Computer Science and Application 计算机科学与应用, 2022, 12(1), 135-146
Published Online January 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.121015

Hans汉斯

稀疏环境下基于假轨迹的轨迹隐私保护方法

黄　景，柳　毅

广东工业大学计算机学院，广东　广州

收稿日期：2021年12月22日；录用日期：2022年1月19日；发布日期：2022年1月26日

摘　要

针对稀疏环境下的移动对象轨迹数据经匿名处理后可用性低的问题，提出一种稀疏环境下基于假轨迹的轨迹隐私保护算法。在本文算法中，考虑了移动对象所处的地理环境，将轨迹的整体方向和轨迹间距作为选择假轨迹的重要依据。此外，还提出了使用访问概率的概念来平衡匿名和数据可用性，从而实现轨迹数据匿名。基于移动对象的轨迹数据集进行实验与分析，实验结果表明，本文算法在满足轨迹数据匿名需求的情况下有更高的数据可用性。

关键词

轨迹数据，*k*-Anonymity，假轨迹，数据可用性，数据发布

Trajectory Privacy Protection Method Based on Dummy Trajectory in Sparse Environment

Jing Huang, Yi Liu

School of Computer, Guangdong University of Technology, Guangzhou Guangdong

Received: Dec. 22nd, 2021; accepted: Jan. 19th, 2022; published: Jan. 26th, 2022

Abstract

Aiming at the problem of low availability of moving object trajectory data in sparse environment after anonymous processing, a trajectory privacy protection algorithm based on dummy trajectories in sparse environment is proposed. In the algorithm of this paper, the geographical environment of the moving object is considered, and the overall direction of the trajectory and the distance between the trajectories are taken as an important basis for selecting dummy trajectories. In addition, the concept of using access probability is proposed to balance anonymity and data

文章引用: 黄景, 柳毅. 稀疏环境下基于假轨迹的轨迹隐私保护方法[J]. 计算机科学与应用, 2022, 12(1): 135-146.
DOI: 10.12677/csa.2022.121015

availability, so as to achieve anonymity of trajectory data. Experiments and analyses are carried out based on the trajectory data set of moving objects. The experimental results show that the algorithm in this paper has higher data availability while meeting the anonymity requirements of trajectory data.

Keywords

Trajectory Data, *k*-Anonymity, Dummy Trajectory, Data Availability, Data Publishing

1. 引言

现如今，基于位置的服务(Location Based Services, LBS)已被广泛应用于各种应用程序，所涉及的领域多种多样，如微博、大众点评、Facebook 以及各种地图导航软件等。通过 LBS，人们可以随时随地发布自己的定位，可以获取附近感兴趣的餐厅、商店等，除此之外，最优路线的推荐也是非常实用的一种服务[1]。随着移动设备与 LBS 的发展，移动用户日常所留下的轨迹数据呈爆炸式增长，海量的轨迹数据难逃于被收集与挖掘[2] [3] [4] [5]。当轨迹数据采集之后不经过处理则直接发布时，基于用户的行为习惯的挖掘可以带来较高的商业效益，但所带来隐私泄露的风险也是巨大的[6] [7]，因此针对轨迹数据发布的隐私保护问题已经成为了一个亟需解决的问题[8] [9]。

轨迹数据隐私保护是基于位置隐私保护技术基础之上的，其目的是保证攻击者无法从包含目标对象信息的匿名轨迹集中推测出目标对象的轨迹信息，从而进行过度的数据挖掘。针对位置隐私保护，Gruteser 等人[10]是第一次将 *k*-anonymity 技术用于位置隐私保护，之后这也成为较为主流的应用技术之一。匿名的实质就是用其他对象的行为来掩盖自己的行为，由位置 *k*-anonymity 到轨迹 *k*-anonymity，Abul 等人[11]将 *k*-anonymity 应用于轨迹隐私保护中，其实质是将用户的基轨迹信息隐藏在一个匿名集中，通过保存至少 *k* 条轨迹记录来隐藏一条轨迹数据，以达到保护隐私的目的。

针对轨迹隐私保护的研究，现如今也取得了一定的成果。最早关于轨迹数据匿名与隐私保护的研究是以 *k*-anonymity 技术为代表。文献[11]提出的(*k*，&)匿名模型，将 *k*-anonymity 技术与聚类方法相结合应用于轨迹隐私保护，其中 *k* 为匿名等级，&为匿名区的半径，先基于欧式距离对轨迹进行聚类再由匿名区中的轨迹求均值或是重建来获得公开发布的轨迹。文献[12]在 *k*-anonymity 技术的基础上做了改进，提出了一种 k^m 匿名方法，*k* 和 *m* 分别为隐私保护强度和攻击者掌握的用户轨迹的位置点信息。在泛化思想下，k^m 匿名方法先是最小化原始轨迹和匿名轨迹之间的欧式距离，再通过归纳法实现轨迹信息的匿名化。文献[13]考虑了真实轨迹与虚假轨迹之间的距离问题，对生成的虚假轨迹做进一步扰动，使攻击者无法从匿名集中区分用户的真实轨迹与虚假轨迹。文献[14]提出了一种基于时空关联性的假轨迹生成方法，对轨迹时空关联性和轨迹相似性的角度进行分析，该方法考虑了真实轨迹的整体方向，将方向斜率作为轨迹相似性的衡量标准，并确保虚假轨迹中每段路径的时空可达性，基于该方法的虚假轨迹扰乱了攻击者的视线，保护了用户的轨迹隐私。在对轨迹数据进行匿名处理时往往还会将轨迹抑制法与 *k*-anonymity 相结合。轨迹抑制法主要思想是通过删除轨迹中敏感位置点来达到隐私保护的效果。文献[15]在轨迹抑制法的基础之上引入了一种分裂思想，通过对局部抑制、全局抑制和轨迹分裂的结合来减少匿名和过度抑制带

