空间众包中的位置隐私保护技术综述*

安 莹, 秦 科, 罗光春
(电子科技大学 计算机科学与工程学院, 成都 611731)

摘 要: 随着移动设备和无线网络的迅速发展,传感器能够更加精确地获取用户的位置、移动速度和方向等信息,空间众包中用户的位置隐私安全问题日益凸显,基于空间众包的位置隐私保护技术成为互联网隐私领域的研究热点。首先系统介绍了空间众包的基本概念、工作流程以及已有空间众包平台;归纳了空间众包中基于差分隐私、空间匿名以及加密技术的三种主流的隐私保护模型,对比分析了三种主流的隐私保护方法。最后总结并展望了未来的研究方向。

关键词: 空间众包; 隐私保护; *k*-匿名; 差分隐私
中图分类号: TP309.2 **文献标志码**: A **文章编号**: 1001-3695(2018)08-2241-04
doi:10.3969/j.issn.1001-3695.2018.08.001

Survey on location privacy preservation technology in spatial crowdsourcing

An Ying, Qin Ke, Luo Guangchun
(*School of Computer Science & Engineering*, *University of Electronic Science & Technology of China*, *Chengdu* 611731, *China*)

Abstract: Due to the rapid development of the mobile devices and the wireless network, sensors could accurately obtain the users' information such as position, moving speed and the direction, location privacy is becoming considerably important. The location privacy protection technology based on spatial crowdsourcing has become a hot research topic in the field of Internet privacy. This paper discussed the basic concept of spatial crowdsourcing, procedure and some application platforms. It analyzed and summarized of the state-of-the-art privacy preservation models based on differential privacy, also included spatial cloaking and encryption technology in spatial crowdsourcing. At last, it presented the future research work.

Key words: spatial crowdsourcing; privacy preservation; *k*-anonymity; differential privacy

美国《连线》杂志的记者 Howe 在 2006 年 6 月首次提出众包(crowdsourcing)这一概念,他将众包定义为:一个公司或机构把过去由雇佣员工或专职的外包人员执行的工作任务,通过公开的网络平台外包给非专职的群体的工作模式。众包模式下的工作任务通常是个体志愿负担完成,也可以由多人合作完成。众包的关键在于充分利用公开的网络平台上的劳动力资源完成简单或复杂的工作任务[1]。众包作为充分利用群体智慧的成功模型,已经广泛应用于个人独立完成或多人协作完成任务的场景中,例如图片标注、自然语言理解、市场预测以及观点挖掘等领域。从 2009 年开始,众包就得到了包括翻译、物流、交通、民宿等行业领域的广泛关注,并逐渐成为新的研究热点之一,面临着诸多理论及应用方面的挑战。

随着移动互联网技术及移动设备的计算和感知能力的提高,一类基于用户位置信息的众包形式应运而生。本文将这类基于用户位置信息的众包形式称为空间众包(spatial crowdsourcing,SC)[2]。空间众包通过雇佣工作者执行空间任务,其作为特殊的众包形式,已在学术界(如 gMission[3] 和 MediaQ[4])和工业界(如 TaskRabbit[5])逐渐兴起[6]。典型的空间众包平台将空间任务分配给就近的工作者,工作者移动到指定的位置并完成已分配的空间任务。通过空间众包平台,人们可以更好地利用群体智慧完成简单或复杂的空间任务。尽管空间众包充分利用了群体智慧并带来极大的效益,但是平台的构建和推广并不容易。空间众包根据用户提交的位置信息进行任务的发布或分配,而用户的位置信息隐含了用户的身份、家庭住址、健康状况和生活习惯等敏感信息[7]。近年来,软、硬件服务不断发展,智能手机可以充当多模式感知器,收集并分享多种类型的数据,包括图片、视频、位置、移动速度、方向以及加速度等信息,空间众包平台通过智能手机获取大量的用户位置数据将会导致用户敏感信息的泄露,严重威胁用户的隐私安全。本文介绍了一种新型的众包模式——空间众包,并通过调研归纳总结了空间众包中的三种主流的位置隐私保护方法。

1 空间众包

1.1 基本概念

定义 1 任务请求者[6,8]。任务请求者通过注册使用空间众包平台完成设计并发布空间任务,拒绝或者接收众包工作者的答案、整理众包工作者的答案等一系列工作。任务请求者通常定义为 $r=\langle \mathrm{loc}_r, T_r\rangle$,其中,$\mathrm{loc}_r$ 表示任务请求者的位置信息;T_r 表示任务请求者发布的任务。

定义 2 空间任务[2,6,8]。一个空间任务通常是具有地理位置和时间属性的特殊任务,一般定义为四元组 $T=\langle \mathrm{loc}_T, t_{\mathrm{start}}, t_{\mathrm{end}}, u_T\rangle$。其中,$\mathrm{loc}_T$ 表示空间任务的位置;t_{start} 表示空间任务的发布时间;t_{end} 表示空间任务的截止时间;u_T 表示完成该任务可以获得的报酬。

