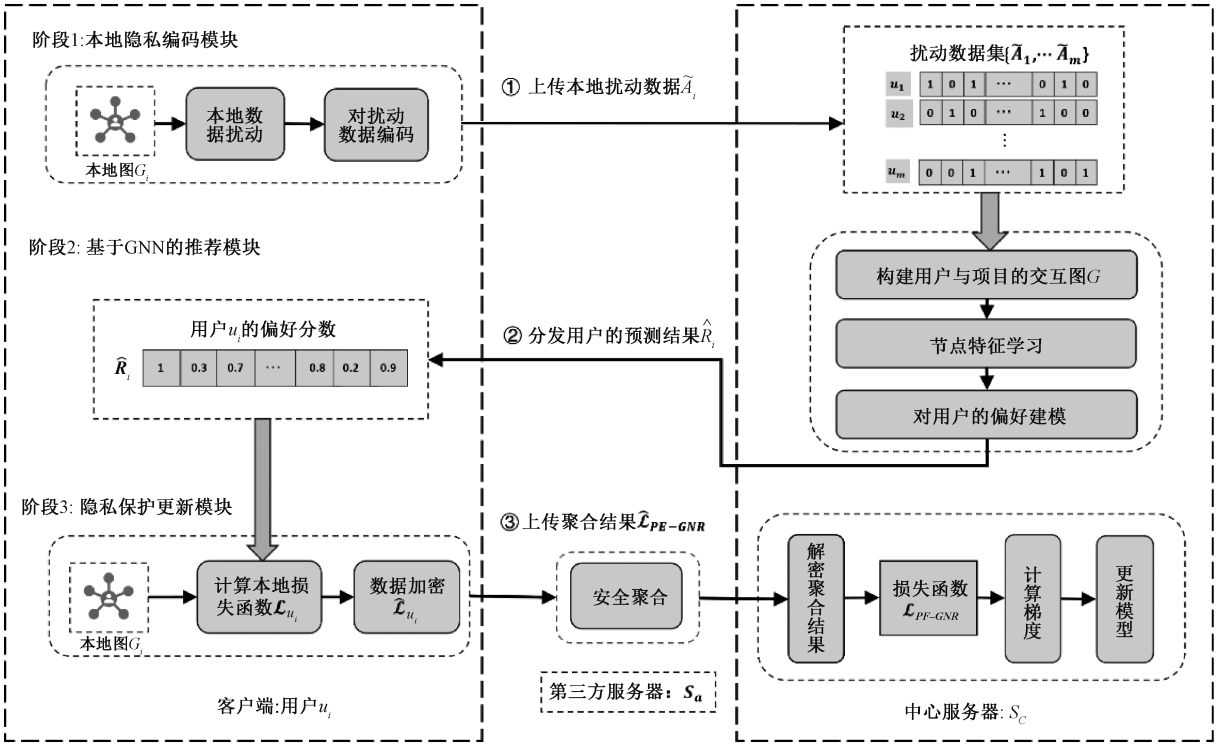

图 1 PF-GNR 框架的整体结构图

分成两个部分：$\varepsilon_a=\delta\varepsilon$ 和 $\varepsilon_d=(1-\delta)\varepsilon$，$\varepsilon_a$ 和 ε_d 将分别用于对邻接表 A_i 和度 D_i 进行扰动，则噪声扰动的结果如下：

$$\widetilde{A}_i=A_i+\left(Lap_1\left(\frac{\Delta f_a}{\varepsilon_a}\right),\cdots,Lap_n\left(\frac{\Delta f_a}{\varepsilon_a}\right)\right) \tag{6}$$

$$\widetilde{D}_i=|N_{u_i}|+Lap\left(\frac{\Delta f_d}{\varepsilon_d}\right) \tag{7}$$

其中，$Lap_j(\Delta f_a/\varepsilon_a)(1\leqslant j\leqslant n)$ 表示服从期望为 0，位置参数为 $\frac{\Delta f_a}{\varepsilon_a}$ 的拉普拉斯分布噪声分量；$Lap\left(\frac{\Delta f_d}{\varepsilon_d}\right)$ 是服从期望为 0，位置参数为 $\frac{\Delta f_d}{\varepsilon_d}$ 的拉普拉斯分布噪声；Δf_a、Δf_d 分别代表敏感度，其大小均为 1。

步骤 2：对扰动数据进行编码。由于每个用户 u_i 的邻接表 A_i 通常是高维度且稀疏的，直接对其添加噪声扰动，会严重损害其可用性[18]。因此，本文考虑利用 $\widetilde{D}_i$ 对扰动后的邻接表 $\widetilde{A}_i$ 进行重新编码，以提高数据的可用性。具体而言，首先从扰动后的邻接表 $\widetilde{A}_i$ 中筛选出前 $|\widetilde{D}_i|$ 个最大的记录；然后，将它们对应位置上的记录值全部设置为 1，而其余位置则设置为 0；最后，将经过重新编码得到的 $\widetilde{A}_i$ 发送给中心服务器。这样，经过编码后得到的邻接表 $\widetilde{A}_i$，会更加接近原始数据 A_i，并且还能保持相似的稀疏度。本地隐私编码模块的详细描述如算法 1 所示。

算法 1. 基于 LDP 的本地隐私编码算法

输入：用户 u_i 的邻接表 $A_i=[a_{i,1},\cdots,a_{i,n}]$，总隐私预算 ε，隐私分配参数 δ

输出：扰动后的邻接表 $\widetilde{A}_i$

1. $\varepsilon_a\leftarrow\delta\varepsilon$，$\varepsilon_d\leftarrow(1-\delta)\varepsilon$；
2. FOR $j\in\{1,\cdots,n\}$ DO
3. $\quad\widetilde{a}_{i,j}\leftarrow a_{i,j}+Lap\left(\frac{\Delta f_a}{\varepsilon_a}\right)$；
4. END FOR
5. $\widetilde{D}_i=\sum_{j=1}^{n}a_{ij}$；
6. $\widetilde{D}_i\leftarrow D_i+Lap\left(\frac{\Delta f_d}{\varepsilon_d}\right)$；
7. 将 $\widetilde{A}_i$ 中所有元素的值设置为 0；
8. $T_{index}\leftarrow\widetilde{A}_i$ 中前 $\lfloor\widetilde{D}_i\rfloor$ 个最大元素的索引集合；
9. FOR each $j\in T_{index}$ DO
10. $\quad\widetilde{a}_{i,j}=1$；
11. END FOR
12. RETURN $\widetilde{A}_i$

算法 1 能够为用户的本地数据提供严格且可量化的隐私保护效果，其隐私性分析见定理 1。

定理 1. 对于给定的总隐私预算 ε 和隐私分配参数 δ，算法 1 满足 ε-边本地化差分隐私。

证明. 假设图 G_u 和 G'_u 是用户 u 的两个边一相邻图,它们的邻接表分别为 $S_u=[s_{u,1},\cdots,s_{u,n}]$,$S'_u=[s'_{u,1},\cdots,s'_{u,n}]$。根据定义 2,$S_u$ 与 S'_u 的汉明距离最大为 $\Delta f=1$,即 S_u、S'_u 在对应位置上不相同的元素仅有 1 个。不失一般性,假设 S_u 和 S'_u 在第 q 个位置上的元素不同,即 $|s_{u,q}-s'_{u,q}|_1=1$。

