差分隐私综述

李效光, 李 晖, 李凤华, 朱 辉

西安电子科技大学网络与信息安全学院, 陕西 西安 710071

摘要 差分隐私是 2006 年由 DWORK 提出的一种新型的隐私保护机制, 它主要针对隐私保护中, 如何在分享数据时定义隐私, 以及如何在保证可用性的数据发布时, 提供隐私保护的问题, 这两个问题提出了一个隐私保护的数学模型。由于差分隐私对于隐私的定义不依赖于攻击者的背景知识, 所以被作为一种新型的隐私保护模型广泛地应用于数据挖掘, 机器学习等各个领域。本文介绍了差分隐私的基础理论和目前的研究进展, 以及一些已有的差分隐私保护理论和技术, 最后对未来的工作和研究热点进行了展望。

关键词 差分隐私; 隐私保护; 数据发布; 数据挖掘; 机器学习

中图法分类号 TP309 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2018.09.08

A Survey on Differential Privacy

LI Xiaoguang, LI Hui, LI Fenghua, ZHU Hui

School of Cyber Engineering, Xidian University, Xi'an 710071, China

Abstract Differential privacy is a privacy preserving mathematical model which was proposed by Dwork at 2006, it aims to solve two mainly problems which are how to define privacy while sharing data and how to publish data to satisfy privacy while provide utility. As the definition of differential privacy doesn't depend on background knowledge of adversaries, it is regarded as a new privacy preserving mechanism to apply in many fields such as data mining and machine learning. In this article, we introduce basic theories and present research progress, besides, we talk about existing theories and technologies of differential privacy. At the end, we discuss directions and underlying research hotspots in the future.

Key words differential privacy; privacy preserving; data publishing; data mining; machine learning

1 引言

在进入 21 世纪以来, 互联网行业发展十分的迅速, 随之而来的是人们通讯与数据共享的便利与快捷。然而, 由此引发的隐私泄露风险也随之日益增长。近年来, 隐私泄露事件时有发生, 另外, 随着计算机技术的发展与网络攻击手段的不断丰富, 保护隐私数据已远远不再是隐藏数据中敏感属性这么简单。例如, 在 2015 年, 俄罗斯约会网站 Topface 有 2000 万访客的用户名和电子邮件地址被盗, 黑客可以使用这些账号来尝试获取银行、病例或其他敏感数据信息。另外, 在 2015 年, 美国大型医疗保险商 CareFirst 表示, 该公司去年六月发现有黑客入侵, 约有 110 万医疗保险客户的个人信息遭泄露, 攻击者窃取了客户姓名、生日、邮箱地址、医疗保险号码等信息。这些事实说明, 隐私的泄露原因可能是多样的, 因此, 隐私保护理论和技术需要能够针对不同的攻击手段提供保护。更为严峻的是, 随着进几年来数据挖掘等数据分析技术的快速发展, 使得攻击者可以从海量数据中发掘出与用户隐私相关的信息, 这又对隐私保护提出了新的挑战。因为攻击者可以应用数据挖掘技术对海量的数据进行分析, 从而得到数据中深层蕴含的用户信息, 而不是通过访问数据直接获取。因此, 传统的加密, 访问控制等技术对这样的攻击方式并没有太好的效果。因此, 隐私保护是一门结合多个学科和领域的交叉技术。

要对隐私进行保护, 首先需要对隐私信息进行定义和度量。李凤华等人[1]提出了隐私信息的全生命周期模型, 如图 1 所示, 其中包括 9 个部分。分别为: 隐私信息产生, 隐私感知, 隐私保护, 隐私发布, 隐私信息存储, 隐私交换, 隐私分析, 隐私销毁, 隐私接收者。该模型如图 1 所示。隐私保护所研究的问

通讯作者: 李效光, 学历: 本科, Email: l1060572040@163.com。

本课题得到国家自然科学基金——广东省联合基金(No. U1401251)资助。

收稿日期: 2017-09-11; 修改日期: 2018-02-22; 定稿日期: 2018-08-20

题, 主要在隐私保护, 隐私发布/存储/交换, 隐私分析这 3 个部分。

隐私保护的方式主要分为以下三种: 数据失真, 加密, 以及访问控制。而目前很多隐私保护技术结合其中多种技术。k-匿名[2-3]是由 Latanya Sweeney 和 Pierangela Samarati 在 1998 提出的一种匿名化数据的技术, 它通过混淆数据的属性, 解决了“给定一个原始数据集, 生成一个匿名化的数据集, 它可以在保证数据的实验可用性的条件下, 保证其中的个体身份不会被恢复出来”。k-匿名对数据集提供了一个良好的性质, 它可以使得包含在匿名化数据集中的每一个个体信息都不能从其他 k-1 个个体信息中区分开。但是, k-匿名中不包含任何的随机化属性, 所以攻击者依然可以从满足 k-匿名性质的数据集中推断出与个体有关的隐私信息。另外, k-匿名还容易遭受到一致性攻击与背景知识攻击[4], 一致性攻击主要是指, 在数据集中, 如果有 k 个相同属性的敏感值, 那么即使数据集符合 k-匿名的要求, 那么这 k 个敏感属性也可能被推断出来。而背景知识攻击是指攻击者可以通过找出一个或多个准身份信息属性和敏感属性之前的关联, 以此来缩小对敏感属性猜测的范围。由于 k-匿名存在以上缺陷, Machanavajjhala 等人[5]在 2007 对 k-匿名提出了一个改进的方案, l-多样性。l-多样性是指, 一个等价类中最少有 l 个可以很好代替的敏感属性的值, 对于数据表来说, 则需要每一个等价类中都满足 l-多样性。但是, l-多样性也并不能完全的保护用户隐私不被泄露, 因为在一个真实的数据集中, 属性值很有可能是偏斜的或者语义相近的, 而 l-多样性只保证了多样性, 没有认识到在属性值上语义相近的情况。因此 l-多样性会受到相似性攻击[6]。李宁辉等人[7]在提出的 t-closeness 方案弥补了 l-多样性, t-closeness 指一个等价类中的属性分布和整个表中的属性分布之间的距离不超过门限 t。如果一个数据表中的每个等价类都满足 t-closeness, 则称这个数据表满足 t-closeness。

虽然已有的隐私保护方案层出不穷, 但是它们有一个共同的缺点, 都依赖于攻击者的背景知识, 没有对攻击模型做出合理的假设。

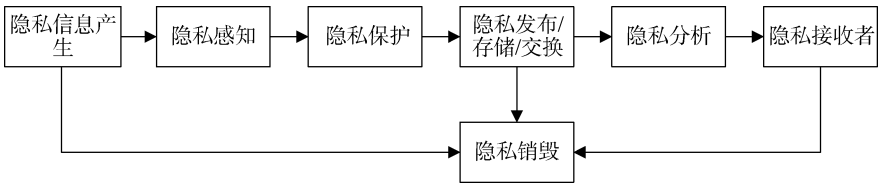


图 1 隐私信息的全生命周期

Figure 1 The full life cycle of privacy information

而 2006 年 Dwork 等人提出的差分隐私[8-14]模型解决了这个问题, 差分隐私的概念来自于密码学中语义安全的概念, 即攻击者无法区分出不同明文的加密结果。在差分隐私中, 要求攻击者无法根据发布后的结果推测出哪一条结果对应于那一个数据集。该模型通过加入随机噪声的方法来确保公开的输出结果不会因为一个个体是否在数据集中而产生明显的变化, 并对隐私泄露程度给出了定量化的模型。因为一个个体的变化不会对数据查询结果有显著的影响, 所以攻击者无法以明显的优势通过公开发布的结果推断出个体样本的隐私信息, 所以差分隐私模型不需要依赖于攻击者所拥有多少背景知识, 而且对隐私信息提供了更高级别的语义安全, 因此被作为一种新型的隐私保护模型而广泛使用。

2 差分隐私基础

差分隐私并不是要求保证数据集的整体性的隐私, 而是对数据集中的每个个体的隐私提供保护。它的概念要求每一个单一元素在数据集中对输出的影响都是有限的。从而使得攻击者在观察查询结果后无法推断是哪一个个体在数据集中的影响使得查询返回这样的结果, 因此, 也就无法从查询结果中推断有关个体隐私的信息。换言之, 攻击者无法得知某一个个体是否存在于这样的一个数据集中。

