数据发布中的个性化隐私匿名技术研究

王 波[1,2] 杨 静[1]

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)[1] (哈尔滨理工大学自动化学院 哈尔滨 150080)[2]

摘 要 个性化隐私保护是目前数据发布中隐私泄露控制技术研究的热点问题之一。对这方面的研究现状进行综述。首先,在分析不同类型个性化服务需求的基础上,建立相应的个性化隐私匿名模型;其次,根据采用技术的不同,对已有的个性化隐私保护匿名技术进行总结,并对各类技术的基本原理、特性进行概括性的阐述。同时,根据算法所采用信息度量的差异,给出现有个性化隐私度量的方法与标准。最后,在对比分析已有研究的基础上,总结全文并展望了个性化隐私保护匿名技术的进一步研究方向。

关键词 数据发布,个性化,隐私保护,匿名化

中图法分类号 TP309 **文献标识码** A

Research on Anonymity Technique for Personalization Privacy-preserving Data Publishing

WANG Bo[1,2] YANG Jing[1]

(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)[1]

(School of Automation, Harbin University of Science and Technology, Harbin 150080, China)[2]

Abstract Privacy-preserving data publishing for personalization is a hot research topic in technique for controlling privacy disclosure in data publishing in recent years. An overview of this area was given in this survey. First, on the basis of analyzing types of different personalization service, relevant anonymization models of personalization privacy were built. Second, according to the different adopting techniques, a summarization of the art of personalization privacy-preserving techniques was given, and the fundamental principles and characteristics of various techniques were generally described. In addition, some existing privacy measure methods and standards were provided, according to the differences of information measure of adopted algorithms. Finally, based on the comparison and analysis of existing researches, future research directions of privacy-preserving data publishing for personalization were forecasted.

Keywords Data publishing, Personalization, Privacy-preserving, Anonymization

1 引言

以信息共享与数据挖掘为目的的数据发布过程,往往伴随着个人隐私的泄露风险,数据发布者在发布数据信息前需要对数据集进行敏感信息的保护处理工作。从攻击者的角度来看,攻击的对象主要是敏感信息与数据拥有者之间的关联关系,从而达到其商业或个人利益的目的。因此,破坏这种关联关系,是数据发布中隐私保护的主要研究工作。

近几年,围绕着这一中心任务,国内外研究者开展了大量的工作,其中匿名技术便是一种隐私保护的有效方法。自Samarati和Sweeney首次提出匿名化概念之后[1],该技术得到了广泛的关注,涌现了大量的研究成果[2-25],同时相继出现了匿名化隐私保护的总结性文献[26,27]。然而,目前大部分匿名化方法都没有考虑不同个体的隐私保护需求。事实上,我们都希望自己被视为一个个性的个体,而不仅仅是一个数字。数据拥有者有权利限定自身隐私信息的保护程度和使用方式,要求在隐私保护的同时存在个性化服务[28]。自Xiao等人在2006年的VLDB会议上首次展开对个性化隐私保护的研究后,研究者越来越重视这一方面的研究,提出了一些相关的解决方法[29-34],个性化隐私保护也成了数据发布中隐私泄露控制技术的研究热点之一。

本文立足于隐私保护数据发布中的个性化服务需求,首次对个性化隐私匿名技术的最新进展进行综述。首先,根据个性化需求表现形式的不同,分类建立了不同的个性化隐私匿名模型;其次,对各类个性化隐私匿名技术进行分析、总结,并给出隐私匿名度量的标准;最后,对全文进行总结并展望了个性化隐私匿名技术进一步的研究方向。

2 个性化隐私匿名模型

在本文的研究中,待发布的数据集为二维表。令 T 为待发布数据集,包含4类属性:

(a)个体标识符属性(identifier attribute, ID), $ID=$

到稿日期:2011-05-17 返修日期:2011-09-02 本文受国家自然科学基金(61073041,61073043),黑龙江省自然科学基金(F200901)和哈尔滨市学科带头人专项基金(2011RFXXG015)资助。

