基于属性权重的隐私泄露风险评估方法

罗 康[1,2],彭长根[1,2,3],谭伟杰[3],杨 训[2]
(1. 贵州大学 计算机科学与技术学院,贵州 贵阳 550025;
2. 贵州大学 贵州省公共大数据重点实验室,贵州 贵阳 550025;
3. 贵州大学 大数据产业发展应用研究院,贵州 贵阳 550025)

摘 要:数据隐私泄露风险评估在数据发布中起着关键的作用。当前在数据发布场景中,避免数据隐私泄露的方法主要是通过数据加噪、混淆、匿名等隐私保护方法进行数据隐私保护,然后进行数据发布共享。但是依然存在数据接收者利用已有数据或者背景知识对发布数据进行攻击,获取个人隐私信息的情况发生。现有的方法对于数据隐私保护后的数据隐私泄露风险却没有进行较好的定量评估。针对该问题,提出了一种面向数据发布场景的隐私泄露风险量化评估方法。首先将数据发布者的数据进行数值化处理,经过映射得到敏感数据矩阵;然后获取需求者对数据的具体需求进行隐私等级标记,计算得到字段属性隐私权重;接着结合隐私权重利用矩阵范数计算得出隐私风险值;最后,通过公开数据集进行评估方法的验证对比分析,验证了评估方法的准确性与效用性。该方法对于降低发布数据的隐私泄露风险,提高数据的共享度具有一定意义。

关键词:等级标记;隐私评估;泄露风险;数据发布;隐私保护
中图分类号:TP309 **文献标识码**:A **文章编号**:1673-629X(2022)10-0082-06
doi:10.3969/j.issn.1673-629X.2022.10.014

Privacy Disclosure Risk Assessment Method Based on Attribute Weight

LUO Kang[1,2], PENG Chang-gen[1,2,3], TAN Wei-jie[3], YANG Xun[2]
(1. School of Computer Science & Technology, Guizhou University, Guiyang 550025, China;
2. Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China;
3. Guizhou Big Data Academy, Guizhou University, Guiyang 550025, China)

Abstract: The risk assessment of data privacy disclosure is a highly key link in data release. At present, in data publishing scenarios, the main methods to avoid data privacy disclosure are to protect data privacy through privacy protection methods such as adding noise, obfuscation and anonymity, and then publish and share data. However, there are still cases in which the data receiver uses the existing data or background knowledge to attack the released data and obtain personal privacy information. Therefore, it is still a difficult problem to quantitatively evaluate the risk of data privacy disclosure. To solve this problem, a quantitative evaluation method of privacy disclosure risk for data publishing scenario is proposed. Firstly, the published data will be digitized to obtain the sensitive data matrix through mapping. Secondly, the specific needs of the demander for the data is obtained to mark the privacy level, which can calculate the privacy weight of the field attribute. Then, combined with the privacy weight, the privacy risk value is calculated by using the matrix norm. Finally, the evaluation methods are verified, compared and analyzed through the public data set to verify the accuracy and effectiveness of the evaluation methods. The proposed method has a certain reference value for reducing the risk of privacy disclosure of published data and enhancing data sharing.

Key words: grade marking; privacy assessment; risk of leakage; data release; privacy protection

0 引 言

移动互联网技术促使数字经济的快速发展,各个行业产生了海量的数据。随着数据的沉淀与积累,数据的研究和利用价值日渐突显。基于人工智能等技术对大数据的分析,导致个人隐私泄露变得越来越严重[1-3]。由闪捷《2020 年度数据泄漏态势分析报告》可

收稿日期:2021-11-08 **修回日期**:2022-03-10
基金项目:国家自然科学基金(U1836205)
作者简介:罗 康(1994-),男,硕士研究生,研究方向为隐私保护、隐私评估;通信作者:彭长根(1963-),男,博士,教授,博导,CCF 高级会员(48309S),研究方向为密码学、信息安全、大数据隐私保护等。

