几类经典隐私定义间的关系

郑智润[1],黄 橙[2],王 萍[3*],李成新[4],许 雯[5],李哲涛[5]
(1. 亚洲大学人工智能系,韩国水原 16499;2. 复旦大学计算机科学技术学院,上海 200433;3. 兰州理工大学计算机与人工智能学院,甘肃兰州 730050;4. 湘潭大学数学与计算科学学院,湖南湘潭 411105;5. 暨南大学信息科学技术学院,广东广州 510632)

摘 要: 针对现有基于不同隐私定义设计的扰动机制在理论上难以比较优劣的问题,本文从理论层面深入分析了中心化场景和本地化场景下可辨识性、差分隐私和互信息隐私这三类经典隐私定义之间的关系,构建了一个完备的隐私定义框架. 具体而言,给定由真实数据先验概率分布决定的常数(当先验概率分布为均匀分布时,常数$\sigma_{\min}=0$),可得以下结论:满足ε_i-可辨识性的隐私保护机制必然也同时满足$(\varepsilon_i-\sigma_{\min})$-差分隐私和$2(\varepsilon_i-\sigma_{\min})$-互信息隐私;满足$\varepsilon_d$-差分隐私的隐私保护机制必然也同时满足$(\varepsilon_d+\sigma_{\min})$-可辨识性和$2\varepsilon_d$-互信息隐私;但是,满足$\varepsilon_m$-互信息隐私的隐私保护机制却不一定满足$\varepsilon_i$-可辨识性和$\varepsilon_d$-差分隐私(在$\varepsilon_m$有限的情况下,$\varepsilon_i$和$\varepsilon_d$可能会趋于无穷大). 此外,所提隐私定义框架能一致地推导出隐私定义间的关系,使得对隐私预算上界的估计更加准确.

关键词: 隐私保护;数据发布;数据隐私;差分隐私;互信息隐私;可辨识性
基金项目: 国家自然科学基金国际合作重点项目(No.W2411053)
中图分类号: TP3-05;O211.1 **文献标识码:** A **文章编号:** 0372-2112(2025)10-3781-13
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250078

Relations Between Several Classical Privacy Notions

ZHENG Zhi-run[1], HUANG Cheng[2], WANG Ping[3*], LI Cheng-xin[4], XU Wen[5], LI Zhe-tao[5]
(1. *Department of Artificial Intelligence*, *Ajou University*, *Suwon* 16499, *South Korea*;
2. *School of Computer Science*, *Fudan University*, *Shanghai* 200433, *China*;
3. *School of Computer Science and Artificial Intelligence*, *Lanzhou University of Technology*, *Lanzhou*, *Gansu* 730050, *China*;
4. *School of Mathematics and Computational Science*, *Xiangtan University*, *Xiangtan*, *Hunan* 411105, *China*;
5. *College of Information Science and Technology*, *Jinan University*, *Guangzhou*, *Guangdong* 510632, *China*)

Abstract: We address the challenge of theoretically evaluating various perturbation-based privacy-preserving mechanisms designed under different privacy notions. By analyzing the relationships among three classical privacy notions, namely identifiability, differential privacy, and mutual-information privacy, in both centralized and local settings, we propose a complete privacy notion framework that establishes theoretical relations among them. Specifically, given a constant $\sigma_{\min}$ determined by the prior probability distribution of the real data (the constant $\sigma_{\min}=0$ when the prior distribution is uniform), the following theorems are formally proved in both central and local settings. First, the mechanism satisfying ε_i-identifiability must also satisfy ε_d-differential privacy with $\varepsilon_d=\varepsilon_i-\sigma_{\min}$. Second, the mechanism satisfying ε_d-differential privacy must also satisfy ε_i-identifiability with $\varepsilon_i=\varepsilon_d+\sigma_{\min}$. Third, the mechanism satisfying ε_i-identifiability must also satisfy ε_m-mutual-information privacy with $\varepsilon_m=2(\varepsilon_i-\sigma_{\min})$. Fourth, the mechanism satisfying ε_m-mutual-information privacy does not guarantee ε_d-differential privacy. Fifth, the mechanism satisfying ε_d-differential privacy must also satisfy ε_m-mutual-information privacy with $\varepsilon_m=2\varepsilon_d$. Sixth, the mechanism satisfying ε_m-mutual-information privacy does not guarantee ε_d-differential privacy. The proposed framework systematically derives the theoretical relationships among identifiability, differential privacy, and mutual-information privacy, enabling a more precise estimation of privacy budget bounds. Furthermore, the framework provides a theoretical foundation for achieving privacy preservation in data-driven applications such as crowdsensing systems, location-based services, and large language models.

Key words: privacy preserving; data publishing; data privacy; differential privacy; mutual-information privacy; identifiability
Foundation Item(s): National Natural Science Foundation of China (No.W2411053)

收稿日期:2025-01-22;录用日期:2025-09-25;责任编辑:宋甲英
*通讯作者:王萍

1 引言

数据驱动的服务已经深入渗透到社会生活的各个领域,例如对话生成模型ChatGPT[1]、无人驾驶服务平台萝卜快跑[2]和网约车平台滴滴出行[3]等,极大提升了人们日常生活的便利性. 然而,用户数据尽管能为各行业赋能,但因其包含大量敏感信息,若未加以规范使用,会对个人隐私构成严重威胁[4~11]. 例如,血压和心率数据包含用户健康状况等敏感信息[12],以及位置轨迹数据可能会泄露用户的宗教信仰、工作地址和社交关系等敏感信息[6~8]. 近些年,为了确保数据在使用、传输和存储过程中的安全性与合法性,各国政府相继出台相关法律法规,如欧盟于2018年发布《通用数据保护条例》[13]、瑞士于2023年发布《联邦数据保护法》[14]、我国则分别于2017年和2021年正式颁布和实施《网络安全法》[15]和《个人信息保护法》[16]. 尽管违反这些法律法规会面临严重惩处,但隐私泄露事件仍屡见不鲜. 例如,2022年蔚来汽车因其用户信息和车辆销售数据遭窃而被勒索225万美元[17],2023年Meta因将欧洲Facebook用户数据传输到美国服务器而被罚12亿欧元[18]. 此外,用户出于对隐私的考虑可能会放弃共享数据,阻碍数据驱动服务的发展. 因此,在当今数据要素时代,保护用户隐私显得尤为重要.

为了防止用户隐私泄露,基于扰动的隐私保护技术已广泛应用于数据驱动的服务中[6~11,19~23]. 在用户共享的数据中添加适量噪声,使得敌手无法观察到用户真实数据,进而无法推测出用户的敏感信息. 直观上,添加更多的噪声能提升隐私保护程度,但不恰当的噪声不仅无法有效保护隐私,反而会破坏数据的固有特征,降低数据质量,甚至引发新的隐私泄露问题[6~8]. 例如,不合适的噪声将在高速公路上正常行驶的用户突然扰动到偏离且垂直该道路的国道上,显著降低了基于位置服务的质量. 此外,攻击者还可以利用位置轨迹的时空特征,如可达性等,发动更强的隐私攻击[24,25]. 因此,隐私定义对基于扰动的隐私保护技术尤为重要,它量化了保护用户隐私所需的噪声量,并为隐私保护提供了严格可证明的理论基础.

尽管已有可辨识性[26,27]、差分隐私[28~30]和互信息隐私[27,31]等多种隐私定义被提出,但这些多样且繁杂的隐私定义却缺乏系统的框架,其关联性并不清晰. 隐私保护机制设计者往往因如何选择合适的隐私定义而困扰. 特别地,隐私保护可从概率论角度视为确保相邻数据集输出分布的不可区分性(如差分隐私和可辨识性等),也可从信息论角度视为控制信息熵减少(如互信息隐私等). 尤其是差分隐私,近些年被视为隐私保护的"黄金标准",在学术界和工业界均受到广泛关注. 例如,苹果公司自2016年起便应用该技术,在保护用户隐私的同时收集系统使用信息. 因此,本文考虑可辨识性、差分隐私和互信息隐私这三种经典隐私定义,试图通过揭示它们之间的关系,理解不同理论视角下隐私定义的一致性与差异性. 详细地,可辨识性关注的是通过扰动后数据推测真实数据的后验概率,差分隐私关注的是发布扰动后数据导致个人隐私的额外泄露量,而互信息则度量的是扰动后数据包含真实数据的平均信息量. 虽然这三种隐私定义是从不同角度给出的,但它们之间仍存在基本联系. 本文从严格的理论证明出发,分别在中心化场景和本地化场景下系统分析了可辨识性、差分隐私和互信息隐私之间的相互关系,并进一步将其归纳为完备的隐私定义框架.

本文的主要贡献是提出了完备的隐私定义框架,在理论层面刻画了中心化场景和本地化场景下可辨识性、差分隐私和互信息隐私间的关系. 详细地,给定依赖于真实数据先验概率分布的常数$\sigma_{\min}$,严格证明了:

(1)满足ε_i-可辨识性的隐私保护机制一定也同时满足$(\varepsilon_i-\sigma_{\min})$-差分隐私和$2(\varepsilon_i-\sigma_{\min})$-互信息隐私;

(2)满足ε_d-差分隐私的隐私保护机制一定也同时满足$(\varepsilon_d+\sigma_{\min})$-可辨识性和$2\varepsilon_d$-互信息隐私;

(3)满足ε_m-互信息隐私的隐私保护机制不一定满足ε_i-可辨识性和ε_d-差分隐私(ε_i和ε_d可能趋于无穷大).