王 波(1982—),男,博士生,讲师,主要研究方向为隐私保护、数据库安全等,E-mail:hust_wb@126.com;杨 静(1962—),女,教授,博士生导师,主要研究方向为数据库理论与应用、数据挖掘技术、知识库系统、软件理论等。

$\{A^{ID}\}$；

(b)准标识符属性(quasi-identifier attribute, QI)，$QI=\{A_1^{QI},A_2^{QI},\cdots,A_m^{QI}\}$；

(c)敏感属性(sensitive attribute, SA)，$SA=\{A_1^S,A_2^S,\cdots,A_n^S\}$；

(d)其他属性(other attributes, OA)。

一般地，假设指定敏感属性 SA 为分类型数据，准标识符属性 QI 可以是分类型或数值型数据，个体标识符 ID 能惟一确定当前个体。表 1 是一个待发布的个人医疗原始信息记录表，其中属性列[Name]为个体标识符属性，[Gender]、[Age]和[Zip code]为准标识符属性，[Disease]为敏感属性。

表 1　待发布的个人医疗原始信息记录表

No.	Name	Age	Sex	Zip code	Disease
1	Fred	24	M	13000	Gastric ulcer
2	Andy	28	M	12000	Dyspepsia
3	James	27	M	17000	Pneumonia
4	Allen	29	M	19000	Bronchitis
5	Jacky	31	M	21000	Pneumonia
6	Jeff	38	M	24000	Pneumonia
7	Emily	42	F	56000	Flu
8	Judy	45	F	38000	Gastritis
9	Rita	47	F	37000	Pneumonia
10	Wendy	76	F	32000	Flu

根据个性化服务表达方式的不同，大致可以将个性化隐私保护模型分为两类：一类是面向个体的个性化隐私匿名模型；另一类是面向敏感值的个性化隐私匿名模型。下面将分别介绍这两类模型。

2.1　面向个体的个性化隐私匿名模型

面向个体的个性化隐私匿名模型是数据拥有者通过公开的敏感属性继承分类信息，对自身敏感值设定其隶属属性值来表达自身对敏感属性值的个性化需求，从而实现隐私匿名的个性化服务。模型中相关概念介绍如下。

定义 1(继承分类树)　对于分类型属性 A，记 $Dom(A)$ 为 A 的有限值域，属性 A 的继承分类树定义为四元组 $ITax(A)=\{r_A,L_A,IN_A,R_A\}$，其中，

(a)r_A 表示分类树的根节点；

(b)L_A 表示分类树叶子节点的集合，$L_A=Dom(A)$；

(c)IN_A 表示分类树中间节点的集合，集合中的元素代表了属性 A 中各取值继承分类的不同程度；

(d)R_A 表示分类树中各节点间的隶属关系。

图 1 给出了表 1 中分类型敏感属性[Disease]的继承分类树。其中，叶子节点包含了属性列[Disease]中全部的取值，根节点代表所有疾病，从根节点到叶子节点的中间节点代表疾病的若干分类。一般地，继承分类树是通用的分类信息，需要公开发布。

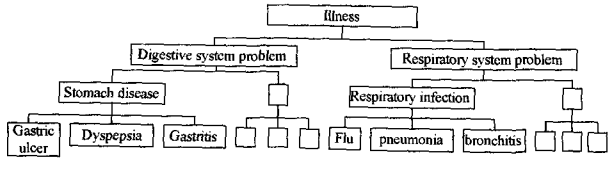

图 1　敏感属性[Disease]的继承分类树

定义 2(继承分类子树)　对于继承分类树 $ITax(A)$ 中的任意节点 N，称以该节点为根节点，包含该节点所有子孙节点的子树为 $ITax(A)$ 的继承分类子树，记为 $SUB\_ITax(N)$。根据上述定义，继承分类子树也可表示为四元组 $SUB\_ITax(N)=\{N,L_N,IN_N,R_N\}$。