定义 3 空间众包工作者[2,6,8]。空间众包工作者是执行空间任务的移动设备使用者,空间众包工作者通过注册使用空间众包平台完成选择空间任务、接受任务分配、提交位置信息、提交任务结果等一系列工作。空间众包工作者通常定义为一个三元组 $w=\langle \mathrm{loc}_w, R_w, \max T\rangle$,其中,$\mathrm{loc}_w$ 表示空间众包工作者当前的位置信息;R_w 表示空间众包工作者可以接受任务的空间域;$\max T$ 表示空间众包工作者在空间域 R_w 内可以接受的最大任务数量。

定义 4 空间众包。完整的空间众包包括任务请求者、空间众包任务、空间众包平台以及空间众包工作者。空间众包通常是指任务请求者设计空间众包任务并将其发布到空间众包平台,空间众包平台实现任务分配等工作,空间众包工作者通过空间众包平台接受空间任务并到指定地点完成空间任务的

收稿日期: 2017-06-08; **修回日期**: 2017-08-31 **基金项目**: 电子科技大学中央高校基本科研业务费资助项目(ZYGX2016J083)
作者简介:安莹(1992-),女,河北定州人,硕士研究生,主要研究方向为机器学习、隐私保护(18200112719@163.com);秦科(1980-),男,副教授,博士,主要研究方向为信息安全、机器学习、大数据;罗光春(1974-),男,教授,博士,主要研究方向为计算机网络、云计算.

过程。空间众包基础模型如图 1 所示。

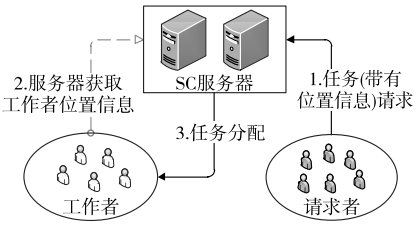

图1 空间众包基础模型

1.2 工作流程

空间众包平台作为空间众包工作的核心,为任务请求者和空间众包工作者建立基于空间任务的合作关系,并负责任务请求者和众包工作者提交的空间任务、个人位置等信息的综合处理。图 2 展示了空间众包的工作流程。一般地,空间众包平台先收集任务请求者设计提交的任务信息以及众包工作者的位置等信息,并由数据处理模块预处理后提交至任务分配模块,由任务分配模块完成任务的分配,最后由众包工作者完成空间任务并提交结果至质量控制模块。

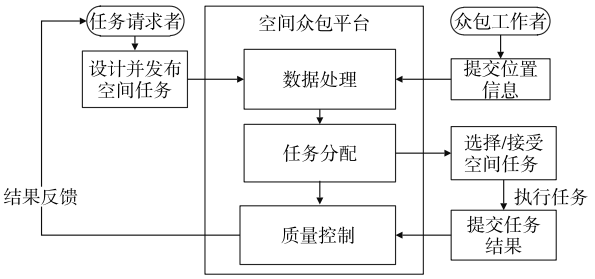

图2 空间众包的工作流程

根据空间任务的分配方式,空间众包可分为 WST(worker selected tasks,工作者选择任务模式)和 SAT(server assigned tasks,服务器分配任务模式)两种工作模式[2]。下面从工作者选择任务模式和服务器分配任务模式两个角度阐述空间众包的工作流程。

a)WST 模式工作流程。众包工作者根据自身的空间位置信息主动查找空间众包平台上发布的任务,并自主选择合适的空间任务去执行。

b)SAT 模式工作流程。众包工作者首先将其空间位置信息提交给空间众包平台,并由空间众包平台对任务位置数据和众包工作者提交的位置信息进行处理完成任务的分配,由众包工作者决策是否接受已分配任务。

1.3 应用平台

空间众包的任务发布、分配和提交等工作通过空间众包平台完成,如表 1 所示,介绍了 gMission[9]、MediaQ 以及与人们日常生活密不可分的滴滴出行[10]、百度外卖[11]四种常见的空间众包应用,并分析其存在的用户位置隐私泄露问题。

表 1 空间众包应用平台对比分析表

应用	gMission	MediaQ	滴滴出行	百度外卖
平台	智能手机	智能手机 Web 平台	智能手机	智能手机
用户	注册用户	注册用户	注册用户 认证司机	注册用户 认证骑手
功能	请求和采集地理信息,通过多种位置感知方案为用户进行任务推荐	注册用户可以发出特殊内容请求,该请求会自动提示在任务位置附近的工作者接受任务	注册用户发出乘车请求;滴滴出行派单系统派单给认证司机	用户在线完成外卖订单;派单系统派单给骑手,骑手选择是否接单
隐患	gMission 服务器能够获取工作者精确的位置信息,存在位置隐私泄露隐患	MediaQ 服务器需要获取精确的位置信息,存在位置隐私泄露的隐患	乘客以及司机的位置信息被滴滴服务器获取,存在位置隐私泄露的隐患	用户以及骑手的位置信息被百度服务器获取,存在位置隐私泄露的隐患