2 相关工作

可辨识性[26,27]、差分隐私[28~30]和互信息隐私[27,31]作为经典的隐私定义为基于扰动的隐私保护技术提供了严格可证明的理论基础[6~8,32,33]. 例如,Andres等人[33]基于差分隐私提出了地理不可区分性,用以保护用户的位置数据隐私;Zheng等人[7]基于差分隐私提出了轨迹数据合成机制,用以保护用户的社交关系隐私;Zheng等人[6]基于互信息隐私提出了轨迹数据共享机制,用以同时保护用户的轨迹数据隐私和语义隐私. 然而,众多隐私定义给隐私保护机制的设计和评估带来了巨大挑战,即难以从理论上比较基于不同隐私定义设计扰动机制的优劣. 因此,分析隐私定义间的理论关系显得尤为重要. 近些年,已有不少研究尝试建立可辨识性、差分隐私和互信息隐私之间的理论关系,但尚未得出完备的隐私定义框架,即无法从该框架中一致地推导出隐私定义间的关系. 换言之,现有研究对可辨识性、差分隐私和互信息隐私中的隐私预算估计并不准确. 例如,Sun等人[34]证明了满足ε-可辨识性的隐私保护机制一定满足$(\varepsilon+\delta_X)$-差分隐私(详见文献[34]中的定理1-xiv),而满足ε-差分隐私的隐私保护机制也一定满足$(\varepsilon+\delta_X)$-可辨识性(详见文献[34]中的定理1-xii). 但是,应用一次该定理可得满足ε-可辨识性的隐私保护

机制一定满足$(\varepsilon+2\delta_X)$-可辨识性,而应用两次该定理却得该隐私保护机制一定满足$(\varepsilon+4\delta_X)$-可辨识性. 显然,隐私预算的误差在此过程中被逐步放大. 此外,Wang等人[27]分析了在可辨识性、差分隐私和互信息隐私下,最优隐私保护与数据质量之间的权衡关系. 然而,该研究关注的是隐私保护与数据质量之间的权衡问题,忽略了隐私定义间的基本关系. 甚至,随着数据质量度量指标的变化,所建立的隐私关系不一定成立. 综上所述,建立完备的隐私定义框架是至关重要且刻不容缓的.

3　问题描述和相关知识

3.1　问题描述

数据驱动的服务供应商(如百度地图和滴滴出行等)虽然能为用户提供高质量服务,但同时也伴随着严重的隐私泄露问题. 这是因为服务供应商在提供服务的过程中需要采集用户个人数据,而这些数据却包含大量敏感信息. 例如,位置轨迹数据包含工作地址和社交关系等敏感信息[6~8],以及血压和心率数据包含健康状况和身体状态等敏感信息[35~38]. 此外,本文考虑服务供应商是诚实且好奇的,即服务供应商诚实地为用户提供高质量服务,但可能会受到利益的驱使滥用甚至出售用户共享的数据. 因此,为了在保护隐私的前提下获得高质量服务,用户首先通过隐私保护机制对原始数据进行脱敏,然后再与服务供应商共享脱敏后的数据,如图1和图2所示. 详细地,在中心化场景中,用户首先将个人数据上传到可信服务器中,然后由可信服务器执行基于扰动的隐私保护技术对原始数据脱敏,最后将脱敏后的数据上传至服务供应商以获取相应服务,如图1所示. 相比之下,在本地化场景中,用户直接在本地端对原始数据进行脱敏,无需任何可信第三方参与,如图2所示. 上述基于扰动的隐私保护技术能够广泛应用在数据共享服务中,得益于可辨识性、差分隐私和互信息隐私等隐私定义所提供的理论基础. 然而,多样且繁杂的隐私定义也给隐私机制设计者们造成了更大挑战. 为此,本文致力于研究可辨识性、差分隐私和互信息隐私之间的关系,旨在构建完备的隐私定义框架.

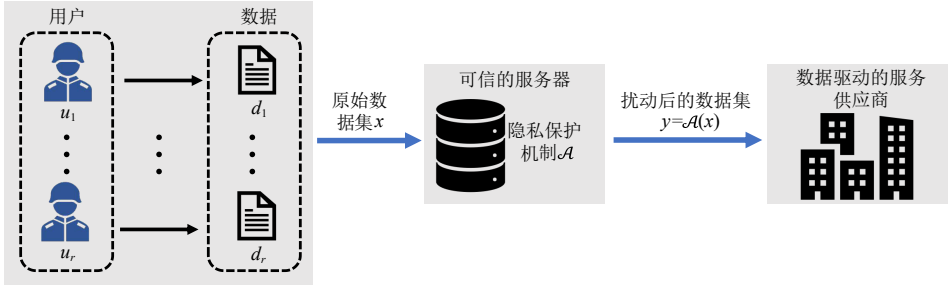

图1　中心化场景中保护隐私的数据共享服务

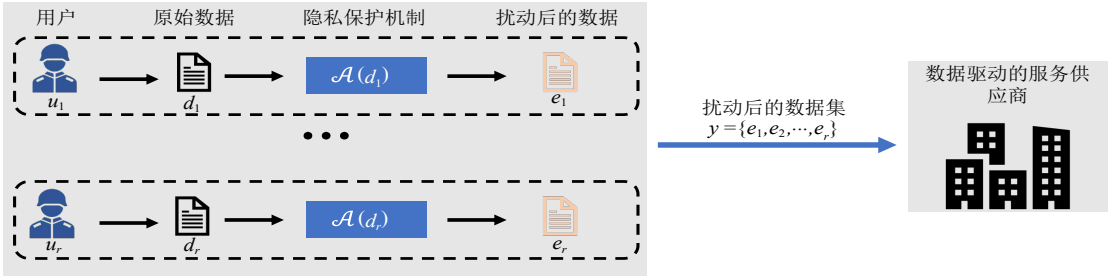

图2　本地化场景中保护隐私的数据共享服务

令用户数据集x由r个取自有限域$\mathcal{D}$的值构成,以及有限域$\mathcal{D}^r, r>1$由所有可能的数据集x构成,即

$$x=\{d_1,d_2,\cdots,d_r\},\ \forall d_i\in\mathcal{D},\forall x\in\mathcal{D}^r \tag{1}$$

在中心化保护隐私的数据共享服务中,如图1所示,可信服务器在收集到所有用户上传的数据后,通过隐私保护机制$\mathcal{A}$将其扰动为数据集y,即

$$y=\{e_1,e_2,\cdots,e_r\},\ \forall e_i\in\mathcal{D},\forall y\in\mathcal{D}^r \tag{2}$$

相应地,如图2所示,在本地化保护隐私的数据共享服务中,用户u_i的个人数据d_i通过隐私保护机制$\mathcal{A}$在本地被扰动为e_i,即$e_i=\mathcal{A}(d_i),\ \forall d_i,e_i\in\mathcal{D}$. 此外,$x$和$y$对应的离散随机变量记为$X$和$Y$,以及$e$和$d$对应的离散随机变量记为$E$和$D$. 本文分别考虑了中心化和本地化场景下的可辨识性、差分隐私和互信息隐私这几类经典隐私定义,并将它们在中心化场景下的隐私预算

分别记为ε_i^c,ε_d^c和ε_m^c,以及在本地化场景下的隐私预算分别记为ε_i^l,ε_d^l和ε_m^l. 特别地,在无需区分中心化和本地化场景时,隐私预算可去掉上标简化为ε_i,ε_d和ε_m. 详细的符号定义如表1所示.

表1 关键符号及其定义

符号	定义	符号	定义
ε_i^c	中心可辨识性的隐私预算	ε_i^l	本地可辨识性的隐私预算
ε_i	可辨识性的隐私预算	ε_d	差分隐私的隐私预算
ε_d^c	中心差分隐私的隐私预算	ε_d^l	本地差分隐私的隐私预算
ε_m^c	中心互信息隐私的隐私预算	ε_m^l	本地区信息隐私的隐私预算
ε_m	互信息隐私的隐私预算	$\sigma_{\min}$	相邻数据集间的最小先验概率差
$\sigma_{\min}^c$	中心化场景中相邻数据集间的最小先验概率差	$\sigma_{\min}^l$	本地化场景中个人数据间的最小先验概率差
$\mathcal{D}$和$\mathcal{D}^r$	有限域,且$r>1$	$\mathcal{A}$	隐私保护机制,且$e_k=A(d_k)$
X和Y	x和y对应的离散随机变量	E和D	e和d对应的离散随机变量
d_k	用户u_k上传的真实数据,且$\forall d_k\in\mathcal{D}$	e_k	真实数据d_k的扰动后版本,且$\forall e_k\in\mathcal{D}$
$x=\{d_1,d_2,\cdots,d_r\}$	用户的原始数据集,且$\forall d_k\in\mathcal{D}$, $\forall x\in\mathcal{D}^r$	$y=\{e_1,e_2,\cdots,e_r\}$	用户的扰动后数据集,且$\forall e_k\in\mathcal{D}$, $\forall y\in\mathcal{D}^r$

3.2 相关知识

在介绍可辨识性、差分隐私和互信息隐私等隐私定义之前,本节首先给出相邻数据集的定义,其详细定义如下.

定义1 相邻数据集. 令$x=\{d_1,d_2,\cdots,d_r\}$表示用户数据集且$\forall x\in\mathcal{D}^r$, $r>1$. 据此,取自有限域$\mathcal{D}^r$的任意两个数据集x和x'是相邻数据集,即$x\sim x'$, $\forall x,x'\in\mathcal{D}^r$,当且仅当它们仅有一条数据不一致.

3.2.1 可辨识性

可辨识性[26,27]能应用于中心化和本地化隐私保护场景,分别称为中心可辨识性和本地可辨识性. 详细地,中心可辨识性从后验概率的角度定义了任意相邻数据集的不可区分性,而本地可辨识性则从后验概率的角度定义了任意个人数据的不可区分性. 接下来,本节分别给出了中心化和本地化隐私保护场景中可辨识性的定义.

