

分类号：TP391

密 级：公开

学校代码：10363

学 号：2210910101



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 面向社交推荐的隐私保护方法
研究

论 文 作 者 汪玉洁

指 导 教 师 刘涛 教授 潘正高 教授

学 科（专业） 软件工程

研 究 方 向 领域软件工程

论文提交日期： 2024 年 5 月 23 日

分类号：TP391

单位代码：10363

密 级：公开

学 号：2210910101

题 目 面向社交推荐的隐私保护方法研究

英文并列

题 目 Research on Privacy Protection Methods for
Social Recommendation

学生姓名： 汪玉洁

指导教师： 刘涛 教授 潘正高 教授

专 业： 软件工程

研究方向： 领域软件工程

论文答辩日期： 2024 年 5 月 24 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：汪玉洁

日期：2024 年 5 月 23 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在_____年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：汪玉洁

日期：2024 年 5 月 23 日

指导教师签名：刘伟

日期：2024 年 5 月 23 日

面向社交推荐的隐私保护方法研究

摘 要

随着互联网技术的迅速发展,在线社交媒体不仅成为了信息传播、获取、共享的重要平台,而且积累了海量的用户数据,包括基本属性、社交信息以及交互记录等。这些数据量的激增为用户带来了信息过载的问题,使得筛选和获取高质量信息变得愈发困难。社交推荐的出现有助于人们快速获取自己所需要的信息,它利用先进的算法和模型,协助用户从海量的信息中筛选出高质量、符合个人偏好的内容。然而,社交推荐在为用户提供个性化推荐的同时,也引发了用户隐私泄露的风险。因此,如何在确保提供准确推荐的前提下,保证用户的隐私安全,已成为当前社交推荐领域的研究热点。本文针对社交推荐隐私保护面临的问题与挑战,提出改进算法,具体工作如下:

1. 深入研究社交推荐模型的基本框架及其各类隐私保护技术,对相关方法进行了代码复现和实验验证,并详细分析了这些方法的优缺点,归纳总结在实现社交推荐隐私保护过程中可借鉴的经验和需要关注的问题。

2. 针对当前社交推荐方法通常忽视在处理原始数据时的隐私保护问题,提出了一种基于差分隐私保护的社交推荐隐私保护方法。该方法采用差分隐私保护技术对数据进行处理,将噪声划分为两部分,分别添加到数据选取与数据扰动的过程中,确保算法满足 ϵ -差分隐私;并且根据用户关联度和用户相似度计算得到社交网络用户的综合相

似度来确定目标用户的最相似用户集合，以提供个性化的推荐结果。实验结果表明，将所提出的算法应用于三种现有的社交推荐算法后，能够在实现隐私保护的情况下，其推荐的准确性也较好。

3. 针对当前社交推荐方法往往忽略用户个人信息的隐私保护问题，提出了一种基于社区划分的隐私保护社交推荐方法。该方法首先在传统社区划分的基础上，考虑用户在社交网络中的联系，然后利用社区数据集合代替个人数据集合进行推荐，以保护用户的隐私信息。此外，通过寻找与目标用户所处社区最相似的社区集合，为目标用户提供符合期望的推荐结果。实验结果显示，与目前基于矩阵分解的社交推荐方法相比，所提出的方法不但保护了用户的个人隐私，而且其推荐精度损失也较少。

本文旨在探索面向社交推荐的隐私保护策略，并根据已有的理论和技术设计新的方法。实验结果显示，所提出的方法能够有效平衡用户的隐私性和推荐结果的准确性。

关键词：社交推荐，隐私保护，差分隐私，社区划分，矩阵分解

RESEARCH ON PRIVACY PROTECTION METHOD FOR SOCIAL RECOMMENDATION

ABSTRACT

With the rapid development of Internet technology, online social media has not only become an important platform for information dissemination, acquisition and sharing, but also accumulated massive user data, including basic attributes, social information and interaction records. The proliferation of these data volumes creates the problem of information overload for users, making it increasingly difficult to sift through and obtain high-quality information. The emergence of social recommendation helps people to quickly obtain the information they need. It uses advanced algorithms and models to help users screen out high-quality content in line with personal preferences from massive information. However, while providing users with personalized recommendations, social recommendation also raises the risk of user privacy disclosure. Therefore, how to ensure the privacy security of users on the premise of providing accurate recommendation has become a research hotspot in the field of social recommendation. In view of the problems and challenges faced by social recommendation privacy protection, this paper proposes an improved algorithm. The specific work is as follows:

1. In-depth study of the basic framework of social recommendation model and various privacy protection technologies, code repetition and experimental verification of relevant methods, detailed analysis of the advantages and disadvantages of these methods, and summary of the experience that can be used for reference and problems that need attention in the process of realizing social recommendation privacy protection.

2. In view of the fact that current social recommendation methods usually ignore the privacy protection problem when processing raw data, a social recommendation privacy protection method based on differential privacy protection is proposed. The method uses differential privacy protection technology to process the data, divides the noise into two parts, and adds the noise to the process of data selection and data disturbance respectively to ensure that the algorithm meets the ϵ -differential privacy. Moreover, the comprehensive similarity of social network users is calculated according to user association degree and user similarity degree to determine the most similar user set of target users to provide personalized recommendation results. The experimental results show that the proposed algorithm can be applied to three existing social recommendation algorithms, and its recommendation accuracy is good in the case of privacy protection.

3. Aiming at the problem that the current social recommendation

methods often ignore the privacy protection of users' personal information, a social recommendation method for privacy protection based on community division is proposed. Based on the traditional community division, the method first considers the user's contact in the social network, and then uses the community data set instead of the personal data set to recommend, so as to protect the user's privacy information. In addition, by finding the community set that is most similar to the target user's community, the recommendation results that meet the expectation of the target user are provided. The experimental results show that compared with the current social recommendation methods based on matrix decomposition, the proposed method not only protects the user's personal privacy, but also has less loss of recommendation accuracy.

The dissertation aims to explore the privacy protection strategy for social recommendation, and design a new method based on the existing theories and technologies. The experimental results show that the proposed method can effectively balance the privacy of users and the accuracy of recommendation results.

KEY WORDS: social recommendation, privacy protection, differential privacy, community division, matrix factorization

目 录

摘 要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.3 论文的主要研究内容	5
1.4 论文的组织结构	6
第 2 章 相关理论与技术介绍	7
2.1 社交推荐的基本框架	7
2.2 社交推荐模型	9
2.2.1 基于矩阵分解的社交推荐模型	10
2.2.2 基于图学习的社交推荐模型	13
2.2.3 基于因子分解机的社交推荐模型	15
2.3 社交推荐面临的隐私问题	17
2.4 社交推荐中使用的隐私保护方法	17
2.4.1 差分隐私保护	18
2.4.2 匿名化	19
2.4.3 联邦学习	19
2.5 本章小结	21
第 3 章 社交推荐中的差分隐私保护方法	22
3.1 方法的提出	22
3.2 方法的描述	23
3.2.1 寻找相似用户	24

3.2.2 保护用户真实评分	27
3.2.3 SRDP 算法的隐私性分析	29
3.3 实验结果与分析	30
3.3.1 实验数据集与对比算法	30
3.3.2 实验结果与分析	33
3.4 本章小结	36
第 4 章 基于社区划分的社交推荐隐私保护方法	37
4.1 方法的提出	37
4.2 方法的描述	38
4.2.1 划分社区	39
4.2.2 分组聚合	41
4.2.3 寻找相似社区	43
4.3 实验结果与分析	45
4.3.1 实验数据集与对比算法	45
4.3.2 实验结果与分析	46
4.4 本章小结	50
第 5 章 总结与展望	51
5.1 工作总结	51
5.2 未来展望	52
参考文献	53
攻读学位期间发表的学术论文目录	59
攻读学位期间参与的科研项目	59
攻读学位期间参与的学科竞赛项目	59
致 谢	60

第 1 章 绪论

1.1 研究背景及意义

现如今虽然信息随处可得，但是想要快速找到有价值的信息却不那么容易。推荐系统^[1]的出现帮助人们做出快速且满意的选择，例如饮食推荐^[2]、电影推荐^[3]等。为了给用户提供准确的推荐结果，推荐系统需要大量收集用户的个人信息以及深入挖掘用户的交互数据来展开分析。但是，这些信息和交互数据蕴含着丰富的个人隐私，可能会给用户带来隐私泄露的风险^[4]。并且伴随着在线社交网络的迅速发展，以社交关系为核心的社交网络受到了越来越多的关注，人们开始思索怎样借助社交关系来生成更为准确的推荐。社交推荐^[5]是在传统推荐的基础上，将用户的各种社交信息作为关键的影响因素融入到推荐的生成过程中，以此提高推荐的准确性。然而，这种做法也增加了用户隐私泄露的风险^[6]。强化个人信息保护，规范个人信息的获取与运用，这不单和个人权益的维护紧密相连，对于数字经济的健康发展也有着重要的影响。在 2021 年 8 月，我国已颁布《中华人民共和国个人信息保护法》，并于同年 11 月开始执行。在“十四五”规划中也明确表明了要加速推动数据安全、个人隐私保护等领域基础性立法的需要。

随着国家在立法层面上针对隐私安全方面的逐步完善以及公民隐私安全意识的不断提高，大家更加关注自身的隐私安全问题。如果社交推荐存在隐私泄露的风险，那么用户将会对其产生不信任，进而导致用户不愿意向社交推荐系统提供自身的相关数据，最终致使社交推荐的质量下降。所以，想要提升社交推荐系统的推荐服务质量，该系统就必须要为用户提供有力的隐私安全保障。如果社交推荐给予了强有力的隐私安全保障，那么用户便会没有隐私安全顾虑，在社交平台上会更乐意去分享数据。由此可见，为了有力推动推荐系统的良性发展，深入研究社交推荐中用户个人隐私的保护是极为必要的。目前已经有许多隐私保护模型和技术被提出，然而如何化解数据可用性与隐私保护之间的矛盾，依旧是一个亟待解决的难题。在这样的背景之下，本章着重阐述社交推荐中隐私保护技术的现况，以及我们的研究内容与结构安排。

1.2 国内外研究现状

针对不同的用户隐私保护需求，多种隐私保护技术被应用于社交推荐中。为了梳理最新研究进展，本节主要介绍差分隐私、匿名化和联邦学习的应用情况：

(1) 差分隐私保护技术：由于该技术不关心攻击者所拥有的背景知识且能对隐私保护水平量化，人们开始将差分隐私与推荐系统相结合。针对社交推荐中的隐私保护问题，彭慧丽等^[7]将用户按照社交网络结构进行分组，并利用加权路径度量推荐效用。同时，在社交图的权重中注入拉普拉斯噪声以实现扰动，但在将用户进行分组时缺少对用户相似度的考虑。鲜征征等^[8]提出了一种利用差分隐私技术扰动目标函数的融合信任关系的奇异值分解模型，然而该方法没有考虑对原始数据的保护问题。Zhou 等^[9]提出了一种针对移动用户的差分隐私社交推荐系统，可以阻止好奇的用户推断其他人的评分值或社交关系的存在。旨为保护用户的原始评分，Dou 等^[10]提出了一种改进的加权噪声注入方法。该方法通过根据每个核心用户对目标用户的不同影响，添加不同级别的噪声来实现保护。此外，他们还使用部分用户的相关数据来替换所有用户的相关数据，减少了用于推荐的数据量。Meng 等^[11]提出了一种社交推荐框架，该框架允许用户在私下中建模用户反馈和社交关系。为了保护用户免受不受信任的推荐人和朋友的干扰，该框架使用不同的噪声幅度来区分个性化的敏感和非敏感反馈，能确保底层推荐系统的有效性。Liu 等^[12]通过引入图嵌入算法生成用户潜在特征，然后与异构信息网络进行信息融合，最后使用符合差分隐私的深度神经网络进行隐私保护推荐。Uplavikar 等^[13]研究了隐私泄露可能发生的多种方式，并提出了一项综合的解决办法。Li 等^[14]研究了用户专用设备中基于不可信服务器的加密干扰，为用户数据设置差分隐私保护机制。孙道柱等^[15]针对现有算法没有充分挖掘数据的特性，提出了一种在矩阵分解过程中加入隐式反馈的并基于差分隐私保护的推荐算法。Huo 等^[16]提出了两种隐私保护方法，分别为地理位置和社会关系提供隐私保障，并提供了一种差分熵方法来证明其隐私增益。Zhou 等^[17]提出了一种隐私保护的分布式推荐系统，采用联邦学习框架来帮助在大量分布式代理上训练高质量的隐私保护集中式模型，但是这些代理在网络连接相对较慢的情况下可能不可靠。

Macwan 等^[18]提出了一种基于同类群体效应的隐私保护系统,用于为社交网络用户生成推荐,并且运用差分隐私技术来保护朋友的敏感属性信息。Cai 等^[19]提出了一种联邦跨域社交推荐算法,应用了随机响应机制实现用户连接的稀疏维护差分隐私。

(2) 匿名化技术:对数据实施匿名化处理,是保护隐私的一种简便且直接的方式,其旨在避免数据使用方知晓数据的来源,进而有效地保护用户的隐私。通常采用两种方法来达成匿名化^[20]:假名化与分组聚合。假名化的核心要义,是令用户在提交数据时,无需提供个人属性信息。例如,Checco 等^[21]提供了一种“隐匿于人群”的推荐服务模式,通过运用带有公共组身份的 web 服务,用户得以在人群中隐匿自身,在获取一定程度的匿名性的同时,依然能够获得个性化推荐。而分组聚合方法的核心思想是以集体数据取代个人数据,就是在收集用户数据时,将用户进行分组,并以组为单位进行信息收集。通过将个人信息隐藏在分组的集体信息中,从而保护用户的隐私。鲜英等^[22]提出了一种敏感属性匿名化方法,该方法首先对敏感属性值进行分组,接着再展开匿名化处理,然而并未考虑动态信息的处理。王海艳等^[23]提出了一种基于群组推荐的隐私保护模型。该模型首先收集团组中各用户的数据,然后通过随机扰动前 k 个用户来保护隐私。彭丽寻等^[24]提出了一种采用 k 匿名技术的隐私推荐算法,该算法首先将用户的隐私属性进行泛化,而后允许同一等价类中的用户相互转发数据以保护用户的隐私,但并未考虑多用户同时转发时的资源问题。Yi 等^[25]利用联邦学习框架,首先将用户进行分组,接着按组收集梯度信息,使得服务器无法区分用户与梯度之间的对应关系,以保护用户隐私。Bassem 等^[26]提出了一种协作群感知模型,在该模型中,对等方相互协作以混淆其轨迹信息,接着通过协作推荐机制和行程分割算法,在不牺牲系统的推荐性能的同时,还保护了轨迹信息。Li 等^[27]提出了一种基于安全聚合的差分隐私联邦推荐模型,该模型不但能够确保数据的隐私性,还提升了集中式训练性能。在推荐过程中,Luo 等^[28]运用聚类方法实现了用户的隐私保护。首先采用同态加密对用户数据进行保护,然后通过聚类技术对数据进行分区后再进行推荐。Sahoo 等^[29]提出了一种用于推荐的扰动隐私保护方法。在该方法中,用户评分是模糊的,接着确保分组在同一聚类中的用户相似。

(3) 联邦学习：作为一种新兴的隐私保护范式，这种范式采用了一种协同学习的方式，将所有的原始数据保存在用户的终端设备上，通过模型参数或者梯度等信息的共享来实现全局模型的学习，以更好地保护用户的隐私。相较于传统的集中式存储与训练模式，这种新兴的范式在保护用户隐私方面有着显著的优势。传统的集中式存储与训练模式需要将用户的数据集中存储在一个服务器上，这样可能导致用户隐私被泄露的风险。而新兴的隐私保护范式则避免了将用户的原始数据集中存储，仅通过模型参数或者梯度等信息的共享来进行模型的学习，从根源上保护了用户的隐私。这种方式不仅可以保护用户的个人隐私，还可以降低数据泄露的风险。Perifanis 等^[30]提出了一种保护隐私的联邦学习方法，通过对加密数据进行操作来整合用户朋友的偏好，以生成前 n 名的兴趣点推荐。联邦学习过程符合数据传播和最小化法规，因为本地数据永远不会离开数据拥有者的设备。中间计算被安全地聚合，以防止被动实体推断出其他信息。社交功能在联合过程之后集成到全局模型中，其中服务器对加密数据进行操作，以使用完全同态加密方案的属性融合学习的用户嵌入。Cai 等^[31]提出了一种联邦跨域隐私保护社交推荐算法。主要创新是应用随机响应机制实现用户连接的稀疏差分隐私，并提出矩阵混淆方法实现高效的加密用户特征向量更新。Zhou 等^[32]提出了一个隐私保护的分布式推荐系统，然后将每个服务提供商建模为具有上下文感知的分布式在线学习者。并采用联邦学习框架来帮助在大量分布式代理上训练高质量的隐私保护集中式模型，这在网络连接相对较慢的情况下可能不可靠。为了处理大数据场景，构建了一个项目集群树来处理不断增加的数据集。并进一步考虑了社交网络的结构，并提出了一种有效的算法，以自适应地避免更多的性能损失。

总的来说，在目前的社交推荐领域中，还没有找到一种完全行之有效的方法来同时保障用户的隐私性和推荐结果的准确性。因此，目前的研究重点之一是在保护用户个人信息的前提下，寻找一种方法来平衡数据隐私和数据可用性。这需要综合考虑各种隐私保护技术，并与推荐系统的设计相结合，以确保在保护用户隐私的同时能够提供准确有用的推荐结果。

1.3 论文的主要研究内容

本课题主要面向社交推荐领域的隐私保护研究，先对其进行分析综述，然后根据不同的隐私保护需求提出两种隐私保护方法，分为基于差分隐私的社交推荐隐私保护方法（SRDP）和基于社区划分的社交推荐隐私保护方法（SRCD）。前者通过融入差分隐私保护技术来保证原始数据的隐私安全问题，同时通过寻找目标用户的相似用户来保证社交推荐的准确度。后者通过引入社区划分算法，同时考虑社交关系的影响，把社交网络中的用户划分到不同的社区，并将用户的个人信息隐藏在社区数据集合中，同时通过寻找目标用户所在社区的相似社区来保证社交推荐的准确性。论文主要研究内容如下：

