局部差分隐私约束的关联属性不变后随机响应扰动

杨高明,朱海明,方贤进,苏树智
(安徽理工大学计算机科学与工程学院,安徽淮南 232001)

摘 要: 本文研究敏感属性与部分准标识符属性存在相关时,如何有效减小重构攻击导致的隐私泄漏风险. 首先,用互信息理论寻找原始数据集中对敏感属性具有强依赖关系的准标识符属性,为精确扰动数据属性提供理论依据;其次,针对关联属性和非关联属性,应用不变后随机响应方法分别对某个数据属性或者属性之间的组合进行扰动,使之满足局部 ε-差分隐私要求,并理论分析后数据扰动对隐私泄露概率和数据效用的影响;最后,实验验证所提算法的有效性和处理增量数据的能力,理论分析了数据结果. 由实验结果可知,算法可以更好地达到数据效用和隐私保护的平衡.

关键词: 局部差分隐私; 不变后随机响应; 数据重构; 数据扰动; 隐私保护
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2019)05-1079-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.05.015

Invariant Post-Random Response Perturbation for Correlated Attributes Under Local Differential Privacy Constraint

YANG Gao-ming, ZHU Hai-ming, FANG Xian-jin, SU Shu-zhi
(*School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui* 232001, *China*)

Abstract: We investigate in this paper how to effectively reduce the risk of privacy leakage caused by refactoring attacks when the sensitive attributes and some quasi-identifier attributes are correlated. Firstly, the mutual information theory is used to find the quasi-identifier attributes which have strong dependence on the sensitive attributes in the original dataset, which provides a theoretical basis for accurately perturbing the data attributes. Secondly, for the correlated attributes and the non-correlated attributes, the invariant random response method is applied to perturb a certain data attribute or a combination of data attributes to satisfy the local ε-differential privacy requirement. Theoretical analysis of the impact of data perturbations on privacy leakage probability and data utility is also conducted. Finally, the experiment verifies the effectiveness of the proposed algorithm and its ability to process incremental data. The experimental results demonstrate that the algorithm can achieve a better balance between data utility and privacy protection.

Key words: local differential privacy; invariant post-random response; data reconstruction; data perturbation; privacy protection

1 引言

随着社会的发展和技术的进步,越来越多的服务和产品以用户为中心建立;服务商为提高服务质量,需要收集、使用和发布用户数据,这不可避免的导致用户隐私泄漏,为避免用户信息隐私泄露[1],许多隐私保护模型被提出,如 k-匿名[2]及其扩展模型对数据进行概化、隐匿处理[3,4],可以某种程度上保护用户隐私信息,但恶意的攻击者可以通过准标识符属性等背景信息,识别某个用户的身份或者敏感属性,从而导致隐私泄漏. 另外,匿名化隐私保护模型主要处理关系数据,涉及到身份标识符属性(ID)、准标识符(QI)属性和敏感

收稿日期:2018-09-14;修回日期:2018-11-21;责任编辑:蓝红杰
基金项目:国家自然科学基金(No. 61572034, No. 61806006);安徽省高校自然科学基金(No. KJ2018A0083, No. KJ2014A061);安徽省重大科技专项(No. 18030901025)

(SA)属性[5].

由于匿名化方法自身的缺陷,Dwork 提出差分隐私保护模型,它具有理论上的严谨性,无需考虑攻击者背景知识. 进一步,随机响应(Randomized Response,RR)被引入局部差分隐私保护技术中[6],这方面的初始研究主要关注构造不同的扰动矩阵. 如 Xiao 等人[7]提出多层次扰动的解决方案,避免不同的接收者通过共享数据获取超越权限的隐私信息从而导致隐私泄漏问题. 在后随机响应(Post Randomization Method,PRAM)方面,Nayak 等人[8]针对发布数据的关键分类变量可能导致隐私泄漏问题,提出一种新的方法衡量识别风险并通过无偏的后随机化方案来降低隐私泄漏的风险.

以上方法主要考虑属性相互独立或者完全相关的情况,忽略了部分属性相关在重构敏感属性中的隐私风险. 实际上属性之间往往存在部分依赖关系,当攻击者获知 QI 属性对 SA 属性进行推理披露时,并非所有的属性都会导致隐私泄露,仅仅属性之间的强依赖关系才会导致敏感属性的泄露. 为有效保护用户隐私,增加数据效用,在考虑数据集 QI 属性与 SA 属性存在依赖关系的基础上,本文提出使用不变后随机响应方法对数据集进行扰动,以满足局部差分隐私要求,具体贡献为:

(1)为针对性的对属性进行扰动,提出根据原始数据集 QI 属性与 SA 属性关系强度进行属性划分的方法,把属性划分为不同的级别,以实现不同的保护.

(2)针对 QI 属性与 SA 属性的不同依赖关系,设计不同的后随机扰动方法,理论分析攻击者重构数据导致的隐私泄漏风险.

(3)实验验证所提方法的有效性,并分析其结果,主要测试数据的发布质量和执行效率,与随机响应的其他方法对比,说明所提方法能更好地保护敏感属性,保持数据效用.