定义2 ε_i^c-中心可辨识性. 对$\forall\varepsilon_i^c\in\mathbb{R}^+$,隐私保护机制$\mathcal{A}$满足$\varepsilon_i^c$-中心可辨识性当且仅当任意相邻数据集$x\sim x'$, $\forall x,x'\in\mathcal{D}^r$, $r>1$和任意扰动后数据集y, $\forall y\in\mathcal{D}^r$, $r>1$满足:

$$p_{X|Y}(x|y)\leqslant \mathrm{e}^{\varepsilon_i^c}\cdot p_{X|Y}(x'|y) \tag{3}$$

定义3 ε_i^l-本地可辨识性. 对$\forall\varepsilon_i^l\in\mathbb{R}^+$,隐私保护机制$\mathcal{A}$满足$\varepsilon_i^l$-本地可辨识性当且仅当任意个人数据$d_j,d_k$, $\forall d_j,d_k\in\mathcal{D}$和任意扰动后数据$e$, $\forall e\in\mathcal{D}$满足:

$$p_{D|E}(d_j|e)\leqslant \mathrm{e}^{\varepsilon_i^l}\cdot p_{D|E}(d_k|e) \tag{4}$$

3.2.2 差分隐私

与可辨识性不同,满足差分隐私[28~30]的隐私保护机制$\mathcal{A}$仅由条件概率分布$p_{Y|X}$(或$p_{E|D}$)决定,而不依赖于先验概率分布p_X(或p_D). 此外,如果某种隐私泄露事件以确定概率发生,差分隐私保证在共享机制$\mathcal{A}$的输出时,该隐私泄露的概率只会增加一个乘法因子[29]. 简而言之,差分隐私不能完全消除隐私泄露的可能性,但却限制了共享数据所带来的额外风险. 接下来,本节分别给出了中心化和本地化隐私保护场景中差分隐私的定义.

定义4 ε_d^c-中心差分隐私. 对$\forall\varepsilon_d^c\in\mathbb{R}^+$,隐私保护机制$\mathcal{A}$满足$\varepsilon_d^c$-中心差分隐私当且仅当任意相邻数据集$x\sim x'$, $\forall x,x'\in\mathcal{D}^r$, $r>1$和任意扰动后数据集y, $\forall y\in\mathcal{D}^r$, $r>1$满足:

$$p_{Y|X}(y|x)\leqslant \mathrm{e}^{\varepsilon_d^c}\cdot p_{Y|X}(y|x') \tag{5}$$

定义5 ε_d^l-本地差分隐私. 对$\forall\varepsilon_d^l\in\mathbb{R}^+$,隐私保护机制$\mathcal{A}$满足$\varepsilon_d^l$-本地差分隐私当且仅当任意个人数据$d_j,d_k$, $\forall d_j,d_k\in\mathcal{D}$和任意扰动后数据$e$, $\forall e\in\mathcal{D}$满足:

$$p_{E|D}(e|d_j)\leqslant \mathrm{e}^{\varepsilon_d^l}\cdot p_{E|D}(e|d_k) \tag{6}$$

3.2.3 互信息隐私

互信息隐私[27,31]是基于信息论的隐私定义. 以中心化场景中的互信息$I(X;Y)$为例,$I(X;Y)$量化的是随机变量X包含在随机变量Y中的平均信息量,即共享数据集y引发的平均隐私泄露量. 当X和Y相互独立时,互信息$I(X;Y)$取最小值0;而当$Y=X$时,互信息$I(X;Y)$取最大值,即信息熵$H(X)=H(Y)$. 因此,互信息$I(X;Y)$越小,隐私保护程度越强. 接下来,本节分别给出了中心化和本地化隐私保护场景中互信息隐私的定义.

定义6 ε_m^c-中心互信息隐私. 对$\forall\varepsilon_m^c\in\mathbb{R}^+$,隐私保护机制$\mathcal{A}$满足$\varepsilon_m^c$-中心互信息隐私当且仅当随机变量$X$(原始数据集)和$Y$(扰动后数据集)之间的互信息满足$I(X;Y)\leqslant\varepsilon_m^c$,其中:

$$I(X;Y)=\sum_{x,y\in\mathcal{D}^r}p_{X,Y}(x,y)\cdot\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \tag{7}$$

定义7 ε_m^l-本地互信息隐私. 对$\forall\varepsilon_m^l\in\mathbb{R}^+$,隐私保护机制$\mathcal{A}$满足$\varepsilon_m^l$-本地互信息隐私当且仅当随机变量$D$

和E之间的互信息满足$I(D;E)\leqslant\varepsilon_m^l$，其中：

$$I(D;E)=\sum_{d,e\in\mathcal{D}}p_{D,E}(d,e)\cdot\log\frac{p_{D,E}(d,e)}{p_D(d)p_E(e)}\tag{8}$$

4　经典隐私定义间关系概述

在概述所提出的完备隐私定义框架之前，本节定义了中心化场景中相邻数据集间的最小先验概率差（即定义8）和本地化场景中个人数据间的最小先验概率差（即定义9），并分别在引理1和引理2中给出了它们的等价形式，其详细定义和引理如下.

定义8　中心化场景中相邻数据集间的最小先验概率差. 令任意取自有限域$\mathcal{D}^r$, $r>1$中的相邻数据集是$x\sim x'$, $\forall x,\ x'\in\mathcal{D}^r$，以及随机变量$X$（原始数据集）的先验概率分布是$p_X$. 据此，相邻数据集间的最小先验概率差可以被定义为

$$\sigma_{\min}^c=\min_{x,x'\in\mathcal{D}^r:x\sim x'}\log\frac{p_X(x)}{p_X(x')}\tag{9}$$

定义9　本地化场景中个人数据间的最小先验概率差. 令用户个人数据d对应的随机变量是D，以及关于D的先验概率分布是p_D. 据此，个人数据间的最小先验概率差可以被定义为

$$\sigma_{\min}^l=\min_{d_j,d_k\in\mathcal{D}}\log\frac{p_D(d_j)}{p_D(d_k)}\tag{10}$$

引理1　中心化场景中相邻数据集间的最小先验概率差$\sigma_{\min}^c$，即式（9），和下式等价：

$$\sigma_{\min}^c=\min_{x,x'\in\mathcal{D}^r:x\sim x'}\log\frac{p_X(x')}{p_X(x)}\tag{11}$$

特别地，当随机变量X（原始数据集）服从均匀分布时，有$\sigma_{\min}^c=0$.

证明　相邻数据集$x\sim x'$, $\forall x,\ x'\in\mathcal{D}^r$意味着数据集$x$和$x'$相邻的同时，数据集$x'$也和$x$相邻. 因此，中心化场景中相邻数据集间的最小先验概率差$\sigma_{\min}^c$具备以下等价形式：

$$\sigma_{\min}^c=\min_{x,x'\in\mathcal{D}^r:x\sim x'}\log\frac{p_X(x)}{p_X(x')}\tag{12}$$

$$\sigma_{\min}^c=\min_{x,x'\in\mathcal{D}^r:x\sim x'}\log\frac{p_X(x')}{p_X(x)}\tag{13}$$

额外地，如果随机变量X服从均匀分布，即$p_X(x)=p_X(x')$, $\forall x,x'\in\mathcal{D}^r$，则有$\sigma_{\min}^c=0$.

引理2　本地化场景中个人数据间的最小先验概率差$\sigma_{\min}^l$，即式（10），和下式等价：

$$\sigma_{\min}^l=\min_{d_j,d_k\in\mathcal{D}}\log\frac{p_D(d_k)}{p_D(d_j)}\tag{14}$$

特别地，当随机变量D（原始数据）服从均匀分布时，有$\sigma_{\min}^l=0$.

证明　因为个人数据d_j和d_k任意取自有限域$\mathcal{D}$，所以本地化场景中个人数据间的最小先验概率差$\sigma_{\min}^l$具备以下等价形式：

$$\sigma_{\min}^l=\min_{d_j,d_k\in\mathcal{D}}\log\frac{p_D(d_j)}{p_D(d_k)}\tag{15}$$

$$\sigma_{\min}^l=\min_{d_j,d_k\in\mathcal{D}}\log\frac{p_D(d_k)}{p_D(d_j)}\tag{16}$$

额外地，如果随机变量D服从均匀分布，即$p_D(d_j)=p_D(d_k)$, $\forall d_j,\ d_k\in\mathcal{D}$，则有$\sigma_{\min}^l=0$.

基于相邻数据集间的最小先验概率差$\sigma_{\min}^c$（详见定义8），本文得到了中心化场景中可辨识性、差分隐私和互信息隐私之间的关系，如图3所示. 详细地，如果隐私保护机制$\mathcal{A}=p_{X|Y}(x|y)$满足ε_i^c-中心可辨识性，那么该机制同时也满足$(\varepsilon_i^c-\sigma_{\min}^c)$-中心差分隐私和$2(\varepsilon_i^c-\sigma_{\min}^c)$-中心互信息隐私（详见定理1和定理5）. 类似地，如果机制$\mathcal{A}$满足ε_d^c-中心差分隐私，那么该机制同时也满足$(\varepsilon_i^c+\sigma_{\min}^c)$-中心可辨识性和$2\varepsilon_d^c$-中心互信息隐私（详见定理2和定理9）. 此外，本文还证明了满足ε_m^c-中心互信息隐私的机制$\mathcal{A}$并不一定满足ε_i^c-中心可辨识性（详见定理6）和ε_d^c-中心差分隐私（详见定理10）.

基于个人数据间的最小先验概率差$\sigma_{\min}^l$（详见定义9），本文得到了本地化场景中可辨识性、差分隐私和互信息隐私之间的关系，如图3所示. 详细地，如果隐私保护机制$\mathcal{A}=p_{D|E}(d|e)$满足ε_i^l-本地可辨识性，那么该机制同时也满足$\left(\varepsilon_i^l-\sigma_{\min}^l\right)$-本地差分隐私和$2\left(\varepsilon_i^l-\sigma_{\min}^l\right)$-本地互信息隐私（详见定理3和定理7）. 类似地，如果机制$\mathcal{A}$满足ε_d^l-本地差分隐私，那么该机制同时也满足$\left(\varepsilon_d^l+\sigma_{\min}^l\right)$-本地可辨识性和$2\varepsilon_d^l$-本地互信息隐私（详见定理4和定理11）. 此外，本文还证明了满足ε_m^l-本地互信息隐私的机制$\mathcal{A}$不一定满足ε_i^l-本地可辨识性（详见定理8）和ε_d^l-本地差分隐私（详见定理12）.