（1）差分隐私在社交推荐隐私保护中的研究

研究将差分隐私保护方法更好的应用在社交推荐系统中，以实现在保护用户评分数据安全的同时，仍能给用户提供的满意的推荐结果。隐私预算是满足差分隐私的社交推荐中的一个重要参数，隐私预算的高低影响着用户数据隐私保护的程序，同时也对推荐结果的准确性产生影响。因为噪声的加入会使数据不再准确，从而在一定程度上降低社交推荐的效果，所以使用差分隐私时需要对隐私保护和推荐效果加以权衡。差分隐私通过为数据集或算法结果添加噪声来实现隐私保护，但如果使用不恰当，可能会导致噪声过大，从而降低数据可用性和推荐准确性。因此，在将差分隐私引入社交推荐算法时，需要与算法进行有效融合，以确保在提供隐私保护的同时，仍然能够产生准确的推荐结果。

（2）社区划分在社交推荐隐私保护中的研究

研究在进行社区划分时使得社区内相似度提高，不同社区之间的相似度降低。当运用这样的社区数据集来替代个人数据进行推荐时，用户不但能够获取到更为优质的推荐结果，还能够达到隐藏个人信息的目的，而且能够通过对社区里的用户数目进行调整，达成不同程度的隐私保护。常用的社区划分算法有 Louvain 算法、Infomap 算法等，但是它们仅考虑了社交图对社区划分的影响，并未考虑用户偏好图对社区划分的影响。因此，本研究的主要思想是将社交网络里的用户划分成多个结构紧凑（社区里的用户间相似度大，而社区之间的相似度比较小）的

小社区，然后用这样的社区数据来代替用户个人数据进行推荐，从而保护用户的个人隐私。

1.4 论文的组织结构

本论文主要聚焦于社交推荐中的隐私保护问题，并对此展开深入研究。全文共分为五个章节：

第一章，介绍社交推荐隐私保护的研究背景和意义、国内外研究现状、主要研究内容以及章节安排。

第二章，对社交推荐的基本框架进行了简述，并介绍了三种主要的社交推荐模型。阐述了社交推荐所面临的问题，以及现有的社交推荐隐私保护方法和模型。

第三章，针对当前社交推荐方法大多数只考虑了如何提高推荐的准确性，而忽略了对用于研究的原始数据进行隐私保护的问题，提出了一种基于差分隐私的社交推荐隐私保护方法（SRDP）。该方法融入差分隐私保护技术来保证原始数据的隐私安全，同时通过寻找目标用户的最相似用户来保证社交推荐的准确度

第四章，针对当前社交推荐方法大多数只考虑了如何提高推荐的准确性，而忽视了对用户个人信息隐私保护的问题，提出了一种基于社区划分的社交推荐隐私保护方法（SRCD）。该方法引入社区划分算法同时考虑社交关系的影响，将用户个人信息隐藏在社区数据集合中，并通过寻找目标用户所在社区的相似社区来做推荐。

第五章，概括工作内容、总结全文，指出研究中存在的问题与不足，对未来的发展趋势进行展望，并进一步阐述未来的研究方向。

第2章 相关理论与技术介绍

随着社交媒体的普及，人们与家人、朋友和同事之间的社交活动变得更加丰富多彩，社交关系也日益丰富，例如抖音中的关注关系、在腾讯 QQ 上的好友关系等。在线社交关系为个人提供了一种不同的数字交流方式，并允许在线用户与相关联的用户分享想法和意见。用户做出的选择往往与他们在社交圈内的朋友相似或受其影响，这种现象类似于现实世界中的用户在做出购买前会征求朋友的建议。所以，社交推荐正是基于这一思想，提高了传统推荐系统的性能。但与此同时，随着这些数据的收集和分析，用户的隐私也更容易受到侵犯。本章首先介绍社交推荐的基本框架和三种主要的社交推荐模型并进行了模型对比分析，然后阐述社交推荐面临的一些隐私安全问题，最后讨论几种隐私保护技术和模型。

2.1 社交推荐的基本框架

传统推荐主要依赖于对用户基本属性信息和历史交互数据的分析，以协助用户在众多信息中发现他们感兴趣的内容。与之相比，社交推荐则更进一步，将用户的社交信息，如朋友关系、信任程度等纳入考虑。这些信息为推荐模型提供了更丰富的背景信息，使之能够更准确地捕捉用户的兴趣和偏好，从而显著提升推荐系统的整体性能。本节根据社交推荐如何生成推荐结果的过程，给出如下的基本框架图^[33]，如图 2-1 所示：

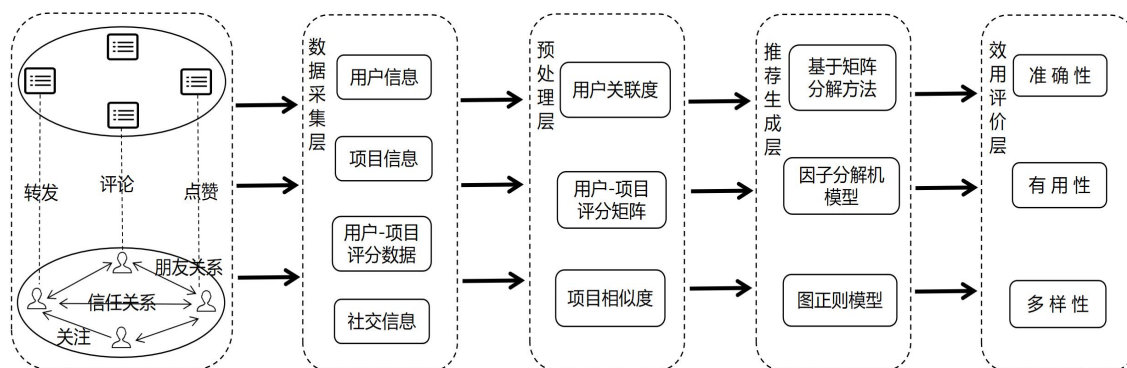


图 2-1 社交推荐的基本框架

Fig.2-1 The basic framework of social recommendation

数据采集层：该层的目标是利用网络爬虫等各种信息技术从社交媒体中获取一系列社交推荐数据。一般情况下，这些数据可以分为三大类别^[34]：基础信息包括年龄、性别、职业等方面的内容；交互信息是由用户对项目的点击、转发、评分等操作所产生的数据；而社交信息则涵盖了用户之间的朋友关系、信任程度等内容。通过收集和分析这些不同类别的数据，可以更全面地了解用户和项目的特征，从而为社交推荐用户提供更精准的推荐服务^[35]。同时，这些数据也有助于揭示用户行为和社交关系的模式，为研究人员提供有关用户行为和社交网络的深入洞察。需要注意的是，为了保护用户隐私，收集和处理这些数据时需要遵守相关的隐私保护法规和规定。同时，数据的使用和共享也需要经过用户的明确授权和合法合规的途径^[36]。对于这些数据的收集和使用，需要保证数据安全性和隐私保护的同时，提供有效的用户选择和控制机制。

预处理层：该层的目标是对上一层获取的数据进行预处理，并将经过处理的结构化数据实时输入到下一层的模型中。在该层中，涉及用户-项目交互矩阵的构建和项目之间的相似度计算，这些计算都非常影响社交推荐的准确性。首先，对用户项目的点击、评分等操作进行分析，可以构建出用户与项目之间的关系^[37]。然后，计算项目相似度^[38]，例如基于内容的相似度计算^[39]、协同过滤算法^[40]等。这些预处理步骤的目的是对原始数据进行整理和转换，以便在下一层的模型中进行进一步的分析和推荐。通过有效的数据预处理，可以提高推荐系统的性能和准确度，从而提供更好的推荐结果。

推荐生成层：该层属于核心层，其主要任务是综合考虑用户与项目的相似度，并将各种类型的社交网络信息引入社交推荐模型中，以直观地体现推荐过程中它们的作用。这是社交推荐系统研究的核心内容^[41]。根据不同的推荐方法特点，社交推荐模型可以进一步划分^[42]，比如基于图学习、矩阵分解和因子分解机等推荐模型。在构建以矩阵分解为基础的推荐系统时，当收集相关的特征向量时，社交网络关系往往对矩阵分解过程中涉及的用户和项目产生影响。通过将社交网络信息与推荐模型相结合，可以更好地利用用户与项目之间的社交关系提升推荐的准确性和个性化程度。这些模型的设计和优化过程中，需要充分考虑社交网络的细节，并将其纳入模型的参数学习和预测过程中。

效用评价层：该层的任务是对得到的推荐结果进行科学合理的综合评价。如果已经向用户提供了推荐结果，那么可以使用一系列相关指标^[43]来评估社交推荐系统的性能，例如准确性、有用性和多样性等。在综合评价中，准确性指标用于衡量推荐结果与用户实际喜好的匹配程度。这可以通过计算推荐物品在用户反馈中的比例或点击率等指标来衡量。有用性指标用于评估推荐结果对用户信息需求的满足程度，即推荐结果是否能够帮助用户发现新颖、有趣的内容。多样性指标则衡量推荐结果的多样性程度，即推荐结果是否能够涵盖不同类型和风格的物品，以提供更广泛的选择。通过对这些指标的评估，可以全面了解社交推荐系统的性能，并根据评估结果进行必要的改进和优化。这可以包括算法调整、数据增强、个性化策略的优化等措施，以提高推荐系统的效果和用户满意度。总之，该层负责科学评估推荐结果，通过使用合适的指标对社交推荐系统的性能进行全面评价，以促进系统的进一步完善和优化。

2.2 社交推荐模型

在社交推荐中，不仅要分析用户的基本信息和交互数据，用户与用户之间的社交关系也会被捕获，以提升推荐效果。将用户之间的社交关系和用户偏好，分别构建成一个社交图 $G_s = (U, E_s)$ 和一个偏好图 $G_p = (U, I, E_p, R)$ ，其中 $U = \langle u_1, u_2, \dots, u_m \rangle$ 表示用户集合，社交边 $(u, v) \in E_s$ 代表两个用户 $u, v \in U$ 之间的社会联系（例如朋友关系）； $I = \langle i_1, i_2, \dots, i_n \rangle$ 表示项目集合，偏好边 $(u, i) \in E_p$

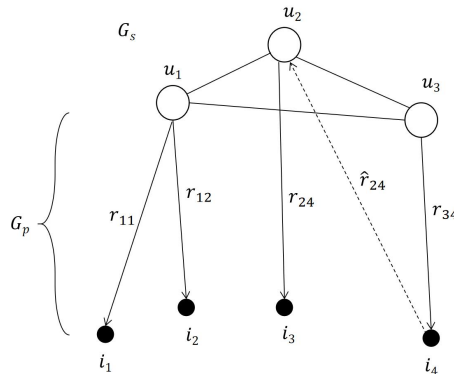


图 2-2 由社交图和偏好图构成的图模型

Fig.2-2 The graph model of social graph and preference graph

表示用户 $u \in U$ 对项目 $i \in I$ 的积极偏好表示， $R = \langle r_1, r_2, \dots, r_n \rangle$ 表示用户对项

目的评分值^[44]。如图 2-2 所示，这两个图构成了社交推荐的输入。

2.2.1 基于矩阵分解的社交推荐模型

在社交推荐算法中，应用矩阵分解技术的目的在于预测用户对未看过的项目的评分^[45]，然后根据评分值给用户做推荐。为清晰起见，保留符号 u 和 v 表示用户， i 和 j 表示项目。 $R = [r_{ui}]_{m \times n}$ 表示用户-项目评分矩阵， r_{ui} 表示用户 u 对项目 i 给出的评分， m 为用户数量， n 为项目数量。基本思想是将矩阵 R 分解成两个低维的小矩阵 U 和 I 的乘积^[46]，使得 $R \approx U^T I$ 。基本原理如图 2-3 所示：

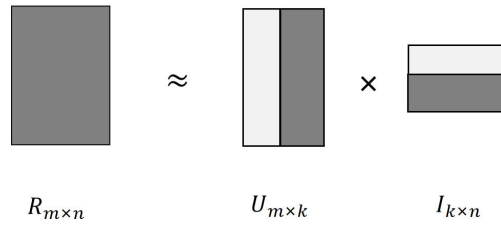


图 2-3 矩阵分解基本原理

Fig.2-3 Basic principles of matrix factorization

也就是最小化下式的值：

$$\frac{1}{2} \|R - U^T I\|_F^2 \quad (2-1)$$

式中： $U = [U_1, U_2, \dots, U_m] \in \mathbb{R}^{d \times m}$ 为用户特征矩阵， U_u 表示用户 u 的特征向量。 $I = [I_1, I_2, \dots, I_n] \in \mathbb{R}^{d \times n}$ 为项目特征矩阵， I_i 表示项目 i 的特征向量。 d 表示特征矩阵维度。 $\|\cdot\|_F^2$ 为 Frobenius 范数。 T 表示矩阵转置符号。

由于评分矩阵 R 通常非常稀疏，即其中包含的评分值项的很少，因此实际上矩阵分解的目标是对评分矩阵中存在的评分值项进行分解。也就是说，要找到合适的用户矩阵 U 和项目矩阵 I ，使得下述目标函数取得最小值：

$$\frac{1}{2} \sum_{u=1}^m \sum_{i=1}^n I_{ui}^R (r_{ui} - U_u^T I_i)^2 \quad (2-2)$$

式中： r_{ui} 表示用户 u 对项目 i 的评分。 I_{ui}^R 是用户 u 评价项目 i 时等于 1，否则等于 0 的指示函数。

在实际操作中，会将正则化项添加到矩阵分解的目标函数里，以此来减轻训练数据的过拟合现象^[47]。正则化项将对参数大小进行惩罚，使得优化过程更倾向

于选择较小的参数值，从而达到良好的泛化性能。同时，寻找合适的用户和项目表示来保持参数的简洁性。在上述目标函数的基础上，加入正则化项后，矩阵分解的目标函数可以表达为：

$$J_{MF} = \frac{1}{2} \sum_{u=1}^m \sum_{i=1}^n I_{ui}^R (r_{ui} - U_u^T I_i)^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_I}{2} \|I\|_F^2 \quad (2-3)$$

式中： λ_U 为用户正则化参数， λ_I 为项目正则化参数。

总之，矩阵分解的目标是通过找到合适的用户和项目表示，最小化评分矩阵中存在的评分值项与预测值之间的差异，以实现稀疏评分矩阵的填补和推荐。

基于矩阵分解的社交推荐，按照矩阵分解的形式，大概可以分成两类^[48]：

(1) 用户特征矩阵共享表示

现有的很多推荐方法在处理非常大的数据集和处理少量评分或无评分用户方面存在着困难^[49]。随着在线社交网络信息的指数级增长，使得社交网络分析技术^[50]在众多 Web 应用中愈发重要。这些技术能够挖掘用户在社交网络中的关系、兴趣和行为，从而实现更精细的推荐过程建模和个性化。SoRec 模型^[51]在矩阵分解时，会融合用户的社交关系信息，同步分解用户的社会关系和项目偏好，进而预测评分。其最终目标函数如下：

$$J_{SoRec} = \frac{1}{2} \sum_{u=1}^m \sum_{i=1}^n I_{ui}^R (r_{ui} - g(U_u^T I_i))^2 + \frac{\lambda_C}{2} \sum_{u=1}^m \sum_{v=1}^m I_{uv}^C (c_{uv} - g(U_u^T Z_v))^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_I}{2} \|I\|_F^2 + \frac{\lambda_Z}{2} \|Z\|_F^2 \quad (2-4)$$

式中： $Z = [Z_1, Z_2, \dots, Z_m] \in \mathbb{R}^{d \times m}$ 表示社交特征矩阵， Z_u 表示用户 u 的社交特征向量。 $C = \{c_{uv}\}$ 表示 G_s 的 $m \times m$ 矩阵，称其为社交信任矩阵。对于一对顶点 p_u 和 p_v ， $c_{uv} \in (0,1]$ 表示与从 p_u 到 p_v 的边相关的权值，否则 $c_{uv}=0$ 。 c_{uv} 表示用户 u 信任或了解用户 v 的程度。 I_{uv}^C 是用户 u 信任或认识用户 v 时等于 1，否则等于 0 的指示函数。 $g(x) = 1/(1 + \exp(-x))$ 是 logistic 函数。 λ_Z 为社交关系正则化参数； λ_C 是平衡来自用户-项目评分矩阵和用户社交信息的参数。

(2) 用户特征矩阵重表示

RSTE 模型^[52]将用户-项目矩阵中的评分 R_{ui} 解释为用户 u 和他/她信任的朋友对项目 i 的偏好的混合表示。其最终目标函数如下：

$$J_{RSTE} = \frac{1}{2} \sum_{u=1}^m \sum_{i=1}^n I_{ui}^R (r_{ui} - g(\lambda_C U_u^T I_i + (1 - \lambda_C) \sum_{v \in \mathcal{T}(u)} C_{uv} U_v^T I_i))^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_I}{2} \|I\|_F^2 \quad (2-5)$$

式中： $\mathcal{T}(u)$ 为用户 u 信任的朋友集合， $|\mathcal{T}(u)|$ 为集合 $\mathcal{T}(u)$ 中用户 u 信任的朋友的个数。

Jamali 等人在之前的工作基础上，提出了一个名为 SocialMF^[53]的社交评分网络推荐模型。在该模型中，用户的潜在向量会受到社交关系的作用。其最终目标函数如下：

$$J_{SocialMF} = \frac{1}{2} \sum_{u=1}^m \sum_{i=1}^n I_{ui}^R (r_{ui} - g(U_u^T I_i))^2 + \frac{\lambda_U}{2} \sum_{u=1}^m U_u^T U_u + \frac{\lambda_I}{2} \sum_{i=1}^n I_i^T I_i + \frac{\lambda_C}{2} \sum_{u=1}^m ((U_u - \sum_{v \in \mathcal{T}(u)} C_{uv} U_v)^T (U_u - \sum_{v \in \mathcal{T}(u)} C_{uv} U_v)) \quad (2-6)$$

Ma 等人提出 SoReg^[54]矩阵分解模型，该模型将用户间的社交信息进行融合，使得用户的潜在向量与他们信任的朋友相似。其最终目标函数如下：

$$J_{SoReg} = \frac{1}{2} \sum_{u=1}^m \sum_{i=1}^n I_{ui}^R (r_{ui} - U_u^T I_i)^2 + \frac{\lambda_C}{2} \sum_{u=1}^m \sum_{v \in \mathcal{T}(u)} C_{uv} \|U_u - U_v\|_F^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_I}{2} \|I\|_F^2 \quad (2-7)$$