定义 1. 差分隐私. 对于一个随机算法 M, P_m 为算法 M 可以输出的所有值的集合。如果对于任意的一对相邻数据集 D 和 D', P_m 的任意子集 S_m, 算法 M 满足:

$$Pr[M(D)\in S_m]\leqslant e^{\epsilon}Pr[M(D')\in S_m] \qquad (1)$$

则称算法 M 满足 ϵ-差分隐私, 其中参数 ϵ 为隐私保护预算。

从上式中可以看出, 当参数 ϵ 越小时, 作用在一对相邻数据集上的差分隐私算法返回的查询结果的概率分布越相似, 攻击者就越难以区分这一对相邻

数据集, 保护程度就越高, 极端情况下, 当 $\epsilon=0$ 时, 攻击者无法区分这一对相邻数据集保护程度最高。反之, 参数 ϵ 越大时, 保护程度越低。

图 2 说明了差分隐私概念的性质。差分隐私机制将一个正常的查询函数 $f(\bullet)$ 的查询结果, 映射到一个随机化的值域上, 并以一定的概率分布来给用户返回一个查询结果。通过参数 ϵ 来控制一对相邻数据集上的概率分布的接近程度, 达到在一对相邻数据集上, 输出结果几乎一致的目的。从而使得攻击者无法区分这一对相邻数据集, 实现保护数据集中个体隐私信息的目的。

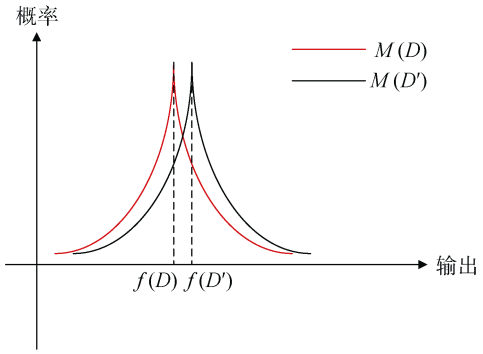


图 2 差分隐私算法在邻近数据集上的输出概率
Figure 2 Output probability of DP algorithm on neighbor dataset

McSherry 等人[15]在 2010 年又对差分隐私提出了 2 个重要的性质。分别为顺序合成性质和平行合成性质。

性质 1. 顺序合成. 对于任意 k 个算法, 分别满足 ϵ_1-差分隐私, ϵ_2-差分隐私, …, ϵ_k-差分隐私。将他们作用于同一个数据集上时, 满足 $(\sum_{i=1}^{k}\epsilon_i)$-差分隐私。

这个性质说明了, 当有一个算法序列同时作用在一个数据集上时, 最终的差分隐私预算等价于算法序列中所有算法的预算的和。

性质 2. 平行合成. 把一个数据集 D 分成 k 个集合, 分别为 $D_1, D_2, \cdots, D_k$, 令 $A_1, A_2, \cdots, A_k$ 是 k 个分别满足 $\epsilon_1, \epsilon_2, \cdots, \epsilon_k$ 的差分隐私算法, 则 $A_1(D_1), A_2(D_2), \cdots, A_k(D_k)$ 的结果满足 $\max_{i\in[1,2,\cdots,k]}\epsilon_i$-差分隐私。

这个性质说明了, 当有多个算法序列分别作用在一个数据集上多个不同子集上时, 最终的差分隐私预算等价于算法序列中所有算法预算的最大值。

这两个性质在设计差分隐私机制时有重要的作用, 它们可以被用来控制一个差分隐私机制在使用中所需要的隐私预算。控制隐私预算的目的在于, 如果在一个较低隐私预算参数 ϵ 的情况下, 攻击者对一个数据集进行了多次查询, 那么根据性质 1, 攻击者实际上获得的隐私预算就相当于获得了多次查询的隐私预算的和, 而这就破坏了原本设定的隐私预算。所以需要控制隐私预算的上限, 来通过上述的性质来计算合适的隐私预算上限。

另外, Daniel Kifer 等人[16]在 2010 年对差分隐私又提出了另外 2 个性质。

性质 3. 变换不变性. 给定任意一个算法 A_1 满足 ϵ-差分隐私, 对于任意算法 A_2 (A_2 不一定是满足差分隐私的算法), 则有 $A(\bullet)=A_2(A_1(\bullet))$ 满足 ϵ-差分隐私。

这个性质说明了差分隐私对于后处理算法具有免疫性, 如果一个算法的结果满足 ϵ-差分隐私, 那么在这个结果上进行的任何处理都不会对隐私保护有所影响。

性质 4. 中凸性. 给定 2 个算法 A_1 和 A_2 满足 ϵ-差分隐私。对于任意的概率 $p\in[0,1]$, 用符号 A_p 表示为一种机制, 它以 p 的概率使用 A_1 算法, 以 $1-p$ 的概率使用 A_2 算法, 则 A_p 机制满足 ϵ-差分隐私。

这个性质说明, 如果有 2 个不同的差分隐私算法, 都提供了足够的不确定性来保护隐私, 那么可以通过选择任意的算法来应用到数据上实现对数据的隐私保护, 只要选择的算法和数据是独立的。

3 差分隐私模型

差分隐私可以通过在查询结果上加入噪声来实现对用户隐私信息的保护, 而噪声量的大小是一个关键的量, 要使加入的噪声既能保护用户隐私, 又不能使数据因为加入过多的噪声而导致数据不可用。函数敏感度是控制噪声的重要参数。Dwork 等人[17]在 2006 年, 提出了全局敏感度以及拉普拉斯机制的概念, 通过全局敏感度来控制生成的噪声大小, 可以实现满足差分隐私要求的隐私保护机制。

定义 2. 全局敏感度. 对于一个查询函数 f, 它的形式为: $f:D\to R$, 其中 D 为一数据集, R 是查询函数的返回结果。在一对任意的相邻数据集 D 和 D' 上, 它的全局敏感度定义如下:

$$S(f)=\max_{D,D'}\|f(D)-f(D')\|_1 \tag{2}$$

其中, $\|f(D)-f(D')\|_1$ 是 $f(D)$ 与 $f(D')$ 之间的曼哈顿距离。

全局敏感度反映了一个查询函数在一对相邻数据集上进行查询时变化的最大范围。它与数据集无关, 只由查询函数本身决定。

拉普拉斯机制是一种简单, 而且广泛用于数值型查询的隐私保护机制。对于数值型的查询结果, 拉普拉斯机制通过在返回一个在查询结果上加入一个满足 $Lap(0,\frac{\Delta f}{\epsilon})$ 分布的噪声的结果来实现差分隐私保护。即 $R(D)=f(D)+x$, 其中 f 为查询函数, x 为满足拉普拉斯分布的噪声。另外, 所加入的拉普拉斯噪声的均值要求为 0, 这样输出的 $R(D)$ 才是 $f(D)$ 的无偏估计。

图 3 展示了不同参数 ϵ 下的拉普拉斯噪声的概率密度函数。从图中可以看出, ϵ 越小, 所加入的拉普拉斯噪声的概率密度越平均, 所加入的噪声为 0 的概率就越小, 对输出的混淆程度就越大, 保护程度也就越高。

但是当全局敏感度较大时, 根据全局敏感度生成的噪声往往会对数据提供过度的保护, 针对这一问题, Nissim 等人[18]提出了一个局部敏感度以及平滑敏感度等新的概念来解决这一问题。