需要强调的是，图3所示的隐私定义框架是完备的. 换言之，从定理1、定理2、定理5、定理6、定理9和定理10能一致地推导出中心可辨识性、中心差分隐私和中心互信息隐私间的关系，以及能从定理3、定理4、定理7、定理8、定理11和定理12一致地推导出本地可辨识性、本地差分隐私和本地互信息隐私间的关系. 以中心化场景为例阐述隐私定义框架的完备性. 首先，如果隐私保护机制$\mathcal{A}=p_{X|Y}(x|y)$满足ε_d^c-中心差分隐私，那么由定理2和定理5可得机制$\mathcal{A}$满足$2(\varepsilon_d^c+\sigma_{\min}^c-\sigma_{\min}^c)$-中心互信息隐私，以及由定理9可得机制$\mathcal{A}$满足$2\varepsilon_d^c$-中心互信息隐私，显而可知$2(\varepsilon_d^c+\sigma_{\min}^c-\sigma_{\min}^c)=2\varepsilon_d^c$. 其次，如果机制$\mathcal{A}$满足$\varepsilon_i^c$-中心可辨识性，那么由定理1和定理9可得机制$\mathcal{A}$满足$2(\varepsilon_i^c-\sigma_{\min}^c)$-中心互信息隐私，也可由定理5推得机制$\mathcal{A}$满足$2(\varepsilon_i^c-\sigma_{\min}^c)$-中心互信息隐私. 最后，如果机制$\mathcal{A}$满足$\varepsilon_i^c$-中心可辨识性，那么由定理1

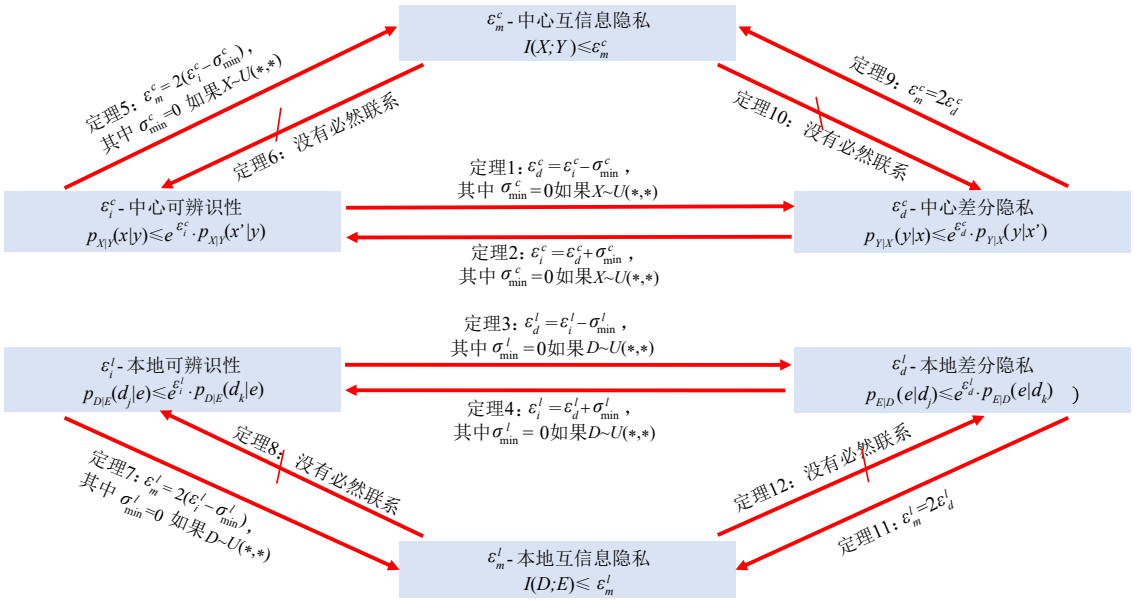

图 3 中心化和本地化场景下的完备隐私定义框架

和定理 2 可得机制 $\mathcal{A}$ 依旧满足 $(\varepsilon_i^c-\sigma_{\min}^c+\sigma_{\min}^c)$-中心可辨识性. 综上所述,在中心化场景中,基于本文提出的隐私定义框架推导出的隐私定义关系是一致的,称之为完备性.

本文提出的完备隐私定义框架通过引入由原始数据集计算得到的最小先验概率差 $\sigma_{\min}$ 这一常数,系统揭示了基于概率论或信息论构建的多种隐私定义(如可辨识性、差分隐私和互信息隐私)之间的潜在联系. 理论上,基于不同隐私定义设计的扰动机制可以统一转换至同一隐私定义下,并在相同数据质量约束下比较其隐私预算大小,从而判断哪种隐私定义对应的扰动机制最优,以此选取合适的隐私定义. 鉴于该过程相对烦琐,本文建议隐私保护机制设计者可根据下述经验规则初步选择隐私定义,再结合所提的隐私定义框架加以调整. 直观上,可辨识性和差分隐私的隐私预算对偏态分布高度敏感,而互信息隐私的隐私预算则相对稳定. 换言之,当原始数据的先验概率分布越偏态(可用偏态系数度量)时,为实现同级别的隐私保护,相较于互信息隐私,可辨识性和差分隐私需添加更多噪声. 需要注意的是,本文所提的完备隐私定义框架适用于对数据质量要求不是特别严格的数据采集、数据共享和数据发布场景,用于评估基于扰动的隐私保护机制的性能优劣.

5 可辨识性、差分隐私和互信息隐私之间的关系

在证明可辨识性、差分隐私和互信息隐私之间的关系前,本节首先简要阐述可辨识性和差分隐私的区别. 从可辨识性和差分隐私的相关定义可以看出,两者非常相似:可辨识性基于后验概率 $p_{X|Y}$(或 $p_{D|E}$)定义了有限域 $\mathcal{D}^r$ 中任意相邻数据集(或有限域 $\mathcal{D}$ 中任意个人数据)的不可区分性,而差分隐私则基于条件概率 $p_{Y|X}$(或 $p_{E|D}$)定义了有限域 $\mathcal{D}^r$ 中任意相邻数据集(或有限域 $\mathcal{D}$ 中任意个人数据)的不可区分性. 尽管可辨识性和差分隐私的定义如此相似,但它们仍存在以下区别:可辨识性中的最小隐私预算 ε_i^c(或 ε_i^l)受先验概率分布 p_X(或 p_D)的约束,而差分隐私中的最小隐私预算 ε_d^c(或 ε_d^l)与先验概率分布 p_X(或 p_D)无关且提供的是相对的隐私保护.

详细地,根据定义 2 和定义 3 可推得可辨识性中的最小隐私预算 ε_i^c(或 ε_i^l)受先验概率分布 p_X(或 p_D)的约束,即

$$\varepsilon_i^c\geqslant\log\frac{p_X(x)\cdot p_{Y|X}(y|x)}{p_X(x')\cdot p_{Y|X}(y|x')},\ \forall x,x'\in\mathcal{D}^r:x\sim x' \tag{17}$$

$$\varepsilon_i^l\geqslant\log\frac{p_D(d_j)\cdot p_{E|D}(e|d_j)}{p_D(d_k)\cdot p_{E|D}(e|d_k)},\ \forall d_j,d_k\in\mathcal{D} \tag{18}$$

此外,根据定义 4 和定义 5 可得差分隐私中的最小隐私预算 ε_d^c(或 ε_d^l)与先验概率分布 p_X(或 p_D)无关,即

$$\varepsilon_d^c\geqslant\log\frac{p_{Y|X}(y|x)}{p_{Y|X}(y|x')},\ \forall x,x'\in\mathcal{D}^r:x\sim x' \tag{19}$$

$$\varepsilon_d^l\geqslant\log\frac{p_{E|D}(e|d_j)}{p_{E|D}(e|d_k)},\ \forall d_j,d_k\in\mathcal{D} \tag{20}$$

接下来通过简单的例子直观展示二者区别.

假定敌手知悉隐私保护机制 $\mathcal{A}$ 及先验概率分布 p_X(或 p_D),并通过观察机制 $\mathcal{A}$ 的输出试图揭露用户的原始数据. 以本地化场景中有限域 $\mathcal{D}=\{1,2\}$ 和隐私保护

机制 $p_{E|D}(1|1)=p_{E|D}(2|2)=0.65$，$p_{E|D}(2|1)=p_{E|D}(1|2)=0.35$ 为例. 本地可辨识性：当先验概率分布 $p_D(1)=0.95$，$p_D(2)=0.05$ 时，由式(18)可得机制 $\mathcal{A}=p_{E|D}(e|d)$ 满足 ε_i^l-本地可辨识性且 $\varepsilon_i^l\approx 3.56$，$\varepsilon_i^l\geqslant\log(0.95/0.05)\approx 2.94$. 换言之，当先验概率分布 $p_D(1)=0.95$，$p_D(2)=0.05$ 时，不存在机制 $\mathcal{A}$ 能满足隐私预算 $\varepsilon_i^l\leqslant 2.94$ 的 ε_i^l-本地可辨识性. 类似地，当先验概率分布 $p_D(1)=0.55$，$p_D(2)=0.45$ 时，由式(18)可得机制 $\mathcal{A}$ 满足 ε_i^l-本地可辨识性且 $\varepsilon_i^l\approx 0.42$，$\varepsilon_i^l\geqslant\log(0.55/0.45)\approx 0.20$. 从上述例子可以直观看出，本地可辨识性中的最小隐私预算 ε_i^l 受先验概率分布 p_D 的约束. 本地差分隐私：由式(20)可得机制 $\mathcal{A}$ 满足 ε_d^l-本地差分隐私且 $\varepsilon_d^l=\log(0.65/0.35)\approx 0.62$ 并与先验概率分布 p_D 无关. 此外，当先验概率分布 $p_D(1)=0.95$，$p_D(2)=0.05$ 时，敌手在没有观察到机制 $\mathcal{A}$ 的输出时，能正确推测用户原始数据 $d=1$ 的概率是 0.95，而在观察到机制 $\mathcal{A}$ 的输出 $e=1$ 时，能正确推测原始数据 $d=1$ 的概率是 0.97，即

$$p_{D|E}(1|1)=\frac{p_{E|D}(1|1)p_D(1)}{p_{E|D}(1|1)p_D(1)+p_{E|D}(1|2)p_D(2)} \tag{21}$$

$$=0.97 \tag{22}$$

换言之，本地差分隐私并没有试图逆转原始数据 $d=1$ 被敌手推测出的概率，而是将共享扰动后数据 $e=1$ 对原始数据泄露增加的风险限定在一定范围内，即 $p_{D|E}(1|1)\approx \mathrm{e}^{0.02}\cdot p_D(1)$.