来的信息损失。文献[16]提出一种基于用户暴露位置的轨迹隐私保护算法，使用真实有效的位置点来形成虚拟轨迹并对不符合匿名条件的轨迹点进行抑制，有效预防了针对暴露位置检索真实轨迹的攻击。文献[17]考虑用户的行为习惯与起终点的关联，生成安全的起终点候选集，根据候选集生成 $k-1$ 匿名轨迹，该方法保持较高轨迹相似性的同时，又能够保证轨迹隐私的安全性。

以上研究均在一定程度保护了用户的轨迹数据信息，但却忽略轨迹数据在真实地理环境中的分布情况，当轨迹中的点所处于稀疏环境或难以抵达的位置时，通常采用抑制法，但直接抑制会造成位置信息损失较大。为了降低用户轨迹被识别的风险以及保证轨迹数据发布后的效用需求，本文提出了一种稀疏环境下基于假轨迹的轨迹隐私保护方法，该方法考虑了地理特征和轨迹的整体方向，在保证隐私保护效果的同时提高轨迹数据的发布质量，减小轨迹数据由匿名化带来的信息损失。

2. 相关概念

定义 1 基轨迹和轨迹数据集移动对象的位置点按时间排序所得到的序列为该对象的轨迹，需要做匿名处理的轨迹即为基轨迹。位置点 loc 为三维空间中的一点，包括二维坐标和时间维，表示为 $loc=(x,y,t)$。因此，基轨迹 tr 则为三维空间中的一条折线，记为 $tr=\{loc_1,loc_2,loc_3,\cdots,loc_n\}$，其中，$loc_i=(x_i,y_i,t_i)$ 表示基轨迹 tr 在 t_i 时刻的位置为 (x_i,y_i)，n 为基轨迹 tr 的位置点数，即 tr 的长度。轨迹数据集是一组轨迹组成的序列集合，记为 $T=\{tr_1,tr_2,tr_3,\cdots,tr_m\}$，$m$ 为轨迹数据集 T 中轨迹的数目。

定义 2 访问概率访问概率是判断一个区域为访问密集还是稀疏的关键因素，用 Q 来表示。图 1 是将一条基轨迹所处的地图数据进行网格划分，每个网格的访问概率不同，网格表示为 $lattice$。将每个网格访问概率定义为式(1)，其中，n 为每条轨迹位置点数，m 为轨迹数据集中的轨迹数目，$\left|lattice_{i,j}\right|$ 为网格内的位置点数。

$$Q_i=\frac{\left|lattice_{i,j}\right|}{\sum_1^m\sum_1^n loc_i} \tag{1}$$

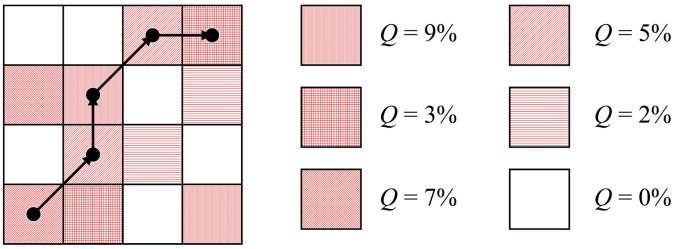

Figure 1. Map data gridding and access probability of base trajectory
图 1. 基轨迹的地图数据网格化及访问概率

定义 3 假轨迹与可供发布的匿名轨迹集基于 k-anonymity 模型的基础，在隐私保护需求下由基轨迹 tr 经过匿名处理得到的轨迹称为假轨迹 tr^d，由假轨迹组成匿名轨迹集，且攻击者无法从该匿名轨迹集中轻易识别出基轨迹，则称该匿名轨迹集为可供发布的匿名轨迹集 TRS。

$$TRS=\{tr,tr_1^d,tr_2^d,\cdots,tr_{k-1}^d\} \tag{2}$$

$$tr^d=\{loc_1^d,loc_2^d,loc_3^d,\cdots,loc_n^d\} \tag{3}$$

式(2)中 k 为匿名等级，假轨迹 tr^d 表示为式(3)。表 1 列出了基轨迹 tr 与可供发布的匿名轨迹集 TRS 的对应关系，此时 $k=3$。

Table 1. Base trajectory and publishable anonymous trajectory dataset
表 1. 基轨迹和可供发布的匿名轨迹数据集