综上所述,空间众包已经深入人们的日常生活当中,同时,现有的空间众包平台只提供了基本的奖励机制、任务分配等服务,基本没有空间众包平台在满足基本服务的同时保护用户的隐私。然而,用户的位置数据能够反映用户的家庭住址、生活习惯等敏感信息,这显然泄露了空间众包平台用户的隐私。因此,基于空间众包的用户的隐私保护问题亟待解决。

2 空间众包中的隐私保护模型

调研发现,空间众包中典型的隐私保护模型主要有基于差分隐私技术的保护模型、基于空间匿名技术的保护模型和基于加密技术的隐私保护模型。其中,基于差分隐私技术的保护模型和基于空间匿名技术的保护模型较为常用。

2.1 基于差分隐私技术的保护模型

微软的 Dwork 等人在 2006 年首次提出差分隐私技术[12],2016 年 6 月份苹果公司在 WWDC 大会上提出使用差分隐私技术帮助从用户的大量数据中发现用户行为而不会泄露用户个体的隐私。本文提出基于位置数据保护的差分隐私定义。

定义 5 ε-位置数据差分隐私[13]。设随机算法 M 为随机查询函数,range(M)表示算法 M 所有可能的输出集合。D_1 和 D_2 为任意两个邻近数据集(即 D_1 和 D_2 两个数据集至多有一条位置记录不同),S 为 range(M)的一个子集,即 $S\subseteq$range(M)。若算法 M 满足

$$\ln\frac{\Pr[M(D_1)\in S]}{\Pr[M(D_2)\in S]}\leqslant\varepsilon$$

其中:概率 $\Pr[M(D_1)\in S]$ 和 $\Pr[M(D_2)\in S]$ 分别表示输出为 $M(D_1)$ 和 $M(D_2)$ 为 S 的概率;参数 ε 用于衡量隐私保护的强度,ε 越小意味着两种概率密度函数相似度越高,也就提供了高的隐私保护强度。由于 $\Pr[M(D_1)\in S]$ 与 $\Pr[M(D_2)\in S]$ 的近似程度与 ε 的取值有直接联系,所以在 ε 取值适当的情况下,对于一个特定输出 S,难以判定原位置数据集是 D_1 或 D_2,最终达到隐私保护的目的。

差分隐私保护机制是一种基于严格的数学背景,实现隐私保护程度可量化、可评估、可证明的方法。差分隐私保护技术通过在原始数据上添加随机噪声干扰敏感个人数据,保证原始数据中的敏感数据在获得保护的同时依然保持其原有的统计性质,以便分析人员执行良性的聚合分析。将差分隐私技术运用到空间众包中的位置数据发布中,能够有效防止基于背景知识的恶意攻击[12]。根据调研,本文总结并给出了空间众包环境下的基于差分隐私技术的位置数据发布基本模型,如图 3 所示。该模型通常先由一个可信的数据存储服务器收集位置数据,再采用差分隐私技术对原始位置数据进行隐私保护处理,最后为查询方提供可发布、可分析的安全数据。

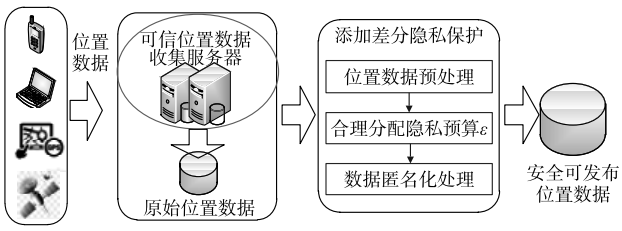

图3 基于差分隐私技术的位置数据发布基本模型

差分隐私技术的实质就是在查询函数的返回结果中添加可控数量的随机噪声,常用的噪声机制有 Laplace 机制[14]和指数机制[15]。

1)Laplace 机制 Laplace 机制一般用于查询结果为数值型的情况,是目前用于实现差分隐私常用的机制,其实质是向真实的查询请求结果 $f(D_1)$ 中添加噪声 η,得到最后的查询结果 $M(D_1)=f(D_1)+\eta$。其中 η 为满足 Laplace 分布的连续性随机变量,其概率密度函数为

$$p(\eta)=\frac{1}{2\lambda}e^{-\frac{|\eta|}{\lambda}}$$

参数 λ 与添加噪声的大小及隐私保护程度有关,λ 越大,噪声数量就越多,最终达到的隐私保护水平就会越高。除参数 λ 之外,敏感度也是影响隐私保护程度的重要因素,下面给出 Laplace 机制敏感度定义。

