移动商务推荐系统中的一种基于P2P的隐私保护策略

王利娥 许元馨 李先贤 刘 鹏
(广西师范大学广西多源信息挖掘与安全重点实验室 桂林 541004)
(广西师范大学计算机科学与信息工程学院 桂林 541004)

摘 要 近年来,移动推荐系统已成为推荐系统研究领域最活跃的课题之一。但由于移动终端的私人性和移动网络的复杂性,在保证高精度推荐的同时如何保护用户隐私已经成为移动商务发展的主要挑战。传统推荐系统中的隐私保护技术由于移动终端的计算能力差、无线网络的带宽弱等局限无法适用于移动商务推荐系统。针对以上问题,面向移动商务推荐提出一种基于P2P的隐私保护策略,通过构建P2P好友圈,采用基于k-匿名的代理转发的增量数据更新方式,实现不对增量数据进行任何修改以保证高精度推荐,同时保护用户隐私安全。最后通过实验验证了基于P2P的隐私保护策略的可行性和推荐服务的有效性。

关键词 P2P,移动商务推荐系统,隐私保护,k-匿名
中图法分类号 TP309.2 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.09.034

P2P-based Privacy Protection Strategy in Mobile-commerce Recommender System

WANG Li-e XU Yuan-xin LI Xian-xian LIU Peng
(Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin 541004, China)
(College of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China)

Abstract Mobile recommender systems have recently become one of the hottest topics in the domain of recommender systems. How to provide high-precision recommendations and privacy protection has become the main challenge in the development of mobile-commerce, since mobile device is privacy and mobile network is complex. Due to its weakness of computation power and bandwidth, recommender system of mobile-commerce is not able to use these privacy-preserving techniques which are initially designed for traditional recommender systems. To address above problems, a P2P-based privacy protection approach specifically for mobile-commerce recommender system was proposed in this paper. Our approach keeps incremental data intactly for guaranteeing high-precision recommendations while preserving privacy by constructing friends' circles and forwarding data anonymously based on the model of k-anonymity. In the end, the experiment shows that the P2P-based privacy protection approach is feasible and effective.

Keywords P2P, Mobile-commerce recommender system, Privacy protection, K-anonymity

1 引言

电子商务方兴未艾,随着智能手机和以iPad为标志的平板电脑等智能移动设备的广泛应用,以及移动通信网络技术的飞速发展,移动电子商务已成为当前流行的一种商务模式,改变着人们的工作和生活方式。移动电子商务(M-Commerce)[1]是指手机、PDA及掌上电脑等无线终端与无线上网技术相结合而进行的B2B、B2C或C2C的电子商务体系。狭义上讲,是指以手机为终端,通过移动通信网络连接互联网所进行的电子商务活动。自从进入3G时代,智能手机发展迅猛,移动电子商务可以真正使任何人在任何时间、任何地点进行网购,实现随时随地、线上线下的购物与交易,比如在等车、睡前等碎片时间瞬间购买商品、参加团购等,彻底摆脱时间和地域的限制,实现有弹性的购物。然而,随着移动互联网信息内容的日益增长,加之移动设备的显示屏幕小、内存处理和输入/输出等能力有限,这会给移动用户带来沉重的"移动信息过载"问题,从而导致移动网络资源利用率和用户体验受到严重影响。个性化推荐系统(Personalized Recommender Systems)[1-4]可以有效缓解此难题,它通过分析用户的兴趣,提供个性化的推荐建议来处理信息的过载问题。但在某种程度上,个性化推荐系统是一把双刃剑:它可以在海量信息空间中根据用户的兴趣偏好为用户过滤信息,推荐其所需要的商品

到稿日期:2016-12-01 返修日期:2017-02-19 本文受国家自然科学基金(61662008,61672176,61502111),广西区域多源信息集成与智能处理协同创新中心,广西自然科学基金(2015GXNSFBA139246),"八桂学者"工程专项经费资助。
王利娥(1981-),女,硕士,副教授,CCF会员,主要研究领域为对等网络和信息安全;许元馨(1991-),女,主要研究领域为数据安全;李先贤(1969-),教授,博士生导师,CCF会员,主要研究领域为网络安全和隐私保护;刘 鹏(1979-),男,博士生,副教授,CCF会员,主要研究领域为网络安全和数据分析,E-mail:liupeng@gxnu.edu.cn(通信作者)。