以看出隐私泄露的案例屡见不鲜,针对隐私信息的保护和隐私泄露风险的评估成为当前的紧迫需求。云服务实现了数据的计算存储与发布,科研机构、银行、医疗机构等单位在数据安全的情况下愿意共享一些数据用作数据的挖掘与技术的研发[4]。在数据共享发布时,为了防止泄露数据中的隐私信息,经常利用一些隐私保护方法[5-7]对数据隐私采取脱敏处理,但是数据需求者获得发布数据后容易发生与已有数据进行数据融合,链接推理出完整的数据信息的情况,造成隐私泄露。因此如何评估数据发布后隐私泄露的风险和控制数据隐私保护的强度成为数据发布领域亟需解决的痛点问题。

2016 年彭长根等人[8]针对基于信息熵进行隐私度量存在的理论体系零散和缺乏统一模型基础的问题,探讨将隐私保护系统视为一种通信模型,用于搭建一个可行的体系基础解决隐私保护系统的量化问题。2017 年 Yeh 等人[9]提出了一个名为 AppLeak 的分析框架,以有效评估信息丢失并检测 Android 应用程序运行期间的隐私泄露。2018 年晏燕等人[10]运用集对分析的五元偏联系数理论,建立隐私风险待评估指标评估系统,此方法消除了隐私泄露风险中的模糊、不确定等因素的影响,实现了对风险评估指标的动态体现。2020 年,周旭晨等人[11]为了促进原始数据的共享,给数据开放者提供隐私风险定量化的评估,提出了一种利用矩阵计算来评估数据隐私泄露风险的方法。2021 年谢小杰等人[12]针对隐私评估大多关注隐私数据脱敏效果,对于社交网络中的隐私问题研究却相对较少等问题,提出了社交网络用户中的隐私泄露量化评估方法,对减少社交网络的隐私风险起到良好的预防效果。

根据 ISO/IEC 29100(2011)信息技术-安全技术-隐私架构框架国际标准[13],风险管控是隐私保护过程中最为核心的关键方法,并且特别提到隐私控制的识别也应该包含在组织信息安全管理框架中。通过 ISO/IEC 29100(2011)隐私架构框架可以看出,建立数据的隐私风险评估方法是必要的,对于数据的保护具有至关重要的作用。

隐私评估方法从早期的专家评分为主导的主观评价慢慢转向通过贝叶斯、攻击模型等客观的隐私泄露风险度量方法,评估隐私泄露风险的方法更加多元化。但是目前而言,针对数据共享领域的隐私风险探索相对较少,但是可以从其他的隐私评估方法中找到借鉴。针对数据发布共享中的隐私风险问题,目前多数方法只是针对敏感数据进行隐私度量,然后进行评估,然而结合数据需求方与数据发布方具体应用目标的评估方法却很少。因此提出一种新的评估方法,将引入数值化映射和隐私标记方法,通过对属性字段,操作进行隐私等级标记,然后对数据进行数值化映射形成敏感数据矩阵,最后对隐私标记与敏感数据矩阵作融合,计算出数据隐私泄露风险值。从而实现数据发布方与数据需求方相结合的隐私泄露风险评估方法,实现了隐私泄露风险的预防,为提高数据共享安全效能带来一定的支撑。

1　基础知识

1.1　矩阵范数

矩阵范数是线性代数、矩阵论中的基本概念,表示矩阵到实数的映射/函数。矩阵的 1 范数,即:$\|\boldsymbol{A}\|_1=\max\limits_j\sum\limits_{i=1}^{m}|a_{i,j}|$ 将矩阵中的各列中的元素取绝对值相加,从中获取最大值;∞ - 范数,即:$\|\boldsymbol{A}\|_\infty=\max\limits_j\sum\limits_{i=1}^{n}|a_{ij}|$ 将矩阵中的各行中的元素取绝对值相加,从中获取最大值;F-范数,即:$(\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|^2)^{1/2}$ 计算方法为将矩阵中的每个元素取绝对值的平方,然后求和再进行开方运算。

