

基于隐私风险评估的脱敏算法自适应方法

祖立军^{1,2}, 曹雅琳³, 门小骅², 吕智慧¹, 叶家炜¹, 李泓一¹, 张亮³

(1. 复旦大学金融科技研究院, 上海 200433; 2. 中国银联股份有限公司, 上海 201210; 3. 华为技术有限公司, 江苏 南京 210012)

摘要: 金融行业业务处理中含大量的敏感数据。金融业务的快速发展导致数据种类激增, 目前绑定待脱敏金融数据以及脱敏算法的方式效率较低, 依靠安全专家经验对脱敏数据进行人工检查与风险评估耗时长。脱敏算法的选择不当导致脱敏处理后的金融数据存在潜在隐私泄露风险。国内外研究大多侧重脱敏方法的实现和隐私保护技术, 极少从自动化角度对脱敏算法进行研究。为提高脱敏效率和隐私保护透明度, 通过梳理现有隐私保护技术特点、业务场景对数据质量需求、金融机构对安全风险要求、数据属性等因素, 搭建自适应选择脱敏策略推荐框架, 建立通用隐私风险和ang数据质量双目标评估函数, 基于多决策因素体系和脱敏效果评估, 实现脱敏算法和参数的自适应选择。相较于传统数据脱敏方式, 所提方法能有效解决人工干预带来的脱敏数据可用性差和个人数据隐私保护性不足等问题, 在多类金融机构数据测试集实验下, 所提方法推荐准确率达到95%以上, 脱敏后的隐私风险非常接近预期隐私风险水平, 差距小于10%, 推荐效率相较于专家人工处理时间提升100倍。

关键词: 自动脱敏; 隐私风险评估; 人工智能; 金融敏感数据

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023037

Adaptive selection method of desensitization algorithm based on privacy risk assessment

ZU Lijun^{1,2}, CAO Yalin³, MEN Xiaohua², LYU Zhihui¹, YE Jiawei¹, LI Hongyi¹, ZHANG Liang³

1. School of Financial Technology, Fudan University, Shanghai 200433, China

2. China UnionPay Co., Ltd, Shanghai 201210, China

3. Huawei Technologies Co., Ltd, Nanjing 210012, China

Abstract: The financial industry deals with a vast amount of sensitive data in its business operations. However, the conventional approach of binding financial data for desensitization and using desensitization algorithms is becoming inefficient due to the fast-paced growth of financial businesses and the proliferation of data types. Additionally, manual verification and assessment of desensitized data by security experts are time-consuming and may carry

收稿日期: 2023-01-29; 修回日期: 2023-05-30

通信作者: 叶家炜, jwye@fudan.edu.cn

基金项目: 国家重点研发计划 (2021YFC330060)

Foundation Item: The National Key R&D Program of China (2021YFC330060)

引用格式: 祖立军, 曹雅琳, 门小骅, 等. 基于隐私风险评估的脱敏算法自适应方法[J]. 网络与信息安全学报, 2023, 9(3): 49-59.

Citation Format: ZU L J, CAO Y L, MEN X H, et al. Adaptive selection method of desensitization algorithm based on privacy risk assessment[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 49-59.

potential privacy risks due to the improper selection of desensitization algorithms. While prior research has emphasized desensitization methods and privacy-preserving technologies, limited work has been conducted on desensitization algorithms from the perspective of automation. To address this issue, an adaptive recommendation framework was proposed for selecting desensitization strategies that consider various factors, such as existing privacy protection technologies, data quality requirements of business scenarios, security risk requirements of financial institutions, and data attributes. Specifically, a dual-objective evaluation function was established for privacy risk and data quality to optimize the selection of desensitization algorithm parameters for different algorithms. Furthermore, the desensitization algorithm and parameters were adaptively selected by considering the data attributes through a multi-decision factor system and desensitization effect evaluation system. Compared to traditional approaches, the proposed framework effectively tackle issues of reduced data usability and inadequate personal data privacy protection that derive from manual intervention. Testing on a dataset with multiple financial institution types, the experiments show that the proposed method achieves a recommendation accuracy exceeding 95%, while the desensitized privacy risk level differed by less than 10% from the expected level. Additionally, the recommendation efficiency is 100 times faster than expert manual processing.

Keywords: automatic data desensitization, privacy risk assessment, artificial intelligence, financial sensitive data

0 引言

信息时代,数据已成为重要的基础资源,数据安全愈发被重视。近年来,国际上数据安全事件频发,国际安全研究中心 Ponemon Institute 和 IBM Security 联合发布的《2021 年数据泄露成本报告》中披露数据泄露单个事件平均造成 424 万美元损失。

层出不穷的安全事件及其带来的重大危害和潜在隐患,驱使国家、行业、企业等各层面聚焦数据安全,并从法规、标准、制度等方面开展相应措施。2018 年,欧盟颁布《通用数据保护条例》(GDPR),其中阐明企业对欧盟公民个人数据的收集、使用和处理规定^[1]。2019 年 8 月,全国人民代表大会通过《个人信息保护法》,明确要求对于个人信息的保存和提供要经过去标识化处理,以切实降低在数据应用中个人信息的潜在泄露风险^[2]。

现代银行是资本、数据、信息密集型部门,拥有大量的敏感数据,在存储、使用、分享等过程中均需积极考量金融敏感数据的隐私处理。随着政策的推动和引导、银行业务模式升级、实体经济对金融服务需求增加以及数字技术的发展,以金融服务为内容、数字技术为手段、应用场景为接入点的开放银行正不断推动自身的数字化转型,促使其更有效地与金融科技公司践行开放合

作,带动及赋能产业的数字化发展。金融数据在跨实体的传递,必然需要重点关注个人信息的保护,实现数据信息共享下的“最小使用原则”,在此过程中离不开高效的数据脱敏处理。

