DOI:10.7544/issn1000-1239.20220504
59(10):2261-2274,2022

效用优化的本地差分隐私集合数据频率估计机制

曹依然[1] 朱友文[1,2] 贺星宇[1] 张 跃[1]
[1](南京航空航天大学计算机科学与技术学院 南京 211106)
[2](广西可信软件重点实验室(桂林电子科技大学) 广西桂林 541004)
(caoyiran@nuaa.edu.cn)

Utility-Optimized Local Differential Privacy Set-Valued Data Frequency Estimation Mechanism

Cao Yiran[1], Zhu Youwen[1,2], He Xingyu[1], and Zhang Yue[1]
[1](*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing* 211106)
[2](*Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin, Guangxi* 541004)

Abstract In recent years, local differential privacy has received much attention because of its advantages of not requiring trusted third parties, less interaction, and high efficiency. However, the existing frequency estimation mechanism under local differential privacy for set-valued data fails to take into account the privacy sensitivity differences of inputs, and treats all data equally, which will over-protect the non-sensitive data and lead to low accuracy of estimation results. To address this problem, the set-valued data utility-optimized local differential privacy (SULDP) model is defined. SULDP considers the case that the original data domain contains both sensitive and non-sensitive values, and allows for a reduction in the protection of non-sensitive values without weakening the protection of sensitive values. Further, five frequency estimation mechanisms suGRR, suGRR-Sample, suRAP, suRAP-Sample and suWheel are proposed under the SULDP model. Theoretical analysis confirms that the proposed schemes can achieve exactly the same protection on sensitive data compared with local differential privacy mechanisms, and improve the accuracy by loosening the protection of non-sensitive data. Finally, the new schemes are evaluated on real and simulated datasets, and the experimental results demonstrate that the proposed five mechanisms can effectively reduce the estimation error and improve the data utility, among which suWheel mechanism achieves best performance.

Key words local differential privacy; frequency estimation; set-valued data; privacy protection; utility optimization

摘 要 本地差分隐私具有不需要可信第三方、交互少、运行效率高等优点,近年来受到了广泛关注.然而,现有本地差分隐私集合数据频率估计机制未能考虑数据的隐私敏感度差异,将所有数据同等对待,

收稿日期:2022-06-11;**修回日期**:2022-08-09
基金项目:国家重点研发计划项目(2021YFB3100400);国家自然科学基金项目(62172216);江苏省自然科学基金项目(BK20211180);广西可信软件重点实验室开放课题(KX202034)
This work was supported by the National Key Research and Development Program of China (2021YFB3100400), the National Natural Science Foundation of China (62172216), the Natural Science Foundation of Jiangsu Province of China (BK20211180), and the Research Fund of Guangxi Key Laboratory of Trusted Software (KX202034).
通信作者:朱友文(zhuyw@nuaa.edu.cn)

这会对非敏感数据保护过强,导致估计结果准确度低.针对这一问题,定义了集合数据效用优化本地差分隐私(set-valued data utility-optimized local differential privacy, SULDP)模型,考虑了原始数据域同时包含敏感值和非敏感值的情况,在不减弱对敏感值保护的前提下,允许降低对非敏感值的保护.进一步,提出了符合 SULDP 模型的 5 种频率估计机制 suGRR,suGRR-Sample,suRAP,suRAP-Sample 和 suWheel,理论分析证实,相对于现有的本地差分隐私机制,所提方案能够对敏感数据实现完全相同的保护效果,并通过降低非敏感数据的保护效果,实现了频率估计结果的准确度提升.最后,在真实和模拟数据集上评估了新的方案,实验结果证明了所提的 5 种机制能够有效降低估计误差,提升数据效用,其中 suWheel 机制表现最优.

关键词 本地差分隐私;频率估计;集合数据;隐私保护;效用优化

中图法分类号 TP309.2

随着经济科技的飞速发展和信息技术的普及应用,人们时时刻刻都在不断地产生大量的数据信息.作为一种重要的数据形式,集合数据在日常生活中有着广泛的应用场景,如购买过的物品、近期到过的地点等,都可以表示为集合数据.通过对这些数据进行收集、记录和分析,可以挖掘出它们中的隐藏信息,对学术、工业、社会服务等多个领域都有重要意义.例如通过分析用户的购买记录,可以得出用户购买需求,从而提高交易成功率;通过分析车联网系统中的 GPS 数据,可以得出实时路况和拥堵情况,从而提供交通流量控制服务.但这些数据中往往包含着大量隐私信息,如购物数据会反映个人生活习惯和财务状况、轨迹数据会包含家庭和工作地址,如果直接将这些数据提供给其他人使用,不仅会对个人的人身安全、财产安全造成极大的威胁,也会使得用户不再愿意共享数据.目前,区块链[1]、边缘计算[2]、机器学习[3-4]等领域已经关注到了隐私保护问题的重要性,并针对特定问题提出了隐私保护的解决方案.然而,如何在保护用户隐私的前提下对数据进行收集,仍是学术界亟待解决的问题.

当今,针对隐私保护,主要的解决方案可以分为 3 类:匿名化技术[5-6]、密码学技术[7-8],以及差分隐私[9-10].差分隐私拥有严格的形式化安全模型,并具有高效低开销的特点,是当今研究的热点方向.作为差分隐私的一大变种,本地差分隐私(local differential privacy, LDP)[11]在继承了差分隐私优点的基础上,摒弃了对可信第三方的需求,大大提高了模型的实用性.自从 2013 年被正式提出至今,本地差分隐私技术已经有了长足的发展与改进,其应用场景也十分宽泛,如机器学习[12-13]、网络服务[14]、数据的统计与优化[15-17],等等.同时,本地差分隐私也已经广泛部署到工业界实际应用中,苹果公司将其应用到手机上来保护用户的使用数据[18-19],谷歌公司也设计了基于本地差分隐私的组件 RAPPOR[20]来采集用户行为数据.

但是,现有的本地差分隐私机制大多并未考虑到,针对实际应用中不同数据的隐私保护需求差异,可以通过降低对非敏感数据的保护程度来减小估计误差.为此,Murakami 等人提出了效用优化本地差分隐私(utility-optimized local differential privacy, ULDP)模型[21],通过对不同的数据处以不同的扰动方法,从而在保障敏感数据隐私安全的前提下提高数据效用.然而,ULDP 模型仅仅可以处理用户数据均为敏感值或均为非敏感值的情况,当用户数据为集合数据,并且既包含敏感值又包含非敏感值时,ULDP 模型难以直接适用.在很多现实应用中,用户的集合数据会同时包含敏感值和非敏感值,比如,一位用户的购物订单记录为{梳子,洗发水,卫生纸,胃药},此时梳子、卫生纸和洗发水是非敏感值,而胃药则是敏感值,ULDP 模型无法直接处理该类数据.因此,需要提出一种新的模型,来处理用户数据且既包含敏感值又包含非敏感值的情况,保证在不降低对敏感数据保护效果的前提下,提高频率估计结果准确度.

本文首先定义了针对集合数据的效用优化本地差分隐私模型,该模型是 ULDP 模型在集合数据上的理论拓展,可以处理用户数据且既包含敏感值又包含非敏感值的情况,具有更广泛的适用性.随后,本文基于集合数据效用优化本地差分隐私(set-valued data utility-optimized local differential privacy, SULDP)模型提出了 5 种频率估计机制,并通过理论分析证明了所提出的 5 种机制能够对敏感数据实现与现有的本地差分隐私机制完全相同的保护效果,同时通过降低对非敏感数据的保护效果,提高频率估计结果的准确度.最后,本文在多个真实数据集和模拟数

据集上展开了实验,结果表明本文所提机制可以有效降低估计误差,提升整体数据效用.

概括地说,本文的主要贡献包括 3 个方面:

1) 定义了 SULDP 模型;

2) 基于传统频率估计机制,本文提出了符合 SULDP 模型的 5 种频率估计机制 suGRR,suGRR-Sample,suRAP,suRAP-Sample 和 suWheel;

3) 通过理论分析和实验对比了提出的 5 种机制,证明了这 5 种机制相较于原始机制均在效用方面有所提升,其中 suWheel 机制表现最优.

1 相关工作

针对用户集合数据进行的统计分析在推荐系统等领域都有着重要的研究意义,直观地,我们可以将针对类别数据的方法应用到集合数据的每一个条目上,但同时为了满足差分隐私的定义,还需要根据条目数量对隐私预算进行划分,当数量较大时,会导致数据效用急剧降低.

文献[22]基于 RAPPOR 实现了集合数据频率估计,虽然比直接划分隐私预算的效果要好,但是在进行扰动时仍然需要将隐私预算除以集合数据条数,集合变大时数据效用仍然很低,并且通信代价也很高.文献[23]提出了 PrivSet 机制,它是基于指数机制实现的,输出域为数据域的大小固定为 k 的所有子集,根据与输入集合是否相交来确定各个子集的输出概率,数据效用相较于文献[22]有所提升,但是 k 的最优取值是需要根据理论分析的均方误差(*MSE*)来确定的,由于分析结果非常复杂,无法直接计算出最优的 k,只能通过顺序查找来找到,当数据域较大时就会造成很高的计算代价.文献[24]提出了 Wheel 机制,这也是目前效果最好的集合数据频率估计机制.Wheel 机制通过哈希将原始集合数据映射到[0,1)上,然后以一定的概率密度从[0,1)上取值作为输出,Wheel 机制的通信代价比 RAPPOR 和 PrivSet 都要低,并且文献[24]还证明了 Wheel 机制的 *MSE* 达到了集合数据频率估计 *MSE* 的理论最优下界.但是 Wheel 等机制都是在 LDP 的定义下实现的,LDP 模型为所有用户和所有输入提供同等的隐私保护.