Guo 等人提出一种基于信任的矩阵分解模型 TrustSVD^[55]，其最终目标函数如下：

$$J_{TrustSVD} = \frac{1}{2} \sum_{u=1}^m \sum_{i \in I_u} (\hat{r}_{ui} - r_{ui})^2 + \frac{\lambda_C}{2} \sum_{u=1}^m \sum_{v \in \mathcal{T}_u} (\hat{c}_{uv} - c_{uv})^2 + \frac{\lambda_I}{2} \sum_{u=1}^m |I_u|^{-\frac{1}{2}} b_u^2 + \frac{\lambda_U}{2} \sum_{i=1}^n |U_i|^{-\frac{1}{2}} b_i^2 + \sum_{u=1}^m (\frac{\lambda_I}{2} |I_u|^{-\frac{1}{2}} + \frac{\lambda_C}{2} |CT_u|^{-\frac{1}{2}}) \|u_u\|_F^2 + \frac{\lambda_U}{2} \sum_{i=1}^n |U_i|^{-\frac{1}{2}} \|i_i\|_F^2 + \frac{\lambda_U}{2} \sum_{i=1}^n |U_i|^{-\frac{1}{2}} \|y_i\|_F^2 + \frac{\lambda_Z}{2} |T_v^+|^{-\frac{1}{2}} \|w_v\|_F^2 \quad (2-8)$$

其中： $\hat{r}_{ui} = b_u + b_i + \mu + q_i^T (u_u + |I_u|^{-\frac{1}{2}} \sum_{i \in I_u} y_i)$ 表示预测的用户 u 对项目 i

的评分： $\hat{c}_{uv} = w_v^T u_u$ 表示用户 u 与 v 之间的预测信任。

两种矩阵分解模型方法分析对比结果如表 2-1 所示：

表 2-1 基于矩阵分解的社交推荐模型的方法对比

Table 2-1 Comparison of methods of social recommendation models based on matrix factorization

推荐方法	优势之处	待完善之处
用户特征矩阵共享表示	通过利用朋友的偏好信息，可以提高推荐的效果，提供更加准确和有针对性的推荐结果。	由于推荐涉及因素复杂多样，可能无法真实反映推荐过程。
用户特征矩阵重表示	在一定程度上提升了推荐效果。	需要学习的参数多。

2.2.2 基于图学习的社交推荐模型

本节将对基于图学习的社交推荐方法进行系统地归纳，主要包含以下三种：

(1) 图表示学习方法

在融合社交信息的推荐系统中，通过图表示学习^[57]，能够利用信任关系图揭示用户之间的信任程度，并结合用户交互图中的行为数据进行建模，以更好地理解用户的喜好和兴趣。Liu 等^[58]提出了一种基于图表示学习的多视角社交推荐方法（MPSR）。该方法通过从不同角度构建分层用户偏好，并分配具有不同信任水平的朋友对用户的推荐结果产生影响，提高了推荐的个性化程度。为了准确衡量朋友的信任程度，MPSR 利用用户历史行为的统计信息构建了多视角的社交网络，以便更好地理解用户之间的关系和信任。通过这些手段，MPSR 能够综合考虑用户的多个方面，提供更精准的推荐结果。Liu 等^[59]提出了一种称为 SMEF 的新型社会感知多维边缘特征的图表示学习方法。该方法将所有用户和项目表示为图形，并利用深度学习技术学习多维边缘特征，以明确描述每个用户-项目对之间的任务特定关系。具体而言，SMEF 方法关注两种不同类型的用户行为模式，即社交朋友和互动项目，通过探索这些行为模式中潜在的异质关系线索，学习到的多维边缘特征能够从社交和交互两个方面对用户信息进行编码。通过这种方式，SMEF 能够更准确地捕捉用户之间的关联性和用户与项目之间的特定关系，从而

提高推荐系统的性能和个性化程度。

（2）随机游走方法

在引入社交信息的推荐系统里,随机游走^[60]是通过模拟用户在图上的行为轨迹,来获取用户和物品间的邻接关系,并且借助用户之间的社交关系图来增强推荐效果。Yang 等^[61]提出了一种将反向搜索和随机游走相结合的好友推荐方法。首先,基于 FP-Growth 重构转移概率矩阵,来呈现直观的社交信息。然后,定义了包含反向社会影响的反向搜索策略,并与随机游走共同作用来推荐朋友。Guo 等^[62]提出了一种基于多因素随机游走的协同过滤推荐方法,该方法以评分时间、用户属性偏好、共同好友数等指标为基础,更精确地计算出当前用户对其他用户的综合信任值,进而确定当前用户的信任邻域。如此,便能便捷、准确地获取与当前用户偏好相似的可信用户作为相邻用户,通过多个相邻用户对目标项目或类似项目的评分,得到目标项目的最终预测评分。此外,为了避免在游走过程中产生影响预测准确性的评级,会根据当前游走的综合信任值、项目相似度和深度来做出停止走动的决策,进一步提升推荐准确率。

（3）图神经网络方法

图神经网络^[63]是一种新型的拓展神经网络,受到卷积神经网络和图嵌入思想的启发,专门用于图领域数据的特征提取和表示。作为一种高效且易于扩展的神经网络结构,图神经网络在学习图数据方面展示出强大的能力。近年来,由于图神经网络提供了更好的表示方式,能够帮助人们分析推荐系统中用户之间的关系,它得到了长足的发展。Tallapally 等^[64]提出了一种基于 Diffnet++ 的新型社交推荐神经影响和扩散算法 (RelationalNet)。该算法对 Diffnet++ 进行了改进,增加了项目-项目和项目-用户交互,这些交互建模为图神经网络中使用的图形。不仅通过以社交图谱的形式整合用户-用户关系 (类似于其他社交推荐系统),而且还整合了独立于任何用户交互的项目-项目关系。此外,该算法利用多层扩散网络,使用图注意力来组合图和节点级表示,设法捕获每个图的潜在特征。Wei 等^[65]提出了一种名为社交驱动的图分层注意网络 (SGHAN),它是一个基于深度学习的模型。SGHAN 具备两个核心组件,它们相互作用以推动模型的学习过程。首先,服务级注意力编码器捕捉用户在朋友的 mashup 中对哪些服务感兴趣。其次,朋友级图形注意力网络选择具有信息价值的朋友,并传播朋友的社会影响。

通过这种方式，SGHAN 能够利用社交关系来影响用户的行为和决策，提供更加个性化和准确的服务推荐。该模型使得 mashup 开发过程更加智能化和自适应，提高了用户的满意度和体验。

表 2-2 基于图学习的社交推荐模型的方法对比

Table 2-2 Comparison of methods of social recommendation models based on graph learning

推荐方法	优势之处	待完善之处
图表示学习方法	能够与已有的分类算法和复杂网络分析算法进行结合。	由于图结构的复杂性和噪声的存在，图表示学习方法对数据中的变化和干扰比较敏感，可能导致嵌入表示的不稳定性和可靠性下降。
随机游走方法	在推荐系统中具有灵活性和通用性，它允许结合多样化的分类算法和监督学习算法。	需要考虑到所有用户朋友对目标物品的评分，计算复杂度较高，可能会导致性能下降或者无法扩展到大规模数据集。
图神经网络方法	具有学习非线性特征的优势，同时，也能够处理图的拓扑结构信息，从而更好地理解节点之间的关系。	在实际社交推荐场景中，社交图和用户-物品交互图往往包含着数十亿个节点和边，因此计算成本很高。

2.2.3 基于因子分解机的社交推荐模型

因子分解机 FM 模型^[66]是在逻辑回归 LR 和矩阵分解的基础上发展的结果。因此，首先介绍一下逻辑回归，它是将所有特征进行线性加权，再加入 Sigmoid 逻辑函数：

$$\hat{y}(x) = w_0 + w_1x_1 + \dots + w_nx_n = w_0 + \sum_{i=1}^n w_ix_i \quad (2-9)$$

式中： w_0 为全局偏置， w_i 是模型第 i 个变量的权重。

而 POLY2 模型则在 LR 的基础上将所有特征两两相交，则变成：

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_ix_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij}x_ix_j \quad (2-10)$$

式中： w_{ij} 是特征组合 (i, j) 的交叉权重。

FM 就是在 POLY2 模型的基础上，融合了矩阵分解的思想，即通过矩阵分

解的方式来调整二阶交叉特征的系数:

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \quad (2-11)$$

式中: $w_0 \in \mathbb{R}$, $w_i \in \mathbb{R}$ 。 $V \in \mathbb{R}^{n \times k}$, V 是一个 $n \times k$ 的向量矩阵, n 是特征 x 的个数, v_i 是第 i 维特征的隐向量。 k 为隐向量的长度。 $\langle \cdot, \cdot \rangle$ 是长度为 k 的两个向量的点乘, 公式如下:

$$\langle v_i, v_j \rangle = \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (2-12)$$

对比 FM 与 POLY2, 可以看出两者的主要区别为 FM 用两个向量的内积 $\langle v_i, v_j \rangle$ 代替了单一的权重系数 w_{ij} 。好处在于: 为每个特征学习了一个隐向量权重, 而这种隐向量的计算方式可以更好地解决数据的稀疏问题, 从而提高模型的泛化能力。并且可将权重参数从 POLY2 的 n^2 级别下降到 nk 级别 (k 为隐向量的维度, $n \gg k$)。Ding 等^[67]提出了一种名为 SCFM 的方法, 它是一种改进的因子分解机模型, 结合了社交和人群信息。该方法将社交计算和群体计算融入因子分解机。通过将社交计算和群体计算相结合, SCFM 可以更好地捕捉用户之间的关系并获取潜在主题的信息, 从而提高推荐系统的性能和准确性。

三种主要的社交推荐模型分析对比结果如表 2-3 所示:

表 2-3 三种主要的社交推荐模型对比

Table 2-3 Comparison of three main social recommendation models

社交推荐模型	核心算法	适用推荐场景
基于矩阵分解的社交推荐模型	矩阵分解、信任推荐、最小二乘	基于项目评分预测的推荐应用
基于图学习的社交推荐模型	图的遍历、分类算法、随机漫步算法	基于项目评分预测的推荐应用
基于因子分解机的社交推荐模型	因子分解、最小二乘、贝叶斯推理	基于项目评分预测的推荐应用、各种类型的社交网络推荐应用

2.3 社交推荐面临的隐私问题

为了提供更准确的个性化推荐,社交推荐需要收集更多信息,如年龄、性别、行为数据、社交关系信息等。然而,如果这些数据被处理不当或未经用户同意的情况下与第三方共享,可能会带来隐私风险。用户往往低估或忽视与个人数据收集和处理相关的隐私风险。许多用户不知道自己的数据是否被收集,被收集了多少数据,以及这些数据是否被出售给第三方,或者其存储安全性和存储时间如何。尽管社交推荐系统能够有效地帮助用户在海量数据中筛选出有价值的信息,但其潜在的信息泄露风险却不容忽视。用户的购物记录、评分历史等敏感数据都需要上传至推荐系统服务器。这些数据若被泄露,攻击者极有可能轻松获取并滥用用户的个人隐私。因此,保护用户隐私信息成为社交推荐系统面临的一项重大挑战。隐私风险^[68]可能包括服务提供商未经授权访问用户信息、非法数据泄露、将用户信息出售给第三方或黑客事件等。因此,需要通过设计和实施隐私保护技术来确保用户信息的安全,并有效防止这些隐私风险的发生。

现有的隐私保护社交推荐方法通常允许推荐者完全控制用户的信息。然而,这可能存在问题,因为推荐者本身就可能不被信任,导致严重的隐私泄露。此外,建立社交关系通常需要与他人共享兴趣和其他私人信息,这可能进一步增加隐私泄露的风险。虽然有时允许用户使用隐私设置来隐藏敏感的私人数据,但仍然有可能共享的数据被攻击者滥用来推断敏感的私人信息,并且还可能导致推荐服务质量变差。因此,为了平衡推荐性能和用户隐私保护,必须在社交推荐的设计和 implement 中重视用户的隐私权,并采取相应的隐私保护措施来保证用户信息的安全。

2.4 社交推荐中使用的隐私保护方法

通常,差分隐私、匿名化和联邦学习等方法被用于社交推荐的隐私保护。差分隐私通过引入噪声或扰动的方式来保护用户隐私。匿名化使得在推荐系统中无法直接将个人身份与推荐结果相关联。联邦学习则是一种通过在本地设备上进行模型训练,将个别用户数据保留在本地而非上传到云端的方法来保护用户隐私。

2.4.1 差分隐私保护

(1) 差分隐私的定义

差分隐私保护技术^[69]，其基本思想是：无法通过删除数据集中的任意一条数据的输出差异来推测个体信息。形式化定义如下：

$$\frac{P_r[A(D) \in S]}{P_r[A(D') \in S]} \leq e^\epsilon \quad (2-13)$$

式中： $P_r[\cdot]$ 是隐私泄露的概率； A 是随机算法； D 和 D' 是相邻数据集； S 是任意查询集合； ϵ 是隐私保护的参数。

(2) 敏感度

噪声的加入量与敏感度有关。敏感度的值是删除数据集中任意一条数据后查询结果的最大改变量。

(3) 实现机制

定义 1 (Laplace 机制^[70]) 给定数据集 D ，设有函数 f ，其敏感度为 Δf ，那么随机算法 $A(D) = f(D) + \text{Lap}(\Delta f/\epsilon)$ 提供 ϵ -差分隐私保护。

定义 2 (指数机制^[71]) 给定数据集 D ，设有随机算法 A ，输出 $r \in \text{Range}$ ， Δq 为函数 $q(D, r)$ 的敏感度。若算法 A 以正比于 $\exp(\frac{\epsilon \times q(D, r)}{2\Delta q})$ 的概率从 Range 中选择并输出 r ，那么算法 A 提供 ϵ -差分隐私保护。

(4) 性质

为了解决一个复杂的隐私保护问题一般需要多次应用差分隐私保护算法，所以需要将隐私预算 ϵ 合理地分配到各个算法步骤中。

性质 1 (序列组合性质) 设有隐私算法 $A = \{A_1, A_2, \dots, A_n\}$ ，分别满足 ϵ_i -差分隐私 ($1 \leq i \leq n$)，那么对于同一数据集，由算法 A 构成的组合算法提供 $(\sum_{i=1}^n \epsilon_i)$ -差分隐私保护。

性质 2 (并行组合性质) 设有隐私算法 $A = \{A_1, A_2, \dots, A_n\}$ ，分别满足 ϵ_i -差分隐私 ($1 \leq i \leq n$)，那么对于一组不相交的数据集，由算法 A 构成的组合算法提供 $(\max \epsilon_i)$ -差分隐私保护。

2.4.2 匿名化

对数据实施匿名化^[72]处理的目的是为了让数据的使用方没办法知晓数据的来源，进而对用户的隐私予以保护。它有三种常用的技术变体： k -anonymity^[73]是一种常用的匿名化技术，通过确保每条发布的记录与至少 $k-1$ 条其他记录没有区别（如果这些记录直接链接到外部信息），以此来达成隐私保护的目。实现 k -anonymity 有两种主要的方式，即抑制和泛化。 k -anonymity 可以有效防止链接攻击，攻击者无法通过将准标识符属性与外部信息联系起来唯一识别用户。 l -diversity^[74]是另一种匿名化技术，它确保在每个匿名化类别中存在至少有 l 个不同值的敏感属性，以便对每个类别当中用户相互之间的隐私进行保护。当需要匿名化的相关数据集具备敏感属性时，可以使用 t -closeness^[75]。 t -closeness 可以防止属性的同质性和接近性，这是 l -diversity 的一种局限。它通过将类别中敏感属性的分布限制在接近全局分布的因子 t 上，提高推荐系统的准确性。

2.4.3 联邦学习

联邦学习提供了一种新颖的设计思路，在许多领域都能发挥重要作用。它允许多方参与协作训练机器学习模型，而无需共享原始数据。同时，参与方将经过脱敏处理的数据发送到中心服务器进行聚合，以实现学习目标^[76]。联邦学习能够和推荐系统联合运用，进而达成联邦推荐隐私保护算法的实现^[77]。在这种情况下，联邦推荐算法可以在模型构建过程中一直将用户数据保留在客户端本地，从而保护用户隐私。随着对用户隐私保护需求的不断增加，联邦学习有望在更多应用场景中得到广泛应用。

（1）横向联邦学习

横向联邦学习^[78]，适用于特征重叠多而用户重叠少的场景。例如，在跨地区的银行业务合作中，由于各银行的业务类型和特征可能非常相似，但用户群体并不完全相同，通过横向联邦学习，这些银行可以在不必直接共享原始用户数据的情况下，共同训练和优化推荐模型。

（2）纵向联邦学习

纵向联邦学习^[79]，适用于用户重叠较多而特征重叠较少的场景。例如，同一

地区的银行和电商，虽然他们的业务不同，但服务的用户群体却存在大量重叠。通过纵向联邦学习，银行和电商可以合作共享用户特征信息，从而为用户提供更丰富、更个性化的推荐服务，满足用户多样化的需求和偏好。

(3) 联邦迁移学习

联邦迁移学习^[80]主要适用于物品相同且用户重叠数量较少的场景。例如，一家位于中国的电商和一家位于美国的银行，他们的业务和地域都不同，通过引入迁移学习，这两家机构可以在保护各自数据隐私的同时，共享和迁移部分共同特征的知识。这样，即使在没有大量共同数据的情况下，也能提升模型的性能和准确性。以预测用户信用评分为例，中国银行可以通过借鉴美国电商的数据和模型，来改善其信用评分模型的性能，而无需公开或共享敏感的用户数据。这不仅增强了模型的泛化能力，也促进了不同机构间的合作和知识共享。