定义 3(保护节点)　给定元组 $t\in T$，其敏感属性值为 $t.s$，则元组 t 的保护节点 $t.GN$ 定义为继承分类树 $ITax(A^S)$ 从根节点 r_{A^S} 到 $t.s$ 所在叶子节点路径上的任何一个节点。

表 2 是对表 1 中各敏感属性值根据图 1 的继承分类树，加入相应保护节点的个人医疗原始信息记录表。属性列[Guarding node]中各值即为属性列[Disease]中各敏感值的保护节点。

表 2　指定保护节点的个人医疗原始信息记录表

No.	Name	Age	Sex	Zip code	Disease	Guarding node
1	Fred	24	M	13000	Gastric ulcer	Stomach disease
2	Andy	28	M	12000	Dyspepsia	Dyspepsia
3	James	27	M	17000	Pneumonia	Respiratory infection
4	Allen	29	M	19000	Bronchitis	Bronchitis
5	Jacky	31	M	21000	Pneumonia	Pneumonia
6	Jeff	38	M	24000	Pneumonia	Pneumonia
7	Emily	42	F	56000	Flu	Ø
8	Judy	45	F	38000	Gastritis	Gastritis
9	Rita	47	F	37000	Pneumonia	Respiratory infection
10	Wendy	76	F	32000	Flu	Flu

数据拥有者 o_t 通过指定保护节点 $t.GN$ 来表达对敏感属性值的个性化需求，保护节点 $t.BN$ 意味着 o_t 不希望数据攻击者能联想到以 $t.GN$ 为根节点的继承分类树 $SUB\_ITax(t.GN)$ 中任何一个叶子结点。也就是说，当攻击者推断出 $t.s$ 所在的继承分类子树 $SUB\_ITax(t.BN)$ 中任何一个叶子节点与数据拥有者的关联关系 $\langle o_t,A_i^S\rangle$，其中 A_i^S 为 $SUB\_ITax(t.GN)$ 的叶子节点，则数据拥有者就认为他/她的隐私受到了侵犯。表 1 中属性列[Guarding node]给出了各元组按照图 1 的公开继承分类树对各自敏感属性取值的保护节点。例如，对于表 1 中的记录 1，Fred 患有疾病[Gastric ulcer]，他指定的保护节点为[Stomach disease]，则如果攻击者推断出 Fred 患有[Gastric ulcer]、[Dyspepsia]以及[Gastritis]中的任何一种疾病，他都认为自身的隐私受到了侵犯。

综上所述，面向个体的个性化隐私匿名模型是从个体的隐私保护需求出发，对敏感属性进行泛化处理，从而扩大敏感属性与数据拥有者的关联范围，体现了隐私保护的个性化服务。

2.2　面向敏感值的个性化隐私匿名模型

面向敏感值的个性化隐私匿名模型是由数据拥有者提供自身对敏感属性值的个性隐私敏感因子，数据发布者根据敏感因子来计算个性隐私保护需求度，从而实现隐私匿名的个性化服务。下面给出这类模型的一个例子。

定义 4(个性隐私敏感因子)　称个体 i 对敏感属性值 S_j 的重视程度为个性隐私敏感因子，记作 psf_{ij}。

$$psf_{ij}=\begin{cases}1, & \text{个体 } i \text{ 对敏感值 } S_j \text{ 非常重视，需要严格保护}\\ -1, & \text{个体 } i \text{ 对敏感值 } S_j \text{ 毫不重视，无需进行保护}\\ 0, & \text{其它情况}\end{cases} \tag{1}$$

式中，当个体 i 对自身的敏感属性值 S_j 非常重视，表明该个体不希望泄露 S_j 值，取 $psf_{ij}=1$；当个体 i 对自身的敏感属性值 S_j 毫不重视，表明该个体的 S_j 值可以不用保护，取 $psf_{ij}=-1$；当个体 i 对自身的敏感属性值 S_j 的保护程度不能确定或无所谓的情形下，取 $psf_{ij}=0$。在表 1 的待发布的数据集