文献[21]将输入域分为敏感数据和非敏感数据,在保证敏感数据的不可区分性的前提下,以一定的概率降低对非敏感数据的保护效果,提高了整体的数据效用.类似的机制还有文献[25-26]提出的个性化本地差分隐私,为用户提供了个性化的隐私需求,但并未考虑到数据层面.而文献[27]提出的地理位置不可区分性则主要根据不同数据之间的距离来确定它们输出同一结果的概率,距离越近则概率越大,这也可以在一定程度上提高数据效用,但是由于距离度量需要满足三角不等式,因此针对某些数据类型是不适用的.

但是 ULDP 是基于类别数据提出的,如果想直接应用到集合数据就需要划分隐私预算,这会对数据效用产生极大的影响.综上,现有方法均存在不足之处,本文针对集合数据的效用优化频率估计机制进行了研究.

2 背景知识

2.1 本地差分隐私

在本地差分隐私中,由用户自己在本地对数据进行扰动,然后将扰动后的数据发送给服务器,服务器再利用这些数据计算得到所需的统计信息.由于服务器无法接触到用户的原始数据,因而无法获得用户的隐私信息.本地差分隐私的形式化定义如下.

定义 1. ε-LDP[11].给定隐私预算 $\varepsilon\geqslant 0$,对输入域为 X、输出域为 Y 的扰动机制 $M:X\to Y$,该扰动机制 M 满足 ε-LDP,当且仅当对于任意输入 x,$x'\in X$,得到任意输出 $y\in Y$ 的概率满足

$$M(y|x)\leqslant e^{\varepsilon}M(y|x'). \tag{1}$$

不难看出,ε-LDP 保证了任意攻击者无法从输出结果推断出确切的原始输入,并且当隐私预算 ε 趋近于 0 时,X 中所有数据都以几乎相同的概率输出同一结果,即隐私预算 ε 越小,对用户隐私保护程度越强.

2.2 效用优化本地差分隐私

LDP 并未考虑用户不同数据的敏感度差异,如统计用户身体状况时,对于“癌症”和“感冒”都使用了同样的扰动方法,这会使得对一些非敏感数据“感冒”保护效果过强,使得数据效用减弱.效用优化本地差分隐私就是针对这一问题进行了改进.ULDP 将原始数据集 X 划分成敏感数据集 X_{sen} 和非敏感数据集 X_{non},将输出集划分为保护数据集 Y_{pro} 和可逆数据集 Y_{inv},它的形式化定义如下.

定义 2. $(X_{\text{sen}},Y_{\text{pro}},\varepsilon)$-ULDP[21].给定 $X_{\text{sen}}\subseteq X$,$Y_{\text{pro}}\subseteq Y$,$\varepsilon\geqslant 0$,对输入域为 X、输出域为 Y 的扰动机制 $M:X\to Y$,当且仅当扰动机制 M 满足以下性质时,M 满足 $(X_{\text{sen}},Y_{\text{pro}},\varepsilon)$-ULDP.

1) 对于任意 $y\in Y_{inv}$,有且仅有一个 $x\in X_{non}$,

$$M(y|x)>0, \tag{2}$$

且对于任意 $x\neq x'$,有

$$M(y|x')=0. \tag{3}$$

2) 对于任意输入 $x,x'\in X$,得到任意给定输出 $y\in Y_{pro}$ 的概率满足

$$M(y|x)\leqslant e^{\varepsilon}M(y|x'). \tag{4}$$

2.3 现有的频率估计机制

我们介绍了 3 种常见的本地差分隐私频率估计机制 GRR(generalized randomized response)机制、RAPPOR 机制和 Wheel 机制的扰动过程.这 3 种机制均可应用于保护隐私的集合数据频率估计,然而都难以应对集合数据中同时包含敏感值和非敏感值的情况.

2.3.1 GRR 机制

在 GRR 机制[28]中,输出域与输入域相同,即 $X=Y$,给定 $\varepsilon\geqslant 0$,则 GRR 以表达式(5)所示的概率将 x 扰动成 y,

$$M(y|x)=\begin{cases}\dfrac{e^{\varepsilon}}{e^{\varepsilon}+|X|-1}, & 若\ y=x,\\ \dfrac{1}{e^{\varepsilon}+|X|-1}, & 若\ y\neq x.\end{cases} \tag{5}$$

式(5)表示了 GRR 处理类别数据的方式,若要处理集合数据,则需要先对数据进行抽样,将其转化为类别数据,或者根据集合数据条数划分隐私预算,再对集合中每条数据分别处理.

2.3.2 RAPPOR 机制

RAPPOR 机制[20]需要先对原始数据进行编码,通过独热编码将 x 映射为向量 $\boldsymbol{c}$,然后再对 $\boldsymbol{c}$ 进行扰动,当用户数据为类别数据时,$\boldsymbol{c}$ 中只有 1 位为 1,这一位以 $p=e^{\varepsilon/2}/(e^{\varepsilon/2}+1)$ 的概率保持不变,其余为 0 的位则以 $q=1/(e^{\varepsilon/2}+1)$ 的概率反转为 1.

随后工作[22]又对 RAPPOR 进行了改进,使之可以处理集合数据.设每个用户有 m 条数据,则编码后的向量 $\boldsymbol{c}$ 中有 m 个 1,为了满足 ε-LDP,这些位保持不变的概率为

$$p=\frac{e^{\varepsilon/2m}}{e^{\varepsilon/2m}+1}, \tag{6}$$

相应地,为 0 的位反转为 1 的概率为

$$q=\frac{1}{e^{\varepsilon/2m}+1}. \tag{7}$$

此外,与 GRR 类似,RAPPOR 也可以直接从集合中抽样出一条数据,然后直接使用原始的 RAPPOR 进行处理.

2.3.3 Wheel 机制

Wheel 机制[24]是由 Wang 等人在 2020 年提出的用于集合数据频率估计的机制,具有通信代价低、估计误差小的优点.Wheel 机制用户端的扰动过程主要分为 2 步:第 1 步是将 x 通过哈希函数映射到 $v\in[0,1)$,第 2 步则是根据式(8)所示的概率密度得到扰动结果 y

$$Q(y|v)=\begin{cases}\dfrac{e^{\varepsilon}}{\Omega}, & 若\ y\in C_v,\\ \dfrac{\Omega-l\times e^{\varepsilon}}{(1-l)\Omega}, & 若\ y\notin C_v.\end{cases} \tag{8}$$

其中 $v=\{v_1,v_2,\cdots,v_m\}$ 是用户数据哈希后的结果;$C_v=\{t\,|\,t\in[v_i,v_i+p)$ 或 $[0,v_i+p-1),i\in\{1,2,\cdots,m\}\}$ 表示总体覆盖区域;参数 $p=1/(2m-1+me^{\varepsilon})$ 表示覆盖长度;l 是 C_v 的长度,即总覆盖长度;参数 $\Omega=mpe^{\varepsilon}+1-mp$ 则是正则化因子.注意扰动结果 $y\in[0,1)$,也就是输出域是 $[0,1)$.

3 针对集合数据的效用优化本地差分隐私模型

3.1 符号描述

本文中所用到的主要符号描述如表 1 所示:

Table 1 Notations Description

表 1 符号描述

符号	说明
ε	隐私预算
X	输入数据域
X_{sen}	敏感数据域
X_{non}	非敏感数据域
Y	输出数据域
Y_{pro}	保护数据域
Y_{inv}	可逆数据域
n	用户数量
d	数据域大小,$\|X\|=d$
s	用户数据,$s\subseteq X$
S	用户数据域,$s\in S$
s_{non}	用户手中的非敏感数据,$s_{non}\subseteq X_{non}$
m	用户手中数据条数,$\|s\|=m$
f_x	元素 x 的真实频率
$\hat{f}_x$	元素 x 的估计频率

3.2 效用优化本地差分隐私模型的建立

ULDP 模型虽然就数据敏感性问题进行了研究,但是仅能处理用户输入均为敏感值或均为非敏

感值的情况.当用户集合数据中既包含敏感值,又包含非敏感值时,ULDP 模型就无法直接处理了.例如统计用户购物数据时,假设某一位用户的购物记录为{香蕉,牛奶,洗衣液,心脏病药物},此时心脏病药物为敏感值而其他数据则为非敏感值,ULDP 模型就难以直接适用了.因此我们提出了 SULDP 模型,通过降低对集合中非敏感数据的保护效果,来提高统计结果的准确性.这里我们首先给出 SULDP 的形式化定义.

