SDLDP：一种支持数据敏感分级的本地差分隐私框架

陈亚青[1,2], 叶宇桐[1,3], 张 敏[1], 舒波文[1,2]

[1] 中国科学院软件研究所可信计算与信息保障实验室 北京 中国 100190
[2] 中国科学院大学计算机科学与技术学院 北京 中国 100049
[3] 中关村实验室 北京 中国 100094

摘要 在当今大数据时代, 人们在日常生活中产生的数据规模空前庞大。基于用户数据的分析与应用为各行各业的发展提供了有力支持, 同时也引发了公众对隐私泄露的担忧。本地差分隐私模型常用于数据统计任务中保护用户的隐私数据, 通过为真实数据添加随机噪声, 降低隐私泄露风险。然而本地差分隐私模型的高可用性伴随着对大规模数据以及较高隐私预算的依赖, 隐私性和可用性之间更优的平衡仍待挖掘。本文根据数据的取值自然拥有不同敏感级别的特性, 提出了一种支持数据敏感分级的本地差分隐私框架SDLDP, 通过对不同取值的数据提供不同程度的隐私保护, 针对性地降低低敏感数据的本地差分隐私噪声添加量, 实现更高的数据可用性。进一步地, 本文提出了基于该框架的两种机制: SDGRR 和 SDPM。SDGRR 优化了本地差分隐私的经典离散型扰动机制 GRR, 适用于频率估计任务。SDPM 对本地差分隐私的连续型扰动机制 PM 进行优化, 经过 EM 算法后处理, 可高效地估计数据均值。实验结果表明, 与原始 LDP 机制相比, 本文提出的两种机制显著提高了频率估计和均值估计结果的准确性。

关键词 本地差分隐私; 隐私保护; 均值估计; 频率估计; EM 算法
中图法分类号 TP309.2 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2025.07.10

SDLDP: A Local Differential Private Framework for Multi-level Sensitive Data

CHEN Yaqing[1,2], YE Yutong[1,3], ZHANG Min[1], SHU Bowen[1,2]

[1] Trusted Computing and Information Assurance Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
[2] School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Zhongguancun Laboratory, Beijing 100094, China

Abstract In the era of Big Data, the scale of data generated by the public in their daily lives is enormous as never before. The analysis and applications based on the users' data have supported the development of various industries and have also raised public concerns about privacy violations. The local differential privacy model is commonly used in statistical tasks to protect users' private data. It reduces the risk of privacy leakage by adding random noise to the real data. However, the high utility of the local differential privacy model accompanies the reliance on large-scale data as well as the high privacy budget, and thus a better balance between privacy and utility remains to be explored. In this paper, we exploit the property that data values have different sensitivity levels and propose a novel local differential privacy framework called SDLDP which supports the data with different sensitivity levels. The basic idea of this framework is to provide different levels of privacy protection for data with different sensitivity levels and reduce the amount of noise added for less sensitive data to improve the utility of noisy data. In addition, two mechanisms based on this framework are proposed in this paper: SDGRR and SDPM. SDGRR optimizes the classical discrete perturbation mechanism GRR in local differential privacy and it is applied to frequency estimation tasks. SDPM optimizes the continuous perturbation mechanism PM in local differential privacy and it is post-processed by the EM algorithm to efficiently estimate the mean values. The experimental results demonstrate that the two mechanisms proposed in this paper significantly enhance the accuracy of the frequency estimation results and the mean estimation results respectively compared with the state-of-the-art based on the local differential privacy model.

Key words local differential privacy; privacy protection; mean estimation; frequency estimation; EM algorithm

通讯作者： 张敏, 博士, 研究员, Email: zhangmin@iscas.ac.cn。
本课题得到国家重点研发计划 (No. 2022YFB4501500, No. 2022YFB4501503) 资助。
收稿日期: 2023-10-24; 修改日期: 2024-01-17; 定稿日期: 2025-06-19

1 引言

随着移动手机、车载导航、可穿戴设备等智能移动终端的普及, 用户数据源源不断产生。因其蕴含的丰富信息, 收集和分析用户数据对众多机构提升其产品或服务质量具有重要意义, 但这同时也引发了大众对隐私泄露问题的担忧, 因此各种隐私保护技术应运而生。本地差分隐私(简称 LDP, Local Differential Privacy)[1]是一种去中心化的差分隐私保护框架, 允许用户在本地对真实数据作扰动处理, 而服务器接收并分析用户的噪声数据, 完成多种统计任务, 如直方图统计[2]、均值估计[3]、热点发现[4]等。LDP 框架基于严格数学定义, 要求任意两个输入经噪声扰动得到相同输出值的概率比值满足隐私预算 ϵ 的约束, 其中 ϵ 越小, 隐私保护水平越高。另外, LDP 框架的安全性与敌手的背景知识和计算能力无关, 且与中心化的隐私保护方法相比, 其安全性不依赖可信第三方, 这显著降低了用户隐私泄露的风险。因此, LDP 框架不仅受到学术界的广泛关注[2-4], 而且在工业界也得以应用, 例如: 苹果公司应用 LDP 协议计算用户中最流行的 emoji 表情[5]; 谷歌公司提出 RAPPOR 算法[6]应用于 Chrome 浏览器中隐私地收集与分析用户浏览网页的行为。

为提高 LDP 协议应用于统计任务的准确性, 尤其针对数据集规模受限或隐私要求较高的情形, 研究人员一方面考虑为不同的统计任务优化 LDP 协议, 另一方面针对不同应用场景的特殊需求, 对 LDP 框架进行隐私保护的“松弛化”改造, 如 PLDP (Personalized Local Differential Privacy)[7], ULDP (Utility-Optimized Local Differential Privacy) [8]等。与经典 LDP 框架不同, “松弛化”不严格要求任意两个输入扰动后得到相同输出的概率比值都满足相同程度约束, 而是结合场景的隐私保护需求, 仅对部分数据进行严格约束, 而对其他数据不予限制或采取更松弛的约束定义, 从而降低为原始数据添加的噪声规模, 实现较高的统计精度。

基于数据敏感度的“松弛化”是“松弛化”LDP 框架的一类方法。Murakami 等人于 2019 年首次提出了一种 ULDP(Utility-Optimized Local Differential Privacy)框架[8]以及基于该框架的离散型数据频率估计机制。ULDP 框架如图 1(a)所示, 其将真实数据区分为敏感型 $\mathcal{X}_S$ 与非敏感型 $\mathcal{X}_N$ 两类。敏感型数据受标准 LDP 保护, 经扰动只可能得到敏感型输出 $\mathcal{Y}_P$ (即箭头(1)所示)。而非敏感型数据由于不需要保护隐私, 它们或被直接披露(即箭头(3)所示), 或被随机扰动至敏感型输出 $\mathcal{Y}_P$ (即箭头(2)所示), 以防止攻击者通过输出类别直接识别真实数据敏感类别。ULDP 模型通过限定、缩减敏感数据的扰动范围, 大幅降低了噪声规模, 在频率估计任务上显著降低了分析误差。然而, 在更多实际场景中, 往往难以简单地区分数据“是否”敏感, 而是数据的敏感程度有所不同。例如, 在统计患者疾病时, 某些特殊疾病如“艾滋病”相较于“发烧”、“腹泻”等常见疾病, 其敏感度应更高; 类似地, 在体重等连续型数据采集与统计场景中, 过高或过低的体重应比正常范围内取值受到更强的隐私保护。此外, ULDP 可能暴露真实数据的敏感类型, 导致额外的隐私泄露。由于 ULDP 对非敏感型真实数据的直接披露(即图 1(a)中的箭头(3)), 攻击者可以确定非敏感型输出 $\mathcal{Y}_I$ 的原始真实数据为非敏感型, 且攻击者若推测敏感型输出 $\mathcal{Y}_P$ 的原始真实数据为敏感型则大概率是正确的, 因此 ULDP 可能暴露真实数据的敏感类型。而在部分场景中, 真实数据的敏感类型本身就是隐私信息, ULDP 将带来额外的隐私泄露。

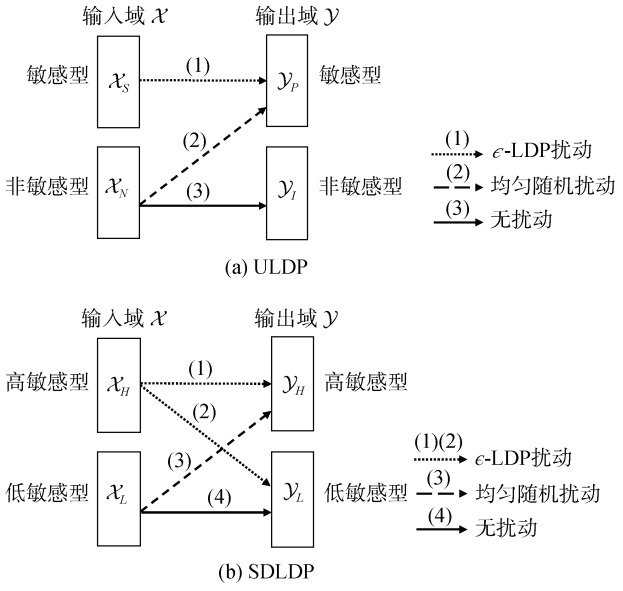


图 1 ULDP 与 SDLDP 的对比

Figure 1 Comparison between ULDP and SDLDP

针对上述问题, 本文提出了一种新型的支持数据敏感分级的本地差分隐私框架 SDLDP(Sensitivity Discriminant Local Differential Privacy), 如图 1(b)所示。SDLDP 将输入数据区分为“高”、“低”两类敏感级别, 即 $\mathcal{X}_H$ 和 $\mathcal{X}_L$, 为两类数据提供不同程度的隐私保护。与 LDP 框架相比, SDLDP 缩减了低敏感数据的扰动输出范围, 以提高噪声数据可用性。与 ULDP 相比, SDLDP 将高敏感数据的扰动输出范围

扩大为整个输出域, 增强了隐私保护效果。对于低敏感数据, 由于噪声数据的干扰(即图 1(b)中的箭头(2)), 其不再有直接泄露的风险。对于高敏感数据, 我们借鉴文献[9-11]采用的从攻击角度度量 LDP 机制隐私性的方法, 采用贝叶斯敌手模型猜测噪声数据的真实取值, 猜测的成功率越低说明机制的隐私保护性能越强。理论和实验的分析结果均表明敌手对 SDLDP 机制的攻击成功率与 LDP 机制相同, 而远低于 ULDP 机制, 从而表明 SDLDP 的隐私保护性能优于 ULDP。

