基于差分隐私的活动模式保护与时空轨迹发布方法

曾 卓,汪成亮,马 飞
(重庆大学计算机学院,重庆 400044)

摘 要: 为了解决用户轨迹数据发布时的活动模式泄露问题,本文提出了一种基于差分隐私的活动模式保护与时空数据发布方法 DPAP-STTP(Differentially Private Activity Pattern and Spatial-Temporal Trajectory Publication),该方法即保护了用户时空数据中活动模式的隐私,又可以保证所发布时空轨迹在服务建议生成上的有效性. 在 DPAP-STTP 中,用户的活动模式表示为个人代表性轨迹的动静态信息,包括代表性轨迹的时空密度分布、时空路径分布、移动模式以及时空跨度. 另外,DPAP-STTP 通过隐私保护预算与隐私保护阈值对该动静态信息进行调控,然后根据调控后的动静态信息依次划分时空网格、重构轨迹所处时空区间、时空轨迹点随机采样,最终生成满足群体差分隐私的时空轨迹进行发布. 本文的实验比较了 DPAP-STTP 与 DP-STAR(Differential Private Synthetic Trajectory Publisher)、BNA(Bounded Noise-Adding)所生成的轨迹在特定时空范围内的有效性,证明 DPAP-STTP 不但可重构服从群体差分隐私的时空轨迹,而且在时空网格上维持了时空轨迹的有效性.

关键词: 活动模式;群体差分隐私;时空轨迹;动静态信息

基金项目: 国家自然科学基金(No.61672115);重庆市技术创新与应用发展专项重大主题专项(No.cstc2020jscx-dxwtBX0055)

中图分类号: TP309.2 **文献标识码:** A **文章编号:** 0372-2112(2023)03-0552-12

电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210631

Differentially Private Activity Pattern and Spatial-Temporal Trajectory Publication

ZENG Zhuo, WANG Cheng-liang, MA Fei
(*Computer School*, *Chongqing University*, *Chongqing* 400044, *China*)

Abstract: In order to solve activity pattern leakage problems while user trajectory data publishing, the paper proposes the DPAP-STTP (Differentially Private Activity Pattern and Spatial-Temporal Trajectory Publication) method to publish spatial-temporal trajectories for achieving required services suggestions in support of users while preserving the privacy of activity patterns. In DPAP-STTP, users' activity patterns are represented as dynamic and static information of personal representative trajectories, including spatial-temporal density distribution, spatial-temporal trip distribution, mobility pattern and spatial-temporal span. Additionally, according to allocated privacy budget and specific privacy-preserving threshold, DPAP-STTP preserves the privacy of dynamic and static information, and uses perturbed information to divide spatial-temporal grids, reconstruct spatial-temporal passing grids, randomly select spatial-temporal point, and finally generate spatial-temporal trajectories with group differential privacy satisfied. The experiment in this paper compares the DPAP-STTP with DP-STAR (Differential Private Synthetic Trajectory Publisher) and BNA (Bounded Noise-Adding) for presenting the utility of DPAP-STTP trajectories. Consequently, the DPAP-STTP method is proved to generate spatial-temporal trajectories which follow the group differential privacy and maintain their utility in some spatial-temporal scopes.

Key words: activity pattern; group differential privacy; spatial-temporal trajectories; dynamic and static information

Foundation Item(s): National Natural Science Foundation of China (No.61672115); Chongqing Technology Innovation & Application Development Key Project (No.cstc2020jscx-dxwtBX0055)

收稿日期:2021-05-18;修回日期:2022-04-16;责任编辑:梅志强

1　引言

随着5G技术的发展,智慧城市中的用户们可以在信号覆盖的任何区域从具有高速信息处理能力和大规模专家知识的第三方服务提供商SPs(Service Providers)获取各种实时服务(例如定位服务,医疗健康服务).这些第三方服务提供商(如一些行政机构,医疗机构,信息技术公司或研究所)不仅能够直接为用户们提供服务,还可以产生并反馈服务建议来指导如何基于用户们的活动模式(即用户在特定时空范围下的移动规律)来提供服务.例如,交通机构可以基于城市市民的活动模式合理规划车流与人流来减轻人群拥挤,并且降低市民的出行成本[1].

但是,普适计算的盛行使得智慧城市中部署的移动或固定IoT(Internet of Things)传感器增加,需要实时服务的居民的日常生活更有可能受到监视.SPs不仅可以收集或分析用户活动轨迹来提供服务建议,还能通过轨迹侵犯用户的隐私.当前主流的隐私保护方法主要可分为两方面:

(1) 身份隐私保护,例如k-匿名,l-多样性,t-贴近性等防止用户身份信息泄露的匿名技术,可用于防止以用户的活动模式作为准标识符来识别用户身份,然而却难以防止背景知识的攻击.

(2) 数据隐私保护,例如ε-差分隐私,可用于防止适应性查询所带来的用户活动数据记录的泄露.虽然克服了基于背景知识的攻击并保护活动数据隐私,然而即便保护了原始活动数据的隐私,半可信的SPs依旧可以利用用户的活动模式来侵犯隐私.另外,传统的活动数据隐私保护方法主要用于保护原始数据中的空间信息,难以防止带有时空信息的活动模式泄露.

如果SPs从用户的活动轨迹中获取了用户的活动模式,则可通过三方面侵犯用户的隐私:(1)获取用户身份,将用户的活动模式作为准标识符来重识别用户的身份以及活动[2].(2)预测未来活动,将轨迹聚类的移动模式用做不确定性轨迹预测[3],可预测用户未来的移动轨迹.(3)关联用户活动习惯,活动模式可以用来探索用户活动的时间规律[4].故而时空轨迹发布的同时也需要维持活动模式隐私以防止轨迹聚类的时空规律泄露.

为了在活动数据发布的同时防止活动模式的泄露,本文并未采用传统的且只面向空间信息的原始轨迹隐私保护方法,而着重于对活动模式的保护.用户的活动模式可以由时空轨迹聚类而来的带计数的代表性轨迹集RTS(Representative Trajectories Set)表示,所以本文通过保护带轨迹计数的RTS的隐私来实现用户活动模式隐私保护.其主要贡献包括:

(1) 将ε/h-群体差分隐私与隐私保护阈值h:$h\geq c$用于保护RTS.该方式既保护了轨迹计数为c的时空轨迹簇的隐私,又保护了轨迹计数大于h的轨迹簇中任意h条时空轨迹的隐私不被泄露.

(2) 本文结合ε/h-群体差分隐私保护的活动模式、轨迹的时空特性和DP-Star[5]轨迹生成体系结合,提出了算法DPAP-STTP以生成满足群体差分隐私的时空轨迹(见图1),即保护了RTS的隐私,也维持了其在部分时空区间上的有效性.

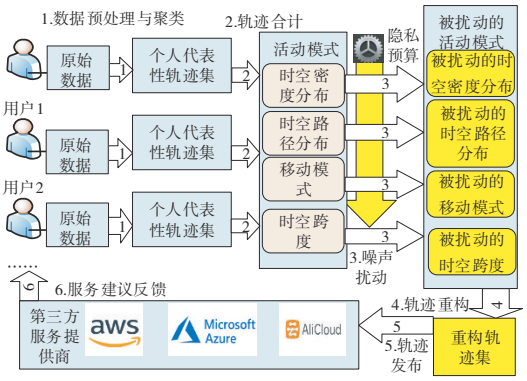


图1　基于差分隐私的活动模式保护与时空轨迹发布

2　相关研究

本研究源自于对智慧城市下服务建议反馈的探究,即使SPs未获取原始数据,也可以面向活动模式提供服务建议.例如,SPs可将服务建议通过IF-THEN规则(前件为活动模式,后件为服务建议)反馈给本地服务器,本地服务器可将规则置入以Rete算法为核心的推理引擎并进行推理[6].当用户的当前活动满足规则中的活动模式时,本地服务区便基于规则指导本地智能设备来服务用户.由于活动模式的泄漏有害于人的隐私,所以需要考虑活动模式的隐私保护.活动模式的隐私保护可以通过保护轨迹聚类的隐私来实现,所以活动模式隐私保护相关工作主要分为三部分,即时空轨迹聚类与轨迹的隐私保护,以及面向轨迹聚类的隐私保护.

2.1　时空轨迹聚类

用户的活动模式可从相似时空轨迹聚类形成的轨迹簇获得[7].目前许多有关时空轨迹聚类的相关著作.经典的轨迹聚类算法traclus[8]基于最小描述长度MDL(Minimum Description Length)将轨迹划分为轨迹片段,然后通过DBSCAN(Density-based Spatial Clustering of Applications with Noise)算法聚类轨迹子段,该算法只考虑了轨迹在空间上的聚类.