定义 3. $(X_{\text{sen}},Y_{\text{pro}},\varepsilon)$-SULDP.给定 $X_{\text{sen}}\subseteq X$,$Y_{\text{pro}}\subseteq Y$,$Y_{\text{inv}}\subseteq Y$,$\varepsilon\geqslant 0$,其中 X_{sen} 表示敏感值集合,X_{non} 表示非敏感值集合,且 $X_{\text{sen}}\cap X_{\text{non}}=\varnothing$,$X_{\text{sen}}\cup X_{\text{non}}=X$,对输入域 $S=\{s\mid s\subseteq X\}$,输出域为 $Y_{\text{pro}}\times Y_{\text{inv}}^{t}$ 的扰动机制 $M:S\to Y_{\text{pro}}\times Y_{\text{inv}}^{t}$,$M(s)=(y_0,y_1,\cdots,y_t)$,$t\leqslant|s_{\text{non}}|$,当且仅当 $y_i(0\leqslant i\leqslant t)$ 满足以下性质时,M 满足 $(X_{\text{sen}},Y_{\text{pro}},\varepsilon)$-SULDP.

1) 对于任意 $y_i\in Y_{\text{inv}}$,其中 $1\leqslant i\leqslant t$,有且仅有一个 $x\in X_{\text{non}}$ 可以映射到 y_i,即

当 $x\in s$ 时,

$$M(y_i\mid s)>0, \tag{9}$$

当 $x\notin s$ 时,

$$M(y_i\mid s)=0. \tag{10}$$

2) 对于任意输入集合 $s,s'\subseteq X$,得到任意输出 $y_0\in Y_{\text{pro}}$ 的概率满足

$$M(y_0\mid s)\leqslant e^{\varepsilon}M(y_0\mid s'). \tag{11}$$

在定义 3 中,式(11)保证了对敏感值的保护程度不会降低;式(9)和式(10)则对应非敏感值,允许通过降低对其保护效果来提高统计结果的准确性.同时,如果限定每个输入集合大小为 1 时,即 $|s|=1$,那么 SULDP 将退化到 ULDP 模型.因此,ULDP 模型是我们 SULDP 模型的一种特殊情况.SULDP 模型是 ULDP 模型在集合数据上的理论拓展,具有更广泛的适用性.

3.3 问题定义

本文专注于在保证用户敏感数据保护效果的前提下,提升频率估计结果准确性.我们考虑的系统模型中包含 n 个用户和 1 个服务器.其中,每个用户 i 持有一个私有的集合 $s^i=\{x_1,x_2,\cdots,x_m\}$,$s\subseteq X$,即 X 为原始数据的全集,令 $|X|=d$.我们假设 X 中既包含敏感值,又包含非敏感值,所有敏感值构成的集合记为 X_{sen},所有非敏感值构成的集合记为 X_{non},$X_{\text{sen}}\cup X_{\text{non}}=X$.

服务器期望尽可能准确地估计出 X 中每个数据 x 的真实频率 $f_x=|\{s^i\mid i\in[1,n],x\in s^i\}|/n$.同时,为了实现对 s^i 中敏感数据的有效保护,用户 i 需要根据 s^i 中的数据敏感与否将其分成 s^i_{sen} 和 s^i_{non},然后使用不同的方式进行扰动,使得扰动结果满足 SULDP 模型,并将扰动结果 $(y_0,y_1,\cdots,y_t)$ 发送给服务器.在收集到扰动后数据后,服务器计算得出 X 中所有数据 x 所对应的估计频率 $\hat{f}_x$.我们依据估计结果 $\hat{f}_x$ 与真实频率 f_x 之间的偏离程度,来评估估计结果准确性.

定义 3 中我们假设用户手中数据条数是一定的,但通过简单扩展可以处理用户手中数据条数不同的情况.即通过填充采样[29]的方式,将用户数据条数统一为 m.具体扩展方式为:首先由服务器根据前期调研或实际情况指定 m.然后由用户在本地对自己数据进行预处理.若数据条数小于 m,则使用虚假数据补齐到 m 条;若数据条数大于 m,则从中随机抽取 m 条,然后再对数据进行后续处理.

3.4 效用评价标准

本文采用 *MSE* 对机制的效用进行评估.*MSE* 表达式为

$$MSE[\hat{f}_x]=Var[\hat{f}_x]+(E[\hat{f}_x]-f_x)^2. \tag{12}$$

MSE 可以看成是方差和偏差平方的和,当 $\hat{f}(x)$ 是无偏估计时,偏差为 0,*MSE* 就等于方差,且 *MSE* 越小,估计值越准确,效果也就越好.

4 机制设计及理论分析

4.1 基于 GRR 的机制设计

4.1.1 方案描述

我们首先基于 GRR 提出满足 SULDP 模型的 suGRR(set-valued utility-optimized GRR)机制.在 suGRR 中,保护数据域 Y_{pro} 是敏感数据域 X_{sen} 子集构成的集合,可逆数据域 Y_{inv} 与非敏感数据域 X_{non} 相等,令 $p=e^{\varepsilon/m}/(e^{\varepsilon/m}+|X_{\text{sen}}|-1)$,$q=1/(e^{\varepsilon/m}+|X_{\text{sen}}|-1)$,$r=(e^{\varepsilon/m}-1)/(e^{\varepsilon/m}+|X_{\text{sen}}|-1)$,给定集合数据 s,则 suGRR 以表达式(13)所示的概率对 s 中每一条数据 x 进行扰动,然后将属于 X_{sen} 的输出 z 看作一个集合,即为 suGRR 输出部分的 y_0,余下属于 X_{non} 的输出 z 则分别对应 $y_1,y_2,\cdots,y_t$,

$$Pr(z\mid x)=\begin{cases}p, & \text{若 } x\in X_{\text{sen}},z=x,\\ q, & \text{若 } x\in X_{\text{sen}},z\in X_{\text{sen}}/\{x\},\\ q, & \text{若 } x\in X_{\text{non}},z\in X_{\text{sen}},\\ r, & \text{若 } x\in X_{\text{non}},z=x.\end{cases} \tag{13}$$

相应地,当服务器收到用户发来的数据后,首先

需要对 X 中所有数据出现的次数进行统计,设 x 出现次数为 F_x,然后再以表达式(14)所示的方法计算出其频率估计值,

$$\hat{f}_x=\begin{cases}\dfrac{F_x/n-mq}{p-q}, & 若\ x\in X_{sen},\\ \dfrac{F_x}{nr}, & 若\ x\in X_{non}.\end{cases}\tag{14}$$

不难发现,suGRR 本质上是对隐私预算 ε 进行了划分,当 m 增大时,分配给每项数据的隐私预算会很小,使得误差变大.因此我们基于抽样的思想又提出了 suGRR-Sample,它的扰动和估计过程与 suGRR 类似,只不过每个用户只抽出一条数据进行扰动并提交,所以 $p=e^{\varepsilon}/(e^{\varepsilon}+|X_{sen}|-1)$,$q=1/(e^{\varepsilon}+|X_{sen}|-1)$,$r=(e^{\varepsilon}-1)/(e^{\varepsilon}+|X_{sen}|-1)$,相应地,服务器估计频率的公式也需要改变成式(15):

$$\hat{f}_x=\begin{cases}\dfrac{F_x/n-q}{p-q}\times m, & 若\ x\in X_{sen},\\ \dfrac{mF_x}{nr}, & 若\ x\in X_{non}.\end{cases}\tag{15}$$

4.1.2 理论分析

定理 1. suGRR 和 suGRR-Sample 均符合 SULDP 模型.

证明. suGRR 中,对任意 $y\in Y_{inv}$,有且仅有一个 $x\in X_{non}$ 可扰动至该可逆数据,即当且仅当 $x\in s$ 时,以 r 的概率输出该可逆数据,因此满足 SULDP 定义中式(9)、式(10)所规定的第 1 条性质.

对于任意 $s,s'\subseteq X$,输出同一结果 y_0 的概率为

$$\frac{M(y_0|s)}{M(y_0|s')}=\prod_{x\in s}\frac{Pr(y|x)}{Pr(y|x')}\leqslant$$

$$\prod_{x\in s}\frac{e^{\varepsilon/m}/(e^{\varepsilon/m}+|X_{sen}|-1)}{1/(e^{\varepsilon/m}+|X_{sen}|-1)}=(e^{\varepsilon/m})^m=e^{\varepsilon}.$$

因此满足式(11)所规定的第 2 条性质.

综上,suGRR 符合 SULDP 模型,同理,我们也可以证明 suGRR-Sample 符合 SULDP 模型. 证毕.

定理 2. suGRR 和 suGRR-Sample 频率估计结果均为无偏估计.

证明. 在 suGRR 中,$x\in X_{sen}$ 时,F_x 可以看作是 $nf(x)$ 个成功概率为 p 和 $n(m-f(x))$ 个成功概率为 q 的伯努利变量之和,因此,

$$E[\hat{f}_x]=E\left[\frac{F_x/n-mq}{p-q}\right]=$$

$$\frac{(npf(x)+nq(m-f(x)))/n-mq}{p-q}=f_x.$$

$x\in X_{non}$ 时,由扰动过程可知

$$E[\hat{f}_x]=E\left[\frac{F_x}{nr}\right]=\frac{nrf(x)}{nr}=f_x.$$

因此,suGRR 的频率估计结果为无偏估计,同理,suGRR-Sample 的频率估计结果也是无偏估计.