基轨迹	可供发布的匿名轨迹数据集
$loc_1 \to loc_2 \to loc_3$	$loc_1 \to loc_2 \to loc_3$
	$loc_1^{d_1} \to loc_2^{d_1} \to loc_3^{d_1}$
	$loc_1^{d_2} \to loc_2^{d_2} \to loc_3^{d_2}$

定义 4 位置点间间距与轨迹间距给定两个位置点 $loc_i\left(x_i,y_i,t_i\right)$ 和 $loc_j\left(x_j,y_j,t_j\right)$，采用欧几里得距离(Euclidean distance)定义其间距为式(4)。

$$dist\_loc\left(loc_i,loc_j\right)=\sqrt{\left(x_i-x_j\right)^2+\left(y_i-y_j\right)^2} \tag{4}$$

若是给定两条轨迹 $tr=\left\{loc_1,loc_2,loc_3,\cdots,loc_n\right\}$ 和 $tr'=\left\{loc_1',loc_2',loc_3',\cdots,loc_n'\right\}$，则轨迹间距定义为

$$dist\_tr\left(tr,tr'\right)=\frac{1}{n}\sum_{i=1}^{n}dist\left(loc_i,loc_i'\right);\left(1\le i\le n\right) \tag{5}$$

定义 5 轨迹方向夹角每一条轨迹都可看作为空间中的一条有方向的折线，方向是由起点指向终点向量来表示。每一条轨迹都有一个对应的向量表示，轨迹方向夹角 α 由两个向量间的夹角来定义[18]。

如图 2 所示，α 为轨迹 tr 和 tr' 的夹角，若 tr 和 tr' 的起点分别为 $loc_1\left(x_1,y_1,t_1\right)$ 和 $loc_1'\left(x_1',y_1',t_1'\right)$，终点分别为 $loc_n\left(x_n,y_n,t_n\right)$ 和 $loc_n'\left(x_n',y_n',t_n'\right)$，则 α 定义为式(6)

$$\alpha\left(tr,tr'\right)=\arccos\left(\frac{\left(x_1-x_n\right)\times\left(x_1'-x_n'\right)+\left(y_1-y_n\right)\times\left(y_1'-y_n'\right)}{\sqrt{\left(x_1-x_n\right)^2+\left(y_1-y_n\right)^2}\times\sqrt{\left(x_1'-x'\right)^2+\left(y_1'-y'\right)^2}}\right) \tag{6}$$

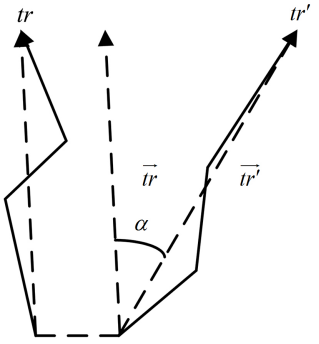

Figure 2. Definition of the angle between two trajectories
图 2. 两条轨迹夹角的定义

3. 稀疏环境下基于假轨迹的轨迹隐私保护算法

为解决过度抑制而导致的轨迹中位置信息损失较大的问题，本文提出了一种稀疏环境下基于假轨迹的轨迹隐私保护(Trajectory Protection based on Dummy Trajectory in Sparse Environment, TPDTSE)算法，该算法主要包含三个步骤：1) 筛选出基轨迹中处于最小访问概率的位置点；2) 将基轨迹中位置点基于访问概率的判断下进行抑制或匿名等级调整来得到匿名位置集；3) 通过特定的方法拟合假轨迹，并筛选出满足匿名需求的假轨迹与基轨迹组成可供发布的匿名轨迹集。TPDTSE 算法的流程图如图 3 所示。