数据脱敏是当前绝大多数企业在数据安全治理与建设过程中的必选技术与措施。从个人数据层面而言,数据脱敏是对去标识化、匿名化的法律效能的实现过程。当前数据脱敏的研究大多侧重于脱敏方法的实现及其效率^[3-5],如以 k -匿名^[6-7]为代表的基于数据匿名的隐私保护技术、以差分隐私为代表的基于数据失真的隐私保护技术^[8-9]、以同态加密^[10-12]为代表的基于密码学的隐私保护技术。Zhang 等^[13]采用分布式计算模型设计并实现了大数据匿名系统,Mohammadian 等^[14]使用多线程技术对匿名技术进行并行化处理,大幅度提高大数据的匿名效率,但缺乏统一的脱敏效果评判标准。另外,大部分业界脱敏算法研究选择采用人工指定和自定义配置,在大数据场景下数据维度和种类不断增长,人工干预的方式效率低。因此,通过对敏感数据自动识别进行智能脱敏是当前热点研究方向^[15],Madan 等^[16]、Choromanski 等^[17]针对不同云大数据提出匿名化自适应算法以提升隐私保护并获取高效用值,但这些研究仅限于单一类型的脱敏算法。同时,对于脱敏后的数据进行标准化的效果评估是业界技术发展的主要趋势。总体而言,目前国内外尚未出现关于自

适应智能脱敏系统与评估体系的系统研究。

为提高脱敏效率和隐私保护透明度,本文构建二阶段自适应脱敏算法以及参数选择,具体工作如下。

1) 构建自适应选择脱敏算法:结合现有隐私保护技术特点、业务场景对数据质量需求、企业对安全风险要求、数据属性等因素自适应选择脱敏算法。

2) 构建通用隐私风险和数据质量双目标评估函数,在给定脱敏算法情况下自适应脱敏算法参数选择。

对比传统数据脱敏方式,所提方法针对人工干预带来的脱敏数据可用性差和个人数据隐私保护性不足问题,提供一种智能化设计数据脱敏的新思路。经过多类银行客户测试集的验证及多名数据安全专家的研判打标,该方法推荐准确率可达95%以上。

1 相关工作

数据脱敏是指在保留数据原始特征的情况下,按需制定脱敏策略,对敏感数据进行修改、变换的技术方法,旨在解决敏感数据在非安全环境下使用问题^[18]。确保数据安全性的同时提供数据可用性是数据脱敏的实现难点,其关键就是脱敏算法选择。

1.1 脱敏算法分类

为解决基于敏感数据删除/替换的脱敏方法存在隐私攻击和数据不可用等问题,国内外学者积极对隐私保护技术展开研究。现有脱敏算法分类为基于泛化的技术、基于密码的技术、基于数据合成技术以及基于扰动技术,如表1所示。不同算法分类都有其特性,如何选择恰当的隐私指标成为一项挑战工作^[19]。

目前,有很多成熟的商用解决方案,如Oracle的Data Masking组件^[20]、IBM的InfoSphere Optim Data Privacy产品^[21]、Informatica的Informatica Data Masking产品^[22]等,这些方案实现数据脱敏往往需要金融机构数据处理员人工选择相应脱敏算法,随后导入数据进行数据脱敏处理。

1.2 挑战

在大数据场景下,数据维度和种类不断增长,

手动绑定待脱敏金融数据以及脱敏算法的方式效率不高^[1]。另外,缺少脱敏残余风险评估,基于专家经验的脱敏策略选择主观性强、脱敏效果参差不齐,容易带来数据可用性差或个人隐私保护不足的问题。自动推荐脱敏算法面临以下两个问题。

表1 脱敏算法分类
Table 1 Desensitization algorithms classification

脱敏算法分类	算法模型	参数
基于泛化的技术	取整	无
	顶部和底部编码	无
	掩码	掩码位数
	截断	截断位数
	K 匿名以及扩展模型 ^[23]	k 、 l 、 t
基于密码的技术	保序加密	密钥
	保留格式加密	密钥
	同态加密	密钥
基于数据合成技术	—	无
基于扰动技术	随机响应	随机概率 p
	交换	随机概率 p
	差分隐私 ^[8]	epsilon

1) 如何构建数据脱敏选择因素。现有方法大多是从模糊的定性角度考虑脱敏算法的选择。乔宏明等^[24]提出从可用性、关联性、真实性、时效性等方面对脱敏算法进行选择,但可用性等因素难以量化进行应用,无法实现相关因素的客观度量。

2) 如何评估不同脱敏算法组合下的数据隐私风险。通过度量指标或度量方法表示存在的信息泄露风险称为隐私度量,其表示隐私保护方法的保护强度。当前已出现上百种隐私度量指标,但是每种指标仅能解决特定场景问题,如何选择合适的隐私指标也成为一项挑战工作^[19]。隐私度量起源于 K 匿名技术,其中 K 直观反映隐私保护强度^[5],但该技术难以对抗同质化攻击、背景知识攻击等。已有学者提出 l -多样性^[6]、 t -接近^[7],但均具有已知攻击假设前提,无法抵抗未知背景攻击,故 k 、 l 、 t 等指标仅能在一定程度上反映隐私风险。此外,国外学者致力于研究与脱敏模型无关的通用型评估技术。

例如, 3 种常见攻击场景和一系列统计评估指标被提出, 但是难以捕捉更多的攻击场景, 且攻击场景难以量化。

1.3 问题分析

本文解决的问题为“给定待脱敏数据, 采用自动推荐脱敏算法, 旨在通过自动脱敏, 平衡数据安全和数据使用”。更进一步, 本文提出一套脱敏算法选择和效果评估量化方案, 在自动脱敏后, 给出脱敏后风险评估, 增加隐私保护透明度。