证毕.

定理 3. suGRR 的 MSE 表达式如式(16):

$$MSE[\hat{f}_x]=$$

$$\begin{cases}\dfrac{m(e^{\varepsilon/m}+|X_{sen}|-2)}{n(e^{\varepsilon/m}-1)^2}+\dfrac{|X_{sen}|-2}{n(e^{\varepsilon/m}-1)}\times f_x,\\ 若\ x\in X_{sen},\\ \dfrac{|X_{sen}|}{n(e^{\varepsilon/m}-1)}\times f_x, 若\ x\in X_{non}.\end{cases}\tag{16}$$

suGRR-Sample 的 MSE 表达式如式(17):

$$MSE[\hat{f}_x]=$$

$$\begin{cases}\dfrac{m^2(e^{\varepsilon}+|X_{sen}|-2)}{n(e^{\varepsilon}-1)^2}+\dfrac{m^2(|X_{sen}|-2)}{n(e^{\varepsilon}-1)}\times f_x,\\ 若\ x\in X_{sen},\\ \dfrac{m^2|X_{sen}|}{n(e^{\varepsilon}-1)}\times f_x, 若\ x\in X_{non}.\end{cases}\tag{17}$$

证明. 由定理 2 可得,式(14)为无偏估计,因此 MSE 就等于方差,当 $x\in X_{sen}$ 时,

$$MSE[\hat{f}_x]=Var[\hat{f}_x]=Var\left[\frac{F_x/n-mq}{p-q}\right]=$$

$$\frac{np(1-p)f_x+n(m-f_x)q(1-q)}{n^2(p-q)^2}.$$

当 $x\in X_{non}$ 时,

$$MSE[\hat{f}_x]=Var[\hat{f}_x]=Var\left[\frac{F_x}{nr}\right]=\frac{nr(1-r)f_x}{(nr)^2}.$$

将 p,q,r 带入,即可得到式(16).同理,也可以证明 suGRR-Sample 的 MSE 表达式为式(17).

证毕.

4.2 基于 RAPPOR 的机制设计

4.2.1 方案描述

基于 GRR 的扰动过程虽然简单、易于实现,但是当数据域很大时,p,q,r 会趋向于 0,会使得数据扰动概率过大,数据效用降低,而 RAPPOR 则不会受此影响.因此,基于 RAPPOR,我们提出满足 SULDP 模型的 suRAP(set-valued utility-optimized RAPPOR)机制.

在 suRAP 中,首先会对数据进行独热编码,即将用户手中的集合数据 s 编码为一个 d 比特的向量 $\boldsymbol{a}$,每一位都与原始数据域中的一条数据相对应,若 s 中包含有 x_i,则令 $\boldsymbol{a}$ 的第 i 位为 1,即 $\boldsymbol{a}$ 中有 m 个 1,其余位为 0.

不失一般性,我们假设$\{x_1,x_2,\cdots,x_{|X_{sen}|}\}$为敏感数据$X_{sen}$,$\{x_{|X_{sen}|+1},x_{|X_{sen}|+2},\cdots,x_{|X|}\}$为非敏感数据$X_{non}$,则$Y_{pro}=\{(y_1,y_2,\cdots,y_{|X|})\mid y_1,y_2,\cdots,y_{|X|}\in\{0,1\}\}$,$Y_{inv}=X_{non}$.令$p=e^{\varepsilon/2m}/(e^{\varepsilon/2m}+1)$,$q=1/(e^{\varepsilon/2m}+1)$,$r=(e^{\varepsilon/2m}-1)/e^{\varepsilon/2m}$,则suRAP以表达式(18)所示的概率对$\boldsymbol{a}$中每一位$a_i$进行扰动得到$b_i$,

$$Pr(b_i\mid a_i)=\begin{cases}p,\text{ 若 }a_i=1,b_i=1,x_i\in X_{sen},\\1-p,\text{ 若 }a_i=1,b_i=0,x_i\in X_{sen},\\q,\text{ 若 }a_i=0,b_i=1,x_i\in X_{sen},\\1-q,\text{ 若 }a_i=0,b_i=0,x_i\in X_{sen},\\r,\text{ 若 }a_i=1,b_i=1,x_i\in X_{non},\\1-r,\text{ 若 }a_i=1,b_i=0,x_i\in X_{non},\\1,\text{ 若 }a_i=0,b_i=0,x_i\in X_{non},\\0,\text{ 若 }a_i=0,b_i=1,x_i\in X_{non}.\end{cases}\tag{18}$$

则y_0就等于向量$\boldsymbol{b}$,同时,若$x_i\in X_{non}$且$b_i=1$,则表示输出了可逆数据x_i,即我们可以直接使用$\boldsymbol{b}$来表示$(y_0,y_1,\cdots,y_t)$,$t\leqslant|s_{non}|$.

服务器端收到用户发来的数据后,首先需要对$\boldsymbol{b}$中所有位出现1的次数进行统计,设x对应位出现1的次数为F_x,则其估计频率为

$$\hat{f}_x=\begin{cases}\dfrac{F_x/n-q}{p-q},\text{ 若 }x\in X_{sen},\\\dfrac{F_x}{nr},\text{ 若 }x\in X_{non}.\end{cases}\tag{19}$$

类似地,为了避免划分隐私预算ε,我们也基于采样提出了suRAP-Sample,此时$p=e^{\varepsilon/2}/(e^{\varepsilon/2}+1)$,$q=1/(e^{\varepsilon/2}+1)$,$r=(e^{\varepsilon/2}-1)/e^{\varepsilon/2}$,用户从$s$中随机抽取一条数据并编码为$\boldsymbol{a}$,注意这时$\boldsymbol{a}$中只有一位为1,然后同样使用式(18)进行扰动并提交结果给服务器,服务器统计出x对应位出现1的次数F_x后,即可根据式(20)计算出x的估计频率,

$$\hat{f}_x=\begin{cases}\dfrac{F_x/n-q}{p-q}\times m,\text{ 若 }x\in X_{sen},\\\dfrac{mF_x}{nr},\text{ 若 }x\in X_{non}.\end{cases}\tag{20}$$

4.2.2 理论分析

定理4. suRAP和suRAP-Sample均符合SULDP模型.

证明. suRAP中,对于任意$s,s'\subseteq X$,输出同一结果y_0的概率为

$$\frac{M(y_0\mid s)}{M(y_0\mid s')}=\prod_{i\in d}\frac{Pr(b_i\mid a_i)}{Pr(b_i\mid a'_i)}\leqslant\prod_{i\in m}\frac{p}{q}\times\frac{1}{1-r}=\left(\frac{e^{\varepsilon/2m}/(e^{\varepsilon/2m}+1)}{1/(e^{\varepsilon/2m}+1)}\times\frac{1}{1/e^{\varepsilon/2m}}\right)^m=e^{\varepsilon}.$$

因此满足式(11)所规定的第2条性质.

对任意$y\in Y_{inv}$,有且仅有一个$x\in X_{non}$可扰动至该可逆数据,即当且仅当$x\in s$时,以r的概率输出该可逆数据,因此满足SULDP定义中式(9)、式(10)所规定的第1条性质.

综上,suRAP符合SULDP模型,同理,我们也可以证明suRAP-Sample符合SULDP模型. 证毕.

定理5. suRAP和suRAP-Sample频率估计结果均为无偏估计.

证明. 当$x\in X_{sen}$时,有

$$E[\hat{f}_x]=E\left[\frac{F_x/n-q}{p-q}\right]=\frac{pf(x)+q(1-f(x))-q}{p-q}=f_x.$$

$x\in X_{non}$时,由扰动过程可知:

$$E[\hat{f}_x]=E\left[\frac{F_x}{nr}\right]=\frac{nrf(x)}{nr}=f_x.$$

综上,suRAP的频率估计结果为无偏估计,同理,suRAP-Sample的频率估计结果也是无偏估计. 证毕.

定理6. suRAP的MSE表达式如式(21):

$$MSE[\hat{f}_x]=\begin{cases}\dfrac{e^{\varepsilon/2m}}{n(e^{\varepsilon/2m}-1)^2},\text{ 若 }x\in X_{sen},\\\dfrac{1}{n(e^{\varepsilon/2m}-1)}\times f_x,\text{ 若 }x\in X_{non}.\end{cases}\tag{21}$$

suRAP-Sample的MSE表达式如式(22):

$$MSE[\hat{f}_x]=\begin{cases}\dfrac{m^2e^{\varepsilon/2}}{n(e^{\varepsilon/2}-1)^2},\text{ 若 }x\in X_{sen},\\\dfrac{m^2}{n(e^{\varepsilon/2}-1)}\times f_x,\text{ 若 }x\in X_{non}.\end{cases}\tag{22}$$

证明. 与定理3证明过程类似.