另外, 本文分别针对离散型数据的频率估计任务和连续型数据的均值估计任务提出 SDLDP 框架下的新机制 SDGRR(Sensitivity Discriminant General Random Response)和 SDPM(Sensitivity Discriminant Piecewise Mechanism)。我们通过实验验证了两个新机制的表现, 在不同隐私预算设置下对比了新机制和 LDP 先进机制在相应任务上的均方误差, 结果表明: 本文提出的新机制的误差显著低于 LDP 先进机制, 尤其在隐私预算较小时($\epsilon \leqslant 0.3$), SDGRR 的均方误差比 GRR(General Random Response)低约 1 个数量级, SDPM 的均方误差低于 PM(Piecewise Mechanism)约 2 个数量级。

本文的贡献总结如下:

(1) 提出了一种支持数据敏感分级的本地差分隐私框架 SDLDP, 对高低敏感级别数据提供不同程度的隐私保护;

(2) 提出了 SDLDP 框架下的频率估计新机制 SDGRR, 理论分析和实验结果证明了其频率估计方差显著低于 GRR;

(3) 首次在松弛化的 LDP 框架下考虑均值估计问题, 提出 SDLDP 框架下的新机制 SDPM。与 PM 相比, SDPM 降低了扰动方差, 并通过 EM(Expectation Maximization)算法进一步降低了均值估计误差。

2 相关工作

针对特定数据分析任务的协议优化是本地差分隐私领域的重要研究内容。其中, 对于频率估计任务, 研究者基于随机响应思想[12], 提出了 GRR[13]、RAPPOR[6]、OUE[2]和 OLH[2]等机制, 文献[2]提出的通用 LDP 框架表明, GRR 在数据取值域较小时表现最优。针对均值统计任务, Wang 等人[3]提出分段机制(Piecewise Mechanism, PM), 其思想是将数据的扰动输出域限定在有限的连续区间内, 并且令输出值以高概率携带输入数据的有关信息, 其准确性显著优于 LDP 下的其他均值估计机制; Li 等人[4]基于与 PM 相似的思想提出了用于估计连续型数据分布的方波机制(SquareWave mechanism, SW), 其在均值估计任务上的性能表现与 PM 接近。另外, 针对键值对收集[15-16]、频繁项挖掘[17]等任务也有相应的优化协议。

另一方面, 考虑到对任意输入提供相同程度的隐私保护在许多场景下是不必要的, 许多研究工作关注本地差分隐私框架的松弛定义。目前, 依赖数据内容的松弛化研究包括两大类, 一类是设定动态隐私预算, 定义任意两个数据之间的输出概率比值与数据之间的相对距离相关, 例如: Local d-Privacy[18]和 CLDP[11]均将隐私预算表示为数据距离的函数, 前者仅适用于测度数据, 如位置数据; 后者覆盖定序数据与非定序数据。此类方法使输出噪声数据的相似度与输入数据间的距离相关, 输入数据距离越大则越容易被区分。另一类松弛方法与本文出发点相似, 考虑数据存在敏感度差异, 为不同敏感度的数据分配不同的隐私预算。前文提到的 ULDP 模型属于此类, 通过将数据划分为敏感与非敏感数据, 为其分别设定不同的扰动机制, 大幅降低噪声规模, 从而显著降低频率估计任务的统计误差。MinID-LDP[19]在此基础上提出了一种更通用的表达, 支持每个数据依据敏感度从隐私预算集合中选择一个隐私预算, 安全性定义为: $\Pr[\mathrm{F}(x)=y] \leqslant e^{\min\{\epsilon_x,\epsilon_{x'}\}}\Pr[\mathrm{F}(x')=y]$, 即对于任意两个输入数据 x 和 x', 经随机化机制 F 扰动输出任意数据 y 的概率相似度由两个输入数据的隐私预算较小值 $\min\{\epsilon_x,\epsilon_{x'}\}$ 约束。虽然 MinID-LDP 声明可以将 ULDP 归纳为其一种特例, 但事实上 MinID-LDP 更为严格。根据其定理[19], MinID-LDP 满足 $2\min\{\epsilon_x|x\in\mathcal{X}\}$-LDP, 因而其事实安全性比上述定义更为严格, 相较于 ULDP 难以达到提升可用性的效果。此外, 有研究者从用户的角度考虑不同的隐私需求。例如 PLDP[7]允许用户依据自身偏好选择合适的隐私预算, 为用户提供个性化的隐私保护。然而只有当每个用户都对 LDP 具体协议有充分理解并能选择适当的参数时, PLDP 才能提升统计结果的可用性。本文提出的 SDLDP 在 ULDP 的基础上增强了安全性, 不仅要求任意两个输入和任意高敏感输出满足 ϵ-LDP(即 ULDP 的式(6)和 SDLDP 的式(10)所示), 而且要求任意两个高敏感输入和任意输出满足 ϵ-LDP(即 SDLDP 的式(8)所示)。

3 预备知识

本章节首先给出问题与符号定义, 其次介绍本地差分隐私的定义, 最后介绍了两个本地差分隐私

协议, 即 General Random Response(GRR)和 Piecewise Mechanism (PM)。

3.1 问题与符号定义

表 1 列出了本文出现的重要符号定义。考虑有 n 名用户, 第 i 名用户的本地隐私数据为输入空间 $\mathcal{X}$ 上的一维随机变量 x_i, 假设所有用户的数据相互独立, 且服从同一概率分布 f。数据收集者的目的是对全体用户数据做统计分析: 若用户数据为连续型则估计总体均值, 若用户数据为离散型则估计各离散数值的频率。

表 1 符号定义

Table 1 Definition of symbols

符号	定义
n	用户总数
$\mathcal{X}$	用户隐私数据的取值域
$\mathcal{Y}$	随机化数据的取值域
f	用户隐私数据的真实分布
x_i	第 i 名用户的隐私数据
x_i^*	第 i 名用户的本地随机化数据
$\tilde{f}$	用户随机化数据的统计分布
$\hat{f}$	对 f 的估计

为了保护数据隐私, 各用户对原始数据 x_i 进行本地随机扰动, 记随机化的噪声数据为输出空间 $\mathcal{Y}$ 上的一维随机变量 x_i^*。各用户将噪声数据 x_i^* 发送给数据收集者, 随后数据收集者基于噪声数据的统计分布 $\tilde{f}$ 执行均值估计或频率估计任务。我们的目标是为用户数据提供隐私保护, 同时保证数据收集者统计分析的准确性。

3.2 本地差分隐私

本地差分隐私(Local Differential Privacy, LDP)[1]是一种基于严格数学定义的隐私保护模型, 其定义如下:

定义 1. ϵ -LDP. 对于 $\forall\epsilon\in\mathbb{R}^+$, 随机化函数 $\mathrm{F}:\mathcal{X}\to\mathcal{Y}$ 满足 ϵ -LDP, 当且仅当对于 $\forall x,x'\in\mathcal{X}$ 和 $\forall y\in\mathcal{Y}$, 满足:

$$\Pr[\mathrm{F}(x)=y]\leqslant e^{\epsilon}\Pr[\mathrm{F}(x')=y]$$

满足 ϵ -LDP 的随机化函数为所有输入数据平等地提供相同水平的隐私保护。对于任意两个输入数据, 扰动输出为相同值的概率是相近的。这种相近程度由隐私预算 ϵ 衡量, ϵ 越小, 则由不同输入得到相同输出的概率越为相近, 即隐私保护更强。

3.3 LDP 扰动机制

本节介绍两种满足 ϵ -LDP 的扰动机制, 即 General Random Response(GRR)[13] 和 Piecewise Mechanism (PM)[3]。GRR 来源于随机响应[12], 是一种经典的 ϵ -LDP 机制, 适用于频率估计任务。在数据取值域较小时($|\mathcal{X}|<3e^{\epsilon}+2$)表现最优。GRR 的主要思想是对输入数据 x_i 做随机化扰动, 使其大概率输出 x_i 本身。

General Random Response(GRR)[13] 记 GRR 的输入域为离散集合 $\mathcal{X}$, 其输出域同样为 $\mathcal{X}$。GRR 根据以下概率密度函数扰动输入值 $x_i\in\mathcal{X}$, 输出 $x_i^*\in\mathcal{X}$:

$$\Pr(x_i^*\mid x_i)=\begin{cases}p, & \text{if } x_i^*=x_i,\\ q, & \text{otherwise.}\end{cases}\tag{1}$$

其中 $p=\dfrac{e^{\epsilon}}{|\mathcal{X}|+e^{\epsilon}-1}$, $q=\dfrac{1}{|\mathcal{X}|+e^{\epsilon}-1}$, $|\mathcal{X}|$ 表示离散集合 $\mathcal{X}$ 的元素个数, ϵ 为隐私预算。

根据式(1), 对于 $\forall x,x'\in\mathcal{X}$ 以及 $\forall y\in\mathcal{X}$, 有 $\dfrac{\Pr(y\mid x)}{\Pr(y\mid x')}\leqslant\dfrac{e^{\epsilon}}{|\mathcal{X}|+e-1}/\dfrac{1}{|\mathcal{X}|+e-1}=e^{\epsilon}$, 从而 GRR 满足 ϵ -LDP。

为了估计以 x_i 为真实值的用户频率 f_i, 数据收集者首先统计噪声数据中 x_i 出现的频率 $\tilde{f}_i$, 随后估计 x_i 的真实频率 $\hat{f}_i$:

$$\hat{f}_i=\frac{\tilde{f}_i-q}{p-q}$$

根据文献[2], $\hat{f}_i$ 是 f_i 的无偏估计, 即 $\mathrm{E}\left[\hat{f}_i\right]=f_i$. 另外, 根据文献[2], 大多数应用场景中稀疏数据占大多数并且目标是识别频繁项, 提高准确性的关键在于降低低频项的估计误差。对于真实频率较低的数据, 其估计方差的近似值为[2]:

$$\mathrm{Var}^*[\hat{f}_i]=\frac{|\mathcal{X}|-2+e^{\epsilon}}{n(e^{\epsilon}-1)^2}\tag{2}$$

为了保证隐私性, 通常 ϵ 的取值较小, 因此可将 e^{ϵ} 视为低阶项, 同时视数据取值域规模 $|\mathcal{X}|$ 或数据库规模 n 为主项, 因此:

$$\mathrm{Var}^*[\hat{f}_i]=\mathrm{O}\left(\frac{|\mathcal{X}|}{n}\right)\tag{3}$$

