文章编号:1007-791X(2016)03-0263-07

基于差分隐私的空间分割研究综述

彭慧丽[1,2],张啸剑[1,*]

(1.河南财经政法大学 计算机与信息工程学院,河南 郑州 450046;2. 河南广播电视大学,河南 郑州 450008)

摘 要:随着基于空间数据应用的广泛出现,如何保护空间数据中的隐私位置信息成为当前面临的重大挑战。针对差分隐私下空间数据分割问题展开综述,该研究对基于位置的服务、交通管理、个性化推荐等具有重要意义。首先介绍了差分隐私保护技术的理论基础;其次详细介绍基于差分隐私的空间分割方法的研究进展,并进行对比分析;最后对未来研究方向给予展望。

关键词:隐私保护;空间分割;差分隐私;数据无关性分割;数据相关性分割

中图分类号: TP309.2 **文献标识码**: A DOI:10.3969/j.issn.1007-791X.2016.03.013

0 引言

随着信息时代的飞速发展,空间数据的获取与收集变得尤为容易,例如移动用户位置、GPS 位置、家庭住址等数据。通过对空间数据进行分割与分析,许多基于此类数据应用(如交通监控、导航系统、位置推荐等)的服务质量得以提高。然而,空间数据通常蕴含着丰富的个人敏感信息,在提供给第三方应用的同时,个人的敏感信息有可能被泄露。例如,在移动用户轨迹中,轨迹数据通常蕴含着丰富的个人敏感信息,比如就医地点,出租车上车、下车地点等。空间数据隐私是指个人实体不愿意被外部知晓的位置或轨迹信息(例如,家庭住址、就医地点等)。在过去十几年中,如何分割空间数据而又不披露数据中所蕴含的隐私是数据库领域的研究热点。虽然出现了多种基于 k-匿名[1-2]及其变种(例如 l-diversity[3]等)的分割方法,但是这些方法通常对攻击背景知识与攻击模型给出特定的假设,而现实中这些假设并不成立。差分隐私[4-8]技术的出现和广泛认可使得空间数据分割变为可能。该技术可以确保在某一数据集中插入或者删除一条记录的操作不会影响任何查询的输出结果,从而保证了每条记录删除或者加入该数据集不会对其隐私造成危险[9]。相比传统的隐私保护模型,差分隐私具有众多显著优点,最重要的是所提供的隐私保护不依赖于攻击者的背景知识,同时对于数据分割与查询也能提供严格的可量化的隐私保证,这使得差分隐私对数据发布与查询中的隐私保护非常合适。本文着眼于基于差分隐私的空间数据分割研究,对于差分隐私保护模型的理论基础进行介绍,然后从当前常用的空间分割方法的优缺点、关键技术以及适用范围进行综合对比分析。

1 差分隐私保护模型

1.1 差分隐私定义

差分隐私保护的核心思想是:给定一个数据集 D,r 是 D 所包含的记录。设 f 是 D 上任意的查询或者分析操作,操作结果为 $f(D)$。如果把 r 从 D 中删除(或者添加另外一条新的记录),然后进行 f 操作,如果操作结果仍然是 $f(D)$,则认为 r 的敏感信息在 D 中不会有额外风险。基于此分析,给出空间数据分割上的差分隐私形式化定义。

定义1 给定一个空间数据分割方法 A,

收稿日期:2016-01-02 **基金项目**:国家自然科学基金资助项目(61502146);河南省科技攻关项目(162102310411);河南省科技厅基础与前沿技术研究项目(152300410091);河南省教育厅高等学校重点科研项目(16A520002)

作者简介:彭慧丽(1981-),女,河南郑州人,硕士,讲师,主要研究方向为数据挖掘、隐私保护;*通信作者:张啸剑(1980-),男,河南周口人,博士,讲师,主要研究方向为数据挖掘、隐私保护,Email:xjzhang82@ ruc.edu.cn。

