doi:10.11959/j.issn.1000-436x.2015053

基于图划分的个性化轨迹隐私保护方法

杨静，张冰，张健沛，谢静
（哈尔滨工程大学 计算机科学与技术学院，黑龙江 哈尔滨 150001）

摘 要：针对用户对轨迹匿名数据的个性化需求，提出一种基于轨迹间夹角和位置重合的(s,λ)-覆盖个性化轨迹间关联构建方法，并根据轨迹间距离和方向度量轨迹间边权，以构造规模可变的个性化轨迹图模型。同时，将轨迹 k-匿名集的构建转化为轨迹图划分问题，提出了一种基于贪心策略寻找近似最优的 k 条轨迹构建轨迹 k-匿名集的方法。通过对比实验，在合成轨迹数据集上验证了所提算法的有效性和合理性。
关键词：隐私安全；个性化；图划分；(s,λ)-覆盖；轨迹间夹角
中图分类号：TP309.2 **文献标识码**：A

Personalized trajectory privacy preserving method based on graph partition

YANG Jing, ZHANG Bing, ZHANG Jian-pei, XIE Jing
(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: A (s,λ)-overlap was proposed which based on the angle and location between two trajectories to construct the relationship between them, and also measure the weight between two trajectories by trajectory angle and distance, together to construct the personal trajectory graph model. Then a greedy partition method to was proposed construct trajectory k-anonymization sets by finding the approximate optimal k trajectories. Experiment results on synthetic dataset show the effectiveness and reasonableness of proposed method.
Key words: privacy security; personalized; graph partition; (s,λ)-overlap; trajectory angel

1 引言

随着无线定位技术和移动设备的发展，基于位置的服务（LBS, location based services）已逐渐深入人们的日常生活。LBS 是由位置服务供应商为用户提供基于用户当前位置的增值服务[1]，用户可主动获取路线导航、商业搜索等服务，也可接受附近的商家提供的广告服务。由服务器搜集的用户位置和轨迹数据含有丰富的时空信息，对其分析和挖掘能够更好地帮助科研单位理解用户行为模式，也能够有效地帮助位置服务供应商制定个性化的用户服务。然而，用户的位置和轨迹数据中含有大量的隐私内容，如用户的家庭地址、单位地址、个人偏好及用户的反常行为，用户在享受 LBS 提供便捷服务的同时，也面临着极大的隐私泄露威胁。

近年来，用户位置和轨迹数据的隐私安全逐渐被人们所关注[2~3]。攻击者可通过用户间交流、相互观察和发布数据泄露等途径获得目标对象的位置信息[4]，对用户位置造成隐私威胁。文献[5]在用户请求服务时，采取将用户真实位置转移的方法来保护用户隐私。文献[6]将关系数据中的 k-匿名模型[7]引入位置隐私保护中，以至少覆盖当前用户和其周边 $k-1$ 个用户位置的区域来替代用户真实位置，保护用户的位置隐私。随后，人们基于位置 k-匿名的思

收稿日期：2013-09-18；**修回日期**：2013-11-25
基金项目：国家自然科学基金资助项目（61370083，61073043，61073041）；高等学校博士学科点专项科研基金资助项目（20112304110011，20122304110012）；哈尔滨市科技创新人才研究专项基金资助项目（2011RFXXG015）
Foundation Items: The National Natural Science Foundation of China (61370083, 61073043, 61073041); The National Research Foundation for the Doctoral Program of Higher Education of China (20112304110011,20122304110012); The Harbin Outstanding Academic Leader Foundation of Heilongjiang Province of China (2011RFXXG015)