2 基本概念

假设原始数据集 DS 有 n 条记录,划分为 QI 属性和 SA 属性,即$\{QI,SA\}$的形式,其中 QI 属性表示为$\{A_i \mid i=1,\cdots,m\}$,SA 属性表示为$\{S_i \mid ,i=1,\cdots,k\}$,$m$ 和 k 分别为准标识符属性和敏感属性的个数. 属性 $A_i(S_i)$ 有 d_i (d_s)个不同值,用 φ_i 表示 A_i 的域 $\varphi_i=\{\varphi_i^1,\cdots,\varphi_i^z,\cdots,\varphi_i^{di}\}$,$1\leqslant z\leqslant d_i$,则 QI 属性的域为 $\varphi_{QI}=\varphi_1\times\cdots\times\varphi_m$,$\varphi_s$ 表示敏感属性域. 为简单起见,本文主要以单敏感属性为例讨论,即$\{S_i\}$取 $i=1$,对多敏感属性的处理可以按照 QI 属性的处理方式进行扩充.

2.1 后随机响应

在随机扰动基础上,Kooiman 等人提出 PRAM,在数学上 PRAM 与 RR 类似[9],都在保护用户隐私信息基础上研究如何提高统计精度. 为避免 PRAM 需要向外部的用户提供数据转移概率矩阵,简化数据分析,不变 PRAM[10]也被提出. 其应用场景如图 1 所示.

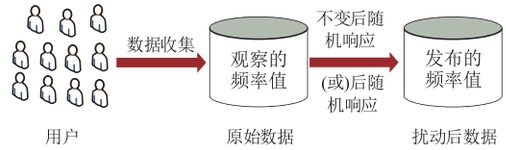


图1 后随机响应总体框架

为便于形式化描述,用 X 表示原始数据集 DB 中的一个属性变量(分类属性),其可以是 QI 属性也可以是 SA 属性(如 A_i,S_i),并用 $\bar{X}$ 表示对应的扰动变量,用 φ_x 表示 X 的域,用 d_x 表示不同属性值个数$\varphi_x=\{\varphi_x^1,\cdots,\varphi_x^z,\cdots,\varphi_x^{dx}\}$,$1\leqslant z\leqslant d_x$. 设转移概率矩阵为 $\boldsymbol{P}=(p_{ij})$,其中 $1\leqslant i,j\leqslant d_x$,$\sum_j p_{ij}=1$,矩阵 $\boldsymbol{P}$ 的元素表示以概率 p_{ij} 随机将原始属性变量的值 φ_x^i 改变为 φ_x^j,即 $p_{ij}=\Pr(\bar{X}=\varphi_x^j \mid X=\varphi_x^i)$.

不变 PRAM 对矩阵 $\boldsymbol{P}$ 的选择要施加额外条件,要满足马尔可夫矩阵以及方程(1),若用 T_X 表示原始数据中属性 X 的频率,则有:

$$\boldsymbol{P}T_X=T_X \tag{1}$$

2.2 局部差分隐私

局部差分隐私(Local Differential Privacy,LDP)继承了中心化差分隐私的组合特性,并利用随机响应机制抵御来自不可信第三方的隐私攻击,其定义如下.

定义 1 ε-局部差分隐私 给定多个用户,每个用户对应一条数据记录,给定隐私保护算法 F 及其定义域 Dom(F)和值域 Ran(F),若算法 F 在任意两条用户记录 l 和 l'($l,l'\in$ Dom(F))上得到相同输出结果 l^*($l^*\in$ Ran(F)),则 F 满足 ε-局部差分隐私:

$$\Pr[F(l)=l^*]\leqslant e^{\varepsilon}\times\Pr[F(l')=l^*] \tag{2}$$

2.3 互信息

设原始数据的分类属性 X 值域为$\varphi_x=\{\varphi_x^1,\cdots,\varphi_x^z,\cdots,\varphi_x^{dx}\}$,$1\leqslant z\leqslant d_x$,每个分类属性值在该分类属性域的概率为 $\Pr(\varphi_i^z)$,则 X 的熵定义为:

$$H(X)=-\sum_{z=1}^{d_x}\Pr(\varphi_x^z)\log\Pr(\varphi_x^z) \tag{3}$$

就具体原始数据的 QI 属性 A_i,$i\in\{1,2,\cdots,m\}$ 和 SA 属性 S 而言,两者间的互信息可定义为

$$I(A_i;S)=\sum_{z=1}^{d_i}\sum_{z=1}^{d_s}\Pr(\varphi_i^z,\varphi_s^z)\log\frac{\Pr(\varphi_i^z,\varphi_s^z)}{\Pr(\varphi_i^z)\Pr(\varphi_s^z)} \tag{4}$$

2.4 隐私保护与数据效用度量

本文使用期望比衡量 PRAM 隐私泄露的风险,其定义为扰动数据中观察值等于原始数据值预期记录数,和观察值不等于原始数据值预期记录数的比,形式化描述为:

$$\mathrm{ER}(\varphi_x^i) = \frac{p_{ii}T_X(\varphi_x^i)}{\sum_{i\neq j} p_{ij}T_X(\varphi_x^j)}, \text{for} \quad i,j = 1,\cdots,d_x \quad (5)$$

$\mathrm{ER}(\varphi_x^i)$的值越小,$\tilde{X} = \varphi_x^i$ 的扰动记录越不可能属于该值,扰动数据越安全.