中加入个体 i 对敏感属性值 S_j 的个性隐私敏感因子，就形成带 psf_{ij} 的个人医疗信息记录表，如表 3 所列。

表 3　带 psf_{ij} 的个人医疗原始信息记录表

No.	Name	Age	Sex	Zip code	Disease	psf_{ij}
1	Fred	24	M	13000	Gastric ulcer	1
2	Andy	28	M	12000	Dyspepsia	1
3	James	27	M	17000	Pneumonia	0
4	Allen	29	M	19000	Bronchitis	0
5	Jacky	31	M	21000	Pneumonia	−1
6	Jeff	38	M	24000	Pneumonia	0
7	Emily	42	F	56000	Flu	−1
8	Judy	45	F	38000	Gastritis	1
9	Rita	47	F	37000	Pneumonia	1
10	Wendy	76	F	32000	Flu	−1

根据定义 4，可以定义各个敏感属性值在表中的个性隐私保护需求度 prf_j。

定义 5(个性隐私保护需求度)　在一个待发布表 T 中，称记录集对各个敏感属性值隐私保护程度的统计概率为敏感属性值 S_j 的个性隐私保护需求度，记为 prf_j。令 $|psf_{ij}=0|$，$|psf_{ij}=1|$ 和 $|psf_{ij}=-1|$ 分别表示敏感属性值 S_j 的隐私敏感因子取 0，1 和 −1 的个数，$|psf_{ij}|$ 表示敏感属性值 S_j 的总个数，显然有 $|psf_{ij}|=|psf_{ij}=0|+|psf_{ij}=1|+|psf_{ij}=-1|$。

在这里，对于 $psf_{ij}=0$ 的个体，由于个体自身对敏感属性值的重视程度不能确定，我们认为这些个体以 50% 的概率要求保护其隐私，则敏感属性值 S_j 的个性隐私保护需求度 prf_j 的计算方法定义为

$$prf_j=\frac{|psf_{ij}=1|+|psf_{ij}=0|/2}{|psf_{ij}|}=\frac{2|psf_{ij}=1|+|psf_{ij}=0|}{2|psf_{ij}|} \tag{2}$$

根据式(2)，对表 3 中的各个敏感属性值的个性隐私保护需求度的计算结果如表 4 所列。

表 4　表 3 各个敏感属性值的个性隐私保护需求度

S_j	Gastric ulcer	Dyspepsia	Pneumonia	Bronchitis	Flu	Gastritis
psf_{ij}	1.0	1.0	0.5	0.5	0	1.0

综上所述，面向敏感值的个性化隐私匿名模型是以记录拥有者对自身敏感值的保护需求为出发点，对敏感值设定个性隐私保护需求度，从而实现隐私匿名的个性化。

3　个性化隐私匿名技术

个性化隐私匿名技术实质上是在传统匿名技术的基础上，融入了个性化服务需求的匿名方法。具体地说，就是对数据表中的每条记录或敏感属性提供不同的隐私保护需求粒度，而不是传统匿名策略只提供整个数据表的保护粒度。在实现技术上，与传统的匿名技术大体相同。下面对这些技术做一个总结和介绍。

3.1　基于泛化与限制的匿名技术

匿名化技术最早使用的方法便是泛化与限制待发布数据表中的相关属性值。泛化，就是用更概括、更抽象的数据或数据区域来代替原始数据值；限制，顾名思义，就是将待发布数据表中的某些属性值或记录直接删除。根据泛化对象的不同，可将其分为“全域泛化”和“局部泛化”。

“全域泛化”，就是对数据表中的整个属性域进行泛化。这一类算法对每一个准标识符属性构建泛化层次树，然后根据匿名原则对不满足隐私要求的取值用泛化值(域)来代替当前值。经典的算法有 Datafly 算法[2]、MinGen 最小泛化算法[4]以及 Incognito 算法[10]等。全域泛化虽然在一定程度上增强了隐私的安全性，但常常也会导致一些不必要的信息损失。针对这一缺陷，研究者提出了“局部泛化”的思想，对其进行改进，如 Iyengar 提出的基于遗传算法的不完全搜索策略[7]，Wang 等提出的“自底向上”[8]和“自顶向下”[9]泛化的贪婪算法，Lefevre 等提出的基于多维空间划分的 k-匿名方法[13]等。