为便于表述,本文将算法 1 表示为机制 $\mathcal{M}(S_u,D_u)=(\widetilde{S}_u,\widetilde{D}_u)$,其中 $D_u=\sum_{j=1}^{n}s_{u,j}$;将算法 1 中第 2－4 行对邻接表的噪声扰动过程表示为机制 $\mathcal{R}_1(S_u)=\widetilde{S}_u$,将第 6 行对度数的随机扰动表示为机制 $\mathcal{R}_2(D_u)=\widetilde{D}_u$。其中,$\mathcal{R}_1$ 的隐私预算为 $\varepsilon_a=\delta\varepsilon$,$\mathcal{R}_2$ 的隐私预算为 $\varepsilon_d=(1-\delta)\varepsilon$。则:

$$\frac{\Pr[\mathcal{M}(S_u,D_u)=(\widetilde{S}_u,\widetilde{D}_u)]}{\Pr[\mathcal{M}(S'_u,D'_u)=(\widetilde{S}_u,\widetilde{D}_u)]}$$

$$=\frac{\Pr[\mathcal{R}_1(S_u)=\widetilde{S}_u]\Pr[\mathcal{R}_2(D_u)=\widetilde{D}_u]}{\Pr[\mathcal{R}_1(S'_u)=\widetilde{S}_u]\Pr[\mathcal{R}_2(D'_u)=\widetilde{D}_u]} \tag{8}$$

$$=\frac{\prod_{p=1}^{n}\Pr[\mathcal{R}_1(s_{u,p})=\widetilde{s}_{u,p}]}{\prod_{p=1}^{n}\Pr[\mathcal{R}_1(s'_{u,p})=\widetilde{s}_{u,p}]}\cdot\frac{\Pr[\mathcal{R}_2(D_u)=\widetilde{D}_u]}{\Pr[\mathcal{R}_2(D'_u)=\widetilde{D}_u]} \tag{9}$$

$$=\frac{\Pr[\mathcal{R}_1(s_{u,q})=\widetilde{s}_{u,q}]}{\Pr[\mathcal{R}_1(s'_{u,q})=\widetilde{s}_{u,q}]}\cdot\frac{\Pr[\mathcal{R}_2(D_u)=\widetilde{D}_u]}{\Pr[\mathcal{R}_2(D'_u)=\widetilde{D}_u]} \tag{10}$$

$$=\frac{\frac{\varepsilon_a}{2}\cdot\exp\left[-\frac{|s_{u,q}-\widetilde{s}_{u,q}|}{\Delta f_a}\cdot\varepsilon_a\right]}{\frac{\varepsilon_a}{2}\cdot\exp\left[-\frac{|s'_{u,q}-\widetilde{s}_{u,q}|}{\Delta f_a}\cdot\varepsilon_a\right]}\cdot\frac{\frac{\varepsilon_d}{2}\cdot\exp\left[-\frac{|D_u-\widetilde{D}_u|}{\Delta f_d}\cdot\varepsilon_d\right]}{\frac{\varepsilon_d}{2}\cdot\exp\left[-\frac{|D'_u-\widetilde{D}_u|}{\Delta f_d}\cdot\varepsilon_d\right]} \tag{11}$$

$$=\exp(-\varepsilon_a(|s_{u,q}-\widetilde{s}_{u,q}|-|s'_{u,q}-\widetilde{s}_{u,q}|))\cdot\exp(-\varepsilon_d(|D_u-\widetilde{D}_u|-|D'_u-\widetilde{D}_u|)) \tag{12}$$

$$\leqslant\exp(\varepsilon_a|s_{u,q}-s'_{u,q}|)\cdot\exp(\varepsilon_d|D_u-D'|) \tag{13}$$

$$=\exp(\varepsilon_a)\cdot\exp(\varepsilon_d) \tag{14}$$

$$=\exp(\varepsilon_a+\varepsilon_d)=\exp(\varepsilon) \tag{15}$$

其中,根据独立性原理可得式(8)和(9);根据已知条件,当 $p\neq q$ 时,$s_{u,p}=s'_{u,q}$,通过化简可得式(10);根据定义 5,由拉普拉斯机制变换可得式(11);根据定义 2,$|s_{u,q}-s'_{u,q}|_1=1$ 和 $|D_u-D'_u|_1=1$,可得公式(14)。　　证毕。

4.3　基于 GNN 的推荐模块

基于 GNN 的推荐模块部署在中心服务器上,利用消息传播机制来学习每个节点的嵌入表示,进而生成用户对项目的偏好分数。该模块包含 3 个部分:用户－项目交互图的构建、基于图的特征学习以及预测用户与项目的偏好分数。

4.3.1　用户－项目交互图的构建

中心服务器首先根据各用户上传的扰动数据,构建全局的用户－项目交互图 $G=(\mathcal{N},\varepsilon)$。其中,$\mathcal{N}=U\cup V$ 代表由用户和项目组成的节点集合;ε 代表边集合。如果边 $(u_i,v_j)\in\varepsilon$,表示用户 u_i 与项目 v_j 之间存在关联关系,否则边 $(u_i,v_j)\notin\varepsilon$。图 G 中节点之间的关联关系可以表示成交互矩阵 $\widetilde{\boldsymbol{A}}=[\widetilde{a}_{i,j}]_{m\times n}$,则 $\widetilde{\boldsymbol{A}}$ 中每个元素 $\widetilde{a}_{i,j}$ 的取值定义:

$$\widetilde{a}_{i,j}=\begin{cases}1, & (u_i,v_j)\in E\\ 0, & (u_i,v_j)\notin E\end{cases} \tag{16}$$

4.3.2　基于图的节点特征学习

受文献[1]的启发,本文将 LightGCN 作为基础网络模型来学习节点的嵌入表示。下面以用户－项目对 $\langle u_i,v_j\rangle$ 为例,介绍节点特征的学习过程。

首先,中心服务器对模型的参数进行初始化,分别得到用户节点和项目节点的初始嵌入矩阵 $\boldsymbol{E}_U=[e_{u_1},e_{u_2},\cdots,e_{u_m}]^{\mathrm{T}}$ 和 $\boldsymbol{E}_V=[e_{v_1},e_{v_2},\cdots,e_{v_n}]^{\mathrm{T}}$,其中,$e_{u_i}$、$e_{v_j}\in\mathbb{R}^d$ 表示用户 u_i 和项目 v_j 的初始嵌入向量,d 代表嵌入维度。本文中,$\boldsymbol{E}_U$ 和 $\boldsymbol{E}_V$ 是可学习的模型参数,由中心服务器负责维护,并以端到端的方式进行优化。

其次,将节点的初始嵌入矩阵 $\boldsymbol{E}_U$、$\boldsymbol{E}_V$ 以及交互矩阵 $\widetilde{\boldsymbol{A}}$ 作为输入,采用 LightGCN 的图卷积来学习各层节点的嵌入向量,则第 l 层图卷积运算的定义如下:

$$\begin{cases}\boldsymbol{z}_{u_i}^{(l)}=\sum\limits_{v_j\in N_{u_i}}\dfrac{1}{\sqrt{|N_{u_i}|}\sqrt{|N_{v_j}|}}z_{v_j}^{(l-1)}\\ \boldsymbol{z}_{v_j}^{(l)}=\sum\limits_{u_i\in N_{v_j}}\dfrac{1}{\sqrt{|N_{v_j}|}\sqrt{|N_{u_i}|}}z_{u_i}^{(l-1)}\end{cases} \tag{17}$$

其中,$\boldsymbol{z}_{u_i}^{(l)}$、$\boldsymbol{z}_{v_j}^{(l)}$ 分别表示第 l 层的用户 u_i 和项目 v_j 的嵌入向量,且 $\boldsymbol{z}_{u_i}^{(0)}=e_{u_i}$、$\boldsymbol{z}_{v_j}^{(0)}=e_{v_j}$;$N_{ui}$ 和 N_{vj} 分别代表用户 u_i 和项目 v_j 的邻居节点集合。

最后,在完成全部 L 层的图卷积运算后,可以分别得到用户 u_i 和项目 v_j 在各层的嵌入向量,这

里将其记为 $\{\boldsymbol{z}_{u_i}^{(0)},\dots,\boldsymbol{z}_{u_i}^{(L)}\}$ 和 $\{\boldsymbol{z}_{v_j}^{(0)},\dots,\boldsymbol{z}_{v_j}^{(L)}\}$。为完整表达每个图卷积层所传递的高阶连接信息，本文将各层输出的嵌入信息进行融合，作为节点的全局嵌入表示：

$$\begin{cases}\boldsymbol{z}_{u_i}^{*}=\alpha_0\boldsymbol{z}_{u_i}^{(0)}+\alpha_1\boldsymbol{z}_{u_i}^{(1)}+\dots+\alpha_L\boldsymbol{z}_{u_i}^{(L)}\\ \boldsymbol{z}_{v_j}^{*}=\alpha_0\boldsymbol{z}_{u_i}^{(0)}+\alpha_1\boldsymbol{z}_{u_i}^{(1)}+\dots+\alpha_L\boldsymbol{z}_{u_i}^{(L)}\end{cases} \tag{18}$$

其中，$\boldsymbol{z}_{u_i}^{*}$、$\boldsymbol{z}_{v_j}^{*}$ 表示用户 u_i 和项目 v_j 的全局嵌入向量；α_k 表示第 l 层嵌入的权重参数；L 表示 GNN 的最大层数。