想提出了轨迹 k-匿名[8,9]的移动轨迹隐私保护方法。轨迹 k-匿名要求每条移动轨迹至少与其他 $k-1$ 条轨迹不可区分，以保护轨迹信息的隐私安全。

轨迹匿名方法的最大难点在于如何在保证轨迹数据匿名化后的隐私安全和数据质量的前提下，构建轨迹匿名集合。目前，主要通过基于泛化的方法和基于聚集的方法 2 种途径实现轨迹匿名隐私保护[10]。基于泛化的轨迹匿名方法[11]首先将轨迹集合中的轨迹泛化到多个 k-匿名集中，然后通过在 k-匿名集中重新构建 k 条轨迹，以实现轨迹信息的隐私保护。基于聚集的轨迹匿名方法[12]需要进行轨迹预处理、轨迹聚集和空间转换 3 个步骤。其中，轨迹预处理阶段需将轨迹同步化处理并生成轨迹等价类，轨迹聚集阶段将符合一定条件的轨迹聚集在一起生成轨迹 k-匿名集，最后空间转换阶段将每个 k-匿名集中的轨迹以聚集中心的形式发布。然而，以上方法在实现轨迹匿名保护时并未限制轨迹 k-匿名集的空间规模和轨迹间的关系，发布的匿名轨迹数据可用性较低。随着研究的不断深入，轨迹 k-匿名集构建的优化方法引起了学者们的重视。文献[13]使用图模型表示移动轨迹间的关系，并通过图划分方法构建轨迹匿名集合，但其在构建轨迹关联时仅考虑移动轨迹在位置上的关联，并未考虑轨迹的方向性。文献[14]通过在方向相同的轨迹间建立关联，并以轨迹间距离和方向来分析轨迹间权重来构建轨迹图模型。

目前，已有的轨迹匿名方法大多忽略了敏感信息的个性化需求问题，而现实生活中的不同轨迹所需的隐私保护程度不同，为所有轨迹制定较高的隐私保护等级会造成大量的信息损失，而为所有轨迹制定较低的隐私保护等级则无法满足高度敏感轨迹的隐私保护需求。因此，本文在个性化服务的前提下，研究基于图模型的轨迹隐私保护问题。本文的主要贡献如下。

1) 提出了一种基于轨迹间夹角和轨迹间位置重合的(s,λ)-覆盖个性化轨迹关联构建方法，并根据轨迹间距离和方向度量轨迹间边权，针对用户对轨迹匿名的个性化需求构造规模可变的个性化轨迹图模型。

2) 将轨迹 k-匿名集的构建转化为轨迹图划分问题，使用贪心策略寻找近似最优的 k 条轨迹构建轨迹 k-匿名集，以实现匿名轨迹数据的隐私保护和数据可用性间的平衡。

2 相关概念

定义 1 轨迹。移动对象 O 的轨迹 T 由 O 的身份信息和一组时空坐标构成，记作

$$T=\{ID,(x_1,y_1,t_1),(x_2,y_2,t_2),\cdots,(x_n,y_n,t_n)\}$$

其中，ID 为 O 的身份标识，$t_1<t_2<\cdots<t_n$，(x_i,y_i)为 O 在时刻 t_i 的位置坐标。

定义 2 同步轨迹[13]。T_p 和 T_q 为 2 条不同的轨迹，如果 T_p 和 T_q 具有相同的时间坐标序列，那么 T_p 和 T_q 是同步的；如果轨迹集合 S 中的轨迹两两同步，那么轨迹集合 S 是同步的。

由于移动终端对移动对象运行轨迹的采样时间不同，同时移动对象向服务器请求服务的时间也不同，现实生活中的轨迹通常是不同步的。本文假设轨迹在 2 个样本时间点内做匀速直线运动，对于轨迹 T_p 和 T_q 间不同的时间坐标，通过在轨迹中插入相应时刻的位置坐标来实现轨迹 T_p 和 T_q 的同步。

定义 3 轨迹同步化。T_p 和 T_q 是时间坐标序列为 S_p 和 S_q 的 2 条不同步的轨迹，设轨迹在每个时间间隔内做匀速直线运动，轨迹的同步化过程如下。

1) 如果$\exists t_p\in S_p$ 且 $t_p\notin S_q$，则计算 t_p 时刻轨迹 T_q 的位置坐标(x_q,y_q)，$T_q=(x_q,y_q,t_p)\cup T_q$。

2) 如果$\exists t_q\in S_q$ 且 $t_q\notin S_p$，则计算 t_q 时刻轨迹 T_p 的位置坐标(x_p,y_p)，$T_p=(x_p,y_p,t_q)\cup T_p$。

例如，图 1 为二维欧式空间中的 2 条不同轨迹 T_p 和 T_q，其中 T_p 的时间坐标序列为$(t_1,t_3,\cdots,t_8)$，T_q 的时间坐标序列为$(t_1,\cdots,t_3,t_5,\cdots,t_8)$，通过在 T_p 中插入 t_2 时刻对应的坐标，在 T_q 中插入 t_4 时刻对应的坐标来实现轨迹 T_p 和 T_q 的同步化。