Range(A)为 A 的输出范围,若 A 在 $D=\{d_1, d_2, \cdots, d_n\}$ 与 $D'=\{d_1, d_2, \cdots, d_{r-1}, d_{r+1}, \cdots, d_n\}$ 上任意分割结果 $T(T\in \text{Range}(A))$ 满足下列不等式,则 A 满足 ε-差分隐私。

$$\Pr[A(D)=T]\leqslant \exp(\varepsilon)\cdot \Pr[A(D')=T], \quad (1)$$

其中,D 与 D' 为相差一个空间数据点的近邻关系,ε 表示隐私预算,其值越小则算法 A 的隐私保护程度越高。

从定义 1 可以看出,ε-差分隐私限制了任意一个空间数据点对算法 A 输出结果的影响。实现差分隐私保护需要噪音机制的介入,拉普拉斯机制[10]与指数机制[11]是实现差分隐私的主要技术。而所需要的噪音大小与其响应查询或者分析函数 f 的全局敏感性密切相关。

定义 2 设 f 为某一个查询,且 $f: D\rightarrow \mathbf{R}^d$,$f$ 的全局敏感性为

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

其中,$\mathbf{R}$ 为映射的实数空间,d 为 f 的查询维度。

文献[10]提出的拉普拉斯机制可以取得差分隐私保护效果,该机制利用拉普拉斯分布产生噪音,进而使得发布方法满足 ε-差异隐私,如定理 1 所示。

定理 1[10] 设 f 某一个查询函数,且 $f: D\rightarrow \mathbf{R}^n$,若方法 A 符合下列等式,则 A 满足 ε-差异隐私,

$$A(D)=f(D)+\langle Lap(\lambda)\rangle^n, \quad (3)$$

其中,$Lap(\lambda)$ 为相互独立的拉普拉斯噪音变量,$\lambda\geqslant \Delta f/\varepsilon$。噪音通常由拉普拉斯分布产生,即 $\Pr(Lap(\lambda)=x)=\frac{1}{2\lambda}e^{-|x|/\lambda}$,噪音量大小与 Δf 成正比,与 ε 成反比。因此,查询 f 的全局敏感性越大,所需的噪音越多。

文献[11]提出的指数机制主要处理抽样算法的输出为非数值型的结果,例如在空间结点中寻找分割线。该机制的关键技术是如何设计打分函数 $u(D, d_i)$。设 A 为指数机制下的某个隐私方法,则 A 在打分函数作用下的输出结果为

$$A(D,d_i)=\left\{d_i: \left| \Pr[d_i\in \Omega]\propto \exp\left(\frac{\varepsilon u(D,d_i)}{2\Delta u}\right)\right.\right\}, \quad (4)$$

其中,Δu 为打分函数 $u(D, d_i)$ 的全局敏感性,Ω 为采用算法的输出域。由该式可知,d_i 的打分函数越高,被选择输出的概率越大。

1.2 查询误差度量

满足差分隐私的空间数据分割的最终目的是能够比较精确地响应查询,或者给出比较精确的分析结果。然而,有拉普拉斯噪音的引入,不可避免地会带来噪音误差,而噪音误差通常采用绝对误差($\sqrt{2}\Delta f/\varepsilon$)或者方差($\sqrt{2}(\Delta f/\varepsilon)^2$)来度量。目前差分隐私下的分割方法可以分为两类:基于数据无关的空间数据分割技术、基于数据相关的空间数据分割技术。如果查询来自于数据无关的空间数据分割方法,查询误差通常由噪音误差与均匀假设误差来度量;如果查询来自于数据相关的空间数据分割方法,查询误差由噪音误差与分割误差来度量,或者有噪音误差与数据转换误差来度量。接下来主要介绍数据无关的空间数据分割方法与基于数据相关的空间数据分割方法。

2 基于差分隐私的空间数据分割方法

2.1 基于数据无关的空间数据分割方法

基于数据无关的空间分割常用的索引技术是网格结构与树结构(Quad-树、b-ary 树等)。这类分割方法通常不考虑底层空间数据的实际分布,而是采用数据结构对空间数据进行逻辑上的划分,然后对每个分割单元进行噪音处理。HKD-Tree[12]是空间数据无关分割的早期代表,该方法利用网格分割原始数据,并为每个网格单元添加噪音,利用 KD-树进行索引,该方法只有在数据均匀分布的前提下才有效。UG[13]采用均匀网格划分二维空间数据,并为每个划分单元添加噪音。虽然 UG 能够比较合理地设定划分粒度,但是却没有考虑数据分布的偏斜和稀疏性。若某个单元过于稀疏,甚至计数为零会导致过大的噪音误差;反之一个单元过于密集,该单元划分不够彻底,会导致过大的均匀假设误差。DP-Where 方法[14]同样采用均匀网格对移动人群的工作位置与家庭位置进行分割,但该方法的不足与 UG 类似。针对 UG 方法的不足,AG[13]根据高层划分单元的粒度不同,自适应地自顶向下划分。虽然 AG 方法能够根据数据稀疏性自适应地设置空间数据的划分粒度,然而该方法却没有给出相应的启发式规则来