脱敏算法的选择取决于数据属性和使用场景。例如, 使用场景是数据发布展示, 数据类型是准标

识符, 可以使用 K 匿名。故从上述 2 个角度和脱敏算法性质出发建立对应定性关系。由于隐私保护技术具有局限性, 其隐私度量模型难以适用所有脱敏算法。为弥补这一缺陷, 本文考虑以隐私信息在数据中的唯一性确定脱敏后的隐私风险。最后, 根据预期风险和脱敏后隐私风险评估结果确定脱敏算法参数, 进一步帮助建立脱敏算法参数优化器。

2 智能脱敏算法

本节介绍智能脱敏算法总体设计和具体技术细节, 算法总体设计如图 1 所示。

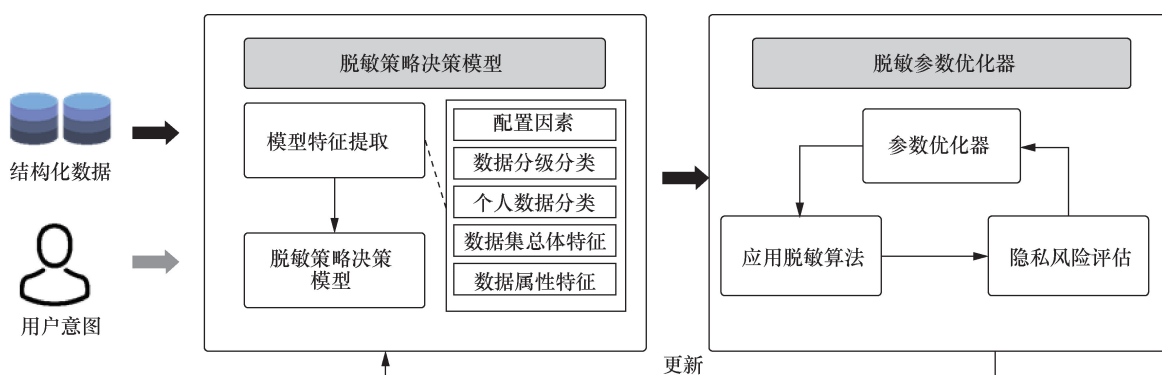


图 1 算法总体设计
Figure 1 Overall Algorithm Design

1) 用户意图: 包括数据使用场景、数据分享类型以及预期风险阈值。其分别确认数据有用性的最低要求、数据共享范围、数据风险的最低要求。由于不同数据处理者对数据风险的容忍度、数据使用方式、数据共享范围不同, 上述 3 种参数由用户配置。参考国内标准《信息安全技术 个人信息去标识化效果分级评估指南》^[25], 数据风险最低要求默认是 0.05。

2) 构建脱敏策略决策模型: 从多个维度构建脱敏策略决策因素体系 X_{rec} , 包括配置因素、数据分级分类、个人数据分类、数据集总体特征、数据属性特征。标签 Y_{rec} 代表脱敏算法及其参数, 使用有监督模型训练建立脱敏策略决策模型, 通过脱敏策略决策模型确定初始化脱敏方式和参数以保证数据可用性。

3) 建立脱敏参数优化器: 基于隐私风险评估模型, 构建脱敏参数优化器, 当隐私风险小于预设阈值时, 停止搜索并确定最终脱敏参数。

2.1 问题分析脱敏策略决策模型

(1) 决策模型因素体系

脱敏策略决策模型训练过程如图 2 所示。脱敏算法的选择难点在于平衡数据使用需求与隐私保护, 它们之间的关系目前只能做定性分析^[24]。结合具体应用场景, 本文重点考虑以下几个因素。

1) 数据使用场景。确认数据有用性的最低要求。一般可分为作为展示使用、作为键值使用、作为分析使用、作为计算使用。

2) 数据分级分类。敏感数据安全级别确定数据保护程度。敏感安全级别越高, 脱敏程度越深。例如,《JR/T 0223—2021 金融数据安全 数据生命周期安全规范》明确规定对处于不同生命周期的数据脱敏要求, 在数据分享阶段, 要求对 3 级以上数据进行脱敏。另外, 不同类型的数据处理方式也不一样。

3) 个人信息分类。不同类型个人信息使用脱敏策略不同, 故针对数据主体的直接标识

符、准标识符、敏感数据、非敏感数据分类。例如，敏感数据需作为统计使用，标识符无须使用，则尽可能在满足风险预期情况下提升脱敏程度。

4) 数据类型属性。不同数据类型选择脱敏算法也不尽相同。例如，作为统计使用，连续性变量使用基于 Laplace 的差分隐私，而针对分类变量使用基于指数机制的差分隐私。

5) 数据整体特征。记录准标识符的个数、一条记录是否对应一个实体等。例如，在 K 匿名技术中，若准标识符个数超过 10 个，则 K 匿名的效果不佳。

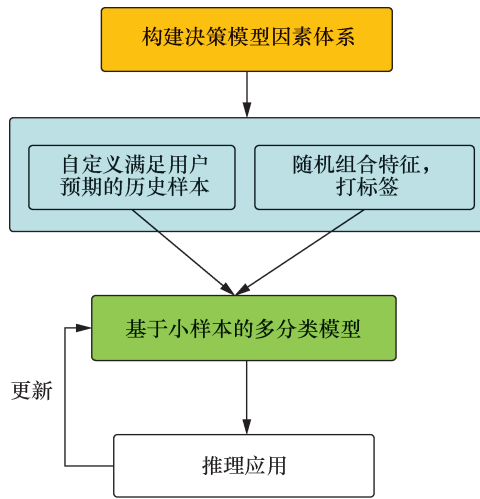


图2 脱敏策略决策模型训练过程
Figure 2 Desensitization strategy decision model training process

(2) 基于小样本多分类模型

基于以上脱敏策略决策因素体系 X_{rec} 和标签 Y_{rec} 建立一一对应关系。标签 Y_{rec} 是脱敏算法及其参数，见表 1。训练样本包括满足用户预期的历史样本和决策因素 X_{rec} 的随机组合，通过专家经验打标签获取标签 Y_{rec} 。