T-DBSCAN(Time-sequential DBSCAN)[9]通过解决轨迹中多个停留点和非连续点序列的问题,实现了连续轨迹片段的聚类.ST-DBSCAN(Spatial-temporal DBSCAN)[10]可以根据时空数据[11]间的时间距离阈值与空间距离阈值实现聚类.时空轨迹聚类的核心是度

量任意两个轨迹之间的时空距离，TDSP（Time-Dependent Shortest Path）[12]与轨迹间隔距离估计法[13]被提出来测量轨迹间的时空距离．按照时空距离度量标准搜索轨迹簇中与其它轨迹的平均距离最小的中心轨迹，中心轨迹可用作轨迹簇的代表性轨迹来代表用户的活动模式．按照轨迹簇中心进行分区[14]的聚类方法可以根据相似性指标找出k个中心轨迹来代表轨迹簇．CBSCAN（Cell-based Spatial Clustering of Applications with Noise）[15]方法将基于数据点密度的聚类扩展为基于cell的密度聚类，可减少高密度区域中轨迹间的距离计算．

2.2 轨迹的隐私保护

就轨迹的隐私保护工作而言，群体差分隐私可用于保护轨迹数据集D，使得任意两个仅相差c条轨迹的子轨迹集$D_1 \subset D$和$D_2 \subset D$满足[16]：

$$\Pr\left(A\left(D_1\right) \in S\right) \leq \mathrm{e}^{c\varepsilon} \times \Pr\left(A\left(D_2\right) \in S\right)$$

其中A是一个随机算法，S是A的像的所有子集，ε为隐私预算．

差分隐私通过拉普拉斯机制与指数机制实现．对于任意的函数$f: D \rightarrow R^d$，如果有$A(D)=f(D)+\mathrm{Lap}\left(\Delta f \times c/\varepsilon\right)$，则$A$满足$\varepsilon/c$-群体差分隐私．$D$中每条数据都被$\varepsilon/c$-差分隐私保护，其中敏感度$\Delta f=\max\limits_{D_1, D_2 \subseteq D}\left|f\left(D_1\right)-f\left(D_2\right)\right|$．对于任意的函数$u:(D \times \tau) \rightarrow R$，如果一个指数机制输出$r \in \tau$的概率服从$\Pr\left(u(d)=r\right) \sim \mathrm{e}^{\frac{\varepsilon u(d, r)}{2c}}$，则该机制满足$\varepsilon/c$-群体差分隐私．He[17]基于拉普拉斯机制和指数机制使用分层参考系统HRS（Hierarchical Reference Systems）生成差分隐私轨迹，Gursoy[5]通过差分隐私调控空间上轨迹的统计信息再重构轨迹，即保护了轨迹的隐私又保持了轨迹的有效性．Terrovitis[18]将位置抑制与轨迹分割相结合来防止轨迹隐私泄漏，同时维持轨迹的统计值．霍峥[19]通过噪声R-树的轨迹数据发布方法，按层次分割隐私预算，并对移动对象在自由空间的计数值添加噪声．这些对轨迹隐私保护的研究主要关注轨迹本身的隐私，而忽略了对于轨迹聚类隐私的保护．

2.3 面向轨迹聚类的隐私保护

目前，面向轨迹聚类的隐私保护的工作主要有两种，一种是通过改变轨迹聚类来实现对轨迹的保护．赵晓东[20]将差分隐私用于轨迹的聚类来实现轨迹的隐私保护，通过在轨迹地点与轨迹聚类中心添加半径约束的拉普拉斯噪声以防御对轨迹的连续查询攻击．Zhang[21]提出了多划分差分隐私的k中心点聚类算法，通过优化初始中心点选择以及为聚类过程添加噪声的方式，既保护了数据又维护了数据聚类的有效性．

另一种是研究对轨迹聚类的保护，Ma[22]提出随机采样的差分隐私方法，在不增加信息损失的前提为轨迹聚类中的轨迹计数添加噪声．Wang[23]提出聚类不可区分法，在聚类算法中加入服从特定概率密度函数的拉普拉斯噪声，实现对轨迹聚类的隐私保护．Liu[24]提出了Bounded Noise-adding算法可扰动轨迹节点聚类的中心点以及该聚类中轨迹的数量．然而这些面向轨迹聚类的隐私保护方法主要通过扰动轨迹的空间位置与数量来保护轨迹聚类，没有结合轨迹聚类的时空信息实现对于时空轨迹聚类的隐私保护．

3 活动模式

本章通过带计数的代表性轨迹集RTS（Representative Trajectories Set）定义了活动模式，活动模式描述了用户在时空范围内各种频繁移动的规律，它由时空轨迹聚类得来的个人代表性轨迹prts（Personal Representative Trajectories）与该聚类的轨迹计数表示（文中参数的定义可见表1）．

表1 参数的定义

参数	定义
prt	个人代表性轨迹
RTS	带计数的代表性轨迹集
$\left(\hat{O}, \hat{T}^m\right)$	适应性时空网格
$\left(\mathrm{eps}_1, \mathrm{eps}_2\right)$	轨迹聚类的时空距离阈值
h	隐私保护阈值
ε	差分隐私预算
$(\mathrm{ss}, \mathrm{ts})$	时空跨度
trajs $\left(N_1, N_2\right),(M, S)$	DPAP-STTP时空轨迹 两级时空粒度大小

表1中，每条代表性轨迹描述了该用户某种频繁移动的时空规律．

轨迹：用户群体U中任意一个人$u_i \in U$的轨迹数据集$D\left(u_i\right)$由带有时间戳的位置点序列$p^j=x^j, y^j, t^j \in D\left(u_i\right)$组成．

时间周期：轨迹的时间戳t^j可以映射到一个时间周期$T=\left[T^{\mathrm{start}}, T^{\mathrm{end}}\right]$形成相对时间，用以衡量轨迹之间的时间相似度．例如，时间戳2020-01-02-23:50:42被映射到$23:50:42 \in [00:00:00, 23:59:59]$．另外时间周期$T$可以进一步划分成$M$个$\left\{T^m \mid 1 \leq m \leq M\right\}$子周期．

个人代表性轨迹prts：用来代表时空轨迹聚类形成的轨迹簇的中心轨迹．轨迹数据集$D\left(u_i\right)$通过轨迹预处理生成时间范围T^m内可聚类的轨迹片段集$\mathrm{Ts}\left(u_i, T^m\right)$．例如MDL（Minimum Description Length）法可将每条轨迹划分成多个轨迹片段．然后基于轨迹片段之间的时

空相似度以及时空聚类距离阈值$(\mathrm{eps}_1,\mathrm{eps}_2)$[11]，DBSCAN，ST-DBSCAN[10]或K-medoids[21]算法等可将$\mathrm{Ts}(u_i,T^m)$聚类生成多个轨迹簇$\mathrm{RTS}(u_i,T^m)$. 簇中与其它轨迹片段平均距离最小的轨迹片段是轨迹簇的中心轨迹可作为个人代表性轨迹，每条代表性轨迹用来代表性具有条轨迹片段的簇，见图2.

$$\mathrm{Ts}(u_i,T^m)\xrightarrow{\text{聚类}}\mathrm{RTS}(u_i,T^m) \tag{1}$$

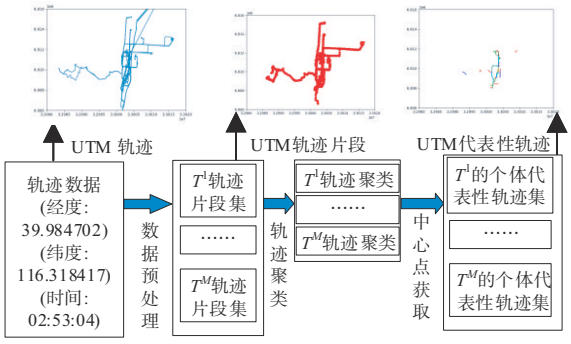


图2　代表性轨迹集的生成过程

代表性轨迹集RTS：用户u_i在T^m的代表性轨迹集$\mathrm{RTS}(u_i,T^m)$具有k_i^m条带计数的代表性轨迹$\{\mathrm{prt}(u_i,T^m,v)|1\le v\le k_i^m\}$，每条代表性轨迹$\mathrm{prt}(u_i,T^m,v)$用于代表轨迹数为$c_i^{mv}$的轨迹簇. 在本文中，用户群体$U$在任意时间$t\in T$的代表性轨迹集通过RTS=$\mathrm{RTS}(U,T)$表示.

$$\mathrm{prt}(u_i,T^m,v)=\{\bar{p}^1,\bar{p}^2,\cdots,\bar{p}^{\mathrm{length}}\} \tag{2}$$

$$\mathrm{RTS}(u_i,T^m)=\{\mathrm{prt}(u_i,T^m,v)c_i^{mv}|1\le v\le k_i^m\} \tag{3}$$

4　基于差分隐私的活动模式保护

因为活动模式的泄露会影响用户隐私，所以代表活动模式的RTS是不能直接泄露的. 但是为了获取所需的服务，用户可授权第三方服务提供商查询RTS的动静态信息：(1)时空密度分布STD(Spatial-Temporal density Distribution)，(2)时空路径分布STTD(Spatial-Temporal Trip Distribution)，(3)移动模式MP(Mobility Pattern)，(4)时空跨度STS(Spatial-Temporal Span). 活动模式隐私保护的目的是通过群体差分隐私扰动面向$\mathrm{RTS}(u_i,T^m)$动静态信息的查询，防止轨迹数量$c\le h$的轨迹簇或者轨迹簇的任意h条轨迹不被泄露，同时保证基于动静态信息获取的$\mathrm{RTS}(u_i,T^m)$的有效性.