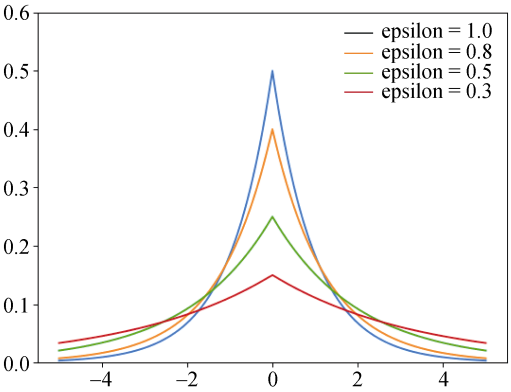


图 3 不同 ϵ 的拉普拉斯噪声的概率密度

Figure 3 Probability density of Laplace noise with different ϵ

定义 3. 局部敏感度. 对于一个查询函数 f, 它的形式为: $f:D\to R$, 其中 D 为一数据集, R 是查询函数的返回结果。在一给定的数据集 D 和与它相邻的任意数据集 D' 上, 它的局部敏感度定义如下:

$$LS_f(D)=\max_{D'}\|f(D)-f(D')\|_1 \tag{3}$$

其中, $\|f(D)-f(D')\|_1$ 是 $f(D)$ 与 $f(D')$ 之间的曼哈顿距离。

与全局敏感度不同, 局部敏感度是由查询函数和给定的数据集共同决定, 因为局部敏感度只是对于一个数据集做变化。

因为局部敏感度限制了一对相邻数据集中的一个数据集, 所以如果在局部敏感度中, 给定的数据集和全局敏感度中使 $\|f(D)-f(D')\|_1$ 达到最大的数据集相同时, 局部敏感度等于全局敏感度。所以, 局部敏感度和全局敏感度的关系可以表示为:

$$S(f)=\max_{D}\{LS_f(D)\} \tag{4}$$

因为根据局部敏感度所产生的噪声和数据集本身相关, 所以直接使用局部敏感度生成噪声会泄露数据集信息, Nissim 等人[18]提出了根据平滑敏感度来生成噪声的方案, 它们首先提出了平滑上界的概念。

定义 4. 平滑上界. 给定一个 $\beta>0$, 对于一个函数 $F:D\to R$, 在查询函数 f 上, 如果它满足如下条件:

$$\forall D:F(D)\geqslant LS_f(D) \tag{5}$$

$$\forall D,D':F(D)\leqslant e^{\beta}\times LS_f(D') \tag{6}$$

则称函数 F 是一个在查询函数 f 上的 β-平滑上界。

定义 8. 平滑敏感度. 对于一个 $\beta>0$, 一个查询函数 f 的 β-平滑敏感度为:

$$S_f^*(D)=\max_{y\in D^n}\{LS_f(y)\times e^{-\beta\times d(D,y)}\} \tag{7}$$

其中, y 是和给定数据集 D 维度相同的任意数据集, 函数 d 返回数据集 D 和 y 中的不同元素的个数。实际上, 平滑敏感度就是可以满足平滑上界条件的最小函数。

根据这一方案, Nissim 等人[18]还提出了 Sample-Aggregate 框架, 对原有的隐私保护框架做出了改进, 在 Sample-Aggregate 框架中, 所添加的噪声不仅仅和查询函数有关, 还和数据集本身有关。而使用平滑敏感度, 保证了添加的噪声虽然与数据集有关, 但不会泄露有关数据集的相关信息。对于很多查询函数来说, 它的平滑敏感度可能是难以有效计算的, 甚至是 NP-困难的, 而且对于不同的查询函数, 平滑敏感度的计算不能自动进行。Sample-Aggregate 框架很好的解决了这一问题, 它可以自动的进行, 并且对于大多数查询函数都适用, 而且不需要精确的计算出查询函数的平滑敏感度。Sample-Aggregate 框架通过把一个查询函数 f 转化为一个平滑敏感度较低相关函数 $\overline{f}$, 再加入合适的噪声来作为查询结果。图 4 说明了 Sample-Aggregate 框架的结构。

Sample-Aggregate 框架首先将一个数据集随机取样划分为 m 个小子集, m 是框架中设定好的参数,

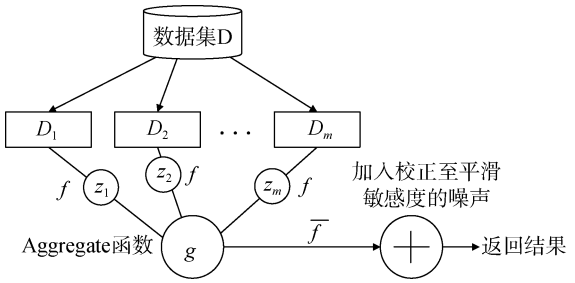


图 4 Sample-Aggregate 框架

Figure 4 Framework of Sample-Aggregate

然后对每个子集上执行查询函数 f 来生成一个在 f 的输出空间上的值 z_k, 最后通过聚合函数生成 $\overline{f}$ 来替代原始查询函数 f, 加入校正至平滑敏感度的噪声来得到查询结果。

对于批量线性查询的问题, Li 等人[19]提出了一种矩阵机制, 优化了大量线性查询中噪声量过大的问题。该方案通过将批量的线性查询转化为一查询负载 W, 该 W 矩阵中包含了一系列不同的线性查询。该方案使用一个不同的矩阵 A 来进行查询, 矩阵 A 称为查询策略。在这里, 我们把可以线性表示查询负载的矩阵 A 称为查询负载 W 的查询策略。严格的说, 即存在解矩阵 X, 使得 $W=XA$ 成立。矩阵机制通过在查询策略上加入合适的噪声来实现差分隐私保护。Li 等人[19]对矩阵机制 $M_{k,A}(W,x)$ 给出了如下定义:

$$M_{k,A}(W,x)=WA^{+}K(A,x) \tag{8}$$

其中 $K(A,x)$ 为作用于数据集 x 和查询策略 A 的差分隐私机制, A^{+} 为查询策略 A 的广义逆矩阵。对于矩阵机制中使用的差分隐私机制 $K(A,x)$, 可以使用拉普拉斯机制来实现。具体来说, $K(A,x)=Ax+\tilde{b}_A$, 其中 Ax 是在查询策略下的查询结果, $\tilde{b}_A$ 是一个噪声向量, 用来提供满足拉普拉斯分布的噪声。图 5 说明了矩阵机制的模型结构。

但是矩阵机制的缺点在于, 当给定一个查询负载时, 求解其最优的查询策略是一个半正定最优问题, 当查询负载在一个有 m 个数据格的直方图上时, 求解该问题的复杂度为 $O(m^6)$, 这使得矩阵机制对于大型的数据是难以使用的。

由于拉普拉斯机制只能针对数值型数据进行隐私保护, 对于非数值型数据, 例如实体对象, McSherry 等人[20]提出了指数机制。

假设所有可能的输出集合为 O, 指数机制的目的是使输出结果满足一定的概率分布。用可用性函数 q 来衡量每一个输出项的价值。q 定义为 $q:(D\times o)\to R$, 其中 D 和 o 为输入的数据集和可能的输出集合中的项, 函数 q 返回一个实数用来表示这一项的价值。当 q 的值越高时, 这一项的价值越大, 被输出的概率也就越大。

定义 9. 指数机制. 对于任意一个可用性函数 q 和一个差分隐私预算 ϵ, 如果随机算法 M 以正比于 $e^{\frac{\epsilon\times q(D,o)}{2\times\Delta q}}$ 的概率输出一个 $o\in O$ 作为结果, 其中 Δq 为可用性函数 q 的全局敏感度, 则随机算法 M 满足 ϵ-差分隐私。

指数机制的意义在于防止了攻击者对数据集中个体的投票情况的推测。例如在一次投票活动中, 一共有 3 个候选人(用编号 1 到 3 表示), 10 位选民, 攻击者控制了除了选民 A 以外的其他 9 个选民(B, C, …, J), 现在他的目的是推断选民 A 的投票情况。假设在 A 没有投票时, 每个候选人的得票数都为 3, 也就是说 B, C, …, J 分别给每个候选人各投了一票, 那么如果攻击者想要推断 A 的投票情况, 就可以通过最终的选举结果看哪位候选人胜出来判断 A 的投票结果。但是, 如果加入指数机制, 就可以抵御这种攻击。

在指数机制中, 参数 ϵ 越小, 每一项输出的概率就越接近, 相应的, 输出三个候选人的概率也就越

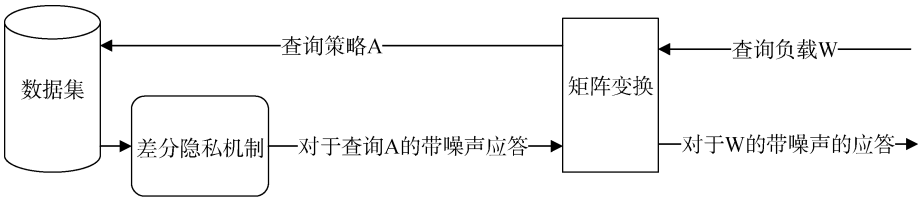


图 5 矩阵机制模型

Figure 5 Matrix mechanism

平均, 从而让攻击者难以判断 A 的投票情况。当参数 ϵ 选择为 0 时, 隐私保护级别最高, 攻击者完全无法判断 A 的投票情况。表 1 以一个图表说明了指数机制如何抵御这种攻击。