此外,文中还采用 KL-散度比较两个离散变量的概率分布接近程度,设分类变量 X 扰动前后的离散概率分布为 $\Pr(\varphi_x^i)$ 和 $\Pr(\widehat{\varphi_x^i})$,则其 KL-散度计算如下:

$$\mathrm{D_{KL}} = \sum_{i=1}^{d_x} \Pr(\varphi_x^i)\log\frac{\Pr(\varphi_x^i)}{\Pr(\varphi_x^i)} \quad (6)$$

3　敏感属性重构攻击保护

PRAM 将用户的属性值按一定的概率随机化以保护个体隐私,而攻击者为查找目标 R 的敏感信息,可通过已有信息对个人数据记录进行连接并重构相关数据. 如果这种敏感属性的分布倾斜,个人的隐私信息就会被泄漏,这种攻击被称为"重构攻击"[11].

3.1　扰动的选择

为在保护用户敏感属性的同时尽可能保持数据效用,需要先按照 QI 属性与 SA 属性的依赖度对 QI 属性进行划分,选取 SA 属性与具有强依赖关系的 QI 属性进行扰动,此处依赖度以属性间的平均互信息计算,定义如下.

定义 2　依赖度　数据集 DS 的准标识符属性 A_i 对敏感属性 S 的依赖度为:

$$\mathrm{I}_A(A_i;S) = \sum_{z=1}^{d_{A_i}} \Pr(\varphi_{A_i}^z|\varphi_s^z)\mathrm{I}(\varphi_{A_i}^z|\varphi_s^z) \sum_{z=1}^{d_s} \Pr(\varphi_s^z) \quad (7)$$

为选择依赖度最大值对应的 A_i 属性进行扰动,首先依次计算 QI 与 SA 属性间的依赖度,当 QI 属性存在数值型属性时,要先对该属性进行离散化. 之后再用式(7)计算出 QI 与 SA 属性间所有的依赖度,并相互比较,返回具有最大依赖度的属性,具体过程见算法 1.

算法 1　选择对敏感属性 S 依赖度最大的 QI 属性

输入:原始数据集 DS{QI,SA},设置区间划分的数目C_i
输出:获得最大依赖度对应的 QI 属性
1. 统计数据集 DS 的属性值φ_i,φ_s 数目,计算相应频率
2. for ($j=0;j<ds;j++$)
3.　　for ($i=0;i<d_{A_i};i++$)
4.　　if φ_i 是数值型
5.　　$\varphi_i = \mathrm{Splitting}(\varphi_i, C_i)$;　//对数值型属性离散化
6.　　endif
7.　　使用式(7)计算 $\mathrm{I}_A(A_i;S)$;
8.　endfor
9. endfor
10. 将依赖度值按大小排列,输出最大依赖度对应的 QI 属性

算法 1 的行 1 主要对数据进行统计工作,获得每个属性域值的频率,行 2 - 8 是将准标识符属性与敏感属性两两组合用式(7)求得属性变量的依赖度,若是数值型属性则先对其离散化.

3.2　失真矩阵的构造

离散属性变量有二值属性和多值属性之分,此处先以二值属性为例解释失真矩阵的构造,再给出多值属性扰动矩阵的构造. 对二值属性变量用 u 和 v 表示属性的两个值,用 p_u,p_v 表示对应值的转移概率,则相应转移概率矩阵可构造为以下形式.

$$\boldsymbol{P}_B = \begin{bmatrix} p_u & 1-p_V \\ 1-p_u & p_v \end{bmatrix}$$

此处矩阵 $\boldsymbol{P}_B$ 满足马尔科夫矩阵. 文中使用二次后随机响应方式实现局部差分隐私,主要思想是:首先对原始数据中属性变量 X 使用矩阵 $\boldsymbol{P}$ 扰动,由第一次扰动后的数据估计原始数据中 X 的概率分布,然后用这个概率分布构造出第二次扰动需要的转移概率矩阵$\boldsymbol{P}$,在一次扰动后的数据上使用$\boldsymbol{P}$将干扰数据转换回来. 经过两次扰动后的数据与原始数据并不完全相同,可以看作是应用不变 PRAM 的结果[10]. 为保证扰动满足 ε-局部差分隐私,对原始数据 X 进行首次扰动时,矩阵 $\boldsymbol{P}$ 采用阶梯机制(staircase mechanism)[12],其中对二值属性 $d_x=2$,p_u,p_v的形式为:$p_u = p_v = \mathrm{e}^{\varepsilon}/(1+\mathrm{e}^{\varepsilon})$,则有:

$$\boldsymbol{P}_B \cdot \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p_u \cdot u + (1-p_v)\cdot v \\ (1-p_u)\cdot u + p_v \cdot v \end{bmatrix} = \begin{bmatrix} \dfrac{\mathrm{e}^{\varepsilon}(u+v)}{1+\mathrm{e}^{\varepsilon}} \\ \dfrac{u+v}{1+\mathrm{e}^{\varepsilon}} \end{bmatrix}$$

设第一次扰动后对应的数据记为 $\tilde{X}_1$,结合扰动矩阵 $\boldsymbol{P}_B$的逆,可以估计出原始数据集的概率分布. 即对原始数据 X 的估计值 $\hat{T}_X$ 可以表示为:

$$\hat{T}_X = \boldsymbol{P}_B \cdot \begin{bmatrix} \mathrm{e}^{\varepsilon}(u+v)/(1+\mathrm{e}^{\varepsilon}) \\ (u+v)/(1+\mathrm{e}^{\varepsilon}) \end{bmatrix}$$