就$\mathrm{RTS}(u_i,T^m)$而言，任意$c\le h$或h条轨迹的隐私保护程度由隐私总预算$\varepsilon=\sum_{i=1}^{4}\varepsilon_i$与隐私保护阈值$h$决定. 隐私总预算决定了RTS动静态信息的最大公开程度. 第三方服务提供商访问四部分动静态信息的时候，隐私保护机制需要消耗隐私预算来扰动访问结果. ε值越小则动静态信息公开程度就越低，其隐私保护效果就越好. 但是当隐私保护机制中剩余的隐私预算$\varepsilon=0$，第三方服务提供商就不能获取动静态信息. 目前，基于差分隐私的轨迹保护方法[5,17,22~25]通常设定隐私预算$\varepsilon=\{0.5,1,2\}$验证轨迹的隐私保护效果与生成轨迹有效性.

另外，隐私保护阈值h也影响了$\mathrm{RTS}(u_i,T^m)$隐私与有效性的权衡，h越大，每条代表性轨迹受ε/h-差分隐私保护的轨迹所分到的隐私保护预算就越小，则扰动每条代表性轨迹的噪声$\mathrm{Lap}(\Delta f\times c/\varepsilon)$也越大，所以隐私保护效果也越好，受扰动的代表性轨迹的有效性会越低. 在实际应用中，h值可以设置为任意正整数，它决定了轨迹计数不大于h的轨迹集的隐私不被泄露. h值设置得越大可保护越高轨迹计数的轨迹集.

4.1　被扰动的时空密度分布的获取

RTS的时空密度分布$\mathrm{STD}(\mathrm{RTS},T^m,O)$通过统计时空网格上任意网格的轨迹点数量得到. 根据统一网格划分法，轨迹所处的空间范围可划分成网格空间$O=\{O^{r_1}|1\le r_1\le N_1\times N_1\}$，时间范围划分为$T=\{T^m|1\le m\le M\}$；$T^m=\{t^{ms}|1\le s\le S\}$. 在时空范围$(T^m,O)$中，$\mathrm{RTS}(U,T^m)$的时空密度分布$\mathrm{STD}(\mathrm{RTS},T^m,O)$可表示为：

$$\mathrm{STD}(\mathrm{RTS},T^m,O)=\gamma(\mathrm{RTS},T^m,O^{r_1})|O^{r_1}\in O \tag{4}$$

$$\gamma(\mathrm{RTS},T^m,O^{r_1})=\sum_{(\mathrm{prt},c)\in\mathrm{RTS}}\frac{\sum_{\bar{p}\in\mathrm{prt}\wedge\bar{p}\,\mathrm{in}\,(T^m,O^{r_1})}c}{|\mathrm{prt}|} \tag{5}$$

其中，$\gamma(\mathrm{RTS},T^m,O^{r_1})$统计了$(T^m,O^{r_1})$的标准化轨迹点数量. 例如，当只有一条轨迹时空点数量为5的代表性轨迹的三个时空点位于(T^m,O^{r_1})，则其$\gamma(\mathrm{RTS},T^m,O^{r_1})=0.6$.

为了在更细的时空粒度上获取RTS的时空密度分布，需要根据$\gamma(\mathrm{RTS},T^m,O^{r_1})$划分适应性网格. 轨迹点数量越多，则该时空网格划分得就越细. 时间范围被进一步划分为$\hat{T}^m=\{t^{ms}|1\le s\le S\}$相同时长的时间间隔. 而空间范围需要基于$\gamma(\mathrm{RTS},T^m,O^{r_1})$进行划分.

然而如果直接根据$\mathrm{STD}(\mathrm{RTS},T^m,O)$划分网格，SPs可通过获取网格大小来推断轨迹点数量$\gamma(\mathrm{RTS},T^m,O^{r_1})$，并进一步对$\mathrm{STD}(\mathrm{RTS},T^m,O)$进行推断. 故而需要先扰动$\mathrm{STD}(\mathrm{RTS},T^m,O)$，然后再进一步按照$\gamma(\mathrm{RTS},T^m,O^{r_1})$适应性地划分空间范围$O^{r_1}=\{o^{r_2}|1\le r_2\le N_2\times N_2\}$，划分后的空间范围通过$o^{r_2}\in\hat{O}$表示：

$$\hat{\gamma}_{T^m,O^{r_1}}=\sum_{(\mathrm{prt},c)\in\mathrm{RTS}}\gamma((\mathrm{prt},c),T^m,O^{r_1})+\mathrm{Lap}(h/\varepsilon_1) \tag{6}$$

$$N_2=\max\left(\min N_2,\sqrt{\beta\times\hat{\gamma}_{T^m,O^{r_1}}}\right) \tag{7}$$

设置N_2的最小值$\min N_2$是为了防止N_2值过小从而导致划分后的网格空间粒度过大.

$\hat{\gamma}_{T^m,O^{r_1}}$在原本的标准化轨迹点数量上添加了噪声$\mathrm{Lap}\left(h/\varepsilon_1\right)$,意味着$\mathrm{STD}\left(\mathrm{RTS},T^m,O\right)$的敏感度$\Delta\mathrm{STD}=h$. 对于任意两个仅相差轨迹计数为$c$的代表性轨迹集$\mathrm{RTS}_1\subset\mathrm{RTS}$与$\mathrm{RTS}_2=\mathrm{RTS}_1\cup\left(\mathrm{prt},c\le h\right)$,可证$\Delta\mathrm{STD}=h$:

$$\begin{aligned}\Delta\mathrm{STD}&=\max_{\mathrm{RTS}_1,\mathrm{RTS}_2}\left\{\mathrm{STD}\left(\mathrm{RTS}_1,T^m,O\right)-\mathrm{STD}\left(\mathrm{RTS}_2,T^m,O\right)\right\}\\&=\max_{\mathrm{RTS}_1,\mathrm{RTS}_2}\sum_{O^{r_1}\in O}\gamma\left(\left(\mathrm{prt},c\right),T^m,O^{r_1}\right)\\&=h\end{aligned} \tag{8}$$

所以,适应性网格划分服从ε_1/h-差分隐私. 而且带计数的代表性轨迹$\left(\mathrm{prt},c\le h\right)$服从$\varepsilon_1$-差分隐私,轨迹簇中任意一条轨迹$\left(\mathrm{prt},1\right)$服从$\varepsilon_1/h$-差分隐私.

4.2 被扰动的时空路径分布的获取

$$\begin{aligned}&\mathrm{STTD}\left(\mathrm{RTS},\hat{O},\hat{O},\hat{T}^m\right)\\&=\left\{\frac{\theta\left(\mathrm{RTS}_{\left(o^x,t^a\right)\to\left(o^y,*\right)}\right)}{\theta\left(\mathrm{RTS}\right)}\middle|\,o^x\times o^y\times t^a\in\hat{O}\times\hat{O}\times\hat{T}^m\right\}\end{aligned} \tag{9}$$

时空路径分布$\mathrm{STTD}\left(\mathrm{RTS},\hat{O},\hat{O},\hat{T}^m\right)$描述了RTS在$\left(\hat{O},\hat{T}^m\right)$上的时空路径分布,即轨迹具有起点$\left(o^x,t^a\right)$与终点$\left(o^y,*\right)$的概率. 因为$\theta\left(\mathrm{RTS}\right)$计算了RTS的所有轨迹数量,轨迹进行发布的时候SPs能够知道$\theta\left(\mathrm{RTS}\right)$,所以对时空路径分布的调控需要对具有起点$\left(o^x,t^a\right)$与终点$\left(o^y,*\right)$的轨迹数量$\theta\left(\mathrm{RTS}_{\left(o^x,t^a\right)\to\left(o^y,*\right)}\right)$进行调控,即添加预算为$\varepsilon_2$,敏感度为$h$的噪声,以生成被扰动的时空路径分布:

$$\frac{\theta\left(\mathrm{RTS}_{\left(o^x,t^a\right)\to\left(o^y,*\right)}\right)+\mathrm{Lap}\left(h/\varepsilon_2\right)}{\theta\left(\mathrm{RTS}\right)} \tag{10}$$

对于任意$\mathrm{RTS}_1\subset\mathrm{RTS}$与$\mathrm{RTS}_2=\mathrm{RTS}_1\cup\left(\mathrm{prt},c\le h\right)$,仍可证$\Delta\mathrm{STTD}=h$:

$$\begin{aligned}&\Delta\mathrm{STTD}\\&=\max_{\mathrm{RTS}_1,\mathrm{RTS}_2}\left\{\left|\theta\left(\mathrm{RTS}_{1\left(o^x,t^a\right)\to\left(o^y,*\right)}\right)-\theta\left(\mathrm{RTS}_{2\left(o^x,t^a\right)\to\left(o^y,*\right)}\right)\right|\right\}\\&=h\end{aligned} \tag{11}$$

所以,被扰动的时空路径分布提取服从ε_2/h-差分隐私.