5.1 可辨识性和差分隐私间的关系

可辨识性和差分隐私分别基于后验概率 $p_{X|Y}$（或 $p_{D|E}$）和条件概率 $p_{Y|X}$（或 $p_{E|D}$）定义了有限域 $\mathcal{D}^r$ 中任意相邻数据集（或有限域 $\mathcal{D}$ 中任意个人数据）的不可区分性. 据此，本节研究满足可辨识性的隐私保护机制 $\mathcal{A}$ 是否同时满足差分隐私，以及满足差分隐私的机制 $\mathcal{A}$ 是否也同时满足可辨识性. 通过严格的理论分析发现：在中心化场景中，机制 $\mathcal{A}$ 满足 ε_i^c-中心可辨识性，那么它必定也满足 $(\varepsilon_i^c-\sigma_{\min}^c)$-中心差分隐私（详见定理1）；而当机制 $\mathcal{A}$ 满足 ε_d^c-中心差分隐私时，它也一定满足 $(\varepsilon_d^c+\sigma_{\min}^c)$-中心可辨识性（详见定理2）. 类似地，在本地化场景中，满足 ε_i^l-本地可辨识性的机制 $\mathcal{A}$ 也一定满足 $(\varepsilon_i^l-\sigma_{\min}^l)$-本地差分隐私（详见定理3），以及满足 ε_d^l-本地差分隐私的机制 $\mathcal{A}$ 也一定满足 $(\varepsilon_d^l+\sigma_{\min}^l)$-本地可辨识性（详见定理4）. 可辨识性与差分隐私间关系的定理及证明如下所示.

定理1　令任意取自有限域 $\mathcal{D}^r, r>1$ 中的相邻数据集是 $x\sim x'$, $\forall x,x'\in\mathcal{D}^r$. 据此，当隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_i^c-中心可辨识性且隐私预算 $\varepsilon_i^c\in\mathbb{R}^+$ 时，机制 $\mathcal{A}$ 也一定满足 $(\varepsilon_i^c-\sigma_{\min}^c)$-中心差分隐私.

证明　根据定义8，即中心化场景中相邻数据集间的最小先验概率差 $\sigma_{\min}^c$，可得：

$$\frac{1}{p_X(x)}\leqslant \mathrm{e}^{-\sigma_{\min}^c}\cdot\frac{1}{p_X(x')},\ \forall x,x'\in\mathcal{D}^r:\ x\sim x' \tag{23}$$

此外，由隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_i^c-中心可辨识性，详见定义2，可得：

$$p_{X|Y}(x|y)\leqslant \mathrm{e}^{\varepsilon_i^c}\cdot p_{X|Y}(x'|y),\ \forall x,x'\in\mathcal{D}^r: x\sim x' \tag{24}$$

根据式(23)和(24)，可得对任意取自有限域 $\mathcal{D}^r$ 中的相邻数据集，即 $\forall x,x'\in\mathcal{D}^r: x\sim x'$，有

$$p_{X|Y}(x|y)\frac{1}{p_X(x)}\leqslant \mathrm{e}^{\varepsilon_i^c}\cdot p_{X|Y}(x'|y)\frac{1}{p_X(x)} \tag{25}$$

$$\leqslant \mathrm{e}^{\varepsilon_i^c-\sigma_{\min}^c}\cdot p_{X|Y}(x'|y)\frac{1}{p_X(x')} \tag{26}$$

又因为：

$$\frac{p_{X|Y}(x|y)}{p_X(x)}=\frac{p_{Y|X}(y|x)}{p_Y(y)} \tag{27}$$

$$\frac{p_{X|Y}(x'|y)}{p_X(x')}=\frac{p_{Y|X}(y|x')}{p_Y(y)} \tag{28}$$

所以式(26)可以重新被形式化为对任意取自有限域 $\mathcal{D}^r$ 中的相邻数据集，即 $\forall x,x'\in\mathcal{D}^r: x\sim x'$，满足：

$$\frac{p_{Y|X}(y|x)}{p_Y(y)}\leqslant \mathrm{e}^{\varepsilon_i^c-\sigma_{\min}^c}\cdot\frac{p_{Y|X}(y|x')}{p_Y(y)} \tag{29}$$

综上，由定义4可得隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 $(\varepsilon_i^c-\sigma_{\min}^c)$-中心差分隐私.

定理2　令任意取自有限域 $\mathcal{D}^r, r>1$ 中的相邻数据集是 $x\sim x'$, $\forall x,x'\in\mathcal{D}^r$. 据此，当隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_d^c-中心差分隐私且隐私预算 $\varepsilon_d^c\in\mathbb{R}^+$ 时，机制 $\mathcal{A}$ 也一定满足 $(\varepsilon_d^c+\sigma_{\min}^c)$-中心可辨识性.

证明　由引理1，即中心化场景中相邻数据集间最小先验概率差 $\sigma_{\min}^c$ 的等价形式，可得：

$$\sigma_{\min}^c=\min_{x,x'\in\mathcal{D}^r:x\sim x'}\log\frac{p_X(x')}{p_X(x)} \tag{30}$$

$$=\max_{x,x'\in\mathcal{D}^r:x\sim x'}\log\frac{p_X(x)}{p_X(x')} \tag{31}$$

由式(31)显而易得：

$$p_X(x)\leqslant \mathrm{e}^{\sigma_{\min}^c}\cdot p_X(x'),\ \forall x,x'\in\mathcal{D}^r: x\sim x' \tag{32}$$

又因隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_d^c-中心差分隐私，详见定义4，所以有

$$p_{Y|X}(y|x)\leqslant \mathrm{e}^{\varepsilon_d^c}\cdot p_{Y|X}(y|x'),\ \forall x,x'\in\mathcal{D}^r: x\sim x' \tag{33}$$

根据式(32)和(33)可得对任意取自有限域 $\mathcal{D}^r$ 中的相邻数据集，即 $\forall x,x'\in\mathcal{D}^r: x\sim x'$，满足：

$$p_{Y|X}(y|x)p_X(x)\leqslant \mathrm{e}^{\varepsilon_d^c}\cdot p_{Y|X}(y|x')p_X(x) \tag{34}$$

$$\leqslant \mathrm{e}^{\varepsilon_d^c+\sigma_{\min}^c}\cdot p_{Y|X}(y|x')p_X(x') \tag{35}$$

此外，因为下式成立：

$$p_{Y|X}(y|x)p_X(x)=p_{X|Y}(x|y)p_Y(y) \tag{36}$$

$$p_{Y|X}(y|x')p_X(x')=p_{X|Y}(x'|y)p_Y(y) \tag{37}$$

所以式(35)可以被重新形式化为

$$p_{X|Y}(x|y)p_Y(y)\leqslant \mathrm{e}^{\varepsilon_d^c+\sigma_{\min}^c}\cdot p_{X|Y}(x'|y)p_Y(y) \tag{38}$$

综上所述,隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 $(\varepsilon_d^c+\sigma_{\min}^c)$-中心可辨识性,即对任意取自有限域 $\mathcal{D}^r$ 中的相邻数据集满足:

$$p_{X|Y}(x|y)\leqslant \mathrm{e}^{\varepsilon_d^c+\sigma_{\min}^c}\cdot p_{X|Y}(x'|y) \tag{39}$$

定理3 如若隐私保护机制 $\mathcal{A}=p_{E|D}(e|d)$ 满足 ε_i^l-本地可辨识性且隐私预算 $\varepsilon_i^l\in\mathbb{R}^+$,那么机制 $\mathcal{A}$ 一定满足 $(\varepsilon_d^l-\sigma_{\min}^l)$-本地差分隐私.

证明 与定理1的证明类似,此处不再赘述.

定理4 如若隐私保护机制 $\mathcal{A}=p_{E|D}(e|d)$ 满足 ε_d^l-本地差分隐私且隐私预算 $\varepsilon_d^l\in\mathbb{R}^+$,那么机制 $\mathcal{A}$ 一定满足 $(\varepsilon_d^l+\sigma_{\min}^l)$-本地可辨识性.

证明 与定理2的证明类似,此处不再赘述.

5.2 可辨识性和互信息隐私间的关系

可辨识性基于后验概率 $p_{X|Y}$(或 $p_{D|E}$)定义了有限域 $\mathcal{D}^r$(或 $\mathcal{D}$)中任意相邻数据集(或任意个人数据)之间的不可区分性,而互信息则衡量了由共享扰动后数据集 y(或扰动后数据 e)所导致用户隐私泄露的平均信息量. 显然,从"平均"推导到"任意"是不可能的. 因此,本节通过举特例的方式证明了:在中心化场景中,隐私保护机制 $\mathcal{A}$ 满足 ε_m^c-中心互信息隐私并不能保证其一定也满足 ε_i^c-中心可辨识性(详见定理6);以及在本地化场景中,满足 ε_m^l-本地化信息隐私的机制 $\mathcal{A}$ 并不一定满足 ε_i^l-本地可辨识性(详见定理8). 此外,本节通过严格的理论分析发现:在中心化场景中,机制 $\mathcal{A}$ 满足 ε_i^c-中心可辨识性,那么该机制也一定满足 $2(\varepsilon_i^l-\sigma_{\min}^c)$-中心互信息隐私(详见定理5);以及在本地化场景中,满足 ε_i^l-本地可辨识性的机制 $\mathcal{A}$ 一定也满足 $2(\varepsilon_i^l-\sigma_{\min}^l)$-本地互信息隐私(详见定理7). 有关可辨识性与互信息隐私间关系的定理及证明如下所示.