区分数据稠密与稀疏之间的边界。此外,该方法同样没有考虑原始数据的实际分布。文献[15]在空间数据众包应用中,采取 AG 方法进行空间分割,然后利用拉普拉斯机制保护工作者的位置信息。Quad-post[16]方法采用完全 Quad-树对二维空间数据进行自顶向下划分,完全 Quad-树需要满足所有的叶子至根路径具有相同长度,以及所有中间结点具有相同的扇出。为了提高分割精度,Quad-post 利用几何分配技术划分隐私代价,利用最小二乘法无偏估计对最终噪音响应结果进行了后置处理。该方法的优点在于能够合理地分配隐私预算,噪音误差较低;缺点在于利用树的深度控制噪音值的大小,如果树的深度比较大,则 Quad-树每层添加的噪音都非常大,致使最终的查询精度很低。此外,该方法没有考虑原始数据分布,仅对叶子结点计数添加相应噪音,均匀假设误差比较高。QuadTree[17]利用 Quad-树与卡尔曼滤波技术分割动态空间数据,该方法首先利用启发式阈值判断每个划分单元是稀疏还是密集,若密集则继续分割。QuadTree 方法的不足与 Quad-post 方法相似,均依赖树深度控制噪音值的大小。不同于 Quad-post 方法与 QuadTree 方法,PrivTree[18]方法结合完全 Quad-树划分空间数据,通过发布叶子结点噪音计数以及非叶子结点域信息,来响应范围查询。该方法完全不依赖于树深度,通过结点计数的偏移值来减少噪音,进而用一个噪音常量来判断某个结点是否分割。同时,该方法利用稀疏向量技术[19-20]计算结点的分割阈值。Boost2 方法[21]采用 b-ary 树对数据进行层次分割,利用噪音对每个结点中的计数进行扰动,然后以直方图的形式发布每层的统计信息。然而该方法也是用树深度控制噪音。H_b 方法[22]利用 b-ary 树对数据进行层次分割,该方法讨论了树的深度、树的扇出、数据维度之间的关系,并利用约束推理对查询结果进行后置处理。GS[23]方法利用抽样方法处理空间数据,然后把空间计数分成大小相同的组,并为每组的均值添加噪音,然而该方法的最终分割精度比较低。DP-Tree[24]方法利用嵌入树分割多维空间数据,并支持范围计数查询,但是该方法利用树深度控制噪音,并容易受树的扇出影响。综上,表 1 分别对上述方法进行分析对比。

表 1　基于数据无关的空间数据分割技术对比

Tab.1　Comparsion of data-independent spatial decomposition methods

方法名称	主要优点	主要缺点	计算开销	实际精度
HKD-Tree[12]	支持范围查询	分割误差大,可用性低	高;涉及额外分割操作	差;取决数据实际分布
UG[13]	支持长度范围查询	没有考虑数据的稀疏性	中;网格划分操作	一般;取决划分粒度
AG[13]	支持范围查询,均衡误差	采用启发式划分	中;网格划分操作	好;取决两层划分粒度
Quad-post[16]	支持长范围查询	利用树深度控制噪音量	高;额外后置处理操作	一般;取决后置处理
QuadTree[17]	支持范围查询	利用树深度控制噪音量	高;额外卡尔曼过滤	一般;取决卡尔曼过滤
PrivTree[18]	支持范围计数查询	无法应对大规模数据	中;额外阈值操作	好;取决噪音常量设置
Boot2[21]	支持范围计数查询	利用树深度控制噪音量	高;额外后置处理操作	一般;取决后置处理
H_b[22]	支持较长范围查询	利用树深度控制噪音量	高;额外后置处理操作	好;取决后置处理与扇出
GS[23]	支持较长范围计数查询	分割误差大	中;额外的抽样操作	差;取决于抽样效果
DP-Tree[24]	支持范围计数查询	利用树深度控制噪音量	高;数据维度影响	一般;取决后置处理