表 1　投票数据集及公布结果概率分布

Table 1　ballot dataset and probability distribution

	得票数	$\epsilon=5$	$\epsilon=0.5$	$\epsilon=0$
1	4	0.9	0.39	0.33
2	3	0.045	0.3	0.33
3	3	0.045	0.3	0.33

另外, Blum 等人[21]提出了网络机制, 可以计算比较可能的新数据集与原始数据集的误差来最优的生成新数据集来实现差分隐私。Hardt 等人[22]提出了隐私乘法权重调整机制, 可以动态地调整数据集真实分布与人为猜测分布间的误差使其满足差分隐私。

综上所述, 在一个隐私信息的全生命周期内, Sample-Aggregate 框架和矩阵机制主要被设计为交互式的模型应用于隐私保护和隐私发布/存储/交换的部分, 而拉普拉斯机制和指数机制是差分隐私方法的基础模型, 由这 4 种机制设计的其他差分隐私方案被广泛的用于隐私信息的全生命周期内的隐私保护, 隐私发布/存储/交换以及隐私分析部分。

4　差分隐私在数据发布中的应用

由于差分隐私在隐私信息保护方面的诸多性质, 使得差分隐私被用于了各个数据发布的技术领域。直方图是一种常用的数据分布的表示形式, 经常被用于数据发布上, 在各个领域都有广泛的应用。除了矩阵机制, 差分隐私还在直方图上有广泛的应用。

Dwork 等人[23]将设计的拉普拉斯机制应用在了直方图发布上, 他设计了一种简单的机制。

因为在数据集对应的直方图中, 每一个数据集中元素的变化最多影响直方图中一个数据格上一个元素的变化, 所以该方法使用拉普拉斯机制, 在每个数据格上添加服从 $Lap(\frac{1}{\epsilon})$ 分布的噪声。Dwork 指出, 在这种机制下, 对直方图进行范围查询的 MSE 为 $|r|\times\frac{2}{\epsilon^2}$, 其中 $|r|$ 为范围查询的长度。Dwork 还提出, 对于各种范围的查询, 这种方案的平均 MSE 为 $\frac{N+2}{3}\times\frac{2}{\epsilon^2}$, 其中 N 为数据格数。

对于提高范围查询精度的问题, Qardaji 等人[24]提出了一种基于分层的差分隐私方法。该方法将直方图转化为一颗 b 叉树, 叶子结点为数据格中的数据值, 每个叶子结点的父节点为其所有叶子结点的和。图 6 显示了分层方法的结构。

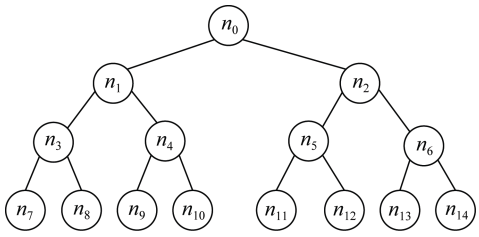


图 6　分层方法结构

Figure 6　Hierarchical mechanism

分层方法的一个优势是, 对于任何范围查询, 所使用的结点个数不会超过 $(b-1)h$ 个, Qardaji 等人[24]还提出, 这种分层方法对于各种范围的查询的平均 MSE 为 $\frac{N}{N+1}((b-1)h^3-\frac{2(b+1)h^2}{3}+O(\frac{bh}{N}))\times\frac{2}{\epsilon^2}$, 其中, h 为整个树不包含根节点的层数, 从叶子层数开始计数, 根节点为 $h+1$ 层。b 为树的分支因数。与 Dwork 的方案相比, 精度有了明显的提高。为了优化分层方法的精度, Qardaji 等人[24]把 Hay 等人[25]之前提出的一种 Constrained Inference 方法应用于分层方法。这种方法对树形结构中每个结点进行了 2 步细化的处理, 分别为 Weighted Averaging 和 Mean Consistency。首先执行 Weighted Averaging 步骤, 从叶子结点开始向上遍历直到根节点, 并用结点的原始噪声值的加权平均值与其所有子节点的和来更新每个结点上的值。

用 $n[v]$ 来表示加入噪声的结点 v。Weighted Averaging 具体过程如下:

$$z_i[v]=\begin{cases}n[v], i=1\\ \dfrac{b^i-b^{i-1}}{b^i-1}n[v]+\dfrac{b^{i-1}-1}{b^i-1}\displaystyle\sum_{u\in child(v)}z_{i-1}[u], i>1\end{cases}\tag{9}$$

Weighted Averaging 的意义在于, 从叶子结点向上更新每个结点上的计数, 以降低每个结点上的方差。

而 Mean Consistency 与之相反, 从根节点开始向下遍历, 更新每个结点的值, 目的是使得加入噪声后, 父节点上的值和其所有子节点上的值的和能保证相等。具体的过程如下:

$$\overline{n}_i[v]=\begin{cases}z_i[v], v\text{是根节点}\\ z_i[v]+\dfrac{1}{b}(\overline{n}_{i+1}[u]-\displaystyle\sum_{v\in child(u)}z_i[v])\end{cases}\tag{10}$$

Qardaji 等人[24]还分析了 Constrained Inference 方法对于分层方法的影响。得出结论, 在加入 Constrained Inference 方法后, 每个结点的方差最小

为 $\frac{b-1}{b+1}\times h^2\times\frac{2}{\epsilon^2}$。

Xiao 等人[26]在分层方法上提出了一种哈尔小波变换的机制, 该方案在哈尔系数上添加噪声来实现隐私保护, 而且只适用于二叉树。该方案中, 根节点不再是所有子节点的和, 而是其所有子节点的平均值。对于所有的非叶子结点 v, 哈尔系数 $h_v=\frac{(a_l-a_r)}{2}$, 其中 a_l 和 a_r 分别为结点 v 的左子树的所有叶子结点的平均值和右子树的所有叶子结点的平均值。在进行查询时, 对于任意一个结点 v, 如果已知它的所有叶子结点的平均值 a, 则 a_l 和 a_r 可由如下方式计算 $a_l=a+h_v$, $a_r=a-h_v$。基于哈儿小波变换的分层方法结构如图 4 所示。

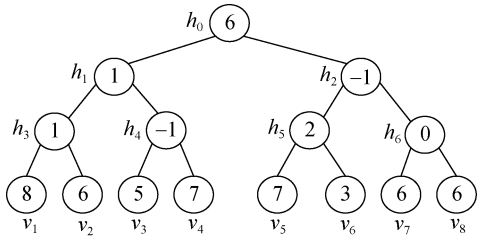


图 7 基于哈尔小波变换的分层方法结构

Figure 7 Hierarchical mechanism based Haar wavelet transform

另一种有效的降低噪声的方法是, 将所有数据格合并为若干个分区, 每个分区的频数为其中全部频数的平均值。然后在每个合并的分区中添加噪声, 这样就可以减少添加的噪声数。但是如何划分使得分区最优, 是一个新的研究问题。Xu 等人[27]采用平方误差和来衡量一种分区方案的优劣程度:

$$Error(H,D)=\sum_{i=1}^{n}\sum_{j=1}^{k}(c_i-x_j)^2 \tag{11}$$

根据这一度量方法, Xu 等人提出了 NoiseFirst 方案用于在直方图上实现差分隐私保护。

NoiseFirst 基于拉普拉斯机制, 首先加入噪声, 然后遍历得到最优的分割数据格方案, 再对直方图进行分割合并。

但是 NoiseFirst 方案只适用于一维的 V-优化直方图发布。针对这一问题, Xiao 等人[28]提出了 DPCube 方案, 结合单元划分与 kd-树的思想, 用以获得多维 V-优化直方图, 来支持多维直方图的数据发布。