定理5 令任意取自有限域 $\mathcal{D}^r, r>1$ 中的相邻数据集是 $x\sim x'$, $\forall x,x'\in\mathcal{D}^r$. 据此,当隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_i^c-中心可辨识性且隐私预算 $\varepsilon_i^c\in\mathbb{R}^+$ 时,机制 $\mathcal{A}$ 也一定满足 $2(\varepsilon_i^c-\sigma_{\min}^c)$-中心互信息隐私.

证明 由隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_i^c-中心可辨识性,详见定义2,可得:

$$\frac{p_{X|Y}(x|y)}{p_{X|Y}(x'|y)}\leqslant \mathrm{e}^{\varepsilon_i^c},\ \forall x,x'\in\mathcal{D}^r:\ x\sim x' \tag{40}$$

对式(40)两边取对数可得:

$$\log\frac{p_{X|Y}(x|y)}{p_{X|Y}(x'|y)}\leqslant \varepsilon_i^c,\ \forall x,x'\in\mathcal{D}^r:\ x\sim x' \tag{41}$$

因此,

$$\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \tag{42}$$

$$\leqslant \varepsilon_i^c-\log\frac{p_X(x)}{p_X(x')}-\log\frac{\sum\limits_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')} \tag{43}$$

$$\leqslant \varepsilon_i^c-\sigma_{\min}^c-\log\frac{\sum\limits_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')} \tag{44}$$

$$\leqslant \varepsilon_i^c-\sigma_{\min}^c-\log\frac{\mathrm{e}^{-\varepsilon_i^c+\sigma_{\min}^c}\cdot\sum\limits_{x\in\mathcal{D}^r}p_{Y|X}(y|x')p_X(x)}{p_{Y|X}(y|x')} \tag{45}$$

$$=\varepsilon_i^c-\sigma_{\min}^c-\log\left(\mathrm{e}^{-\varepsilon_i^c+\sigma_{\min}^c}\cdot\sum_{x\in\mathcal{D}^r}p_X(x)\right) \tag{46}$$

$$=2(\varepsilon_i^c-\sigma_{\min}^c) \tag{47}$$

其中,式(43)成立是因为:

$$\log\frac{p_{X|Y}(x|y)}{p_{X|Y}(x'|y)} \tag{48}$$

$$=\log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\cdot\frac{p_X(x)}{p_X(x')}\cdot\frac{p_X(x')p_Y(y)}{p_{X,Y}(x',y)}\right) \tag{49}$$

$$=\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}+\log\frac{p_X(x)}{p_X(x')}+\log\frac{p_X(x')p_Y(y)}{p_{X,Y}(x',y)} \tag{50}$$

$$=\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}+\log\frac{p_X(x)}{p_X(x')}+\log\frac{\sum\limits_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')} \tag{51}$$

式(44)成立是因为定义8,即中心化场景中相邻数据集间的最小先验概率差:

$$-\log\frac{p_X(x)}{p_X(x')}\leqslant-\sigma_{\min}^c,\ \forall x,x'\in\mathcal{D}^r:\ x\sim x' \tag{52}$$

以及式(45)成立是因为满足 ε_i^c-中心可辨识性的机制 $\mathcal{A}$ 一定也满足 $(\varepsilon_d^c-\sigma_{\min}^c)$-中心差分隐私(详见定理1),即对任意取自有限域 $\mathcal{D}^r$ 中的相邻数据集,即 $\forall x,x'\in\mathcal{D}^r$: $x\sim x'$,均有:

$$\mathrm{e}^{-\varepsilon_i^c+\sigma_{\min}^c}\cdot p_{Y|X}(y|x')\leqslant p_{Y|X}(y|x) \tag{53}$$

综上所述,可得:

$$\sum_{x,y\in\mathcal{D}^r}p_{X,Y}(x,y)\cdot\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \tag{54}$$

$$\leqslant 2(\varepsilon_i^c-\sigma_{\min}^c)\sum_{x,y\in\mathcal{D}^r}p_{X,Y}(x,y) \tag{55}$$

$$=2(\varepsilon_i^c-\sigma_{\min}^c) \tag{56}$$

因此,隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 $2(\varepsilon_i^c-\sigma_{\min}^c)$-中心互信息隐私.

定理6 令任意取自有限域 $\mathcal{D}^r, r>1$ 中的相邻数据集是 $x\sim x'$, $\forall x,x'\in\mathcal{D}^r$. 据此,满足 ε_m^c-中心互信息隐私的机制 $\mathcal{A}=p_{Y|X}(y|x)$ 不一定也满足 ε_i^c-中心可辨识性.

证明　考虑有限域 $\mathcal{D}^2=\{(1,1),(1,2),(2,2)\}$ 且联合概率分布 $p_{X,Y}$ 是:

$$p_{X,Y}(x,y)=\begin{cases}\beta, & x=y=(1,1)\\ 0, & x=(1,1),\ \forall y\in\mathcal{D}^2\\ 0, & \forall x\in\mathcal{D}^2,\ y=(1,1)\\ \dfrac{1-\beta}{4}, & \text{其他}\end{cases}\tag{57}$$

据此,随机变量 X 和 Y 之间的互信息 $I(X;\ Y)$ 可以通过下式计算:

$$I(X;\ Y)=\sum_{x,y\in\mathcal{D}^2}p_{X,Y}(x,y)\cdot\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\tag{58}$$

$$=\beta\cdot\log\frac{1}{\beta}+(1-\beta)\cdot\log\frac{1}{1-\beta}\tag{59}$$

根据定义 6 可得隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$(可由式(57)推得)满足 ε_m^c-中心互信息隐私且 $\varepsilon_m^c=-\beta\cdot\log\beta+(\beta-1)\cdot\log(\beta-1)$. 然而,机制 $\mathcal{A}=p_{Y|X}(y|x)$ 并不满足 ε_i^c-中心可辨识性,因为隐私预算 $\varepsilon_i^c=+\infty$,即

$$\varepsilon_i^c=\max_{x,x',y\in\mathcal{D}^2:x\sim x'}\log\frac{p_{X|Y}(x|y)}{p_{X|Y}(x'|y)}\tag{60}$$

$$=\max_{x,x',y\in\mathcal{D}^2:x\sim x'}\log\frac{p_{X,Y}(x,y)}{p_{X,Y}(x',y)}\tag{61}$$

$$=+\infty\tag{62}$$

由本特例可得,满足 ε_m^c-中心互信息隐私的机制 $\mathcal{A}=p_{Y|X}(y|x)$ 不能保证其一定满足 ε_i^c-中心可辨识性.

定理 7　如若隐私保护机制 $\mathcal{A}=p_{E|D}(e|d)$ 满足 ε_i^l-本地可辨识性且隐私预算 $\varepsilon_d^l\in\mathbb{R}^+$,那么机制 $\mathcal{A}$ 一定满足 $2(\varepsilon_i^l-\sigma_{\min}^l)$-本地互信息隐私.

证明　与定理 5 的证明类似,此处不再赘述.

定理 8　满足 ε_m^l-本地互信息隐私的机制 $\mathcal{A}=p_{E|D}(e|d)$ 不能保证其一定满足 ε_i^l-本地可辨识性.

证明　考虑有限域 $\mathcal{D}=\{1,2\}$ 且联合概率分布 $p_{D,E}$ 是:

$$p_{D,E}(d,e)=\begin{cases}\beta, & d=e=1\\ 1-\beta, & d=e=2\\ 0, & \text{其他}\end{cases}\tag{63}$$

根据定义 7 可得隐私保护机制 $\mathcal{A}=p_{E|D}(e|d)$(可由式(63)推得)满足 ε_m^l-本地互信息隐私,且隐私预算 ε_m^l 为

$$I(D;\ E)=\sum_{d,e\in\mathcal{D}}p_{D,E}(d,e)\cdot\log\frac{p_{D,E}(d,e)}{p_D(d)p_E(e)}\tag{64}$$

$$=\beta\cdot\log\frac{1}{\beta}+(1-\beta)\cdot\log\frac{1}{1-\beta}\tag{65}$$

然而,根据定义 3 可得机制 $\mathcal{A}$ 并不满足 ε_i^l-本地可辨识性,因为隐私预算 $\varepsilon_i^l=+\infty$,即

$$\varepsilon_i^l=\max_{d^j,d^k,e\in\mathcal{D}}\log\frac{p_{D|E}(d_j|e)}{p_{D|E}(d_k|e)}\tag{66}$$

$$=+\infty\tag{67}$$

综上所述,隐私保护机制 $\mathcal{A}$ 满足 ε_m^l-本地互信息隐私且隐私预算 $\varepsilon_m^l=-\beta\cdot\log\beta+(\beta-1)\cdot\log(\beta-1)$,但该机制并不满足 ε_i^l-本地可辨识性.

5.3　差分隐私和互信息隐私间的关系

差分隐私定义的是有限域 $\mathcal{D}^r$(或 $\mathcal{D}$)中任意相邻数据集(或任意个人数据)之间的不可区分性,而互信息隐私衡量的却是扰动后数据集 y(或扰动后数据 e)包含原始数据集 x(或与原始数据 d)的平均信息量. 显然,从基于"平均"的隐私定义推导出基于"任意"的隐私定义是不可能的. 因此,本节通过举特例的方式证实了:在中心化场景中,即使隐私保护机制 $\mathcal{A}$ 满足 ε_m^c-中心互信息隐私也不一定能确保其也满足 ε_d^c-中心差分隐私(详见定理 10);以及,在本地化场景中,满足 ε_m^l-本地互信息隐私的机制 $\mathcal{A}$ 不一定也满足 ε_d^l-本地差分隐私(详见定理 12). 此外,本节通过严格的理论分析发现:在中心化场景中,隐私保护机制 $\mathcal{A}$ 满足 ε_d^c-中心差分隐私,那么它也一定满足 $2\varepsilon_d^c$-中心互信息隐私(详见定理 9),以及在本地化场景中,满足 ε_d^l-本地差分隐私的机制 $\mathcal{A}$ 一定满足 $2\varepsilon_d^l$-本地互信息隐私(详见定理 11). 有关差分隐私与互信息隐私间关系的定理及证明如下所示.