用 $p_{\tilde{u}}$表示属性变量 $\tilde{X}_1$ 对应的原始数据值为 u 的概率,则有:

$$p_{\tilde{u}} = \Pr(X=u|\tilde{X}_1=u) = \frac{p_u\hat{T}_X(u)}{p_u\hat{T}_X(u)+(1-p_v)\hat{T}_X(v)}$$

得到对二值属性第二次扰动的转移概率矩阵 $\boldsymbol{P}$:

$$\boldsymbol{P} = \begin{bmatrix} p_{\tilde{u}} & 1-p_{\tilde{v}} \\ 1-p_{\tilde{u}} & p_{\tilde{v}} \end{bmatrix} = \begin{bmatrix} \dfrac{p_u\hat{T}_X(u)}{p_u\hat{T}_X(u)+(1-p_v)\hat{T}_X(v)} & \dfrac{(1-p_u)\hat{T}_X(u)}{p_v\hat{T}_X(v)+(1-p_u)\hat{T}_X(u)} \\ \dfrac{(1-p_v)\hat{T}_X(v)}{p_u\hat{T}_X(u)+(1-p_v)\hat{T}_X(v)} & \dfrac{p_v\hat{T}_X(v)}{p_v\hat{T}_X(v)+(1-p_u)\hat{T}_X(u)} \end{bmatrix}$$

再将此矩阵应用于第一次扰动后的数据,即可完成对二值属性的不变后随机扰动. 用 $\tilde{X}_2$ 表示两次扰动后数

据中对应 X 的变量.

$$\tilde{X}_2 = \boldsymbol{P} \cdot \begin{bmatrix} p_u \cdot u + (1-p_v) \cdot v \\ (1-p_u) \cdot u + p_v \cdot v \end{bmatrix}$$

下面给出对多值属性变量扰动矩阵的构造,对于多值属性变量而言有$d_x > 2$,设其扰动矩阵为 $\boldsymbol{P}_m$:

$$\boldsymbol{P}_m = \begin{bmatrix} p_{v_1 v_1} & p_{v_1 v_2} & \cdots & p_{v_1 v_{d_x}} \\ p_{v_2 v_1} & p_{v_2 v_{12}} & \cdots & p_{v_2 v_{d_x}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{v_{d_x} v_1} & p_{v_{d_x} v_2} & \cdots & p_{v_{d_x} v_{d_x}} \end{bmatrix} \tag{8}$$

多值属性扰动与二值属性处理方式类似,此时扰动候选值为多个,需要扩展二值不变后随机响应算法. 对于任意属性变量值 $\varphi_x^j \in X$,其响应输出的方式如下:

$$\Pr(\tilde{X} \mid X) = \begin{cases} e^{\varepsilon}/(d_x - 1 + e^{\varepsilon}) & \text{if } \tilde{X} = X \\ 1/(d_x - 1 + e^{\varepsilon}) & \text{if } \tilde{X} = X \end{cases} \tag{9}$$

即按照 $e^{\varepsilon}/(d_x - 1 + e^{\varepsilon})$ 的概率响应输出真实值,以 $1/(d_x - 1 + e^{\varepsilon})$ 的概率响应输出剩下 $d_x - 1$ 个结果的任意一种. 令 $p_{ii} = \dfrac{e^{\varepsilon i}}{d_x - 1 + e^{\varepsilon i}}$, $i = 1, 2, \cdots, d_x$, 带入式(8)可得:

$$\boldsymbol{P}_m = \begin{bmatrix} \dfrac{e^{\varepsilon i}}{d_x - 1 + e^{\varepsilon i}} & \dfrac{1}{d_x - 1 + e^{\varepsilon i}} & \cdots & \dfrac{1}{d_x - 1 + e^{\varepsilon i}} \\ \dfrac{1}{d_x - 1 + e^{\varepsilon i}} & \dfrac{e^{\varepsilon i}}{d_x - 1 + e^{\varepsilon i}} & \cdots & \dfrac{1}{d_x - 1 + e^{\varepsilon i}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{1}{d_x - 1 + e^{\varepsilon i}} & \dfrac{1}{d_x - 1 + e^{\varepsilon i}} & \cdots & \dfrac{e^{\varepsilon i}}{d_x - 1 + e^{\varepsilon i}} \end{bmatrix}$$

用 $P_{j\bar{i}}$ 表示多值属性变量 X 原始值为 φ_x^j 的概率,则有:

$$P_{j\bar{i}} = \Pr(X = \varphi_x^j \mid \tilde{X}_1 = \varphi_x^i) = \frac{p_{ij} \hat{T}_X(\varphi_x^i)}{\sum_j p_{ij} \hat{T}_X(\varphi_x^j)} \tag{10}$$

此时得到对多值属性第二次扰动的转移概率矩阵 $\boldsymbol{P}_*$($\boldsymbol{P}_*$ 是 $P_{j\bar{i}}$ 元素构成的矩阵),则 $\tilde{X}_2 = \boldsymbol{P}_* \cdot \boldsymbol{P}_m$. 下面证明 $\tilde{X}_2$ 是属性变量 X 应用不变 PRAM 的结果,因为:

$$\Pr(\tilde{X}_2 = \varphi_x^j) = \sum_{i=1}^{d_x} \sum_{j=1}^{d_x} \Pr(\tilde{X}_2 = \varphi_x^j \mid \tilde{X}_1 = \varphi_x^i) \cdot \Pr(\tilde{X}_1 = \varphi_x^i \mid X = \varphi_x^j) \Pr(X = \varphi_x^j)$$

替换转移概率并重新排列项得到 $\Pr(\tilde{X}_2 = \varphi_x^j) = \Pr(X = \varphi_x^j)$,容易推断出 $\tilde{X}_2$ 的概率分布与 X 的概率分布相同. 这相当于使用一个满足 ε-局部差分隐私的不变 PRAM 转移概率矩阵对原始数据进行扰动.