基于泛化与限制的匿名技术，其不足之处是在匿名过程中会造成一些不必要的信息损失。同时，在预定义属性的泛化层次结构时，没有一个统一的标准，对不同类型的数据也会有所差异。这些都是泛化匿名需要解决的问题。

3.2　基于聚类的匿名技术

聚类是数据挖掘中应用甚广的数据分析方法。在数据发布时，基于聚类匿名的基本思想是：首先将待发布的数据集划分为若干簇，其中簇内记录相关，簇间记录不相关；然后将每个簇内记录的准标识符泛化为相同的属性值，生成等价类，实现数据集的匿名化。

对数据集进行聚类处理，在一定程度上增强了匿名等价类的同质性，使发布的数据集具有更高的安全性。由于同样需要对簇内记录进行泛化，因此在对原始数据集进行聚类前需要设定匿名模型，例如基于 k-匿名的聚类[15]、基于 l-多样性的聚类[17]。Aggarwal 等对聚类匿名做了进一步的研究，提出了 r-gather 和 r-cellular 方法[19]。王智慧等对传统的聚类方法进行改进，并渗入 l-多样性的思想，提出了一种基于聚类的数据匿名方法 L-clustering[21]。近年来，国内外研究学者基于聚类的思想，将 SDC(Statistical Disclosure Control)技术中的微聚集(Micmaggregation)方法引入到数据表的匿名化模型上，相继提出了许多相关的方法。韩建民等对这一技术做了总结，并对其以后的研究方向做了展望[22]。

基于聚类的匿名技术在数据的安全性以及信息的损失量上都优于传统的泛化匿名方法，但聚类本身需要解决的问题在匿名处理的过程中仍然是需要进一步研究的方向。

3.3　基于交换与分割的匿名技术

数据交换与分割技术是基于隐私保护的数据挖掘领域中常用的方法。在数据发布匿名化过程中，不对原始数据进行泛化，而是将数据表中的某些数据项按照一定的规则进行交换，以达到保护隐私的目的。

最常用的方法是对待发布的数据表进行拆分，然后对各个发布的数据子表进行有损连接，从而达到数据匿名的效果。这类算法包括，在满足 l-多样性的基础上，Xiao 等提出的基于有损连接的 Anatomy 分解匿名方法[24]；Wong 等提出的基于有损连接的 (α,k)-匿名隐私保护方法[16]；杨晓春等提出的面向多敏感属性的多维桶分组技术[25]等。

通过交换或分组使等价类中记录的准标识符属性与敏感属性以多对多关系的形式发布，在很大程度上保持了数据的完整性，增强了匿名数据的可用性，能够很好地满足隐私保护的要求。但由于需要对待发布的数据表进行分割处理，相应也增加了系统的空间开销。

4　个性化隐私匿名度量评价

隐私匿名需要在保护敏感信息隐私不被泄露的同时，考

虑发布数据的可用性以及信息的损失程度等。在个性化隐私匿名的度量过程中，可以借鉴传统的匿名度量标准。本节将结合个性化隐匿匿名模型的特征，从发布数据的隐私侵犯率以及信息损失两个方面对隐私匿名技术进行度量与评价。

4.1 隐私侵犯率度量

隐私侵犯率在一定程度上反映了发布数据的安全程度。Xiao 等针对第一类个性化匿名模型，提出了隐私侵犯率的概念以及计算方法[29]。

定义 6(隐私侵犯率) 给定元组 $t\in T$，其隐私侵犯率 $P_{\text{breach}}(t)$ 定义为攻击者从发布数据表 PT 中推断出存在于原始数据表 T 中的关联关系 $\langle o_t, A_i^S\rangle$ 的概率，其中 A_i^S 为 $SUB\_ITax(t.GN)$ 的叶子节点。