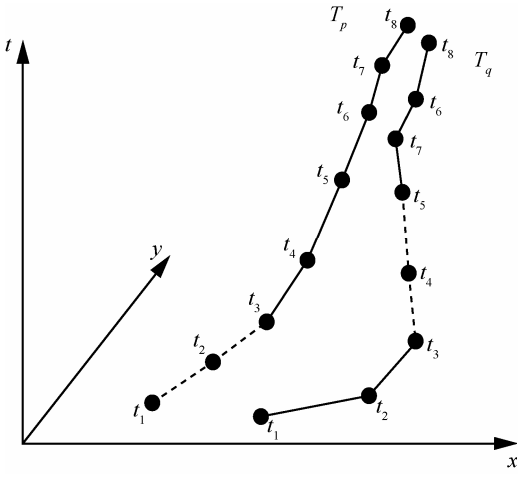


图 1 轨迹同步化构造

定义 4　轨迹间夹角。T_p 和 T_q 为 2 条同步的轨迹，$T_p=(ID_p,(x_p^1,y_p^1,t_1),(x_p^2,y_p^2,t_2),\cdots,(x_p^n,y_p^n,t_n))$，$T_q=(ID_q,(x_q^1,y_q^1,t_1),(x_q^2,y_q^2,t_2),\cdots,(x_q^n,y_q^n,t_n))$，$T_p$ 和 T_q 在时间段$[t_i,t_{i+1}]$的移动向量分别为 $\overrightarrow{p_{[i,i+1]}}$ 和 $\overrightarrow{q_{[i,i+1]}}$，则$[t_i,t_{i+1}]$时间内轨迹 T_p 和 T_q 间的夹角余弦定义为

$$\cos\theta_{[t_i,t_{i+1}]}=\frac{\overrightarrow{p_{[t_i,t_{i+1}]}}\overrightarrow{q_{[t_i,t_{i+1}]}}}{\left|\overrightarrow{p_{[t_i,t_{i+1}]}}\right|\left|\overrightarrow{q_{[t_i,t_{i+1}]}}\right|}$$

$$=\frac{\left(x_p^{i+1}-x_p^i\right)\left(x_q^{i+1}-x_q^i\right)+\left(y_p^{i+1}-y_p^i\right)\left(y_q^{i+1}-y_q^i\right)}{\sqrt{\left(x_p^{i+1}-x_p^i\right)^2+\left(y_p^{i+1}-y_p^i\right)^2}\sqrt{\left(x_q^{i+1}-x_q^i\right)^2+\left(y_q^{i+1}-y_q^i\right)^2}}$$

轨迹 T_p 和 T_q 间的夹角余弦定义为

$$\cos(T_p,T_q)=\frac{1}{n-1}\sum_{i=1}^{n-1}\cos\theta_{[t_i,t_{i+1}]}$$

定义 5　轨迹的距离。T_p 和 T_q 为 2 条同步的轨迹，$T_p=(ID_p,(x_p^1,y_p^1,t_1),(x_p^2,y_p^2,t_2),\cdots,(x_p^n,y_p^n,t_n))$，$T_q=(ID_q,(x_q^1,y_q^1,t_1),(x_q^2,y_q^2,t_2),\cdots,(x_q^n,y_q^n,t_n))$，$T_p$ 和 T_q 间的距离定义为

$$dis(T_p,T_q)=\frac{\sum_{i=1}^{n}\sqrt{\left(x_p^i-x_q^i\right)^2+\left(y_p^i-y_q^i\right)^2}}{n}$$

定义 6　轨迹 k-匿名。D 为移动轨迹的集合，当且仅当 D 的匿名集合 D'中的每条轨迹都至少与其他 k–1 条轨迹不可区分时，匿名集合 D'是满足轨迹 k-匿名的。