将三种隐私保护技术的适用场景和优缺点进行了对比，如表 2-4 所示：

表 2-4 社交推荐的隐私保护技术比较

Table 2-4 Comparison of privacy protection techniques in social recommendation

隐私保护技术	实现方法	主要特点		
		使用场景	优点	缺点
差分隐私保护	拉普拉斯机制、高斯机制、指数机制	数据对外发布或提交给数据收集者	基于严谨的数学原理、隐私保护能力强、适用面广	保护隐私的同时会降低数据的可用性
匿名化	不署名或用假名、成组提供数据	大量用户需要提交数据的场景	实现简单	隐私保护能力弱、适用面窄
联邦学习	一种学习框架，在这种框架中，多方协作训练模型，而不共享原始数据	需要大量用户数据支撑的推荐场景	应对某种特定攻击的能力强	用户与服务器之间的交流技术量大，上传的梯度信息仍有可能被攻击

2.5 本章小结

在本章中，首先深入且详细地阐述了与社交推荐有关的知识，内容涵盖了社交推荐的基本框架、相关模型以及其他方面。然后对社交推荐所面临的隐私问题进行了探讨，并介绍了在社交推荐中应用的多种隐私保护技术。特别介绍的是：差分隐私保护技术、匿名化技术以及联邦学习这三种关键技术。

第3章 社交推荐中的差分隐私保护方法

社交推荐主要通过对用户基本信息和社交信息的分析，以实现准确推荐。然而，这可能会导致用户隐私的泄露。因此，在确保推荐准确的前提下，如何保护用户的隐私成为一个急需解决的问题。针对这一问题，提出了一种基于差分隐私的社交推荐隐私保护方法（SRDP）。

3.1 方法的提出

为了在社交推荐中实现对用户原始数据的保护，提高数据的安全性，并减少数据可用性的损失，本章提出了一种基于差分隐私的社交推荐隐私保护方法。该方法首先通过同时考虑社交关系和用户相似度来寻找目标用户的相似用户。然后，将目标用户和相似用户所看过的项目及其评分构造成用户-项目评分矩阵。针对矩阵的每一列评分，采用差分隐私技术中的指数机制来选择中位数，并在其基础上添加拉普拉斯噪声，从而得到添加了噪声的用户-项目评分矩阵。将这个模糊化的用户-项目评分矩阵作为社交推荐算法的输入，并将矩阵分解的结果作为给目标用户的预测评分，以此来评估算法的性能。研究框架如图 3-1 所示：

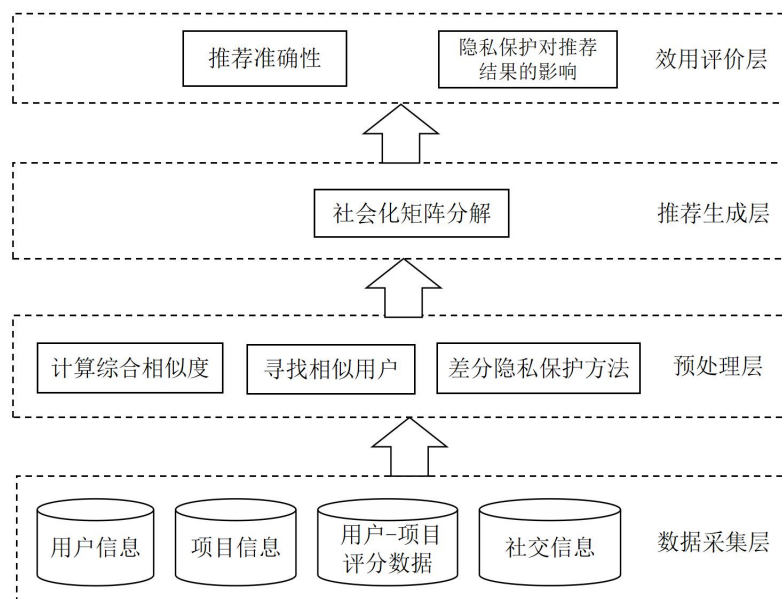


图 3-1 社交推荐中的差分隐私保护方法研究框架

Fig.3-1 Research framework of differential privacy protection in social recommendation

3.2 方法的描述

SRDP 算法包括两个主要的步骤。其中，步骤 1：通过考虑社交网络中用户之间的社交关系和用户相似度，来计算用户之间的综合相似度，从而为目标用户寻找 k 个与其最为相似的用户；步骤 2：利用实现差分隐私保护方法中的拉普拉斯机制和指数机制，来保护用户的真实评分。具体做法是：将目标用户以及步骤 1 中找到的 k 个相似用户，还有他们对项目的评分，构建成用户-项目评分矩阵。对于该矩阵的每一列项目评分，使用指数机制选择中位数，然后利用拉普拉斯机制对中位数添加噪声，并用添加噪声后的中位数替换该列项目的评分，最终得到模糊化的用户-项目评分矩阵。总体算法流程图如图 3-2 所示：

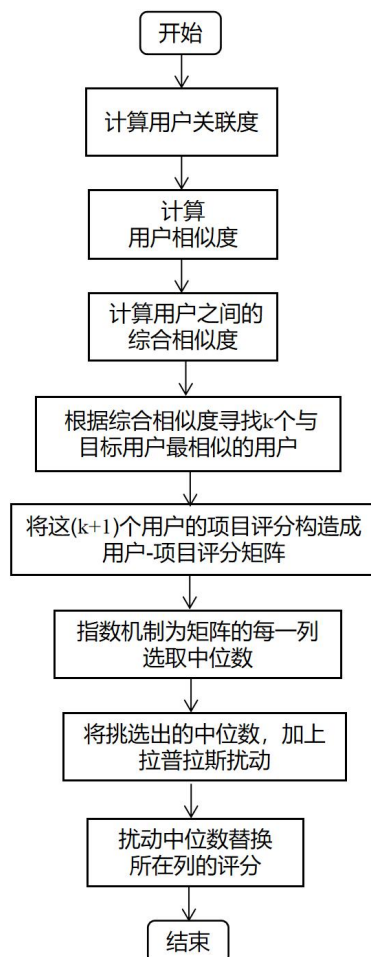


图 3-2 SRDP 算法流程图

Fig.3-2 SRDP algorithm flow chart

3.2.1 寻找相似用户

步骤 1 是利用社交关系和用户相似度来计算用户之间的综合相似度，然后利用用户之间的综合相似度来寻找目标用户的 k 个最相似用户。本节设计如下算法，如表 3-1 所示：

表 3-1 寻找相似用户算法

Table 3-1 Find similar users algorithm

Input: 社交图 G_s ，偏好图 G_p ，相似用户数 k
Output: 目标用户和 k 个与之最相似的用户列表U
1.for u in $G_s.nodes$:
2. for v in $G_s.nodes$:
3. if $u \neq v$:
4. 利用式 (3-1) 计算关联度 $sim(u, v)$
5. if u, v 没有共同项目:
6. 综合相似度 = $sim(u, v)$
7. else:
8. 利用式 (3-2) 计算用户相似度 $sim_i(u, v)$
9. 综合相似度 = $sim(u, v) + sim_i(u, v)$
10.for u in $G_s.nodes$:
11. U 是 u 的 k 个最大综合相似度用户集合
12.return U

为了向目标用户提供准确的推荐，寻找 k 个与目标用户最相似的用户。具体做法如下：首先，依据式 (3-1) 在社交图中计算用户 u, v 之间的关联度：

$$sim(u, v) = \begin{cases} \frac{|t_{u,v}|}{|t_u + t_v - 2|}, & \text{节点 } u, v \text{ 相连} \\ \frac{|t_{u,v}|}{|t_u + t_v|}, & \text{节点 } u, v \text{ 不相连} \end{cases} \quad (3-1)$$

式中： t_u 为用户 u 的邻居， t_v 为用户 v 的邻居； $t_{u,v}$ 表示用户 u 与用户 v 的共同邻居。

例 3-1. 如图 3-3 所示，每个节点表示一个用户。假设目标用户是用户 1，用户 1 与用户 2 之间的关联度为 $\text{sim}(1,2) = \frac{2}{3+3-2} = \frac{1}{2}$ ；用户 1 和用户 5 之间的关联度为 $\text{sim}(1,5) = \frac{1}{3+3} = \frac{1}{6}$ 。

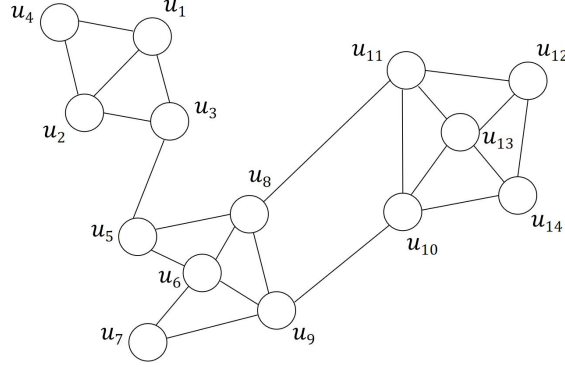


图 3-3 社交图模型

Fig.3-3 Social graph model

然后，计算用户之间的相似度。通常会采用欧几里得距离^[81]、余弦相似度^[82]等方法来计算用户之间的相似度。然而需要考虑这样一种情形：即便两个用户看过同一个项目，可针对该项目给出的评分却不相同。

例3-2. 假设用户1看了项目1，且给出的评分是1分，用户2同样看了项目1不过评分高达5分。如此情形之下，项目1便无法被纳入用户1与用户2的用户相似度计算里。

因此，本章考虑了用户评分分布范围对用户相似度的影响，并利用式(3-2)来计算用户相似度：

$$\text{sim}_i(u, v) = \frac{|d_{uv}|}{|d_u \cap d_v|} \quad (3-2)$$

式中： d_u 为用户看过的项目， d_v 为用户 v 看过的项目； d_{uv} 为用户 u 与用户 v 的相似项目。

例3-3. 假设用户1看过的项目是项目1、5、7、9、11、12，用户2看过的项目是项目1、5、11、12。那么用户1和用户2看过的共同项目是项目1、5、12。用户1和用户2的给项目的评分情况如表3-2所示：

表 3-2 用户 1 和用户 2 的对项目的评分数据

Table 3-2 Items ratings data of user 1 and user 2 on the item

用户/项目	1	5	7	9	11	12
1	1	1	3	2		4
2	2	5			3	5

可以看到，用户1和用户2给项目1的评分在[1,2]区间里，给项目12的评分在[3,5]区间里，但是他们给项目5的评分既不在[1,2]区间里也不在[3,5]区间里，因而用户1和用户2看过的相似项目数量为2，那么用户1与用户2之间的项目相似度为 $sim_i(u, v) = \frac{2}{5+4-3} = \frac{1}{3}$ 。

最后，将两个用户之间的关联度与相似度进行相加计算，得到的结果就是两个用户之间的综合相似度。那么，用户1与用户2之间的综合相似度最终就是 $\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ 。

按照以上方法，计算出各用户之间的综合相似度。找出与目标用户综合相似度最高的 k 个用户，他们就是与目标用户最相似的 k 个用户。

例3-4. 假设给用户1做推荐且 $k=10$ ，即寻找10个与用户1最相似的用户。假设用户5是与用户1的综合相似度最大的一个用户，那么就先将用户1和用户5划分在一个社区。然后继续寻找与这个社区综合相似度最大的第二个用户，即比较用户1与第二个用户的综合相似度加上用户5与第二个用户的综合相似度，相加值最大的那个用户，即是要寻找的第二个用户。按照以上方法，依次寻找，直到找到10个用户，并将这10个用户划分在同一个社区。

时间复杂度：假设有 m 个用户节点，算法1首先计算社交网络中用户之间的综合相似度，运行时间为 $O(m^2)$ ；然后寻找目标用户的 k 个最相似的用户，时间复杂度为 $O(mk)$ 。 k 远远小于 m ，所以算法1的时间复杂度为 $O(m^2)$ 。

3.2.2 保护用户真实评分

步骤2 首先将目标用户和步骤1中寻找的 k 个相似用户及其项目评分构造成用户-项目评分矩阵，然后采用差分隐私保护技术来确保用户的真实评分得到保护。本节设计如下算法，如表3-3所示。

表 3-3 保护用户真实评分算法

Table 3-3 Protect users' real ratings algorithm

输入：	用户列表 U ，用户-项目评分 r_{ui} ，隐私预算 ϵ
输出：	模糊化的用户-项目评分 $\tilde{r}_{ui}$
1.	$\epsilon \leftarrow \epsilon_1 + \epsilon_2$; $I(U)$ 为 U 的项目集合;
2.	for i in $I(U)$:
3.	for u in U :
4.	以 $\exp(\frac{-\epsilon_1 R(\tilde{r}_{ui}) - R(r_{ui}) }{2})$ 的概率在 R_i 中抽取中位数，记为 $r_{r_{ui}}$;
5.	$\tilde{r}_{r_{ui}} = r_{r_{ui}} + Lap(1/\epsilon_2)$
6.	用 $\tilde{r}_{r_{ui}}$ 替换 r_{ui}
7.	return $\tilde{r}_{ui}$

具体做法：首先将目标用户和步骤1中寻找的 k 个相似用户及其项目评分，构造成用户-项目评分矩阵。然后采用指数机制对矩阵的每一列项目评分抽取评分的中位数（第4行），再采用拉普拉斯噪声扰动抽取出的中位数（第5行），最后将噪声中位数替换该列项目的所有评分（第6行）。本文不考虑差分隐私保护参数 ϵ 的分配对结果的影响，使 $\epsilon_1 = \frac{\epsilon}{2}$ ， $\epsilon_2 = \frac{\epsilon}{2}$ 。

例3-5. 假设在算法1找到用户1的9个最相似的用户，将用户1和这9个相似用户及其对项目的评分，构造成如下表所示的矩阵：

表 3-4 用户-项目评分矩阵

Table 3-4 Users-items ratings matrix

用户/项目	1	2	3	5	9	11	20	...
1	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$			$r_{1,11}$		...
3	$r_{3,1}$	$r_{3,2}$		$r_{3,5}$	$r_{3,9}$		$r_{3,20}$	...
4					$r_{4,9}$	$r_{4,11}$		...
51			$r_{51,3}$	$r_{51,5}$	$r_{51,9}$	$r_{51,11}$	$r_{51,20}$	...
828	$r_{828,1}$		$r_{828,3}$			$r_{828,11}$	$r_{828,20}$	...
876								...
897	$r_{897,1}$			$r_{897,5}$		$r_{897,11}$		...
1197			$r_{1197,3}$				$r_{1197,20}$	...
1213	$r_{1213,1}$	$r_{1213,2}$			$r_{1213,9}$			...
1346	$r_{1346,1}$	$r_{1346,2}$		$r_{1346,5}$	$r_{1346,9}$	$r_{1346,11}$		...

运用差分隐私技术中的指数机制，选取矩阵各列的项目评分中位数。假设用户 1 和与之最相似的 9 个用户对项目 1 的评分分别为 $R_1 = \langle 0, 0, 0, 0, r_{1213,1}, r_{3,1}, r_{1,1}, r_{9,1}, r_{897,1} \rangle$ ，并且已经按照升序排列。 $\bar{r}_{ui}$ 为 R_i 中的真实评分中位数，在 R_i 中任取一个评分，记为 $r_{r_{ui}}$ ，根据指数机制， $r_{r_{ui}}$ 被选取的概率为 $P_r(x) \propto \exp(-\varepsilon_1 |R(\bar{r}_{ui}) - R(r_{r_{ui}})|/2)$ ，其中 $R(r_{r_{ui}})$ 表示 $r_{r_{ui}}$ 在 R_i 中的排名。因此，使用指数机制可以选择与 $\bar{r}_{ui}$ 最接近的值，也可以保护 $\bar{r}_{ui}$ 的隐私。再将 $r_{r_{ui}}$ 采用拉普拉斯噪声扰动，得到扰动评分中位数 $\tilde{r}_{r_{ui}}$ ，用 $\tilde{r}_{r_{ui}}$ 替换用户 u 对项目 i 的评分。这样就得到了模糊化的用户-项目评分矩阵，如表 3-5 所示：

表 3-5 模糊化的用户-项目评分矩阵

Table 3-5 Fuzzy users-items ratings matrix

用户/项目	1	2	3	5	9	11	20	...
1	$\tilde{r}_{1,1}$	$\tilde{r}_{1,2}$	$\tilde{r}_{1,3}$	$\tilde{r}_{1,5}$	$\tilde{r}_{1,9}$	$\tilde{r}_{1,11}$	$\tilde{r}_{1,20}$	...
3	$\tilde{r}_{3,1}$	$\tilde{r}_{3,2}$	$\tilde{r}_{3,3}$	$\tilde{r}_{3,5}$	$\tilde{r}_{3,9}$	$\tilde{r}_{3,11}$	$\tilde{r}_{3,20}$	...
4	$\tilde{r}_{4,1}$	$\tilde{r}_{4,2}$	$\tilde{r}_{4,3}$	$\tilde{r}_{4,5}$	$\tilde{r}_{4,9}$	$\tilde{r}_{4,11}$	$\tilde{r}_{4,20}$	...
51	$\tilde{r}_{51,1}$	$\tilde{r}_{51,2}$	$\tilde{r}_{51,3}$	$\tilde{r}_{51,5}$	$\tilde{r}_{51,9}$	$\tilde{r}_{51,11}$	$\tilde{r}_{51,20}$	...
828	$\tilde{r}_{828,1}$	$\tilde{r}_{828,2}$	$\tilde{r}_{828,3}$	$\tilde{r}_{828,5}$	$\tilde{r}_{828,9}$	$\tilde{r}_{828,11}$	$\tilde{r}_{828,20}$	...
876	$\tilde{r}_{876,1}$	$\tilde{r}_{876,2}$	$\tilde{r}_{876,3}$	$\tilde{r}_{876,5}$	$\tilde{r}_{876,9}$	$\tilde{r}_{876,11}$	$\tilde{r}_{876,20}$	...
897	$\tilde{r}_{897,1}$	$\tilde{r}_{897,2}$	$\tilde{r}_{897,3}$	$\tilde{r}_{897,5}$	$\tilde{r}_{897,9}$	$\tilde{r}_{897,11}$	$\tilde{r}_{897,20}$	...
1197	$\tilde{r}_{1197,1}$	$\tilde{r}_{1197,2}$	$\tilde{r}_{1197,3}$	$\tilde{r}_{1197,5}$	$\tilde{r}_{1197,9}$	$\tilde{r}_{1197,11}$	$\tilde{r}_{1197,20}$	...
1213	$\tilde{r}_{1213,1}$	$\tilde{r}_{1213,2}$	$\tilde{r}_{1213,3}$	$\tilde{r}_{1213,5}$	$\tilde{r}_{1213,9}$	$\tilde{r}_{1213,11}$	$\tilde{r}_{1213,20}$	...
1346	$\tilde{r}_{1346,1}$	$\tilde{r}_{1346,2}$	$\tilde{r}_{1346,3}$	$\tilde{r}_{1346,5}$	$\tilde{r}_{1346,9}$	$\tilde{r}_{1346,11}$	$\tilde{r}_{1346,20}$	...