对于元组 $t\in T$，令敏感属性与其数据拥有者可能的关联组合数为 N_r，其中违反隐私规则的组合总数为 N_b，则根据定义 6，元组 t 隐私被侵犯的概率为

$$P_{\text{breach}}(t)=N_b/N_r \tag{3}$$

4.2 隐私披露风险度量

隐私披露风险是与隐私侵犯相对应的一个标准，反映了攻击者通过发布的数据表与其它相关背景知识链接后得到隐私信息的概率。

对于元组 $t\in T$，其数据拥有者为 o，敏感属性值为 s，令 S_k 表示“攻击者在背景知识 K 下获得关联关系 $\langle o,s\rangle$”的事件，则隐私披露风险 $r(s,K)$ 可表示为

$$r(s,K)=P_r(S_k) \tag{4}$$

在传统的匿名模型中，数据发布者对整个数据集设定阈值 $\alpha\in[0,1]$，要求所有敏感数据的披露风险都小于等于 α，则称 α 是该数据集的披露风险。引入个性化服务后，要求对数据集中的每一条记录设定相应的披露风险值。正如第二类个性化隐私匿名模型，根据对个性隐私保护需求度的定义，要求等价类 φ_i 中任一敏感属性值 S_j 在该等价类中出现的频率都不超过 α_j，其中 $\alpha_j=1-prf_j$，则个性化的隐私披露风险可定义为向量 $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)$。

4.3 信息损失度量

无论是面向敏感值或是面向个体的个性化隐私匿名模型，在发布匿名数据表的过程中，都必然会产生信息的损失。由于各种匿名技术实现方式以及应用场景的不同，不可能用惟一的信息损失标准来衡量全部的匿名算法，只能说在某一特定的应用中，某一种方法优于其它另外的方法。表 5 对各种信息损失度量标准进行了对比分析。

表 5 各种信息损失度量方法对比分析

度量机制	度量标准描述	应用示例
数据匿名前后的可辨识度	通过对比原始数据和匿名发布后数据的可辨识程度	Mondrian[13]、Incognito[10]
基于数据泛化层次结构	通过对属性泛化树或泛化关系的高度，计算匿名的代价	Generalization Hierarchy[4]
基于限制单元总数	通过匿名前后的记录行之间的相似程度，计算相应的信息损失	Hamming distance[35]、Diameter[36]
基于分块	通过分类的惩罚系数之和来定义表的信息损失度	Classification[7]、C_{AVG}[13]
基于熵	通过属性的泛化熵损益来定义记录或表的信息损失度	Datafly[2]、L-diverity[14]

如表 5 中所列，由于匿名实现方式的不同，因此出现了不同信息损失度量标准。基于数据泛化层次结构和基于熵的度量机制以匿名前、后的属性为研究对象，而另外的 3 类度量机制以匿名前、后的记录为度量对象。这些度量方法在其特定的应用环境中都存在着各自的合理性，所以在选择信息损失度量标准的时候，需要考虑匿名方法的实现方式。

结束语 本文对数据发布中的个性化隐私匿名技术进行了总结和综述。首先简要介绍了个性化隐私匿名技术的研究背景，然后再根据个性化服务实现方式的不同，对个性化隐私匿名模型进行了分类。第 3 节中，在传统隐私匿名技术的基础上，详细介绍了个性化隐私匿名的相关技术，并指出需要进一步研究的内容。第 4 节结合个性化隐私匿名模型的特征，给出了个性化隐私匿名的度量标准与方法。

隐私保护中的个性化服务研究具有非常广泛的应用前景。随着数据发布形式与内容的不断更新，对隐私保护的个性化需求也必将越来越重视。个性化系统固有的复杂性、动态性、分散性，以及原始数据日益呈现的高维特征，都是隐私保护的个性化需要亟待解决的问题。

另外，目前的各种方法大部分只停留在理论研究阶段，并没有在实际中得到应用。因此，个性化隐私匿名进一步研究的重点问题不仅需要更好地平衡个性化服务质量和隐私保护的效果，而且需要考虑如何提高算法的效率，使之应用于大规模的实际系统中。