定义 6　Laplace 机制敏感度。对于任意的函数 f,f 的敏感度为

$$S(f) = \max_{(D_1,D_2)} \| f(D_1) - f(D_2) \|$$

其中:$\| f(D_1) - f(D_2) \|$ 为 $f(D_1)$ 和 $f(D_2)$ 之间的一阶范数距离。

定义 7　Laplace 机制。对于随机算法 M,若向函数 f 的输出添加的独立噪声为参数值取 $\frac{S(f)}{\varepsilon}$ 的 Laplace 分布,则算法 M 满足 ε-差分隐私;若该噪声为参数值取 λ 的 Laplace 分布,则算法 M 满足 $\frac{S(f)}{\varepsilon}$-差分隐私。也就是说参数 λ 的取值由敏感度 $S(f)$ 以及差分隐私参数 ε 决定。

2)指数机制　Laplace 机制要求算法 M 的输出必须是一个实数,只适用于查询结果为数值型的情况,具有一定的局限性。因此,McSherry 等人提出了指数机制,用于查询返回值为非数值型的情况。指数机制通常采用满足特定分布的随机抽样来实现差分隐私,相比添加噪声的方法,拓宽了差分隐私的适用范围。

定义 8　指数机制敏感度。给定一个效用估值函数 u,那么 u 的敏感度定义为

$$S(q) = \max_{T_1,T_2,r} \| q(D_1,r) - q(D_2,r) \|$$

其中:$r \in \text{range}$,range 为查询函数的输出域。效用估值函数 u 用来评估输出值的实用性,函数值越大,对应的查询函数的输出越容易被发布,从而保证发布数据的质量。

定义 9　指数机制。假设有一个效用估值函数 $u(D,r)$,随机算法 M 的输入为位置数据集 D,输出为 r,若算法 M 的输出为 r 的概率正比于 $\exp(\varepsilon u(D,r)/2S(q))$,则算法 M 满足 ε-差分隐私保护。

目前,基于差分隐私技术的空间众包隐私保护研究相对较少,现有的研究基本是采用 Laplace 机制来实现差分隐私保护。空间众包平台要求众包参与者实时提交真实的位置信息,而空间众包平台并不是一个可以完全信任的第三方。因此,基于差分隐私技术的空间众包隐私保护通常是将众包工作者的真实位置信息发送到一个可信的第三方,空间众包平台向该可信的服务器请求位置信息时,可信服务器会对众包参与者的真实位置信息进行差分隐私处理以供空间众包平台查询。通常基于差分隐私技术的空间众包隐私保护模型如图 4 所示。

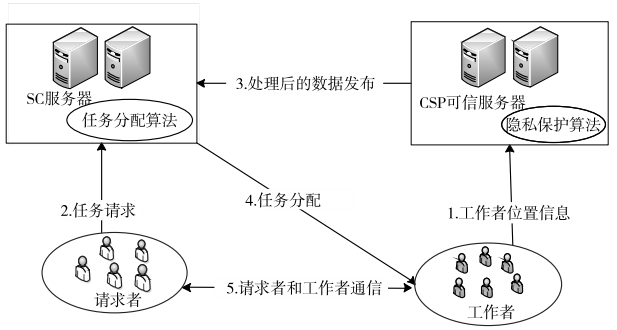

图4　基于差分隐私技术的空间众包隐私保护模型

To 等人[16]提出基于差分隐私的保护框架。空间众包中的工作者首先将自身的真实位置信息提交给可信的移动服务商,移动服务商对原始的位置信息构建 PSD(private spatial decompositions),并采用 Laplace 机制实现差分隐私保护。当任务请求者发送任务到空间众包平台时,空间众包服务器会根据任务的地理位置信息查询 PSD 并构建多播区域,最后将适宜的工作者推送给空间众包的任务请求者。Xiong 等人[17]同样是使用(cellular service provider)来存储工作者的确切位置信息,并通过添加 Laplace 噪声处理工作者的位置数据达到位置隐私保护的目的,然后使用处理后的位置数据与空间众包服务器进行交互。

两者的整体架构基本相同,都是使用 CSP 作为第三方服务器来存储工作者的位置信息,并使用差分隐私技术处理获取的真实数据,以达到隐私保护的目的。不同之处在于为达到差分隐私保护的目的,在构建 PSD 时使用了不同的技术方案,前者使用基于网格的方法 AG[18](adaptive grid)技术;后者则是构建满足差分隐私保护的工作者位置等高线来展示工作者的位置分布情况,空间众包服务器可以根据位置等高线来估算给定点工作者的密度及其变化趋势。相比基于网格的方法,采用等高线方法更能精确地估算给定点附近工作者的位置密度。