时间复杂度：假设目标用户的相似用户数为 k ，项目数为 p ，则构造用户-项目评分矩阵的时间复杂度为 $O(kp)$ 。用户-项目评分矩阵的大小：假设用户的数量为 m ，项目的数量为 p ，则用户-项目评分矩阵的大小为 np 。对于每一列项目评分，需要抽取中位数并添加拉普拉斯噪声，时间复杂度为 $O(m\log(m))$ 。因此，整个过程的时间复杂度为 $O(pm\log(m))$ 。所以，该算法的复杂度主要取决于相似用户和项目的数量，以及用户-项目评分矩阵的大小。

综上所述，算法 1 的时间复杂度为 $O(m^2)$ ，算法 2 的时间复杂度为 $O(pm\log(m))$ 。将两个部分相加，得到整个 SRDP 算法的时间复杂度为： $O(m^2) + O(pm\log(m))$ 。其中， m 为用户的数量， k 为相似用户的数量， p 为项目的数量。

3.2.3 SRDP 算法的隐私性分析

下面分析 SRDP 算法是否满足 ϵ -差分隐私保护：

在 SRDP 算法中，只有算法 2 涉及到用户-项目评分的隐私保护，而算法 1 并没有泄露隐私，所以无需为该操作分割隐私代价。算法 2 首先使用指数机制抽取中位数，然后对抽取的中位数应用拉普拉斯噪声进行干扰（以上两步严格按照

差分隐私实现机制进行添加)。显然,两次添加噪声的过程是串行的。 ϵ 分为 ϵ_1 和 ϵ_2 ,根据 2.4.1 节中差分隐私保护算法的组合性质中的性质 1,算法 2 满足差分隐私,即满足 ϵ -差分隐私。

3.3 实验结果与分析

在本节中,进行实验评估来验证 SRDP 算法的有效性。首先,将对实验数据集、对比算法、实验参数设置以及所采用的评估指标进行介绍,随后进行实验评估和结果分析。

3.3.1 实验数据集与对比算法

本实验使用 Film Trust 电影评分数据集作为实验数据集。数据集的详细统计数据如表 3-6 所列:

表 3-6 Film Trust 实验数据集统计分析

Table 3-6 Statistical information of Film Trust dataset

信息	特征	指标值
评分信息	用户数	1508
	项目数	2071
	评分数	35497
	稀疏度	1.14%
社交关系	信任者数	609
	被信任者数	732
	信任数	1853
	密度	0.42%

为了验证所提算法的有效性,首先将数据集中的评分数据随机分为训练集和测试集,占比分别为80%和20%。然后用SRDP方法模糊化后的用户-项目评分 $\hat{s}_{ui}$ 作为社会化矩阵分解算法的输入,并将社会化矩阵分解的结果作为预测的用户-项目评分 $\hat{s}_{ui}$,实验步骤如图3-4所示:



图 3-4 实验步骤

Fig.3-4 Experimental procedure

为了验证 SRDP 算法应用在社交推荐算法的有效性,分三组实验进行分析比较:

(一) SoRec 算法^[51]:经典的以矩阵分解技术为基础的社交推荐算法,不过该算法没有考虑到用户的隐私保护方面的问题。

SoRec_DP 算法:是将 SRDP 应用于 SoRec 的一种算法,通过实验将 SoRec_DP 算法与 SoRec 算法进行对比,比较其推荐效果,证明所提出的算法能保证推荐的准确性。

SoRec_DP_N 算法:是将 SoRec_DP 算法去掉差分隐私保护,通过实验将 SoRec_DP_N 与 SoRec_DP 算法进行对比,比较在差分隐私保护下对推荐效果的影响。

(二) SocialMF 算法^[53]:在 SoRec 算法的基础上改进的社交推荐算法,同样,算法中未考虑用户的隐私问题。

SocialMF_DP 算法:是将 SRDP 应用于 SocialMF 的算法,通过实验将 SocialMF_DP 算法与 SocialMF 算法进行对比,比较其推荐效果。

SocialMF_DP_N 算法:是将 SocialMF_DP 去掉差分隐私保护的算法,通过实验将 SocialMF_DP_N 与 SocialMF_DP 算法进行对比,比较在差分隐私保护下对推荐效果的影响。

(三) SoReg 算法^[54]:该算法也是经典的基于矩阵分解的社交推荐算法之一,其在设计时也并未将用户的隐私问题纳入考量。

SoReg_DP 算法:将 SRDP 应用于 SoReg 的算法,通过实验将 SoReg_DP 算法与 SoReg 算法进行对比,比较其推荐效果。

SoReg_DP_N 算法:将 SoReg_DP 去掉差分隐私保护的算法,通过实验将 SoReg_DP_N 与 SoReg_DP 算法进行对比,比较在差分隐私保护下对推荐效果的影响。

实验中的基本参数如表 3-7 所列：

表 3-7 实验参数设置

Table 3-7 Experimental parameter setting

变量名	描述	默认值
λ_U	用户正则化参数	10^{-3}
λ_I	项目正则化参数	10^{-3}
λ_Z	社交特征正则化参数	10^{-3}
γ	偏置项计算学习速率	10^{-5}
n	冷启动用户对项目的评分数量	5
d	特征矩阵维度	10^1
$steps$	迭代次数	10^2

为了降低模型的复杂度，沿用 Ma 等人在文献[50]中设定的参数值，即在本文的所有实验中都设置成 $\lambda_U = \lambda_I = \lambda_Z = 10^{-3}$ 。实验不考虑隐私预算的分配对实验结果的影响，对于总隐私预算 ε ，在所有算法中 $\varepsilon_1 = \frac{\varepsilon}{2}$ ， $\varepsilon_2 = \frac{\varepsilon}{2}$ 。实验也不考虑用户综合相似度的权重分配对实验结果的影响，用户综合相似度 $= \frac{1}{2} \times$ 用户关联度 $+ \frac{1}{2} \times$ 用户相似度。

为了评估将 SRDP 应用于三种社交推荐算法后的推荐效果和进行隐私保护后的推荐效果，采用平均绝对误差（MAE）和均方根误差（RMSE）进行评估。

平均绝对误差定义为：

$$MAE = \frac{1}{N} \sum_{u,i \in N} |r_{ui} - \hat{r}_{ui}| \quad (3-3)$$

均方根误差定义为：

$$RMSE = \sqrt{\frac{1}{N} \sum_{u,i \in N} |r_{ui} - \hat{r}_{ui}|^2} \quad (3-4)$$

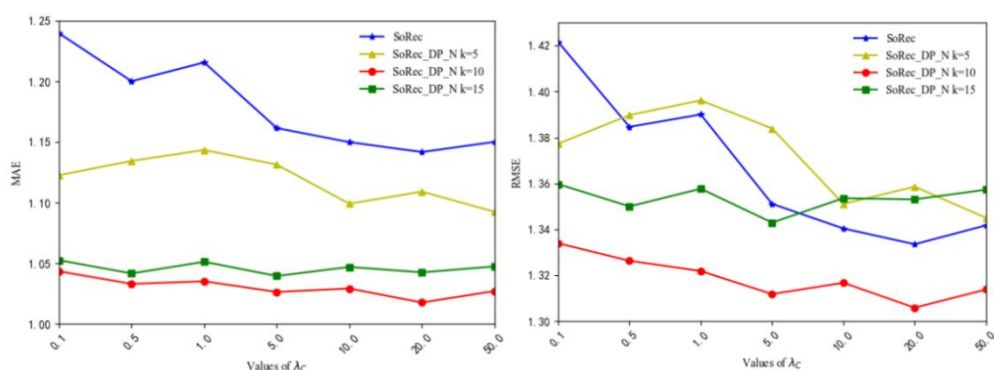
式中： N 为测试集中的评分条数； r_{ui} 是用户 u 对项目 i 的真实评分； $\hat{r}_{ui}$ 是算法为用户 u 对项目 i 的预测评分。MAE 和 RMSE 的值越小，表示真实评分与预测评分之间的误差越小，即推荐效果越好。

3.3.2 实验结果与分析

实验一：推荐效果对比

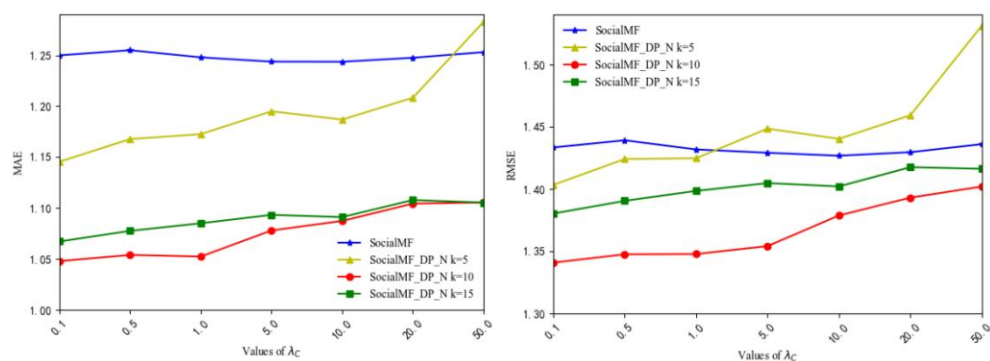
本次实验的目的是对本节所提出的 SRDP 算法的推荐效果进行考察。因此，将 SoRec 算法与 SoRec_DP_N 算法、SocialMF 算法与 SocialMF_DP_N 算法、SoReg 算法与 SoReg_DP_N 算法进行对比实验。算法中要给目标用户寻找 k 个与之最相似的用户， k 分别取 5、10、15，通过对比实验来寻找推荐准确度最优情况下的 k 值。

社交推荐算法与传统推荐算法相比的优点是它结合了用户之间的社交信息，有助于预测用户的评分。参数 λ_c 是用来平衡了来自用户-项目评分矩阵和用户社交关系的信息。 λ_c 在不同取值的情况下对 MAE 和 RMSE 的影响，如图 3-5 所示：



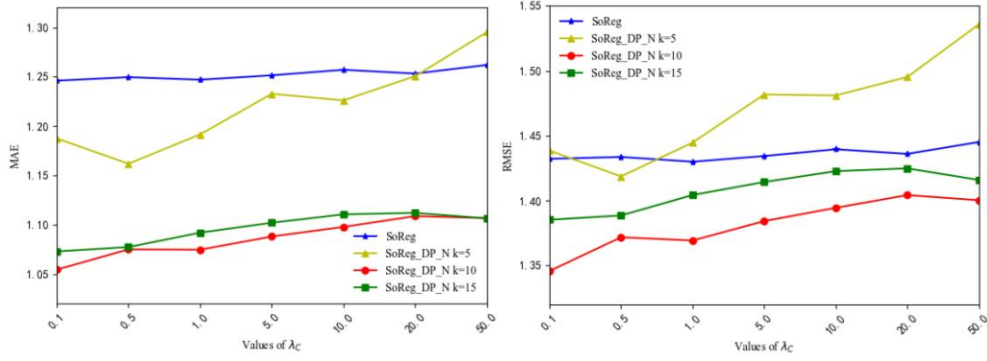
(a) 将 SRDP 算法应用于 SoRec 算法后的推荐效果

(a) The recommendation effect after SRDP algorithm is applied to SoRec algorithm



(b) 将 SRDP 算法应用于 SocialMF 算法后的推荐效果

(b) The recommendation effect of SRDP algorithm applied to SocialMF algorithm



(c) 将 SRDP 算法应用于 SoReg 算法后的推荐效果

(c) The recommendation effect after SRDP algorithm is applied to SoReg algorithm

图 3-5 推荐效果对比

Fig.3-5 Comparison of recommendation effect

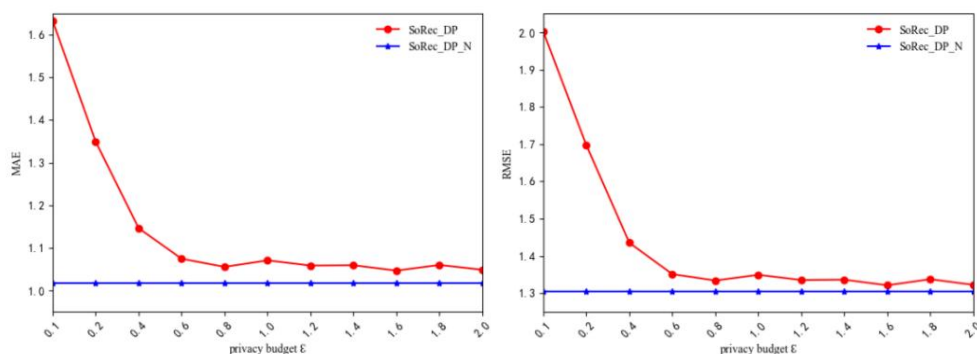
可以看出 λ_c 值对推荐结果有影响，这表明将用户-项目评分矩阵与用户社交信息融合在一起可以提高推荐准确性。从图 3-5 可以看出，SoRec_DP_N、SocialMF_DP_N、SoReg_DP_N 中的 k 取 10 和 15 时，MAE 值都分别比 SoRec、SocialMF、SoReg 的小，表明 SRDP 提供了有效的推荐结果。当 k 取 5 时，SoRec_DP_N、SocialMF_DP_N、SoReg_DP_N 的 MAE 值都比 k 取 10 和 15 时的 MAE 值小，说明相似用户个数 k 的大小影响了推荐效果；当 k 取 10 时，SoRec_DP_N、SocialMF_DP_N、SoReg_DP_N 算法明显有着更低的 MAE 值，这时推荐的准确性最高，说明相似用户个数 k 并非越大推荐效果就越好。同时从图 3-5 可以观察到，当 k 取 5 时，RMSE 值并非一直小于各自对比的社交推荐算法。当 k 取 15 时，SocialMF_DP_N、SoReg_DP_N 算法的 RMSE 值都分别比 SocialMF、SoReg 的小。但是当 k 取 10 时，SoRec_DP_N、SocialMF_DP_N、SoReg_DP_N 算法的 RMSE 值都分别比 SoRec、SocialMF、SoReg 的小，此时推荐效果最好。

总之，从图3-5可以看出当 λ_c 取20、 k 取10时，SoRec_DP_N算法推荐效果最好；当 λ_c 取0.1、 k 取10时，SocialMF_DP_N、SoReg_DP_N算法推荐效果最好。

实验二：隐私保护对推荐结果的影响

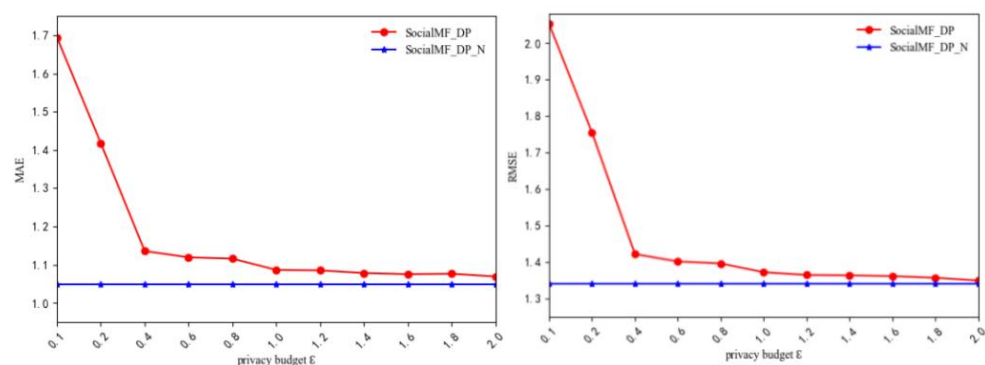
本组实验的目的是考察本文提出的 SRDP 算法在对原始数据进行差分隐私保护的同时对推荐效果的影响。因此，将 SoRec_DP 与 SoRec_DP_N、SocialMF_DP 与 SocialMF_DP_N、SoReg_DP 与 SoReg_DP_N 分别进行对比实验。本组实验根

据实验一的实验结果固定 λ_c 和 k 的取值。隐私预算是差分隐私保护算法中的重要参数，为评估不同隐私预算对推荐结果的影响，分别设置 $\epsilon=0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 2.0$,对 MAE 和 RMSE 的影响如图 3-6 所示：



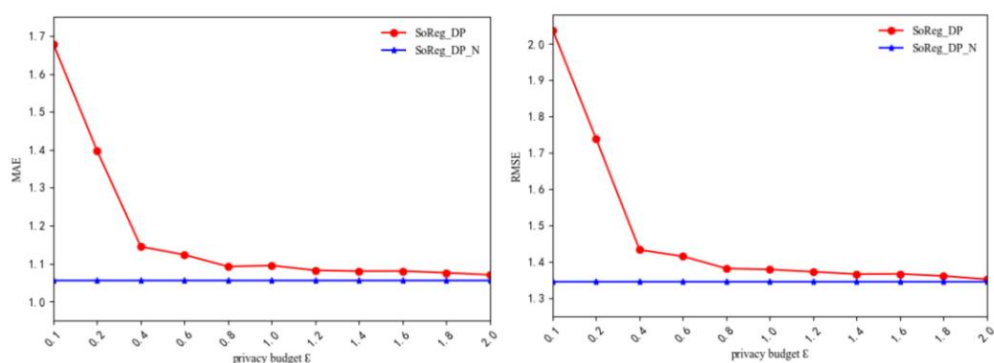
(a) SoRec_DP 算法的推荐效果

(a) Recommendation effect of SoRec_DP algorithm



(b) SocialMF_DP 算法的推荐效果

(b) Recommendation effect of SocialMF_DP algorithm



(c) SoReg_DP 算法的推荐效果

(c) The recommendation effect of SoReg_DP algorithm

图 3-6 隐私保护对推荐结果的影响

Fig.3-6 The impact of privacy protection on recommendation results

可以观察到,对于SoRec_DP、SocialMF_DP、SoReg_DP算法随着隐私预算 ϵ 从0.1变化到2.0, MAE和RMSE的值由大变小,其主要原因是隐私预算 ϵ 越大,所需的噪声量就越小,数据的可用性越高,那么推荐的准确性就会越高;而SoRec_DP_N、SocialMF_DP_N、SoReg_DP_N算法是未加差分隐私保护的,所以它的MAE和RMSE的值不随隐私预算 ϵ 的变化而变化。通过观察图3-6,我们可以发现当隐私预算小于1.0时,SRDP算法的MAE和RMSE值逐渐增大。尽管这意味着隐私保护能力增强,但是推荐准确性却明显降低。而当隐私预算大于等于1.0时,加差分隐私保护的算法和对应的去掉差分隐私保护的算法的MAE和RMSE值趋于接近。这表明在这种情况下,隐私保护力度和推荐准确性达到了一个比较合理的平衡点。换句话说,这时的算法既能实现对原始数据的差分隐私保护,又能接近未经隐私保护的算法的推荐准确性。此外,在实验一的实验中已经证实,在 $k=10$ 时, SoRec_DP_N、SocialMF_DP_N、SoReg_DP_N算法的推荐准确性分别优于SoRec_DP、SocialMF_DP、SoReg_DP算法。即证实:本文提出的SRDP算法不仅可以给用户提供的推荐结果,而且实现了对用户真实数据的隐私保护。