式(3)表明, 数据取值域规模越大, GRR 的统计精度越差, 而增加用户数量可以减小误差。

PM 是适用于均值估计的 ϵ -LDP 机制, 其主要思

想与随机响应类似, 以较高的概率将输入值 x_i 扰动为邻近数据, 在噪声数据中保留较多原始数据的有关信息, 从而提升噪声数据的可用性。

Piecewise Mechanism (PM)[3] 对于输入值 $x_i \in [-1,1]$, PM 依据以下概率密度函数, 扰动输入值 x_i, 输出 $x_i^* \in [-C,C]$:

$$\Pr(x_i^* \mid x_i)=\begin{cases} p, & \text{if } x_i^* \in [l(x_i), r(x_i)], \\ \dfrac{p}{e^{\epsilon}}, & \text{if } x_i^* \in [-C, l(x_i)) \cup (r(x_i), C]. \end{cases} \tag{4}$$

其中 $C=\dfrac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}$, $p=\dfrac{e-e^{\epsilon/2}}{2e^{\epsilon/2}+2}$, $l(x_i)=\dfrac{C+1}{2}, x_i-\dfrac{C-1}{2}$, $r(x_i)=l(x_i)+C-1$, ϵ 为隐私预算。

根据式(4), 对于任意两个输入 $x,x' \in [-1,1]$, 以及任意输出 $y \in [-C,C]$, 都有 $\dfrac{\Pr(y \mid x)}{\Pr(y \mid x')} \leqslant \dfrac{p}{p/e^{\epsilon}}=e^{\epsilon}$, 从而 PM 满足 ϵ-LDP。由文献[3]可知, 给定输入值 x_i, PM 输出的噪声值 x_i^* 是 x_i 的无偏估计, 即 $\mathrm{E}[x_i^*]=x_i$, 并且方差满足:

$$\mathrm{Var}[x_i^*]=\frac{x_i^2}{e^{\epsilon/2}-1}+\frac{e^{\epsilon/2}+3}{3(e^{\epsilon/2}-1)^2} \tag{5}$$

数据收集者计算噪声值 x_i^* 的均值, 作为均值 X 的估计结果, 即:

$$\hat{X}=\frac{1}{n}\sum_{i=1}^{n} x_i^*$$

因为噪声值 x_i^* 是真实值 x_i 的无偏估计, 数据收集者估计的均值 $\hat{X}$ 也是真实均值 X 的无偏估计。

3.4 ULDP

ULDP(Utility-Optimized Local Differential Privacy)框架[8]如图 1(a)所示, 其将真实数据区分为敏感型 $\mathcal{X}_S$ 与非敏感型 $\mathcal{X}_N$ 两类, 扰动输出也被划分为受保护的敏感型输出 $\mathcal{Y}_P$ 和可逆推出真实数据的非敏感型输出 $\mathcal{Y}_I$。其形式化定义如下:

定义 2. $(\mathcal{X}_S, \mathcal{Y}_P, \epsilon)$-ULDP. 给定 $\mathcal{X}_S \in \mathcal{X}$ 和 $\mathcal{Y}_P \in \mathcal{Y}$, 对于 $\forall \epsilon \in \mathbb{R}^+$, 随机化函数 $\mathrm{F}: \mathcal{X} \to \mathcal{Y}$ 满足 $(\mathcal{X}_S, \mathcal{Y}_P, \epsilon)$-ULDP, 当且仅当以下两个条件成立:

1. 对于 $\forall y \in \mathcal{Y}_I$, 存在 $x \in \mathcal{X}_N$ 使得:

$\Pr[\mathrm{F}(x)=y]>0$ 且对于 $\forall x' \neq x$ 有 $\Pr[\mathrm{F}(x')=y]=0$

2. 对于 $\forall x,x' \in \mathcal{X}, \forall y \in \mathcal{Y}_P$ 满足:

$$\Pr[\mathrm{F}(x)=y] \leqslant e^{\epsilon}\Pr[\mathrm{F}(x')=y] \tag{6}$$

由条件 2 的式(6)可知, ULDP 对任意两个输入和 $\mathcal{Y}_P$ 中的输出提供 ϵ-LDP 保护。另外, 由条件 1 可知, 扰动输出到 $\mathcal{Y}_I$ 的相应输入是能够唯一确定的, 即 $\mathcal{Y}_I$ 中的噪声数据能被逆推出原真实数据。

URR(Utility-Optimized Randomized Response)[8] 是 ULDP 框架下针对离散型数据的随机化机制, 其输入域与输出域相同, 即 $\mathcal{X}=\mathcal{Y}$, 并且满足 $\mathcal{X}_S=\mathcal{Y}_P$ 和 $\mathcal{X}_N=\mathcal{Y}_I$, 依据以下概率密度函数扰动输入 $x_i \in \mathcal{X}$, 输出噪声值 $x_i^* \in \mathcal{Y}$:

$$\Pr(x_i^* \mid x_i)=\begin{cases} c_1, & \text{if } x_i \in \mathcal{X}_S \text{ and } x_i^*=x_i, \\ c_2, & \text{if } x_i \in \mathcal{X}_S \text{ and } x_i^* \in \mathcal{X}_S \setminus \{x_i\}, \\ c_2, & \text{if } x_i \in \mathcal{X}_N \text{ and } x_i^* \in \mathcal{X}_S, \\ c_3, & \text{if } x_i \in \mathcal{X}_N \text{ and } x_i^*=x_i, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

其中: $c_1=\dfrac{e}{|\mathcal{X}_s|+e-1}, c_2=\dfrac{1}{|\mathcal{X}_s|+e-1}, c_3=\dfrac{e-1}{|\mathcal{X}_s|+e-1}$, ϵ 为隐私预算。

3.5 贝叶斯敌手模型

假定敌手只能观测到用户端发送给数据收集端的本地扰动输出值 y, 贝叶斯敌手(记为 $\mathcal{A}$)[9-11]的攻击目标是寻找使得后验概率 $\Pr[x \mid y]$ 最大的输入数据 x, 作为对扰动输出值 y 的真实值的猜测。形式化贝叶斯敌手 $\mathcal{A}$ 的攻击输出 $\mathcal{A}(y)$ 为:

$$\begin{aligned} \mathcal{A}(y) &\triangleq \arg\max_{x \in \mathcal{X}} \Pr[x \mid y] \\ &= \arg\max_{x \in \mathcal{X}} \frac{\Pr[y \mid x]\pi(x)}{\Pr[y]} \\ &= \arg\max_{x \in \mathcal{X}} \Pr[y \mid x]\pi(x) \end{aligned}$$

其中: $\pi(\bullet)$ 表示输入数据的先验概率; $\Pr[y \mid x]=\Pr[\mathrm{F}(x)=y]$, 表示输入数据 x 经本地隐私化扰动机制 $\mathrm{F}(\bullet)$ 扰动后输出 y 的概率。第一步推导服从贝叶斯定理, 第二步推导来自 $\Pr[y]$ 是一个与输入数据 x 无关的常数。若输入数据的先验分布未知, 贝叶斯敌手 $\mathcal{A}$ 的攻击输出 $\mathcal{A}(y)$ 可简化为:

$$\mathcal{A}(y)=\arg\max_{x \in \mathcal{X}} \Pr[y \mid x]$$

即猜测扰动输出值 y 的真实值为最大概率扰动输出 y 的输入值。不失一般性地, 本文考虑没有输入数据先验分布的贝叶斯敌手。

对于 LDP 机制 $\mathrm{F}(\bullet)$, 可采用贝叶斯敌手攻击的成功率 *ACC* 度量其隐私保护水平[9-10], 其合理性在于: LDP 机制的目标是阻止敌手从观测的扰动输出

值成功推测出用户的真实值, 因此隐私保护水平越高的机制使贝叶斯敌手攻击的成功率越低。理论上, 贝叶斯敌手攻击的成功率 ACC 的期望为:

$$\mathrm{E}[ACC]=\Pr[\mathcal{A}(\mathrm{F}(x_{in}))=x_{in}]$$

其中, $\mathcal{A}(\mathrm{F}(x_{in}))$ 表示敌手观测到扰动输出值 $\mathrm{F}(x_{in})$ 后对真实值的猜测, x_{in} 表示用户的真实值。

4 敏感分级的本地差分隐私

4.1 定义

支持数据敏感分级的本地差分隐私框架 SDLDP (Sensitivity Discriminant Local Differential Privacy)的基本思想如图 1(b)所示。我们将真实数据取值域 $\mathcal{X}$ 划分为高敏感型数据 $\mathcal{X}_H$ 和低敏感型数据 $\mathcal{X}_L$。SDLDP 对 $\mathcal{X}_H$ 的保护等同于本地差分隐私, 缩减了 $\mathcal{X}_L$ 内数据的扰动范围, 其具体定义如下。

定义 3. $(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$ -SDLDP. 假定随机化函数为 $\mathrm{F}:\mathcal{X}\to\mathcal{Y}$,其中输入域 $\mathcal{X}=\mathcal{X}_H\cup\mathcal{X}_L$ 且 $\mathcal{X}_H\cap\mathcal{X}_L=\varnothing$, 输出域 $\mathcal{Y}=\mathcal{Y}_H\cup\mathcal{Y}_L$, $\mathcal{Y}_H\cap\mathcal{Y}_L=\varnothing$ 且 $\mathcal{Y}_L=\mathcal{X}_L$。对于 $\forall\epsilon\in\mathbb{R}^+$, F 满足$(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$ -SDLDP, 当且仅当以下三个条件成立:

1. 对于 $\forall x,x'\in\mathcal{X}_H$, $\forall y\in\mathcal{Y}$ 满足:

$$\Pr[\mathrm{F}(x)=y]\leqslant e^{\epsilon}\Pr[\mathrm{F}(x')=y] \tag{8}$$

2. 对于 $\forall x\in\mathcal{X}_L$ 以及 $\forall x'\in\mathcal{X}_L\setminus\{x\}$ 满足:

$$\Pr[\mathrm{F}(x)=x]>0 \text{ 且 } \Pr[\mathrm{F}(x')=x]=0 \tag{9}$$

3. 对于 $\forall x,x'\in\mathcal{X}$, $\forall y\in\mathcal{Y}_H$ 满足:

$$\Pr[\mathrm{F}(x)=y]\leqslant e^{\epsilon}\Pr[\mathrm{F}(x')=y] \tag{10}$$

条件 1 表明, $\mathcal{X}_H$ 中的任意两个高敏感输入经 F 扰动后输出为 $\mathcal{Y}$ 中同一值的概率相似, 这种相似程度由隐私预算 ϵ 衡量, 这意味着对高敏感输入的隐私保护程度等同于 ϵ-LDP。而条件 2 缩减了低敏感输入的扰动输出范围: 若低敏感输入 $x\in\mathcal{X}_L$ 经 F 扰动后输出到 $\mathcal{Y}_L$, 则只有可能输出 x 本身。结合条件 3 可知, 随机化函数对低敏感输入的随机输出限制为 $\mathcal{Y}_H\cup\{x\}$。另外, 高敏感输入经扰动可能输出到 $\mathcal{Y}_L$, 这不仅隐藏了低敏感输入的直接输出, 避免完全泄露低敏感输入, 也使得敌手在观察噪声数据时, 无法以较高的概率区分高敏感输入与低敏感输入的扰动输出, 保证了高敏感输入的隐私。