参 考 文 献

[1] Samarati P, Sweeney L. Generalizing data to provide anonymity when disclosing information [C]// Proceedings of the 17th ACM Sigact-Sigmod-Sigart Symposium on Principles of Database Systems. New York: ACM Press, 1998: 188

[2] Sweeney L. Computational disclosure control: a primer on data privacy protection [D]. Massachusetts Institute of Technology, 2001: 67-82

[3] Sweeney L. k-anonymity: A model for protecting privacy [J]. International Journal on Uncertainty, Fuzziness and Knowledge Based Systems, 2002, 10(5): 557-570

[4] Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression [J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10 (5): 571-588

[5] Aggarwal G, Feder T, Kenthapadi K, et al. k-anonymity: algorithms and hardness [R]. Stanford University, 2004

[6] Meyerson A, Williams R. On the complexity of optimal k-anonymity [C]// Proceedings of the ACM SIGMOD-SIGACT-SIGART Principles of Database Systems. ACM Press, 2004: 223-228

[7] Iyengar V. Transforming data to satisfy privacy constraints [C]// Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2002: 279-288

[8] Wang K, Yu P, Chakraborty S. Bottom-up generalization: a data mining solution to privacy protection [C]// Proceedings of the 4th IEEE International Conference on Data Mining. 2004: 249-256

[9] Fung B, Wang K, Yu P. Top-down specialization for information and privacy preservation [C]// Proceedings of the 21st International Conference on Data Engineering. 2005: 205-216

[10] Lefevre K, Dewitt D J, Ramakrishnan R. Incognito: Efficient full-domain k-anonymity [C]// Proceedings of the 24th ACM SIGMOD International Conference on Management of Data. 2005: 49-60

(下转第 200 页)

Computer Vision, Graphics, and Image Processing-Image Understanding, 1993, 57(2): 131-154

[3] Matas J. Robust detection of line using the progressive probabilistic hough transform [J]. Computer Vision and Image Understanding, 2000, 78(1): 119-137

[4] 谢维波，王永初，郑艺泉. 基于多粒度数据融合的直线检测算法[J]. 计算机科学，2007，34(9)：213-217

[5] 徐胜华，朱庆，刘纪平，等. 基于预存储权值矩阵的多尺度Hough变换直线提取算法[J]. 测绘学报，2008，37(1)：83-88

[6] Dahyot R. Statistical Hough transform [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009, 31(8): 1502-1509

[7] Guo S Y, Prodmore T, Kong Y G, et al. Improved Hough transform voting scheme utilizing surround suppression [J]. Pattern Recognition Letters, 2009, 30(13): 1241-1252

[8] 魏志强，孙亚兵，纪筱鹏，等. 数字城市中矩形建筑物区域的自动获取[J]. 计算机科学，2009，36(1)：211-215

[9] Hubert M, Rousseeuw P J, van Aelst S. High-breakdown robust multivariate methods [J]. Statistical Science, 2008, 23(1): 92-119

[10] Hawkins D M, Olive D J. Inconsistency of resampling algorithms for high breakdown regression estimators and a new algorithm [J]. Journal of the American Statistical Association, 2002, 97(457): 136-159

[11] 张祥德，王琪，黄亚平，等. 采用最小均方误差筛选参数的Hough变换及其应用[J]. 东北大学学报：自然科学版，2007，28(9)：1270-1273

(上接第 171 页)

[11] 刘向宇，杨晓春，于戈. 一种基于特征类的高精度隐私保护数据发布方法[J]. 计算机科学，2005，32(7A)：368-373

[12] 杨晓春，刘向宇，王斌，等. 支持多约束的K-匿名化方法[J]. 软件学报，2006，17(5)：1222-1231

[13] Lefevre K, Dewitt D, Ramakrishnan R. Mondrian multidimensional k-anonymity [C]// Proceedings of the 22nd International Conference on Data Engineering. 2006: 25-34