基于差分隐私技术的保护模型基本可以用于空间众包的各种应用场景,该技术的关键在于 PSD 的构建,针对不同的应用场景,选择不同的 PSD 构建方式将会达到需求的效果。现阶段,PSD 的构建主要有 *kd*-树、四分树以及网格结构三种方式。基于 *kd*-树的方法可以获得较高的精度,但是对二维空间数据的划分不稳定,容易出现划分结构不均匀的问题;由于四分树与结构数据无关,采用几何分配的方法分配隐私预算能够有效地提升查询精度,但是该方法同样没有考虑数据可能会存在不均匀的情况,导致结果出现误差;基于网格的构建方式则考虑了二维空间数据的不均匀性问题,合理设定了数据的划分粒度,但是该方法未能启发式地考虑数据的分布情况,可能会导致部分区域噪声过多,影响查询精度。

2.2　基于空间匿名技术的保护模型

除了使用差分隐私技术来保护空间众包中用户的位置信息以外,采用一个空间匿名区域来代替用户的精确位置信息的空间匿名技术也是常用的一种隐私保护技术。

k-匿名是由 Sweeney[19]首次提出的避免个人敏感数据泄露的一种隐私保护技术。他提出:若数据表中的每一条记录至少与其他 $k-1$ 条记录在准标志符上相一致,则该数据表满足 *k*-匿名。空间众包中的基于空间匿名技术的隐私保护通常采用 *k*-匿名模型,空间众包位置 *k*-匿名定义如下。

定义 10　空间众包位置 *k*-匿名。在空间众包中,工作者的位置属性为准标志符;在空间匿名区域中,空间众包中的任一工作者的位置不能从其他至少 $k-1$ 个工作者的位置中区分。其中,准标志符为联合其他外部信息、以较大概率辨识目标位置的最小属性集合[20]。

如图 5 所示,某空间众包工作者的真实位置为 l,空间匿名的核心思想即为将位置点 l 扩充为一个隐匿区域 R 来代替工作者的准确位置信息,在这个空间隐匿区域中每一个工作者被隐匿在其他至少 $k-1$ 个工作者中。从而,攻击者只能判断出工作者在隐匿区域内,而无法判断出工作者在隐匿区域内的准确位置。

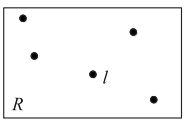

图5　空间*k*-匿名示意图

传统的位置数据保护研究大多是采用 *k*-匿名以及 *l*-多样性等方法模糊掉用户的位置信息,将用户的位置扩展成一个模糊的位置范围,从而实现对用户位置信息的保护。但是,模糊的用户位置信息将会导致空间众包中任务分配效率降低。因此,基于空间匿名技术的空间众包隐私保护模型一般通过结合其他区域划分等方法在实现模糊用户位置信息的同时保证任务分配质量。Hu 等人[21]研究了在 P2P 通信环境下的空间众包位置隐私保护问题,并采用 P2P 空间 *k*-匿名算法[22]实现空间众包工作者的位置隐私保护。P2P 空间 *k*-匿名算法的核心思想是任意的空间众包工作者能够通过多跳与至少 $k-1$ 位其他工作者通信。每一位工作者决定自己的空间匿名区域,若空间匿名区域包含工作者自身以及至少 $k-1$ 位最近邻的其他工作者,则这个匿名区域满足 *k*-匿名,因为在该空间匿名区域中,攻击者不能将该空间众包工作者从包含自身的 k 位工作者中

区分出来。在满足 k-匿名隐私需求的前提下,众包工作者将空间匿名区域扩展至大小至少为 A_{min} 的区域,以满足最小区域隐私需求。算法分为两步:

a)最近邻工作者的查找。

(a)用户 U 以 1 跳的距离向其邻居广播一个请求,每一个邻居响应请求并向用户 U 发送自己的身份和位置信息;

(b)用户 U 将接收到的邻居信息存放在一个列表 list 中,若 U 有至少 $k-1$ 个邻居,则 U 已获得足够的信息,执行步骤 b);否则,以步长 1 增加跳距不断查找邻居并将邻居的身份和位置信息存放到列表 list 中,直到找到 $k-1$ 个邻居。

b)匿名区域大小的选择[23,24]。

以步骤 a)的输出 list 作为输入,决策隐匿区域 A,该隐匿区域同时满足 k-匿名和最小区域隐私要求,即满足 numUser$(A)\geqslant k$ 和 area$(A)\geqslant A_{min}$。其中,numUser(A)代表隐匿区域中的用户数量,area(A)代表隐匿区域的大小,A_{min} 为满足要求的最小隐匿区域。