4.2 离散型数据的敏感分级随机响应 SDGRR

针对离散型数据, 本文提出了满足$(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$ -SDLDP 的随机响应(Sensitivity Discriminate General Random Response, SDGRR), 描述如下。

Sensitivity Discriminate General Random Response (SDGRR) 记输入域为离散集合 $\mathcal{X}=\mathcal{X}_H\cup\mathcal{X}_L$ 且 $\mathcal{X}_H\cap\mathcal{X}_L=\varnothing$, 输出域 $\mathcal{Y}=\mathcal{Y}_H\cup\mathcal{Y}_L$ 且满足 $\mathcal{Y}_H=\mathcal{X}_H$ 和 $\mathcal{Y}_L=\mathcal{X}_L$。SDGRR 依据以下概率密度函数对输入 $x_i\in\mathcal{X}$ 进行扰动, 输出噪声值 $x_i^*\in\mathcal{Y}$:

$$\Pr(x_i^*\mid x_i)=\begin{cases}c_1, & \text{if } x_i\in\mathcal{X}_H \text{ and } x_i^*=x_i,\\ c_2, & \text{if } x_i\in\mathcal{X}_H \text{ and } x_i^*\in\mathcal{Y}\setminus\{x_i\},\\ c_2, & \text{if } x_i\in\mathcal{X}_L \text{ and } x_i^*\in\mathcal{Y}_H,\\ c_3, & \text{if } x_i\in\mathcal{X}_L \text{ and } x_i^*=x_i,\\ 0, & \text{otherwise.}\end{cases} \tag{11}$$

其中: $c_1=\dfrac{e^{\epsilon}}{|\mathcal{X}|+e-1}, c_2=\dfrac{1}{|\mathcal{X}|+e-1}, c_3=\dfrac{|\mathcal{X}_L|+e^{\epsilon}-1}{|\mathcal{X}|+e-1}$, ϵ 为隐私预算。

为了直观说明 SDGRR 的思想, 这里描述一个简单的例子。假定在统计疾病频率的场景中, $\mathcal{X}$= {“乙肝”, “糖尿病”, “脂肪肝”}, 其中高敏感数据为 $\mathcal{X}_H$ = {“乙肝”}, 低敏感数据为 $\mathcal{X}_L$ = {“糖尿病”, “脂肪肝”}。则 $\mathcal{Y}_H$ = {“乙肝”}, $\mathcal{Y}_L$ = {“糖尿病”, “脂肪肝”}。若输入为高敏感数据, 即“乙肝”, SDGRR 将以较高的概率 c_1 输出“乙肝”本身, 而以较低的概率 c_2 输出“脂肪肝”或“糖尿病”。若输入为低敏感数据, 例如“糖尿病”, 则 SDGRR 以概率 c_3 输出为“糖尿病”本身, 或者以较低的概率 c_2 输出为高敏感数据“乙肝”, 而不可能输出“脂肪肝”。

定理 1. SDGRR 满足$(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$ -SDLDP.

证明. 1. 对于 $\forall x,x'\in\mathcal{X}_H$ 和 $\forall y\in\mathcal{Y}$, 由式(11)可知, $\dfrac{\Pr(y\mid x)}{\Pr(y\mid x')}\leqslant\dfrac{c_1}{c_2}=e^{\epsilon}$, 故式(8)成立.

2. 对于 $\forall x\in\mathcal{X}_L$ 以及 $\forall x'\in\mathcal{X}_L\setminus\{x\}$, 由式(11)可知, $\Pr(x\mid x)=c_3$, 则有 $\Pr(x\mid x')=0$, 故满足式(9).

3. 对于 $\forall x,x'\in\mathcal{X}$ 和 $\forall y\in\mathcal{Y}_H$, 由式(11)可知, 有 $\dfrac{\Pr(y\mid x)}{\Pr(y\mid x')}\leqslant\dfrac{c_1}{c_2}=e^{\epsilon}$, 故式(10)成立.

综上所述, SDGRR 满足$(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$ -SDLDP.

证毕。

为了估计 x_i 的频率 f_i, 首先统计 n 位用户经 SDGRR 扰动的噪声数据中 x_i 出现的频率 $\tilde{f}_i$。对于高敏感数据, 即 $x_i\in\mathcal{X}_H$, 噪声数据中 x_i 出现的概率为:

$$\Pr[x_i^*=x_i]=f_ic_1+(1-f_i)c_2$$

由最大似然估计得 x_i 的频率估计值为:

$$\hat{f}_i = \frac{\tilde{f}_i - c_2}{c_1 - c_2}$$

同理, 可得 $x_i \in \mathcal{X}_L$ 的频率的最大似然估计值为:

$$\hat{f}_i = \frac{\tilde{f}_i - \sum_{x_k \in \mathcal{X}_H} \hat{f}_k c_2}{c_3},$$

易知 $\hat{f}_i$ 是 f_i 的无偏估计量, 即 $\mathrm{E}\left[\hat{f}_i\right] = f_i$.

因为 SDGRR 对高敏感数据的频率估计方法与 GRR 相同, 所以其方差的近似值如式(2)所示。定理 2 给出了 SDGRR 中低敏感数据的频率估计的近似方差。

定理 2. SDGRR 中低敏感数据的频率估计方差为 $\mathrm{Var}^*\left[\hat{f}_i\right] = \mathrm{O}\left(\frac{1}{\alpha^2 n|\mathcal{X}|}\right)$, 其中 $\alpha = \frac{|\mathcal{X}_L|}{|\mathcal{X}|}$.

证明: 对于 $x_i \in \mathcal{X}_L$, $\hat{f}_i$ 的方差满足:

$$\mathrm{Var}\left[\hat{f}_i\right] = \frac{\mathrm{Var}\left[\tilde{f}_i\right]}{c_3^2} = \frac{\mathrm{Var}[Z_i]}{n c_3^2},$$

其中随机变量 Z_i 服从 $\Pr[Z_i = 1] = f_i c_3 + \sum_{x_k \in \mathcal{X}_H} f_k c_2$ 的伯努利分布, 故:

$$\mathrm{Var}\left[\hat{f}_i\right] = \frac{\left[f_i c_3 + \sum_{x_k \in \mathcal{X}_H} f_k c_2\right]\left[1 - f_i c_3 - \sum_{x_k \in \mathcal{X}_H} f_k c_2\right]}{n c_3^2}$$

$$= \frac{\left(\sum_{x_k \in \mathcal{X}_H} f_k c_2\right)\left(1 - \sum_{x_k \in \mathcal{X}_H} f_k c_2\right)}{c_3^2 n}$$

$$+ \frac{f_i\left(1 - 2\sum_{x_k \in \mathcal{X}_H} f_k c_2\right)}{c_3 n} - \frac{f_i^2}{n}$$

当 f_i 较小时, 带入 c_2, c_3 的值, 得到 $\mathrm{Var}\left[\hat{f}_i\right]$ 的近似值为:

$$\mathrm{Var}^*\left[\hat{f}_i\right] = \frac{\left(|\mathcal{X}| + e^{\epsilon} - 1 - \sum_{x_k \in \mathcal{X}_H} f_k\right)\sum_{x_k \in \mathcal{X}_H} f_k}{n\left(|\mathcal{X}_L| + e^{\epsilon} - 1\right)^2} \quad (12)$$

当隐私保护程度较高, 即 ϵ 较小时, 将 e^{ϵ} 视为低阶项, 以及视数据取值域规模 $|\mathcal{X}|$ 或数据库规模 n 为主项, 同时令 $\alpha = \frac{|\mathcal{X}_L|}{|\mathcal{X}|}$ 则有:

$$\mathrm{Var}^*\left[\hat{f}_i\right] = \mathrm{O}\left(\frac{|\mathcal{X}|}{n|\mathcal{X}_L|^2}\right) = \mathrm{O}\left(\frac{1}{\alpha^2 n|\mathcal{X}|}\right) \quad (13)$$

证毕。

图 2 对比了在不同隐私预算下, GRR(即 SDGRR 的高敏感数据扰动方式)与 SDGRR 的低敏感数据扰动方式下频率估计的近似方差, 即可视化了式(2)与式(12)。其中 SDGRR_L/H 表示 SDGRR 的低/高敏感数据扰动方式, 后缀数字表示低敏感取值 $\mathcal{X}_L$ 占取值域 $\mathcal{X}$ 的比例。根据文献[20], 均匀分布是最难估计的分布形式, 因此假设真实数据服从均匀分布。由图可知, SDGRR 中低敏感数据的频率估计方差显著低于 GRR, 且低敏感数据的占比越高, 优势越明显, 这与式(13)的观察一致。

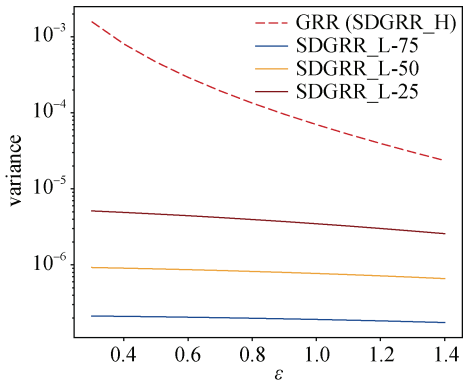


图 2 GRR 与 SDGRR 频率估计方差对比

Figure 2 Comparison of variances between GRR and SDGRR for frequency estimation

4.3 连续型数据的敏感分级扰动机制 SDPM

针对连续型数据的均值估计任务, 本文在 PM 的基础上, 提出了满足 $(\mathcal{X}_H, \mathcal{X}_L, \epsilon)$-SDLDP 的分段机制(Sensitivity Discriminate Piecewise Mechanism, SDPM), 具体描述如下。