1.2　差分隐私

定义 1　ε - 差分隐私[14]:设具有属性结构相同的数据集 D 与 D',两数据集间存在最多一条有差别的数据元组,则称 D 与 D' 为邻近数据集或者兄弟数据集。

给定随机算法 K,PK 是 K 输出的所有可能组成的集合。对于任意兄弟数据集 D 和 D' 以及 PK 的任意子集 SK,如果算法 K 符合:

$$\Pr[K(D)\in \mathrm{SK}]\leqslant \exp(\varepsilon)*\Pr[K(D')\in \mathrm{SK}] \tag{1}$$

那么称 K 提供 ε - 差分隐私保护。ε 为隐私保护预算,反映出隐私保护的强度,ε 数值越大,隐私保护强度越弱。

1.3　拉普拉斯噪声

定义 2　Laplace 分布:以 η 变量作为满足 Laplace 分布的连续型随机变量,Laplace 分布满足期望为 0,方差为 $2\lambda^2$,它的概率密度函数为:

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

1.4　隐私标记

字段隐私等级[11]标记:某医疗机构疾病患者信息数据集包括年龄、医生、疾病等字段,机构对发布数据集字段进行标记,如表 1 所示($N=5$,最高隐私等级为 5)。

操作等级隐私标记:某医疗机构数据集中提供的

操作有取平均、取值、求最值等,医疗机构分别对操作设置相应等级标记,如表 2 所示($N=3$,最高隐私操作等级为 3)。

表 1 某医院机构数据字段隐私等级标记

字段	隐私等级
年龄	3
医生	3
疾病	4

表 2 某医疗机构数据集操作隐私等级标记

操作	隐私等级
取平均	3
取值	3
求最值	2

2 隐私标记与隐私度量

2.1 属性字段权重计算

根据属性的特点,按照属性元素敏感度越敏感,数值映射越大的原则对数据的每个敏感属性进行数值化处理,进而建立敏感数据矩阵。但是敏感数据矩阵中,每种敏感属性在实际中,敏感性却存在差异。例如年龄和诊断结果,很明显诊断结果对于个人的隐私显得更加重要。数据使用者的不同操作需求也会增大数据隐私泄露的风险,例如取值操作明显就比求平均值风险更高,取值将会对具体的某个用户值进行操作,提高了隐私泄露的风险程度。因此通过用户具体的数据使用需求建立数据字段的隐私泄露风险权重,对于更加准确高效地评估数据的隐私风险显得尤为重要。

定义 3 隐私权重向量:设每个敏感属性的隐私权重为 p_i ,则 $p=(p_1,p_2,\cdots,p_n)$ 。

获取数据开放者拟支持的 m 种操作和拟开放的 n 种字段,对操作建立隐私等级标记,记为 $q_1,q_2,\cdots,q_m$ ($N=3$,等级依次提高);对字段隐私等级进行标记,记为 $f_1,f_2,\cdots,f_n$ ($N=5$,等级依次提高)。然后以字段为行,操作为列建立 $n\times m$ 的矩阵 $\boldsymbol{M}$,根据使用需求,将第 i 个字段和第 j 个操作的值 t_{ij} ($t_{ij}=q_i\times f_j$)赋予矩阵中,未涉及到的需求使用 0 填充。

$$\boldsymbol{M}=\begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,m} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,m} \\ \vdots & \vdots & & \vdots \\ t_{n,1} & t_{n,2} & \cdots & t_{n,m} \end{bmatrix}$$

对 $\boldsymbol{M}$ 矩阵行进行求和得到向量 $\boldsymbol{p}=(p_1,p_2,\cdots,p_n)$, $p_i=\sum_{j=1}^{m}t_{i,j}$, $\boldsymbol{p}$ 即为隐私敏感属性字段的隐私权重向量。