该方法解决了文献[16]研究的差分隐私空间分解方法中未考虑每个众包工作者具有空间区域属性的问题。另外,改进传统的 k-匿名技术并应用到 P2P 众包场景中,结合 V-Cover 问题决策众包工作者的匿名区域,保护众包工作者的位置隐私,解决了传统的空间匿名技术由于过强的隐私保护而影响任务分配质量的问题。文献[25]提出了基于局部敏感哈希[26]的隐私保护机制,将用户的位置进行分组,每组包含至少 k 位用户以实现空间匿名。由于局部敏感哈希具有很好的位置保持特性,很适合用来设计空间匿名。所提出的基于局部敏感哈希的位置分区算法,在参与式感知应用场景下保护用户的身份同时保护用户的位置信息。该算法的步骤如下:

a)使用 L 哈希方程针对用户的位置数据集 S 的哈希值构建一个排序列表 $l_1,l_2,\cdots,l_L$;

b)将每一个列表划分为多个桶,每一个桶包含 k 个元素,由于最后一个桶包含的元素个数可能多于 k 个,算法通常开始于 l_1 的第一个可用位置点 q;

c)依次从 $l_1,l_2,\cdots,l_L$ 中选取包含 q 的桶,作为 q 的 $k-1$ 最近邻点;

d)重复步骤 c)直到排序列表为空。

文献[25]采用基于局部敏感哈希的方法,解决了简单划分导致的不均匀问题,实现了在较低的时间复杂度下空间数据的理想分区,并且能够在攻击者已知用户的匿名算法的情况下较好地防止用户隐私被攻击。

现有基于空间匿名技术的模型的研究认为:相对于基于差分隐私技术的保护模型,采用空间匿名技术的保护模型的应用场景十分受限,并且容易出现隐私保护强度过大导致任务分配质量降低,因此,需要根据实际的应用需求,综合其他的空间数据处理方法来实现空间众包中的隐私保护。同时,该模型的性能一般低于基于差分隐私技术的保护模型。

2.3 基于加密技术的隐私保护模型

此外,还有部分研究人员将传统的数据隐私保护技术,如基于数据加密的技术等,改进并应用到空间众包平台以达到保护空间众包工作者位置隐私的目的。

Shen 等人[27]认为现实中不存在像 CSP 一样完全可信的数据存储服务器,并在基于差分隐私技术的空间众包架构上进行了改进,提出了使用 PSP(privacy service provider)作为第三方服务器来存储空间众包用户的位置信息。PSP 与空间众包服务器相同都是半可信的,但是彼此之间没有任何利益关系。由于 PSP 为半可信服务器,工作者将精确的位置信息发送给 PSP 就会存在位置隐私泄露的安全隐患,空间众包用户的位置信息需要进行加密处理之后发送给该服务器。于是,Shen 等人首次提出一种基于加法同态加密技术并采用乱码电路[28]比较加密值的新型加密协议来构建安全加密数据库用以存储工作者的位置信息;然后,PSP 向空间众包服务器提供处理过的数据以供空间众包服务器查询位置信息并进行任务分配。

由于只有极少数的研究人员在空间众包中应用加密技术保护用户的位置隐私,在此不作赘述。

3 空间众包中隐私保护技术对比

由于无线通信技术和空间众包模式不断发展创新,空间众包中的位置隐私保护技术正处于并在未来的一段时间内持续处于研究的高峰期。本章对比了已有的隐私保护模型和方法,如表 2 所示。

表 2 空间众包中隐私保护技术对比表

技术分类	隐私保护水平	优点	缺点
差分隐私	高	提供可证的隐私保护,能够对抗背景知识攻击	噪声数量难以控制,容易导致数据可用性降低
空间匿名	中	具有较好的数据安全性和可用性	不能抵抗背景知识攻击
加密技术	高	具有较强的隐私保护水平	运行开销大,需要构建专门的数据库

4 展望

空间众包的位置隐私保护技术作为一个新兴的研究领域,还有很多方面值得研究人员深入研究。下面从三方面简述笔者未来的研究方向。

a)传统 LBS 隐私保护技术在空间众包中的改进与应用。现阶段,基于空间匿名技术的隐私保护通常采用传统的 k-匿名方法进行建模,如何改进该方法,甚至采用以 k-匿名为基础的隐私保护模型(例如,(α,k)-匿名[20],(k,e)-匿名[29]等)保护空间众包中的位置隐私将是未来空间众包的隐私保护研究领域的一大挑战。

b)挖掘空间众包中其他隐私保护。由于空间众包中的用户提供的是动态的位置数据,在某些特殊的空间众包中,研究如何保护空间众包用户的轨迹隐私以及发掘其他的敏感数据并实现隐私保护将是未来空间众包的一个研究热点。

c)空间众包位置数据建模问题[30]。目前,空间众包中用户的位置数据大多是基于网格坐标建模,如何构建更切合实际生活的网格坐标模型或构建新型的位置数据模型,并选择合适的建模参数是未来非常值得研究人员深入研究的问题。

5 结束语

本文主要阐述了空间众包位置隐私保护技术的基本方法和研究进展,同时详细介绍了空间众包的基本概念、工作流程,并列举了部分典型的应用平台。总的来说,空间众包的位置隐私保护研究已经取得了一些成果,但是作为新兴的研究领域仍然有很多问题亟待解决。

参考文献:

[1] Howe J. The rise of crowdsourcing[J]. Wired Magazine, 2006, 14(6): 1-4.