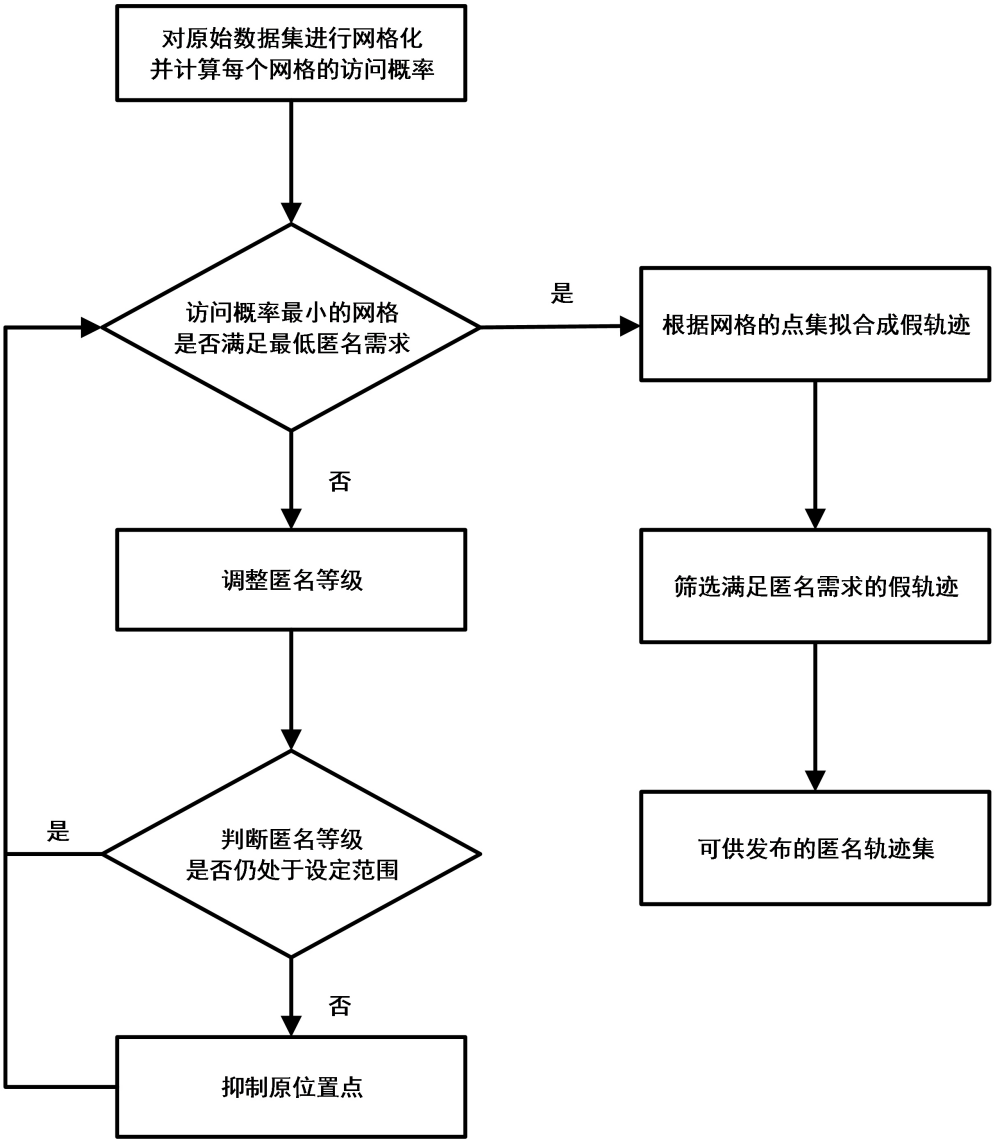

Figure 3. Algorithm flow diagram
图 3. 算法流程图

3.1. 基于访问概率的匿名位置集生成算法

网格划分是用来处理空间数据的一种常用且高效的方法，算法 1 使用网格划分法，输入原始轨迹数据集 T、用户的基轨迹 tr、最大距离阈值 μ、匿名等级 k 和最低匿名等级 $k_{\min}$，以二维表的形式输出匿名位置集 $Dlocs$。基于访问概率的匿名位置集生成算法的伪代码如算法 1 所示。

算法 1 基于访问概率的匿名位置集生成算法

输入：轨迹数据集 T，基轨迹 tr，最大距离阈值 μ， k， $k_{\min}$；

输出：匿名位置点 $Dlocs$

1) $Dlocs \leftarrow \varnothing$;

2) Divide the T into lattices with a size of 2μ;

3) Calculate the Q of each lattice;

4) $p=\sqrt[n]{k}$;

5) for each loc_i of tr in lattice do;

6) find $Q_{\min}$ of tr;

7) if ($\left|lattice_{Q_{\min}}<p\right|$)

8) $Dlocs_i=lattice_i$;

9) else

10) $k=k-1$;

11) if ($k\geq k_{\min}$)

12) back to 4);

13) else

14) delete loc_i in tr;

15) $k=k+1$ and back to 4);

16) return $Dlocs$;

在算法1中，首先根据轨迹数据集划分为单元网格大小为2μ的网格，并计算每个网格的访问概率Q，其中的μ为轨迹间距的最大距离阈值。对于每条基轨迹tr，查询其每个位置点loc_i的所在的网格$lattice_i$，从访问概率最小的网格开始判断该网格中的点的数量是否满足最低匿名需求p，若满足，该网格则作为loc_i对应的匿名位置集；若不满足，则调整匿名等级并判断是大于最低匿名等级，是则返回第四步，否则将loc_i抑制，因为该点所在的网格内没有足够的匿名位置来隐匿基轨迹tr中的真实位置，只能采取抑制。其中最低匿名需求p的设定，与匿名等级k相关，假设基轨迹tr中有n个位置点，每个位置点对应的匿名位置集中有p个匿名位置，在没有抑制的情况下，则有n个匿名位置集，理论上通过枚举的方法可得到p^n条假轨迹，所以最低匿名需求p定义为$p=\sqrt[n]{k}$。算法1最终得到的匿名位置集的形式与tr中的真实位置点的对应关系如表2所示，在没有抑制的情况下，此时基轨迹长度为3。