Sensitivity Discriminate Piecewise Mechanism(SDPM) 令输入域为 $\mathcal{X} = [-1,1]$, 其中 $\mathcal{X}_L = [l,r]$, $\mathcal{X}_H = [-1,l) \cup (r,1]$, 输出域 $\mathcal{Y} = [-C,C]$, 其中 $\mathcal{Y}_L = \mathcal{X}_L = [l,r]$ 且 $\mathcal{Y}_H = \mathcal{Y} \setminus \mathcal{Y}_L = [-C,l) \cup (r,C]$。SDPM 依据以下算法对输入数据 $x_i \in \mathcal{X}$ 进行扰动, 输出噪声值 $x_i^* \in \mathcal{Y}$:

1. 若 $x_i \in \mathcal{X}_H$, 根据式(4)扰动 x_i, 输出 $x_i^* \in [-C,C]$;

2. 若 $x_i \in \mathcal{X}_L$, 根据以下概率密度函数扰动 x_i, 输出 $x_i^* \in \mathcal{Y}_H \cup \{x_i\}$:

$$\Pr(x_i^* \mid x_i) = \begin{cases} p', & \text{if } x_i^* = x_i, \\ \frac{p}{e^{\epsilon}}, & \text{if } x_i^* \in \mathcal{Y}_H. \end{cases} \quad (14)$$

其中 $p = \frac{e - e^{\epsilon/2}}{2e^{\epsilon/2} + 2}$, $p' = 1 - (2C + l - r)\frac{p}{e^{\epsilon}}$, ϵ 为隐私预算。

图 3 展示了对于不同的输入 x_i, SDPM 的输出服从的概率分布。当输入 x_i 为高敏感型数据 (如图(a)和(d)所示), 输出值的取值范围为 $\mathcal{Y}=[-C,C]$, 且输出值的概率分布与 PM 相同, 即以高概率 p 输出真实值附近的数据(即式(4)中 $[l(x_i),r(x_i)]$ 内的数据), 以低概率 p/e^{ϵ} 输出其他离真实值较远的数据; 当输入 x_i 为低敏感型数据 (如(b)和(c)所示), 输出值的取值范围为 $\mathcal{Y}_H \cup \{x_i\}$, 即以高概率 p' 输出真实值 x_i, 以低概率 p/e^{ϵ} 输出 $\mathcal{Y}_H$ 中的数据。

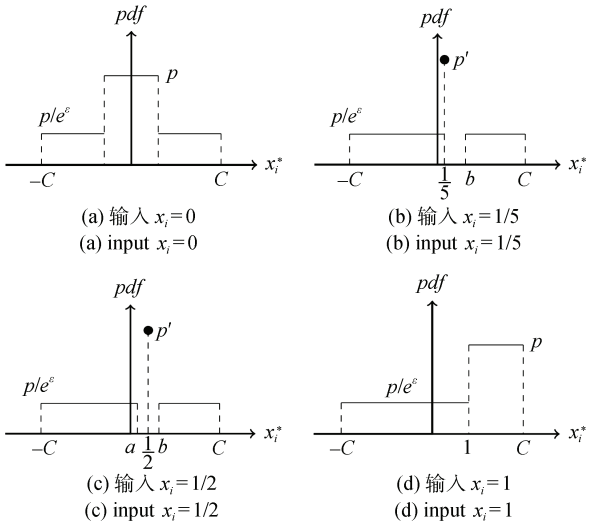


图 3 SDPM 输出的概率分布, 假定低敏感区间为 $\left[\frac{1}{5},\frac{4}{5}\right]$

Figure 3 The Probability Density Functions (*pdf*) of SDPM's Outputs with the Low Sensitive Range as $\left[\frac{1}{5},\frac{4}{5}\right]$

定理 3. SDPM 满足$(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$-SDLDP.

证明. 1. 对于 $\forall x,x'\in\mathcal{X}_H$ 和 $\forall y\in\mathcal{Y}$, 由式(4)可知, $\frac{\Pr(y\mid x)}{\Pr(y\mid x')}\leqslant\frac{p}{p/e^{\epsilon}}=e^{\epsilon}$, 故式(8)成立.

2. 对于 $\forall x\in\mathcal{X}_L$ 以及 $\forall x'\in\mathcal{X}_L\setminus\{x\}$, 由式(14)可知, 若 $\Pr(x\mid x)=p'$, $\Pr(x\mid x')=0$, 故满足式(9).

3. 对于 $\forall x,x'\in\mathcal{X}$ 和 $\forall y\in\mathcal{Y}_H$, 由式(4)和式(14)可知, $\frac{\Pr(y\mid x)}{\Pr(y\mid x')}\leqslant\frac{p}{p/e^{\epsilon}}=e^{\epsilon}$, 故式(10)成立.

综上所述, SDPM 满足$(\mathcal{X}_H,\mathcal{X}_L,\epsilon)$-SDLDP.

证毕。

SDPM 对高敏感数据的扰动方式与 PM 相同, 所以其扰动输出的方差为式(5)。下面我们考虑 SDPM 对低敏感数据的扰动输出的方差。不失一般性地, 假设低敏感区间为$[-c,c]\subseteq[-1,1]$, 可得定理 4.

定理 4. SDPM 的低敏感数据的扰动输出的方差为 $\mathrm{Var}[x_i^*]=\mathrm{O}\left(\frac{e^{\epsilon/2}+3}{3\left(e^{\epsilon/2}-1\right)^2}\right)$.

证明: $\mathrm{Var}[x_i^*]=\mathrm{E}[(x_i^*)^2]-(\mathrm{E}[x_i^*])^2=\int_{-C}^{-c}\frac{px^2}{e}\mathrm{d}x+\int_{x_i}^{x_i}p'x^2\mathrm{d}x+\int_{c}^{C}\frac{px^2}{e}\mathrm{d}x-0=\frac{\left(C^3-c^3\right)(e^{\epsilon}-e^{\epsilon/2})}{3e^{\epsilon}\left(e^{\epsilon/2}+1\right)}$

带入 $C=\frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}$, 得:

$$\begin{aligned}\mathrm{Var}[x_i^*]&=\frac{e^{\epsilon/2}+2+e^{-\epsilon/2}}{3(e^{\epsilon/2}-1)^2}-\frac{(e^{\epsilon/2}-1)c^3}{3e^{\epsilon/2}(e^{\epsilon/2}+1)}\\&=\mathrm{O}\left(\frac{e^{\epsilon/2}+3}{3(e^{\epsilon/2}-1)^2}\right)\end{aligned}\tag{15}$$

图 4 可视化了在不同的隐私预算下, SDPM 的低敏感扰动方式(记为 SDPM_L)与 PM 扰动方式(也是 SDPM 的高敏感扰动方式, 记为 SDPM_H)的最大方差。可以看出 SDPM_L 的方差始终小于 PM, 且随着隐私预算的增加, SDPM_L 的优势越来越显著。

为了估计均值, 数据收集者得到 n 位用户经 SDPM 扰动的数据 $x_1^*,x_2^*,\ldots,x_n^*$ 后, 首先采用期望最大化算法 (Expectation Maximization, 简称为 EM)[14] 估计原始数据的分布, 然后计算分布的均值。

具体而言, 首先将输入域 $\mathcal{X}=[-1,1]$ 均匀地划分为 d 个子区间 $\mathcal{X}_d=\left\{S_1^{in},S_2^{in},\ldots,S_d^{in}\right\}$。同样地, 将输出域 $\mathcal{Y}=[-C,C]$ 划分为 $\mathcal{Y}_d=\left\{S_1^{out},S_2^{out},\ldots,S_d^{out}\right\}$。依据扰动函数, 即式(4)和式(14), 可得转移概率矩阵 $M_{d\times d}$, 其中矩阵元素 m_{ij} 代表输入 $x\in S_i^{in}$ 经 SDPM 扰动输出 $x^*\in S_j^{out}$ 的概率, 即:

$$m_{ij}=\Pr(x^*\in S_j^{out}\mid x\in S_i^{in})=\int_{x^*\in S_j^{out}}\Pr(x^*\mid x\in S_i^{in})\mathrm{d}x^*$$

其中 $\Pr(x^*\mid x\in S_i^{in})$ 由 SDPM 的扰动函数确定。

随后, 基于转移概率矩阵 $M_{d\times d}$ 和用户报告的噪声数据 $x_1^*,x_2^*,\ldots,x_n^*$ 在 $\mathcal{Y}_d$ 上服从的分布 $\tilde{f}$, EM 算法将通过迭代最大化以下对数似然函数 $\mathrm{LL}(f)$:

$$\mathrm{LL}(f)=\ln[\Pr(\tilde{f}\mid f)]=\ln\prod_{k=1}^{n}\Pr(x_k^*\mid f)$$

$$= \sum_{k=1}^{n} \ln[\sum_{i\in[d]} f_i \Pr(x_k^* \mid x_k \in S_i^{in})] \tag{16}$$

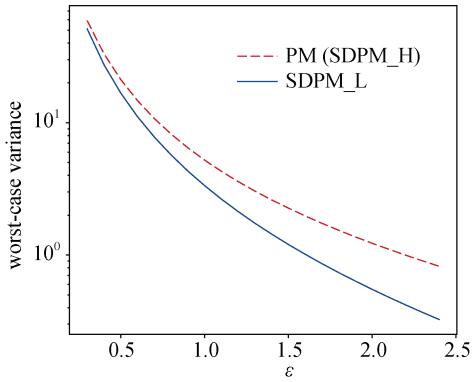


图 4　PM 与 SDPM 的方差对比

Figure 4　Comparison of variances between PM and SDPM

得到输入数据在 $\mathcal{X}_d$ 上的最大似然分布估计值 $\hat{f}$，其中，x_k^* 为第 k 名用户的扰动数据，f_i 为第 i 个输入子区间 S_i^{in} 的概率值。具体的迭代过程见算法 1.

最后，在算法 1 返回输入数据的分布 $\hat{f}$ 后，即可估计输入数据的均值 $\hat{X}$：

$$\hat{X} = \sum_{i=1}^{d} [\text{middle}(S_i^{in}) \cdot \hat{f}_i]$$

其中，$\text{middle}(S_i^{in})$ 为第 i 个输入子区间 S_i^{in} 的中值，$\hat{f}_i$ 为算法 1 返回的分布 $\hat{f}$ 中第 i 个输入子区间 S_i^{in} 的概率估计值。

算法 1.EM 算法.