4.3 基于Wheel的机制设计

4.3.1 方案描述

虽然基于RAPPOR处理集合数据相较于基于GRR的效果有所提升,但是其通信代价很高,并且效果也并非目前最优的,因此我们在本节提出满足SULDP模型的suWheel(set-valued utility-optimized wheel),具体的扰动算法如算法1所示.

算法1. suWheel用户端扰动算法.

输入:集合数据$s=\{x_1,x_2,\cdots,x_m\}$、敏感数据域X_{sen}、非敏感数据域X_{non}、隐私预算ε、用户数据条数m、覆盖长度$p=1/(2m-1+me^{\varepsilon})$、正则化因子$\Omega=mpe^{\varepsilon}+1-mp$、哈希函数$h:X\to[0.0,1.0)$;

输出:三元组 $z=\langle z_0,z_1,h\rangle$,其中 $z_0=y_0$ 表示保护数据,$z_1=\{y_1,y_2,\cdots,y_t\}(t\leqslant|s_{\text{non}}|)$表示可逆数据,$h$ 为该用户使用的哈希函数.

① $v=\{h(x_1),h(x_2),\cdots,h(x_m)\}=\{v_1,v_2,\cdots,v_m\}$;/* 将原始数据映射到[0.0,1.0)上 */

② $C_v=\{t\mid t\in[v_i,v_i+p)$或$[0,v_i+p-1),i\in\{1,2,\cdots,m\}\}$;

③ $Q(y_0\mid v)=\begin{cases}e^{\varepsilon}/\Omega, & y\in C_v,\\ (\Omega-l\times e^{\varepsilon})/((1-l)\Omega), & y\notin C_v,\end{cases}$ 以 $Q(y_0\mid v)$所示概率密度得到扰动结果 y_0,其中 l 是 C_v 的长度,令 $z_0=y_0$;

④ $z_1=\varnothing$;

⑤ for $i=1$ to m:

⑥ if $x_i\in X_{\text{non}}$:

⑦ if $y_0\notin[v_i,v_i+p)$且$[0,v_i+p-1)$

⑧ add x_i into z_1;

⑨ end if

⑩ end if

⑪ end for

⑫ 输出 $z=\langle z_0,z_1,h\rangle$.

其中 z_0 对应的是 y_0,z_1 则对应 $y_i(1\leqslant i\leqslant t)$. 令 C_{v_x} 为元素 x 对应的覆盖区域,当 $x\in s$ 时,输出的 z_0 属于该区域的概率为 $P_{\text{true}}=pe^{\varepsilon}/(mpe^{\varepsilon}+1-mp)$,当 $x\notin s$ 时,概率则为 $P_{\text{false}}=p$.

算法 2. suWheel 服务器端聚合算法.

输入:敏感数据域 X_{sen}、非敏感数据域 X_{non}、隐私预算 ε、用户手中数据条数 m、覆盖长度 $p=1/(2m-1+me^{\varepsilon})$、$n$ 个用户的扰动结果$\{z^1,z^2,\cdots,z^n\}$,其中 $z^i=\langle z_0^i,z_1^i,h^i\rangle$;

输出:任意 $x(x\in X)$的频率的估计结果 $\hat{f}_x$.

① $F_x=0$;

② if $x\in X_{\text{sen}}$

③ for $i=1$ to n

④ $v=h_i(x)$;

⑤ if $z_0^i-p<v\leqslant z_0^i$ 或 $z_0^i-p+1<v<1$

⑥ $F_x=F_x+1$;

⑦ end if

⑧ end for

⑨ $\hat{f}_x=\dfrac{F_x/n-P_{\text{false}}}{P_{\text{true}}-P_{\text{false}}}$;

⑩ end if

⑪ if $x\in X_{\text{non}}$

⑫ for $i=1$ to n

⑬ if $x\in z_1^i$

⑭ $F_x=F_x+1$;

⑮ end if

⑯ end for

⑰ $\hat{f}_x=F_x\Big/\left(n\left(1-\dfrac{e^{\varepsilon}p}{\Omega}\right)\right)$

⑱ end if

⑲ 输出 $\hat{f}_x$.

4.3.2 理论分析

定理 7. suWheel 符合 SULDP 模型.

证明. suWheel 中,对任意 $y\in Y_{\text{inv}}$,有且仅有一个 $x\in X_{\text{non}}$可扰动至该可逆数据,即当且仅当 $x\in s$ 时,以 $r=1-e^{\varepsilon}p/\Omega$ 的概率输出该可逆数据,因此满足 SULDP 定义中式(9)、式(10)所规定的第 1 条性质.

对于任意集合 $s,s'\subseteq X$,输出同一结果 y_0 的概率为

$$\frac{M(y_0\mid s)}{M(y_0\mid s')}=\frac{e^{\varepsilon}/\Omega}{(\Omega-l\times e^{\varepsilon})/((1-l)\Omega)}=\frac{e^{\varepsilon}(1-l)}{\Omega-l\times e^{\varepsilon}},$$

令

$$\frac{e^{\varepsilon}(1-l)}{\Omega-l\times e^{\varepsilon}}\geqslant\frac{1-l}{\Omega-l\times e^{\varepsilon}}\geqslant e^{-\varepsilon},\tag{23}$$

化简得 $e^{\varepsilon}(1-mp)\geqslant 1-mp$,$mp=\dfrac{1}{2-1/m+e^{\varepsilon}}<1$,所以式(23)恒成立.令

$$\frac{e^{\varepsilon}(1-l)}{\Omega-l\times e^{\varepsilon}}\leqslant e^{\varepsilon},\tag{24}$$

化简得 $l(e^{\varepsilon}-1)\leqslant mp(e^{\varepsilon}-1)$,因为 l 为覆盖区域的总长度,并且各个元素覆盖区域之间可能会存在相交的情况,因此 $l\leqslant mp$,又有 $\varepsilon\geqslant 0$,所以式(24)恒成立.即满足 SULDP 定义中式(11)所规定的第 2 条性质.

综上,suWheel 符合 SULDP 模型. 证毕.

定理 8. suWheel 频率估计结果为无偏估计.

证明. 当 $x\in X_{\text{sen}}$时,F_x 可以看作是 $nf(x)$个成功概率为 P_{true}和 $n(1-f(x))$个成功概率为 P_{false}的伯努利变量之和,因此

$$E[\hat{f}_x]=E\left[\frac{F_x/n-P_{\text{false}}}{P_{\text{true}}-P_{\text{false}}}\right]=\frac{P_{\text{true}}f(x)+P_{\text{false}}(1-f(x))-P_{\text{false}}}{P_{\text{true}}-P_{\text{false}}}=f_x.$$

$x\in X_{\text{non}}$时,由扰动过程可知,输出可逆数据 x 的概率 $r=1-\dfrac{e^{\varepsilon}p}{\Omega}$,因此

$$E[\hat{f}_x]=E\left[\frac{F_x}{nr}\right]=\frac{nrf(x)}{nr}=f_x.$$

综上,suWheel 的频率估计结果为无偏估计.

证毕.

定理 9. 当 $\varepsilon=O(1)$ 时,suWheel 的 *MSE* 表达式如式(25):

$$MSE[\hat{f}_x]=\begin{cases}O\left(\dfrac{(3m-1)^2(3m-2)}{n(e^{\varepsilon}-1)^2(me^{\varepsilon}+m-1)^2}\right),\\ \text{若 } x\in X_{\text{sen}},\\ \dfrac{e^{\varepsilon}}{(2me^{\varepsilon}+m-e^{\varepsilon}-1)n}\times f_x,\ \text{若 } x\in X_{\text{non}}.\end{cases}\tag{25}$$

证明. 因为算法 2 中的频率估计为无偏估计,因此$(E[\hat{f}_x]-f_x)^2=0$,此时 *MSE* 就等于方差.即,当 $x\in X_{\text{sen}}$ 时,

$$MSE[\hat{f}_x]=Var[\hat{f}_x]=Var\left[\frac{F_x/n-P_{\text{false}}}{P_{\text{true}}-P_{\text{false}}}\right]=\frac{P_{\text{true}}(1-P_{\text{true}})f_x+(1-f_x)P_{\text{false}}(1-P_{\text{false}})}{n(P_{\text{true}}-P_{\text{false}})^2}=O\left(\frac{(3m-1)^2(3m-2)}{n(e^{\varepsilon}-1)^2(me^{\varepsilon}+m-1)^2}\right).$$

同样地,当 $x\in X_{\text{non}}$ 时,非敏感数据保持不变的概率 $r=1-\dfrac{e^{\varepsilon}p}{\Omega}$,此时

$$MSE[\hat{f}_x]=Var[\hat{f}_x]=Var\left[\frac{F_x}{nr}\right]=\frac{nr(1-r)f_x}{(nr)^2}=\frac{e^{\varepsilon}}{(2me^{\varepsilon}+m-e^{\varepsilon}-1)n}\times f_x.$$

证毕.

4.4 理论结果对比

我们首先对 suGRR,suGRR-Sample,suRAP,suRAP-Sample,suWheel 等 5 种机制的效用进行评估.因为

$$MSE[\hat{f}_x]=\sum_{x\in X_{\text{sen}}}MSE[\hat{f}_x]+\sum_{x\in X_{\text{non}}}MSE[\hat{f}_x],$$

根据定理 3、定理 6 和定理 9,我们可以计算得到 $\varepsilon=O(1)$ 时本文机制与现有本地差分隐私机制分别所对应的 *MSE*,结果如表 2 所示.