信息，以解决信息超载问题；但同时也需要收集用户的个人信息、兴趣偏好等，存在隐私泄露的安全隐患。传统的个性化推荐系统通常是定期对数据进行分析，然后对模型进行更新，进而利用新的模型进行个性化推荐，由于其是离线计算，通常可以采用数据聚合、伪装和加密等技术来保护数据隐私。

移动电子商务不同于传统的电子商务模式，主要区别在于其服务对象的移动性、服务要求的即时性、服务终端的私人性和服务方式的方便性。服务终端的私人性主要体现在手机用户的唯一性。移动互联网的安全环境比传统互联网复杂，威胁来源和易被攻击的范围更加广泛，包含大量个人信息和机密信息的移动数据更容易引起黑客关注，因此移动电子商务的隐私安全显得尤为重要，甚至成为移动电子商务发展的瓶颈。同时由于移动终端的内存空间和计算能力有限、显示屏幕小、输入输出能力差、依靠电池工作等特性，使得移动商务在个性化、实时性方面对推荐系统提出了更高的要求，使得移动商务推荐系统[5]中的隐私保护更具挑战性，主要体现在以下几个方面：

(1)安全性要求更高。由于移动终端的私人性、便携性及移动数据网络的复杂性，移动商务推荐系统对安全性提出了更高的要求，保护个人隐私和敏感的移动数据显得尤为重要。

(2)实时性要求更高。移动终端的移动性强，使其移动信息需求和对推荐的需要受上下文影响更大，因此移动推荐对实时性的要求更高。比如某南方用户在东北出差，希望购买保暖的羽绒服，而推荐系统推荐的依然是适合南方穿的薄衣服，这就未能很好地符合用户的需求。如果推荐系统能根据位置信息做出推荐反馈，推荐周围的人都在购买的羽绒服则可收到较好的效果，从而能够为用户提供更准确的推荐。移动商务推荐的实时性要求，移动商务推荐不能采取离线计算方式。

(3)推荐精度要求高。移动终端的屏幕小、输入输出能力差和无线网络的带宽弱等因素，使其需要更高的推荐精确度，以确保用户在有限的设备终端界面接收最需要和最感兴趣的信息，否则将会大大降低用户体验，从而丧失推荐的必要性。而传统的聚合或添加噪音等隐私保护技术修改了用户数据，大大降低了推荐质量，因此移动商务推荐应尽量保证用户数据的准确性以达到高精度推荐。

(4)处理能力差。移动终端的内存空间和计算能力有限、依靠电池工作等特性，使得隐私保护算法本身不能太复杂，而采用传统的加密等隐私保护技术要求较强的计算能力，可能会导致移动终端性能降级，甚至电池耗尽、可用性丧失。

本文针对移动商务推荐系统的隐私保护中存在的困难，提出一种面向移动电子商务实时推荐系统的基于 P2P 的隐私保护策略。该策略首先结合移动商务的特性，构建基于 P2P 网络模型的好友圈，不对增量数据进行任何修改，而采用 k-匿名的代理转发方式获得推荐的隐私保护机制。本文将 P2P 网络中的匿名技术引入推荐系统中，研究在移动商务推荐中如何构建基于 P2P 的隐私保护模型，使其在保证个性化推荐系统的精确性、实时性、轻量级计算的前提下，保证用户的隐私安全。

2　相关研究基础

2.1　个性化推荐系统

随着电子商务技术的不断发展，个性化推荐系统已经逐渐成为网络营销的一种策略和手段。它利用电子商务网站向客户提供商品信息和建议，帮助用户决定应该购买何种产品，模拟销售人员帮助客户完成购买过程。因此个性化推荐系统首先要采集用户的兴趣、需求等个性化信息，以分析用户的个性化偏好，从而做出有效的、准确的个性化推荐。目前推荐系统的主流推荐算法主要有以下几种[1-5]：1)基于项目的推荐。根据用户喜欢或已购买的项目，按照项目关联规则选择相类似或相关的项目向用户进行推荐。2)协同过滤推荐。利用与目标用户兴趣相似的其他用户的偏好来预测目标用户的偏好，即根据目标用户与其他用户购买项目的相似性，推测目标用户可能也喜欢其他相似用户所购买的其他项目，从而进行推荐。3)混合推荐。由于各种推荐方法都存在优缺点[3]，实际使用时可以按照不同策略将不同的推荐技术进行组合并完成推荐[5]。在移动推荐系统中，研究和应用最多的是将协同过滤推荐与基于项目的推荐相结合的方法[6-7]。