2.2 结构化数据处理

从隐私保护的角度出发,结构化数据属性中主要可以分为四类,即标志性属性、准标志属性、敏感属性与非敏感属性。由于标志性属性会进行删除,非敏感属性不属于隐私保护的范围,准标志已有相应的处理方法,因此敏感属性将作为处理的重点。敏感属性结构化数据见表 3。表中 D_i 表示用户个体, SA_j 表示感属性, $\boldsymbol{d}_{i,j}$ 表示敏感属性值。

表 3 敏感属性结构化数据

	SA_1	SA_2	…	SA_n
D_1	$d_{1,1}$	$d_{1,2}$	…	$d_{1,n}$
D_2	$d_{2,1}$	$d_{2,2}$	…	$d_{2,n}$
⋮	⋮	⋮	…	⋮
D_m	$d_{m,1}$	$d_{m,2}$	…	$d_{m,n}$

定义 4 非负数值化映射[15]:设存在映射 f ,非负数值元素集 E ,若 $\forall\ e\in E$, $f(e)\geqslant 0$,则 f 为非数值化映射。

针对表 3 的敏感属性进行非负值数值化映射,敏感属性映射值随着属性敏感性的增大而增大。将表中敏感属性分别对应 $f_1,f_2,\cdots,f_n$ 进行数值化映射。通过以下计算方法:

$$a_{i,j}=f(d_{i,j})\quad(1\leqslant i\leqslant m,1\leqslant j\leqslant n,\text{且 } m,n\in N^*) \tag{3}$$

将敏感数据进行非负数值化处理,得到 $D=(a_{i,j})_{m\times n}$ 的敏感数据矩阵。

2.3 数据效用

数据效用是指敏感数据经过隐私保护技术手段脱敏处理后,与原来未经处理后的数据相比,所具有的真实性或者相同性的程度。当隐私保护后,往往会降低数据的使用性。因此通过效用的度量来评价隐私脱敏处理后数据的使用性。数据效用性越高数据的真实性越好,数据的使用价值就越高。

定义 5 数据效用[16]:设敏感数据矩阵为 $\boldsymbol{D}$,将 $\boldsymbol{D}$ 经过隐私脱敏处理后的敏感数据矩阵为 $\boldsymbol{D}'$, $\boldsymbol{D}$ 与 $\boldsymbol{D}'$ 结构相同。$U(\boldsymbol{D})$ 与 $U(\boldsymbol{D}')$ 分别表示 $\boldsymbol{D}$ 和 $\boldsymbol{D}'$ 的数据量,数据效用计算公式如下:

$$R=U(\boldsymbol{D}')/U(\boldsymbol{D}) \tag{4}$$

其中, R 为数据脱敏处理后数据的效用, $U(\boldsymbol{D})=\boldsymbol{D}\odot\boldsymbol{D}_F$, $U(\boldsymbol{D}')=\boldsymbol{D}\odot\boldsymbol{D}'_F$,符号 $\odot$ 表示:

$$\boldsymbol{A}\odot\boldsymbol{B}=(c_{i,j})_{m\times n},c_{i,j}=\begin{cases} a_{i,j}/b_{i,j} & b_{i,j}\neq 0 \\ 1 & b_{i,j}=0 \end{cases} \tag{5}$$

3 属性权重的隐私泄露风险评估方法

3.1 隐私风险泄露评估系统的方案设计

隐私评估系统设计如图 1 所示。该系统主要是建

立在知道具体需求的前提下,首先建立隐私等级标记矩阵;然后对发布数据采取数值化函数处理创建敏感数据矩阵;接着将隐私等级标记矩阵与敏感数据矩阵相结合进行矩阵范数计算得到隐私风险值。通过上述步骤建立数据发布的泄露风险评估系统。