$f(X_{\text{non}})$ 表示非敏感数据的真实频率总和,因此 *MSE* 的前半部分,即敏感数据造成的误差在实际应用中占主导地位.而在实际应用中,敏感数据在总体数据中的占比往往很小,即 $|X_{\text{sen}}|<d$,因此本文机制的 *MSE* 均要小于现有机制.

同时计算可得,当 $\varepsilon<m^2$ 时,suWheel 的 *MSE* 一定是最小的,并且 suWheel 的敏感数据部分的 *MSE* 为 $O\left(\dfrac{m|X_{\text{sen}}|}{n\varepsilon^2}\right)$,也是最低的.因此,suWheel 在数据效用方面是 5 种机制中最优的.

Table 2 Comparison of *MSE* when $\varepsilon=O(1)$

表 2 $\varepsilon=O(1)$ 时针对 *MSE* 的对比

机制	*MSE*
GRR	$O\left(\dfrac{m^3d^2}{n\varepsilon^2}\right)$
GRR-Sample	$O\left(\dfrac{m^2d^2}{n\varepsilon^2}\right)$
RAPPOR	$O\left(\dfrac{m^2d}{n\varepsilon^2}\right)$
RAPPOR-Sample	$O\left(\dfrac{m^2d}{n\varepsilon^2}\right)$
Wheel	$O\left(\dfrac{md}{n\varepsilon^2}\right)$
suGRR	$O\left(\dfrac{m^3\lvert X_{\text{sen}}\rvert^2}{n\varepsilon^2}+\dfrac{m\lvert X_{\text{sen}}\rvert f(X_{\text{non}})}{n\varepsilon}\right)$
suGRR-Sample	$O\left(\dfrac{m^2\lvert X_{\text{sen}}\rvert^2}{n\varepsilon^2}+\dfrac{m^2\lvert X_{\text{sen}}\rvert f(X_{\text{non}})}{n\varepsilon}\right)$
suRAP	$O\left(\dfrac{m^2\lvert X_{\text{sen}}\rvert}{n\varepsilon^2}+\dfrac{mf(X_{\text{non}})}{n\varepsilon}\right)$
suRAP-Sample	$O\left(\dfrac{m^2\lvert X_{\text{sen}}\rvert}{n\varepsilon^2}+\dfrac{m^2f(X_{\text{non}})}{n\varepsilon}\right)$
suWheel	$O\left(\dfrac{m\lvert X_{\text{sen}}\rvert}{n\varepsilon^2}+\dfrac{f(X_{\text{non}})}{mn}\right)$

除了数据效用,通信代价也是评价一个机制好坏与否的重要标准,假设 suWheel 中使用的哈希函数是从集合 H 中选取的,表 3 给出了相应的结果.可以看出 suWheel 虽然比 suGRR-Sample 略高,但是仍然是可以接受的.

Table 3 Comparison of Comunication Cost

表 3 通信代价的对比

机制	通信代价
GRR	$O(m\log d)$
GRR-Sample	$O(\log d)$
RAPPOR	$O(d)$
RAPPOR-Sample	$O(d)$
Wheel	$O(\log(me^{\varepsilon})+\log\lvert H\rvert)$
suGRR	$O(m\log d)$
suGRR-Sample	$O(\log d)$
suRAP	$O(d)$
suRAP-Sample	$O(d)$
suWheel	$O(\log(me^{\varepsilon})+t\log d+\log\lvert H\rvert)$

5 实 验

5.1 实验设置

我们的实验环境设置如下:操作系统为 Windows

10,处理器为 Inter i7-6700 CPU,内存为 32 GB,实验所用的语言为 Python 3.6.

我们在 4 个数据集上进行了实验,表 4 展示了它们的参数信息.销售数据集[30]包含了英国一家商店一年的在线交易信息,该商店主要销售成人、儿童和家居用品,我们将敏感数据设置为儿童用品;动漫数据集[31]则描述了用户对一些动漫的评分,我们将每位用户评分的动漫作为一条集合数据,并将类别为成人、惊悚、恐怖的动漫作为敏感数据;对电影数据集[32]也是类似的处理.

Table 4 Description of Dataset Parameters

表 4 数据集参数描述

数据集	n	d	$\lvert X_{sen}\rvert$	m
销售数据集[30]	19 789	3 752	201	54
动漫数据集[31]	71 662	12 994	1 748	231
电影数据集[32]	5 641	3 883	799	300
模拟数据集	100 000	256	64	8

我们是将常规意义上可能会泄露用户隐私的数据设置为敏感值.比如销售数据集中儿童用品会泄露用户孩子的性别和年龄;动漫数据集和电影数据集中给成人、惊悚等类别的动漫、电影评价的信息则会泄露用户的观影偏好,而针对此类特殊类别的影视的喜好信息对于用户而言往往也是较为敏感的.在实际应用中,针对敏感值和非敏感值的划分,可以由服务器根据常识和专业知识决定,同时根据用户的使用反馈来进行补充,也可以通过问卷调查等调研方式收集各个用户认为的敏感数据,然后由服务器取并集作为敏感值划分结果.

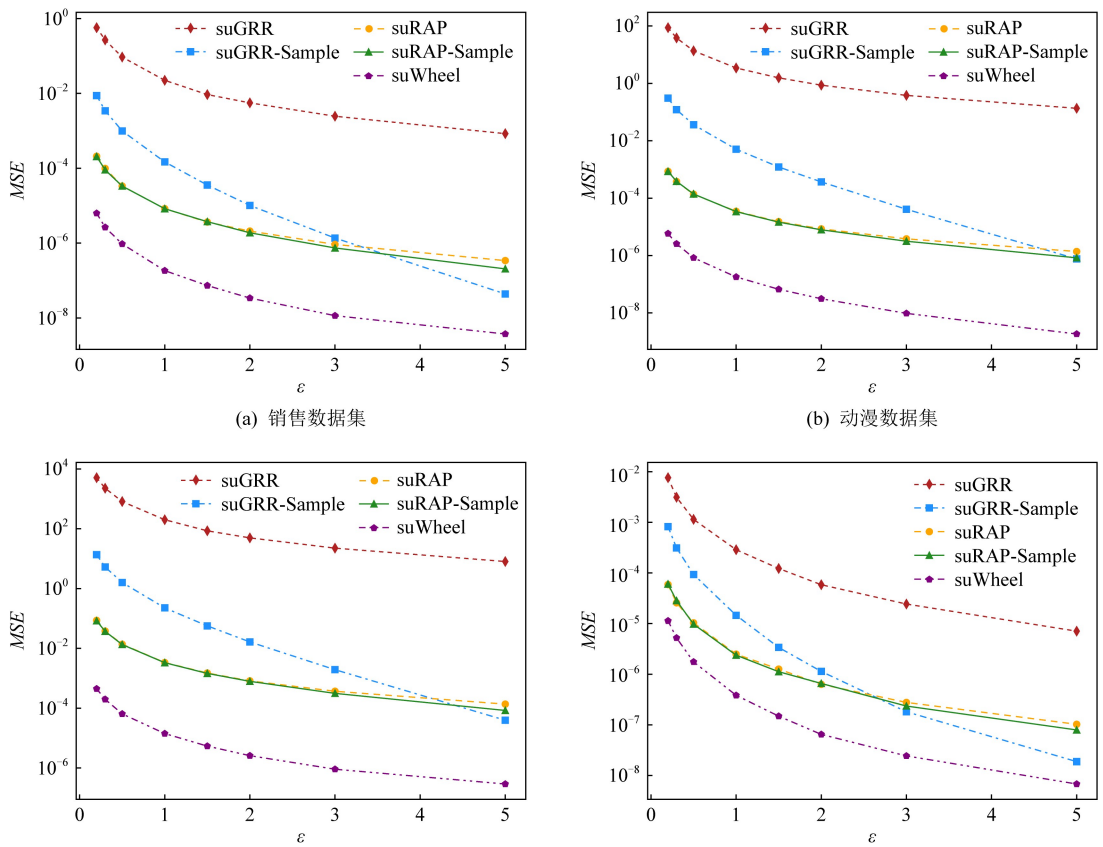
此外,我们所使用的 3 个真实数据集都是不定长的,因此需要对它们进行预处理,通过填充采样的方法将之转换为定长数据.同时,我们还使用蓄水池抽样的方法生成了模拟数据集.

实验采用 *MSE* 作为评估标准,用以评价频率估计准确性.同时,为避免随机性影响实验结果,本文中每一项实验重复进行 10 次,并取平均值作为结果.

5.2 实验结果

5.2.1 ε 对 *MSE* 的影响

本节对比了 5 种机制在不同隐私预算 ε 下的性能差异,实验结果如图 1 所示.可以看出,随着隐私



(a) 销售数据集 (b) 动漫数据集

(c) 电影数据集 (d) 模拟数据集

Fig. 1 Effect of ε on *MSE*

图 1 ε 对 *MSE* 的影响

预算 ε 的增大,5 种机制的 *MSE* 都在减小,相应地,频率估计结果会更加准确,数据效用也就越高.即 ε 越大,隐私保护程度越低,*MSE* 越小,数据效用越高.