4.3 被扰动的移动模式的获取

通过轨迹中任意具有两个连续且相同时间间隔时空点$\left(o^x,o^y\right)$的数量$\mathrm{MC}\left(\mathrm{RTS},o^x,o^y\right)$可获得RTS中的移动模式$\mathrm{MP}\left(\mathrm{RTS}\right)$,它计算了轨迹在两个空间$\left(o^x,o^y\right)$的转移概率:

$$\mathrm{MP}\left(\mathrm{RTS}\right)=\frac{\mathrm{MC}\left(\mathrm{RTS},o^x,o^y\right)}{\sum_{o\in\hat{O}}\mathrm{MC}\left(\mathrm{RTS},o^x,o\right)} \tag{12}$$

已知起点位置o^x的轨迹总量,则对$\mathrm{MP}\left(\mathrm{RTS}\right)$的扰动需要通过为始终点为$\left(o^x,o^y\right)$的轨迹数量$\mathrm{MC}\left(\mathrm{RTS},o^x,o^y\right)$添加敏感度为$h$且隐私预算$\varepsilon_3$为的噪声$\mathrm{Lap}\left(h/\varepsilon_3\right)$:

$$\mathrm{MC}'\left(\mathrm{RTS},o^x,o^y\right)=\mathrm{MC}\left(\mathrm{RTS},o^x,o^y\right)+\mathrm{Lap}\left(h/\varepsilon_3\right) \tag{13}$$

以下可证,敏感度:

$$\begin{aligned}\Delta\mathrm{MC}&=\max_{\left(\mathrm{prt},c\right)\in\mathrm{RTS}}\left\{\sum_{o^x\in\hat{O}}\sum_{o^y\in\hat{O}}\mathrm{MC}\left(\left(\mathrm{prt},c\right),o^x,o^y\right)\right\}\\&=\max_{\left(\mathrm{prt},c\right)\in\mathrm{RTS}}\left\{c\right\}=h\end{aligned} \tag{14}$$

所以获取被扰动的移动模式服从于ε_3/h-差分隐私.

4.4 被扰动的时空跨度的获取

RTS的时空跨度统计具有始终位置$\left(o^x,o^y\right)$的轨迹的时空跨度,它通过递增集合$\mathrm{STS}_{o^x\to o^y}\left\{\mathrm{RTS}\right\}=\left\{\left(\mathrm{ss}_j,\mathrm{ts}_j\right)\middle|1\le j\le\delta\right\}$来表示.

空间跨度ss_j,RTS中$\left(\mathrm{prt},c\right)$的空间跨度计算了轨迹通过了几个空间网格.

时间跨度ts_j,RTS中$\left(\mathrm{prt},c\right)$的时间跨度测量了该轨迹起点到终点的时间跨度. 当轨迹被拟合成任意两个连续节点具有相同时间间隔时,则轨迹的节点数即可用于表示轨迹的时间跨度.

集合的序列,通过集合中时空跨度的秩$\sqrt{w_s\left(\mathrm{ss}_j\right)^2+w_t\left(\mathrm{ts}_j\right)^2}$来决定集合的顺序,$w_s$为空间跨度权重,$w_t$为时间跨度权重.

将轨迹点所在的空间网格$o^{r_2}\in\hat{O}$映射为一阶Markov链的状态$l\in\Omega$(Ω为状态空间). 一条$\mathrm{prt}=\left\{\left(l^i,t^i\right)\middle|1\le i\le\varsigma\right\}$会被转化为$\varsigma$个时空状态与$\varsigma-1$个状态转移. 状态转移可分为两种:

停留状态 $\mathrm{stay}=\left\{\left(l^i,t^i\right),\left(l^{i+1},t^{i+1}\right)\right\}\subseteq\mathrm{prt}$具有关系$l^i=l^{i+1}$,意味着两个状态都由同一个空间网格$o^x$映射而来.

移动状态 $\mathrm{move}=\left\{\left(l^i,t^i\right),\left(l^{i+1},t^{i+1}\right)\right\}\subseteq\mathrm{prt}$具有关系$l^i\ne l^{i+1}$,意味着两个状态由不同的空间网格$o^x$与$o^{y\ne x}$映射而来.

所以$\mathrm{ts}_j=\mathrm{num}\left(\mathrm{stay}\right)+\mathrm{num}\left(\mathrm{move}\right)+1$,其中,$\mathrm{num}\left(\mathrm{stay}\right)$

为停留状态数量,$\mathrm{ss}_j=\mathrm{num}(\mathrm{move})+1$.

以 $\mathrm{STS}_{o^x\to o^y}\{\mathrm{RTS}\}$ 中值作为 RTS 的时空跨度估计并对输出中值的概率进行扰动. 时空跨度估计满足 ε_4/h-差分隐私当且仅当输出其中 $(\mathrm{ss}_j,\mathrm{ts}_j)$ 的概率满足:

$$\Pr\left(\psi\left[\mathrm{STS}_{o^x\to o^y}\{\mathrm{RTS}\}\right]=(\mathrm{ss}_j,\mathrm{ts}_j)\right)\propto\exp\left(-\frac{\varepsilon_4}{h}\left|j-(\delta+1)/2\right|\right)\quad(15)$$

5　基于差分隐私的时空轨迹发布

所发布的时空轨迹由扰动后的活动模式重构. 就轨迹重构算法而言,DP-Star算法基于轨迹静态(只带有空间信息)的统计性信息去重构特定长度的轨迹. 为了使得生成的轨迹具备动态信息,轨迹生成过程需要服从时空跨度的. DP-Star算法没有考虑时空跨度,需要结合回溯方法去生成轨迹的节点. 假设需要生成的轨迹具有 n 个节点,因为节点所处网格的邻近网格数量由轨迹点数量决定,所以每次节点位置生成可以从附近 $\Theta(n)$ 个网格中选择. 如果轨迹的第 i 个节点的生成不满足时空跨度,则需要进行回溯,重新生成第 $i-1$ 个节点. DP-Star结合回溯法生成服从时空跨度的轨迹,其时间复杂度为 $\Theta(n^2)$.

本文提出了 DPAP-STTP 算法,该算法基于差分隐私活动模式重构时空轨迹. 在 DPAP-STTP 算法重构时空轨迹之前,需要对 RTS 的动静态信息进行调控:

(1) 时空轨迹聚类:将时空轨迹划分成轨迹片段,然后进行聚类形成轨迹簇,再生成 RTS 用于代表轨迹簇.

(2) 隐私预算分配:根据差分隐私的组合性,根据隐私保护需求把隐私总预算 ε 进行划分 $[\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4]$.

(3) 适应性时空网格划分:根据隐私预算 ε_1,获取适应性空间网格 $\hat{O}$ 和统一时间网格 $\hat{T}^m$.

(4) 动静态信息的扰动:基于隐私保护阈值 h 与隐私预算 $[\varepsilon_2,\varepsilon_3,\varepsilon_4]$ 扰动时空路径分布、移动模式、时空跨度.

DPAP-STTP 算法基于扰动的动静态信息重构了每条轨迹的时空点,且保证这些轨迹满足时空跨度的限制:

(1) 初始时空轨迹生成:根据需要生成的轨迹数 tn 与被扰动的时空路径分布 STTD′,采样生成轨迹的始终点所在网格.

(2) 时空跨度生成:按照扰动的时空跨度 STS′输出时空跨度 $(\mathrm{ss},\mathrm{ts})$.

(3) 空间网格生成:按照时空跨度 $(\mathrm{ss},\mathrm{ts})$ 生成 ss 个空间网格以及 ts−2 个轨迹点(非始终点). 第 j 个空间网格 o^j 的生成需要保证轨迹可以经由该网格,从前一个网格 o^{prev} 到达终点网格 o^{end}. 所以当前一个网格 o^{prev} 已生成的情况下,选择 o^k 作为 o^j 的概率为 $X^{\mathrm{ss}-j}_{k,\mathrm{end}}\times X^1_{\mathrm{prev},k}$ $(o^k\neq o^{\mathrm{prev}})$ 或者 $0(o^k=o^{\mathrm{prev}})$. 其中 $X^{\mathrm{ss}-j}_{k,\mathrm{end}}$ 测量了 $\mathrm{ss}-j$ 移动后从 o^k 到 o^{end} 的概率. 空间网格的生成,需要保证每次选择的网格 o^k 与上一个网格 o^{prev} 不同.

(4) 停留时长的确定:按照 $X^1_{\mathrm{prev},\mathrm{prev}}/X^1_{\mathrm{prev},k\neq\mathrm{prev}}$ 的概率确定轨迹在每个网格的停留时长,即该网格有多少个时空点.