经过上述的讨论,我们可以描述扰动算法如下:

算法 2 满足局部差分隐私的不变 PRAM 扰动算法

输入:数据集 DS,最大依赖对应的 QI 属性,差分隐私预算 ε

输出:扰动数据集 DS*

1. 统计数据集的记录数:n
2. 由式(9)计算扰动概率P_x
3. while $n \neq 0$
4. 按顺序提取每一行记录,$n = n - 1$;
5. for 为每个属性值生成一个随机数 R_1
6. if $R_1 \in [0, P_x]$
7. $\tilde{X}_1^i = X_1^i$;
8. else $\tilde{X}_1^i = X_1^{j \neq i}$;
9. end for
10. end while
11. 由式(10)计算扰动概率 $P_{\tilde{x}}$;
12. while $n \neq 0$
13. 按顺序提取每一行记录,$n = n - 1$;
14. for 为每个属性值生成一个随机数 R_2
15. if $R_2 \in [0, P_{\tilde{x}}]$
16. $\tilde{X}_2^i = X_2^i$;
17. else $\tilde{X}_2^i = X_2^{j \neq i}$;
18. end for
19. end while
20. 输出 DS*

算法 2 的行 1 ~ 2 是先对数据集进行初步统计,获取相应参数并由式(9)计算出第一次的扰动概率;行 3 ~ 9 是对原始数据集进行第一次扰动,依次遍历数据并进行扰动;行 10 是利用式(10)计算二次扰动概率;行 11 ~ 17 的操作类似于行 3 ~ 9 的扰动过程.

4 理论分析

本节先介绍如何估计 PRAM 的统计参数,再对不变 PRAM 数据的重构风险进行分析,理论证明满足局部差分隐私的不变 PRAM 在重构攻击时具有较低的属性泄露风险.

4.1 PRAM 估计

对属性变量 X,令 $\pi_i = \Pr(X = \varphi_x^i)$, $i = 1, \cdots, d_x$,其中 $\boldsymbol{\pi} = (\pi_1, \cdots, \pi\, d_x)^t$,用 n 表示数据集样本数,T_i 表示原始数据类别 φ_x^i 频率,则有 $T = (T_1, \cdots, T\, d_x)^t \sim \text{Mult}(n, \boldsymbol{\pi})$,但注意 $T_1, \cdots, T_{d_x}$ 不能从扰动数据中获得. 令 $\tilde{T}_j$ 表示扰动后类别 φ_x^j 的频率,$\lambda_i = \Pr(\tilde{X} = \varphi_x^i)$, $i = 1, \cdots, d_x$ 且 $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda\, d_x)^t$. 那么,对于任何固定的(预先指定的)扰动概率矩阵 $\boldsymbol{P}$,$\tilde{T} = (\tilde{T}_1, \cdots, \tilde{T}_{d_x})^t \sim \text{Mult}(n, \boldsymbol{\lambda})$,其中

$$\boldsymbol{\lambda} = \boldsymbol{P}\boldsymbol{\pi} \tag{11}$$

由于差分隐私预算参数 $\varepsilon > 0$,扰动矩阵 $\boldsymbol{P}$ 的对角元素 $p_{ii} = e^{\varepsilon}/(d_x - 1 + e^{\varepsilon}) > 0.5$,很容易找到相关元素使扰动矩阵 $\boldsymbol{P}$ 为非奇异矩阵,此时,$\boldsymbol{\lambda}$ 的任何无偏估计量 $\hat{\boldsymbol{\lambda}}$ 产生 $\boldsymbol{\pi}$ 的无偏估计量由 $\hat{\boldsymbol{\pi}} = \boldsymbol{P}^{-1}\hat{\boldsymbol{\lambda}}$ 给出. $\boldsymbol{\lambda}$ 的最大似然估计量是 $\hat{\boldsymbol{\lambda}} = \tilde{T}/n$,其产生 $\boldsymbol{\pi}$ 的估计量:

$$\hat{\boldsymbol{\pi}} = \boldsymbol{P}^{-1}\hat{\boldsymbol{\lambda}} = \boldsymbol{P}^{-1}(\tilde{T}/n) \tag{12}$$

4.2 重构风险分析

数据属性 X 经过随机化并发布以后,攻击者得不到原始数据集,但可以尝试根据观察到的数据和发布的随机化参数估计原始值. 与前文相同,以 $\hat{X}$ 表示攻击者对 X 的原始值估计,$\tilde{X}$ 表示扰动后数据对应的 X. X 的随机化过程可能取 $\varphi_x^1,\cdots,\varphi_x^{d_x}$ 中任意值,对于任何 $u,v\in\varphi_x^1,\cdots,\varphi_x^{d_x}$,后验概率的计算采用以下策略:

$$\hat{X}=u \text{ with probability } \Pr(X=u\mid\tilde{X}=v) \tag{13}$$

其中 $\Pr(X=u\mid\tilde{X}=v)$ 表示观察值 $\tilde{X}=v$ 时原始值 $X=u$ 的后验概率. 由于不变 PRAM 的数据发布者往往不公布扰动矩阵 $\boldsymbol{P}$,下面分析失真矩阵未知情况下攻击者获得重构数据的概率.