同时,基于 GRR 的 2 个机制的效用要明显低于另外 3 个,这是因为实验中数据域都很大,suGRR 和 suGRR-Sample 中数据被扰动的概率会很大,频率估计效果就会很差,这也与 GRR 的特性一致.

除此之外,文献[21]提出了基于 ULDP 的 uRR 和 uRAP 机制,并建议通过划分隐私预算来实现对集合元素的频率估计需求,划分隐私预算的 uRR 实际上就等同于本文所提出的 suRR,而 suRAP 则是基于改进的 RAPPOR[22] 实现的,效果比划分隐私预算的 uRAP 要更好.因此,无论是与本文所提出的 4 种机制比较,还是与文献[21]比较,suWheel 都是表现最优的机制.

5.2.2 $|X_{sen}|$ 对 *MSE* 的影响

本节通过随机抽样的方法选取敏感数据,来调整敏感数据域在全体数据域中的占比,进而评估 $|X_{sen}|$ 对 *MSE* 的影响.我们将隐私预算 ε 固定为 1,结果如图 2 所示.可以看出,随着 $|X_{sen}|$ 的增大,5 种机制的 *MSE* 都会增大,即敏感数据越多,需要的扰动就越多,估计结果的准确度就越低,数据的整体效用就会越差.

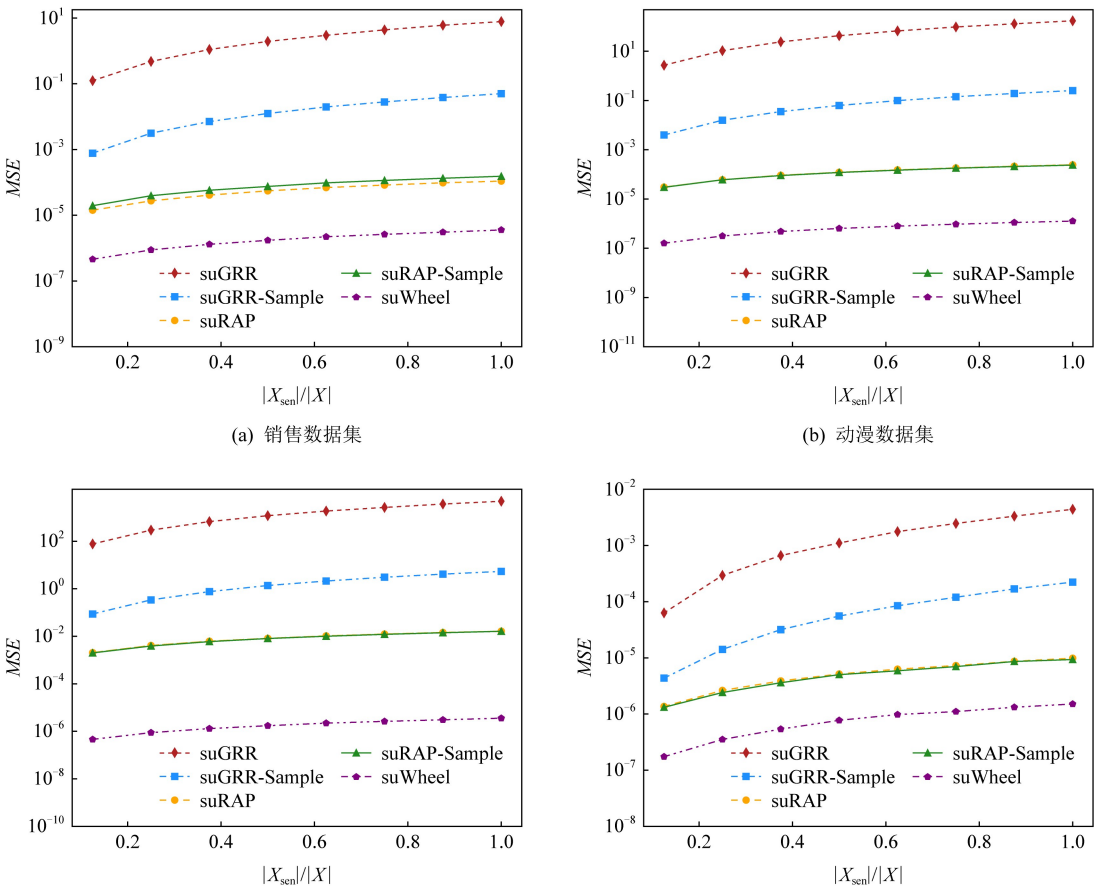


Fig. 2 Effect of the percentage of $|X_{sen}|$ in $|X|$ on *MSE*

图 2 $|X_{sen}|$ 在 $|X|$ 中的占比对 *MSE* 的影响

此外,当 $|X_{sen}|/|X|=1$,即用户数据均为敏感值时,等同于直接使用原始的 GRR,RAPPOR 等机制进行处理,此时效果是最差的.这也证明了通过对数据进行划分,然后根据敏感与否加以不同的处理方式,确实可以在一定程度提高整体数据效用.

5.2.3 d 对 *MSE* 的影响

数据域大小 d 也对数据效用有一定的影响.由于真实数据集的数据域是固定的,因此本节通过不同大小的模拟数据集来进行评估.实验设置的 d 的取值范围为{16,32,64,…,1 024},并且敏感数据占

比为 0.25,隐私预算 $\varepsilon=1$,$m=8$,结果如图 3 所示:

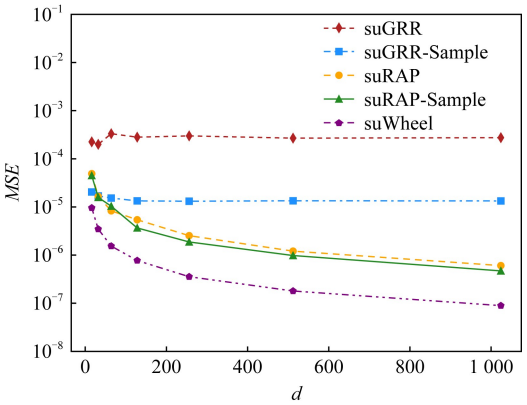


Fig. 3 Effect of d on MSE

图 3 d 对 MSE 的影响

一方面,因为是通过蓄水池抽样生成模拟数据,因此数据服从均匀分布,d 越大,数据对应的真实频率就越低,估计频率偏离程度一定时,$E[\hat{f}_x-f_x]$ 的值就会越小,即 MSE 就会越低.另一方面,根据 4.4 节的结果可知,5 种机制的敏感数据部分的 MSE 都随着 $|X_{sen}|$ 的增大而变大,即随着 d 的增大而变大.

根据实验结果来看,随着 d 的增大,suGRR 和 suGRR-Sample 的 MSE 呈现不变的趋势,这是因为基于 GRR 的 2 种方法的敏感数据部分的 MSE 与 $|X_{sen}|^2$ 正相关,这使得 2 方面的影响相互抵消了.而其余 3 种则是与 $|X_{sen}|$ 成正相关,因此前者的影响占了主导地位,MSE 呈下降趋势.

5.2.4 m 对 MSE 的影响

本节则主要评估了 m 对 MSE 的影响,同样是通过构造不同模拟数据集来进行实验.令敏感数据占比为 0.25,隐私预算 $\varepsilon=1$,数据域大小 $d=256$,设置 m 的取值范围为{1,2,4,8,…,64}.

与 5.2.3 节的分析类似,m 对 MSE 的影响也包括 2 方面:1)其余参数一定,m 越大,每条数据对应的真实频率就越低,相应的 MSE 也就越大;2)如 4.4 节所示,suWheel 的敏感部分的 MSE 与 m 呈正相关,其余 4 种机制更是与 m^2 呈正相关,因此 m 越大,MSE 也越大.

实验结果如图 4 所示,随着 m 的增大,5 种机制的数据效用都会变低,并且 suWheel 的 MSE 增大的幅度要明显小于另外 4 个机制.

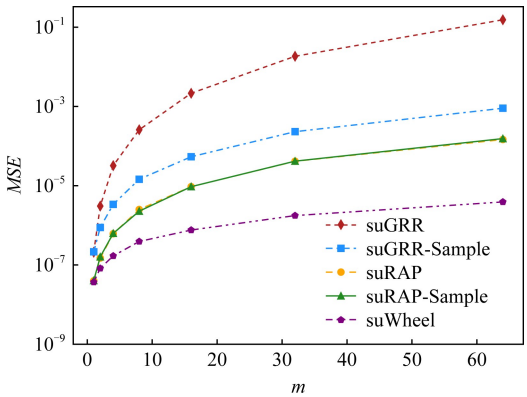


Fig. 4 Effect of m on MSE

图 4 m 对 MSE 的影响

6 结 论

现有本地差分隐私集合数据频率估计机制在设计时并未考虑数据的敏感性差异,本文针对这一问题,首先定义了针对集合数据的效用优化本地差分隐私模型,并提出了符合 SULDP 模型的 suGRR,suGRR-Sample,suRAP,suRAP-Sample 和 suWheel 机制,在保证敏感数据隐私保护效果不降低的前提下,提高了整体的数据效用.从理论和实验的结果来看,suWheel 具有低通信代价和高数据效用的特点,是整体表现最优的机制.