4.3.3 预测偏好分数

通过以上步骤获得所有节点的全局嵌入向量后，我们以内积作为预测函数，来计算每个用户 u_i 对项目 v_j 的偏好分数：

$$\hat{r}_{i,j}=\boldsymbol{z}_{u_i}^{*\mathrm{T}}\times\boldsymbol{z}_{v_j}^{*} \tag{19}$$

4.4 隐私保护更新模块

由于在基于 GNN 的推荐模块中，中心服务器仅能访问扰动后的用户数据，如直接使用这些扰动数据进行训练，将会严重影响模型的准确性。因此，为了降低噪声扰动带来的负面影响，本文采用客户－服务器的协作方式进行模型训练。具体而言，在每次迭代中，中心服务器 S_c 使用分批的方式随机选择部分本地用户参与训练，同时将预测结果分享给各参与方；其次，每个参与方 u_i 根据接收到的预测结果 $\hat{R}_i$ 来计算其本地损失函数 $\mathcal{L}_{u_i}$（中间结果），并将 $\mathcal{L}_{u_i}$ 上传给中心服务器 S_c；最后，中心服务器 S_c 收到各参与方发送的中间结果后进行聚合计算，获得完整的损失函数 $\mathcal{L}_{PF-GNR}$，并以此更新模型参数，从而完成本次迭代。

虽然采用上述方式训练模型，可有效降低差分噪声带来的负面影响，提高模型的可用性，但由于训练过程中需要交换中间参数，使得各参与方的数据面临着隐私泄露的风险。这是因为参与方 u_i 上传的中间结果 $\mathcal{L}_{u_i}$ 与其原始数据 A_i 直接相关，中心服务器或其他敌手能够利用 $\mathcal{L}_{u_i}$ 发起攻击，推断某条数据记录是否与目标用户 u_i 有关联，从而导致用户的隐私泄露。为解决该问题，本文采用同态加密机制 Paillier 为各参与方上传的中间结果提供保护，以确保模型训练过程的隐私性。隐私保护更新模块主要包括以下步骤。

步骤 1：每一轮迭代中，中心服务器 S_c 首先根据给定的安全参数 λ 和 τ，生成公钥和私钥的密钥对 (sk,pk)。然后，采用分批的方式随机选择部分本地用户参与本轮训练，同时将公钥 pk 和预测结果 $\hat{R}$ 分发给各参与方。密钥对的生成过程如下：

$$(sk,pk)=KeyGen(1^{\lambda},1^{\tau}) \tag{20}$$

其中，$KeyGen(\cdot)$ 是密钥生成函数。

步骤 2：训练开始后，每个参与方 u_i 根据接收到的预测结果 $\hat{R}_i$，计算其本地损失函数。本文根据推荐任务的要求，采用贝叶斯个性化排名损失函数 BPRLoss[34] 作为模型的学习目标，因此，用户 u_i 的本地损失函数可以定义如下：

$$\mathcal{L}_{u_i}=-\sum_{p\in N_{u_i}}\sum_{q\notin N_{u_i}}In\sigma(\hat{r}_{i,p}-\hat{r}_{i,q}) \tag{21}$$

其中，N_{u_i} 表示与用户 u_i 相关联的项目集合。用户 u_i 在完成 $\mathcal{L}_{u_i}$ 的计算后，使用公钥 pk 进行加密，并将密文 $E(\mathcal{L}_{u_i})$ 发送至第三方服务器 S_a：

$$E(\mathcal{L}_{u_i})=Enc(\mathcal{L}_{u_i},pk) \tag{22}$$

其中，$Enc(\cdot)$ 是加密运算。

步骤 3：安全聚合。第三方服务器 S_a 收到各方发送的密文后，利用 Paillier 算法对密文执行加法同态运算，并将加密运算的结果上传给中心服务器 S_c。该加密运算的过程如下：

$$E(\mathcal{L}_{PF-GNR})=\Pi(E(\mathcal{L}_{u_1}),\cdots,E(\mathcal{L}_{u_B})) \tag{23}$$

其中，Π 表示 Paillier 机制，B 表示批量大小。

步骤 4：更新模型参数。中心服务器 S_c 使用私钥 sk 对 $E(\mathcal{L}_{PF-GNR})$ 进行解密，得到对应的明文信息，即完整的模型目标函数：

$$\mathcal{L}_{PF-GNR}=Dec(E(\mathcal{L}_{PF-GNR}))+\lambda\|\Theta\|_2 \tag{24}$$

其中，$Dec(\cdot)$ 是解密运算；Θ 为模型的参数；λ 为正则化系数，用于防止模型过拟合。最后，本文使用 Adam[35] 优化器对目标函数 $\mathcal{L}_{PF-GNR}$ 进行优化，更新模型的参数。该模块的伪代码如算法 2 所示。

算法 2. 隐私保护模型更新算法

输入：用户集合 U，用户的本地数据集合 $D=\{A_i\}_{i=1}^{m}$，安全参数 λ 和 τ

输出：模型损失函数 $\mathcal{L}_{PF-GNR}$

1. 中心服务器 S_c 从用户集 U 中随机选择一个子集 U_S；
2. 中心服务器 S_c 生成本轮训练的密钥对 $(sk_t,\ pk_t)$；
3. FOR each $u_i\in U_S$ DO
4. 根据公式(21)计算本地损失函数 $\mathcal{L}_{u_i}$；
5. 根据公式(22)对本地损失函数 $\mathcal{L}_{u_i}$ 加密；
6. 用户 u_i 将密文 $\mathcal{L}_{u_i}$ 发送给第三方服务器 S_a；
7. END FOR
8. $E(\mathcal{L}_{u_i})\leftarrow$ 第三方服务器 S_a 对密文执行公式(23)；
9. $\mathcal{L}_{PF-GNR}\leftarrow$ 中心服务器 S_c 执行公式(24)解密；

10. RETURN $\mathcal{L}_{PF-GNR}$

综合上述各小节的内容,下面给出 PF-GNR 框架的完整训练过程,如算法 3 所示。

算法 3. PF-GNR

输入:用户集 U,项目集 V,本地数据集 $D=\{A_i\}_{i=1}^m$,最大迭代次数 T,隐私预算 ε,隐私分配参数 δ

输出:推荐列表 $U_{Rec\ List}$

1. //客户端
2. FOR each $u_i \in U$ DO
3. $\widetilde{A}_i \leftarrow$ 用户 u_i 根据算法 1 对 A_i 进行隐私处理;
4. 用户 u_i 将 $\widetilde{A}_i$ 上传至中心服务器 S_c;
5. END FOR
6. //服务器端
7. S_c 接收用户发送的扰动数据 $\{\widetilde{A}_1,\widetilde{A}_2,\cdots,\widetilde{A}_n\}$,构建全局用户-项目交互图 G
8. 随机初始化用户和项目的嵌入矩阵 E_U,E_V
9. FOR $t=1$to T DO
10. 采用 GNN 推荐模块学习节点的嵌入,并生成当前用户对项目的偏好分数 $\hat{R}$;
11. 根据算法 2 获得模型的损失函数 $\mathcal{L}_{PF-GNR}$;
12. 使用 Adam 优化器更新 Z_U^*、Z_V^*;
13. END FOR
14. $\boldsymbol{U}_{RecList}=Z_U^* \cdot {Z_V^*}^{\mathrm{T}}$
15. RETURN $\boldsymbol{U}_{Rec\ List}$

4.5 算法的理论分析

4.5.1 隐私性分析

本文为 PF-GNR 框架设计了一种两阶段的混合隐私保护机制。该机制由本地隐私编码模块和隐私保护更新模块组成,能够在不可信环境中,为用户的本地数据提供严格的隐私保护。其隐私性分析如下:

首先,本地隐私编码模块为数据采集过程提供了 ε-边本地差分隐私保护。在数据采集过程中,每个用户是独立地利用算法 1 对本地数据进行隐私化处理,由于不再依赖于中心服务器,因此能够有效抵御不可信第三方带来的隐私攻击。同时,由于差分隐私的后处理特性[20],用户数据经过 LDP 保护后,可以在接下来的学习过程中得到持续的保护。此外,4.2 小节中的定理 1 也从理论上严格证明了算法 1 满足 ε -边本地差分隐私。