[14] Machanavajjhala A, Gehrke J, Kifer D. L-diversity: Privacy beyond k-anonymity [C]// Proceedings of the 22nd IEEE International Conference on Data Engineering(ICDE'06). 2006: 24-36

[15] Li Jiu-yong, Wong R C W, Fu A C, et al. Achieving k-anonymity by clustering in attribute hierarchical structures [C]// Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery. 2006: 405-416

[16] Wong R C, Li J, Fu A W, et al. (α, k)-anonymity: an enhanced k-anonymity model for privacy-preserving data publishing [C]// Proceedings of 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2006: 754-759

[17] Byun J-W, Kanira A, Bertino E, et al. Efficient k-anonymity using clustering technique [R]. CERIAS Technical Report 2006-10. West Lafayette, Indiana, USA: Purdue University, 2006

[18] Truta T M, Vinay B. Privacy protection: p-Sensitive k-anonymity property [C]// Proceedings of the 22nd International Conference on Data Engineering Workshops. 2006: 94

[19] Aggarwal G, Feder T, Kenihapadi K, et al. Achieving anonymity via clustering [C]// Proceedings of the 25th ACM SIGMOD SIGACT SIGART Symposium on Principles of Database Systems. 2006: 153-162

[20] Li N, Li T. t-closeness: privacy beyond k-anonymity and l-diversity [C]// Proceedings of IEEE 23rd International Conference on Data Engineering. 2007: 106-115

[21] 王智慧，许俭，汪卫，等. 一种基于聚类的数据匿名方法[J]. 软件学报，2010，21(4)：680-693

[22] 韩建民，岑婷婷，虞慧群. 数据表K-匿名化的微聚集算法研究[J]. 电子学报，2008，36(10)：2021-2029

[23] 童云海，陶有东，唐世渭，等. 隐私保护数据发布中身份保持的匿名方法[J]. 软件学报，2010，21(4)：771-781

[24] Xiao Xiao-kui, Tao Yu-fei. Anatomy: Simple and effective privacy preservation [C]// Proceedings of the 32nd International Conference on Very Large Databases. 2006: 139-150

[25] 杨晓春，王雅哲，王斌，等. 数据发布中面向多敏感属性的隐私保护方法[J]. 计算机学报，2008，31(4)：574-586

[26] 刘喻，吕大鹏，冯建华，等. 数据发布中的匿名化技术研究综述[J]. 计算机应用，2007，27(10)：2361-2364

[27] 王平水，王建东. 匿名化隐私保护技术研究综述[J]. 小型微型计算机系统，2011，32(2)：248-252

[28] 曾春，邢春晓，周立柱. 个性化服务技术综述[J]. 软件学报，2002，13(10)：1952-1960

[29] Xiao Xiao-kui, Tao Yu-fei. Personalized privacy preservation [C]// Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data. New York, NY, USA, 2006: 229-240

[30] Ye X J, Zhang Y W, Liu M. A Personalized (α, k)-Anonymity Model [C]// Proceedings of the 9th International Conference on Web-Age Information Management(WAIM'08). 2008: 341-348

[31] 刘明，叶晓俊. 个性化K-匿名模型[J]. 计算机工程与设计，2008，29(2)：282-286

[32] Shen Y G, Liu Y H, Zhang Y L. Personalized-granular k-Anonymity [C]// Proceedings of International Conference on Information Engineering and Computer Science(ICIECS'09). 2009: 1-4

[33] 申艳光，刘永红. 个性粒化k-匿名隐私保护方法研究[J]. 数学的实践与认识，2010，40(11)：97-104

[34] 韩建民，于娟，虞慧群，等. 面向敏感值的个性化隐私保护[J]. 电子学报，2010，38(7)：1723-1728

[35] Aggarwal G, Feder T. Anonymizing tables for privacy protection [EB/OL]. http://theory.standford.edu/ rajeev/privacy.html, 2004

[36] Meyerson A, Williams R. On the complexity of optimal k-anonymity [C]// Proceedings of the ACM SIGMOD-SIGACT-SIGART Principles of Database Systems. ACM Press, 2004: 223-228