3.4 本章小结

社交推荐通过融合用户的各类社交信息,提升了推荐的准确性,使得社交推荐逐渐变成推荐领域的热门之一。不过,当下的社交推荐方式多数仅考虑了怎样提升推荐的精确性,却忽视了针对用于研究的原始数据展开隐私保护的问题。为处理此问题,本章提出了一种基于差分隐私保护技术的社交推荐方式。该方式通过采用差分隐私保护机制来处理数据,以达到保护用户隐私的目的;与此同时,通过找寻和目标用户最为相似的 k 个用户,为目标用户提供予令人满意的推荐结果。实验结果显示,本章所提出的算法能够在保护用户隐私的同时,为用户提供满意的推荐结果。

第4章 基于社区划分的社交推荐隐私保护方法

针对当下的社交推荐方式多数仅仅关注怎样增进推荐的准确性，却忽略了有关用户个人信息隐私保护的问题。本章提出了一种以社区划分为基础的社交推荐隐私保护方法（SRCD）。该方法将用户的个人数据隐藏在社区数据集中，并使用社区数据代替个人数据进行推荐。社区数据集可以在不泄露用户个人隐私的情况下公开披露，并且依旧可以为用户提供令人满意的推荐服务。

4.1 方法的提出

为了保护用户的个人信息，SRCD 方法考虑了用户之间的评分分值的分布范围给用户相似度带来的影响，且结合用户彼此间的社交关系，给社交网络里的用户划分社区，并算出每个社区中用户针对所看过项目的评分的均值，以这一评分均值替代用户的评分数据。接着依照社区划分的结果，找寻与目标用户所在社区最为相似的其余社区。如此，便能够构建出一个社区-项目评分均值矩阵。而且针对实际场景中评分均值矩阵稀疏的状况，采用中位数去填充矩阵的缺失元素。最终利用矩阵分解的结果来预测用户对项目的评分，以此评估算法的性能。研究框架如图 4-1 所示：

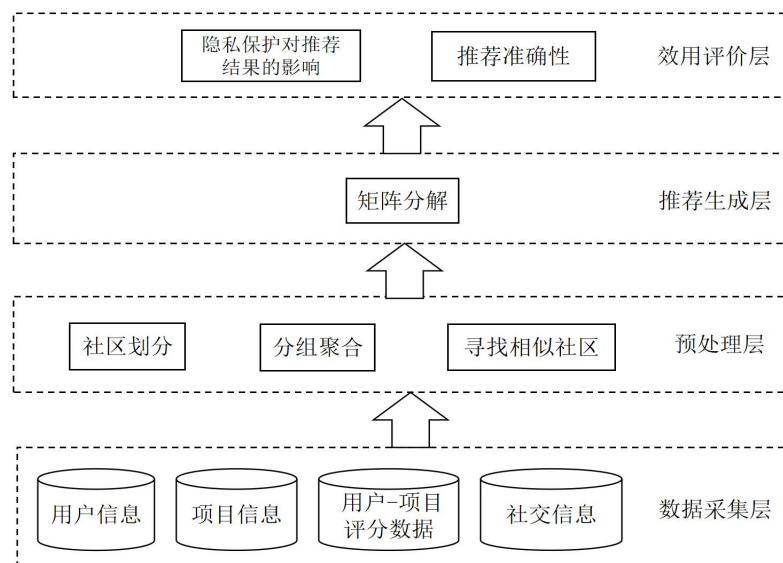


图 4-1 基于社区划分的社交推荐隐私保护方法研究框架

Fig.4-1 Research framework of social recommendation privacy protection methods based on community division

4.2 方法的描述

SRCD 算法主要包括 3 个步骤。其中，步骤 1：通过同时考虑社交关系和用户相似度来划分社交网络中的用户，不同社区间的用户可重叠；步骤 2：计算每个社区里所有用户看过的项目评分均值；步骤 3：寻找 n 个与目标用户所在社区最相似的其他社区。具体做法：首先对社交网络中的用户进行社区划分，以社区用户的项目平均评分代替用户个人对项目的评分。接着找到目标用户所在社区的 k 个相似社区，将这些社区的评分均值和看过的项目构建成社区-项目评分均值矩阵。最后利用该矩阵进行社交推荐。总体算法流程图如图 4-2 所示：

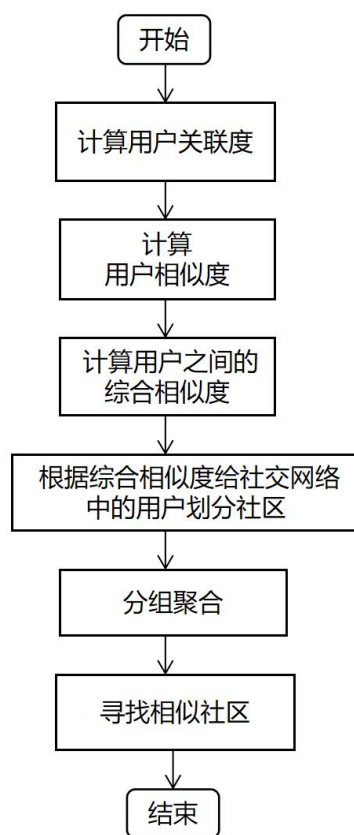


图 4-2 SRCD 算法流程图

Fig.4-2 SRCD algorithm flow chart

4.2.1 划分社区

步骤 1 通过综合考虑社交关系和用户相似度来计算用户间的综合相似度，并利用综合相似度将社交网络中的用户进行社区划分。本节设计如下算法，如表 4-1 所示：

表 4-1 划分社区算法

Table 4-1 Community division algorithm

Input: 用户列表 $users$ ，项目列表 $items$ ，用户-项目评分字典 $user_ratings$ ，社区里的用户数 k

Output: 社区列表 $communities$

```

1.初始化空列表 $communities$ 用于存放划分好的社区
2.初始化空字典 $user\_similarities$ 用于记录每对用户之间的综合相似度
3.for each  $user1$  in  $users$ :
4.    for each  $user2$  in  $users$ :
5.        if  $user1 \neq user2$ :
6.            利用式（4-1）计算关联度 $sim(user1, user2)$ 
7.            if  $user1, user2$ 没有相似项目:
8.                 $user\_similarity = sim(user1, user2)$ 
9.            else:
10.                利用式（4-2）计算用户相似度 $sim\_i(user1, user2)$ 
11.                 $user\_similarity = sim(user1, user2) + sim\_i(user1, user2)$ 
12.                 $similarities.append(user1, user2, user\_similarity)$ 
13. $similarity.sort(reverse = True, key = lambda x: x[2])$ 
14.for  $i = 1$  to  $k$  do
15.     $community = find\_user\_by\_index(similarity[i][2])$ 
16.     $communities.append(community)$ 
17.return  $communities$ 

```

为了保护用户的个人隐私，本章使用社区数据集合代替个人数据进行推荐。

算法 1 首先遍历每对用户节点：如果这两个用户没有观看过相同的项目，则仅考虑他们之间的关联度作为他们之间的综合相似度；如果这两个用户观看过相同的项目，则考虑他们之间的关联度和用户相似度作为他们之间的综合相似度。计算出每对用户之间的综合相似度后，找到与目标用户综合相似度最大的前 $(k-1)$ 个用户，并将目标用户与这 $(k-1)$ 个用户划分在同一个社区。

具体做法：利用本论文中 3.2.1 节的方法计算得到每两个用户间的综合相似度。通过比较目标用户与其他用户的综合相似度，来寻找 $(k-1)$ 个与目标用户最相似的用户，然后将这 k 个用户划分在一个社区。

例 4-4.假定要给用户 1 做推荐，且设定社区划分的参数 k 为 10，也就是每个社区包含 10 个用户。为保证处于同一社区内的用户彼此间的相似性达到最大，首先使用户 1 找到与其综合相似度最高的一个用户，假定是用户 5。接着，把用户 1 和用户 5 划分至一个社区。而后，继续寻找综合相似度与当前社区中全体用户之间相加数值最大的第二个用户。也就是，比较用户 1 与第二个用户的综合相似度加上用户 5 与第二个用户的综合相似度，进而选取相加数值最大的那个用户。依照一样的办法，逐步进行找寻。最终，将这 10 个用户划分进同一个社区。按照这样的方法，能够切实保证同一社区里的用户具有相似性。

时间复杂度：假设有用户节点有 m 个，首先计算用户之间的综合相似度，运行时间为 $O(m^2)$ ；然后给目标用户寻找 $(k-1)$ 个最相似的用户，时间复杂度为 $O(mk)$ 。 k 远远小于 m ，所以算法 1 的时间复杂度为 $O(m^2)$ 。

4.2.2 分组聚合

步骤 2 是计算每个社区里的项目评分均值。本节设计如下算法，如表 4-3 所示：

表 4-3 分组聚合算法

Table 4-3 Grouping aggregation attributes

Input: 社区 *communities*，用户-项目评分字典 *user_ratings*

Output: 社区对项目的评分均值字典 *communities_avgs*

1. 初始化空字典 *communities_avgs* 用于存放每个社区对项目的评分均值
2. for each *community* in *communities*:
3. 初始化空字典 *community_items* 用于存放该社区的用户看过的项目和评分
4. for each *user* in *community*:
5. for each *item, score* in *user_ratings[user]*:
6. *community_item* = *user_ratings[user]*
7. *community_items.append(community_item)*
8. for each *item, scores* in *community_items*:
9. 利用式 (4-3) 计算社区 *community* 对项目 *item* 的评分均值 *avg*
10. *community_avgs[item] = avg*
11. *communities.append(community_avgs)*
12. return *communities_avgs[item]*

算法 2 首先系统地遍历整个社区，找到该社区中所有用户观看过的项目以及对应的评分信息；然后将该社区中的用户看过的项目进行整理，得到该社区中所有用户看过的项目的并集；最后为该并集中的所有项目计算各自的评分平均分。根据式 (4-3) 计算每个社区的项目评分平均分。

$$avg_{ci} = \frac{sum_{ci}}{k} \quad (4-3)$$

其中： k 为社区 c 里的用户数； sum_{ci} 为社区 c 里的用户对项目 i 的评分总和。

例 4-5. 假设存在一个包含 100 个用户的社交网络，当采用算法 1 进行社区划

分时，设定社区里的用户数量 k 为 10，其划分结果如表 4-4 所示：

表 4-4 社区划分结果
Table 4-4 Community division outcome

社区		用户								
1	1	3	7	11	13	15	19	27	36	45
2	2	4	12	19	23	56	59	62	67	85
3	3	2	4	17	19	26	45	54	62	91
4	4	2	12	19	23	31	59	62	67	97
5	5	3	19	20	31	36	56	75	80	97
...										

其中，社区 1 里的用户给项目的评分情况如表 4-5 所示：

表 4-5 社区 1 里用户的项目评分数据
Table 4-5 Items ratings data of users in Community 1

用户/项目	3	7	15	19	25	31	...
1			1		5		
3	2	5	2	1	4	1	
7				3		3	
11	2			2	3	2	
13		2	1		3	3	
15	3			3			
19		3		2	5	2	
27	5		3	1			
36	2	5			4	2	
45		3	2	1			

则社区 1 对项目 3 的评分均值为 $(2 + 2 + 3 + 5 + 2)/5 = 2.8$ ，…。依次计算，那么社区 1 的项目评分均值如表 4-6 所示：

表 4-6 社区 1 的项目评分均值
Table 4-6 Average items ratings for Community 1

社区/项目	3	7	15	19	25	31	...
1	2.8	3.0	1.8	2.6	4.8	2.6	

按照以上的计算方法，计算每个社区中所用用户看过的项目评分均值。

时间复杂度：算法 2 需要遍历每个社区、每个用户以及每个用户的项目评分。因此，总的时间复杂度为 $O(mks)$ ，其中 m 是社区列表中的社区数量， k 是每个社区的用户数量， s 是平均每个用户评分的项目数量。

4.2.3 寻找相似社区

步骤 3 的目的是计算社区间的相似度，然后寻找与目标用户所在社区相似的其他社区。本节设计如下算法，如表 4-7 所示。

表 4-7 寻找相似社区算法
Table 4-7 Find similar communities algorithm

Input: 社区列表 <i>communities</i> ，相似社区个数 n
Output: 相似社区列表 <i>sim_communities</i>
1. 初始化空列表 <i>sim_communities</i> 用于存放相似社区
2. 初始化空字典 <i>community_similarities</i> 用于记录两个社区之间的相似度
3. for <i>community1</i> in <i>communities</i> :
4. for <i>community2</i> in <i>communities</i> :
5. if <i>community1</i> != <i>community2</i> :
6. 利用式 (4-4) 计算社区之间的相似度
7. <i>community_similarities.append((community1, community2, similarity))</i>
8. <i>community_similarities.sort(reverse = True, key = lambda x: x[2])</i>
9. for $i = 1$ to n do
10. <i>sim_community = find_community_by_index(communities_similarities[i][2])</i>
11. <i>sim_communities.append(sim_community)</i>
12. return <i>sim_communities</i>

算法 3 首先遍历每个社区，然后计算两两社区之间的相似度，两个社区里的相同用户数越多，那么这两个社区越相似。最后找 n 个与目标用户所在社区相似度最大的其他社区。

寻找相似社区的具体做法如下：首先计算社区之间的共同用户数，然后根据式（4-4）计算社区 x, y 之间的相似度：

$$simcommunity_{x,y} = \begin{cases} 0, node_x \cap node_y = 0 \\ \frac{|node_x \cap node_y|}{|2k|}, else \\ 1, node_x \cap node_y = k \end{cases} \quad (4-4)$$

其中： $node$ 代表社区里的用户节点。

例 4-6.社区划分结果如表 4-4 所示，可以看到社区 1 和社区 3 没有共同用户数，则 $simcommunity_{1,3} = 0$ ；社区 4 和社区 5 里共同用户是用户 19、用户 31 和用户 97，所以共同用户数是 3，则 $simcommunity_{3,4} = \frac{3}{20}$ ；而社区 2 和社区 4 里共同用户数等于 k （社区规模的大小），则 $simcommunity_{2,4} = 1$ 。

按照以上方法，首先算出社区间的相似度，接着通过比对目标用户所在的社区跟其他社区之间的相似度，来找寻和目标用户所在社区最为相似的前 n 个社区。这样便能确定与目标用户有着类似兴趣和特征的其他社区。

例 4-7.假设利用算法 2 找到社区 1 的四个最相似的社区，是社区 5、社区 17、社区 42、社区 56。将社区 1 和这四个相似社区及其项目评分，构造成社区-项目评分均值矩阵，如下表所示：

表 4-8 社区-项目评分均值矩阵

Table 4-8 Communities-items scoring matrix

社区/项目	3	7	15	19	25	31	...
1	$avg_{1,3}$		$avg_{1,15}$		$avg_{1,25}$		
5		$avg_{5,7}$		$avg_{5,19}$		$avg_{5,31}$	
17	$avg_{17,3}$		$avg_{17,15}$				
42				$avg_{42,19}$	$avg_{42,25}$	$avg_{42,31}$	
56	$avg_{56,3}$		$avg_{56,15}$		$avg_{56,25}$		

时间复杂度： m 个用户节点，共划分为 m 个社区，算法 3 首先计算社区之间的相似度，运行时间为 $O(m^2)$ ；然后寻找 n 个相似社区，时间复杂度为 $O(n^2)$ 。 m 远远大于 n ，所以算法 3 的时间复杂度为 $O(m^2)$ 。

算法 1 的时间复杂度为 $O(m^2)$ ，算法 2 的时间复杂度为 $O(mks)$ ，算法 3 的时间复杂度为 $O(m^2)$ 。由于用户的评分项目数 s 很大，远大于用户数 m ，所以 SRCD 算法的时间复杂度为 $O(mks)$ 。

4.3 实验结果与分析

4.3.1 实验数据集与对比算法

本实验使用公开的 Film Trust 数据集，并将评分数据随机划分成训练集和测试集，各自占比 80%和 20%。为了验证所提出算法的有效性，用本节提出的 SRCD 算法得到的社区-项目评分均值矩阵作为矩阵分解算法的输入，将矩阵分解的结果作为用户-项目的预测评分 $\hat{s}_{ui}$ ，实验步骤如图 4-3 所示：

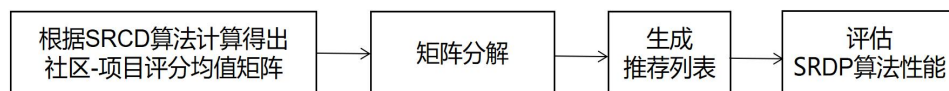


图 4-3 实验步骤

Fig.4-3 Experimental procedure

SRCD 算法是在 Pycharm 平台上实现的，并与三种典型的基于矩阵分解的社交推荐算法进行比较，这三种社交推荐算法的基本介绍如表 4-9 所示：

表 4-9 基于矩阵分解的社交推荐算法

Table 4-9 Social recommendation algorithm based on matrix decomposition

社交推荐算法		特点介绍
基于矩阵分解的社交推荐方法	RSTE ^[52]	将用户的预测评分构建为用户自身偏好和信任用户偏好之间的平衡。
	SocialMF ^[53]	将信任传播机制纳入模型。
	SoReg ^[54]	在矩阵分解目标函数中加入了社会正则化的考虑。