2.2　半分布式 P2P 网络模型

本文结合移动终端和 P2P 网络中节点存在的共性，即高度自治性、匿名性、动态性以及节点之间的地位平等的特性，设计了一种基于 P2P 网络的隐私保护模型，以适应移动终端的移动性和匿名性要求。本节对 P2P 网络模型进行了简单介绍。

对等网络(Peer-to-Peer，P2P)[8]中的节点处于完全对等的地位，不区分客户机和服务器，网络中各个节点之间可以直接进行数据通信。P2P 网络中的节点既可以获取其他节点的资源或服务，也可以为其他节点提供资源或服务，不依赖于集中服务器。本文采用半分布式拓扑结构 P2P 网络，它由提供查询服务的超级节点和其他普通节点组成。超级节点的功能要远远强于普通节点，可以为其他对等节点提供定位、资源检索服务等，所有普通节点都向它发送请求，如图 1 所示。

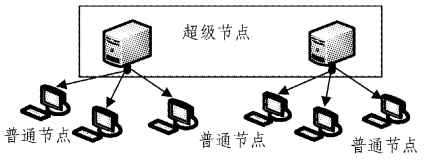

图 1　基于 P2P 的网络模型

3　现有的隐私保护技术

近年来，面对由互联网的不断发展而引发的全球市场竞争，个性化推荐系统已经成为企业能否生存的关键。随着信息泄露事件的频频出现，用户信息安全意识提高，个性化推荐系统中的隐私保护研究逐渐受到大家的重视，不同的隐私保护技术不断被提出。

(1)基于体系结构的隐私保护技术。早期的个性化推荐系统都是基于集中服务器的，因此用户信息的收集、建模以及用户描述文件都存放在服务器上，这对用户的隐私构成了极大的威胁，因为用户无法控制自己的个人数据。文献[9-11]

提出了基于分布式体系结构的隐私保护技术，其重点是将用户的个人信息数据存放在用户个人客户端，使用户可以完全操控自己的数据，用户可以自行决定哪些数据是需要被保护的。

(2)基于匿名模型的隐私保护技术。该技术主要是通过一定的隐私策略对原始数据进行修改，使挖掘方无法从最终发布的数据中提取出原始数据信息或用户与隐私信息的关联，以达到隐私保护的目的。文献[12-19]采用各种匿名技术对原始数据进行修改，使用修改后的值来代替原始值，以达到隐私保护的目的。

(3)基于 P2P 的隐私保护技术。文献[9,15-21]提出了基于 P2P 的推荐系统。在 P2P 的推荐系统中，每个用户的计算机要同时担任客户端和服务器的角色，用户的个人数据存放在用户的计算机中，用户可以完全操控个人数据。

(4)基于密码学的隐私保护技术。基于密码学的隐私保护技术[17-18,21]主要是通过密码机制实现对原始数据的不可见性，从而达到隐私保护和精确推荐的目的。

可以看出，每种隐私保护技术针对不同的应用需求都有各自的特点。我们对现有各种隐私保护技术进行对比分析，比较结果如表 1 所列。从表 1 可以看出，它们的适用范围、性能表现等不尽相同。由于前 3 种隐私保护技术都是基于离线计算模型的，因此不满足实时推荐的要求；而基于 P2P 机制的隐私保护技术是以在线计算的方式进行实时推荐，但由于终端的限制只能执行轻量级计算，如果用户有具体的隐私保护需求时可采用基于体系结构的隐私保护技术。当对推荐质量要求不高时，基于匿名模型的隐私保护技术能以一定的计算开销实现较好的隐私保护。当关注交易数据的完美隐私保护时，使用基于密码学的隐私保护技术更合适，但其计算量代价较高。而对于移动推荐系统，基于 P2P 机制的隐私保护技术在实时性方面具有一定的优势，但是现有的基于 P2P 的隐私保护技术往往存在数据不全或经过修改后推荐质量不高等问题。

表 1 推荐系统中各隐私保护技术的性能比较

	隐私保护度	计算开销	推荐精度	实时性
基于体系结构的隐私保护技术	中	低	中	低
基于匿名模型的隐私保护技术	中高	中	低	低
基于密码学的隐私保护技术	高	高	高	低
基于 P2P 机制的隐私保护技术	中高	低	中	高

4 基于 P2P 的隐私保护策略

针对上述存在的问题，本文面向移动商务推荐系统提出了一种基于 P2P 的匿名转发隐私保护机制，结合移动电子商务的特性，将 P2P 网络模型中的隐私保护技术与 k-匿名思想结合起来，设计了一种混合的隐私保护策略，如图 2 所示。该策略通过构建自适应的 P2P 好友圈，并采用代理匿名转发的增量数据更新方式，不对数据进行修改并保证用户数据包在 k 个数据包中不可区分，即在保证高精度、实时推荐的同时保护用户隐私安全。该机制主要包括两个模块：P2P 好友圈的构建模块和 k-匿名的代理转发数据更新模块。