综上所述, 这 5 种方法主要应用于隐私信息全生命周期中的隐私保护和隐私发布/存储/交换的部分。

另外, 还可以通过先对直方图结构进行重新组织再加入噪声来实现差分隐私。根据聚类方法重新划分直方图中的数据格, Xu 等人[27]提出了 StructureFirst 方案, 采用平方误差和来衡量一种直方图分区方案的优劣程度, 结合指数机制, 对原始直方图进行压缩得到了满足差分隐私的 V-优化直方图。该方案的缺点在于, 没有顾及到重构误差和噪音误差之间的平衡; 对于一些数据格个数比较多的直方图, 该方法效率非常低。Acs 等人[29]提出了 P-HPartation 方案来解决该问题, 该方案结合自适应的层次聚类技术和贪婪二等分策略, 平衡了 StructureFirst 方案中的重构误差与噪音误差并且提高了效率。根据傅里叶变换理论, Rastogi 等人[30]结合离散傅里叶变换和逆离散傅里叶变换以及拉普拉斯机制, 发布满足差分隐私的直方图。

除了直方图发布技术, 还存在基于划分的数据发布技术。Qardaji 等人[31]提出了 UG 方案, 对针对二维空间数据, 结合划分粒度与拉普拉斯机制对数据进行划分, 实现基于差分隐私的数据发布。但是该方案的缺点在于, 没有考虑数据分布的密度和稀疏。针对这一问题, Qardaji 等人[31]又提出了 AG 方案, 该方案是对 UG 方案的改进, 通过自适应划分策略来避免单元过于密集与过于稀疏的问题。

另外, 还可以通过非交互式算法发布合成数据集或加入噪声的数据集来实现数据发布。针对高维数据, Chen 等人[32]提出一种基于取样的框架, 通过联结树对高维数据中每个数据属性取样分析, 生成带有噪声的属性依赖关系图, 再根据关系图合成数据集来进行高维数据发布。对于用户轨迹数据数据集, Chen 等人[33]还结合前缀树结构, 生成带有噪声的前缀树, 再根据此前缀树生成用于发布的净化数据集。Li 等人[34]提出压缩机制, 通过使用压缩技术, 将噪声加入稀疏数据集的取样样本中, 将噪声的大小降低到 $O(\log n)$。Machanavajjhala 等人[35]通过设计带有噪声的统计模型从原始数据集中随机取样以生成净化过的数据集用来发布。Zhou 等人[36]通过随机线性变换或仿射变换来压缩数据集。如表 2 所示, 表 2 总结了本节中说明的差分隐私算法的优缺点。

综上所述, 这 15 种方法主要应用于隐私信息全生命周期中的隐私保护和隐私发布/存储/交换的部分。

5 基于差分隐私保护的数据挖掘和机器学习

由于数据挖掘和机器学习往往要处理海量的用

表 2 差分隐私在数据发布中方案总结

Table 2 Summary of DP in data release

方法名称	方案描述	优点	缺点
Laplace	对每个数据格中加入拉普拉斯噪声	简单而且易于实现	对于直方图上的范围较大的查询精度较低
基于分层的差分隐私方法	将直方图转化为树形结构, 在结点中加入噪声	对于直方图上的范围较大的查询精度较高	在高维直方图上精度不高
Constrained Inference	通过权重平均和均值限制降低直方图误差	敏感度小, 噪声较小	查询次数有限
哈尔小波变换的机制	在哈尔系数上添加噪声	支持较长范围计数查询, 精度较高	实际的可用性比较差
NoiseFirst	先加入噪声, 遍历得到最优分割数据格方案, 再分割合并	支持较长范围计数查询, 查询精度较高	仅适用于一维直方图
DPCube	结合单元划分与 kd-树的思想, 用以获得多维 V-优化直方图	支持多维的单位长度与较长范围计数查询	发布误差大, 可用性低, 查询精度不稳定
StructureFirst	采用平方误差和指数机制, 压缩原始直方图	支持较长范围计数查询, 查询精度较高	无法平衡重构与噪音误差, 无法处理奇异点
P-HPartation	结合自适应层次聚类技术和贪婪二等分策略	支持较长范围计数查询, 效率高	无法处理奇异点, 仅适用一维直方图
傅里叶变换机制	结合离散傅里叶变换和逆离散傅里叶变换以及拉普拉斯机制	支持查询类型多	对于高维度数据集产生较大噪声
UG	结合划分粒度与拉普拉斯机制对数据进行划分	支持范围计数查询, 均衡噪音与均匀假设误差	没有考虑数据分布的稀疏性
AG	自适应划分并结合划分粒度与拉普拉斯机制划分数据	支持数据依赖的范围查询, 自适应均衡两种误差	均衡误差时, 没有采用启发式方法
JTree	生成带有噪声的属性依赖关系图, 根据关系图合成数据集	可以应用于二进制和非二进制数据集, 均有较高精度	对于高维数据集, 隐私预算消耗速度快
STM-Full	通过加入噪声的前缀树合成数据集	适用于多种类型的轨迹数据	对于传统的轨迹数据隐私保护方案时间消耗更大
压缩机制	对稀疏数据集压缩取样加入噪声再生成新的合成数据集	可以以更小的噪声实现差分隐私保护	对于非稀疏数据集降噪效果不明显
噪声统计模型	设计噪声统计模型从原始数据集中随机取样生成净化数据集	对于稀疏数据集可以产生精确的合成数据集	当数据块过长时会导致对数据集过度估计

户数据, 这其中会蕴含有大量的用户隐私信息, 如何在保护用户隐私信息的同时, 还可以挖掘分析出可用的信息, 是隐私保护研究的一个重要课题。因此, 数据挖掘是差分隐私应用的一个重要领域。

Mohammed 等人[42]提出了一种基于泛化技术的非交互模式匿名化数据挖掘算法, 通过在泛化技术中加入指数噪声来实现差分隐私保护。该方案设计了一种非交互式的差分隐私机制。它使用分类树来对原始数据集进行分类, 并对属性信息进行泛化。在该方案中, 采用了信息增益来选择分割属性, 在分类树中, 以信息增益来作为可用性函数, 对每个结点的分割值进行可用性度量, 属性 A 对于数据集 D 的信息增益可表示为:

$$InfoGain(D, A) = H(D) - H(D \mid A) \tag{12}$$

其中 $H(D)$ 表示数据集 D 的经验熵, 而 $H(D \mid A)$ 表示用属性 A 对数据集 D 划分后 D 的经验熵。以此作为可用性函数, 通过指数机制来对属性进行分割, 最后结合拉普拉斯机制实现差分隐私保护。图 8 说明了这种数据挖掘算法的框架。

另一种差分隐私的数据挖掘方案基于朴素贝叶斯分类器。朴素贝叶斯分类器是数据挖掘中一种常用的分类器, 朴素贝叶斯模型发源于古典数学理论, 有稳定的分类效率, 并且对小规模的数据表现很好, 能个处理多分类任务, 适合增量式训练, 尤其是数据量超出内存时, 我们可以一批批的去增量训练, 而且对缺失数据不太敏感, 算法也比较简单, 常用于文本分类。Vaidya 等人[43]基于朴素贝叶斯的概念, 提出了一种基于差分隐私的朴素贝叶斯模型。该方案分别对离散型分类属性和连续型分类属性的敏感度进行的计算。