本实验中的参数设置如表 4-10 所列：

表 4-10 实验参数设置

Table 4-10 Experimental parameter setting

Variable name	Description	Default Values
λ_U	用户正则化参数	10^{-3}
λ_I	项目正则化参数	10^{-3}
λ_Z	社交特征正则化参数	0.5
γ	偏置项计算学习速率	10^{-4}
q	冷启动用户对项目的评分数量	5
d	特征矩阵维度	10^1
$steps$	迭代次数	10^2

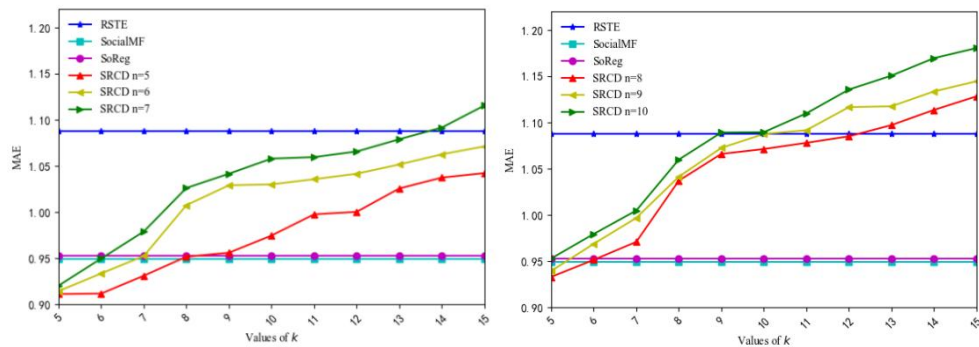
为了降低模型的复杂度，在本文的所有实验中都设置成 $\lambda_U = \lambda_I = 10^{-3}$ 。

为了评估 SRCD 算法的推荐和隐私保护的效果，这里采用与上章相同的两个评估指标：平均绝对误差（MAE）和均方根误差（RMSE）。

4.3.2 实验结果与分析

实验一：隐私保护对推荐结果的影响

本组实验的目的是考察本文提出的 SRCD 算法在对用户隐私进行保护的同时对推荐效果的影响。社区里的用户数 k ，相似社区个数 n 在不同取值的情况下对 MAE 和 RMSE 值的影响如图 4-4 所示：



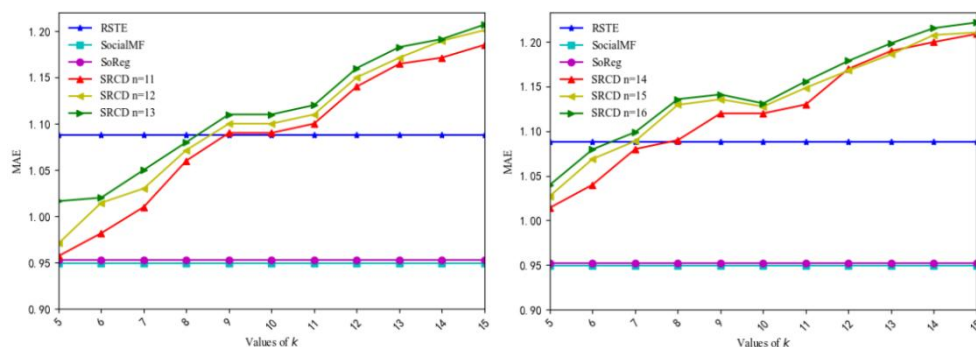
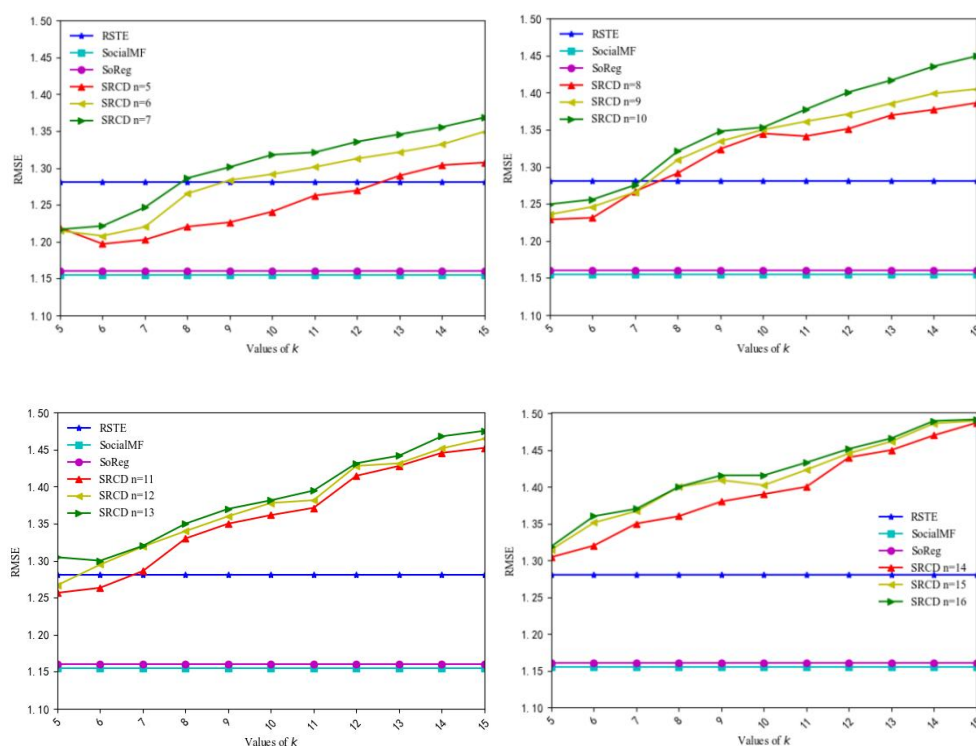
(a) 社区里的用户数 k 变化对 MAE 值的影响(a) The influence of the number of users k in the community on MAE value(b) 社区里的用户数 k 变化对 RMSE 值的影响(b) The influence of the number of users k in the community on the RMSE value

图 4-4 隐私保护对推荐结果的影响

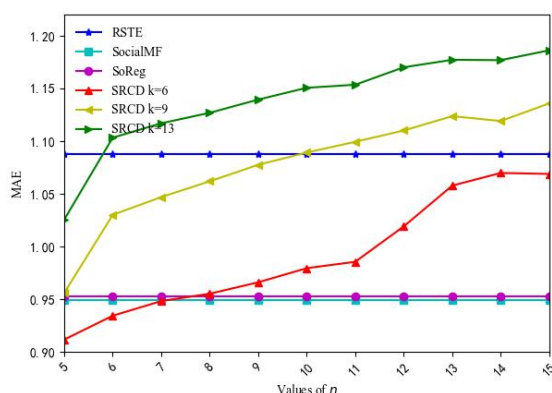
Fig.4-4 The impact of privacy protection on recommendation results

由图 4-4 可以看出, 对于 SRCD 算法而言, 随着社区中用户数量 k 从 5 增加到 15, 不论 n 取 5 至 16 之间的任何数值, MAE 和 RMSE 逐渐增大。这是由于随着社区用户数量 k 的增加, 用户的个人数据变得更加模糊。与此同时, 图表显示社区中用户数量 k 的变化不会改变 RSTE 算法、SocialMF 算法以及 SoReg 算法的 MAE 和 RMSE 值, 原因在于这些算法并未考虑隐私保护问题。

通过观察图 4-4 (a) 可以得知, 当 n 取值为 5 和 6 时, 社区中用户数 k 取 1 到 15 中的任意值, SRCD 算法的 MAE 值都低于 RSTE 算法的 MAE 值, 表示 SRCD 算法提供了有效的推荐。当 n 取值为 7、 k 小于 14 时, SRCD 算法的 MAE 值也都低于 RSTE 算法的 MAE 值。当 n 取值为 8、9 和 10 时, 分别在 k 小于 12、 k 小于 11 和 k 小于 10 时, SRCD 算法的 MAE 值低于 RSTE 算法的 MAE 值。当 n 取值为 11、12 和 13 时, 分别在 k 小于 10、 k 小于 9 和 k 小于 8 时, SRCD 算法的 MAE 值低于 RSTE 算法的 MAE 值。当 n 取值为 14、15 和 16 时, 分别只有在 k 小于 8、 k 小于 7 和 k 小于 7 时, SRCD 算法的 MAE 值才低于 RSTE 算法的 MAE 值。从图 4-4 (b) 可以看出, 当 n 取值为 5、6 和 7 时, 分别在 k 小于 13、 k 小于 9 和 k 小于 8 时, SRCD 算法的 RMSE 值低于 RSTE 算法的 RMSE 值。当 n 取值为 8、9 和 10 时, 分别在 k 小于 8、 k 小于 7 和 k 小于 7 时, SRCD 算法的 RMSE 值低于 RSTE 算法的 RMSE 值。当 n 取值为 11 和 12 时, 分别只在 k 小于 7 和 k 小于 6 时, SRCD 算法的 MAE 值才低于 RSTE 算法的 MAE 值。当 n 取 13、14、15 和 16, 此时隐私保护效果虽然很好, 却损失了推荐的准确性。

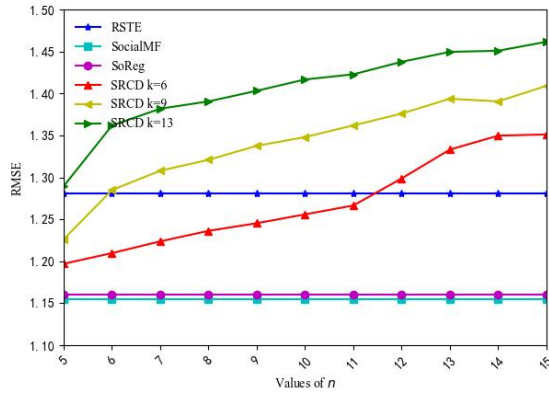
实验二：推荐效果对比

本组实验的目的在于考察本章所提算法的推荐准确性。本组实验根据本节实验一的实验结果选取 k 的值, 当其分别为 6、9 和 13 时, 探究在相似社区个数 n 在不同取值的情况下对 MAE 和 RMSE 值产生的影响, 如图 4-5 所示:



(a) 相似社区个数 n 变化对 MAE 值的影响

(a) Effect of the change of the number of similar communities n on MAE value



(b) 相似社区个数 n 变化对 RMSE 值的影响

(b) Effect of the change of the number of similar communities n on RMSE value

图 4-5 推荐效果对比

Fig.4-5 Comparison of recommendation effect

从图 4-5 的观察结果可知，就 SRCD 算法来说，不管 k 取 6、9 或是 13，伴随着相似社区个数 n 由 5 变动至 15，MAE 与 RMSE 的值都由小变大。这是因为利用社区数据集合替代用户个人数据进行推荐，会使推荐效果降低。与此同时，图表显示相似社区个数 k 的改变不会改变 RSTE 算法、SocialMF 算法以及 SoReg 算法的 MAE 和 RMSE 的值。

从图 4-5 (a) 可以看出，当 k 等于 6 时，无论相似社区个数 n 为 5 至 15 中的任意值，SRCD 算法的 MAE 值都小于 RSTE 算法的 MAE 值，显示出 SRCD 算法提供了有效的推荐。当 k 等于 6 且 n 小于 8 时，SRCD 算法的 MAE 值也小于 SocialMF 算法和 SoReg 算法的 MAE 值；然而，当 n 大于等于 8 时，SRCD 算法的 MAE 值却高于 SocialMF 算法和 SoReg 算法的 MAE 值。另外，在 k 取 9 时，当 n 小于 10 时，SRCD 算法的 MAE 值小于 RSTE 算法的 MAE 值；而在 k 取 13 时，当 n 小于 6 时，SRCD 算法的 MAE 值也小于 RSTE 算法的 MAE 值。在图 4-5 (b) 中可以观察到，当 k 为 6 且 n 小于 12 时，SRCD 算法的 RMSE 值小于 RSTE 算法的 RMSE 值；但当 n 大于等于 12 时，SRCD 算法的 RMSE 值则高于 RSTE 算法的 RMSE 值。此外，在 k 为 9 且 n 小于 6 时，SRCD 算法的 RMSE 值小于 RSTE 算法的 RMSE 值。然而，在 k 取值为 13 时，SRCD 算法的推荐效果较差。

综合这两组实验结果能够知道：当 n 取 5、 k 取 6 时，SRCD 算法的推荐效

果最为理想；当 n 取 5、 k 取 9 时，可以在推荐与隐私保护之间实现较好的平衡；当 n 取 5、 k 取 13 时，既可以较好地保护用户隐私，也能够获得较好的推荐结果。因此，依据本章算法的实验结果能够得出结论：SRCD 算法在给用户提供准确推荐的同时，还能够提供一定程度的隐私保护。

4.4 本章小结

当前的社交推荐方法侧重于提高推荐准确性，但忽略了用户个人信息的隐私问题。因此，本章提出了一种基于社区划分的、用于隐私保护的社交推荐方法。该方法首先综合考虑了社交图和偏好图，对社交网络中用户进行社区划分，然后以社区数据集合替代个人数据集合进行推荐，从而保护用户的个人隐私。随后，通过寻找与目标用户所在社区最相似的其他 n 个社区，为目标用户提供令人满意的推荐结果。实验结果显示，相比现有基于矩阵分解的社交推荐方法，本章提出的方法能够同时提供用户满意的推荐结果并实现一定程度的个人隐私保护。

第5章 总结与展望

本章是对整篇论文的研究工作及成果进行归纳总结，同时在已经完成的工作基础之上，对未来研究的方向进行展望。

5.1 工作总结

社交推荐需要利用大量的用户历史数据和社交信息来训练预测模型，然而用户提供的数据越详细，个人隐私泄露的风险就越高。如果社交推荐系统不能提供良好的隐私保障，可能导致用户对服务提供商的信任下降，从而不再分享相关数据，进而影响系统的准确性和推荐效果。为了保证社交推荐的准确性，需要采取多项措施来保护用户的隐私和数据安全，如差分隐私、匿名化、联邦学习等隐私保护技术手段。然而，过度的隐私保护会对推荐服务的质量产生影响。所以，社交推荐的隐私保护需要同时实现推荐结果的有效性与用户的隐私性。本研究聚焦于社交推荐中的隐私保护议题，针对评分预测过程中可能泄露用户隐私信息的问题，提出两种算法，旨在实现在保护用户隐私的同时仍能保持推荐的效果。工作总结如下：

（1）概述社交推荐隐私保护的研究背景与研究意义，并对国内外在社交推荐中运用的隐私保护方法的成果进行了综述。

（2）概述社交推荐基本框架和社交推荐模型，描述社交推荐面临的隐私保护，并阐述现有的社交推荐隐私保护技术以及社交推荐隐私保护模型。

（3）提出基于差分隐私的社交推荐隐私保护方法（SRDP）。该方法利用差分隐私保护技术来保证用户评分数据的隐私安全，同时通过寻找目标用户的相似用户来保证社交推荐的准确性，最后利用社会化矩阵分解的结果来预测评分，以评估算法性能。

（4）提出基于社区划分的社交推荐隐私保护方法（SRCD）。该方法引入社区划分算法，同时考虑社交关系的作用，把社交网络里的用户划分在不同的社区，将用户个人信息隐匿在社区数据集中，且通过查找和目标用户所属社区相似的社区来进行推荐，最后借助矩阵分解的结果预测评分，以测评算法性能。

5.2 未来展望

本文主要面向社交推荐领域的隐私保护研究，针对其不足设计两种改进算法，并在社交推荐隐私保护中取得相对较好的实验结果，虽然本文的改进算法取得一定的成果，但要想社交推荐隐私保护达到更好的效果，仍然需要进一步研究与改进。在今后的工作中，结合当前社交推荐中的隐私保护技术发展现状，比如利用联邦学习等技术以及本文已经完成的工作，将继续改进现有的社交推荐隐私保护模型，在保护用户隐私的前提下，提高推荐的准确性，预期在以下两个方面深入研究：

（1）本文在提出的基于差分隐私的社交推荐隐私保护方法中对于隐私预算的分配和用户综合相似度的权重分配只是简单的按 1: 1 分配，所以对于下一步的工作，需要进一步研究隐私预算的分配和用户综合相似度的权重分配对推荐结果的影响，以进一步提高推荐结果的准确性和隐私保护水平。

（2）本文主要以静态社交网络为基础展开研究，然而，随着时间的推移，用户的兴趣可能会发生改变。例如在某个时期热衷于动作类电影，而在另一时期则可能对喜剧类电影产生浓厚兴趣。因此，研究如何在这种动态变化的情况下，既能保护用户的隐私，又能提供准确的推荐服务，是极具探究价值的。