定理 9　令任意取自有限域 $\mathcal{D}^r, r>1$ 中的相邻数据集是 $x\sim x',\ \forall x,x'\in\mathcal{D}^r$. 据此,隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_d^c-中心差分隐私且隐私预算 $\varepsilon_d^c\in\mathbb{R}^+$,那么该机制 $\mathcal{A}$ 也一定满足 $2\varepsilon_d^c$-中心互信息隐私.

证明　由隐私保护机制 $\mathcal{A}=p_{Y|X}(y|x)$ 满足 ε_d^c-中心差分隐私,详见定义 4,可得:

$$\frac{p_{Y|X}(y|x)}{p_{Y|X}(y|x')}\leqslant\mathrm{e}^{\varepsilon_d^c},\ \forall x,x'\in\mathcal{D}^r:x\sim x'\tag{68}$$

对式(68)两边取对数可得:

$$\log\frac{p_{Y|X}(y|x)}{p_{Y|X}(y|x')}\leqslant\varepsilon_d^c,\ \forall x,x'\in\mathcal{D}^r:x\sim x'\tag{69}$$

式(69)可以被重新形式化为

$$\log\frac{p_{Y|X}(y|x)}{p_X(x)p_Y(y)}\leqslant\varepsilon_d^c-\log\frac{\sum_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')}\tag{70}$$

因为:

$$\log\frac{p_{Y|X}(y|x)}{p_{Y|X}(y|x')}\tag{71}$$

$$=\log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\cdot\frac{p_X(x')p_Y(y)}{p_{X,Y}(x',y)}\right)\tag{72}$$

$$=\log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\cdot\frac{p_X(x')p_Y(y)}{p_{Y|X}(y|x')p_X(x')}\right)\tag{73}$$

$$=\log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\cdot\frac{\sum_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')}\right)\tag{74}$$

$$=\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}+\log\frac{\sum_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')} \tag{75}$$

再根据ε_d^c-中心差分隐私的等价形式，即对$\forall x,x'\in\mathcal{D}^r: x\sim x'$都有：

$$\mathrm{e}^{-\varepsilon_d^c}\cdot p_{Y|X}(y|x')\leqslant p_{Y|X}(y|x) \tag{76}$$

可得：

$$\log\frac{\sum_{x\in\mathcal{D}^r}p_{Y|X}(y|x)p_X(x)}{p_{Y|X}(y|x')} \tag{77}$$

$$\geqslant\log\frac{\mathrm{e}^{-\varepsilon_d^c}\cdot\sum_{x\in\mathcal{D}^r}p_{Y|X}(y|x')p_X(x)}{p_{Y|X}(y|x')} \tag{78}$$

$$=\log\left(\mathrm{e}^{-\varepsilon_d^c}\cdot\sum_{x\in\mathcal{D}^r}p_X(x)\right) \tag{79}$$

$$=-\varepsilon_d^c \tag{80}$$

根据式(80)，式(70)可以重新被形式化为

$$\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\leqslant 2\varepsilon_d \tag{81}$$

综上可得：

$$\sum_{x,y\in\mathcal{D}^r}p_{X,Y}(x,y)\cdot\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \tag{82}$$

$$\leqslant 2\varepsilon_d^c\cdot\sum_{x,y\in\mathcal{D}}p_{X,Y}(x,y) \tag{83}$$

$$=2\varepsilon_d^c \tag{84}$$

由定义6可知隐私保护机制$\mathcal{A}=p_{Y|X}(y|x)$满足$2\varepsilon_d^c$-中心互信息隐私.

定理10 令任意取自有限域$\mathcal{D}^r, r>1$中的相邻数据集是$x\sim x'$, $\forall x,x'\in\mathcal{D}^r$. 据此，满足$\varepsilon_m^c$-中心互信息隐私的机制$\mathcal{A}=p_{Y|X}(y|x)$不一定满足$\varepsilon_d^c$-中心差分隐私.

证明 考虑有限域$\mathcal{D}^2=\{(1,1),(1,2),(2,2)\}$且联合概率分布$p_{X,Y}$是：

$$p_{X,Y}(x,y)=\begin{cases}\beta, & x=y=(1,1)\\ 0, & x=(1,1),\ \forall y\in\mathcal{D}^2\\ 0, & \forall x\in\mathcal{D}^2,\ y=(1,1)\\ \dfrac{1-\beta}{4}, & \text{其他}\end{cases} \tag{85}$$

根据定义6可得，隐私保护机制$\mathcal{A}=p_{Y|X}(y|x)$(可由式(83)推得)满足ε_m^c-中心互信息隐私，且隐私预算可以通过下式计算：

$$I(X;\ Y)=\sum_{x,y\in\mathcal{D}^2}p_{X,Y}(x,y)\cdot\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \tag{86}$$

$$=\beta\cdot\log\frac{1}{\beta}+(1-\beta)\cdot\log\frac{1}{1-\beta} \tag{87}$$

此外，由定义4可得机制$\mathcal{A}$不满足ε_d^c-中心差分隐私，因为隐私预算$\varepsilon_d^c=+\infty$，即

$$\varepsilon_d^c=\max_{x,x',y\in\mathcal{D}^2:x\sim x'}\log\frac{p_{Y|X}(y|x)}{p_{Y|X}(y|x')} \tag{88}$$

$$=\max_{x,x',y\in\mathcal{D}^2:x\sim x'}\log\left(\frac{p_X(x')}{p_X(x)}\cdot\frac{p_{X,Y}(x,y)}{p_{X,Y}(x',y)}\right) \tag{89}$$

$$=+\infty \tag{90}$$

由本特例可得，满足ε_m^c-中心互信息隐私的机制$\mathcal{A}=p_{Y|X}(y|x)$不满足ε_d^c-中心差分隐私.

定理11 隐私保护机制$\mathcal{A}=p_{E|D}(e|d)$满足ε_d^l-本地差分隐私且隐私预算$\varepsilon_d^l\in\mathbb{R}^+$，那么机制$\mathcal{A}$一定也满足$2\varepsilon_d^l$-本地互信息隐私.

证明 与定理9的证明类似，此处不再赘述.

定理12 满足ε_m^l-本地互信息隐私的机制$\mathcal{A}=p_{E|D}(e|d)$不一定满足ε_d^l-本地差分隐私.

证明 考虑有限域$\mathcal{D}=\{1,2\}$且联合概率分布$p_{D,E}$是：

$$p_{D,E}(d,e)=\begin{cases}\beta, & d=e=1\\ 1-\beta, & d=e=2\\ 0, & \text{其他}\end{cases} \tag{91}$$

根据定义7可得隐私保护机制$\mathcal{A}=p_{E|D}(e|d)$(可由式(91)推得)满足ε_m^l-本地互信息隐私且隐私预算ε_m^l为

$$I(D;\ E)=\sum_{d,e\in\mathcal{D}}p_{D,E}(d,e)\cdot\log\frac{p_{D,E}(d,e)}{p_D(d)p_E(e)} \tag{92}$$

$$=\beta\cdot\log\frac{1}{\beta}+(1-\beta)\cdot\log\frac{1}{1-\beta} \tag{93}$$

此外，由定义5可得隐私保护机制$\mathcal{A}$不满足ε_d^l-本地可辨识性，因为隐私预算$\varepsilon_d^l=+\infty$，即

$$\varepsilon_d^l=\max_{d_j,d_k,e\in\mathcal{D}}\log\frac{p_{D|E}(e|d_j)}{p_{D|E}(e|d_k)} \tag{94}$$

$$=\max_{d_j,d_k,e\in\mathcal{D}}\log\left(\frac{p_D(d_k)}{p_D(d_j)}\cdot\frac{p_{D,E}(d_j,e)}{p_{D,E}(d_k,e)}\right) \tag{95}$$

$$=+\infty \tag{96}$$

由本特例可得，隐私保护机制$\mathcal{A}$满足ε_m^l-本地互信息隐私且隐私预算$\varepsilon_m^l=-\beta\cdot\log\beta+(1-\beta)\cdot\log(1-\beta)$，但该机制并不满足$\varepsilon_d^l$-本地差分隐私.

6 结束语

本文提出了完备的隐私定义框架，从理论上分析了可辨识性、差分隐私和互信息隐私之间的基本联系. 该框架为设计和评价基于扰动的隐私保护机制提供了理论基础，进而推动了此类隐私保护机制的发展. 今后的研究可以将所提的框架应用于具体的隐私保护场景中，如位置隐私等，以评估现有基于扰动的数据隐私保护机制的优劣性.

参考文献

[1] OpenAI. Introducing ChatGPT[EB/OL]. (2022-11-30)[2024-10-23]. https://openai.com/index/chatgpt/.

[2] 萝卜快跑. 萝卜快跑: 自动驾驶出行服务平台[EB/OL]. (2024-06-10)[2024-10-23]. https://www.robotgo.com.
APOLLO GO. Apollo Go: Autonomous driving travel service platform[EB/OL]. (2024-06-10)[2024-10-23]. https://www.robotgo.com. (in Chinese)

[3] 滴滴. 滴滴一下 美好出行[EB/OL]. (2024-07-30)[2024-10-23]. https://www.didiglobal.com.
DIDI. DiDi-better travel[EB/OL]. (2024-07-30)[2024-10-23]. https://www.didiglobal.com. (in Chinese)

[4] 赵景欣, 岳星辉, 冯崇朋, 等. 基于通用数据保护条例的数据隐私安全综述[J]. 计算机研究与发展, 2022, 59(10): 2130-2163.
ZHAO J X, YUE X H, FENG C P, et al. Survey of data privacy security based on general data protection regulation[J]. Journal of Computer Research and Development, 2022, 59(10): 2130-2163. (in Chinese)

[5] 刘雅辉, 张铁赢, 靳小龙, 等. 大数据时代的个人隐私保护[J]. 计算机研究与发展, 2015, 52(1): 229-247.
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. (in Chinese)

[6] ZHENG Z R, LI Z T, JIANG H B, et al. Semantic-aware privacy-preserving online location trajectory data sharing[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 2256-2271.