总体来看,大多数已有的方法大都受到实际的数据分布影响,通常采用树深度控制拉普拉斯噪音量,导致计算开销比较高、实际的可用性比较低。上述这些方法通常无法顾及如何均衡噪音误差与均匀假设误差。一些基于网格结构的分割方法,虽然顾及到了上述两种误差的均衡,但没有考虑到如何利用启发式规则来自适应地设置均衡参数。当空间数据达到百万级别时,上述这些方法通常无法输出精确的分割结果。此外,上述方法大都属于数据无关的方法,优势在于能够给出严谨的数据可用性理论下界。但在实际数据上,这一系列方法均无法获得理想数据可用性和效率。这一缺点引出了数据相关方法的研究。

2.2　基于数据相关的空间数据分割方法

基于数据相关的空间分割常用的索引技术是

树结构(KD-树、Hilbert-R 树、H-树、前缀树、分类树、小波树等)。这类分割方法通常要考虑底层空间数据的实际分布,根据空间数据点的实际位置进行分割。然而,这种分割必须在差分隐私保护下进行,否则会泄露底层数据的隐私。KD-SM[25]是采用 KD-树划分相关数据的早期代表,该方法采用噪音均值代替中位数来划分数据空间。通过噪音和与噪音计数的比率来获得噪音均值,然而这种方法获得的噪音中位数与实际值差别比较大。KD-Stand[16]利用指数机制[34]选择中位数,以免泄露实际的空间数据点,并利用结点中的噪音计数与给定的阈值相比较,来判断该结点是否继续分割。然而,在为树中结点添加拉普拉斯噪音时,采用 KD-树的深度控制每层的噪音量。不同于 KD-Stand 方法,KD-Hybrid[16]首先利用 Quad-树分割部分原始数据,然后再利用 KD-树分割剩余数据,其分割原理与 KD-Stand 相同,分割精度高于 KD-Stand 方法。但是,KD-Hybrid 同样采用树深度控制噪音量。HT-Stand 方法[26]采用两层 H-树分割空间数据,该方法采用较少隐私代价为每个结点添加噪音值,较多隐私代价实现中位数选择。尽管该方法只对隐私代价做两次分割,但是在寻找中位数时,要不断分割隐私代价,进而使得找出来的中位数会越来越不准确。此外,当空间数据达到百万级别时,两层 H-树很难有效索引。

除了上述的 KD-树、H-树之外,还有一些方法依赖分类树与前缀树来分割数据。而这些方法通常可以直接用于空间数据分割。DiffPart[27]采用自顶向下的方式随机地分割基于泛化技术[28]的分类树。按照分类树索引结构,树的根结点存储所有空间数据点,然后逐层向下分割。然而,DiffPart 方法仅支持计数查询,在泛化时没有考虑空间数据之间的相似关系。DiffGen 方法[29]结合指数机制与信息增益来确定决策树中的分割属性,借助于分类树自顶向下地把所有的数据点分割到叶子结点中去,然后对叶子结点中的计数值添加拉普拉斯噪音。Hybrid-Bus[30]借鉴前缀树与分类树对空间轨迹数据进行自顶向下分割,并利用前缀树本身所蕴含的固有约束,设计了一种一致性约束推理策略来增强分割精度。但是该方法却忽视了空间轨迹自身携带的时间戳,导致分割结果的可用性较低。N-Gram[31]方法通过抽取所有变长的 *n*-gram,并结合前缀树索引分割序列数据。N-Gram 方法采用马尔科夫过程来自适应地分配隐私预算。然而,该方法却人为设定 *n*-gram 的最大长度,进而造成隐私预算分割的合理性较差。类似于 N-Gram 方法,DNA-Motif 方法[32]与 M-OMC 方法[33]也是采用最大长度参数控制 DNA 序列数据与轨迹数据,这两种方法具有与 N-Gram 相同的缺陷。DPT[34]方法利用前缀树结构离散化空间轨迹数据,采用方向权重采样技术抽取空间数据,并基于抽样数据构建前缀树。然后利用顺序马尔科夫过程选择合适的且满足差分隐私的前缀树。此外,EFPAG 方法[35]采用 Voronoi 图作为索引结构来分割空间数据,然后针对每个分割单元添加相应的噪音值。综上,表 2 给出了上述方法之间的对比分析。