输入：噪声数据 $x_1^*, x_2^*, \ldots, x_n^*$；子区间个数 d；转移概率矩阵 M；

输出：输入数据的分布估计 $\hat{f}$；

1. 由噪声数据统计得到输出子区间 $\mathcal{Y}_d$ 上的分布 $\tilde{f}$；

2. 初始化 $\hat{f}$ 为均匀分布:

$$\hat{f}^{(1)} = \left\{ \hat{f}_1^{(1)}, \hat{f}_2^{(1)}, \ldots, \hat{f}_d^{(1)} \right\};$$

3. WHILE $t < 10000$ AND $\tau > e^{\epsilon} \cdot 10^{-3}$ DO:

4. 对于每一个输入子区间 S_i^{in}，计算:

$$P_i = \sum_{j\in[d]} \tilde{f}_j \frac{\hat{f}_i^{(t)} m_{ij}}{\sum_{k\in[d]} \hat{f}_k^{(t)} m_{kj}}$$

5. 更新 $\hat{f}^{(t)}$:

$$\hat{f}_i^{(t+1)} = P_i / \sum P_i$$

6. 基于式(16)计算

$$\tau = \left| \text{LL}\left(\hat{f}^{(t+1)}\right) - \text{LL}\left(\hat{f}^{(t)}\right) \right|$$

7. END WHILE

8. RETURN $\hat{f}$

EM 算法的时间复杂度主要来自两方面。一方面，统计噪声数据在输出子区间 $\mathcal{Y}_d$ 上的分布时，需遍历噪声数据 $x_1^*, x_2^*, \ldots, x_n^*$，时间复杂度为 $O(n)$；另一方面，每次更新 $\hat{f}^{(t)}$ 中的一个元素，需要遍历 d 个子区间，因此更新 $\hat{f}^{(t)}$ 中的全部元素 t 次的时间复杂度为 $O(t \cdot d^2)$。因此总时间复杂度为 $O(n + t \cdot d^2)$。

定理 5. EM 算法最终收敛到输入数据分布 f 的最大似然估计。

证明：为了证明 EM 收敛到最大似然估计，只需要证明对数似然函数式(16)是凹函数[21]。因为 $\Pr(x_k^* \mid x_k \in S_i^{in})$ 是由 SDPM 的扰动函数确定的一个常数，所以对数似然函数(14)是一个凹函数。

证毕。

EM 算法得到的均值估计结果是有偏的，但是 EM 算法能够降低结果总体误差。定理 5 说明 EM 算法最终收敛到输入数据分布的最大似然估计，即 EM 算法通过迭代寻找到的是最有可能产生噪声数据的输入数据分布。EM 算法通过隐式地要求输入分布为合理的概率分布形式，控制输入分布的估计误差，从而进一步控制均值估计的误差。文献[8,14,22]也表明，基于 EM 的有偏估计方法与直接计算无偏估计量的误差相当或误差更小。因此 EM 算法虽然不能保证结果的无偏性，但是能保证估计误差维持在较低水平。

4.4　基于贝叶斯敌手模型的隐私分析

为了说明 SDLDP 比 ULDP 在高敏感型数据上具有更强的隐私保护能力，我们采用文献[9-11]中度量 LDP 机制隐私性的贝叶斯攻击方法，对比分析了 SDLDP 与 ULDP 抵抗攻击能力。考虑到 ULDP 框架下只有面向离散型数据的扰动机制，我们采用贝叶斯敌手分别攻击面向离散型数据的 ULDP 机制 URR 和 SDLDP 机制 SDGRR。

对于 SDGRR，根据其输出的概率密度函数(即式(11))，贝叶斯敌手 $\mathcal{A}$ 的最优攻击策略为:

$$\mathcal{A}(x_i^*) = \arg\max_{x\in\mathcal{X}} \Pr[x_i^* \mid x] = x_i^*$$

即猜测噪声数据 x_i^* 的原真实值为 x_i^* 本身。因此，对于高敏感数据 $x_i \in \mathcal{X}_H$，贝叶斯敌手攻击成功率

ACC_{SDGRR} 的期望为:

$$\begin{aligned}&\mathrm{E}[ACC_{SDGRR}]=\Pr[\mathcal{A}(x_i^*)=x_i]\\&=\Pr[x_i^*=x_i]\\&=\Pr[\mathrm{SDGRR}(x_i)=x_i]\\&=\frac{e^{\epsilon}}{|\mathcal{X}|+e^{\epsilon}-1}\end{aligned}\tag{17}$$

另外, 我们注意到贝叶斯敌手对 GRR 的最优攻击也是猜测噪声数据的真实值是其本身, 攻击成功率的期望与 SDGRR 相同[9-10]。

对于 URR, 根据其输出的概率密度函数(即式(7)), 给定噪声数据 x_i^*, 贝叶斯敌手 $\mathcal{A}$ 的最优攻击策略为:

$$\mathcal{A}(x_i^*)=\arg\max_{x\in\mathcal{X}}\Pr[x_i^*\mid x]=x_i^*$$

从而对于敏感型数据 $x_i\in\mathcal{X}_S$, 贝叶斯敌手攻击的成功率 ACC_{URR} 的期望为:

$$\begin{aligned}&\mathrm{E}[ACC_{URR}]=\Pr[\mathcal{A}(x_i^*)=x_i]\\&=\Pr[x_i^*=x_i]\\&=\Pr[\mathrm{URR}(x_i)=x_i]\\&=\frac{e^{\epsilon}}{|\mathcal{X}_s|+e^{\epsilon}-1}\end{aligned}\tag{18}$$

由于 $|\mathcal{X}|>|\mathcal{X}_s|$, 根据式(17)和式(18), 贝叶斯敌手 $\mathcal{A}$ 的攻击成功率满足:

$$\mathrm{E}[ACC_{SDGRR}]<\mathrm{E}[ACC_{URR}]$$

即贝叶斯敌手 $\mathcal{A}$ 对 SDGRR 的攻击成功率低于 URR, 从而 SDGRR 对高敏感数据的隐私保护性能高于 URR, 证明了 SDLDP 框架对高敏感数据的隐私保护性能高于 ULDP 框架。

4.5 个性化 SDLDP 机制

本小节考虑用户自定义划分高低敏感数据的场景。例如, 在统计身高均值时, 不同人群对低敏感身高区间的定义可能不同。我们提出了一种个性化 SDLDP 机制以及数据聚合方法, 允许不同用户根据不同的敏感需求, 个性化地划分高低敏感区间。

在个性化 SDLDP 机制中, 每个用户根据本地设置的高/低敏感输入域, 依据任务类型, 以 SDGRR 或 SDPM 的扰动方式处理原始数据。随后, 所有用户将扰动后的数据与高/低敏感输入域的划分方式共同发送给数据收集方。数据收集方将高/低敏感输入域的划分方式相同的用户划为一类, 并分别聚合同一类用户的数据: 对于频率估计任务, 估计同一类用户中各取值的频率; 对于均值估计任务, 利用 EM 算法估计同一类用户的均值。最后, 根据文献[23-24], 采用最小化方差的方式聚合各类用户的结果:

$$\tilde{v}=\sum_{i=1}^{k}\frac{\tilde{v}_i}{Var_i}\Big/\sum_{i=1}^{k}\frac{1}{Var_i}\tag{19}$$

其中 k 表示用户的类别数目; $\tilde{v}_i$ 表示第 i 类用户的聚合结果, 即估计的频率或均值; Var_i 表示第 i 类用户的方差: 在频率估计任务中为 SDGRR 的频率估计方差, 在均值估计任务中为 SDPM 的扰动方差。

5 实验评估

5.1 实验设计

5.1.1 数据集

为了验证本文提出的 SDGRR 和 SDPM 的表现, 本文在 2 个数据集上分别进行了频率估计和均值估计实验, 其中人口普查数据集包含离散型敏感属性, 用于估计频率, 体重身高数据集为连续型敏感属性, 用于估计均值。

人口普查数据集[25]该数据集共有 2458285 名真实人员的人口普查信息, 其中每名人员包含 18 个离散型属性。我们选择两个敏感属性—学历(Education Attainment, EA)和婚姻状态(Marriage Status, MS)分别进行频率估计实验, 其中 EA 属性具有 18 个类别, MS 属性有 7 个类别。

体重身高数据集[26]该数据集根据 25000 名儿童的成长调查数据模拟而成, 共包含 25000 条有关身高和体重的合成数据。考虑到体重和身高是典型的连续型敏感属性, 我们分别在这两个属性上进行均值估计实验。实验前对数据集进行了归一化处理, 即: 将身高和体重的取值范围缩放至[–1,1]。

5.1.2 评价指标

我们使用均方误差 *MSE*(Mean Square Error)度量估计结果的准确性, 其定义如下:

$$MSE=\frac{1}{m}\sum_{i=1}^{m}(y-y_i)^2\tag{20}$$

其中, m 为实验次数, y 为真实值, y_i 为估计值。*MSE* 越小, 则估计结果越准确。为降低随机误差影响, 我们记录了 10 次重复实验的平均值, 即 $m=10$.

我们使用全体用户中贝叶斯敌手 $\mathcal{A}$ 正确猜测出真实值的比例作为攻击成功率 *ACC* 的度量[9-10], 具体定义如下:

$$ACC=\frac{\sum_{i=1}^{n}\mathbb{I}(\mathcal{A}(x_i^*)=x_i)}{n}\tag{21}$$

其中, n 是用户数量; x_i^* 为第 i 名用户的噪声数据, x_i 为第 i 名用户的真实数据; $\mathbb{I}(\mathcal{A}(x_i^*)=x_i)$ 为指示函数, $\mathcal{A}(x_i^*)=x_i$ 表示敌手猜测正确, 此时 $\mathbb{I}(\mathcal{A}(x_i^*)=x_i)=1$,

否则若 $\mathcal{A}(x_i^*) \neq x_i$，则 $\mathbb{I}(\mathcal{A}(x_i^*) = x_i)$=0．$ACC$ 越高，则表明扰动机制的隐私保护水平越低。

5.1.3 对比方法

GRR 扰动离散型数据的经典 LDP 机制，适用于频率估计任务，具体介绍参见 3.3 节；

URR ULDP 框架下的离散型数据扰动机制，具体介绍参见 3.4 节。URR-25/50/75 分别代表敏感取值域占比为 25%/50%/75%的 URR 方法；

SDGRR 本文提出的 SDLDP 框架下的离散型数据扰动机制，具体介绍参见 4.2 节。SDGRR-25/50/75 分别代表高敏感取值域占比为 25%/50%/75%的 SDGRR, SDGRR-Personalized 代表 4.5 节提出的个性化 SDGRR；

PM 适用于扰动连续型数据的经典 LDP 机制，具体介绍参见 3.3 节；

SDPM 本文提出的满足 SDLDP 的连续型数据扰动机制，具体介绍参见 4.3 节；SDPM-25/50/75 分别代表高敏感取值域占比为 25%/50%/75%的 SDPM, SDPM -Personalized 代表 4.5 节提出的个性化 SDPM。