参考文献

- [1] 黄勃, 严非凡, 张昊, 等. 推荐系统研究进展与应用[J]. 武汉大学学报(理学版), 2021, 67(06): 503-516.
- [2] 耿化聪, 梁宏涛, 刘国柱. 基于知识图谱与协同过滤的饮食推荐算法[J]. 计算机与现代化, 2021, (08): 24-29.
- [3] Kota N, Padmanabhan V, Bhukya W. Content Based Network Representational Learning for Movie Recommendation[C]//Proceedings of 16th Multi-Disciplinary International Conference on Artificial Intelligence. Berlin: Springer, 2023: 112-123.
- [4] 冯晗, 伊华伟, 李晓会, 等. 推荐系统的隐私保护研究综述[J]. 计算机科学与探索, 2023, 17(08): 1814-1832.
- [5] 张岐山, 翁丽娟. 社会化推荐系统综述[J]. 计算机工程与应用, 2020, 56(01): 1-10.
- [6] Macwan K, Imine A, Rusinowitch M, et al. Privacy Preserving Recommendations for Social Networks[C]//Ninth International Conference on Social Networks Analysis. New York: IEEE, 2022: 1-8.
- [7] 彭慧丽, 张啸剑, 金凯忠. 基于差分隐私的社交推荐方法[J]. 计算机科学, 2017, 44(S1): 395-398+423.
- [8] 鲜征征, 李启良, 黄晓宇, 等. 融合显/隐式信任协同过滤算法的差分隐私保护[J]. 电子学报, 2018, 46(12): 3050-3059.
- [9] Zhou X L, Wei L B, Niu Y K, et al. DPSR: A differentially private social recommender system for mobile users[C]//2019 International Conference on Wireless Algorithms, Systems, and Applications. Berlin: Springer, 2019: 639-646.
- [10] Dou K, Guo B, Kuang L. A privacy-preserving multimedia recommendation in the context of social network based on weighted noise injection[J]. Multimedia Tools Applications, 2019, 78(19): 26907-26926.
- [11] Meng X Y, Wang S H, Shu K, et al. Towards privacy preserving social recommendation under personalized privacy settings[J]. World Wide Web, 2019, 22(6): 2853-2881.
- [12] Liu F K, Wu G R, Liu Y. A privacy-preserving friend recommendation mechanism for online social networks[C]//Proceeding of the 4th International Conference on Cryptography, Security and Privacy. New York: ACM, 2020: 63-67.
- [13] Uplavikar N, Vaidya J, Lin D, et al. Privacy-preserving friend recommendation in an integrated social environment[C]//Proceedings of 16th International Conference on Information Systems Security. Berlin: Springer, 2020: 117-136.
- [14] Li M, Zeng Y M, Guo Y, et al. The movie recommendation system based on differential privacy[C]//Proceedings of 6th International Symposium on Security and Privacy in Social Networks and Big Data. Berlin: Springer, 2020: 318-328.
- [15] 孙道柱, 李男, 杜启明, 等. 融合显隐式反馈协同过滤的差分隐私保护算法[J]. 计算机应用研究, 2021, 38(08): 2370-2375.
- [16] Huo Y F, Chen B L, Tang J, et al. Privacy-preserving point-of-interest recommendation based on geographical and social influence[J]. Information

- Sciences, 2021, 543(03): 202-218.
- [17] Zhou P, Wang K H, Guo L K, et al. A Privacy-preserving distributed contextual federated online learning framework with big data support in social recommender systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(3): 824-838.
- [18] Macwan K, Imine A, Rusinowitch M. Privacy preserving recommendations for social networks[C]//Proceedings of the 9th International Conference on Social Networks Analysis, Management and Security. New York: IEEE, 2022: 1-8.
- [19] Cai J P, Liu Y, Liu X M, et al. Privacy-preserving federated cross-Domain social recommendation[C]//International Workshop on Trustworthy Federated Learning. Berlin: Springer, 2022: 144-158.
- [20] 刘生昊, 吴国洋, 邓贤君, 等. 推荐系统与隐私保护研究综述[J]. 华中科技大学学报(自然科学版), 2023, 51(02): 1-9.
- [21] Checco A, Bracciale L, Leith D, et al. OpenNym: Privacy preserving recommending via pseudonymous group authentication[J]. IEEE Security and Privacy, 2022, 5(2): 1-23.
- [22] 鲜英, 于炯, 薛朋强. 基于改进匿名模型的上下文推荐系统研究[J]. 计算机工程, 2019, 44(03): 214-219.
- [23] 王海艳, 陆金祥. 面向群组推荐的个性化隐私保护方法[J]. 通信学报, 2019, 40(09): 106-115.
- [24] 彭丽寻, 刘丰恺. 基于个性化 k 匿名隐私保护的资源推荐算法[J]. 电脑与电信, 2020, 283(06): 66-73.
- [25] Yi J W, Wu F Z, Wu C H, et al. Efficient-FedRec: Efficient federated learning framework for privacy-preserving news recommendation[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 2814-2824.
- [26] Bassem C. On Cooperative Obfuscation for Privacy-Preserving Task Recommendation in Mobile CrowdSensing[C]//17th International Conference on Wireless and Mobile Computing, Networking and Communications. New York: IEEE, 2021: 90-95.
- [27] Li W Q, Chen H Y, Zhao R F, et al. A Federated Recommendation System Based on Local Differential Privacy Clustering[C]//2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation. New York: IEEE, 2021: 364-369.
- [28] Luo J W, Yi X, Han F L, et al. An efficient clustering-based privacy-preserving recommender system[C]//Network and System Security-16th International Conference. Berlin: Springer, 2022: 387-405.
- [29] Sahoo A, Raj S, Pradhan C, et al. Perturbation-based fuzzified k-mode clustering method for privacy preserving recommender system[J]. International Journal of Information Security and Privacy, 2022, 16(1): 1-20.
- [30] Perifanis V, Drosatos G, Stamatelatos G, et al. FedPOIRec: Privacy-preserving federated poi recommendation with social influence[J]. Information Sciences, 2023, 623: 767-790.

- [31]Cai J P, Liu Y, Liu X M, et al. Privacy-Preserving Federated Cross-Domain Social Recommendation[C]//Trustworthy Federated Learning-First International Workshop. Berlin: Springer, 2022: 144-158.
- [32]Zhou P, Wang K H, Guo L K, et al. A Privacy-Preserving Distributed Contextual Federated Online Learning Framework with Big Data Support in Social Recommender Systems[J]. IEEE Transaction on Knowledge and Data Engineering, 2021, 33(3): 824-838.
- [33]王刚, 蒋军, 王含茹. 社会化推荐研究综述[J]. 计算机科学, 2018, 45(S2): 37-42+62.
- [34]Chen J Y, Fan G, Gong Z G, et al. A Topic-Aware Graph-Based Neural Network for User Interest Summarization and Item Recommendation in Social Media[C]//Database Systems for Advanced Applications-28th International Conference. Berlin: Springer, 2023: 537-546.
- [35]Li N, Gao C, Jin D P, et al. Disentangled Modeling of Social Homophily and Influence for Social Recommendation[J]. IEEE Transaction on Knowledge and Data Engineering, 2023, 35(6): 5738-5751.
- [36]Rehman F, Sharif H, Anwar M, et al. Big Data Analytics for Cybersecurity in IoE Networks[C]//Cybersecurity Vigilance and Security Engineering of Internet of Everything-Technology, Communications and Computing. Berlin: Springer, 2024: 163-176.
- [37]Deng J Z, Ran X, Wang Y, et al. Probabilistic Matrix Factorization Recommendation Approach for Integrating Multiple Information Sources[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(10): 6220-6231.
- [38]Leniowski R, Leniowska L, Wronski M, et al. Intelligent candidate recommendation system based on experimental calculation of the similarity model [C]//IEEE International Conference on Fuzzy Systems. New York: IEEE, 2022: 1-7.
- [39]刘宇东, 孙豪, 蒋运承. 融合内容相似度与多特征计算的个性化微博推荐模型[J]. 计算机科学, 2020, 47(10): 97-101.
- [40]任莉君, 刘彦戎. 一种优化的相似度协同过滤推荐算法[J]. 信息技术, 2023, (07): 77-81+86.
- [41]Qiao P P, Zhang Z W, Li Z T, et al. TAG: Joint Triple-Hierarchical Attention and GCN for Review-Based Social Recommender System[J]. IEEE Transaction on Knowledge and Data Engineering, 2023, 35(10): 9904-9919.
- [42]Yuan K, Liu G N, Wu J J, et al. Semantic and Structural View Fusion Modeling for Social Recommendation[J]. IEEE Transaction on Knowledge and Data Engineering, 2023, 35(11): 11872-11884.
- [43]孟祥武, 刘树栋, 张玉洁, 等. 社会化推荐系统研究[J]. 软件学报, 2015, 26(06): 1356-1372.
- [44]Badiy M, Amounas F, Hajar M. A Novel Hybrid Approach for Improving the Accuracy of the Supervised Link Prediction Based on Graph Structure Features in Social Networks[C]//Business Intelligence - 7th International Conference. Berlin: Springer, 2022: 231-242.

- [45]李旭, 李景文, 姜建武, 等. 基于矩阵分解的上下文感知兴趣点推荐方法[J]. 计算机工程与设计, 2022, 43(11): 3108-3115.
- [46]贾俊杰, 刘鹏涛, 陈旺虎. 融合社交信息的矩阵分解改进推荐算法[J]. 计算机工程, 2021, 47(09): 97-105.
- [47]李翔, 陈硕, 杨健. 泛化界正则项: 理解权重衰减正则形式的统一视角[J]. 计算机学报, 2021, 44(10): 2122-2134.
- [48]刘华锋, 景丽萍, 于剑. 融合社交信息的矩阵分解推荐方法研究综述[J]. 软件学报, 2018, 29(02): 340-362.
- [49]Benabbes K, Housni K, Mezouary A, et al. Recommendation System Issues, Approaches and Challenges Based on User Reviews[J]. Journal of Web Engineering, 2022, 21(4): 1017-1054.
- [50]Fu Y, Wen P H, Wu J L. Social Network Analysis-Based Evaluation on Carbon Neutrality Acceptance of Weibo Users[C]//HCI International 2023-Late Breaking Papers-25th International Conference on Human-Computer Interaction. Berlin: Springer, 2023: 254-272.
- [51]Ma H, Yang H X, Lyu M R, et al. SoRec: Social recommendation using probabilistic matrix factorization[C]//Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM, 2008: 931-940.
- [52]Ma H, King I, Lyu M R, et al. Learning to recommend with social trust ensemble[C]//Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2009: 03-210.
- [53]Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks[C]//Proceedings of the fourth ACM conference on Recommender systems. New York: ACM, 2010: 135-142.
- [54]Ma H, Zhou D Y, Liu C, et al. Recommender systems with social regularization[C]//Proceedings of the Forth International Conference on Web Search and Web Data Mining. New York: ACM, 2011: 287-296.
- [55]Guo G B, Zhang J, Yorke-Smith N. TrustSVD: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2015: 123-129.
- [56]程章桃, 钟婷, 张晟铭, 等. 基于图学习的推荐系统研究综述[J]. 计算机科学, 2022, 49(09): 1-13.
- [57]李青, 王一晨, 杜承烈. 图表示学习方法研究综述[J]. 计算机应用研究, 2023, 40(06): 1601-1613.
- [58]Liu H, Zheng C, Li D T C, et al. Multi-perspective social recommendation method with graph representation learning[J]. Neurocomputing, 2022, 468: 469-481.
- [59]Liu X, Meng S M, Li Q M, et al. SMEF: Social-aware Multi-dimensional Edge Features-based Graph Representation Learning for Recommendation[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. New York: ACM, 2023: 1566-1575.
- [60]马铁民, 周福才, 王爽. 基于用户相似度的随机游走社交网络事件推荐算法

- [J]. 东北大学学报(自然科学版), 2019, 40(11): 1533-1538.
- [61] Yang Q, Wang H Y, Bian M Y, et al. Incorporating reverse search for friend recommendation with random walk[J]. International Arab Journal Of Information Technology, 2020, 17(3): 291-298.
- [62] Guo L M, Luan K X, Sun L, et al. Collaborative filtering recommendations based on multi-factor random walks[J]. Engineering Applications of Artificial Intelligence, 2023, 123(B): 106409.
- [63] 肖国庆, 李雪琪, 陈玥丹, 等. 大规模图神经网络研究综述[J]. 计算机学报, 2024, 47(01): 148-171.
- [64] Tallapally D, Wang J, Potika K, et al. Using Graph Neural Networks for Social Recommendations[J]. Algorithms, 2023, 16(11): 515.
- [65] Wei C Y, Fan Y S, Zhang J. Time-Aware Service Recommendation With Social-Powered Graph Hierarchical Attention Network[J]. IEEE Transactions on Services Computing, 2023, 16(3): 2229-2240.
- [66] 赵衍衍, 张良富, 张静, 等. 因子分解机模型研究综述[J]. 软件学报, 2019, 30(03): 799-821.
- [67] Ding Y, Wang D, Xin X. SCFM: Social and crowdsourcing factorization machines for recommendation[J]. Applied Soft Computing, 2018, 66: 548-556.
- [68] Shu J G, Jia X H, Yang K, et al. Privacy-Preserving Task Recommendation Services for Crowdsourcing[J]. IEEE Transactions on Services Computing, 2021, 14(1): 235-247.
- [69] 赵禹齐, 杨敏. 差分隐私研究进展综述[J]. 计算机科学, 2023, 50(04): 265-276.
- [70] Croft W, Sack J R, Shi W. Differential privacy via a truncated and normalized laplace mechanism[J]. Journal of Computer Science and Technology, 2022, 37(2): 369-388.
- [71] Ding Z Y, Wang Y X, Xiao Y T, et al. Free gap estimates from the exponential mechanism, sparse vector, noisy max and related algorithms[J]. The VLDB Journal, 2023, 32(1): 23-48.
- [72] Milani M, Huang Y, Chiang F, et al. Data Anonymization With Diversity Constraints[J]. IEEE Transaction on Knowledge and Data Engineering, 2023, 35(4): 3603-3618.
- [73] Ling C, Zhang W Z, He H. K-anonymity privacy-preserving algorithm for IoT applications in virtualization and edge computing[J]. Cluster Computing, 2023, 26(2): 1495-1510.
- [74] 邹劲松, 李芳. 基于可伸缩 L-多样性的大数据发布隐私保护[J]. 计算机应用研究, 2021, 38(02): 564-566+571.
- [75] Wang M Z, Jiang Z R, Zhang Y, et al. T-Closeness Slicing: A New Privacy-Preserving Approach for Transactional Data Publishing[J]. INFORMS Journal on Computing, 2018, 30(3): 438-453.
- [76] 张洪磊, 李浥东, 邬俊, 等. 基于隐私保护的联邦推荐算法综述[J]. 自动化学报, 2022, 48(09): 2142-2163.
- [77] 肖雄, 唐卓, 肖斌, 等. 联邦学习的隐私保护与安全防御研究综述[J]. 计算机学报, 2023, 46(05): 1019-1044.

- [78]吴文泰, 吴应良, 林伟伟, 等. 横向联邦学习: 研究现状、系统应用与挑战[J/OL].计算机学报, 1-33[2024-01-09].
- [79]陈晋音, 李荣昌, 黄国瀚, 等. 纵向联邦学习方法及其隐私和安全综述[J]. 网络与信息安全学报, 2023, 9(02): 1-20.
- [80]李响, 付春霖, 雷亚国, 等. 保证数据隐私的装备协同智能故障诊断联邦迁移学习方法[J]. 机械工程学报, 2023, 59(06): 1-9.
- [81]Cioaba S, Gupta H, Ihringer F, et al. The least Euclidean distortion constant of a distance-regular graph[J]. Discrete Applied Mathematics, 2023, 325: 212-225.
- [82]Mohana H, Suriakala M. Integrated cosine and tuned co-sine similarity measure to alleviate data sparsity issues for personalized recommendation[C]//Proceedings of the 2018 3rd International Conference on Computational System & Information Technology for Sustainable Solutions. New York: IEEE, 2018: 41-49.

攻读学位期间发表的学术论文目录

- [1] 汪玉洁, 刘涛, 包象琳, 等. 社交推荐中的差分隐私保护方法[J]. 天津理工大学学报. (第一作者, 已录用)
- [2] 汪玉洁, 刘涛, 包象琳, 等. 基于社区划分的社交推荐隐私保护方法[J/OL]. 重庆工商大学学报(自然科学版), 1-10[2024-05-20]. (第一作者)
- [3] Yujie Wang, Zhi Deng, Kai Tao, et al. Intelligent Urban Street Lighting System based on IoT Cloud Platform[C]//Proceedings of 2022 9th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2022: 192-197. (第一作者, EI 收录)
- [4] 王振昕, 刘涛, 汪玉洁, 等. k 度匿名社交网络隐私保护方法[J]. 天津理工大学学报, 2024, 40(01): 92-100. (第三作者)
- [5] Zhenxin Wang, Tao Liu, Yujie Wang, et al. Graph-Clustering Anonymity Privacy Protection Algorithm With Fused Distance-Attributes[C]//Journal of Physics: Conference Series. IOP Publishing, 2023, 2504(1): 012058. (第三作者, EI 收录)

攻读学位期间参与的科研项目

- [1] 基于符号方法的区块链系统形式化模型构建与验证方法研究 (2108085QF264), 安徽省自然科学基金.
- [2] 群智感知用户激励机制与隐私保护研究 (xjky2020120), 校级科研项目.

攻读学位期间参与的学科竞赛项目

- [1] 2021 年安徽省大学生网络与分布式系统创新设计大赛, 作品赛, 一等奖.
- [2] 2021 年安徽省高校物联网应用创新大赛, 创意赛, 二等奖.

致 谢

相逢总有千言，离别仅需二字。现在终于到了我们要说再见的时候了。又是一年毕业季，虽然年纪已经增长许多，踏入社会我却仍然是个新人，今后要学习的地方依然有很多，愿一直有从头再来的勇气和一往无前的干劲。这三年匆匆而过，疫情从我的本科贯穿到了研究生阶段，但是读研期间让我体验到不同的大学生活，青春里的回忆果真永远是最美好的。求学二十载，走过的路，读过的书，昨日种种，皆成今我。

感谢计算机与信息学院所有老师给予我们的关心和指导。感谢导师刘涛教授对我学习和思想上的指导，每次汇报时，都会耐心地帮我梳理研究思路，明确研究方向，让时常感到焦虑的我能够找到当下应该去做的事情，并时刻督促我的研究进度，帮我把控好大局。拨穗正冠，感恩栽培。

感谢父母，读书二十载，辛苦养育我长大成人，而现在我终于可以步入社会，他们终于不用操心我的学习成绩了，却又要开始操心我的工作了，他们无时无刻不在牵挂着我。无以为报，唯有努力不辜负他们的期望。

感谢一路遇见的人，我的同学们和实验室的师兄师姐们都很友好，热心帮助我解决困难。从考研读研到现在，感谢李同学一直倾听我的吐槽，陪我度过一个个难关，不胜其烦地安慰我，好景难遇，如你一般。

生活也许普普通通，也许困难几重。愿我们从容面对，乐在其中。毕业之后，我们将奔赴各自的人生，希望大家都会有好的归属。我亲爱的朋友们，花自向阳开，人终往前走。如果不是终点，我们要好好努力，但最重要的是要平安和开心。跃入人海，祝各位工作顺利，生活愉快，期待和大家的再次相逢。