定义 3　如果 PRAM 是一个不变的 PRAM,则转移概率 $\boldsymbol{P}$ 满足式(14)

$$\boldsymbol{P}T=T \text{ or equivalent } \boldsymbol{P}\hat{\pi}_0=\hat{\pi}_0 \tag{14}$$

其中 $\hat{\pi}_0=T/n$ 是未应用 PRAM 时,即基于原始数据 π 的 MLE. 令 $\boldsymbol{P}=[P_1,\cdots,P_{d_x}]$是一个不变的 PRAM 矩阵,并将式(14)重写为:

$$\sum_{i=1}^{k}T_iP_i=T \tag{15}$$

设 F_x^i 表示属性变量值为 φ_x^i 的数量,$F_i=(F_x^1,\cdots,F_x^{dx})^t$. 扰动后属性变量的频率可以表示为 $\tilde{T}=\sum_i^{d_x}F_i$,其中给定 T 和 $\boldsymbol{P}$,$F_i\sim\text{Mult}(T_i,p_i)$,$i=1,\cdots,d_x$. 在不变 PRAM 估计中将 $\hat{\pi}_*=\tilde{T}/n$ 的统计特性作为 π 的估计量,使用 $\tilde{T}=\sum_{i=1}^{d_x}F_i$和$F_i$,$i=1,\cdots,d_x$的上述条件分布,可以得到:

$$\begin{aligned}\mathrm{E}(\hat{\pi}_*\mid T,\boldsymbol{P})&=\frac{1}{n}\sum_{i=1}^{d_x}\mathrm{E}[F_i\mid T,P]=\frac{1}{n}\sum_{i=1}^{d_x}T_iP_i\\&=\frac{1}{n}T=\hat{\pi}_0\end{aligned} \tag{16}$$

结合式(15)得到

$$\begin{aligned}\mathrm{V}(\hat{\pi}_*\mid T,P)&=\frac{1}{n^2}\sum_{i=1}^{d_x}T_i[\boldsymbol{D}_{P_i}-P_i(P_i)^t]\\&=\frac{1}{n}\Big[D_{\hat{\pi}_0}-\sum_{i=1}^{d_x}\Big(\frac{T_i}{n}\Big)P_i(P_i)^t\Big]\end{aligned} \tag{17}$$

这是不变 PRAM 方差的变化. 其中 $\boldsymbol{D}_{pi}$是对角元素为$p_1,\cdots,p_{dx}$的对角矩阵,$\boldsymbol{D}_{\hat{\pi}_0}$的定义类似. 由 $T\sim\text{Mult}(n,\pi)$和 $\hat{\pi}_0=T/n$,得到 $\mathrm{E}(\hat{\pi}_0)=\pi$ 和 $\mathrm{V}(\hat{\pi}_0)=[\boldsymbol{D}\pi-\pi(\pi)^t]/n$. 由式(16)和式(17)可以看出 $\mathrm{E}(\hat{\pi}_*)=\pi$,此时有

$$\begin{aligned}\mathrm{V}(\hat{\pi}_*)&=\mathrm{V}[\mathrm{E}(\hat{\pi}_*\mid T,P)]+\mathrm{E}[\mathrm{V}(\hat{\pi}_*\mid T,P)]\\&=\mathrm{V}(\hat{\pi}_0)+\frac{1}{n}\Big[\boldsymbol{D}_\pi-\mathrm{E}\Big\{\sum_{i=1}^{d_x}\Big(\frac{T_i}{n}\Big)P_i(P_i)^t\Big\}\Big]\end{aligned} \tag{18}$$

因此,不变 PRAM 数据的相对频率向量 $\hat{\pi}_*$ 是 π 的无偏估计量. 在未知转移概率矩阵 $\boldsymbol{P}$ 情况下,估计数据用户的 $\mathrm{V}(\hat{\pi}_*)$具有难度,但是攻击者一般只需估计出 $\mathrm{V}(\hat{\pi}_*)$的方差上界即可. 假设攻击者采用式(13)的概率策略来确定数据,用 $\Pr(\hat{X}=X=u)$表示重构概率,以评估敏感属性披露的风险,结合式(18)其准确估计出 X 的原始值的概率为

$$\Pr(\hat{X}=X=u)=\frac{1}{n}\Big[D_{\hat{\pi}_0}-\sum_{i=1}^{d_x}\Big(\frac{T_i}{n}\Big)P_i(P_i)^t\Big]$$

5 实验评估

5.1 实验设置

为验证算法的有效性,采用 UCI machine learning repository 人口统计数据集进行测试①. 随机选择 workclass, education, marital-status, occupation, relationship, sex, race, native-country, income 和 hours-per-week 10 个属性,workclass 作为敏感属性,其余属性作为准标识符属性. 实验硬件环境为 Intel(R) Core™ 2.60GHz CPU, 8GB RAM;操作系统为 Windows7 Ultimate,编程语言为 python3.6.

对原始数据清理后,首先用算法 1 选取原始数据集中与敏感属性具有强相互关系的 QI 属性,其中 occupation 属性与 workclass 属性有强依赖关系,其余 QI 属性与 workclass 属性的关联较弱.