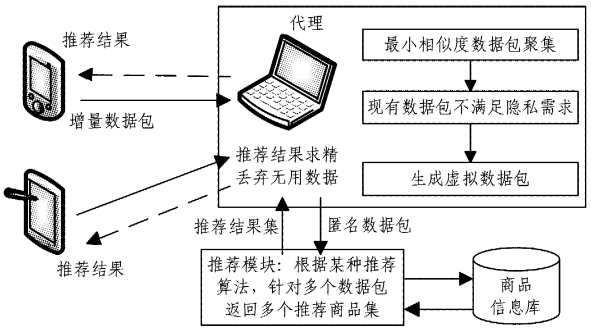

图 2 基于 P2P 的匿名转发隐私保护策略

4.1 P2P 好友圈的构建

P2P 是一种完全分布式的网络模型，每个节点既是客户端又是服务器，每个节点所拥有的权利和义务是对等的。根据移动商务的移动性和不稳定性，本文设计了一种适合移动终端的 P2P 管理机制来构建完全开放式好友圈，管理用户的加入、离开，以及代理的选举。每个用户均有在线、离线两种状态；新的用户可以动态加入新的好友圈，也可以因地理位置的移动或兴趣的侧重不同而退出好友圈。P2P 好友圈的构建主要包括 3 个主要任务：新用户的加入、代理的选举、离开管理。

1)加入管理。好友圈的构建是基于 P2P 网络模型的(见图 1)，每个移动用户节点之间相互平等。用户在需要获得推荐服务时，可主动向推荐服务器发送加入 P2P 好友圈的请求，并根据自己的爱好贴上标签。如果不知道如何贴标签，可以提供历史购物聚合数据，由推荐服务器检索已有好友圈，并反馈最为相似的好友圈代理(相当于 P2P 网络模型中的超级节点)给新用户，新用户可以直接联系并加入。如果拒绝加入或没有足够相似的好友圈(用户可设置一个相似阈值)，则可以自行创建一个新的推荐圈，并自动成为该圈的代理。每个用户只能隶属于一个好友圈。新用户向推荐服务器申请加入好友圈时，可动态地根据想要购买物品的属性权重不同选择不同的相似用户作为好友或邻居，比如若想要购买的物品与地理位置相关，则可以请求地理位置最为相似的在线用户作为好友；若关注购买物品的性价比，则可以选择收入最为相似的在线用户作为邻居等。根据自己的隐私需求就近联系 k 个最邻近好友(比如隐私需求为 k-匿名)，如果联系的 k 个好友中存在一个或多个用户已加入好友圈，则可以选择申请加入其中最相似的好友圈，并通过邻居节点所属代理发布加入信息。如果 k 个邻居中均不存在好友圈，则自动构建好友圈。通常由发起终端担任代理，并向邻居节点发布好友圈的构建信息。

2)代理选举。每个推荐圈都应有一个节点作为代理，即超级节点，负责收集该圈的实时增量数据。一般情况下，代理可由最先发起的终端担任，但如果用户隐私要求高，可由隐私要求高的终端担任，因为隐私要求高也就意味着要求的好友数目较多，代理需要处理的数据多。为防止恶意节点收集数据，代理采取轮流担任的办法，圈内所有节点按照加入时间的先后轮流担任代理。

3)离开管理。通常情况下，在一定的时间间隔内，用户与代理之间进行一次通信，超过阈值未得到回复则视为离线。代理发现普通用户离线后需检查好友圈的数量是否满足匿名

需求，不满足则需要重新构建好友圈。普通用户发现代理离线则需要重新启动加入机制。

4.2　代理匿名转发数据更新模块

P2P 好友圈构建完成后，移动用户即可向代理发送增量数据请求推荐服务。数据更新模块主要完成实时增量数据的匿名更新，为个性化推荐提供用户数据，同时无法唯一确定数据属于某一用户，以达到保护用户隐私的目的。本文设计的基于 P2P 的数据更新方式采取匿名转发方式，由代理统一匿名转发提交增量数据，不修改数据而又使得至少 k 个数据包不可区分，以确保在获得精确推荐的同时保护用户的隐私安全。