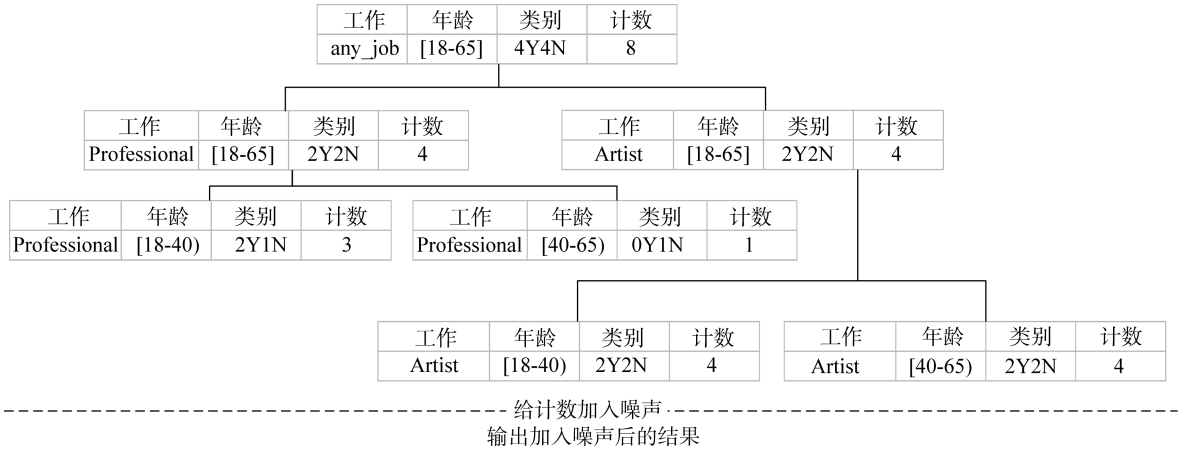


图 8　非交互式差分隐私数据挖掘框架

Figure 8　Non-interactive DP framework for data mining

对于离散型分类属性, 由于一个元素只能影响一条记录, 所以敏感度为 1, 而对于连续型分类属性, 需要通过正态分布来预测数据, 因此需要对均值 μ 和方差 δ 分别计算敏感度。Vaidya 等人[43]指出, 如果数据集中属性 X_j 在值域 $[l_j,u_j]$ 上时, 均值 μ 的敏感度为 $\frac{(u_j-l_j)}{n+1}$, 而方差 δ 的敏感度为 $\sqrt{(n)}\times\frac{(u_j-l_j)}{n+1}$。其中 n 为数据集的大小。在得到对应的参数的敏感度之后, 该方案通过在原始的朴素贝叶斯分类器的参数上加上由对应敏感度生成的噪声来实现差分隐私保护。

在机器学习中, 主要分为有监督学习和无监督学习。其中有监督学习中, 线性回归是一种简单有效的模型, 对于线性回归, Zhang 等人[36]提出了一种函数机制, 通过在线性回归模型中的代价函数的系数上加上噪声, 再通过梯度下降法等方法求解最优模型。该方案的优势在于提高了线性回归模型的精度, 但是一旦生成模型就是不可逆的。

对于分类算法, 支持向量机和 logisitc 回归是常用的方法。对于支持向量机, Chaudhuri 等人[37-38]提出了 2 点假设, 假设求解的参数是通过凸优化得到的, 并且一个样本点的变化会导致代价函数边界变化。在这 2 点假设下, Chaudhuri 等人提出了输出扰动机制和目标扰动机制。输出扰动机制首先训练模型, 然后在模型中加入噪声以实现差分隐私。目标扰动机制通过在代价函数中引入一个线性扰动项对代价函数进行扰动, 再从该代价函数中学习出加入噪声的支持向量机模型。对于 logistic 回归, Zhang 等人[36]还将函数机制应用在了 logistic 回归上, 通过对 logistic 回归的代价函数进行泰勒展开, 将其转化为多项式, 然后再对多项式的每一项系数加入噪声, 再通过梯度下降法等方法求解最优模型。

对于核函数支持向量机, 与线性支持向量机不同的是, 核函数支持向量机包含全部的训练数据, 为了避免公开数据, Rubinstein 等人[39]提出了一种隐私保护的核函数支持向量机, 该方案首先构造一个和隐私数据无关的空间, 然后将原始数据映射到构造的空间, 通过使用核函数来估计样本空间中的原始数据以避免公开发布原始数据。

在无监督学习中, k-均值方法是一种常用的聚类方法。Blum 等人[40]给出了提供差分隐私保护的 k-均值算法. 由于在计算每个记录与质心的距离时会泄露隐私, 因此在 SuLQ 框架下通过发布聚类质心和记录数量的估计值来满足隐私保护的要求, 但查询聚类质心的函数敏感度为聚类的最大直径, 而以此敏感度计算出的噪声量较大, 降低了聚类结果的可用性。针对这一问题, Kobbi 等人[18]提出了基于 Sample-Aggregate 框架的 k-均值聚类方法。但是该方案的问题在于, 只有取样的数据有一定的一致性时聚类结果才有较高可用性, 因此 Dan 等人[41]提出了基于核心集的方法来, 该方案通过从 d 维空间 G 中的 n 个数据点中计算出核心点集 $S_G\subseteq G$, 通过对该核心点集进行聚类能得到近似的结果。如表 3 所示, 表 3 总结了本节中说明的差分隐私算法的优缺点

综上所述, 这 9 种方法主要被应用于数据挖掘领域, 其中, Mohammed 等人提出的基于泛化技术的非交互模式匿名化数据挖掘算法属于隐私信息全生命周期中的隐私发布/存储/交换部分和隐私分析部分, 而基于差分隐私的朴素贝叶斯分类方法属于隐私分析部分。

表 3 差分隐私在数据挖掘中方案总结

Table 3 Summary of DP algorithm in data mining

方法名称	方案描述	优点	缺点
DiffID3	结合指数机制生成 ID3 决策树用于数据分析	容易实现, 分类准确率高	不易控制隐私预算参数分配
DP-Bayes	在朴素贝叶斯分类器中的参数上加入噪声实现差分隐私	模型简单, 易于实现, 能处理多分类任务	依赖于对数据分布的假设, 而且数据集越小噪声越大
Function Mechanism	先在代价函数中加入噪声再通过梯度下降等方法求解参数	准确性高, 模型简单	有可能加入噪声后导致代价函数无法收敛
输出扰动机制	首先训练模型, 然后在模型中加入噪声以实现差分隐私	适合较小数据集, 分类准确性高	SVM 中的系数会引起较大误差, 难以处理较大数据集
目标扰动机制	在代价函数中引入线性扰动项, 再从中学习 SVM 模型	适合较小数据集, 分类准确性高	计算代价较大
隐私保护核函数支持向量机	将原始数据映射到与其无关的空间, 用核函数来估计空间中的原始数据		
Sample-Aggregate	在数据集子集上执行查询, 最后合并结果并加入噪声	比拉普拉斯机制需要的噪声更小	取样的数据有一定一致性时结果才有较高可用性
核心集方法	从高维空间中计算核心点集, 对核心点集进行聚类	模型简单, 容易实现	敏感度高且难以计算, 需较大隐私预算才能保证精度
SuLQ-k-均值	发布聚类质心和记录数量的估计值	能够保护聚类质心位置, 模型简单	敏感度计算出的噪声量较大, 降低了聚类结果的可用性

6 差分隐私参数度量方法

在差分隐私模型中, 隐私预算参数 ϵ 是一个关键的参数, 对于这个参数的选取不能仅仅通过直觉。Lee 等人[44]提出了一种根据后验概率来度量参数 ϵ 所提供的隐私保护程度的方案。该方案定义了一种攻击模型, 攻击者的目的是猜测数据集 D 的相邻数据集 D', 在该攻击模型中, 先验概率定义为攻击者看到差分隐私机制发布的结果之前, 对于数据集 D' 的猜测正确的概率, 后验概率定义为攻击者看到差分隐私机制发布的结果之后, 再对于数据集 D' 猜测正确的概率。Lee 等人定义隐私泄露程度为后验概率与先验概率的差值, 差值越大, 说明保护程度越低。Lee 等人指出, 对于数据集相邻 D_i' 猜测正确的后验概率, 将其定义为 $\beta(D_i')$, 那么后验概率有如下等式:

$$\begin{aligned}\beta(D_i') &= \frac{1}{1+\sum_{k=1,k\neq i}^{n}\frac{e^{-\frac{|\gamma-f(D_k)|}{\lambda}}}{e^{-\frac{|\gamma-f(D_i)|}{\lambda}}}} \\ &\leqslant \frac{1}{1+\sum_{k=1,k\neq i}^{n}e^{-\frac{\Delta v}{\lambda}}} \\ &= \frac{1}{1+(n-1)\times e^{-\frac{\epsilon\Delta v}{\Delta f}}}\end{aligned} \tag{13}$$

其中, Δv 表示 $max|f(D_i')-f(D_k')|$, 指在所有的子数据集中进行查询所产生的最大差值。在给定隐私泄露程度后, 可以唯一的求解出参数 ϵ。该方案的缺陷在于, 这种方法在一定程度上依赖于数据集中数据的分布。因为在推导后验概率时, 经过了一步不等式的缩放, 用 $\sum_{k=1,k\neq i}^{n}\Delta v$ 代替了 $\sum_{k=1,k\neq i}^{n}|f(D_i)-f(D_k)|$, 根据数据集中数据分布的不同, 这样的缩放带来的误差也就不同。因此有时这一步会使得 ϵ 的理论上界远远大于实际上界。这时使用 ϵ 的理论上界会导致“过保护”的情况, 使得数据可用性降低。