假设决策因素 $X_{\text{rec}} = \{X_1, X_2, \dots, X_k\}$ 与对应的多分类标签 $Y_{\text{rec}} = \{y_1, y_2, \dots, y_k\}$ 属于类别 $C = \{C_1, C_2, \dots, C_t\}$ ，即共有 k 个样本，属于 t 类。在算法冷启动阶段，无任何历史样本，通过随机组合特征 X_{rec} 生成少量样本，专家经验获取十几种标签 Y_{rec} 。由于传统的神经网络模型在小样本多分类问题易过拟合，故使用 SVM 解决该问题。首先，构建一个目标函数，使得类内间距变小、类间间距变大。

$$\text{Fitness} = \text{Max} \frac{1}{K} \sum_{i=1}^k (\text{Acc}_i - \text{Dist}_i) \quad (1)$$

其中， Acc_i 代表第 i 类样本的分类准确率， Dist_i 代表第 i 类样本类内距离和类间距离之商，针对小样本学习问题，使用 K 个样本进行 K 折交叉验证。Dist 计算公式如下。

$$\text{Dist} = \sum_{c=1}^C \frac{\sum_{m=1}^M (z_{c,m} - u_{c,m})^2}{\sum_{j \neq c, j=1}^C \sum_{m=1}^M (z_{j,m} - u_{j,m})^2} \quad (2)$$

其中， $z_{c,m}$ 代表第 c 类中测试样本的第 m 维特征向量中心点， $u_{c,m}$ 代表第 c 类中训练样本第 m 维特征向量的中心点。

2.2 隐私风险函数

含有攻击者的单隐私信源隐私保护通信模型如图 3 所示，隐私数据 X_d 通过隐私保护机制实现隐私保护并共享变形后数据 Y_d 。相反，攻击者通过某种手段获取 Y_d 可称为观察数据，结合其自身背景知识 K_d 进行隐私攻击使得隐私数据 X_d 泄露，则隐私数据 X_d 被泄露的风险可表示为

$$R(X_d) = \Pr(K_d, Y_d) \quad (3)$$

$R(X_d)$ 越大时，隐私信息被泄露的风险越大。

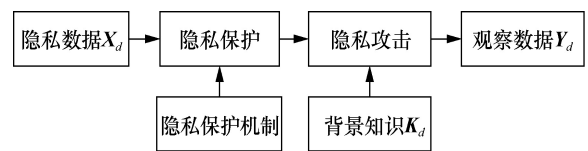


图3 含有攻击者的单隐私信源隐私保护通信模型
Figure 3 A single privacy source privacy-preserving communication model with an attacker

由于隐私泄露风险与攻击者获取背景知识难易程度 p_k 、数据脆弱性 p_y 、数据价值 v 及观察数据泄露风险 p_{content} 有关，进一步的隐私风险可以表示为

$$R(X_d) = \Pr(p_k, p_y, p_{\text{content}}) F(X_d, v) \quad (4)$$

其中， $\Pr(*)$ 表征数据泄露的可能性。 $F(*)$ 表征数据泄露损失，由泄露量 X_d 和敏感等级 v 共同决定，泄露量越大，数据越敏感，损失程度越高。

其中, 观察数据泄露风险与数据发布方式相关, 表示形式如下。

$$P_{\text{content}} = \begin{cases} 1, & \text{完全公开共享数据发布} \\ p_d, & \text{受控公开或领地公开} \end{cases} \quad (5)$$

其中, p_d 是数据泄露概率, 由数据存储的安全基础设置决定。

2.2.1 攻击者背景知识量化

对攻击者具备的背景知识进行量化, 本文的方法是列举所有可能背景知识组合, 通过计算每种可能性对隐私信息泄露的影响以衡量攻击者获取背景知识的强弱。

背景知识是攻击者的重要数据来源, 一般来自公共信息即用户的准标识符集合。准标识符是类似生日、邮编、年龄等个人信息, 难以单独实现对个人信息主体的唯一识别, 需要结合其他属性以识别个人信息唯一主体。背景知识获取越多且内容与隐私信息的关联性越大, 推测出隐私信息的可能性越大。

使用数据集描述攻击者的先验知识, 如用户身份集合 $U = \{u_1, u_2, \dots, u_n\}$ 中的每个子数据集分别表示攻击者的一项知识, 每个用户 $u_i = \{q_1, q_2, \dots, q_m\}$ 由 m 个属性组成, 则攻击者含有背景知识的组合数量为 $2^m - 1$ 个。假设 $m=2$, 攻击者拥有用户 u_i 的背景知识组合分别为 $\{A_1 = q_1\}$ 、 $\{A_2 = q_2\}$ 、 $\{A_3 = q_1, q_2\}$ 。此时, 攻击者对获取每种背景知识 A_i 的概率 p_k 与组合个数 z 、敏感等级大小 v 强相关。敏感等级 v 越大, 组合中的属性个数 z 越多, 获取的概率越小, 影响因素可以表征为

$$x_{vz} = \sum_{k=0}^z v_k \quad (6)$$

p_k 是 x_{vz} 的递减函数, 本文采用 logit 函数将概率 p_k 映射为 x_{vz} 的 sigmoid 反向函数。

$$p_k = 2 \left(1 - \frac{1}{1 + e^{-\sum_{k=0}^z v_k}} \right) \quad (7)$$

2.2.2 通用隐私风险度量机制

建立一种通用的隐私度量机制以量化数据泄露可能性, 依照攻击者拥有的背景知识及获取概

率, 结合隐私信息在数据中确定性计算用户隐私泄露风险。

信息熵表示某种特定信息出现的概率, 一个系统信息熵越低, 越是有秩序, 数据唯一性越低, 是一种量化信息平均不确定性的方法^[26]。但在现实环境中, 数据具有重要价值导致存在隐私攻击, 攻击者一般借助背景知识进行隐私攻击。