基于 P2P 的匿名转发机制如图 3 所示。

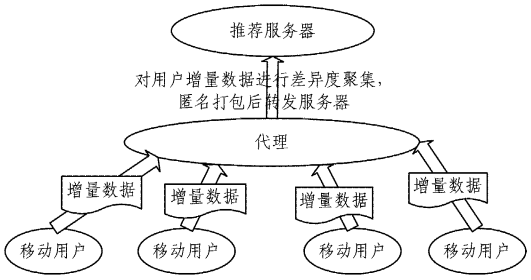

图 3　基于 P2P 的匿名转发机制

其主要过程如下：

步骤 1　数据打包。P2P 好友圈内普通用户为了获得更好的推荐效果，应根据自身的兴趣爱好提供个性化的信息，包括某时效内的历史购物数据或购物兴趣标签、隐私要求等，将其打包发送给代理以请求推荐。如果代理节点自身也需要推荐，则可以产生相应的数据包作为其中一个用户数据包。数据打包格式如表 2 所列。

表 2　增量数据包

ID	历史购物数据	时戳	隐私需求（k-匿名、l-多样性）	时效限制

步骤 2　k-匿名数据处理。为保证用户的隐私，代理收到增量数据包后需要对数据进行匿名处理，为保证隐私安全，至少对 k 个数据包进行匿名打包，切断其与用户个体的联系，使其不可区分后由代理统一转发给推荐服务器以获得推荐。由于移动终端的计算能力有限，本文提出的匿名处理机制不对增量数据本身进行任何处理来获得推荐的高度精确性，而是采用构建匿名数据包的方式切断数据包与用户之间的联系以达到用户匿名性的要求。同时为满足移动商务的时效限制，采取代理在匿名要求和时效限制不能同时满足的情况下，以生成虚拟数据包的方式达到匿名要求，同时满足移动商务推荐的实时性要求。

代理收到用户数据包后（包括自身的增量数据包），为防止一致性攻击，采用最小相似方法聚集用户数据包或虚拟数据包，以达到数据包多样化的目的。聚集数据包时，可采用格雷码序[22]进行排序分组，每组选取一个元素进行聚集以达到最小相似度聚集。当数据包达到匿名要求（k-匿名）时，则将数据进行匿名打包发给推荐服务器。匿名数据包的构建方法如下：去掉数据包 ID，根据数据包到达时间的先后给定序号，虚拟数据包的时效限制为无限，以此来区分有效数据包。匿名后的数据格式如表 3 所列。

表 3　匿名数据包

1	历史购物数据
2	历史购物数据
…	…

步骤 3　获得推荐。推荐服务器收到匿名数据包，根据某种推荐算法（协同过滤＋内容推荐）检索产品信息库，生成相应的推荐商品集，并附上相应序号反馈给代理。推荐算法的选择与好友圈的数目以及用户的兴趣相似程度有关。我们在实验部分验证了协同过滤算法明显优于内容推荐算法，且能有效避免冷启动问题[23]。代理收到推荐结果集后，可根据序号对推荐结果进行求精，丢弃虚拟数据推荐结果集，并分发推荐结果集给移动用户。

至此，整个推荐过程已完成，可以看到在整个推荐过程中，用户的个人信息不需要提交，存储在移动用户终端，且对用户的增量数据未做任何聚合或扰动处理，推荐服务器可以得到准确的增量数据，从而提高推荐精确度。增量数据的更新采取由代理匿名转发的方式实现了移动用户的匿名，使推荐服务器无法区分哪一个增量数据包对应于哪一个移动用户；并且由于移动用户的移动性和兴趣的侧重不同，某个移动用户不可能一直处于某个区域或者一直由某个终端代理转发，因此即使存在恶意代理，鉴于用户的地理位置移动或兴趣侧重不同而离开以及代理的更换机制，代理也只能收集到用户的部分增量数据。本文在实验部分设计了隐私风险率指标来度量存在恶意代理的情况下用户隐私数据的安全程度，实验验证了极端情况下用户隐私仍然得到了较好的保护。

5　实验

本节通过实验设计来验证本文提出的基于 P2P 隐私保护机制的可行性以及推荐效果的评价，包括实验数据集、实验环境以及评估实验效果的标准等。

5.1　数据集

为确保实验的公平性，本文采用文献[20]使用的 MovieLens数据集，它来自 MovieLens 站点（http://movielens.umn.edu）。数据集是由明尼苏达州立大学的 GroupLens 项目组提供的真实数据，为 6040 个用户对 3593 部电影的 10 万条评价，每个用户至少对 20 部电影做出评分。为了进行性能评估，我们将数据集分成训练集和测试集，90％的数据作为训练集，即平均每个用户随机选取约 18 部电影作为训练集，其他的作为测试集，用以检测推荐结果的准确性。