其次,隐私保护更新模块为模型的训练过程提供了强有力的隐私保证。从算法 2 可知,在每一轮迭代中,各参与方首先利用同态加密技术对上传的中间结果进行加密。接着,第三方服务器利用 Paillier 算法执行加法同态运算。最后,中心服务器在加密结果上完成模型训练。由于整个训练过程中不存在明文的传递,中心服务器仅能获得加密后的运算结果,可有效防止其窃取参与方的隐私信息,从而确保模型的安全性和隐私性。

综上所述,本文通过将本地化差分隐私和同态加密技术相结合,形成了对图神经推荐系统从本地数据收集到模型训练的全面隐私保护。

4.5.2 时间复杂度分析

PF-GNR 的时间复杂度主要由 3 个阶段组成:本地隐私编码模块、图神经推荐模块和隐私保护更新模块。本小节从各个阶段的角度出发,对本文方法的时间复杂度进行分析。

(1)第 1 阶段:每个用户利用算法 1 中的公式(6)和(7)分别向其邻接表和度数添加噪声,现总共有 m 个用户,则算法 1 的时间复杂度为 $O_1=O(m\times n)$;

(2)第 2 阶段:中心服务器利用公式(17)对每个节点的嵌入表示进行迭代更新。设用户-项目交互图的邻接矩阵为 $\widetilde{A}$,节点的嵌入向量矩阵为 $Z\in\mathbb{R}^{(m+n)\times d}$,则单层 GNN 模型的时间复杂度为 $O(\widetilde{A}\cdot Z)$。因此,对于一个深度为 L 层的 GNN 模型,其时间复杂度为 $O_2=O(L\cdot\widetilde{A}\cdot Z)$。

(3)第 3 阶段:PF-GNR 采用算法 2 来训练模型,主要包含以下步骤:首先,每轮迭代中需随机选择 B 个用户参与训练且各参与方根据公式(21)和(22)进行计算,其时间复杂度为 $O_{Enc}=O(B\cdot T_{Enc})$,$T_{Enc}$ 表示加密时间开销;其次,第三方服务器 S_a 执行公式(23)完成加密运算,其时间复杂度为 $O_{HE}=O(\Pi(B))$;最后,中心服务器 S_c 根据公式(24)解密,其时间复杂度为 $O_{Dec}=O(T_{Dec})$,T_{Dec} 表示解密时间开销。因此,单次迭代的时间复杂度为 $O_3=O_{Enc}+O_{HE}+O_{Dec}$。综上,假设总共执行了 T 次迭代训练,则 PF-GNR 的整体时间复杂度为 $O_{PF\text{-}GNR}=(O_1+T\cdot(O_2+O_3))$。

通过以上分析可知,与传统的 GNN 推荐方法相比,本文方法由于应用了 LDP 和同态加密技术,带来了更多额外的计算开销,从而导致模型训练时间的成本增加。针对这一问题,本文采取了多种优化措施以提高模型的训练效率,降低时间复杂度。具体而言,对于因 LDP 技术造成的开销,可以利用多方的并行性来实现效率的提升;对于因同态加密

机制造成的开销，通过采用分批的方式执行同态加密运算，可以将时间开销由 $O(m)$ 降低为 $O(B)$，其中 B 为批次大小。

总之，通过上述措施，可以有效降低本文方法的时间复杂度，提高模型训练的可行性。本文第 5.4 节，通过实验也进一步验证了 PF-GNR 在隐私保护和响应时间之间达到了较好的权衡。

4.5.3 空间复杂度分析

PF-GNR 算法的空间复杂度主要包括以下两个方面：

(1)本地客户端：训练前，用户端需要存储本地交互数据，其空间复杂度为 $O(n)$。另外，在训练过程中，每个用户还需要生成中间计算结果，设每个结果占用的内存大小为 f，其空间复杂度为 $O(f)$。因此，用户端的空间复杂度为 $O(n+f)$。

(2)服务器端：中心服务器的存储开销包括两个部分。首先，需要保存全局的用户－项目交互图和全局的模型参数(即节点的嵌入矩阵)，其空间复杂度为 $O(\varepsilon+(m+n)\times d)$；其次，训练过程中，GNN 会生成高阶嵌入表示，设每一层输出的维度与输入相同，那么 L 层 GNN 的空间复杂度为 $O(L\cdot(m+n)\cdot d)$。因此，服务器端的空间复杂度为 $O(\varepsilon+(L+1)\cdot(m+n)\times d))$。

通过上述分析可知，本文方法的隐私保护机制主要作用于客户端，因此其空间复杂度与传统的 GNN 推荐模型相当。

5 实验与结果分析

为了评价本文提出方法的有效性和实用性，在三个真实数据集上进行了多组实验。通过实验分析，本文旨在回答以下研究问题。

RQ1：与典型的推荐方法相比，PF-GNR 的推荐性能如何？

RQ2：与现有隐私保护方法相比，PF-GNR 在隐私－效用的均衡性上表现如何？

RQ3：PF-GNR 是否具有良好的实用性？

RQ4：通过消融实验来检验 PF-GNR 中主要组成部分的有效性？

RQ5：检验超参数对 PF-GNR 性能的影响？

下面，首先介绍实验设置，然后依次回答上述研究问题。

5.1 实验设置

5.1.1 数据集

为了验证本文方法的有效性，我们在 3 个公开真实的基准数据集上开展实验，分别是：MovieLens-1M[36]、Gowalla[2] 和 Yelp[1]。MovieLens 是由 Grouplens 研究小组发布的真实电影推荐数据集，记录了用户对电影的评分以及相关的属性信息。根据参考文献[36]的处理方式，本文将评分信息转化为隐式反馈信息用于模型学习。Gowalla 是一个基于位置服务和社交网络分析的数据集，收集了用户在 Gowalla 社交网络平台上的位置签到记录和社交关系数据。Yelp 是一个常用的商业评论数据集。该数据集包含了用户对商家的评论、评分、信任关系等，其中评论和评分信息可以用来分析用户的偏好及特征。这三个数据集分别代表了不同的应用场景以及数据规模，非常适合评估推荐算法的有效性。

为了与现有工作保持一致，实验中所有数据集均采用 10-core 设置(即仅保留至少有过 10 次交互的用户数据)，以确保数据集的质量。在此基础上，本文随机将每个数据集划分为训练集(80%)、验证集(10%)以及测试集(10%)。关于数据集的统计信息如表 1 所示。

表 1 实验数据信息表

数据集	用户数	商品数	关系数	密度
MovieLens-1M	6040	3952	1000209	0.04468
Gowalla	29858	40981	1027370	0.00084
Yelp	31668	38048	1561406	0.00130

5.1.2 对比方法

为了全面客观地评估本文方法的有效性，本文设置如下两种类型的对比方法。

第 1 类是目前被广泛使用且具有代表性的推荐方法。它们将作为非隐私保护方法，用于评估 PF-GNR 的推荐性能。

(1)BPR-MF[34]：一种经典的矩阵因式分解方法，通过最小化成对排序损失函数实现个性化推荐，常被用作推荐排序算法的基准。

(2)NeuMF[36]：一种先进的基于神经网络的推荐方法，它使用多层感知机代替点乘运算来捕捉隐式反馈中的非线性关系。

(3)NGCF[2]：一种基于图神经网络的推荐方法。通过构建一个用户－项目交互图，并使用图神经网络对用户和项目之间的高阶关系建模。

(4)LigthtGCN[1]：一种轻量级的图卷积协同过

滤方法。它在 NGCF 模型的基础上,删除了特征转换和非线性激活操作,从而使得模型更加简洁和高效,并且更适合于推荐领域。值得注意的是,LightGCN 也被作为 PF-GNR 的基础网络模型,用于对用户与项目节点之间的相关性进行建模。

第 2 类是当前先进的图神经网络隐私保护方法。通过与它们进行比较,可以评估本文方法在隐私与效用均衡性方面的表现。

(1)GAP[20]:一种基于 CDP 的图神经网络模型。为了方便与所提方法进行比较,本文在推荐任务中重新实现了 GAP 方法。该方法利用高斯机制对参数的聚合过程进行随机扰动,使得发布的推荐模型满足中心化差分隐私保护。