Naldi 等人[45]提出了一种基于区间估算理论的度量方法。该方案中, 度量了加入噪声后的变量 $\hat{c}$ 和真实值 c 之间的距离, 用它们之间的距离来度量提供的隐私程度, 以此来计算合适的参数 ϵ。该方案中定义了一种对隐私保护级别的描述,

$$Pr[\hat{c}-wc<c<\hat{c}+wc]=p \tag{14}$$

其中, w 是一个控制 $\hat{c}$ 与 c 之间距离的参数, w 越大则说明 $\hat{c}$ 与 c 之间距离越大, 则隐私保护级别越高, 反之则说明隐私保护级别越低。而 p 则表示真实值落在 $[\hat{c}-wc,\hat{c}+wc]$ 之间的概率, 概率越大说明隐私保护级别越低, 反之说明隐私保护级别越高。Naldi 等人[45]给出了最终的参数 ϵ 的表达式为

$$\epsilon = -\frac{ln(1-p)}{wc} \tag{15}$$

综上所述, 这两种方法都是对差分隐私预算参数 ϵ 进行度量的, 主要应用于隐私信息全生命周期中的隐私分析部分。

7 差分隐私的其他应用

Mcsherry 等人[46]首次将差分隐私应用在推荐系统中, 在分析输入系统中的各项目之间的关系时, 他们先建立项目相似度协方差矩阵, 然后向矩阵中加入 Laplace 噪声, 最后将加入噪声的协方差矩阵提交给推荐系统执行常规推荐算法。Machanavajjhal 等人[47]将差分隐私应用于社交网络中, 在社交网络中往往用图来表示用户之间的关系, 用图中的结点表示用户, 结点之间的边表示用户之间的关系。用户之间的关系往往是敏感信息, 所以 Machanavajjhal 等人以结点上的邻居数为可用性函数, 用指数机制随机生成边, 实现差分隐私保护。

Mcsherry 等人[48]还开发的一套为隐私敏感数据提供差分隐私保护的框架。PINQ 框架提供一系列的 API, 便于可以自己根据需求编写程序来使用 PINQ 框架进行相关的差分隐私保护系统开发, 框架中还提供了丰富的差分隐私数据分析的应用实例。但是 PINQ 框架的缺点在于, 无法防止隐蔽信道攻击等多种攻击。

针对 PINQ 框架的缺点, Haeberlen 等人[49]设计了一个新的差分隐私的工具集模型 Fuzz, 该模型也可以让用户自主编程实现所需功能。它主要由 3 部分组成, 分别为程序语言编译器, 类型检查器以及查询处理器。其中, 类型检查器的作用是在执行一个查询前判断现有的可用预算能否进行查询, 最后将查询交由查询处理器执行, 返回差分隐私查询结果。Fuzz 的框架如图 9 所示。

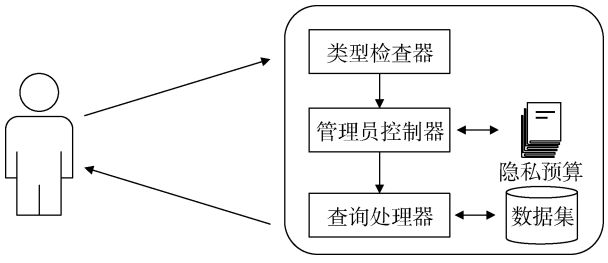


图 9 Fuzz 框架

Figure 9 Framework of Fuzz

综上所述, Mcsherry 等人[48]提出的基于差分隐私的推荐系统和 Machanavajjhal 等人[49]提出的基于差分隐私的社交网络方法, 属于隐私信息的全生命周期的隐私保护和隐私发布/存储/交换部分, 而 Dwork 等人[23]和 Haeberlen 等人[49]设计的差分隐私框架属于隐私保护部分。

8 结束语

差分隐私保护是一种通用、灵活、具有坚实的数学理论支撑的隐私保护方法, 可以用来解决很多传统密码学不适合甚至不可行的问题, 实现了隐私的语义安全, 因此引起越来越多研究者的兴趣。差分隐私保护在近两年取得了飞速的发展, 随着大数据, 人工智能领域的兴起, 差分隐私的被越来越多的应用在这些场景下。

1) 差分隐私分为交互式差分隐私和非交互式差分隐私, 交互式差分隐私主要通过设计接口的方式, 在查询结果上加入噪声来实现隐私保护。而非交互式的差分隐私保护机制需要直接将数据集通过隐私保护算法处理后发布, 以满足用户的查询需求。虽然已经有一些非交互式的差分隐私算法, 但对于如何降低计算复杂度, 如何处理数值型数据以及如何提供更广泛的查询类型等问题上还有待进一步研究。

2) 差分隐私虽然现在已经被用于数据挖掘, 推荐系统等领域, 但是差分隐私对于挖掘数据保护后, 还能对数据分析者提供多少可用信息, Fredrikson 等人[50]通过实验分析了差分隐私数据挖掘算法和常规数据挖掘算法对于药物剂量预测以及病人身体状况的差距, 但是目前还没有一个合理通用的度量方法。

另外, 机器学习中往往需要对大量的数据进行分析, 这些数据中往往包含着大量的用户隐私信息, 而随着机器学习的发展, 差分隐私与机器学习的结合将是未来的一个研究热点。因此, 在差分隐私和机器学习中, 主要有以下问题需要解决。

3) 如何解决样本数据集中缺失数据的问题, 传统机器学习方法不能满足差分隐私的需求。

4) 医疗数据集中, 很多体征数据只是暂时的, 例如化验结果只能代表病人这一时期的体征, 多次的化验结果之间没有相关性, 而且对于数据的扰动很有可能使数据失去重要的信息, 因此需要有应对这种类型数据的差分隐私模型。

5) 隐私是否能在不牺牲机器学习模型可用性的条件下实现。一种已有的思路是, 当噪声大小小于样本抽样的随机性的时候, 噪声对隐私的就不会产生影响。

6) 在正则化的机器学习模型中, 差分隐私是否可以与正则化的想法兼容。

参考文献

[1] FH Li, H Li, Y Jia, NH Yu, J Weng, “Privacy computing: concept, connotation and its research trend”, *Journal on Communications,* vol. 37, no. 4, pp. 1-11, 2016.
(李凤华, 李晖, 贾焰, 俞能海, 翁建. "隐私计算研究范畴及发展趋势." *通信学报* 37.4(2016):1-11.)

[2] Latanyasweeney. “k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY.” *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 10.05(2002):557-570.

[3] Latanyasweeney. “Achieving k-anonymity Privacy Protection Using Generalization, and Suppression.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05(2002): 571-588.

[4] Kumar, P. Mayil Vel, and M. Karthikeyan. “L Diversity on K-Anonymity with External Database for improving Privacy Preserving Data Publishing.”*International Journal of Computer Applications* 54.14 (2012): 7-13.

[5] Machanavajjhala, Ashwin, D. Kifer, and J. Gehrke. “L - diversity: Privacy beyond k -anonymity.” *Acm Transactions on Knowledge Discovery from Data* 1.1(2007):3.

[6] Wang, Qian, Z. Xu, and S. Qu. "An Enhanced K-Anonymity Model against Homogeneity Attack." *Journal of Software* 6.10(2011): 1945-1952.

[7] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity." IEEE, International Conference on Data Engineering IEEE, 2007: 106-115.

[8] Dwork, Cynthia. "Differential privacy." *International Colloquium on Automata, Languages, and Programming* Springer, Berlin, Heidelberg, 2006:1-12.

[9] Dwork, Cynthia. "Differential Privacy: A Survey of Results." *International Conference on Theory and Applications of MODELS of Computation*Springer-Verlag, 2008:1-19.

[10] Dwork, Cynthia. "The Differential Privacy Frontier (Extended Abstract)." *Theory of Cryptography Conference* Springer Berlin Heidelberg, 2009:496-502.