5.2　实验框架

本文设计了一个基于 P2P 模型的移动商务推荐的模拟系统，本实验模拟手机用户通过代理获得商务推荐，系统以 UCI 机器学习数据集里的 Adult 数据集的比例为每一位测试用户设定相应属性（比如地理位置、收入、年龄等），并为每一位用户生成随机位置，且随着时间的推移，随机生成离原始位置的移动距离。根据训练数据集生成历史记录数据，由代理匿名转发后获得模拟推荐商品。为保证数据的差异性，满足用户的隐私要求，生成虚拟数据包的商品信息尽量选择有效数据包中部分出现频度高的商品，再选择部分不包含在有效数据包中但与高频度商品较为相关的商品信息，构成虚拟数

据包，以尽量接近有效数据包的构成。

5.3 评价指标

本文从推荐的精确度、时效性以及隐私安全性 3 个方面来衡量本文的推荐机制是否有效。平均绝对误差 MAE[22-25]是推荐系统中最常用的一种推荐质量度量方法，本文采用 MAE 通过计算用户观看电影的真实值和预测值之间的偏差来度量预测的准确性。显然，MAE 越小，推荐质量越高。采用以下表达式计算 MAE 值：

$$MAE=\frac{\sum_{a\in A}|1-sim_a(u,Q)|}{|T|} \tag{1}$$

其中，T 是测试用户集合，$|T|$ 为测试集合的大小。MAE 的值越小说明预测的准确度越高。式(2)用于计算测试用户 a 的推荐物品与实际购买物品之间的相似度，显然 Sim 值越大，相似度越大。

$$Sim_a(u,Q)=\frac{1}{B}\sum_{i=1}^{B}II(u_i=Q_i) \tag{2}$$

其中，u 为用户 a 推荐的物品列表，Q 为用户 a 的实际购买物品列表，i 为项目列表中的第 i 个项目，函数 $II(x)$ 表示参数为真时函数返回值为 1，否则，返回值为 0。

5.4 实验分析

实验中共有 6000 个移动用户，初始用户数为 1000，以 100 个/s 的速度进入网络，用户节点的网络带宽设置为 100kBps。为了验证基于 P2P 的隐私保护机制的有效性，我们进行了大量的模拟实验，从模拟实验中观察到好友圈的推荐能力依赖于好友圈的用户的相似程度以及参与的移动用户的数量，如果好友圈的相似度不够，或者参与的移动用户数目过少，就会存在冷启动问题。冷启动问题包括新用户问题和新项目问题，新用户问题是由因新加入用户没有任何数据而无法进行推荐，本系统采取用户主动贴数据标签的方式来规避新用户问题；而新项目的冷启动问题则值得关注，由于新项目没有被移动用户浏览或评分，因此不能被推荐。模拟实验还表明终端的移动距离的跨度越大，用户隐私泄露的可能性就越小。同时实验也验证了移动跨度小的终端，根据不同的属性权重选择好友圈的频度越高，越能有效地阻止恶意节点对于隐私数据的攻击等。

由于文献[20]中提出的基于 P2P 的随机邻居选择方法与 k 值无关，因此本文分别采用基于项目的推荐算法和协同过滤推荐算法进行对比分析，实验结果如图 4 所示。本实验中的协同过滤推荐算法在进行推荐时综合考虑了匿名数据包中的其他增量数据，将其视为相似用户的增量数据，以增加推荐的精确性。系统会根据匿名用户的增量数据包将推荐物品按照推荐值大小排序后进行推荐。隐私度 k 值的增加代表用户的隐私需求更加严格，要求更多的用户构成 P2P 好友圈从而保护用户隐私。图 4 所示结果表明，在基于 P2P 的隐私保护策略中，k 的不同取值对基于项目的推荐算法的影响不大；但对协同过滤算法而言，随着用户隐私要求提高，k 值增大，也就意味着 P2P 好友圈中的用户数目越多，推荐效果越好。这是因为 P2P 好友圈主要是由兴趣相同的用户构成，而协同过滤推荐算法同时综合考虑了匿名数据包中其他增量数据包从而进行推荐，k 值越大，匿名数据包中的有效数据包越多，推荐精确度就越高。整体而言，协同过滤算法优于基于项目的推荐算法，因此以下实验中均采用协同过滤推荐算法。