[2] Kazemi L, Shahabi C. GeoCrowd: enabling query answering with spatial crowdsourcing[C]//Proc of the 20th International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2012: 189-198.

[3] gMission[EB/OL]. (2014-01-05)[2017-06-05]. http://gmission.github.io/.

[4] MediaQ[EB/OL]. (2014-04-21)[2017-06-05]. http://mediaq.usc.edu:8080/home/#about.

[5] What TaskRabbit offers[EB/OL]. (2017-08-25). https://support.taskrabbit.com/hc/en-us/articles/204411410-What-TaskRabbit-Offers.

[6] Zhao Yongjian, Han Qi. Spatial crowdsourcing: current state and future directions[J]. IEEE Communications Magazine, 2016, 54(7): 102-107.

[7] 吴振刚,孙惠平,关志,等. 连续空间查询的位置隐私保护综述[J]. 计算机应用研究,2015,32(2):321-325. (下转第 2264 页)

对数据进行预处理,改进最近邻选取方法,根据相似簇和最近邻生成基本预测评分,从而降低了数据的维度,并且有效地减少了计算量;同时建立了用户兴趣模型,从用户评分偏置和用户项目类型偏好两方面出发进行建模;最后通过多元线性回归确定每部分的权重,生成最终推荐结果。算法对不同用户进行针对性分析,充分结合了用户本身的评分偏好,实验结果表明,新算法对推荐系统的准确性起了积极的作用。

协同过滤推荐算法是推荐系统研究的核心内容,随着机器学习和数据挖掘技术的发展,在不同的应用场景下涌现出越来越多的新思路。例如,信任网络和结合时间属性的模型的应用。同时,也有很多研究者通过情感分析和领域知识对推荐算法进行改进,使得推荐算法不断向前发展。

参考文献:

[1] Herlocker L, Konstan A, Borchers S A, *et al*. An algorithmic framework for performing collaborative filtering [C]// Proc of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 1999: 230-237.

[2] Goldberg D, Nichols D, Oki B M, *et al*. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.

[3] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. IEEE Trans on Knowledge & Data Engineering, 2005, 17(6): 734-749.

[4] Goldberg K, Roeder T, Gupta D, *et al*. Eigentaste: a constant time collaborative filtering algorithm [J]. Information Retrieval Journal, 2001, 4(2): 133-151.

[5] You Haipeng, Li Hui, Wang Yunmin, *et al*. An improved collaborative filtering recommendation algorithm combining item clustering and slope one scheme [C]//Lecture Notes in Engineering & Computer Science, vol 2215. 2015: 313-316.

[6] Frémal S, Lecron F. Weighting strategies for a recommender system using item clustering based on genres[J]. Expert Systems with Applications, 2017, 77(7): 105-113.

[7] Guo Guibing, Zhang Jie, Yorke-Smith N. Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems [J]. Knowledge-Based Systems, 2015, 74(1): 14-27.

[8] 邓爱林,左子叶,朱扬勇. 基于项目聚类的协同过滤推荐算法[J]. 小型微型计算机系统,2004,25(9):1665-1670.

[9] 黄创光,印鉴,汪静,等. 不确定近邻的协同过滤推荐算法[J]. 计算机学报,2010,33(8):1369-1377.

[10] Herlocker J. Clustering items for collaborative filtering [C]// Proc of ACM SIGIR Workshop on Recommender Systems. New York: ACM Press, 1999.

[11] Beutel A, Beutel A, Ahmed A, *et al*. ACCAMS: additive co-clustering to approximate matrices succinctly[C]//Proc of the 24th International Conference on World Wide Web. Switzerland: International World Wide Web Conferences Steering Committee, 2014: 119-129.

[12] MovieLens_100K[DB/OL]. https://grouplens.org/datasets/movielens/.

[13] Forsati R, Barjasteh I, Masrour F, *et al*. PushTrust: an efficient recommendation algorithm by leveraging trust and distrust relations [C]//Proc of the 9th ACM Conference on Recommender Systems. New York: ACM Press, 2015: 51-58.

[14] Sarwar B, Karypis G, Konstan J, *et al*. Item-based collaborative filtering recommendation algorithms[C]//Proc of International Conference on World Wide Web. New York: ACM Press, 2001: 285-295.

[15] Altingovde I S, Subakan Ö N, Ulusoy Ö. Cluster searching strategies for collaborative recommendation systems[J]. Information Processing & Management, 2013, 49(3): 688-697.