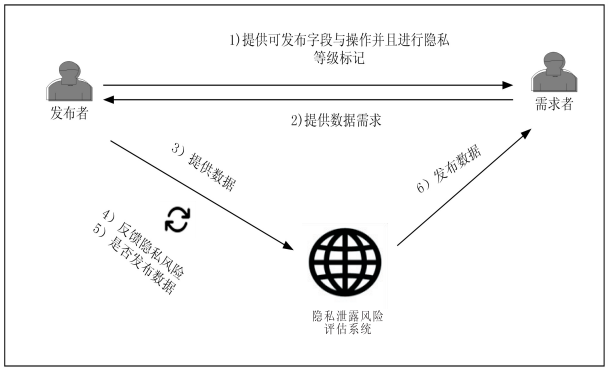

图 1　隐私评估系统整体框架

针对需求数据集隐私泄露风险高的问题,评估系统通过引入加噪、泛化等隐私保护方法,对数据进行脱敏处理,降低隐私泄露的风险,实现数据发布的低风险。

3.2　评估系统详细流程

数据共享成为当前的现实需求,为促进发展,数据拥有者愿意在一定条件下开放某些数据给某些科研机构进行研究,但是隐私信息安全一直是数据拥有者最为关心的问题,因此建立一套隐私泄露风险评估系统是十分必要的。隐私泄露风险评估方法作为该系统的核心,能够有效预防数据发布过程中的隐私泄露。

具体评估流程如下:

(1)数据需求方的隐私泄露风险评估首先是通过获取数据发布方拟支持的 m 种操作和拟开放的 n 种字段,对操作建立隐私等级标记,记为 $q_1,q_2,\cdots,q_m$ ($N=3$,等级依次提高);对字段隐私等级进行标记,记为 $f_1,f_2,\cdots,f_n$ ($N=5$,等级依次提高)。

(2)以字段为行,操作为列建立 $n\times m$ 的矩阵 $\boldsymbol{M}$,遍历数据需求者使用需求集合 G,查询第 i 个字段与第 j 个操作是否存在于需求表中。如果存在,则将 t_{ij} ($t_{ij}=q_i\times f_j$)的值赋予矩阵中。其余矩阵中的值用 0 填充。

(3)对 $\boldsymbol{M}$ 矩阵行进行求和得到向量 $\boldsymbol{p}=(p_1,p_2,\cdots,p_n)$,$p_i=\sum_{j=1}^{m}t_{i,j}$,$\boldsymbol{p}$ 即为隐私敏感属性字段的隐私权重向量。

(4)根据定义 5,将敏感属性数据进行非负值数值化映射,按照敏感属性的特性,敏感数值映射随着敏感度的增加而增大。该文选取学历(education)、工作时长(hours-per-week)、收入(income)敏感属性进行举例说明:

education:Doctor → 0.8, Master → 0.7, Bachelor → 0.6, High School → 0.5, Junior High School → 0.4, Primary School → 0.3, other → 0.1。

hours-per-week:

$$f(\text{hours-per-week})=\begin{cases}\dfrac{|\text{hours-per-week}-40|}{40},[0,80]\\ 1,\text{other}\end{cases}\tag{6}$$

$$\text{income}:f(\text{incomes})=\begin{cases}0.3,[0,50)\\ 0.1,[50,50\ k)\\ 0.8,[50\ k,+\infty)\end{cases}\tag{7}$$

然后通过数值化构建隐私敏感数据矩阵 $\boldsymbol{D}$。

(5)将敏感数据矩阵 $\boldsymbol{D}$ 与隐私权重向量 p 合成带有权重向量的数据矩阵 $\boldsymbol{B}$。合成方法如下:

$$\boldsymbol{B}=\boldsymbol{D}\oplus\boldsymbol{p}=\begin{bmatrix}a_{1,1}*p_1 & a_{1,2}*p_2 & \cdots & a_{1,n}*p_n\\ a_{2,1}*p_1 & a_{2,2}*p_2 & \cdots & a_{2,n}*p_n\\ \vdots & \vdots & \cdots & \vdots\\ a_{m,1}*p_1 & a_{m,2}*p_2 & \cdots & a_{m,n}*p_n\end{bmatrix}\tag{8}$$