5.2 实验分析及结果

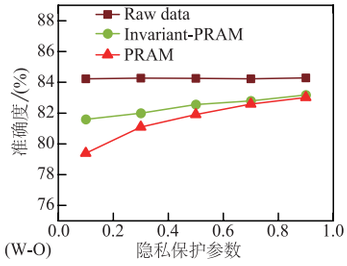
(1)属性关系对数据效用的影响

采用 scikit-learn 开源库②中的决策树算法验证标准可靠性,对未扰动情况,PRAM 和不变 PRAM 进行比较. 横坐标表示不同隐私保护参数 ε,取 0.1,0.3,0.5,0.7,0.9;纵坐标表示用决策树分类的准确度. 图 2(a)表示敏感属性 workclass 与具有强依赖关系的准标识符属性 occupation 的组合扰动结果,图 2(b)表示敏感属性 workclass 与随机选取的准标识符属性 relationship 组合扰动结果,图 2(c)表示敏感属性 workclass 与随机选取的多个准标识符属性 education, occupation, relationship 的组合扰动结果.

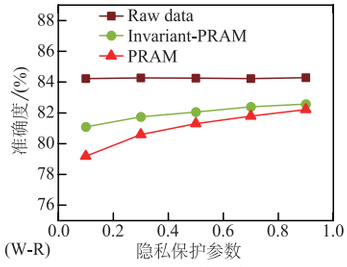
从图 2 中可知,未扰动的原始数据决策树分类的准确度最高,当选择二个属性扰动时,具有较强依赖关系的属性组合有更好的分类准确度,而选择多个属性进行扰动将会显著降低决策树分类准确度. 由此可知,在掌握原始数据集敏感属性与准标识符属性依赖关系的基础上有针对性的扰动,不变后随机响应具有明显的优势.

① http://archive.ics.uci.edu/ml/datasets/Adult

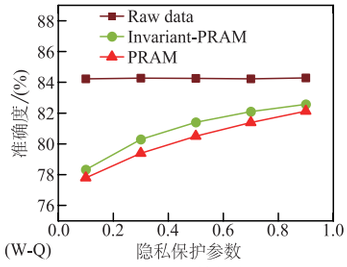
② http://scikit-learn.org/stable/install.html



(*a*) W-O属性组合



(*b*) W-R属性组合



(*c*) W-Q属性组合

图2 决策树分类准确度分析

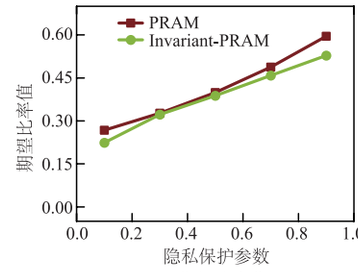
(2)隐私保护参数 ε 对数据效用的影响

隐私保护的数据发布需要考虑数据效用与敏感信息泄露程度的平衡,因此需要验证不变 PRAM 受参数 ε 影响程度,即预算参数 ε 通过随机响应输出原始值的概率对数据偏离程度的影响,隐私保护参数 ε 取 0.1, 0.2,…,0.9.

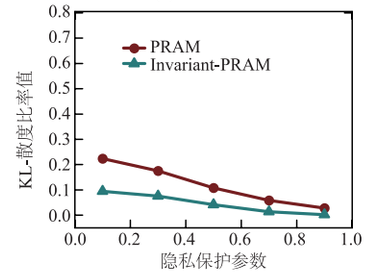
图 3 是 workclass 属性在不同隐私保护参数下 KL-散度与期望比的变化. 由图 3(a)可知,随着隐私保护参数 ε 的增加,PRAM 与不变 PRAM 的期望比也在增加,意味着扰动数据中预期记录数在不断增加,符合实际情况. 从图 3(b)可看出,在同样的隐私保护程度下,不变 PRAM 具有比传统 PRAM 更小的 KL-散度,说明算法的数据效用比传统 PRAM 更好.

(3)算法执行时间分析

表 1 给出算法运行时间,包括对原始数据处理和判断属性间关系并进行扰动的过程,为排除干扰,采用执行 10 次算法取平均值. 取不同的 ε 进行对数据进行扰动,ε 取 0.5 时算法执行的平均时间是 4.8453 秒. 从运行时间可以看出,不变 PRAM 算法与一般 PRAM 运行时间处于用一个数量级,且隐私保护的程度要求越低



(*a*) 期望比变化



(*b*) KL-散度变化

图3 敏感属性workclass的期望比及其 π 和之间的距离 (ε 变化)

扰动算法运行时间越短.

表 1 Adult 数据集运行时间 (秒)

隐私参数	$\varepsilon=0.1$	$\varepsilon=0.2$	$\varepsilon=0.3$	$\varepsilon=0.4$	$\varepsilon=0.5$	$\varepsilon=0.9$	$\varepsilon=1$	$\varepsilon=1.5$	$\varepsilon=2$
时间	4.851	4.850	4.848	4.846	4.845	4.837	4.836	4.831	4.826

(4)数据量增长时算法的稳定性

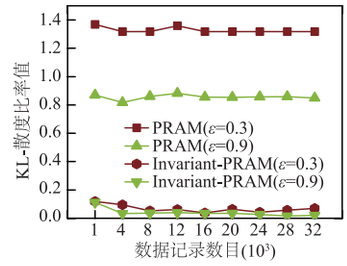
为验证扰动算法在数据量增加时的稳定性,采用不放回抽样随机选择 1000,4000,8000,…,32000 条数据记录作为实验数据,以横坐标表示数据的增量,纵坐标表示扰动后数据集中属性的 KL-散度.