攻击者可以通过借助额外背景知识 K_d 并结合观察数据 Y_d 对隐私数据进行攻击, 引入条件熵

$$H(X_d | Y_d K_d) = - \sum_{i=1}^n \sum_{j=1}^m \sum_{z=1}^l p(k) p(x_i, y_j, k_z) \log p(x_i | y_j, k_z) \quad (8)$$

$H(X_d | Y_d K_d)$ 反映隐私数据 X_d 在攻击者结合观察数据 Y_d 和背景知识 K_d 攻击时存在的不确定性, 可以抽象为在某种隐私攻击下隐私信息的不确定性。

在攻击者没有任何知识背景的情况下, 最大信息熵为

$$H(X_d) = - \sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} \quad (9)$$

因此, 数据脆弱性 p_y 可以用归一化条件熵表示为

$$p_y = \frac{H(X_d | Y_d K_d)}{H(X_d)} \quad (10)$$

2.2.3 基于隐私保护模型的隐私风险度量机制

隐私保护技术的保护强度可衡量隐私风险, 即隐私风险与使用的隐私模型相关。1.2 节将隐私模型分为 4 大类: 基于泛化的技术、基于密码的技术、基于扰动技术及基于数据合成技术, 下面详细介绍不同隐私风险度量机制。

在基于泛化的隐私度量方法中, k -匿名值是隐私度量中的经典指标之一, 表示数据集中的准标识符属性匿名程度。其中, 每个等价类中的记录个数为 k 个, 且等价类的每个记录被识别的概率相等, k 值越小, 攻击者推测出隐私信息的概率越大, 隐私保护能力越小^[23]。若等价类内的敏感属性同质化严重, 攻击者易推测出敏感属性值与用户的对应关系, 因此 k 值难以覆盖所有攻击

场景，度量能力有限。文献[6-7]引入 EDM (earth mover distance) 方法计算数据中敏感属性值的全局分布和任意等价类中同一敏感属性值分布的差异度。但是这种方法未考虑等价类与敏感数据分布的稳定性。本文提出基于 K 、EMD 和 JS 散度的度量方法。

$$R(A) = \frac{1}{|J|} \sum_{j \in J} \left(\frac{1}{f_j} \left(\frac{1}{2} \sum_{i \in m} p_i \log \frac{2p_i}{(p_i + q_i)} + \frac{1}{2} \sum_{i \in m} q_i \log \frac{2q_i}{(p_i + q_i)} \right) \sum_{k \in m} d_{km} f_{km} \right) \quad (11)$$

其中, $R(A)$ 取值在 $[0, 1/k]$, $|J|$ 是等价组大小, f_j 是第 j 等价组大小, $P = \{p_1, p_2, \dots, p_m\}$, $Q = \{q_1, q_2, \dots, q_m\}$ 分别代表全局敏感属性分布和等价组内敏感属性分布。 d_{km} 是 P 中第 k 个元素和 Q 中第 m 个元素的接地距离, f_{km} 是 P 中第 k 个元素到 Q 中第 m 个元素的质量流动。

在基于扰动的隐私度量方法中, 最为代表的是差分隐私保护技术, 其具有严格的数学证明, 故能够提供一种可量化的隐私信息泄露风险评估方法^[8-9]。但是当数据元组相互关联时, 差分隐私保护技术难以反映关联隐私的保护强度。本文在差分隐私中引入互信息, 将差分隐私 ϵ 作为互信息的约束值, 如果互信息计算的值小于 ϵ , 则数据中隐私信息的强度满足用户对隐私保护的预期。

$$R(A) = \mu \log \frac{m_v e^\epsilon}{m_v - 1 + e^\epsilon} / H(X_d) \quad (12)$$

其中, $R(A)$ 取值在 $[0, 1]$, μ 是数据集 X_d 的数量, m_v 是敏感属性的取值个数, $H(X_d)$ 是数据集 X_d 的信息熵。

在基于密码的隐私度量方法中, 密码是衡量隐私保护强度的重要指标, 本文用密钥强度代表基于密码技术的隐私风险 $R(A)$ 。基于数据合成的隐私度量方法中, 数据集通常采用人为合成方法构建, 但保留数据之间的相关性, 所以一般风险是 0。本文用归一化基于隐私模型的隐私风险来表征数据脆弱性 p_y , 进一步可以表示为

$$p_y = \begin{cases} \frac{R(A) - \text{Min}(R(A))}{\text{Baseline} - \text{Min}(R(A))} \alpha, & R(A) < \text{Baseline} \\ \alpha + \frac{R(A) - \text{Baseline}}{\text{Max}(R(A)) - \text{Baseline}} (1 - \alpha), & R(A) \geq \text{Baseline} \end{cases} \quad (13)$$

其中, Baseline 是每个模型的风险阈值最低要求。 α 取值在 $[0, 1]$ 。结合式(10)和式(13)可以综合确定数据脆弱性 p_y 。

2.3 数据质量函数

学术界主要通过隐私保护技术处理后的数据缺损以反馈数据质量^[27], 数据缺损越低, 数据可利用率越高。数据可用性度量方法有: 分辨率度量 DM^[28-29]、数据信息损失 ILOSS^[30-31], 由于这些度量指标不具备单调性和均匀性, 因此无法对数据的有用属性加以区分。为解决这些问题, 在现有信息损失熵的基础上优化, 引入非均匀熵。另外, 现有的数据质量函数难以应对实际场景中不同脱敏算法的组合脱敏问题, 因此, 本文针对不同类型模型引入不同的数据质量评估函数。

$$DQ(X_d, Y_d) = \begin{cases} \sum_{i=1}^n \sum_{j=1}^r -\log \Pr(R_i(j) | R'_i), \text{基于泛化/噪声} & (14) \\ DQ(X_d, X_d), \text{其他} \end{cases}$$

其中, 原始数据 $X_d = \{R_1, R_2, \dots, R_n\}$, 脱敏后数据 $Y_d = \{R'_1, R'_2, \dots, R'_n\}$, r 是数据集 X_d 的属性个数, n 是数据集 X_d 的记录数。