算法1　DPAP-STTP基于差分隐私活动模式保护与时空轨迹发布方法

输入:
适应性时空网格:$\hat{O},\hat{T}^m$
生成轨迹数量:tn
两个连续节点间的时间间隔: val
STTD′,MP′,STS′
输出:时空轨迹:trajs
初始化: trajs=∅
//轨迹样本生成
STTD′$(\hat{O},\hat{O},\hat{T}^m)\to$ tn samples $(o^{\mathrm{start}},o^{\mathrm{end}},t^{\mathrm{start}})$
for $i=1$ to tn do:
　STS′$(o^{\mathrm{start}},o^{\mathrm{end}},t^{\mathrm{start}})\to$ ss, ts
　for $j=2$ to ss−1: do //时空网格生成
　　$o^{\mathrm{prev}}=o^{j-1}$
　　randomly select o^k as $o^j\propto X^{\mathrm{ss}-j}_{k,\mathrm{end}}\times X^1_{\mathrm{prev},k}$
　staylist=∅　//停留点生成
　for $j=2$ to ss−1 do:
　　stay$[j-1]=X^1_{j,j+1:o^j=o^{j+1}}/X^1_{j,j+1:o^j\neq o^{j+1}}$
　if sum(staylist) ≤ ts−ss:
　　staylist$[\mathrm{ss}-1]=\mathrm{round}(\mathrm{ts}-\mathrm{ss}-\mathrm{sum}(\mathrm{stay}[2:\mathrm{ss}-2]))$
　　staylist$[0\cdots\mathrm{ss}-2]=\mathrm{round}(\mathrm{staylist}[0\cdots\mathrm{ss}-2])$
　else:
　　staylist$[\mathrm{ss}-1]=0$
　　staylist$[0\cdots\mathrm{ss}-2]=$ round$((\mathrm{ts}-\mathrm{ss})\times\mathrm{staylist}[0\cdots\mathrm{ss}-2]/\mathrm{sum}(\mathrm{staylist}))$
　ps=∅　// 初始化轨迹点
　time=$t\in t^{\mathrm{start}}$
　for $j=1$ to ss do:
　　add $(o^j,\mathrm{time}\in t^{\mathrm{start}})$ to ps
　　time+=val
　　for $v=1$ to staylist$[j-1]$ do:
　　　add (o^j,time) to ps
　　　time+=val
　for $j=1$ to ts do:
　　基于 (o^j,time) in $p\in$ ps 随机生成地点与时间
　trajs$[i]$.ps=ps
return trajs

（5）时空点生成：根据每个轨迹点所处的时空网格(o^x,t^a)，随机生成一个时空点作为轨迹点$p^i\in(o^x,t^a)$.

对比DP-STAR算法，本文所提出的DPAP-STTP算法用时空跨度取代了轨迹长度以生成带有时空属性的轨迹，同时解决了时空跨度限制．在DPAP-STTP算法中，生成服从时空跨度限制的轨迹的时间复杂度为$\Theta(\mathrm{ts})=\Theta(n)$，优于DP-STAR算法与回溯法结合生成轨迹的时间复杂度$\Theta(n^2)$.

6 时空轨迹有效性

时空轨迹生成不仅需要考虑RTS的隐私，还需要保证所生成时空轨迹trajs的有效性．在本文中，其有效性通过三方面来衡量：(1)RTS与trajs在时空网格上轨迹点分布上的平均相对误差．(2)RTS与trajs的频繁模式平均相对误差．(3)RTS与trajs的频繁模式的KL散度均值．

6.1 平均相对误差

平均相对误差MRE（Mean Relative Error）计算了RTS与trajs在不同的时空区间T上的gridnum个网格的时空轨迹点计数的平均相对误差$\mathrm{MRE}(\mathrm{trajs},\mathrm{RTS},T,\hat{O},\mathrm{gridnum})$：

$$\frac{\sum_{m=1}^{M}\sum_{r_2=1}^{\mathrm{gridnum}}\frac{\left|\gamma(\mathrm{trajs},T^m,o^{r_2})-\gamma(\mathrm{RTS},T^m,o^{r_2})\right|}{\gamma(\mathrm{RTS},T^m,o^{r_2})}}{M\times\mathrm{gridnum}}\tag{16}$$

6.2 频繁模式平均相对误差

频繁模式平均相对误差FPAVE（Frequent Pattern Average Relative Error）计算了RTS与trajs在不同的时空区间T上最频繁的k个频繁模式上的支持度的相对误差．例如．一个频繁模式表示为网格间的移动$o^1\to o^2\to o^3$．在频繁模式中$o^1\to o^1\to o^1\to o^2\to o^3$，连续多个轨迹点在同一个网格$o^1$，则该频繁模式$\mathrm{FPAvRE}(\mathrm{trajs},\mathrm{RTS},T)$可用$o^1\to(o^1,\mathrm{stay})\to o^2\to o^3$表示：

$$\frac{\sum_{m=1}^{M}\sum_{P\in\mathrm{FP(RTS)}^k}\frac{\left|\mathrm{supp}(\mathrm{trajs},P,T^m),-\mathrm{supp}(\mathrm{RTS},P,T^m)\right|}{\mathrm{supp}(\mathrm{RTS},P,T^m)}}{k\times M}\tag{17}$$

6.3 频繁模式KL散度

频繁模式KL散度FPKL（Frequent Pattern Kullback-Leibler Divergence）计算了RTS与trajs在同不同时间区间T上最频繁的k个频繁模式上支持度分布的差值均值，也表示的产生trajs的信息损耗$\mathrm{FPKL}(\mathrm{trajs},\mathrm{RTS},T)$：

$$\frac{\sum_{m=1}^{M}\mathrm{KL}\left(\left\{\mathrm{supp}(\mathrm{trajs},P,T^m),\mathrm{supp}(\mathrm{RTS},P,T^m)\right\}_{P\in\mathrm{FP(RTS)}^k}\right)}{M}\tag{18}$$

7 实验结果与分析

基于真实数据库GEOLIFE[25]中的位移数据，本实验调控了代表性轨迹集RTS的动静态信息，并实现了活动模式隐私保护与生成轨迹有效性验证．

7.1 个人代表性轨迹生成

本实验从GEOLIFE数据库中选取了30个人的位移数据，并把所有位置的经纬度映射为utm坐标系中的x与y值．例如，经纬度(115.395 903 3，41.214 191 1)可以映射为(x:32 900 000，y:6 700 000)．

图3中，带时间戳的地点位于utm坐标系(x:[32 900 000，33 200 000]，y:[6 700 000，7 000 000])．另外，时间被映射到时间周期[00:00:00，23:59:59]．第一级统一空间网格O^{r_1}的空间粒度为(x-10 000，y-10 000)．第一级T^m的时间粒度为4小时，第二级t^{ms}的时间粒度为一刻钟．图4中展示了代表性轨迹集生成的过程．首先GEOLIFE轨迹数据集(图4(a))基于MDL方法进行划分获取轨迹片段(图4(b))．然后，划分所得的轨迹片段通过ST-DBSCAN算法$(\mathrm{eps}_1=1\,000,\mathrm{eps}_2=150)$按照密度进行聚类并形成多个轨迹簇．在ST-DBSCAN算法中，轨迹簇内的各个轨迹片段的空间距离通过DTW算法测量，而其时间距离通过轨迹点平均时间差距来衡量．最后选择每个轨迹簇中最小平均距离的中心轨迹片段作为代表性轨迹(图4(c))．

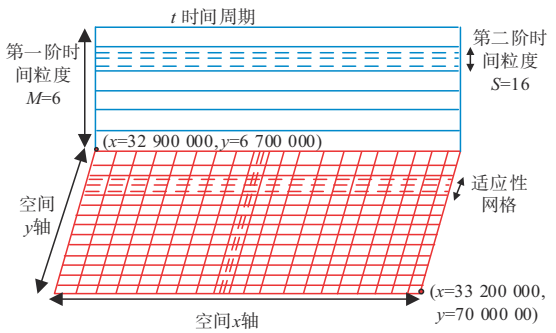
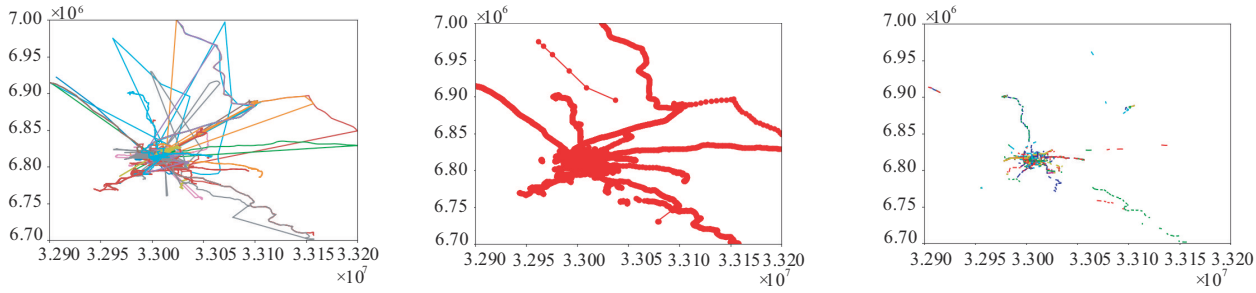