(2)LapGraph[17]:一种基于边本地化差分隐私的保护方法。该方法采用输入扰动策略,在数据上传的阶段通过向原始图的邻接矩阵中添加拉普拉斯噪声,以保证用户的隐私安全。

(3)DPRR[37]:一种基于边本地化差分隐私的图神经网络框架。该框架采用随机响应的方式对原始图中的每一条边进行随机扰动,然后通过构建边采样策略对扰动后的图数据进行校正,以提高数据的可用性。

5.1.3　评价指标

为了保持与现有工作相同的评估设置,本文采用 Top-K 推荐任务中两种常用的评价指标:召回率 Recall 和归一化折损累积增益 NDCG(normalized discounted cumulative gain),来评估推荐方法的性能。

(1)Recall@K 表示在用户真实感兴趣的目标项目中,被正确预测到 Top-K 推荐列表中的比例,常用来衡量推荐列表的查全率。

(2)NDCG@K 则进一步关注目标项目在推荐列表中的位置,是基于排序结果的评价指标,用来衡量推荐列表的排序质量。

Recall@K 和 NDCG@K 的值越高,代表模型的推荐性能越好。默认情况下,本文中推荐列表长度 K 的值设定为 20。

5.1.4　超参数设置

本文所有实验均在如下环境中完成:Python3.9、Pytorch1.12.1、GeForce RTX2060 SUPER GPU。实验的主要参数设置如下:(1)用户和项目的嵌入使用 Xavier 方法[38]进行初始化,且所有方法的嵌入维度固定为 64;(2)使用 Adam[35]优化器对模型进行优化,学习率为 1e-3,L_2 正则化系数 λ 为 1e-4;(3)图神经网络的层数 L 在{1,2,3,4}中选择,并将各层的权重参数 α_l 统一设置为 $1/(1+L)$。默认情况下,L 设置为 3。所有对比方法均参照文献提供的代码实现,以保证实验的公平性。所有方法都进行 300 个周期完整批次的训练,并采取早期停止策略,即如果连续 10 个周期内验证数据集上的评价指标没有增加,则提前终止训练。所有实验均使用不同的随机种子运行 5 次,然后取评价指标的均值作为最终的实验结果。

5.2　推荐性能的评估(RQ1)

为了评估 PF-GNR 的推荐性能,本文将其与具有代表性的推荐方法进行比较,其中,本文方法的隐私预算 ε 设置为 5。表 2 列出了各模型在 3 个数据集上的推荐性能,加粗表示最佳性能(其他表同此)。从表 2 中可以观察到:

表 2　PF-GNR 与各对比方法的性能比较

模型	Movielens-1M		Gowalla		Yelp	
	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
BPRMF	0.2043	0.3074	0.1291	0.1109	0.0433	0.0354
NUMF	0.2175	0.3428	0.1399	0.1210	0.0451	0.0363
NGCF	0.2201	0.3541	0.1570	0.1327	0.0572	0.0477
LightGCN	**0.2521**	**0.3851**	**0.1823**	**0.1554**	**0.0639**	**0.0525**
PF-GNR($\varepsilon=5$)	0.2341	0.3712	0.1545	0.1341	0.0569	0.0470

(1)基于 GNN 的推荐方法(LightGCN、NGCF 和 PF-GNR)的性能要明显高于传统的矩阵分解方法(BPRMF、NeuMF)。这是因为基于 GNN 的推荐方法能够有效地对实体之间的高阶连接信息进行建模,以增强用户嵌入和项目嵌入的表达能力,从而提升模型的推荐性能。

(2)与本文的基础网络模型 LightGCN 相比,PF-GNR 的推荐性能有所下降。例如,在 Gowalla 数据集上,PF-GNR 的 Recall@20 和 NDCG@20 指标比 LightGCN 方法分别下降了 15.25% 和 13.70%。其主要原因是:PF-GNR 中引入了 LDP 和同态加密技术为用户数据提供保护,这些保护措

施不可避免地会给模型训练带来一定的效用性损失，导致本文方法的性能相较于 LightGCN 方法有所下降。

(3)从整体来看，PF-GNR 在 3 个数据集上的推荐效果与 NGCF 相当，并且显著优于 BPRMF 和 NeuMF。例如，在 Gowalla 数据集上，PF-GNR 的 Recall@20 和 NDCG@20 指标分别比 NeuMF 提升了 10.45%和 10.82%。这说明与现有方法相比，PF-GNR 方法的推荐性能具有较强的竞争力，可为用户提供较高质量的推荐服务。

综上所述，虽然隐私保护机制对 PF-GNR 的性能会产生一定的影响，但总体而言，该方法仍然能够充分发挥 GNN 在图表示学习方面的优势，是一个可用性较高的隐私保护方案。

5.3 隐私－效用均衡性的评估(RQ2)

隐私性与效用性的权衡问题是隐私保护推荐系统的研究重点。为了评估 PF-GNR 在隐私－效用均衡性方面的表现，本文将其分别与基于 CDP 的保护方法(GAP)以及基于 LDP 的保护方法(LapGraph、DPRR)进行了比较。实验过程中，共设置了 4 种隐私预算 $\varepsilon \in \{1,3,5,10\}$，它们分别代表了不同的隐私保护级别($\varepsilon$ 越小，表示隐私级别越高)。图 2 展示了各方法在每种隐私保护级别下进行比较的结果，从中可以看出：

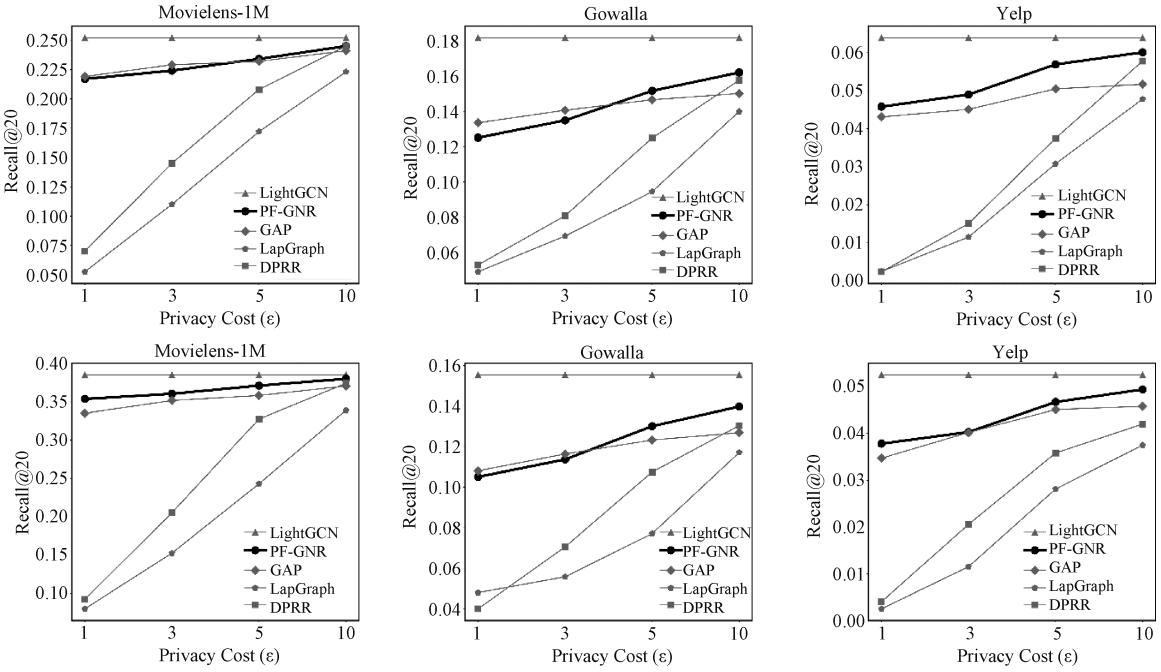

图 2 各模型在不同隐私预算下的性能比较

(1)整体来看，随着隐私预算 ε 减少，所有方法的性能均呈现下降趋势，这符合之前的理论分析。因为隐私预算 ε 与模型可用性呈负相关，当 ε 减少时，意味着添加的噪声量会增加，所造成的负面影响更大，进而导致各方法的性能下降。