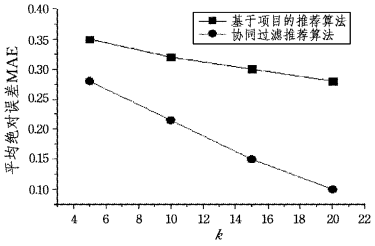

图 4 k 值变化对 MAE 的影响

图 5 示出平均绝对误差 MAE 随训练集大小变化的情况，其中 x 轴是对数刻度。从图 5 中可以看出，训练集的大小对于文献[20]的基于 P2P 的随机邻居选择方法和本文提出的基于 P2P 的代理转发隐私保护方法的影响差异明显。这是因为对基于 P2P 的随机邻居选择隐私保护方法而言，训练集的规模越小，邻居选择空间不大，推荐质量受到的影响小，但随着训练集的增长，邻居选择空间增大，随机选择算法选择的结果差异大，以致影响推荐效果。而本文提出的基于 k-匿名代理转发隐私保护方法中节点根据兴趣爱好的相似性选择好友圈加入，从某种程度上说，好友圈的邻居正是最佳邻居，而训练集的增长对于推荐质量的影响并不明显。我们也注意到在训练集过小时推荐误差明显，这是由推荐系统中新项目的冷启动问题所导致的。

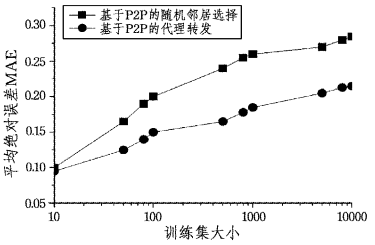

图 5 MAE 随训练集大小变化的情况

综上所述，本文提出的基于 P2P 的匿名转发隐私保护策略构建最优的 P2P 好友圈，采用协同过滤推荐算法进行推荐，能够获得较好的推荐体验。此外，我们也注意到大数目的好友圈会加重代理的系统负担，因此适当地加入鼓励机制和进一步发展终端设备会改善这一瓶颈问题。未来的工作中我们将会对如何选择最优 P2P 好友圈、添加代理的激励机制以及改善冷启动问题展开进一步探究。

结束语 本文研究了移动商务推荐系统中的隐私保护问题，针对其存在的挑战，提出了一种基于 P2P 的代理匿名转发的隐私保护策略，针对移动商务推荐系统的特性，通过将 P2P 网络中的匿名技术与 k-匿名隐私模型结合起来，达到同时满足实时性、高精度性和匿名性的目的。该策略构建了一种完全开放式的 P2P 好友圈，采用由好友圈代理对用户增量数据进行匿名转发处理的数据更新方式，不修改原始数据以保证准确的增量数据从而获得高精度推荐，同时采用代理匿名转发方式使得至少 k 个数据包不可区分以切断移动用户与增量数据之间的一对一关系，从而保护用户隐私安全。本文巧妙地利用移动电子商务系统和 P2P 网络存在的共性，将 P2P 网络中的匿名机制引入推荐系统中，首次将 k-匿名的思

想与 P2P 中代理匿名转发机制相结合，在某种意义上，为后来的研究提供了一种全新的思路。

参 考 文 献

[1] MENG X W, HU X, WANG L C, et al. Mobile recommender systems and their applications[J]. Journal of Software, 2013, 24(1): 91-108. (in Chinese)
孟祥武，胡勋，王立才，等. 移动推荐系统及其应用[J]. 软件学报，2013，24(1)：91-108.

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

[3] XU H L, WU X, LI X D, et al. Comparison study of Internet recommendation system [J]. Journal of Software, 2009, 20(2): 350-362. (in Chinese)
许海玲，吴潇，李晓东，等. 互联网推荐系统比较研究[J]. 软件学报，2009，20(2)：350-362.

[4] RESNICK P, VARIAN H R. Recommender systems [J]. Communications of the ACM, 1997, 40(3): 56-58.

[5] WANG L C, MENG X W, ZHANG Y J. Context-Aware recommender systems [J]. Journal of Software, 2012, 23(1): 1-20. (in Chinese)
王立才，孟祥武，张玉洁. 上下文感知推荐系统[J]. 软件学报，2012，23(1)：1-20.

[6] WÖRNDL W, SCHÜLLER C, WOJTECH R. A hybrid recommender system for context-aware recommendations of mobile applications[C]// Proc. of the Int'l Conf. on Data Engineering (ICDE 2007). Washington: IEEE Computer Society, 2007: 871-878.