[7] ZHENG Z R, LI Z T, LI J, et al. Utility-aware and privacy-preserving trajectory synthesis model that resists social relationship privacy attacks[J]. ACM Transactions on Intelligent Systems and Technology, 2022, 13(3): 1-28.

[8] ZHENG Z R, LI Z T, LONG S Q, et al. Pricing utility vs. location privacy: A differentially private data sharing framework for ride-on-demand services[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22(4): 3497-3513.

[9] ZHENG Z R, LI Z T, HUANG C, et al. Defending data poisoning attacks in DP-based crowdsensing: A game-theoretic approach[J]. IEEE Transactions on Mobile Computing, 2025, 24(3): 1859-1876.

[10] ZHENG Z R, LI Z T, HUANG C, et al. Data poisoning attacks and defenses to LDP-based privacy-preserving crowdsensing[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(5): 4861-4878.

[11] LI Z T, ZHENG Z R, GUO S M, et al. Disguised as privacy: Data poisoning attacks against differentially private crowdsensing systems[J]. IEEE Transactions on Mobile Computing, 2023, 22(9): 5155-5169.

[12] 邱宇, 王持, 齐开悦, 等. 智慧健康研究综述: 从云端到边缘的系统[J]. 计算机研究与发展, 2020, 57(1): 53-73.
QIU Y, WANG C, QI K Y, et al. A survey of smart health: System design from the cloud to the edge[J]. Journal of Computer Research and Development, 2020, 57(1): 53-73. (in Chinese)

[13] Union European. Regulation (EU) 2016/679 of the European Parliament and of the Council[S]. Luxembourg: Official Journal of the European Union, 2016: 1-88.

[14] The Federal Council. Federal act on data protection[S/OL]. (2023-09-01)[2024-10-23]. https://www.fedlex.admin.ch/eli/cc/2022/491/en.

[15] 全国人民代表大会. 中华人民共和国网络安全法[S/OL]. (2016-11-07)[2024-10-23]. http://www.npc.gov.cn/zgrdw/npc/xinwen/2016-11/07/content_2001605.htm.
National People's Congress. Cybersecurity Law of the People's Republic of China[S/OL]. (2016-11-07)[2024-10-23]. http://www.npc.gov.cn/zgrdw/npc/xinwen/2016-11/07/content_2001605.htm. (in Chinese)

[16] 全国人民代表大会. 中华人民共和国个人信息保护法[S/OL]. (2021-08-20)[2024-10-23]. http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html.
National People's Congress. Personal Information Protection Law of the People's Republic of China[S/OL]. (2021-08-20)[2024-10-23]. http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html. (in Chinese)

[17] 澎湃新闻. 大批用户数据泄露, 蔚来致歉[EB/OL]. (2022-12-21)[2024-10-23]. https://www.thepaper.cn/newsDetail_forward_21269411.
The Paper. A large number of user data leaks, NIO apologizes[EB/OL]. (2022-12-21)[2024-10-23]. https://www.thepaper.cn/newsDetail_forward_21269411. (in Chinese)

[18] 澎湃新闻. Meat偷传欧洲Facebook会员数据,被罚12亿欧元[EB/OL]. (2023-05-23)[2024-10-23]. https://www.thepaper.cn/news Detail_forward_23188354 commTag=true.
The Paper. Meat steals European Facebook membership data, fined 1.2 billion euros[EB/OL]. (2023-05-23)[2024-10-23]. https://www.thepaper.cn/news Detail_forward_23188354 commTag=true. (in Chinese)

[19] JIANG H B, LI J, ZHAO P, et al. Location privacy-preserving mechanisms in location-based services: A comprehensive survey[J]. ACM Computing Surveys, 2021, 54(1):

3423165.

[20] 康海燕, 王骁识. 基于数据特征相关性和自适应差分隐私的深度学习方法研究[J]. 电子学报, 2024, 52(6): 1963-1976.
KANG H Y, WANG X S. Research on the deep learning method based on data feature relevance and adaptive differential privacy[J]. Acta Electronica Sinica, 2024, 52(6): 1963-1976. (in Chinese)

[21] 曾卓, 汪成亮, 马飞. 基于差分隐私的活动模式保护与时空轨迹发布方法[J]. 电子学报, 2023, 51(3): 552-563.
ZENG Z, WANG C L, MA F. Differentially private activity pattern and spatial-temporal trajectory publication[J]. Acta Electronica Sinica, 2023, 51(3): 552-563. (in Chinese)

[22] 蒋伟进, 王海娟, 周为, 等. 基于自适应连续时间的群智感知轨迹隐私保护方案[J]. 电子学报, 2023, 51(10): 2894-2901.
JIANG W J, WANG H J, ZHOU W, et al. Track privacy protection scheme based on adaptive continuous time in crowdsensing[J]. Acta Electronica Sinica, 2023, 51(10): 2894-2901. (in Chinese)

[23] 康海燕, 冀源蕊. 基于本地化差分隐私的时序位置发布方案研究[J]. 电子学报, 2022, 50(9): 2222-2232.
KANG H Y, JI Y R. Research on time-serial location data publication based on local differential privacy[J]. Acta Electronica Sinica, 2022, 50(9): 2222-2232. (in Chinese)

[24] ZHAO P, JIANG H B, LI J, et al. Synthesizing privacy preserving traces: Enhancing plausibility with social networks[J]. IEEE/ACM Transactions on Networking, 2019, 27(6): 2391-2404.

[25] LIU H, LI X H, LI H, et al. Spatiotemporal correlation-aware dummy-based privacy protection scheme for location-based services[C]//IEEE INFOCOM 2017 - IEEE Conference on Computer Communications. Piscataway: IEEE, 2017: 1-9.

[26] LEE J, CLIFTON C. Differential identifiability[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 1041-1049.

[27] WANG W N, YING L, ZHANG J S. On the relation between identifiability, differential privacy, and mutual-information privacy[J]. IEEE Transactions on Information Theory, 2016, 62(9): 5018-5029.

[28] DWORK C, MCSHERRY F, NISSIM K, et al. Calibrating noise to sensitivity in private data analysis[C]//Theory of Cryptography. Berlin: Springer, 2006: 265-284.

[29] DWORK C. Differential privacy[M]//Automata, Languages and Programming. Berlin, Heidelberg: Springer, 2006: 1-12.

[30] DUCHI J C, JORDAN M I, WAINWRIGHT M J. Local privacy and statistical minimax rates[C]//Proceedings of the 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. New York: ACM, 2013: 429-438.

[31] SANKAR L, RAJAGOPALAN S R, POOR H V. Utility-privacy tradeoffs in databases: An information-theoretic approach[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 838-852.

[32] CUFF P, YU L Q. Differential privacy as a mutual information constraint[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016: 43-54.

[33] ANDRÉS M E, BORDENABE N E, CHATZIKOKOLAKIS K, et al. Geo-indistinguishability: Differential privacy for location-based systems[C]//Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. New York: ACM, 2013: 901-914.

[34] SUN M, TAY W P. On the relationship between inference and data privacy in decentralized IoT networks[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 852-866.

[35] BIAN M Y, HE G H, FENG G R, et al. Verifiable privacy-preserving heart rate estimation based on LSTM[J]. IEEE Internet of Things Journal, 2024, 11(1): 1719-1731.

[36] KNORR K, ASPINALL D, WOLTERS M. On the privacy, security and safety of blood pressure and diabetes apps[C]//ICT Systems Security and Privacy Protection. Cham: Springer, 2015: 571-584.

[37] 孙祯锋. 运动员可穿戴设备中个人数据隐私的法律保护[J]. 沈阳体育学院学报, 2022, 41(4): 104-110.
SUN Z F. Legal protection of data privacy in athletes' wearable devices[J]. Journal of Shenyang Sport University, 2022, 41(4): 104-110. (in Chinese)

[38] 张明武, 黄嘉骏, 韩亮. 医疗大数据隐私保护多关键词范围搜索方案[J]. 软件学报, 2021, 32(10): 3266-3282.
ZHANG M W, HUANG J J, HAN L. Range-based multi-keyword searchable scheme with privacy protection in e-healthcare cloud systems[J]. Journal of Software, 2021, 32(10): 3266-3282. (in Chinese)

作者简介

郑智润　男,1995年2月出生于江西省上饶市.现为韩国亚洲大学人工智能系博士后,同时兼任该校软件工程系讲师.主要研究方向为群智系统安全、隐私计算和人工智能安全等.
E-mail: zhengzhirun2019@gmail.com

黄　橙　男,1991年8月出生于安徽省淮南市.现为复旦大学青年研究员.主要研究方向为人工智能安全.
E-mail: chuang@fudan.edu.cn

王　萍　女,1994年出生于甘肃省兰州市.现为兰州理工大学讲师.主要研究方向为智能感知、安全计算等.
E-mail: wangping51226@gmail.com

李成新　男,1996年出生于湖北省黄冈市.现为湘潭大学博士研究生.主要研究方向为移动群智感知、数据隐私保护等.
E-mail: snowwhite@uestc.edu.cn

许　雯　女,1997年出生于湖南省邵阳市.现为暨南大学博士研究生.主要研究方向为图数据分析、信息隐私和安全、物联网等.
E-mail: wenxu018@gmail.com

李哲涛　男,1980年1月出生于湖南省邵阳市.现为暨南大学教授.主要研究方向为机器学习、深度学习、智能系统安全、云边端资源管理等.
E-mail: liztchina@hotmail.com