表 2 基于数据相关的空间数据分割技术对比

Tab.2 Comparsion of data-dependent spatial decomposition methods

方法名称	主要优点	主要缺点	计算开销	实际精度
KD-SM[25]	支持计数查询	噪音均值带来的误差大	高;额外的通信代价	差;取决数据分布
KD-Stand[16]	支持范围查询	利用树深度控制噪音量	中;树划分操作	一般;取决数据分布
KD-Hybrid[16]	支持范围计数查询	利用树深度控制噪音量	高;额外网格划分操作	好;取决网格划分粒度
HT-Stand[26]	支持范围查询	无法应对大规模数据	中;额外后置处理操作	一般;取决隐私代价分配
DiffPart[27]	支持计数查询	利用树深度控制噪音量	中;额外后置处理操作	一般;取决隐私代价分配
DiffGen[29]	支持决策树分类	无法应对连续型属性	高;额外信息增益计算	一般;取决数据分布
Hybrid-Bus[30]	支持计数查询	利用树深度控制噪音量	中;额外后置处理操作	一般;取决数据维度
N-Gram[31]	支持较范围查询	利用最大长度控制噪音	高;额外数据重构操作	好;取决最大分割长度
M-OMC[33]	支持较范围查询	利用最大长度控制噪音	中;额外后置处理操作	一般;取决最大分割长度
DPT[34]	支持较范围计数查询	利用马尔科夫选择模型	高;额外的抽样操作	好;取决数据分布

从表 2 可以看出,许多基于 KD-树的空间数据分割方法,通常依赖树的深度控制拉普拉斯噪音量,而设置合适的树深度非常具有挑战性。此外基于其它树结构的分割方法同样是利用树深度控制噪音大小。尽管上述这些研究具有重要的理论价值,但是它们距离解决实际应用中的空间数据分割问题仍然很遥远。就我们所知,目前仍没有能够支持实际应用的空间数据分割方法。

3　结论与展望

本文对差分隐私下的空间数据分割方法做了系统总结,分析了现有分割方法的技术特点与不足。总体来看,基于差分隐私的空间分割方法研究在国内还处于起步阶段,很多挑战性问题有待解决。接下来对一些挑战性问题给予展望,并给出相应的研究方法。

1) 已有的数据无关与数据相关的空间分割方法常处理小规模空间数据。然而,实际生活中的空间数据通常是大规模的与偏斜的。大规模的数据会导致基于树结构与网格结构的分割方法无法实施。偏斜的数据通常会导致树与网格结构出现大量的 0 值结点或者 0 值单元,如果直接对这些 0 值进行噪音扰动,会导致最终查询或者分析结果的可用性非常低。因此,如何分割大规模且偏斜的空间数据是个大的挑战。结合上述分析,可以利用满足差分隐私抽样技术和滤波技术获取足够多的样本,然后基于样本进行分割。

2) 已有的基于树结构的空间分割方法,通常采用树深度控制拉普拉斯噪音大小。然而,如何设置合适的树深度非常困难,如果人为直接调节树深度的大小,则调节过程会违背差分隐私,进而无法保护空间数据中的敏感信息。如果能够在为树中的结点添加噪音时不依赖树深度控制噪音,则是比较理想的选择。综上分析,可以利用稀疏向量技术来设置树中结点的分割条件,而在分割过程中,分割结点只需与稀疏向量技术产生的阈值比较即可,无需记录其噪音值。这样在计算过程中,隐私代价无需重复分配。因此,利用稀疏向量技术是解决树深度控制噪音的途径之一。

3) 已有的空间分割方法通常着眼于静态数据,而在实际应用中,空间数据通常随时间而演化。例如人群的轨迹数据。目前基于静态的空间分割方法无法直接应用动态环境。动态的空间数据分割存在两大挑战:一是如何对动态空间数据进行建模;二是如何在动态环境中分配隐私代价。针对第一个挑战,可以利用基于滑动窗的动态马尔科夫过程对动态空间数据进行建模,在每个滑动窗中分割相应的空间数据。针对第二个问题,可以考虑设计一套动态的隐私参数分配策略,结合滑动窗中空间数据之间的相似性来节省部分隐私代价。

参考文献

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

[2] Sweeney L, Abu A, Winn J. Identifying participants in the personal genome project by name: 1021-1 [R]. Harvard University Data Privacy Lab, 2013.