5.2 实验结果与分析

5.2.1 频率估计

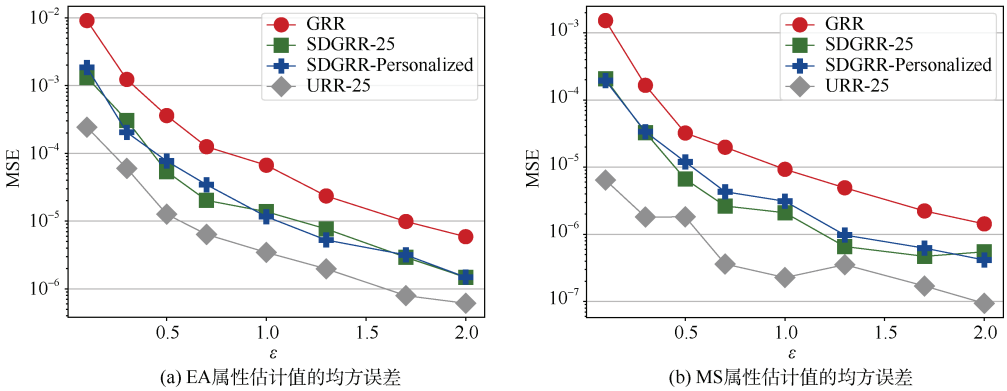
首先我们在人口普查数据集的 EA 和 SA 两个属性上，比较了 LDP 框架下的 GRR，ULDP 框架下的 URR，以及本文提出的 SDGRR 和个性化 SDGRR(记为 SDGRR-Personalized)的频率估计均方误差。我们为 URR 和 SDGRR 设置高敏感取值域占比(即 $|\mathcal{X}_S|/|\mathcal{X}|$ 和 $|\mathcal{X}_H|/|\mathcal{X}|$)为 25%，分别表示为 URR-25 与 SDGRR-25，个性化 SDGRR 算法设置高敏感取值域占比为 25%，50%，75%的用户比例分别为 40%，30%，30%。对比结果如图 5 所示，在不同隐私预算下，SDGRR-25 与 SDGRR-Personalized 的频率估计误差较 GRR 约下降 1 个数量级，同时高于 URR-25 约 1 个数量级。这是因为在相同的隐私预算下，相比于 GRR 对任意输入采取相同的扰动方式，SDGRR 降低了低敏感数据的扰动程度，将低敏感数据的直接输出隐藏在其他噪声数据中，提升了噪声数据整体的可用性，而 URR 忽略了低敏感数据的隐私，依概率直接暴露低敏感数据，因而误差更小。结果也表明，SDGRR-Personalized 几乎能达到和 SDGRR 相同的误差，这证明了 4.5 节提出的个性化机制的有效性。

另外，我们研究了高/低敏感域的划分比例对频率估计精度的影响，对比了高敏感取值域占比不同时 SDGRR 方法的表现，图 6 展示了在 EA 和 SA 两个离散型属性上的结果，其中 SDGRR-25/50/75 表示高敏感取值域占比为 25%/50%/75%的 SDGRR 方法。结果显示，SDGRR 中高敏感取值域占比越高，误差越大，SDGRR-75 表现最接近 GRR；反之，误差下降，SDGRR-25 最低，接近 URR-25。这是因为与 URR 松弛化方案类似，SDGRR 的性能提升主要来自松弛化低敏感数据的隐私保护。

SDGRR 与 URR 的差异在于，URR 相较于 GRR 同时降低了高、低敏感数据的隐私保护程度以换取可用性的提升，而 SDGRR 能够保证高敏感数据的隐私保护程度与 GRR 相同，即优于 URR，同时满足可用性介于两者之间。例如，在高敏感取值域占比为 25%时，SDGRR 的均方误差较 URR 高约一个数量级，而较 GRR 降低了约一个数量级(如图 5 所示)，在隐私性和可用性之间取得了较好的平衡。

5.2.2 均值估计

我们分别在身高体重数据集的“体重”和“身高”属性上对比了均值估计任务上 PM 与 SDPM(即 SDPM-50)以及个性化 SDPM(即 SDPM-Personalized)的均方误差。实验前将数据取值域归一化至[–1,1]。

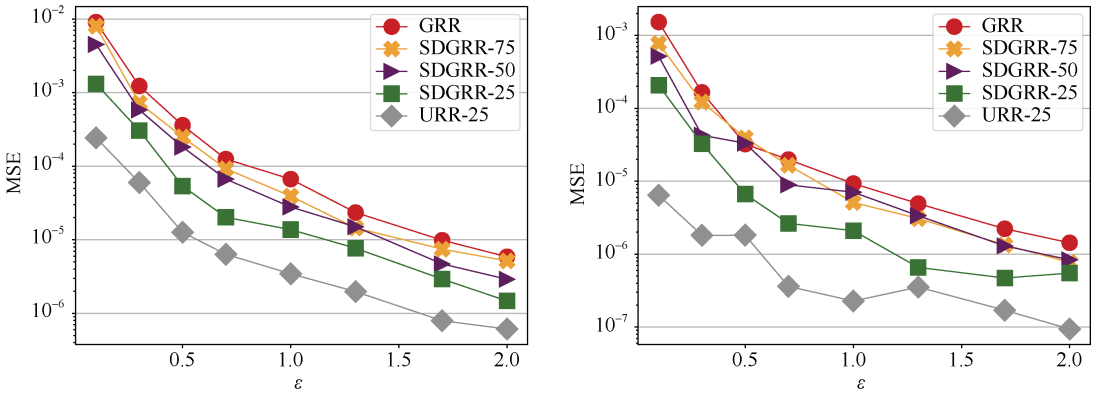


(a) EA属性估计值的均方误差
(a) The MSE of Estimations on EA Attribute

(b) MS属性估计值的均方误差
(b) The MSE of Estimations on MS Attribute

图 5 不同隐私预算下频率估计准确性对比，其中高敏感取值占比 25%

Figure 5 Comparison of frequency estimation accuracy varying epsilon with high sensitive values reaching 25%



(a) EA属性估计值的均方误差
(a) The MSE of Estimations on EA Attribute

(b) MS属性估计值的均方误差
(b) The MSE of Estimations on MS Attribute

图 6　设置不同的高敏感取值占比, SDGRR 的频率估计准确性对比

Figure 6　Comparison of frequency estimation accuracy with different radio of high sensitive values on SDGRR

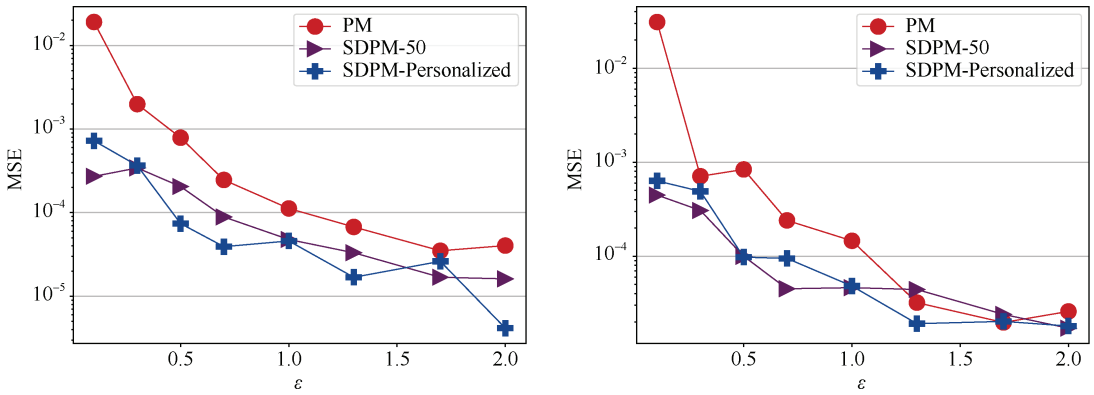
不失一般性地，设置占比为 p 的低敏感区间为 $[-p,p]$，相应地设置高敏感区间为 $[-1,-p)\cup(p,1]$。SDPM-50 中高敏感区间占比设为 50%，SDPM-Personalized 设置高敏感区间占比为 25%，50%，75%的用户比例分别为 40%, 30%, 30%。结果如图 7 所示。在不同隐私预算下，SDPM-50 与 SDPM-Personalized 表现均优于 PM，且 ϵ 越小，效果越显著。当 $\epsilon=0.1$ 时，SDPM-50 相较于 PM 均方误差降低约 2 个数量级。与 SDGRR 类似，SDPM 的高可用性来源于敏感数据的分级保护，通过降低低敏感数据的扰动程度，提高数据可用性。此外，结果表明 SDPM-Personalized 的误差表现与 SDPM-50 接近，总体低于 PM，验证了个性化机制的有效性。

另外，我们对比了 SDPM 在高敏感区间占比分别为 25%, 50%, 75%的三种设置下的均方误差。图 8 分别显示了在“体重”和“身高”两个属性上的对比结果，图中 SDPM 的后缀数字表示高敏感区间占比。结果显示，在所有的隐私预算下, SDPM-25 表现均优于 SDPM-50 与 SDPM-75，高敏感区间越小，则均值估计结果越准确。这是由于 SDLDP 性能提升来自对低敏感数据的松弛，高敏感区间越小代表松弛化程度越高，因而数据可用性越高。

SDPM 对高敏感数据的隐私保护程度与 PM 相同，而可用性有较大提升，并且在高敏感数据的隐私预算较小时提升效果更显著，例如，当隐私预算为 0.1 时，SDPM 相较于 PM 均方误差降低约 2 个数量级(如图 7 所示)。

5.2.3　攻击实验

我们分别在人口普查数据集的 EA 和 SA 两个属性上对比了贝叶斯敌手对 SDGRR、URR 和 GRR 中高敏感型数据的攻击成功率 *ACC* 。图 9 展示了实验结果。在不同的隐私预算下，贝叶斯敌手对 SDGRR