通过合成得到 $\boldsymbol{B}$,将敏感数据矩阵 $\boldsymbol{B}$ 进行范数计算 $|\boldsymbol{B}|=\boldsymbol{B}_F$,对数据进行归一化处理,得到隐私风险泄露系数 $r=\dfrac{|\boldsymbol{B}|_{\text{sum}}-\boldsymbol{B}_{\min}}{|\boldsymbol{B}|_{\max}-|\boldsymbol{B}|_{\min}}$。得到的隐私泄露风险系数范围是[0,1],因此可以根据该范围建立相对应的隐私风险评估等级。

3.3　隐私泄露评估算法实现

结合上面的隐私泄露风险评估方法,该文的隐私评估算法的伪代码如下所示:

算法 1:隐私泄露风险评估算法。

输入:数据使用者需求集合 G,数据发布字段隐私等级集合 S,数据操作隐私等级集合 H,发布数据集 data。

输出:隐私泄露风险系数 r。

(1) matrix = Create(n, m);

//建立 $m\times n$ 的矩阵,默认值为 0

(2) FOR　item $\in$ G;

//遍历使用者需求结合

(3) PL(item, matrix, S, H);

//通过查询字段隐私等级集合 S 与数据操作隐私等级集合 H,对第 i 行与第 j 列进行隐私等级赋值

(4) END FOR

(5) $\boldsymbol{p}$ = CoSum(matrix)

//对矩阵行求和,得到向量 $\boldsymbol{p}$

(6) dataX = init(data);

//对数据进行处理,去除掉标识符等非敏感数据

(7) numDataMatrix = GetNumData(SenData);

//对数据进行非负值映射,得到敏感数据矩阵

(8) weightMatrix = numDataMatrix $\oplus$ $\boldsymbol{p}$

//将敏感数据矩阵与权重向量合成带有权重的敏感数据矩阵

(9) AmountPrivacy = GetPrivacy( );

//通过范数计算,获取数据隐私量

(10) r = CalculatedRisk(AmountPrivacy)

//通过数据归一化,计算风险系数得到 r

(11) output: r

//输出 r

(12) return

4 实验结果与分析

4.1 实验设置

该文选用公开数据集 UCI 中的 Adult 进行实验分析,其中包含 15 个属性字段,32 670 条数据,通过进行实验分析,验证该方法的效用性和正确性。

实验环境具体如下:实验算法采用 Python 进行开发,操作系统采用基于 x64 的 Windows10,内存为 24G,CPU 为 Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 GHz。

4.2 实验分析

实验首先验证在评估数据、操作方式、字段隐私等级不变的状态下,增加字段需求数量,数据隐私泄露风险值的变化。采用 Adult 数据集中的 15 个字段进行实验,通过数值化处理转化为敏感数据矩阵 $\boldsymbol{D}_1$,字段操作隐私等级标记参考表 2,字段隐私等级按照数据发布者的标记,按照字段隐私泄露的风险程度进行隐私标记,等级越高风险越大,分别标记 1 ~ 5,方法参照表 1。通过实验,结果如图 2 所示。