图3 特定时空范围的适应性网格

7.2 差分隐私活动模式设置

差分隐私活动模式包括了从代表性轨迹集RTS中获取的被扰动的动静态信息STD′，STTD′，MP′及STS′．动静态信息的扰动程度由隐私预算和隐私保护阈值所决定．本实验将隐私预算$\varepsilon=\{0.5,1,2\}$分成$(\varepsilon_1=0.1\varepsilon,\varepsilon_2=0.3\varepsilon,\varepsilon_3=0.3\varepsilon,\varepsilon_4=0.3\varepsilon)$．$\varepsilon_1$用于获取扰动的STD′来生成适应性时空网格．其余隐私预算$\varepsilon_{2,3,4}$用于调控适应性网格上RTS计数并获取STTD′，MP′，STS′．由于$\varepsilon_{2,3,4}$会扰动真实的轨迹计数，所以设定ε_1取值小于$\varepsilon_{2,3,4}$使生成轨迹更具有效性．在任意一个时空网格(o^{r_2},T^m)上添加的噪声可按照第二级时间网格的



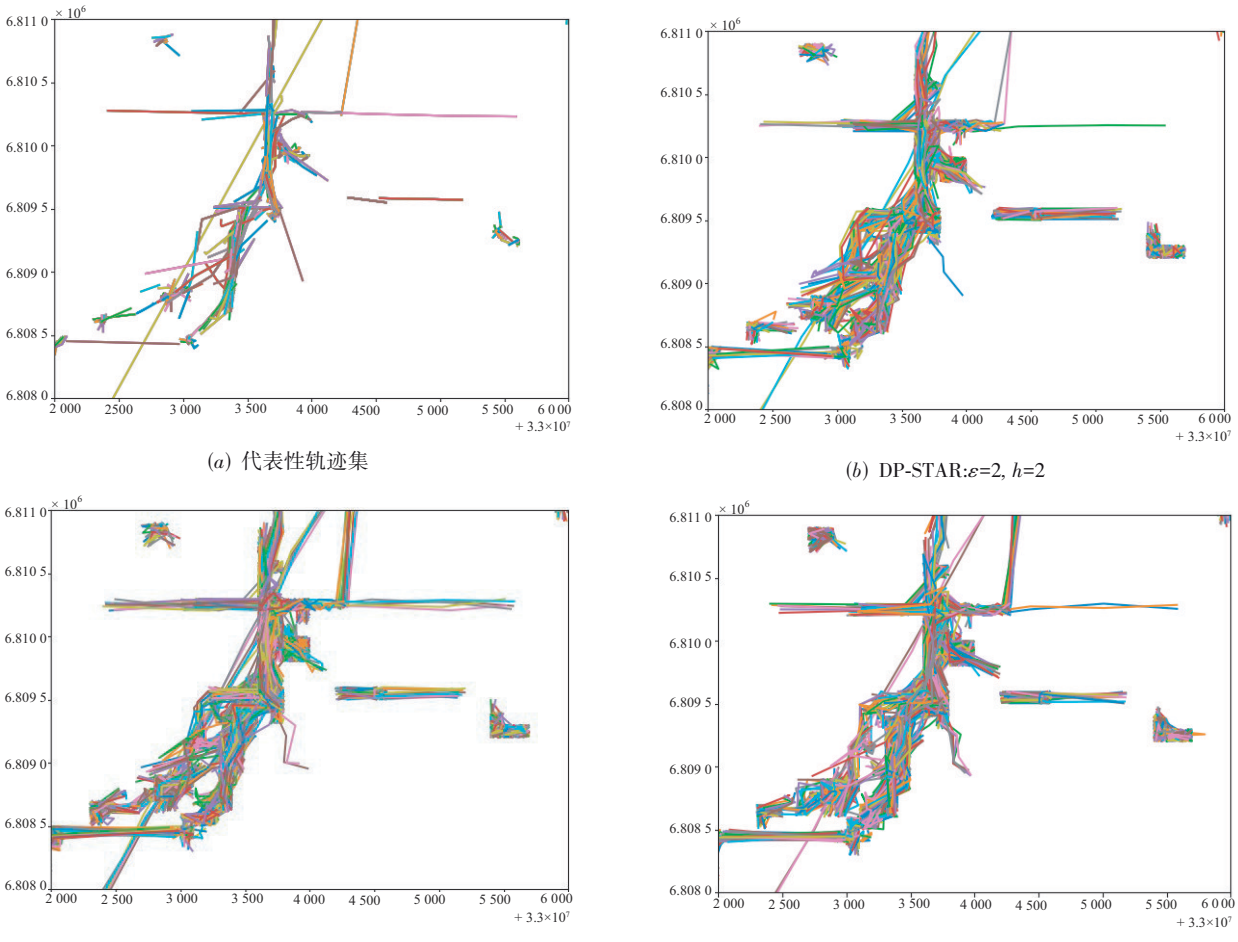
(a) GEOLIFE的轨迹数据集　(b) GEOLIFE的轨迹片段　(c) GEOLIFE的代表性轨迹集生成

图4　GEOLIFE数据的代表性轨迹集生成

时间粒度$S=16$平均地为每个时空网格$\left(o^{r_2},t^{ms}\right)$的轨迹计数添加噪声$\mathrm{Lap}\left(\varepsilon_2/h\right)/S$与$\mathrm{Lap}\left(\varepsilon_2/h\right)/S$. 另外,实验中隐私保护阈值设定为$h=\{1,2\}$,选择$h=1$用于验证RTS中单条代表性轨迹的隐私保护效果. 选择$h=2$用来验证轨迹计数为2的轨迹集的隐私保护效果.

7.3　DPAP-STTP方法

在适应性时空网格下,STTD′,MP′,STS′用于时空轨迹的生成. 所生成的时空轨迹的有效性通过生成轨迹与所选择代表性轨迹的误差来测量. 实验中,所选择的代表性轨迹来自于时间范围T^1,T^2,T^3,且这些代表性轨迹的起点都位于$\hat{O}$中100个密度最高的空间网格o^{r_2}. DPAP-STTP首先根据STS′生成了时空轨迹的起始点. 然后基于MP′和STS′生成所有的时空点. 图5、图6与图7分别展示了DP-STAR算法与DPAP-STTP算法在隐私预算$\varepsilon=\{0.5,1,2\}$和隐私保护阈值$h=2$下所生