Table 2. Real location and anonymous location set
表2. 真实位置与匿名位置集

真实位置	匿名位置集
loc_1	$loc_1, loc_1^{d_1}, loc_1^{d_2}, loc_1^{d_3}, \cdots$
loc_2	$loc_2, loc_2^{d_1}, loc_2^{d_2}, loc_2^{d_3}, \cdots$
loc_3	$loc_3, loc_3^{d_1}, loc_3^{d_2}, loc_3^{d_3}, \cdots$

3.2. 假轨迹生成算法

如何将由基于访问概率的匿名位置集生成算法得到的匿名位置集拟合成假轨迹是本节要介绍的算法的关键。假轨迹生成算法如算法2所示。

算法2 假轨迹生成算法

输入：基轨迹tr，匿名位置点$Dlocs$；

输出：假轨迹集$DTrs$；

1) $DTrs\leftarrow\varnothing$;

2) $G\leftarrow\varnothing$;

3) $DDTrs \leftarrow \varnothing$;
4) $V = Dlocs$;
5) for ($i = 0; i < |tr| - 1; i++$)
6) for each location V_{ij} in V_i do
7) for each location $V_{i+1,j}$ in V_{i+1} do
8) insert $E\langle V_{ij}, V_{i+1,j}\rangle$;
9) $G_{ij} \leftarrow \langle V_{ij}, E\langle V_{ij}, V_{i+1,j}\rangle\rangle$;
10) insert G_{ij} into G_i;
11) insert G_i into G;
12) for each location loc_i in $Dlocs_0$ do
13) $DDTrs \leftarrow DFS(G)$;
14) for each tr_i^d in $DDTrs$ do;
15) calculation $dist\langle loc_1^d, loc_n^d\rangle$;
16) if ($dist\langle loc_1^d, loc_n^d\rangle / U_{\max} \le |t_n - t_1|$)
17) insert tr_i^d into $DTrs$;
18) return $DTrs$;

基轨迹可以看作只有一条路径的有向图，而有多条路径的有向图则可遍历出不同的路径。算法 2 首先将匿名位置集建模为一个有向图 G，起点为 loc_1 对应的 $Dloc_1$ 中所有得匿名位置点，表示为 V_{1i}，相邻匿名位置集的位置点间有一条边，表示为 E，轨迹有向图如图 4 所示。在有向图中通过深度优先遍历得到初步的假轨迹集 $DDTrs$，遍历 $DDTrs$ 中的每一条假轨迹，将满足全局时空特性的假轨迹存入 $DTrs$，最终便可得到一个满足全局时空特性的假轨迹集。全局时空特性基于假轨迹的起点至终点的可达性，假轨迹能在 $|t_1 - t_n|$ 内使用基轨迹的最大速度 $U_{\max}$ 从起点到达终点，则说明该条假轨迹满足全局时空特性。

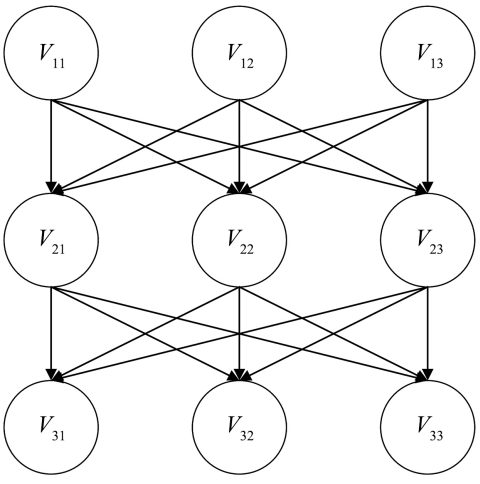

Figure 4. A trajectory directed graph modeled by anonymous location sets
图 4. 由匿名位置集建模的轨迹有向图

3.3. 可供发布的匿名轨迹集生成算法

本节所提出的可供发布的匿名轨迹集生成算法主要思想是将有算法 2 得到的假轨迹集通过特定筛选

方法筛选出 $k-1$ 条假轨迹与基轨迹组成匿名轨迹集。特定的筛选方法主要包含两种。在算法 3 中，首先基于轨迹整体方向的筛选，计算 $DTrs$ 中每一条假轨迹与基轨迹 tr 的夹角，按夹角的升序重新将假轨迹存入 or_TRS；接着是基于轨迹间距的筛选，从 or_TRS 中，按照排序顺序选取 $k-1$ 条与基轨迹的轨迹间距在 (λ,μ) 之间的假轨迹存入 TRS 中，最终返回可供发布的匿名轨迹集 TRS。

算法 3 匿名轨迹集生成算法

输入：基轨迹 tr，假轨迹集 $DTrs$，轨迹间距离阈值 (λ,μ)，匿名等级 k；

输出：可供发布的匿名轨迹集 TRS；

1) $TRS \leftarrow \varnothing$;

2) $or\_TRS \leftarrow \varnothing$;

3) for each tr_i^d in $DTrs$ do

4) calculate $\theta\langle tr, tr_i^d\rangle$;

5) insert tr_i^d into or_TRS;

6) while $|TRS| \neq k-1$ do

7) $temp\_tr =$ the top one from or_TRS;

8) if ($\lambda \le dist\langle tr, temp\_tr\rangle \le \mu$)