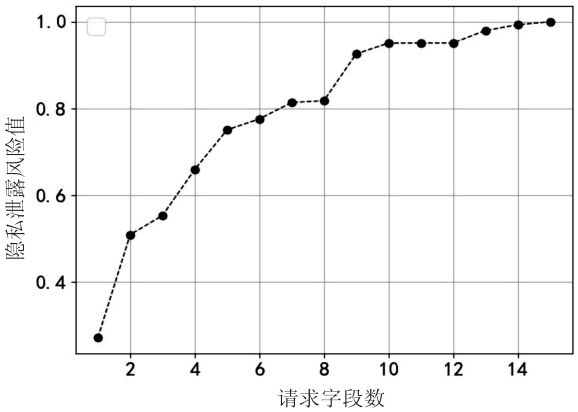

图 2 隐私风险值随请求字段数的变化

根据图 2 的结果表明,在评估数据、字段操作需求、隐私等级不变的情况下,随着字段需求的增加,数据隐私泄露风险值也随字段数量的增加而增大。

接着分析在字段操作、隐私等级、字段需求不变的情况下,对数据添加噪声是否可以降低隐私泄露的风险值,对数据效用的影响情况。为了产生与拉普拉斯同分布的随机数,实验采用了拉普拉斯的逆累积分布函数,随机数据的生成如下:

$$y = \mu - \frac{\delta}{\sqrt{2}}\mathrm{sgn}(\beta) \times \ln(1 - 2|\beta|) \tag{9}$$

其中,β 为在区间[-0.5,0.5]中均匀分布的一个随机数,设均值 $\mu = 0$,δ 为标准差,δ 代表噪声量,δ 越大,噪声越大。

对敏感数据矩阵 $\boldsymbol{D}_1$ 中的元素添加拉普拉斯噪声,首先生成满足 $\boldsymbol{D}_1$ 矩阵大小的噪声矩阵 $\triangle \boldsymbol{D}$,然后添加噪声如下:$D_2 = D_1 - |\triangle D|$,当敏感数据中的元素添加噪声的值为负时,将元素改为零。

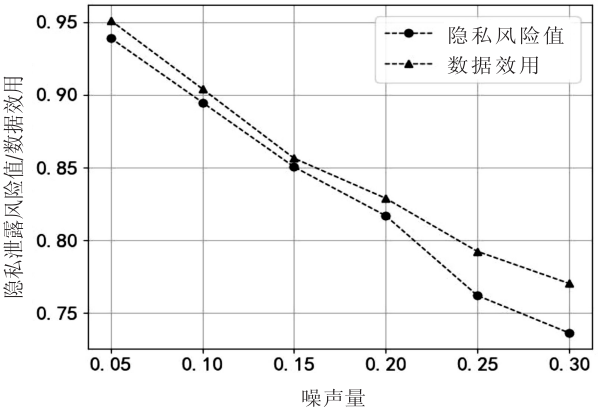

图 3 隐私风险值与隐私效用随噪声增加的变化

通过图 3 的结果分析表明,在字段操作、字段需求不变的情况下,随着对数据添加噪声量的增加,隐私风险值降低,数据效用也随之降低。

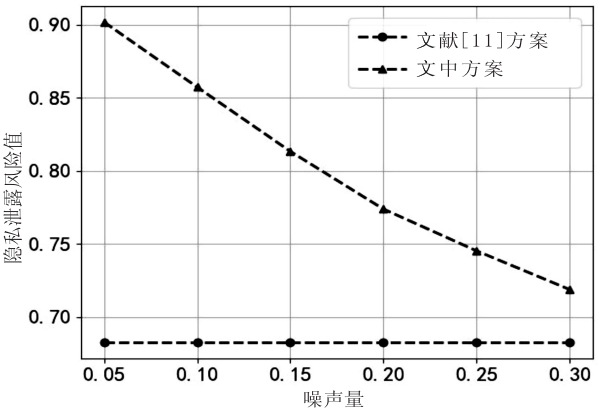

图 4 隐私评估方案对比

最后将文中方案与文献[11]做对照分析,实验结果如图 4 所示。通过对评估数据添加噪声,文中方案的隐私风险值随着噪声的增加而降低,但是文献[11]没有发生任何变化。通过对比分析可以看出,文中方案不仅可以对字段和操作进行风险评估,还加入了对数据本身的数据隐私泄露风险评估,使得评估结果更加全面,也更能看出隐私保护的效果。