图 4 是先选取敏感属性 workclass 进行独立扰动,再选择与其具有强依赖关系的准标识符 occupation 属性一起扰动进行验证,并进行 10 次扰动后的 KL-散度取平均值的结果.

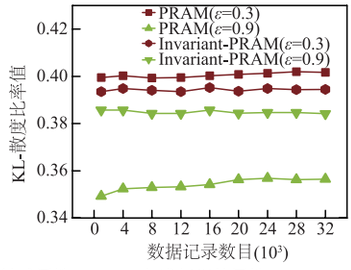
由图 4 可知,采用不变 PRAM 可以更有效的保留数据效用,减少用户信息泄漏,这是由于原始数据和扰动数据之间距离越小,它们之间的差异越小,扰动后的数据效用越好. 对于 PRAM,扰动前后数据间的距离受隐私保护参数影响更加明显;对于不变 PRAM,ε 取不同值时扰动前后数据的距离值都较小,意味着不变 PRAM 可以取得更高的隐私保护程度. 同时可以看出,属性个数的增加对 KL-散度影响很小,特别是数据量达到 2000 以后基本上处于稳定状态.

6 结束语

采用随机扰动模式的隐私保护数据发布方法,假设数据属性独立或者全相关,没有考虑准标识符属性与敏感属性部分相关的情况,导致数据效用降低或者计算复



(*a*) 敏感属性workclass 独立扰动



(*b*) 敏感属性workclass与准标识符属性occupation一起扰动

图4　数据 π 和 $\hat{\pi}$ 之间的KL-散度变化（十次取平均值）

杂度过大,为解决该问题,提出一种根据数据属性相互关系进行不变后随机响应扰动的方法. 重点讨论了原始数据属性关系的判断、扰动属性的选择、以及扰动矩阵的构建等问题. 最后设计实验验证,并对实验结果进行分析讨论. 实验结果表明,在考虑数据属性间的相互关系的基础上,采用满足 ε-局部差分隐私的不变 PRAM 扰动,可以保护隐私的同时具有较好的数据效用.

参考文献

[1] Sei Y, Ohsuga A. Differential private data collection and analysis based on randomized multiple dummies for untrusted mobile crowdsensing[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(4): 926 – 939.

[2] Zhang L, Mu Y, Wu Q. Compact Anonymous hierarchical identity-based encryption with constant size private keys: table 1[J]. The Computer Journal, 2016, 59(4): 452 – 461.

[3] Liu P, Hu C, Guo S, et al. Anonymous hierarchical identity-based encryption with bounded leakage resilience and its application[J]. International Journal of High Performance Computing and Networking, 2017, 10(3): 226 – 239.

[4] 杨静,王超,张健沛. 基于敏感属性熵的微聚集算法[J]. 电子学报,2014,42(7):1327 – 1337.
YANG Jing, WANG Chao, ZHANG Jianpei. Micro-aggregation algorithm based on sensitive attribute entropy[J]. Acta Electronica Sinica, 2014, 42(7): 1327 – 1337. (in Chinese)

[5] Guo L, Ying X, Wu X. On Attribute Disclosure in randomization based privacy preserving data publishing[A]. Proceedings of the International Conference on Data Mining Workshops[C]. Sydney, Australia: IEEE, 2010. 466 – 473.

[6] Holohan N, Leith D J, Mason O. Optimal differentially private mechanisms for randomised response[J]. IEEE Transactions on InformationForensics & Security, 2017, 12(11): 2726 – 2735.

[7] Xiao X, Tao Y, Chen M. Optimal random perturbation at multiple privacy levels[J]. Proceedings of the VLDB Endowment VLDB Endowment Hompage archive, 2010, 2(1): 814 – 825.

[8] Nayak T K, Zhang C, You J, et al. Measuring identification risk in microdata release and its control by post-randomisation [J]. International Statistical Review, 2018, 86(2): 300 – 321.

[9] Ardo V D H, Kooiman P. Estimating the linear regression model with categorical covariates subject to randomized response[J]. Computational Statistics & Data Analysis, 2006, 50(11): 3311 – 3323.

[10] Nayak T K, Adeshiyan S A. On invariant post-randomization for statistical disclosure control[J]. International Statistical Review, 2016, 84(1): 26 – 42.

[11] Wang K, Han C, Fu A W. Randomization resilient to sensitive reconstruction[J]. Annals of Internal Medicine, 2012, 158(2): 397 – 403.

[12] Geng Q, Kairouz P, Oh S, et al. The staircase mechanism in differential privacy[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 9(7): 1176 – 1184.

作者简介

杨高明　男,1974 年出生,安徽临泉人. 2012 年在哈尔滨工程大学获博士学位,现为安徽理工大学副教授,硕士研究生导师,主要研究领域为隐私保护、机器学习.

朱海明　男,1994 年出生,安徽阜阳人. 现为安徽理工大学硕士研究生,主要研究领域为隐私保护.

方贤进(通信作者)　男,1970 年出生,安徽舒城人. 2010 年在安徽大学获博士学位,现为安徽理工大学教授,硕士研究生导师,主要研究领域为信息安全、数据挖掘.
E-mail: xjfang@ aust. edu. cn