2.4 隐私风险函数脱敏参数优化器

基于隐私风险函数和数据质量函数, 本文通过优化器确定最终参数。由于参数组合数庞大, 遍历所有组合不可取, 本文实例中采用深度优先搜索算法实现。脱敏参数优化问题是一个带约束的目标优化问题, 即给定风险预期的情况下, 使得信息损失最小。

$$\begin{aligned} & \text{MAX } DQ(X_d, Y_d) \\ & \text{st. } R(X_d) \leq \text{Expected Risk} \end{aligned} \quad (15)$$

3 实验数据与结果

本文采用 Java 编写自动脱敏程序（代码量 11 000 行），并使用金融机构数据进行实验测试。实验环境为一台 Windows 主机，处理器为 Intel

Core i7-8700 CPU @ 3.20 GHz 3.19 GHz。本文使用 SVM 算法调用 python sklearn 中 SVM 模块。

3.1 实验数据

实验数据集来自某金融机构真实数据和某银行真实交易流水数据，其数据类型如表 2 所示。

表 2 实验数据类型
Table 2 Experimental data type

数据来源	数据类型	技术场景	主要使用场景	预期风险	涉及敏感数据类型
XX 银行 OA 部门	员工信息数据 1	数据内部共享	用于展示	0.2	员工姓名、性别、电话号码、邮编
	员工信息数据 2	对外完全公开	用于展示	0.1	员工姓名、性别、电话号码、邮编
	项目数据	数据内部共享	作为 Key 使用	0.1	姓名、单位名称、金额、链接等
	合同数据	数据内部共享	用于统计	0.3	姓名、合同名称、金额、币种、供应商等
XX 银行 交易部门	个人交易流水 数据	对外领地公开	参与计算	0.1	交易方姓名账户、被交易方姓名账户、交易金额等

（1）员工信息数据

员工信息数据集含有 76 个变量，11 280 条数据，其中包含员工的姓名、工号、性别、手机号、职务、部门、邮箱。OA 部门正在建立数据门户网站，以便内部员工了解数据门户。在该场景下，需要事先将少量数据以脱敏方式展现给所有内部员工。依据技术场景不同，本文将员工信息数据集分为两类：员工信息数据 1 为外部使用，预期风险设置为 0.2；员工信息数据 2 为内部使用，预期风险设置为 0.1。

（2）项目数据

项目数据集是金融机构立项项目的基本信息，含 306 个变量，20 206 条数据，其中包含项目涉及人员及部门信息、项目预算等信息，另外数据存在大量缺失值。这些数据存储在数据中心 API 服务共享中，金融机构内部项目预算系统每隔 20 min 主动调取这些项目数据，但是必须匿名化获取且不影响其作为键值以匹配其他项目相关数据。

（3）合同数据

合同数据集有 37 个变量，4 373 条数据，涉及合同基本信息及供应商信息。内部业务系统期望定期调用这些数据进行分析统计，为保护隐私需要脱敏处理后进行分享。

（4）个人交易流水数据

券商在上市准备阶段，需要核实重要股东的关联交易情况。此时，券商需要银行共享个人交

易流水。但是银行考虑到数据合规要求和客户隐私保护，需要对相关数据进行匿名化处理。

3.2 实验过程

本次实验依照第 2 节提出的自适应脱敏算法选择原理，构建脱敏策略决策模型和脱敏参数优化器自适应不同场景、不同脱敏目标。多个专家进行研判打标，最终使用准确率以及性能评价模型进行评价，其中准确率（ACC）为式(14)。

$$ACC = \frac{TP}{ALL} \quad (14)$$

其中，TP 为脱敏正确的列数，ALL 为所有列数量。

3.3 实验分析

本文以银行个人交易流水场景为例进行结果分析，如表 2 所示，某证券公司分析人员希望对个人交易流水进行关联交易检测。但是交易方姓名账户、被交易方姓名账户、交易金额等均为敏感信息，因此需要在做关联交易检测之前，对其进行脱敏处理。

获取敏感信息后，如何选择合适脱敏算法是脱敏实现的关键，部分字段的脱敏方式结果如表 3 所示。

首先要明确每列数据的使用场景、数据分享类型及期望风险阈值。以 3 级敏感数据交易金额为例，其使用场景是参与关联分析检测，结合其数据类型为连续性数据，为保证数据可用性，可选算法是同态加密、加噪或者不脱敏。假设数据

表3 交易流水—自动脱敏推荐结果

Table 3 Automatic desensitization recommendation result under transaction flow scenario				
列名	敏感类别	敏感级别	使用方式	脱敏算法
incomeExpensesType	交易类型	2	展示使用	掩码
transAt	交易金额	3	统计使用	Laplace 差分隐私
balance	余额	3	展示使用	掩码
...	...	...	...	...
outAcctName	姓名	3	作为 key 使用	基于密码的假名化技术

主体的标识符不需要使用,可以采取基于泛化技术的脱敏技术,若风险已经满足预期风险,对于交易金额,优选的方式是不脱敏。因此数据主体的标识符的脱敏方式也会影响敏感数据的脱敏方式。在该例中,确定主体的准标识符采用基于掩码技术,初始掩码1位,敏感数据交易金额可以不脱敏。

对数据集应用初始化脱敏算法和参数,按照式(2)确定数据泄露风险,如果数据泄露风险小于预期风险且数据质量最大,则停止搜索脱敏参数,从而确定最终脱敏策略。

3.4 实验结论

通过实验分析,基于脱敏策略决策因素和脱敏效果评估体系的自动脱敏技术可有效解决传统基于人工脱敏方法效率低的困境,使得隐私保护透明化。在实验数据集上,推荐脱敏方式的准确率高达95%,平均性能在秒级以内,远远优于人工小时级别,且适合工程化使用。下面具体分析实验结果。