(2)与基于 CDP 的方法 GAP 相比，本文方法在大多数情况下具有更好的表现。例如，当隐私预算 $\varepsilon \geqslant 3$ 时，PF-GNR 在 3 个数据集上的性能均要优于 GAP 方法，并且随着 ε 的增加，这种优势还会进一步扩大。只有在 $\varepsilon < 3$ 时，PF-GNR 的表现才会略逊于 GAP 方法。其主要原因是：GAP 通过中心服务器集中对训练数据进行扰动，其优势在于能够灵活有效地控制添加噪声的规模，从而保持较高的数据效用。但是，该方法需要建立在第三方完全可信的假设基础之上。相比之下，本文方法采用了 LDP 技术，不依赖于可信第三方的假设，允许每个用户独立地添加噪声扰动。因此，在隐私预算较低的情况下，整体上可能会积累更多的噪声，从而对模型可用性产生较大的影响。然而，从实际效果来看，PF-GNR 不仅提供了更全面的隐私保护，而且在大多数情况下能够达到甚至超过 GAP 的性能，这充分说明本文方法可以有效缓解隐私性与效用性之间的矛盾。

(3)与基于 LDP 的方法(LapGraph、DPRR)相比，本文方法在各隐私级别下均展现出显著的优势。具体地，从评估指标的变化幅度分析发现，随着 ε 的减少，PF-GNR 的性能呈缓慢下降趋势，且指标下降幅度较小。而 LapGraph 和 DPRR 的性能则表现

出急剧下降的趋势,同时指标下降幅度也非常大。其主要原因在于:一方面,LapGraph 和 DPRR 都是只利用 LDP 技术对邻接矩阵进行随机扰动,以实现隐私保护。然而,由于邻接矩阵具有高维度的特点,直接应用 LDP 技术对其进行处理,将会带来过大的噪声干扰,导致模型可用性严重下降。因此,它们仅适合在低隐私保护级别的情况下使用;另一方面,针对图神经推荐系统的特点,本文设计了一种两阶段隐私保护机制。该机制将 LDP 和同态加密技术相结合,可有效降低差分噪声带来的影响,从而保持模型的可用性。

(4)为了验证 PF-GNR 与其他对比方法在隐私保护性能上是否存在显著的差异,本文采用配对 t-检验[39]对实验数据进行了详细分析。分析结果如表 3 所示。考虑到篇幅的原因,这里只展示了 Gowalla 数据集中的配对 t-检验结果,其他实验数据集的分析结果与此类似,并显示出一致的趋势。其中,平均值和标准差是根据同一算法独立运行 5 次所获得的统计结果(这里是以 NDCG@20 指标为例);SIG 表示 t-检验的 P 值,在置信度 α 取值为 95%的情况下,若 P 位于 0 至 0.05 之间,表示两种算法的性能存在显著差异,反之则表示没有显著差异;Score 表示 PF-GNR 相较于其他对比方法在显著差异上所取得的净胜得分。从表 3 的统计结果中可以看出,在隐私预算相同的情况下,本文算法在 3 个数据集上的总净胜得分为 8,这说明 PF-GNR 的表现要显著优于各对比方法,也进一步验证了本文方法的有效性。

表 3　PF-GNR 与各对比方法的配对 t-检验结果

模型	Mov				Gowalla				Yelp			
($\varepsilon=5$)	均值	标准差	t-test	SIG	均值	方差	t-test	SIG	均值	方差	t-test	SIG
PF-GNR	**0.3712**	**0.007**	—	—	**0.1341**	**0.005**	—	—	**0.0470**	**0.002**	—	—
配对模型 1:GAP	0.3583	0.008	3.98	0.0163	0.1243	0.003	2.89	0.084	0.0412	0.002	4.13	0.0145
配对模型 2:LapGraph	0.2431	0.008	25.15	0.0004	0.0781	0.005	17.80	0.0001	0.0241	0.003	12.5	0.0002
配对模型 3:DPRR	0.2869	0.01	11.24	0.003	0.1090	0.002	10.23	0.0005	0.0358	0.002	8.73	0.0009
Score	**3**				**2**				**3**			

综上所述,通过多个角度的对比分析,可以发现本文方法在各隐私保护级别下均能保持较高且稳定的综合性能,并且对隐私预算变化的敏感度较低,这充分证明了 PF-GNR 能够在隐私性与效用性之间实现有效平衡。

5.4　实用性的评估(RQ3)

5.4.1　通用性分析

本节旨在探讨 PF-GNR 是否具有良好的通用性。为此,本文选取三种典型的图神经网络推荐方法(NGCF、GAT、GraphSage)作为基础网络模型,然后基于本文的隐私保护机制,分别构造出它们的隐私保护变体(PF-NGCF、PF-GAT、PF-GSage),来进行对比实验。其中,所有方法的隐私预算 ε 统一设置 5。实验结果如图 3 所示,从中可以观察到:

(1)在三个数据集上,PF-NGCF、PF-GAT、PF-GSage 的推荐性能都要低于它们的基础模型 NGCF、GAT 以及 GraphSage,该观察结果与第 4.5 节的理论分析相符。因为在 PF-NGCF、PF-GAT 以及 PF-GSage 中都引入了隐私保护技术来处理用户的数据,必然会给模型可用性带来一定的损失。

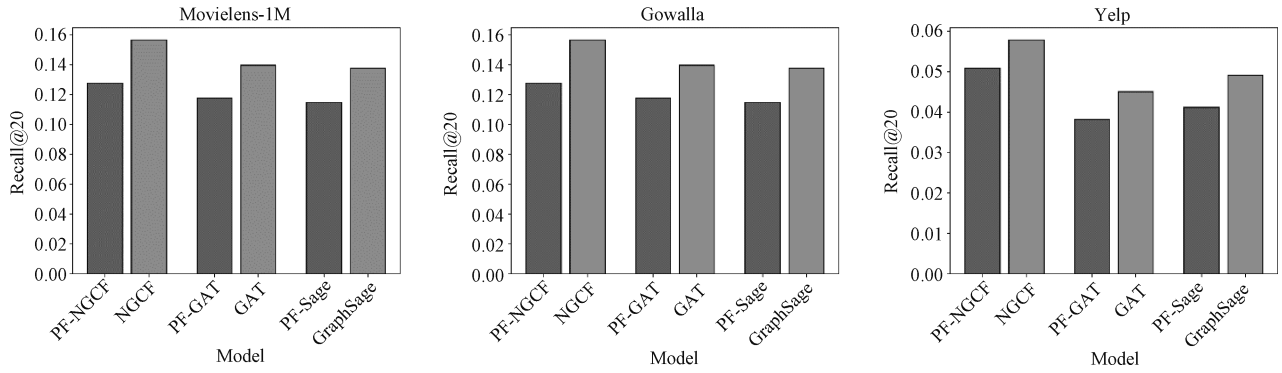

图 3　PF-GNR 方法及其 3 个变体的推荐性能

(2)经过对比发现,PF-NGCF、PF-GAT 以及 PF-GSage 的性能与它们各自的基础网络模型之间差距并不明显。这表明本文提出的隐私保护方案具备良好的通用性,可适用于不同的图神经网络模型。

5.4.2 训练效率分析

对于隐私保护方法来说，训练效率是最直接也是最重要的衡量标准之一。因此，本小节在 3 个公开数据集上，对 PF-GNR 训练效率进行评估。为了确保对比的公平性，所选取的三种对比方法 GAP、LapGraph 以及 DPRR 均以 LightGCN 作为基础网络模型，并在相同的计算环境中对每种方法进行实验，记录它们在每个数据集上的总训练时间。实验结果如表 4 所示。从中可以看出：

表 4 PF-GNR 与各对比方法的训练时间

数据集	训练时间/h				
	PF-GNR ($\varepsilon=5$)	GAP ($\varepsilon=5$)	LapGraph ($\varepsilon=5$)	DPRR ($\varepsilon=5$)	Light-GCN
Movielens-1M	2.7	2.3	2.18	2.15	**2.1**
Gowalla	27.8	26.4	25.7	25.1	**24.5**
Yelp	43.8	40.3	38.2	37.9	**37.2**

(1)与基础网络模型 LightGCN 相比，隐私保护方法(PF-GNR、GAP、LapGraph 和 DPRR)需要花费更多的训练时间。其主要原因在于这些隐私保护方法中都引入了相应的隐私保护机制，不可避免地会带来额外的计算开销，从而影响模型的训练效率，导致训练时间有所增加。