未来的工作包括 2 个方面:1)研究使用本文所提的思想,对本地差分隐私下的其他机制,如均值估计机制等进行改进;2)当前工作只是将数据分为了敏感和非敏感 2 种类型,那么可以通过对敏感程度进一步细分来达到更高的数据效用.

作者贡献声明:曹依然负责完成实验并撰写论文;朱友文提出了算法思路和实验方案;贺星宇和张跃协助设计了实验方案和对比分析.

参 考 文 献

[1] Wang Chenxu, Cheng Jiacheng, Sang Xinxin, et al. Data privacy-preserving for blockchain: State of the art and trends [J]. Journal of Computer Research and Development, 2021, 58(10): 2099-2119 (in Chinese)
(王晨旭,程加成,桑新欣,等.区块链数据隐私保护:研究现状与展望[J].计算机研究与发展,2021,58(10):2099-2119)

[2] Zhou Jun, Shen Huajie, Lin Zhongyun, et al. Research advances on privacy preserving in edge computing [J]. Journal of Computer Research and Development, 2020, 57(10): 2027-2051 (in Chinese)
(周俊, 沈华杰, 林中允, 等. 边缘计算隐私保护研究进展[J]. 计算机研究与发展, 2020, 57(10): 2027-2051)

[3] Wei Lifei, Chen Congcong, Zhang Lei, et al. Security issues and privacy preserving in machine learning [J]. Journal of Computer Research and Development, 2020, 57(10): 2066-2085 (in Chinese)
(魏立斐, 陈聪聪, 张蕾, 等. 机器学习的安全问题及隐私保护[J]. 计算机研究与发展, 2020, 57(10): 2066-2085)

[4] Guo Juanjuan, Wang Qiongxiao, Xu Xin, et al. Secure multiparty computation and application in machine learning [J]. Journal of Computer Research and Development, 2021, 58(10): 2163-2186 (in Chinese)
(郭娟娟, 王琼霄, 许新, 等. 安全多方计算及其在机器学习中的应用[J]. 计算机研究与发展, 2021, 58(10): 2163-2186)

[5] Nergiz M E, Atzori M, Saygin Y. Towards trajectory anonymization: A generalization-based approach [C] //Proc of the SIGSPATIAL ACM GIS 2008 Int Workshop on Security and Privacy in GIS and LBS. New York: ACM, 2008: 52-61

[6] Zhou Bin, Pei Jian. Preserving privacy in social networks against nei-ghborhood attacks [C] //Proc of the 24th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2008: 506-515

[7] Yao Qizhi. How to generate and exchange secrets [C] //Proc of the 27th Annual Symp on Foundations of Computer Science (SFCS 1986). Piscataway, NJ: IEEE, 1986: 162-167

[8] Gentry C. Fully homomorphic encryption using ideal lattices [C] //Proc of the 41st Annual ACM Symp on Theory of Computing. New York: ACM, 2009: 169-178

[9] Dwork C. Differential privacy: A survey of results [C] //Proc of the Int Conf on Theory and Applications of Models of Computation. Berlin: Springer, 2008: 1-19

[10] Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis [C] //Proc of the Theory of Cryptography Conf. Berlin: Springer, 2006: 265-284

[11] Duchi J C, Jordan M I, Wainwright M J. Local privacy and statistical minimax rates [C] //Proc of the 54th Annual Symp on Foundations of Computer Science. Piscataway, NJ: IEEE, 2013: 429-438

[12] Wang Di, Gaboardi M, Xu Jinhui. Empirical risk minimization in non-interactive local differential privacy revisited [C] //Proc of Advances in Neural Information Processing Systems. La Jolla, CA: NIPS, 2018: 973-982

[13] Zheng Kai, Mou Wenlong, Wang Liwei. Collect at once, use effectively: Making non-interactive locally private learning possible [C] //Proc of the Int Conf on Machine Learning. New York: PMLR, 2017: 4130-4139

[14] Banerjee S, Hegde N, Massoulié L. The price of privacy in untrusted recommender systems [J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 9(7): 1319-1331

[15] Duchi J C, Jordan M I, Wainwright M J. Minimax optimal procedures for locally private estimation [J]. Journal of the American Statistical Association, 2018, 113(521): 182-201

[16] Wang Tianhao, Blocki J, Li Ninghui, et al. Locally differentially private protocols for frequency estimation [C] //Proc of the 26th USENIX Security Symp (USENIX Security 17). Berkeley, CA: USENIX, 2017: 729-745

[17] Zhang Xiaojian, Fu Nan, Meng Xiaofeng. Towards spatial range queries under local differential privacy [J]. Journal of Computer Research and Development, 2020, 57(4): 847-858 (in Chinese)
(张啸剑, 付楠, 孟小峰. 基于本地差分隐私的空间范围查询方法[J].计算机研究与发展, 2020, 57(4): 847-858)

[18] Tang Jun, Korolova A, Bai Xiaolong, et al. Privacy loss in Apple's implementation of differential privacy on macos 10.12 [J/OL]. arXiv preprint arXiv: 1709.02753, 2017 [2022-06-10]. https://arxiv.org/pdf/1709.02753.pdf

[19] Greenberg A. Apple's 'differential privacy'is about collecting your data—but not your data [J/OL]. Wired, June, 2016 [2022-06-10]. https://www.wired.com/2016/06/apples-differential-privacy-collecting-data

[20] Erlingsson Ú, Pihur V, Korolova A. RAPPOR: Randomized aggregatable privacy-preserving ordinal response [C] //Proc of the 2014 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2014: 1054-1067

[21] Murakami T, Kawamoto Y. Utility-optimized local differential privacy mechanisms for distribution estimation [C] //Proc of the 28th USENIX Security Symp (USENIX Security 19). Berkeley, CA: USENIX, 2019: 1877-1894

[22] Qin Zhan, Yang Yin, Yu Ting, et al. Heavy hitter estimation over set-valued data with local differential privacy [C] //Proc of the 2016 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2016: 192-203

[23] Wang Shaowei, Huang Liusheng, Nie Yiwen, et al. PrivSet: Set-valued data analyses with locale differential privacy [C] //Proc of the IEEE Conf on Computer Communications (IEEE INFOCOM 2018). Piscataway, NJ: IEEE, 2018: 1088-1096

[24] Wang Shaowei, Qian Yuqiu, Du Jiachun, et al. Set-valued data publication with local privacy: Tight error bounds and efficient mechanisms [J]. Proceedings of the VLDB Endowment, 2020, 13(8): 1234-1247

[25] Wang Shaowei, Huang Liusheng, Tian Miaomiao, et al. Personalized privacy-preserving data aggregation for histogram estimation [C] //Proc of the 2015 IEEE Global Communications Conf (GLOBECOM). Piscataway, NJ: IEEE, 2015: 1-6

[26] Nie Yiwen, Yang Wei, Huang Liusheng, et al. A utility-optimized framework for personalized private histogram estimation [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 31(4): 655-669

[27] Andrés M E, Bordenabe N E, Chatzikokolakis K, et al. Geo-indistinguishability: Differential privacy for location-based systems [C] //Proc of the 2013 ACM SIGSAC Conf on Computer & Communications Security. New York: ACM, 2013: 901-914

[28] Mangat N S. An improved randomized response strategy [J]. Journal of the Royal Statistical Society: Series B (Methodological), 1994, 56(1): 93-95

[29] Wang Tianhao, Li Ninghui, Jha S. Locally differentially private frequent itemset mining [C] //Proc of the 2018 IEEE Symp on Security and Privacy (SP). Piscataway, NJ: IEEE, 2018: 127-143

[30] Kaggle. An online shop business transaction [DB/OL]. [2022-03-25]. https://www.kaggle.com/datasets/gabrielramos87/an-online-shop-business

[31] Kaggle. Anime recommendations database [DB/OL]. [2022-04-14]. https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database

[32] Grouplens. MovieLens 1M Dataset [DB/OL]. [2022-04-14]. https://grouplens.org/datasets/movielens/

Cao Yiran, born in 1997. Master candidate. Her main research interests include differential privacy and privacy protection.

曹依然,1997 年生.硕士研究生.主要研究方向为差分隐私和隐私保护.

Zhu Youwen, born in 1986. PhD, professor. Member of CCF. His main research interests include data security, privacy computing, and applied cryptography.

朱友文,1986 年生.博士,教授.CCF 会员.主要研究方向为数据安全、隐私计算和应用密码学.

He Xingyu, born in 1997. PhD candidate. His main research interests include differential privacy and privacy protection.

贺星宇,1997 年生.博士研究生.主要研究方向为差分隐私和隐私保护.

Zhang Yue, born in 1994. PhD candidate. Her main research interests include information security and data privacy.

张 跃,1994 年生.博士研究生.主要研究方向为信息安全和数据隐私.