9) insert $temp\_tr$ into TRS;

10) update or_TRS;

11) return TRS;

3.4. 算法分析

3.4.1. 算法复杂度分析

假设基轨迹长度为 n，在算法 1 中，时间复杂度的计算主要根据基轨迹 tr 的位置点查找网格，假设网格中的匿名位置点的数量为 $|lattice_i|$，则算法 1 的时间复杂度为 $O(n\times|lattice_i|)$。算法 2 的时间复杂度主要依赖于构建有向图和深度优先遍历图，构建有向图的时间复杂度为 $O\left(n\times|DLocs_i|^2\right)$，图的深度优先遍历的时间复杂度最大为 $O\left(|Dlocs_i|^n\right)$。所以算法 2 时间复杂度为 $O\left(Max\left(n\times|DLocs_i|^2,|DLocs_i|^n\right)\right)$。算法 3 的时间复杂度主要依赖于按轨迹间夹角升序的排序与按轨迹间距的筛选，前者的时间复杂度为 $O\left(|DTrs|^2\right)$，后者只有在最坏的情况下需要花费的时间为 $|DTrs|$；所以算法 3 的时间复杂度为 $O\left(|DTrs|^2\right)$。

3.4.2. 算法安全性分析

定理 1 可供发布的匿名轨迹集满足最低匿名等级 $k_{\min}$。

证明假设基轨迹中有 n 个位置点，匿名位置集的最低匿名需求为 p，匿名等级和最低匿名等级分别为 k 和 $k_{\min}$。在算法 1 中，$p=\sqrt[n]{k}$，此时 p 的定义已为最低匿名需求，而在实际中，每个网格的匿名位置数量超于 p。若在极为稀疏环境下，即 $|lattice_i|<p$，对匿名等级做出调整，又保证 $k\ge k_{\min}$，从而保证 $|lattice_i|\ge p$。综上，可供发布的匿名轨迹集满足最低匿名等级 $k_{\min}$。

4. 实验与分析

本节将从数据发布后的效益需求出发，使用方向相似性、位置信息损失和平均间距作为评价指标来验证本文提出的 TPDTSE 算法的有效性。本文使用的数据集是从 Gowalla 上获的取轨迹数据集，从中选取不同用户在美国内的签到数据作为本次实验数据集。数据集的信息如表 3 所示。本次实验使用 Java 语言在 IDEA 平台上对 TPDTSE 算法加以实现。实验环境为：Intel(R) Core(TM) i5-4570 3.20 GHz，物理内存 16GB，操作系统 Windows 7。

Table 3. Experimental data information
表 3. 实验数据信息

参数	值
轨迹数目	1000
位置点数	8513
最长轨迹	23
最短轨迹	5

4.1. 评价指标

基于轨迹数据发布的主要目的是用于用户行为模式方面的研究，本文将方向相似性、平均间距和位置信息损失率作为评价可供发布的匿名轨迹集的标准，具体如下：

定义 6 位置信息损失 loc_loss [19]若原始轨迹集中基轨迹 tr_i 的长度为 $n_i\left(1\le i\le|T|\right)$，$tr_i$ 经匿名处理后得到的可供发布的轨迹匿名集的平均长度为 $n_i'\left(1\le i\le|T|\right)$，则位置信息损失 loc_loss 的数学定义为式(7)。

$$loc\_loss=\frac{\sum_{i=1}^{|T|}n_i-\sum_{i=1}^{|T|}n_i'}{\sum_{i=1}^{|T|}n_i} \tag{7}$$

定义 7 平均间距 avg_dist 逐条计算基轨迹 tr 与假轨迹 tr^d 间距，则平均间距的计算方法如式(8)所示。

$$avg\_dist=\frac{1}{k-1}\sum_{i=1}^{k-1}dist\_tr\left(tr,tr^{d_i}\right) \tag{8}$$

定义 8 方向相似性 sim_dir 方向相似性是基于定义 5 中的轨迹方向夹角，用来评估基轨迹 tr 与假轨迹 tr^d 的整体方向的相似性，其计算方法如式(9)所示。

$$sim\_dir=\frac{\frac{1}{k-1}\sum_{i=1}^{k-1}\alpha\left(tr,tr^{d_i}\right)}{180} \tag{9}$$

4.2. 结果分析

实验主要从 3 个方面对本文 TPDTSE 算法作对比分析：不同方法对方向相似性、平均间距和位置信息损失情况的影响。参与比较的方法包括安全起点与安全终点算法(记为 SSE) [17]和 DTPP 算法[16]。在测试中，为保证轨迹数据的匿名效果，将 k 与 $k_{\min}$ 的差值设定为 1，网格的大小为 2μ [16]。图 5 所示为 3 种算法在不同 k 值下的轨迹方向相似性情况。DTPP 算法没有具体考虑假轨迹与基轨迹的方向相似性，所以其方向相似性随 k 值的变化没有明显规律，但通过网格化也能将其方向相似性保持在 60%~80%；SSE 算法基于基轨迹的起点和终点，在一定程度上可保证假轨迹与基轨迹有较高的相似度。由图 5 可看出，TPDTSE 算法经匿名处理后得到的假轨迹与基轨迹相比 DTPP 算法和 SSE 算法有较高的方向相似度，且随着的 k 值增大有规律性的变化，提高了轨迹数据匿名后的使用效益。