(2)在所有的方法中，LightGCN 的训练时间最少，其次是 GAP、LapGraph 和 DPRR，最后是本文方法 PF-GNR。这与第 4.5.2 小节的理论分析相一致。其主要原因是：GAP、LapGraph 和 DPRR 三种对比方法中仅使用了差分隐私技术，而本文方法则同时采用了 LDP 和同态加密两种隐私保护技术。特别是，同态加密计算带来的计算开销较大，直接导致了 PF-GNR 训练时间的增加。整体来看，尽管本文方法训练时间是最多的，但与其他对比方法相比，它在推荐性能上取得了显著的改进，同时也将训练时间的增幅控制在较低的水平。例如，在 Gowalla 数据集上，PF-GNR 的训练时间仅比基础网络模型 LightGCN 多出约 13.5%。这充分说明，本文方法 PF-GNR 的隐私保护机制对训练效率产生的影响仍保持在可接受的范围之内。

5.5 消融测试(RQ4)

本节通过消融实验，来验证 PF-GNR 中本地隐私编码模块和隐私保护更新模块的作用。为此，本文设计了 3 个变体方法：GNR+LDP 表示去除了隐私保护更新模块，仅利用 LDP 技术为用户的本地数据提供保护；GNR+HE 表示去除了本地差分隐私编码模块，只借助同态加密技术保护训练过程中的数据隐私；GNR 表示同时去除本地差分隐私编码模块和隐私保护更新模块，即不使用任何隐私保护技术，主要作为其他方法的参考基准。表 5 展示了消融实验的结果。

表 5 消融实验

模型	Movielens-1M		Gowalla		Yelp	
	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
GNR+LDP($\varepsilon=5$)	0.1721	0.2435	0.0943	0.0781	0.0408	0.0354
GNR+HE	0.2504	0.3820	0.1804	0.1210	0.0632	0.051
GNR	0.2521	0.3851	0.1823	0.1554	0.0639	0.0525
PF-GNR($\varepsilon=5$)	**0.2341**	**0.3712**	**0.1545**	**0.1341**	**0.0569**	**0.0458**

从实验结果可以看出：(1)与非隐私保护变体 GNR 相比，GNR+LDP 在 3 个数据集上的推荐效果都有明显下降。这说明单独应用 LDP 技术，虽然能够保证图数据的隐私安全，但容易引入过量噪声，导致训练数据的可用性较差，从而严重影响模型性能。(2)GNR+HE 的表现与 GNR 的相当。这说明使用同态加密技术对模型可用性造成的影响很小，能够在不影响训练效果的前提下，为数据提供严格的隐私保护。但需要注意的是，基于同态加密技术的保护策略只针对数据的内部交换，无法抵御模型发布之后受到的隐私威胁，因此通常需要与其他隐私技术结合应用。(3)与 GNR+LDP 和 GNR+HE 相比，PF-GNR 实现了隐私性和可用性之间更有效的平衡。这说明本文方法在本地隐私编码模块和隐私保护更新模块的共同作用下，不仅能够为用户提供严格的隐私保护，而且还可以有效减少差分噪声带来的负面影响，保持模型的可用性。

5.6 超参数的敏感性分析(RQ5)

本小节通过实验分析 2 个关键参数对 PF-GNR 性能的影响，包括：图神经网络层数 L 和隐私预算分配参数 δ。实验中，本文将 PF-GNR 的隐私预算 ε 固定为 5，并通过分别改变参数 L 和 δ，观察模型性能的变化。

(1)图神经网络层数 L 的影响。对于 GNN 模型来说，通过堆叠图神经网络层，可以使每个节点访问到高阶连接信息，但这也意味着每个节点会接收

到更多的邻域节点信息,这使得图中节点的特征可能变得越来越相似,从而导致过平滑问题。本文在3个数据集上分别测试了图神经网络层数 L ($L \in \{1,2,3,4\}$) 对 PF-GNR 性能的影响,结果如表6所示。可以发现,PF-GNR 的性能随着层数 L 的增加先升后降,大多数情况是在 $L=3$ 时,取得最佳性能。其主要原因在于增加层数获取高阶连接信息的收益已经无法抵消过平滑产生的负面影响,使得模型的性能逐渐下降。因此,在实际应用中,需要综合考虑任务的复杂性、数据规模以及模型结构等因素,来选择合适的图神经网络层数,以减轻过平滑问题带来的影响。

表6　图神经网络层数对 PF-GNR 推荐性能的影响

方法	GNN 的层数 L	Movielens-1M		Gowalla		Yelp	
		Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
PF-GNR ($\varepsilon=5$)	$L=1$	0.2287	0.3530	0.1353	0.1135	0.0528	0.0433
	$L=2$	0.2258	0.3502	0.1409	0.1217	0.0530	0.0435
	$L=3$	**0.2341**	**0.3712**	**0.1459**	**0.1264**	0.0569	0.0468
	$L=4$	0.2314	0.3685	0.1431	0.1240	**0.0575**	**0.0479**

(2)隐私分配参数 δ 的影响。本文在第4.2节中介绍了隐私分配参数 δ 的作用。具体而言,在数据采集阶段,每个用户 u_i 需根据 $\delta \in [0,1]$ 将隐私预算 ε 分成两部分 $\varepsilon_a=\delta\varepsilon$ 和 $\varepsilon_d=(1-\delta)\varepsilon$,分别用于对其邻接表 A_i 和度 D_i 进行随机扰动。本文将隐私预算 ε 设定为5,然后在3个数据集上分别测试 δ($\delta \in \{0.0,0.5,0.7,0.9\}$)对 PF-GNR 性能的影响,结果如表7所示。从中可以看出,随着 δ 值增加,PF-GNR 的性能会有一定提升。这是因为用户邻接表通常具有高纬度且稀疏的特性,对于添加的噪声会非常敏感,将更多的隐私预算分配给邻接列表可以提高它的可用性,从而帮助提升训练效果。因此,通常情况下,需要将 δ 的值设置在0.9以上,以获得性能更好的推荐模型。

表7　隐私分配参数对 PF-GNR 推荐性能的影响

方法	隐私分配参数 δ	Movielens-1M		Gowalla		Yelp	
		Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
PF-GNR ($\varepsilon=5$)	$\delta=0.0$	0.1915	0.3074	0.1091	0.9485	0.0387	0.0314
	$\delta=0.5$	0.2236	0.3610	0.1310	0.1123	0.0490	0.0405
	$\delta=0.7$	0.2278	0.3658	0.1362	0.1173	0.0535	0.0434
	$\delta=0.9$	**0.2341**	**0.3712**	**0.1459**	**0.1264**	**0.0569**	**0.0468**

6　总　结

本文研究了图神经网络推荐系统的隐私保护问题,并提出了一种隐私保护框架 PF-GNR。该框架通过结合本地差分隐私和同态加密技术的优势,能够为用户提供从数据收集到模型发布的全周期保护。首先,在数据收集阶段,采用 LDP 技术对用户原始数据进行隐私化处理,保证了用户数据在传输和使用中的隐私性。其次,在模型训练阶段,借助同态加密技术保证了训练过程中用户数据的安全性和隐私性。此外,PF-GNR 框架是一种通用的隐私保护解决方案,可适用于现有各种图神经网络推荐方法。为了验证本文方法的有效性,在三个公共数据集上进行了综合实验,实验结果表明,相较于现有的隐私保护方法,PF-GNR 框架能实现更好的隐私性与效用性之间的平衡。下一步,将对个性化、轻量级的隐私保护图神经推荐方案展开研究,使保护方案能自适应地处理更加复杂的实际应用场景。

参　考　文　献

[1] He X, Deng K, Wang X, Li Y, Zhang Y, Wang M. LightGCN: Simplifying and powering graph convolution network for recommendation//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2020: 639-648

[2] Wang X, He X, Wang M, Feng F, Chua TS. Neural graph collaborative filtering//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2019: 165-174

[3] Ying R, He R, Chen K, Eksombatchai P, Hamilton WL, Leskovec J. Graph convolutional neural networks for web-scale recommender systems//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA, 2018: 974-983

[4] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks. arXiv preprint arXiv: 1710.10903, 2017

[5] Zhang Y, Wu L, Shen Q, et al. Graph learning augmented heterogeneous graph neural network for social recommendation. ACM Transactions on Recommender Systems, 2023, 1(4): 1-22