[11] Dwork, Cynthia. "Differential privacy in new settings." *Acm-Siam Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, Usa, January* DBLP, 2010:174-183.

[12] Dwork, Cynthia. *A firm foundation for private data analysis*. ACM, 2011.

[13] Dwork, Cynthia. "The Promise of Differential Privacy: A Tutorial on Algorithmic Techniques. " *Foundations of Computer Science* IEEE, 2011:1-2.

[14] Dwork, Cynthia, and J. Lei. "Differential privacy and robust statistics." *ACM Symposium on Theory of Computing* ACM, 2009:371-380.

[15] Mcsherry, Frank D. "Privacy integrated queries: an extensible platform for privacy-preserving data analysis. " *Communications of the Acm*53.9(2010):89-97.

[16] Kifer, Daniel, and B. R. Lin. "Towards an axiomatization of statistical privacy and utility." *Twenty-Ninth ACM Sigmod-Sigact-Sigart Symposium on Principles of Database Systems, PODS 2010, June 6-11, 2010, Indianapolis, Indiana, Usa* DBLP, 2010:147-158.

[17] Dwork, Cynthia, F. Mcsherry, and K. Nissim. "Calibrating Noise to Sensitivity in Private Data Analysis." *Proceedings of the Vldb Endowment* 7.8(2006):637-648.

[18] Nissim, Kobbi, and S. Raskhodnikova. "Smooth sensitivity and sampling in private data analysis." *Thirty-Ninth ACM Symposium on Theory of Computing* ACM, 2007:75-84.

[19] Li, Chao, et al. "The matrix mechanism: optimizing linear counting queries under differential privacy." *Vldb Journal — the International Journal on Very Large Data Bases* 24.6(2015):757-781.

[20] Mcsherry, Frank, and K. Talwar. "Mechanism Design via Differential Privacy." *Foundations of Computer Science, 2007. FOCS '07. IEEE Symposium on* IEEE, 2007:94-103.

[21] Blum, Avrim, K. Ligett, and A. Roth. "A learning theory approach to noninteractive database privacy." *Journal of the Acm* 60.2(2011):1-25.

[22] Hardt, Moritz, and G. N. Rothblum. "A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis." *IEEE, Symposium on Foundations of Computer Science* IEEE Computer Society, 2010:61-70.

[23] Dwork, Cynthia, and A. Roth. "The Algorithmic Foundations of Differential Privacy." *Foundations & Trends® in Theoretical Computer Science*9.3(2014): 211-407.

[24] Qardaji, Wahbeh, W. Yang, and N. Li. "Understanding hierarchical methods for differentially private histograms." *Proceedings of the Vldb Endowment* 6.14(2013): 1954-1965.

[25] Hay, Michael, et al. "Boosting the accuracy of differentially private histograms through consistency." *Proceedings of the Vldb Endowment*3.1-2(2010):1021-1032.

[26] Xiao, Xiaokui, G. Wang, and J. Gehrke. "Differential Privacy via Wavelet Transforms." *IEEE Transactions on Knowledge & Data Engineering* 23.8(2009):1200-1214.

[27] Xu, Jia, et al. "Differentially Private Histogram Publication." *Vldb Journal* 22.6(2013):797-822.

[28] Xiao, Yonghui, et al. "DPCube: Differentially Private Histogram Release through Multidimensional Partitioning." *Transactions on Data Privacy* 7.3(2014):195-222.

[29] Acs, Gergely, C. Castelluccia, and R. Chen. "Differentially Private Histogram Publishing through Lossy Compression." *IEEE, International Conference on Data Mining* IEEE Computer Society,

2012:1-10.

[30] Nath, Suman, and S. Nath. "Differentially private aggregation of distributed time-series with transformation and encryption." *ACM SIGMOD International Conference on Management of Data* ACM, 2010:735-746.

[31] Qardaji, Wahbeh, W. Yang, and N. Li. "Differentially private grids for geospatial data." (2012):757-768.

[32] Chen R, Xiao Q, Zhang Y, et al. Differentially Private High-Dimensional Data Publication via Sampling-Based Inference[J]. 2015:129-138.

[33] Chen, R., Fung, B. C. M., and Desai, B. C. Differentially private trajectory data publication. CoRR (2011), –1–1.

[34] Li Y D, Zhang Z, Winslett M, et al. Compressive mechanism: utilizing sparse representation in differential privacy[C]// ACM, 2011:177-182.

[35] Machanavajjhala A, Kifer D, Abowd J, et al. Privacy: Theory meets Practice on the Map[C]// IEEE, International Conference on Data Engineering. IEEE Computer Society, 2008:277-286.

[36] Zhang Z, Zhang Z, Yang Y, et al. Functional mechanism: regression analysis under differential privacy[J]. Proceedings of the Vldb Endowment, 2012, 5(11):1364-1375.

[37] Chaudhuri K, Monteleoni C. Privacy-preserving logistic regression[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2008:289-296.

[38] Chaudhuri K, Monteleoni C, Sarwate A D. Differentially Private Empirical Risk Minimization.[J]. Journal of Machine Learning Research Jmlr, 2011, 12(2):1069.

[39] Rubinstein B I P, Bartlett P L, Ling H, et al. Learning in a Large Function Space: Privacy-Preserving Mechanisms for SVM Learning[J]. Eprint Arxiv, 2012, 4.

[40] Blum A, Dwork C, Mcsherry F, et al. Practical privacy: the SuLQ framework[C]// Twenty-Fourth ACM Sigmod-Sigact-Sigart Symposium on Principles of Database Systems. ACM, 2005:128-138.

[41] Dan F, Fiat A, Kaplan H, et al. Private coresets[J]. Proceedings of the Annual Acm Symposium on Theory of Computing, 2009: 361-370.

[42] Mohammed, Noman, et al. “Differentially private data release for data mining.” *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* ACM, 2011:493-501.

[43] Vaidya, Jaideep, et al. "Differentially Private Naive Bayes Classification." *Ieee/wic/acm International Joint Conferences on Web Intelligence* IEEE Computer Society, 2013:571-576.

[44] Lee, Jaewoo, and C. Clifton. "How Much Is Enough? Choosing ε for Differential Privacy." *Information Security, International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings* DBLP, 2011: 325-340.

[45] Naldi, Maurizio, and G. D'Acquisto. "Differential Privacy: An Estimation Theory-Based Method for Choosing Epsilon." *Computer Science* (2015).

[46] Mcsherry, Frank, and I. Mironov. "Differentially private recommender systems: building privacy into the net." *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* ACM, 2009:627-636.

[47] Machanavajjhala, Ashwin, A. Korolova, and A. D. Sarma. "Personalized social recommendations: accurate or private." *Proceedings of the Vldb Endowment* 4.7(2011): 440-450.

[48] Mcsherry, Frank D. "Privacy integrated queries." *Communications of the Acm* 53.9(2009).

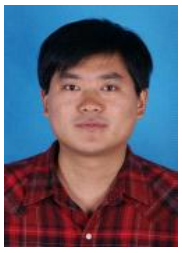
[49] Haeberlen, Andreas, B. C. Pierce, and A. Narayan. "Differential privacy under fire." *Usenix Conference on Security* USENIX Association, 2011:33-33.

[50] Fredrikson, Matthew, et al. "Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing." *Proceedings of the. USENIX Security Symposium. UNIX Security Symposium* 2014(2014): 17.

李效光 于 2017 年在西安电子科技大学信息安全专业获得学士学位。现在西安电子科技大学网络空间安全专业攻读博士学位。研究领域为隐私保护。研究兴趣包括: 差分隐私、数据挖掘。

李晖 于 1998 年在西安电子科技大学通信与电子系统专业获得博士学位。现任西安电子科技大学网络与信息安全学院执行院长。研究兴趣包括: 云计算中的密码理论与安全协议、移动互联网的隐私保护、信息论与编码理论。

李凤华 于 2009 年在西安电子科技大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所副总工程 师、研究员、博士生导师, 主要研究方向为 网络与系统安全、隐私计算、信息保护。



朱辉 于 2009 年在西安电子科技大学信息安全专业获得博士学位。现任西安电子科技大学网络与信息安全学院教授。研究兴趣包括: 数据安全与隐私保护、安全方案及协议设计、网络及应用安全。