(1) 推荐准确率

在不同数据集上保持预期最低风险不变,遍历不同的使用场景以验证推荐准确率。图4展示实验后应用式(8)和式(11)得出的预期可接受最低风险、脱敏前风险、脱敏后风险情况,观察到脱敏后的隐私风险非常接近预期隐私风险水平,差距小于10%。

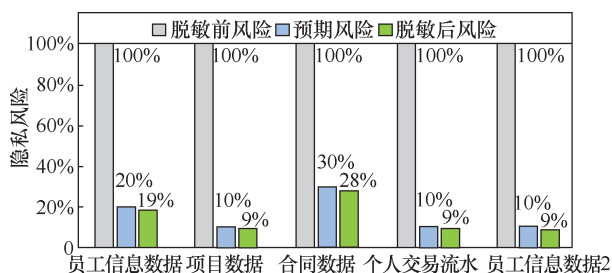


图4 不同数据集下隐私风险评估情况

Figure 4 Privacy risk assessment on different data sets

在满足预期隐私风险情况下,不同数据集下脱敏推荐准确率达95%以上,其中员工信息数据、项目数据接近100%,如图5所示。

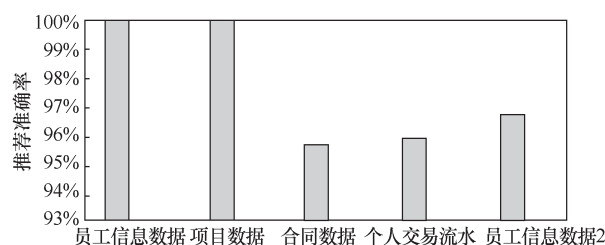


图5 不同数据集下脱敏推荐准确率

Figure 5 Desensitization recommendation accuracy on different data sets

图6展示不同使用场景下脱敏推荐准确率,可以看到用于展示场景和作为Key场景推荐成功率接近100%,参与计算和用于分析场景推荐准确率超过95%。这归因于参与计算和用于分析的场景相对复杂且当前隐私保护领域的保护算法尚未成熟。随着该领域算法趋于成熟,算法推荐准确率将会提升。

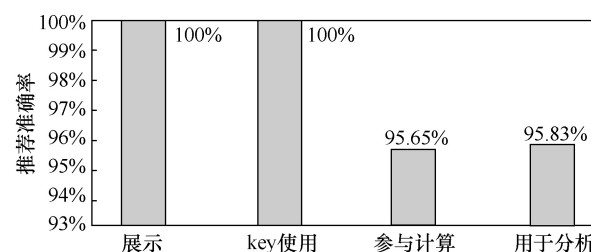


图6 不同使用场景下脱敏推荐准确率

Figure 6 Desensitization recommendation accuracy in different usage scenarios

(2) 推荐效率

从性能角度进行推荐算法评估,如表4所示,分别由10个领域脱敏专家研判并脱敏,专家经验平均耗时1160s,推荐算法运行耗时0.98s。随着文件行数、列数、场景复杂度增加,推荐效率可以提升1000倍。

表4 不同大小的数据集下的性能
Table 4 Performance on different size datasets

数据大小		运行时间/s	专家经验时间/s
列数	行数		
51	9	0.01	1200
25	11 280	0.805	480
38	11 329	0.902	840
155	20 206	3.05	2 400
151	3 002	1.54	2 300
9	4 373	0.03	300
28	1 979	0.552	600

4 结束语

随着法律和金融行业标准对数据安全和隐私保护要求的完善,脱敏技术和工具也越来越重要,将应用在金融数据全生命周期。另外,数据是当今时代的新能源且是国家重要发展战略,数据量和数据类型激增导致数据脱敏需求的增多。指定数据脱敏策略,手动绑定待脱敏金融数据及脱敏算法的方式效率低,会造成应用过程中人工工作量大幅增加,因此在脱敏工具产品中需要实现脱敏算法自动选择等便利化功能。通过结合各类数据分类分级规则、数据脱敏策略以及应用机器学习等技术,实现智能化数据脱敏能力将成为发展方向。本文提供一种基于隐私风险评估的脱敏算法自适应选择技术,在实际金融机构场景下进行实验验证,推荐准确率、自动化效率具备应用条件,后续将在更大范围进行实验验证,优化脱敏算法参数的自动化选择。

参考文献:

- [1] 王卓, 刘国伟, 王岩, 等. 数据脱敏技术发展现状及趋势研究[J]. 信息通信技术与政策, 2020(4): 18-22.
WANG Z, LIU G W, WANG Y, et al. Research on the development and trend of data masking technology[J]. Information and Communications Technology and Policy, 2020, 46(4): 18-22.
- [2] 中华人民共和国个人信息保护法[S]. 2021.
Personal Information Protection Law of The People's Republic of China[S]. 2021.
- [3] 王茂华, 郝云力, 褚亚伟. 具有隐私保护功能的数据加密算法[J]. 计算机工程与应用, 2014, (23): 87-90.
WANG M H, HAO Y L, CHU Y W. Data encryption algorithm with privacy protection function[J]. Computer Engineering and Applications, 2014, 0(23): 87-90.
- [4] 周期律, 张旭春, 蔡仕志. 商业银行客户姓名脱敏技术研究[J]. 中国金融电脑, 2014(6): 65-68.
ZHOU Q L, ZHANG X C, CAI S Z. Research on commercial bank customer name desensitization technology[J]. Financial Computer of China, 2014(6): 65-68.
- [5] LATANYA S, K-ANONYMITY. A model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems: IJUFKS, 2002, 10(5): 557-570.
- [6] WANG M, JIANG Z, ZHANG Y, et al. T-closeness slicing: a new privacy-preserving approach for transactional data publishing[J]. INFORMS Journal on Computing, 2018, 30(3): 438-453.
- [7] SORIA-COMAS J, DOMINGO-FERRER J, SANCHEZ D, et al. t-closeness through microaggregation: Strict privacy with enhanced utility preservation[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11): 3098-3110.
- [8] DONG J, ROTH A, SU W J. Gaussian differential privacy[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2022, 84(1): 3-37.
- [9] CYNTHIA D, JING L. Differential privacy and robust statistics[C]//Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 2009: 371-380.
- [10] BENALOH J. Dense probabilistic encryption[C]//Proceedings of the Workshop on Selected Areas of Cryptography. 1994: 120-128.
- [11] AGGARWAL N, GUPTA C P, SHARMA I. Fully homomorphic symmetric scheme without bootstrapping[C]//Proceedings of 2014 International Conference on Cloud Computing and Internet of Things. 2014: 14-17.
- [12] VENU D, MAYURI A V R, NEELAKANDAN S, et al. An efficient low complexity compression based optimal homomorphic encryption for secure fiber optic communication[J]. Optik, 2022, 252: 168545.
- [13] ZHANG X, LIU C, NEPAL S, et al. A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud[J]. Journal of Computer and System Sciences, 2014, 80(5): 1008-1020.
- [14] MOHAMMADIAN E, Noferesti M, JALILI R. FAST: fast anonymization of big data streams[C]//Proceedings of the 2014 International Conference on Big Data Science and Computing. 2014: 1-8.
- [15] 徐李阳, 李慧芹, 赵文华, 等. 基于卷积神经网络的敏感数据自动脱敏方法[J]. 自动化应用, 2021(10): 86-88.
XU L Y, LI H Q, ZHAO W H, et al. Automatic desensitization method of sensitive data based on convolutional neural network[J]. Automation Application, 2021(10): 86-88.
- [16] MADAN S, GOSWAMI P. Adaptive privacy preservation approach for big data publishing in cloud using k-anonymization[J]. Recent Advances in Computer Science and Communications, 2021, 14(8): 2678-2688.
- [17] CHOROMANSKI K M, JEBARA T, TANG K. Adaptive anonymity via b-matching[J]. Advances in Neural Information Processing Systems, 2013, 26: 3192-3200.
- [18] 佟玲玲, 李鹏霄, 段东圣, 等. 面向异构大数据环境的数据脱敏模型[J]. 北京航空航天大学学报, 2022, 48(2): 249-257.
TONG L L, LI P X, DUAN D S, et al. Data masking model for heterogeneous big data environment[J]. Journal of Beijing University

- of Aeronautics and Astronautics, 2022, 48(2): 249-257.
- [19] WAGNER, ISABEL, ECKHOFF, et al. Technical privacy metrics: a systematic survey[J]. ACM Computing Surveys (CSUR), 2018, 51(3): 1-38.
- [20] ORACLE. Oracle data masking[EB].
- [21] IBM. InfoSphere optim data privacy[EB].
- [22] INFORMATICA. Data masking[EB].
- [23] 熊金波, 王敏荣, 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.
- XIONG J B, WANG M S, TIAN Y L, et al. Research Progress on Privacy Measurement for Cloud Data[J]. Journal of Software, 2018, 29(7): 1963-1980.
- [24] 乔宏明, 梁旻. 运营商面向大数据应用的数据脱敏方法探讨[J]. 移动通信, 2015(13): 17-20, 24.
- QIAO H M, LIANG H. Discussion on operators' data desensitization methods for big data applications[J]. Mobile Communications, 2015(13): 17-20, 24.
- [25] GB/T 42460-2023, 信息安全技术 个人信息去标识化效果分级评估指南[S]. 北京: 中国标准出版社, 2023.
- GB/T 42460-2023, Information security technology—Guide for evaluating the effectiveness of personal information de-identification[S]. Beijing: Standards Press of China, 2023.
- [26] XU W, YUAN K, LI W, et al. An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023, 7(1): 76-88.
- [27] FUNG B, WANG K, CHEN R, et al. Privacy-preserving data publishing: a survey of recent developments[J]. ACM computing surveys, 2010, 42(4): 1-53.
- [28] BAYARDO R J, AGRAWAL R. Data privacy through optimal k-anonymization[C]//21st International Conference on Data Engineering, 2005: 217-228.
- [29] LU Q, WANG C, XIONG Y, et al. Personalized privacy-preserving trajectory data publishing[J]. Chinese Journal of Electronics, 2017, 26(2): 285-291.
- [30] XIAO X, TAO Y. Personalized privacy preservation[C]//Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. 2006: 229-240.
- [31] 姜火文, 曾国荪, 马海英. 面向表数据发布隐私保护的贪心聚类匿名方法[J]. 软件学报, 2017, 28(2): 341-351.
- JIANG H W, ZENG G S, MA H Y. Greedy clustering-anonymity method for privacy preservation of table data publishing[J]. Journal of Software, 2017, 28(2): 341-351.

[作者简介]



祖立军 (1986-), 男, 江苏扬州人, 复旦大学博士生, 中国银联股份有限公司高级工程师, 主要研究方向为开放银行场景下的金融科技研究与应用。



曹雅琳 (1991-), 女, 安徽宿州人, 华为技术有限公司高级工程师, 主要研究方向 AI 通信、数据安全和隐私保护。



门小骅 (1990-), 女, 河北邯郸人, 博士, 中国银联股份有限公司工程师, 主要研究方向为开放银行、金融数据安全。



吕智慧 (1975-), 男, 内蒙古人, 复旦大学教授、博士生导师, 主要研究方向为大数据架构、区块链分布式系统、网络安全技术。



叶家炜 (1977-), 男, 上海人, 复旦大学工程师, 主要研究方向为网络与信息安全、敏感信息保护、金融科技安全。



李泓一 (2000-), 女, 江西南昌人, 复旦大学硕士生, 主要研究方向为金融科技安全。



张亮 (1982-), 男, 山东淄博人, 博士, 华为技术有限公司高级工程师, 主要研究方向为网络 AI 算法、IP 协议与转发。