[7] WANG S L, WU C Y. Application of context-aware and personalized recommendation to implement an adaptive ubiquitous learning system[J]. Expert Systems with Applications, 2011, 38(9): 10831-10838.

[8] ZHOU W L, WU X F. Survey of P2P technologies[J]. Computer Engineering and Design, 2006, 27(1): 76-79. (in Chinese)
周文莉，吴晓非. P2P 技术综述[J]. 计算机工程与设计，2006，27(1)：76-79.

[9] TVEIT A. Peer-to-Peer based recommendations for mobile commerce[C]// Proc of the 1st International Workshop on Mobile Commerce. New York: ACM Press, 2001: 26-29.

[10] BERKOVSKY S, et al. Enhancing privacy and preserving accuracy of a distributed collaborative filtering[C]// Proceedings of the 2007 ACM Conference on Recommender Systems. 2007.

[11] SHOKRI R, PEDARSANI R, THEODORAKOPOULOS G, et al. Preserving privacy in collaborative filtering through distributed aggregation of offline profiles[C]// Proc of the 3rd ACM Conference on Recommender Systems. New York: ACM Press, 2009: 157-164.

[12] POLAT H, DU W. Privacy-preserving collaborative filtering using randomized perturbation techniques[C]// Proc of the 3rd Internation Conference on Data Mining. Washington DC: IEEE Computer Society, 2003: 625-628.

[13] POLAT H, DU W. SVD-based collaborative filtering with privacy[C]// Proc of ACM Symposium on Applied Computing. New York: ACM Press, 2004: 791-795.

[14] ZHANG F Z, LIU T, FENG S S. Improved Privacy-preserving collaborative Filtering Recommendation Algorithm[J]. Computer Engineering, 2010, 36(16): 126-134. (in Chinese)
张付志，刘亭，封素石. 一种改进的隐私保持协同过滤推荐算法[J]. 计算机工程，2010，36(16)：126-134.

[15] BERKOVSKY S, EYTANI Y, KUFLIK T, et al. Privacy-enhanced collaborative filtering [C]// Proc of user Modeling workshop on Privacy-Enhanced Personalization. 2005: 75-83.

[16] BERKOVSKY S, KUFLIK T. Hierarchical neighborhood topology for privacy enhanced collaborative filtering[C]// Proc of CHI Workshop on Privacy-Enhanced Personalization. 2006: 6-13.

[17] CANNY J. Collaborative filtering with privacy [C]// Proc of IEEE Symposium on Security and Privacy. Washington DC: IEEE Computer Society, 2002: 45-57.

[18] CANNY J. Collaborative filtering with privacy via factor analysis [C]// Proc of the 25th Annual International ACM SIGIR Conference on Research and Development in information Retrieval. New York: ACM Press, 2002: 238-245.

[19] FRAN C, J D F. Constantinos Patsakis, Domenec Puig, Agusti Solanas, Privacy Preserving Collaborative Filtering with k-Anonymity through Microaggregation[C]// IEEE 10th International Conference on e-Business Engineering. 2013: 490-497.

[20] BAKKER A, OGSTON E, VAN STEEN M. Collaborative filtering using random neighbours in peer-to-peer networks[C]// Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledgemanagement. 2009.

[21] MILLER B, KONSTAN J, RIEDL J. PockLens: toward a personal recommender system[J]. ACM Trans. on Information Systems, 2004, 22(3): 437-476.

[22] SARWAR B, KARYPIS G, KONSTAN J, et al. Analysis of recommendation algorithms for E-commerce[C]// Proc. 2nd ACM Conf. Electronic Commerce. New York: ACM Press, 2000: 158-167.

[23] SARWAR B, KONSTAN J, BORCHERS A, et al. Using filtering agents to improve prediction quality in the groupLens research collaborative filtering system[C]// Proc. ACM Conf. Computer Supported Cooperative Work (CSCW). New York: ACM Press, 1998: 345-354.

[24] GOOD N, SCHAFER J B, KONSTAN J A, et al. Combing collaborative filtering with personal agents for better recommendations[C]// Proc. 16th National Conf. Artificial Intelligence (AAAI-99). Menlo Park, CA: AAAI/MIT Ptess, 1999: 439-446.

[25] SHARDANAND U, MAES P. Social Information Filtering: Algorithms for Automating "Word of Mouth"[C]// Proceedings 1995 ACM SIGCHI Conference on Human Factors in Computing Systems. Denver, CO, USA, 1995: 210-217.