[3] Machanavajjhala A, Kifer D, Gehrke J, et al. L-diversity: privacy beyond k-anonymity [J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 1-47.

[4] Dwork C. Differential privacy [C] //Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, Venice, Italy, 2006: 1-12.

[5] Dwork C. Differential privacy: a survey of results [C] //Proceedings of the 5th International Conference on Theory and Applications of Models of Computation, Xi′an, China, 2008: 1-19.

[6] Dwork C, Lei J. Differential privacy and robust statistics [C] // Proceedings of the 41th Annual ACM Symposium on Theory of Computing, Bethesda, MD, USA, 2009: 371-380.

[7] Dwork C, Naor M, Reingold O, et al. On the complexity of differentially private data release: efficient algorithms and hardness results [C] //Proceedings of the 41th Annual ACM Symposium on Theory of Computing, Bethesda, MD, USA, 2009: 381-390.

[8] Dwork C. The differential privacy frontier (extended abstract) [C] //Proceedings of the 6th Theory of Cryptography Conference, San Francisco, CA, USA, 2009: 496-502.

[9] 张啸剑, 孟小峰: 面向数据发布和分析的差分隐私保护 [J]. 计算机学报, 2014, 37(4): 927-949.

[10] Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis [C] //Proceedings of the 3th Theory of Cryptography Conference, New York, USA, 2006: 363-385.

[11] McSherry F, Talwar K. Mechanism design via differential privacy [C] //Proc of the 48th Annual IEEE Symposium on Foundations of Computer Science, RI, USA, 2007: 94-103.

[12] Xiao Y, Xiong L, Yuan C. Differentially private data release

through multidimensional partitioning [C] //Proceedings of 7th VLDB Workshop on Secure Data Management, Singapore, 2010: 150-168.

[13] Qardaji W H, Yang W, Li N. Differentially private grids for geospatial data [C] //Proceedings of IEEE 29th International Conference on Data Engineering, Brisbane, Australia, 2013: 757-768.

[14] Mir D J, Isaacman S, Caceres R, et al. DP-Where: differentially private modeling of human mobility [C] //Proceedings of the 2013 IEEE International Conference on Big Data, 2013: 580-588.

[15] To H, Ghinita G, Shahabi C. A framework for protecting worker location privacy in spatial crowdsourcing [C] //Proceedings of the 40th Conference of Very Large Databases, 2014: 919-930.

[16] Cormode G, Procopiuc C M, Srivastava D, et al.. Differentially private spatial decompositions [C] //Proceedings of IEEE 28th International Conference on Data Engineering, Washington, DC, USA, 2012: 20-31.

[17] Fan L, Bonomi L, Xiong L, et al. Monitoring web browsing behavior with differential privacy [C] //Proceedings of the 23th Internation World Wide Web Conference, Seoul, Republic of Korea, 2014: 177-188.

[18] Zhang J, Xiao X, Xie X. PrivTree: a differentially private algorithm for hierarchical decompositions [C] //Proceedings of the 36th ACM International Conference on Management of Data, San Francisco, CA, USA, 2016: 155-170.

[19] Lee J, Clifton C W. Top-k frequent itemsets via differentially private fp-trees [C] //Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2014: 931-940.

[20] Chen R, Xiao Q, Zhang Y, et al. Differentially private high-dimensional data publication via sampling-based inference [C] // Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 2015: 129-138.

[21] Hay M, Rastogi V, Miklau G, et al. Boosting the accuracy of differentially private histograms through consistency [J]. Proceedings of the VLDB Endowment, 2010, 3(1): 1021-1032.

[22] Qardaji W, Yang W, Li N. Understanding hierarchical for differentially private histograms [J]. Proceedings of the VLDB Endowment, 2013, 6(1): 1021-1032.