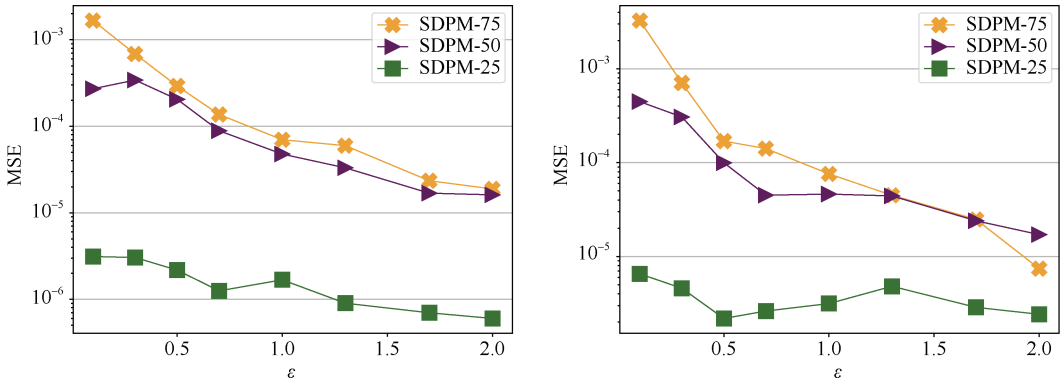


(a) 体重估计值的均方误差
(a) The MSE of Estimations on Weight Attribute

(b) 身高估计值的均方误差
(b) The MSE of Estimations on Height Attribute

图 7　不同隐私预算下的均值估计准确性对比

Figure 7　Comparison of mean estimation accuracy varying epsilon

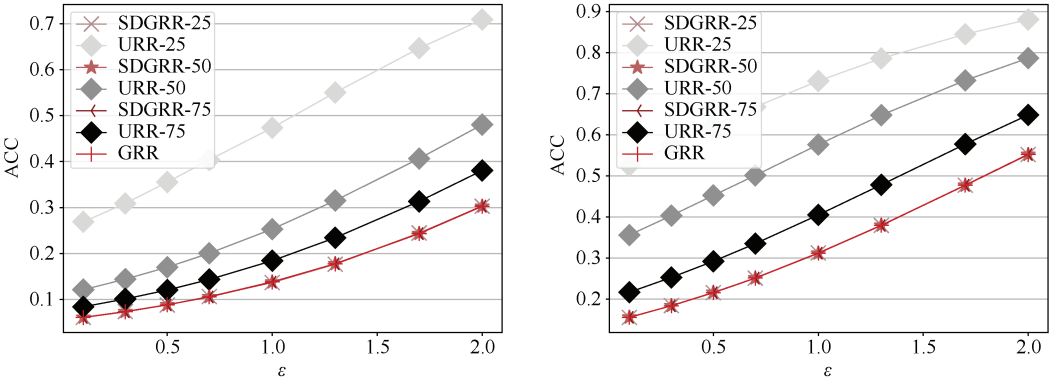


(a) 体重估计值的均方误差
(a) The MSE of Estimations on Weight Attribute

(b) 身高估计值的均方误差
(b) The MSE of Estimations on Height Attribute

图 8　设置不同的高敏感区间占比, SDPM 的均值估计准确性对比

Figure 8　Comparison of mean estimation accuracy with different radio of high sensitive range on SDPM



(a) EA属性的攻击成功率
(a) Adversarial Success Rate on EA Attribute

(b) MS属性的攻击成功率
(b) Adversarial Success Rate on MS Attribute

图 9　贝叶斯敌手对高敏感数据的攻击成功率对比

Figure 9　Comparison of adversarial success rate of bayesian adversary to high sensitive data

的攻击成功率与 GRR 基本相等, 显著低于 URR, 并且高敏感取值域占比越低则 SDGRR 抵抗攻击的优势越明显, 在高敏感取值域占比为 25%时, SDGRR 相较于 URR 的攻击成功率下降约 30%。这表明 SDGRR 对高敏感数据的隐私保护效果优于 URR, 与我们的理论分析一致(参见 4.4 节)。另外, 贝叶斯敌手对 MS 属性的攻击成功率一致高于 EA 属性, 这是因为类别数目越少时, 敌手攻击成功的概率越大。

6　结束语

本文提出了一种支持数据敏感程度分级的松弛化 LDP 框架 SDLDP, 其利用敏感数据的隐私保护需求不尽相同的性质, 通过为低敏感数据降低隐私保护程度, 实现数据可用性与隐私性的更优平衡。另外, SDLDP 同时支持离散型和连续型数据的隐私化收集与统计分析, 本文提出了适用于频率估计任务的 SDGRR, 以及适用于均值估计任务的 SDPM, 理论分析和实验结果均表明新方法的准确性显著优于 LDP 下的先进机制。下一步工作将考虑设计支持复杂数据类型的 SDLDP 机制, 将 SDLDP 框架应用在更多的实际场景中。

参考文献

[1] Duchi J C, Jordan M I, Wainwright M J. Local Privacy and Statistical Minimax Rates[C]. *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 2013: 429-438.

[2] Wang T, Blocki J, Li N, et al. Locally differentially private protocols for frequency estimation[C]. *26th USENIX Security Symposium*, 2017: 729-745.

[3] Wang N, Xiao X K, Yang Y, et al. Collecting and Analyzing Multidimensional Data with Local Differential Privacy[C]. *2019 IEEE 35th International Conference on Data Engineering*, 2019: 638-649.

[4] Qin Z, Yang Y, Yu T, et al. Heavy Hitter Estimation over Set-Valued Data with Local Differential Privacy[C]. *The 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016: 192-203.

[5] Apple. Apple differential privacy team, learning with privacy at scale [EB/OL].[2017-09-08]. https://machinelearning.apple.com/

docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf.

[6] Erlingsson Ú, Pihur V, Korolova A. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response[C]. *The 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014: 1054-1067.

[7] Chen R, Li H R, Qin A K, et al. Private Spatial Data Aggregation in the Local Setting[C]. *2016 IEEE 32nd International Conference on Data Engineering*, 2016: 289-300.

[8] Murakami T, Kawamoto Y. Utility-Optimized Local Differential Privacy Mechanisms for Distribution Estimation[C]. *28th USENIX Security Symposium*, 2019: 1877-1894.

[9] Arcolezi H H, Gambs S, Couchot J F, et al. On the Risks of Collecting Multidimensional Data under Local Differential Privacy[J]. *Proceedings of the VLDB Endowment*, 2023, 16(5): 1126-1139.

[10] Gursoy M E, Liu L, Chow K H, et al. An Adversarial Approach to Protocol Analysis and Selection in Local Differential Privacy[J]. *IEEE Transactions on Information Forensics and Security*, 2022, 17: 1785-1799.

[11] Gursoy M E, Tamersoy A, Truex S, et al. Secure and Utility-Aware Data Collection with Condensed Local Differential Privacy[J]. *IEEE Transactions on Dependable and Secure Computing*, 2021, 18(5): 2365-2378.

[12] Warner S L. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias[J]. *Journal of the American Statistical Association*, 1965, 60(309): 63-69.

[13] Kairouz P, Oh S, Viswanath P. Extremal Mechanisms for Local Differential Privacy[J]. *Advances in neural information processing systems*, 2014, 27.

[14] Li Z T, Wang T H, Lopuhaä-Zwakenberg M, et al. Estimating Numerical Distributions under Local Differential Privacy[C]. *The 2020 ACM SIGMOD International Conference on Management of Data*, 2020: 621-635.

[15] Ye Q Q, Hu H B, Meng X F, et al. PrivKV: Key-Value Data Collection with Local Differential Privacy[C]. *2019 IEEE Symposium on Security and Privacy*, 2019: 317-331.

[16] Gu X L, Li M, Cheng Y Q, et al. PCKV: Locally Differentially Private Correlated Key-Value Data Collection with Optimized Utility[EB/OL]. 2019: 1911.12834. https://arxiv.org/abs/1911.12834v1.

[17] Wang T H, Li N H, Jha S. Locally Differentially Private Frequent Itemset Mining[C]. *2018 IEEE Symposium on Security and Privacy*, 2018: 127-143.

[18] Alvim M S, Chatzikokolakis K, Palamidessi C, et al. Metric-Based Local Differential Privacy for Statistical Applications[EB/OL]. 2018: 1805.01456. https://arxiv.org/abs/1805.01456v1.

[19] Gu X L, Li M, Xiong L, et al. Providing Input-Discriminative Protection for Local Differential Privacy[C]. *2020 IEEE 36th International Conference on Data Engineering*, 2020: 505-516.

[20] Ye M, Barg A. Optimal Schemes for Discrete Distribution Estimation under Locally Differential Privacy[J]. *IEEE Transactions on Information Theory*, 2018, 64(8): 5662-5676.

[21] Bilmes J A. A Gentle Tutorial of the Em Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models[J]. *International computer science institute*, 1998, 4(510): 126.

[22] Murakami T, Hino H, Sakuma J. Toward Distribution Estimation under Local Differential Privacy with Small Samples[J]. *Proceedings on Privacy Enhancing Technologies*, 2018, 2018(3): 84-104.

[23] Ye Y T, Zhang M, Feng D G, et al. Multiple Privacy Regimes Mechanism for Local Differential Privacy[M]. Database Systems for Advanced Applications. Cham: Springer International Publishing, 2019: 247-263.

[24] Zhang Z K, Wang T H, Li N H, et al. CALM: Consistent Adaptive Local Marginal for Marginal Release under Local Differential Privacy[C]. *The 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018: 212-229.

[25] 1990 US Census Data[DB/OL]. 2013. http://kdd.ics.uci.edu/ databases/census1990/USCensus1990.data.txt.

[26] SOCR Data Dinov 020108 HeightsWeights [DB/OL]. http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Dinov_020108_HeightsWeights.

陈亚青 CCF 学生会员, 于 2020 年在西安电子科技大学计算机科学与技术专业获得工学学士学位。现在中国科学院软件研究所可信计算与信息保障实验室网络空间安全专业攻读博士学位。研究领域为大数据安全与隐私保护。研究兴趣包括: 差分隐私、联邦学习。Email: yaqing2021@iscas.ac.cn

张敏 CCF 高级会员, 于 2007 年在中国科学院软件研究所计算机应用技术专业获得博士学位。现任中国科学院软件研究所可信计算与信息保障实验室研究员。研究领域为大数据安全与隐私保护。研究兴趣包括: 差分隐私、可搜索加密、AI 隐私保护。Email: zhangmin@iscas.ac.cn

叶宇桐 CCF 专业会员, 于 2022 年在中国科学院大学计算机应用技术专业获得博士学位。现任中国科学院软件研究所可信计算与信息保障实验室助理研究员。研究领域为大数据安全与隐私保护。研究兴趣包括: 差分隐私、机器学习。Email: yutong2017@iscas.ac.cn

舒波文 CCF 学生会员, 于 2021 年在中国科学院大学计算机科学与技术专业获得工学学士学位。现在中国科学院软件研究所可信计算与信息保障实验室网络空间安全专业攻读博士学位。研究领域为大数据安全与隐私保护。研究兴趣包括差分隐私、访问模式隐藏。Email: bowen2021@iscas.ac.cn