[6] Jiang Y., Yang Y., Xia L., Chao H. DiffKG: Knowledge graph diffusion model for recommendation//Proceedings of the 17th ACMInternational Conference on Web Search and Data Mining. New York, USA, 2024: 313-321

[7] Li X, Xiao G, Chen Y, et al. An explicitly weighted gcn aggregator based on temporal and popularity features for recommendation. ACM Transactions on Recommender Systems, 2023, 1(2): 1-23

[8] Liao X, Liu W, Zheng X, et al. PPGenCDR: A stable and robust framework for privacy-preserving cross-domain recommendation//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA, 2023: 4453-4461

[9] Liu W, Zheng X, Chen C, et al. Reducing item discrepancy via differentially private robust embedding alignment for privacy-preserving cross domain recommendation//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria, 2024:32455-32470

[10] Zhang S., Yuan W., Yin H. Comprehensive privacy analysis on federated recommender system against attribute inference attacks. IEEE Transactions on Knowledge and Data Engineering, 2024, 36: 987-999

[11] Xiao GQ, Li XQ, Chen YD, et al. A survey of large scale graph neural network. Chinese Journal of Computers. 2024, 47(1): 148-171, (in Chinese)
(肖国庆, 李雪琪, 陈玥丹,等.大规模图神经网络研究综述. 计算机学报, 2024, 47(1): 148-171)

[12] Zhang SH, Yin HZ, Chen T, Huang Z. 2021. Graph embedding for recommendation against attribute inference attacks//Proceedings of the Web Conference 2021. New York, USA, 2021: 3002-3014

[13] Zhang ZX, Liu Q, Huang ZY, et al. GraphMI: Extracting private graph data from graph neural networks//Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal, Canada, 2021: 3749-3755

[14] H Jiang, J Pei, D Yu, J Yu, et al. Applications of differential privacy in social network analysis: A survey. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 108-127

[15] He XL, Jia JY, Michael B, et al. Stealing links from graph neural networks//Proceedings of the 30th USENIX Security Symposium. Virtual, 2021: 2669-2686

[16] Zhang Z, Chen M, Backes M, Shen Y, Zhang Y. Inference attacks against graph neural networks. //Proceedings. of the 31st USENIX Security Symposium. Boston, USA, 2022: 4543-4560

[17] Wu F, Long Y, Zhang C, et al. LinkTeller: Recovering private edges from graph neural networks via influence analysis//Proceedings of the 2022 IEEE Symposium on Security and Privacy. San Francisco, USA 2022: 2005-2024

[18] Liu YH, Chen H, Liu YX, et al. State-of-the-art privacy attacks and defenses on graphs. Chinese Journal of Computers. 2022, 45(4):702-734 (in Chinese)
(刘宇涵, 陈红, 刘艺璇等. 图数据上的隐私攻击与防御技术. 计算机学报, 2022, 45(4): 702-734)

[19] Zheng X, Wang Z, Chen C, et al. Decentralized graph neural network for privacy-preserving recommendation//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. New York, USA, 2023: 3494-3504

[20] Sajadmanesh S, Shamsabadi A S, Bellet A, et al. GAP: Differentially private graph neural networks with aggregation perturbation//Proceedings of the 32nd USENIX Security Symposium. CA, USA 2023: 3223-3240

[21] Wu, C., Wu, F., Lyu L., et al. A federated graph neural network framework for privacy-preserving personalization. Nature Communications, 2022, 13(1): 3091

[22] Pei X, Deng X, Tian S, et al. Privacy-enhanced graph neural network for decentralized local graphs. IEEE Transactions on Information Forensics and Security, 2024, 19: 1614-1629

[23] Pei Y, Mao R, Liu Y, et al. Decentralized federated graph neural networks//Proceedings of the International Workshop on Federated and Transfer Learning for Data Sparsity and Confidentiality in Conjunction with IJCAI 2021. Montreal, Canada, 2021

[24] Wang S, Zheng Y, Jia X. Secgnn: Privacy-preserving graph neural network training and inference as a cloud service. IEEE Transactions on Services Computing, 2023, 16(4): 2923-2938

[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] Ye QQ, Meng XF, Zhu MJ, Huo Z. 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)

[27] Imola J, Murakami T, Chaudhuri K. Locally differentially private analysis of graph statistics//Proceedings of the 30th USENIX Security Symposium. Virtual, 2021:983-1000

[28] Menezes AJ, Van Oorschot PC, Vanstone SA. Handbook of Applied Cryptography. CRC Press, 2018

[29] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York, USA, 2016:308-318

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

[31] Zhang Q, Hong kyu Lee, Ma J, et al. DPAR: Decoupled

graph neural networks with node-level differential privacy// Proceedings of the ACM Web Conference 2024. New York, USA, 2024: 1170-1181

[32] Ye Q, Hu H, Au M H, et al. LF-GDPR: A framework for estimating graph metrics with local differential privacy. IEEE Transactions on Knowledge and Data Engineering, 2020, 34 (10): 4905-4920

[33] Li ZY, Gui XL, Gu YJ, Li XS, Dai HJ, Zhang XJ. Survey on homomorphic encryption algorithm and its application in the privacy-preserving for cloud computing. Journal of Software, 2018, 29(7): 1830-1851 (in Chinese)
(李宗育, 桂小林, 顾迎捷等. 同态加密技术及其在云计算隐私保护中的应用. 软件学报, 2018, 29(7): 1827-1851)

[34] Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L. BPR: Bayesian personalized ranking from implicit feedback// Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada, 2009: 452-461

[35] Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv: 1412.6980, 2014

[36] 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

[37] Hidano S, Murakami T. Degree-preserving randomized response for graph neural networks under local differential privacy. Transactions on Data Privacy. 2024, 17(2): 89-121

[38] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks//Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Sardinia, Italy, 2010: 249-256

[39] Douglas C D. Design and Analysis of Experiments. 6th Edition. New York, USA: John Wiley & Sons, 2007

WANG Kun, Ph. D. candidate, senior engineer. His research interests include graph representation learning and data privacy protection.

WANG Yong, Ph. D., professor, Ph. D. supervisor. His research interests include data mining and data privacy protection.

ZHANG Zhi-Qiang, Ph. D., lecturer. His research interests include data mining and recommender system.

LIU Jin-Yuan, Ph. D., lecturer. His research interests include computer networks and network security.

DENG Jiang-Zhou, Ph. D., lecturer. His research interests include data mining and recommender system.

Background

In recent years, graph neural networks have been widely used in the field of recommendation by virtue of their powerful graph representation learning capability, and have become one of the most effective methods for solving recommendation problems. However, it has been proved that due to the uniqueness of its own architecture and training method, GNN-based recommender systems faces more prominent and harmful privacy and security problems, which seriously threaten users' privacy rights and interests.

The current research on privacy protection of GNN-based recommender systems has made some achievements, but it is still in the early exploration stage, and there are still many key problems to be solved. First, the existing researches are mainly based on the security assumption that the central server is trustworthy to design the privacy protection scheme, which has greater limitations in practical applications; second, due to the complex structure and strong correlation of the graph data, the traditional privacy-preserving methods for relational data can no longer satisfy the privacy protection needs of GNN-based recommender systems; finally, the existing methods are designed with little improvement and optimization for the special properties of GNN-based recommender systems, which makes them difficult to achieve an effective balance between privacy and usability.

Focusing on the above problems, this paper proposes a privacy-preserving framework PF-GNR for GNN-based recommender systems. The framework consists of three main parts: a local coding module, a recommendation module based on graph neural networks, and a privacy-preserving model updating module. Firstly, in the data collection phase, each user randomly scrambles and encodes the original data using local differential privacy techniques to ensure the privacy of user's data during transmission and use; second, in the learning phase, the central server uses graph neural networks to model the preference relationship between users and items; and finally, in the training phase, this paper achieves the security and privacy of the model update with the help of homomorphic encryption techniques. Compared with the existing approaches, the PF-GNR can provide users with full-cycle protection from data collection to model release while guaranteeing model availability by combining the advantages of local differential privacy and homomorphic encryption techniques. In addition, PF-GNR is a generalized privacy-preserving scheme suitable for various existing graph neural network recommendation models.

This work was supported by the National Natural Science Foundation of China (Nos. 62272077 and 72301050), and the Major scientific and technological research projects of the Chongqing Municipal Education Commission (No. KJZD-M202400604).