[23] Kellaris G, Papadopoulos S. Practical differential privacy via grouping and smoothing [J]. Proceedings of the VLDB Endowment, 2013, 6(1): 301-312.

[24] Peng S, Yang Y, Zhang Z, et al. DP-tree: indexing multi-dimensional data under differential privacy [C] //Proceedings of the 32th ACM Int Conf on Management of Data, Scottsdale, AZ, USA, 2012: 864.

[25] Inan A, Kantarcioglu M, Ghinita G, et al. Private record matching using differential privacy [C] //Proceedings of the 13th International Conference on Extending Database Technology, Lausanne, Switzerland, 2010: 123-134.

[26] To H, Fan L, Shahabi C. Differentially private h-tree [C] // Proceedings of the 2th Workshop on Privacy in Geographic Information Collection and Analysis In Conjunction with ACM Sigspatial 2015, Seattle, Washington, USA, 2015: 95-82.

[27] Chen R, Fung B C M, Mohammed N, et al. Publishing set-valued data via differential privacy [J]. Proceedings of the VLDB Endowment, 2011, 4(11): 1087-1098.

[28] 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.

[29] Mohammed N, Chen R, Fung B C M, et al. Differentially private data release for data mining [C] //Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 2011: 493-501.

[30] Chen R, Fung B C M, Desai B C, et al. Differentially private transit data publication: a case study on the Montreal transportation system [C] //Proceedings of the 18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Beijing, China, 2012: 213-221.

[31] Chen R, Acs G, Castelluccia C. Differentially private sequential data publication via variable-length n-grams [C] //Proceedings of the 19th ACM Conference on Computer and Communications Security, Raleigh, NC, USA, 2012: 638-649.

[32] Chen R, Peng Y, Choi B, et al. A private DNA motif finding algorithm [J]. Journal of Biomedical Informatics, 2014, 50: 122-132.

[33] Zhang J, Ghinita G, Chow C. Differentially private location recommendations in geosocial networks [C] //Proceedings of the 15th International Conference on Mobile Data Management, Brisbane, Australia, 2014: 1154-1165.

[34] He X, Cormode G, Machanavajjhala A. DPT: differentially private trajectory synthesis using hierarchical reference systems [J]. Proceedings of the VLDB Endowment, 2015, 7(1): 1154-1165.

[35] Acs G, Castelluccia C. A case study: private preserving release of spatio-temporal density in paris [C] //Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, USA, 2014: 1679-1688.

Survey on spatial data decomposition with differential privacy

PENG Hui-li[1, 2], ZHANG Xiao-jian[1]

(1. College of Computer & Information Engineering, Henan University of Economics and Law, Zhengzhou, Henan 450046, China;
2. Henan Radio & Television University, Zhengzhou, Henan 450008, China)

Abstract: As the emergence of applications over spatial data, a big challenge to those applicatons is to protect private location information from disclosure. A survey of spatial decomposition under differential privacy is presented, which is important to location-based service, traffic management and personalized recommendation. The basic theory of differential privacy is firstly introduced. And then the state-of-the-art solutions of spatial decomposition are summarized and compared in detail. Finally, some future research directions are discussed.

Key words: privacy protection; spatial decomposition; differential privacy; data-independent decomposition; data-dependent decomposition

(上接第 245 页)

Numerical Simulation and experimental study on gas-liquid two-phase flow of curved-plate dehydration equipment

XIAO Li-chun, TI Xiao-yu

(School of Environmental and Chemical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract: The complex problem on two-phase flow field in the curved plate dehydration equipment is presented in the paper. The influence of various factors on the limit speed and dehydration effect were discussed by the simulation of droplets′ mechanical behavior and the gas-liquid two-phase flow field in the curved plate dehydration equipment when the structure parameters and operating conditions were changed. The results show that the dehydration equipment has the highest dehydration efficiency and minimum resistance loss when the distance between the plate is 15 mm, the droplets′ average diameter is 160 μm, and the quantity of spraying water is 0.89 m^3/h. At the moment the limit wind speed is 3.7 m/s. The numerical simulation results agree well with the experimental results. It provides a guide for the industrial design of the curved plate dehydration equipment.

Key words: curved plate dehydration equipment; two-phase flow; dehydration efficiency; numerical simulation; limit wind speed