图 6 所示为 2 种算法在不同 k 值下的平均间距，其中轨迹间距离阈值 (λ,μ) 取值范为(3, 6) [16]，由图 6 可看出 2 种算法经过匿名处理后得到的匿名轨迹集的平均间距变化无明显差别，其中，TPDTSE 算法对应的平均间距的波动幅度为 2.14%~42.04%，DTPP 算法对应的平均间距的波动幅度为 3.16%~13.49%，

前者相较于后者在不同的 k 值下的波动更大，基于 TPDTSE 算法经匿名处理后的得到匿名集有较高的方向相似性的前提下，平均间距波动幅度大，也就意味着 TPDTSE 算法生成的匿名集具有较高相似性的同时又具有一定的多样性，使攻击者难以识别出基轨迹。

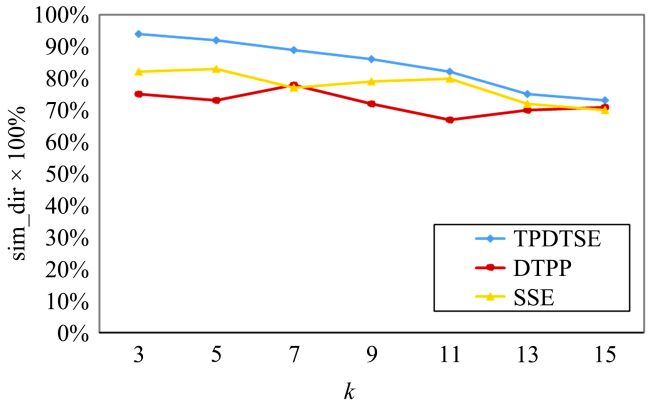

Figure 5. Directional similarity of different methods
图 5. 不同方法的方向相似性

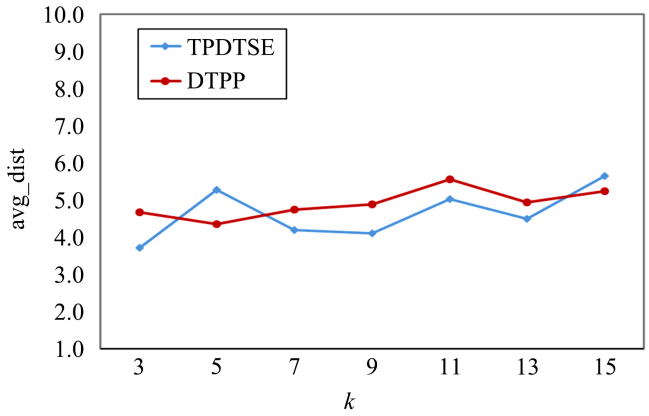

Figure 6. Average distance of different methods
图 6. 不同方法的平均间距

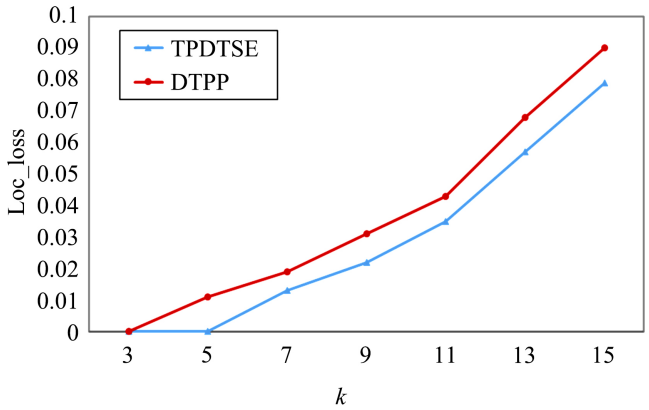

Figure 7. Location information loss by different methods
图 7. 不同方法的位置信息损失

图 7 为 2 种算法在不同 k 值下的位置信息损失情况。由图可以看出，TPDTSE 算法和 DTPP 算法的位置信息损失情况均随着 k 的增大而增加，TPDTSE 考虑轨迹所处稀疏环境时，通过最低匿名等级 k_{min} 避免了直接抑制有效减少了位置信息损失。

5. 总结

本文针对现有的基于 k-anonymity 的轨迹数据隐私保护算法没有充分考虑移动对象所处地理环境及轨迹整体方向等特性可能导致匿名后的数据可用性较低的问题，提出了一种稀疏环境下基于假轨迹的轨迹隐私保护算法(TPDTSE)。TPDTSE 算法考虑了移动对象所处的环境并将轨迹整体方向及轨迹间距作为衡量轨迹相似性的重要指标，基于此对轨迹进行匿名化处理，实验结果表明，TPDTSE 算法满足轨迹匿名需求的同时减少了位置信息的丢失，提高了轨迹数据的可用性。如何平衡数据匿名与数据发布后的可用性一直是轨迹隐私保护算法的研究热点，在接下来的工作中，将进一步研究各类轨迹隐私保护算法，使经匿名处理后的轨迹数据发布后具有更高的使用效益。