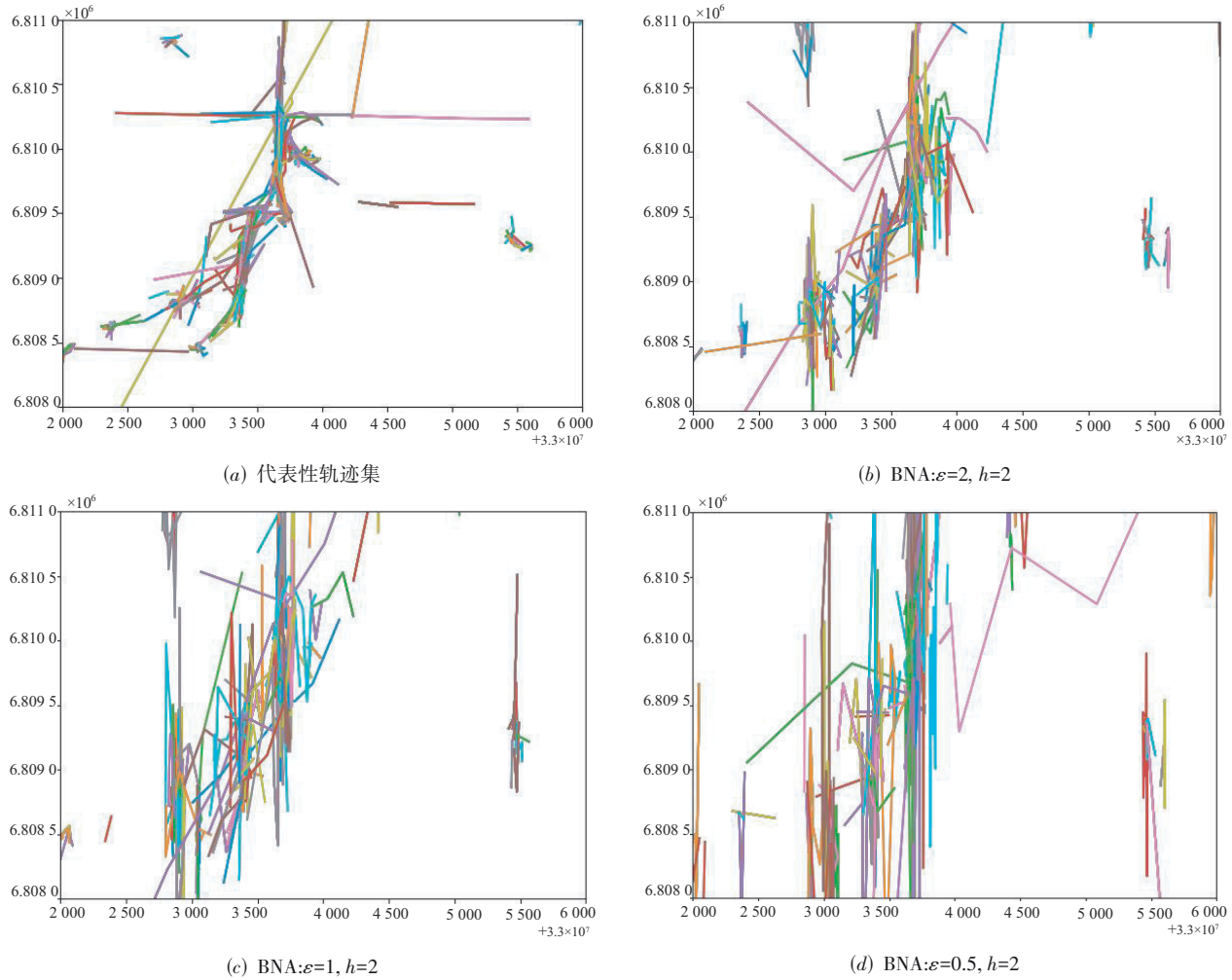


(a) 代表性轨迹集　(b) DP-STAR:ε=2, h=2

(c) DP-STAR:ε=1, h=2　(d) DP-STAR:ε=0.5, h=2

图5　基于DP-STAR算法所生成轨迹

成轨迹的有效性. 图7中,在不同时间范围产生的轨迹用不同的颜色表示(T^1(蓝色),T^2(绿色),T^3(红色)). 由表2和表3可见,DP-STAR方法与DPAP-STTP算法具有相同的特性:隐私总预算ε越小,生成的时空轨迹有效性越小. 隐私保护阈值h越小,有效性则越大. 本实验将DPAP-STTP时空轨迹(图7)与DP-STAR[12]、BNA(Bounded Noise-Adding Algorithm)[25]轨迹在时空轨迹有效性上进行对比(图5、图6).



(a) 代表性轨迹集 (b) BNA:ε=2, h=2

(c) BNA:ε=1, h=2 (d) BNA:ε=0.5, h=2

图6 基于Bounded Noise-adding算法所生成轨迹

DP-STAR算法使用被扰动的轨迹统计性信息,划分空间网格$\hat{O}$,并在该空间网格上按照采样获取轨迹的起始点(o^x,o^y),再根据轨迹长度与轨迹在任意两网格间的转移概率随机生成轨迹中间点来重构轨迹. 由于轨迹中间点是通过转移概率随机采用生成,则容易出现多个连续轨迹点过于集中个别或少数空间网格. 而DPAP-STTP算法不但在时空网格上生成了带有时间属性的轨迹,还设定了轨迹的停留时长与转移次数,可避免生成的连续轨迹点在空间上过于集中.

Bounded Noise-adding方法是通过staircase噪声扰动了轨迹中节点聚类的中心以及轨迹数量,从而影响了各个网格上的节点数量与任意两个网格之间轨迹节点转移次数,最终改变了BNA生成轨迹在适应性时空网格上的轨迹数量查询结果与频繁位置转换的挖掘结果. 而DPAP-STTP算法是直接扰动轨迹节点在适应性时空网格上的密度与轨迹节点的转换概率,再重构时空轨迹. 另外,在BNA算法对于轨迹中节点位置的扰动不考虑中上一个节点位置,容易导致轨迹中两个连续节点的距离过长. 而DPAP-STTP算法限制了轨迹节点可转换的网格范围,降低了轨迹节点在网格上转换时的偏移.

在表2~表4中,DP-STAR算法所生成轨迹在时空网格上的轨迹计数的有效性MRE上略高于DPAP-STTP(见表2、表3). 而频繁模式挖掘结果上,DPAP-STTP生成轨迹的有效性(FPAVE与FPKL)略高于DP-STAR所生成轨迹. 尽管DPAP-STTP轨迹平均有效性不能完全

表2　DP-STAR生成轨迹有效性

DP-STAR	$\varepsilon=0.5$	$\varepsilon=1$	$\varepsilon=2$
MRE($h=1$)	0.154 4	0.090 0	0.066 0
FPAVE($h=1$)	0.326 8	0.293 1	0.045 2
FPKL($h=1$)	0.002 6	0.002 5	0.002 1
MRE($h=2$)	0.224 1	0.207 9	0.112 8
FPAVE($h=2$)	0.536 6	0.425 2	0.042 7
FPKL($h=2$)	0.003 0	0.001 9	0.000 8

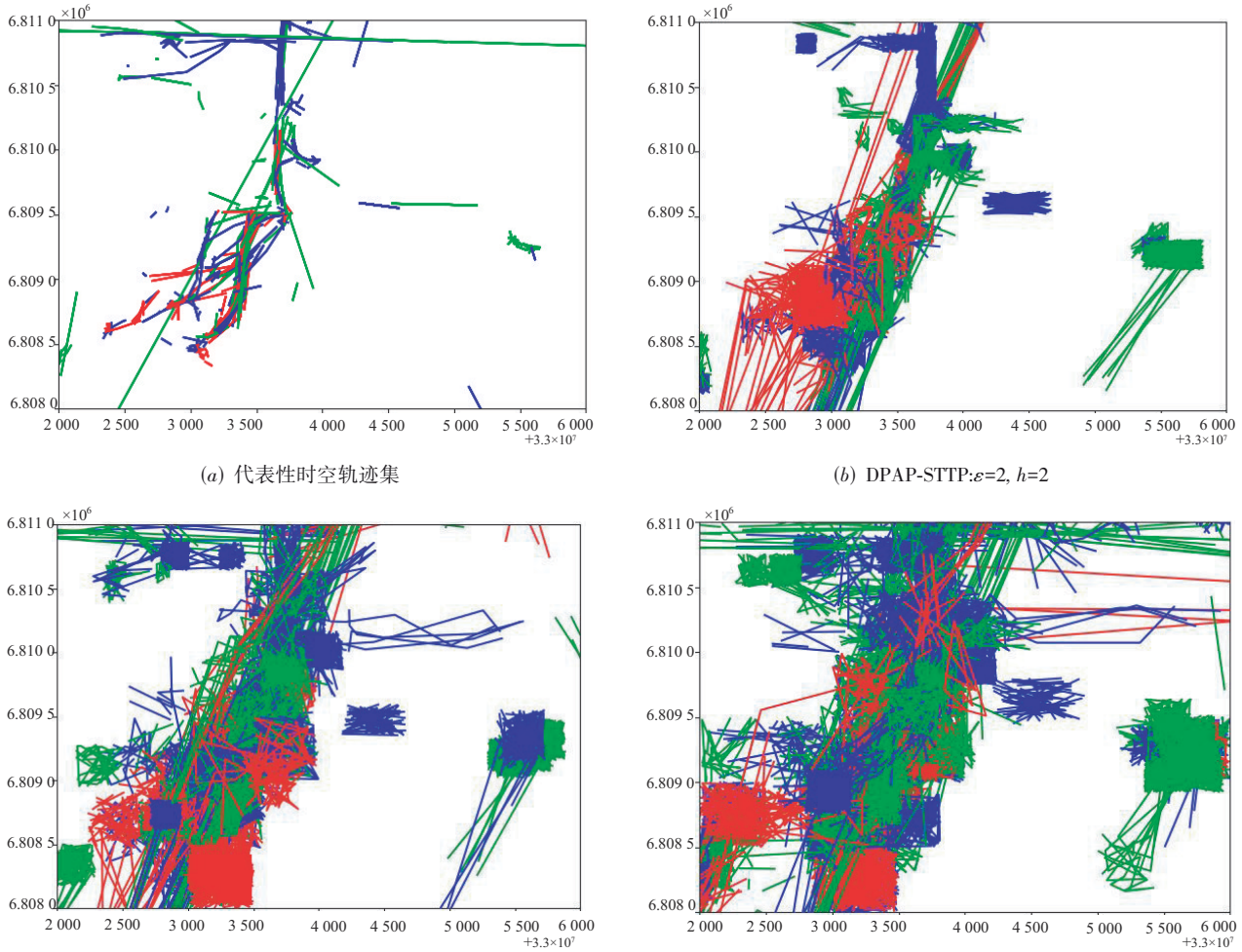
表3　DPAP-STTP生成轨迹有效性

DPAP-STTP	$\varepsilon=0.5$	$\varepsilon=1$	$\varepsilon=2$
MRE($h=1$)	0.385 0	0.325 6	0.265 8
FPAVE($h=1$)	0.141 4	0.044 8	0.039 0
FPKL($h=1$)	0.001 4	0.001 6	0.001 4
MRE($h=2$)	0.414 2	0.330 0	0.253 2
FPAVE($h=2$)	0.081 6	0.110 3	0.044 4
FPKL($h=2$)	0.003 4	0.001 7	0.001 5

超过DP-STAR,但是也在部分时间范围上具有更好的有效性. 以表5中加粗数据结果为例,当隐私保护阈值为2时,DPAP-STTP生成轨迹的FPAVE值比DP-STAR生成轨迹低. 所以DPAP-STTP不仅仅赋予了轨迹的时间信息,还维持了轨迹的有效性. 另外,由表3与表4中可知,在适应性网格上,DPAP-STTP轨迹的轨迹数量查询结果误差MRE与频繁项集挖掘结果误差(FPAVE与FPKL)更小,所以DPAP-STTP轨迹在网格上的有效性高于BNA轨迹.