[16] 王茜,杨莉云,杨德礼. 面向用户偏好的属性值评分分布协同过滤算法[J]. 系统工程学报,2010,25(4):131-138.

[17] 郭均鹏,赵梦楠. 面向在线社区用户的群体推荐算法研究[J]. 计算机应用研究,2014,31(3):696-699.

[18] 崔春生. 推荐系统中显式评分输入的用户聚类方法研究[J]. 计算机应用研究,2011,28(8):2856-2858.

[19] 范波,程久军. 用户间多相似度协同过滤推荐算法[J]. 计算机科学,2012,39(1):23-26.

[20] 黄震华,张佳雯,田春岐,等. 基于排序学习的推荐算法研究综述[J]. 软件学报,2016,27(3):691-713.

(上接第 2244 页)

[8] 冯剑红,李国良,冯建华. 众包技术研究综述[J]. 计算机学报,2015,38(9):1713-1726.

[9] Chen Zhao, Fu Rui, Zhao Ziyuan, *et al*. gMission: a general spatial crowdsourcing platform [J]. Proceedings of the VLDB Endowment, 2014, 7(13): 1629-1632.

[10] DiDi[EB/OL]. [2017-06-05]. http://www.xiaojukeji.com/website/about.html.

[11] BaiDuWaiMai[EB/OL]. [2017-06-05]. http://waimai.baidu.com/waimai?qt=about.

[12] 张琳,刘彦,王汝传. 位置大数据服务中基于差分隐私的数据发布技术[J]. 通信学报,2016,37(9):46-54.

[13] Hassan U U, Curry E. Multi-armed bandit approach to online spatial task assignment[C]//Proc of the 11th International Conference on Ubiquitous Intelligence and Computing. Washington DC: IEEE Computer Society, 2014: 212-219.

[14] McSherry F, Talwar K. Mechanism design via differential privacy [C]//Proc of the 48th Annual IEEE Symposium on Foundations of Computer Science. Washington DC: IEEE Computer Society, 2007: 94-103.

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

[16] To H, Ghinita G, Shahabi C. Framework for protecting worker location privacy in spatial crowdsourcing [J]. Proceedings of the VLDB Endowment, 2014, 7(10): 919-930.

[17] Xiong Ping, Zhang Lefeng, Zhu Tianqing. Reward-based spatial crowdsourcing with differential privacy preservation [J]. Enterprise Information Systems, 2016, 11(10): 1-18.

[18] Qardaji W, Yang Weining, Li Ninghui. Differentially private grids for geospatial data[C]//Proc of the 29th International Conference on Data Engineering. Washington DC: IEEE Computer Society, 2013: 757-768.

[19] Sweeney L. *k*-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[20] 吴英杰,唐庆明,倪巍伟,等. 基于取整划分函数的 *k* 匿名算法[J]. 软件学报,2012,23(8):2138-2148.

[21] Hu Jie, Huang Liusheng, Li Lu, *et al*. Protecting location privacy in spatial crowdsourcing [M]//Web Technologies and Applications. Cham: Springer International Publishing, 2015: 113-124.

[22] Chow C Y, Mokbel M F, Liu Xuan. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments [J]. GeoInformatica, 2011, 15(2): 351-380.

[23] Kazemi L, Shahabi C. A privacy-aware framework for participatory sensing[J]. ACM SIQKDD Explorations Newsletter, 2011, 13(1): 43-51.

[24] Kleinberg J, Tardos É. Algorithm design[M]. Boston: Addison Wesley, 2005.

[25] Vu K, Zheng Rong, Gao Jie. Efficient algorithms for *k*-anonymous location privacy in participatory sensing[C]//Proc of IEEE INFOCOM. Piscataway, NJ: IEEE Press, 2012: 2399-2407.

[26] Datar M, Immorlica N, Indyk D, *et al*. Locality-sensitive hashing scheme based on p-stable distributions[C]//Proc of the 20th Annual Symposium on Computational Geometry. New York: ACM Press, 2004: 253-262.

[27] Shen Yao, Huang Liusheng, Li Lu, *et al*. Towards preserving worker location privacy in spatial crowdsourcing [C]// Proc of IEEE Global Communications Conference. Piscataway, NJ: IEEE Press, 2015: 1-6.

[28] Yao A C. How to generate and exchange secrets [C]// Proc of the 27th Annual Symposium on Foundations of Computer Science. Washington DC: IEEE Computer Society, 1986: 162-167.

[29] Xu Jian, Wang Wei, Pei Jian, *et al*. Utility-based anonymization using local recoding [C]// Proc of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2006: 785-790.

[30] 童咏昕, 袁野, 成雨蓉, 等. 时空众包数据管理技术研究综述[J]. 软件学报, 2017, 28(1): 35-58.