综上,通过对请求字段数、数据效用、隐私风险值和其他方案的对比分析可以得出,该方法符合隐私泄露风险评估的预期,对于有效评估数据的隐私泄露风

险,提高数据的共享价值具有一定的参考意义。

5　结束语

通过对数据进行数值化映射度量数据隐私并建立隐私等级标记,将两者紧密融合,实现隐私泄露风险的多方因素综合评估,提出了一种数据发布的隐私泄露风险评估方法。

该方法首先获取数据使用者数据需求,通过分别对属性字段隐私等级以及操作隐私等级进行等级划分,建立隐私等级矩阵,然后对发布数据进行数值化处理创建敏感数据矩阵,最后将隐私等级矩阵与敏感数据矩阵相结合,计算得出隐私泄露风险值,从而建立数据发布的隐私泄露风险评估方法。该方法为数据发布场景中的隐私泄露风险评估提供了适当的指导作用。

未来的研究中,将会探索更多更好的隐私泄露风险度量方法和评价指标,通过扩展相关度量方法与评价指标,建立更加可靠安全的隐私泄露风险评价体系。

参考文献:

[1] 孟小峰,慈　祥. 大数据管理:概念、技术与挑战[J]. 计算机研究与发展,2013,50(1):146-169.

[2] 罗亦军,刘　强,王　宇. 社会网络的隐私保护研究综述[J]. 计算机应用研究,2010,27(10):3601-3604.

[3] 周水庚,李　丰,陶宇飞,等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报,2009,32(5):847-861.

[4] 熊金波,王敏燊,田有亮,等. 面向云数据的隐私度量研究进展[J]. 软件学报,2018,29(7):1963-1980.

[5] XIONG J B,LI F H,LIU X M,et al. A full lifecycle privacy protection scheme for sensitive data in cloud computing[J]. Peer-to-Peer Networking and Applications,2015,8(6):1025-1037.

[6] LIU Y H,ZHANG T Y,JIN X L,et al. Personal privacy protection in the era of big data[J]. Journal of Computer Research and Development,2015,52(1):229-247.

[7] MEHMOOD A,NATGUNANATHAN I,XIANG Y,et al. Protection of big data privacy[J]. IEEE Access on Theoretical Foundations for Big Data Applications,2016,4:1821-1834.

[8] 彭长根,丁红发,朱义杰,等. 隐私保护的信息熵模型及其度量方法[J]. 软件学报,2016,27(8):1891-1903.

[9] YEH K H,HOU J L,CHEN L C,et al. Privacy risk assessment for SQLite based android applications[J]. Journal of Internet Technology,2017,18(7):1533-1541.

[10] 晏　燕,王万军. 偏联系数隐私风险态势评估方法[J]. 计算机工程与应用,2018,54(10):143-148.

[11] 周旭晨,王智慧,王　宇,等. 一种基于矩阵计算的数据开放隐私泄露评估方法[J]. 计算机应用与软件,2020,37(1):298-303.

[12] 谢小杰,梁　英,王梓森,等. 社交网络用户隐私泄露量化评估方法[J]. 计算机工程与科学,2021,43(8):1376-1386.

[13] ISO B S. Information technology-security techniques-privacy framework:IEC 29100[R]. [s. l.]:British Standard and the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC),2011.

[14] DWORK C. Differential privacy:a survey of results[C]//International conference on theory and applications of models of computation. Berlin:Springer,2008:1-19.

[15] 谢明明,彭长根,吴睿雪,等. 结构化数据的隐私与数据效用度量模型[J]. 计算机应用研究,2020,37(5):1465-1469.

[16] 谢明明. 面向数据发布的隐私泄露理性风险计算研究及应用[D]. 贵州:贵州大学,2019.