表4　BNA生成轨迹有效性

BNA	$\varepsilon=0.5$	$\varepsilon=1$	$\varepsilon=2$
MRE($h=1$)	0.857 6	0.683 0	0.442 6
FPAVE($h=1$)	1.362 2	1.088 0	0.955 4
FPKL($h=1$)	0.008 3	0.009 8	0.014 6
MRE($h=2$)	0.914 0	0.730 6	0.555 4
FPAVE($h=2$)	1.391 5	1.230 7	1.086 5
FPKL($h=2$)	0.010 4	0.010 8	0.006 5



(*a*) 代表性时空轨迹集　　(*b*) DPAP-STTP:ε=2, h=2

(*c*) DPAP-STTP:ε=1, h=2　　(*d*) DPAP-STTP:ε=0.5, h=2

图7　基于DPAP-STTP算法所生成时空轨迹

表5 DPAP-STTP在各个时间范围生成轨迹的有效性

$h=2$	$\varepsilon=0.5$	$\varepsilon=1$	$\varepsilon=2$
MRE(T^1)	0.452 4	0.277 0	0.137 1
MRE(T^2)	0.353 8	0.388 9	0.365 7
MRE(T^3)	0.436 4	0.324 1	0.256 6
FPAVE(T^1)	**0.116 6**	**0.071 9**	**0.096 7**
FPAVE(T^2)	**0.043 4**	**0.115 7**	**0.017 0**
FPAVE(T^3)	**0.084 6**	**0.143 4**	**0.019 5**
FPKL(T^1)	0.004 5	0.000 7	0.001 2
FPKL(T^2)	0.001 9	0.001 4	0.001 0
FPKL(T^3)	0.003 7	0.002 9	0.002 2

8 结论

第三方服务提供商不但具有能够获取用户活动模式的大数据挖掘与分析技术,还可以基于大规模专家知识生成面向用户日常生活的服务建议,这使得用户需要在获得所需服务建议的同时保护自己的活动模式隐私,本文提出了DPAP-STTP方法以实现活动模式的隐私保护,并生成活动模式满足群体差分隐私的时空轨迹发布给第三方服务提供商,该方法通过保护轨迹计数不大于隐私保护阈值的代表性轨迹,保护了活动模式隐私且维持了所发布时空轨迹的有效性.

未来的研究会聚焦于在多个参与者之间分享轨迹数据的动静态信息时防止原始数据的泄露. 尤其是,当某个参与者为攻击者的时候,如何防止攻击者在获取动静态信息后,通过成员推理攻击来推断某类数据是否存在于原始数据集中,或者重构与原始数据相似的数据实现隐私侵犯.

参考文献

[1] JIANG S, FERREIRA J, GONZALEZ M C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore[J]. IEEE Transactions on Big Data, 2017, 3(2): 208-219.

[2] CHANG S, LI C, ZHU H Z, et al. Revealing privacy vulnerabilities of anonymous trajectories[J]. IEEE Transactions on Vehicular Technology, 2018, 67(12): 12061-12071.

[3] AUGUSTIN D, HOFMANN M, KONIGORSKI U. Motion pattern recognition for maneuver detection and trajectory prediction on highways[C]//2018 IEEE International Conference on Vehicular Electronics and Safety. Piscataway: IEEE, 2018: 1-8.

[4] ZHANG D Z, LEE K, LEE I. Mining hierarchical semantic periodic patterns from GPS-collected spatio-temporal trajectories[J]. Expert Systems With Applications, 2019, 122: 85-101.

[5] GURSOY M E, LIU L, TRUEX S, et al. Differentially private and utility preserving publication of trajectory data[J]. IEEE Transactions on Mobile Computing, 2019, 18(10): 2315-2329.

[6] 汪成亮, 黄心田. 智能环境下基于雾计算的推理节点优化分配研究[J]. 电子学报, 2020, 48(1): 35-43.
WANG C L, HUANG X T. Study on optimal allocation of inference nodes for fog computing in smart environment[J]. Acta Electronica Sinica, 2020, 48(1): 35-43. (in Chinese)

[7] SHOU Z Y, DI X. Similarity analysis of frequent sequential activity pattern mining[J]. Transportation Research Part C: Emerging Technologies, 2018, 96: 122-143.

[8] LEE J, HAN J, WHANG K. Trajectory clustering: A partition-and-group framework[C]//Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2007:593-604.

[9] CHEN W, JI M H, WANG J M. T-DBSCAN: A spatiotemporal density clustering for GPS trajectory segmentation [J]. International Journal of Online and Biomedical Engineering (IJOE), 2014, 10(6): 19.

[10] BIRANT D, KUT A. ST-DBSCAN: An algorithm for clustering spatial-temporal data[J]. Data & Knowledge Engineering, 2007, 60(1): 208-221.

[11] ANSARI M Y, MAINUDDIN, AHMAD A, et al. Spatiotemporal trajectory clustering: A clustering algorithm for spatiotemporal data[J]. Expert Systems with Applications, 2021, 178: 115048.

[12] HONG Z H, CHEN Y, MAHMASSANI H S. Recognizing network trip patterns using a spatio-temporal vehicle trajectory clustering algorithm[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(8): 2548-2557.

[13] NADERIVESAL S, KULIK L, BAILEY J. An effective and versatile distance measure for spatiotemporal trajectories[J]. Data Mining and Knowledge Discovery, 2019, 33 (3): 577-606.

[14] XU F L, XIA T, CAO H C, et al. Detecting popular temporal modes in population-scale unlabelled trajectory data [J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 46.

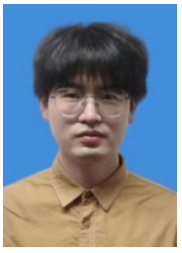
[15] 于彦伟, 贾召飞, 曹磊, 等. 面向位置大数据的快速密度聚类算法[J]. 软件学报, 2018, 29(8): 2470-2484.
YU Y W, JIA Z F, CAO L, et al. Fast density-based clustering algorithm for location big data[J]. Journal of Software, 2018, 29(8): 2470-2484. (in Chinese)

[16] DWORK C, ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science, 2013, 9(3/4): 211-407.
[17] HE X, CORMODE G, MACHANAVAJJHALA A. DPT: differentially private trajectory synthesis using hierarchical reference systems[J]. Proceedings of the VLDB Endowment, 2015, 8(11):1154-1165.
[18] TERROVITIS M, POULIS G, MAMOULIS N, et al. Local suppression and splitting techniques for privacy preserving publication of trajectories[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(7): 1466-1479.
[19] 霍峥, 孟小峰. 一种满足差分隐私的轨迹数据发布方法[J]. 计算机学报, 2018, 41(2): 400-412.
HUO Z, MENG X F. A trajectory data publication method under differential privacy[J]. Chinese Journal of Computers, 2018, 41(2): 400-412. (in Chinese)
[20] ZHAO X D, PI D C, CHEN J F. Novel trajectory privacy-preserving method based on prefix tree using differential privacy[J]. Knowledge-Based Systems, 2020, 198: 105940.
[21] ZHANG Z K, WU T T, SUN X T, et al. MPDP *k*-medoids: Multiple partition differential privacy preserving *k*-medoids clustering for data publishing in the Internet of Medical Things[J]. International Journal of Distributed Sensor Networks, 2021, 17(10): 155014772110425.
[22] MA T H, SONG F G. A trajectory privacy protection method based on random sampling differential privacy[J]. ISPRS International Journal of Geo-Information, 2021, 10(7): 454.
[23] WANG H, XU Z Q, JIA S. Cluster-Indistinguishability: A practical differential privacy mechanism for trajectory clustering[J]. Intelligent Data Analysis, 2017, 21(6): 1305-1326.
[24] LIU Q, YU J, HAN J M, et al. Differentially private and utility-aware publication of trajectory data[J]. Expert Systems With Applications, 2021, 180: 115120.
[25] ZHENG Y, ZHANG L Z, XIE X, et al. Mining interesting locations and travel sequences from GPS trajectories [C]//Proceedings of the 18th International Conference on World Wide Web. New York: ACM, 2009: 791-800.

作者简介

曾　卓　男,1991年5月出生,重庆涪陵人. 重庆大学博士生,主要研究轨迹隐私保护,动态图神经网络安全等.
E-mail: zengz@ cqu.edu.cn

汪成亮(通讯作者)　男,1975年5月出生,四川资阳人. 博士,现为重庆大学计算机学院教授,博士生导师. 主要研究领域为复杂系统智能控制,无线网络及RFID研究与应用等.
E-mail: wangcl@ cqu.edu.cn



马　飞　男,1993年4月出生,河南商丘人. 重庆大学博士生,主要研究方向为语音信号处理、迁移学习.
E-mail: mafei@cqu.edu.cn