基金项目

广州市科技计划项目(201907010025)。

参考文献

[1] Zhang, S.B., Mao, X.J., Choo, K.K.R., Peng, T. and Wang, G.J. (2020) A Trajectory Privacy-Preserving Scheme Based on a Dual-K Mechanism for Continuous Location-Based Services. *Information Sciences*, **527**, 406-419. https://doi.org/10.1016/j.ins.2019.05.054

[2] Feng, Z.N. and Zhu, Y.M. (2017) A Survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access*, **4**, 2056-2067. https://doi.org/10.1109/ACCESS.2016.2553681

[3] Guo, L.M., Huang, G.Y., Gao, X., He, J., Wu, B. and Guo, H.M. (2015) DoSTra: Discovering Common Behaviors of Objects Using the Duration of Staying on Each Location of Trajectories. *Radiation Research*, **161**, 137-142.

[4] Huang, G.Y., He, J., Zhou, W.L., *et al*. (2016) Discovery of Stop Regions for Understanding Repeat Travel Behaviors of Moving Objects. *Journal of Computer & System Sciences*, **82**, 582-593. https://doi.org/10.1016/j.jcss.2015.10.006

[5] Shen, Y., Zhao, L.G. and Fan, J. (2015) Analysis and Visualization for Hot Spot Based Route Recommendation Using Short-Dated Taxi GPS Traces. *Information*, **6**, 134-151. https://doi.org/10.3390/info6020134

[6] Chun-Guang, M.A., Zhang, L. and Yang, S.T. (2015) Review on Location Trajectory Privacy Protection. *Netinfo Security*, **10**, 24-31.

[7] Wernke, M., Skvortsov, P., Duerr, F., *et al*. (2014) A Classification of Location Privacy Attacks and Approaches. *Personal & Ubiquitous Computing*, **18**, 163-175. https://doi.org/10.1007/s00779-012-0633-z

[8] 霍峥, 孟小峰. 轨迹隐私保护技术研究[J]. 计算机学报, 2011, 34(10): 1820-1830.

[9] 雷凯跃, 李兴华, 刘海, 裴卓雄, 马建峰, 李晖. 轨迹发布中基于时空关联性的假轨迹隐私保护方案[J]. 通信学报, 2016, 37(12): 156-164.

[10] Gruteser, M. and Grunwald, D. (2003) Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking. *Proceedings of the* 1*st International Conference on Mobile Systems*, *Applications and Services*, San Francisco, 5-8 May 2003, 31-42. https://doi.org/10.1145/1066116.1189037

[11] Abul, O., Bonchi, F. and Nanni, M. (2008) Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. 2008 *IEEE* 24*th International Conference on Data Engineering*, Cancun, 7-12 April 2008, 376-385. https://doi.org/10.1109/ICDE.2008.4497446

[12] Poulis, G., Skiadopoulos, S., Loukides, G. and Gkoulalas-Divanis, A. (2013) Distance-Based k^m-Anonymization of Trajectory Data. 2013 *IEEE* 14*th International Conference on Mobile Data Management*, Milan, 3-6 June 2013, 57-62. https://doi.org/10.1109/MDM.2013.66

[13] Wu, X.C. and Sun, G.Z. (2015) A Novel Dummy-Based Mechanism to Protect Privacy on Trajectories. 2014 *IEEE International Conference on Data Mining Workshop*, Shenzhen, 14 December 2014, 1120-1125. https://doi.org/10.1109/ICDMW.2014.122

[14] 雷凯跃, 李兴华, 刘海, 裴卓雄, 马建峰, 李晖. 轨迹发布中基于时空关联性的假轨迹隐私保护方案[J]. 通信学

报, 2016, 37(12): 156-164.

[15] Terrovitis, M., Poulis, G., Mamoulis, N., *et al*. (2017) Local Suppression and Splitting Techniques for Privacy Preserving Publication of Trajectories. *IEEE Transactions on Knowledge & Data Engineering*, **29**, 1466-1479. https://doi.org/10.1109/TKDE.2017.2675420

[16] 刘向宇, 陈金梅, 夏秀峰, Singh, M., 宗传玉, 朱睿. 防止暴露位置攻击的轨迹隐私保护[J]. 计算机应用, 2020, 40(2): 479-485.

[17] Zhao, Y.N., Luo, Y.L., Yu, Q.Y., *et al*. (2020) A Privacy-Preserving Trajectory Publication Method Based on Secure Start-Points and End-Points. *Mobile Information Systems*, 2020, Article ID: 3429256. https://doi.org/10.1155/2020/3429256

[18] 许华杰, 吴青华, 胡小明. 基于轨迹多特性的隐私保护算法[J]. 计算机科学, 2019, 46(1): 190-195.

[19] 陈传明, 林文诗, 俞庆英, 罗永龙. 一种基于单点收益的轨迹隐私保护方法[J]. 电子学报, 2020, 48(1): 143-152.