定义 7　匿名域。$T_1,T_2,\cdots,T_k$ 为 k 条样本时间为$(t_1,t_2,\cdots,t_n)$的同步的轨迹，这 k 条轨迹在 t_i 时刻的匿名域定义为 $x=\min(x_{T_j}^{t_i})$、$x=\max(x_{T_j}^{t_i})$、$y=\min(y_{T_j}^{t_i})$ 和 $y=\max(y_{T_j}^{t_i})$ 4 条直线构成的矩形 $S(t_i)(1\leqslant i\leqslant n, 1\leqslant j\leqslant k)$，其中，$\min(x_{T_j}^{t_i})$、$\max(x_{T_j}^{t_i})$ 分别为这 k 条轨迹在 x 轴位置坐标的最小值与最大值，$\min(y_{T_j}^{t_i})$、$\max(y_{T_j}^{t_i})$ 分别这 k 条轨迹在 y 轴位置坐标的最小值与最大值；轨迹 $T_1,T_2,\cdots,T_k$ 在样本时间$(t_1,t_2,\cdots,t_n)$的匿名域 S 定义为 k 条轨迹在 n 个时刻的匿名域之和，即 $S=S(t_1)\cup S(t_2)\cup\cdots\cup S(t_n)$。

图 2 为 k=4 时 4 条同步的轨迹形成的匿名域。轨迹 k-匿名模型能够保证攻击者最高以 1/k 的概率正确推测出轨迹的隐私信息。一般地，阈值 k 越大，说明匿名域中轨迹的数量越多，轨迹产生的匿名域的面积越大，发布的数据隐私保护程度越高，但数据的可用性却越差，反之亦然。

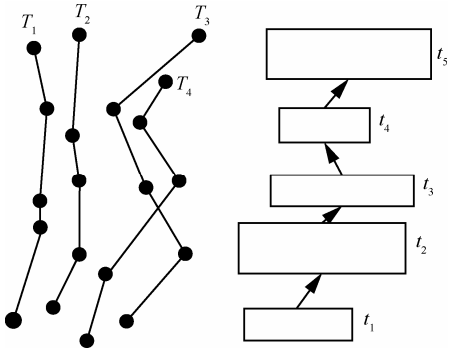


图 2　轨迹 4-匿名

定理 1　$T_1,T_2,\cdots,T_k$ 为 k 条时间坐标序列为$(t_1,t_2,\cdots,t_n)$的同步轨迹，轨迹 $T_1,T_2,\cdots,T_k$ 在样本时间$(t_1,t_2,\cdots,t_n)$的匿名域 S 能够覆盖这 k 条轨迹的所有位置坐标。

证明　根据定义 7，设 t_i (i=1,2,$\cdots$,n)时刻 k 条轨迹的匿名域为 $S(t_i)$，为 $x=\min(x_{T_j}^{t_i})$、$x=\max(x_{T_j}^{t_i})$、$y=\min(y_{T_j}^{t_i})$ 和 $y=\max(y_{T_j}^{t_i})$ 4 条直线构成的矩形，其中，$\min(x_{T_j}^{t_i})$、$\max(x_{T_j}^{t_i})$ 分别为这 k 条轨迹在 x 轴位置坐标的最小值与最大值，$\min(y_{T_j}^{t_i})$、$\max(y_{T_j}^{t_i})$ 分别这 k 条轨迹在 y 轴位置坐标的最小值与最大值。对于轨迹 $T_1,T_2,\cdots,T_k$ 在 t_i 时刻的每一个坐标，即 $\forall(x_{T_j}^{t_i},y_{T_j}^{t_i},t_i)\in T_j$ (j=1,2,$\cdots$,k)，都有 $x_{T_j}^{t_i}\in[\min(x_{T_j}^{t_i}),\max(x_{T_j}^{t_i})]$，$y_{T_j}^{t_i}\in[\min(y_{T_j}^{t_i}),\max(y_{T_j}^{t_i})]$。因此，$(x_{T_j}^{t_i},y_{T_j}^{t_i},t_i)\in S(t_i)$，即 t_i 时刻的匿名域 $S(t_i)$能够覆盖 k 条轨迹在 t_i 时刻的所有位置坐标。由于 $S=S(t_1)\cup S(t_2)\cup\cdots\cup S(t_n)$，因此，匿名域 S 能够覆盖这 k 条轨迹在样本时间$(t_1,t_2,\cdots,t_n)$的所有位置坐标。

3　基于图划分的个性化轨迹匿名方法

3.1　轨迹预处理

轨迹预处理阶段的任务是构建具有相近的起始时刻和结束时刻的轨迹等价类，并生成轨迹等价类的距离矩阵。文献[14]使用集合中轨迹的最大起始时刻和最小结束时刻作为轨迹等价类的时间约束，但如果轨迹间的起始时刻或结束时刻的差异很大，构建轨迹等价类时会损失很多轨迹信息，轨迹匿名时也会产生较大的信息损失。为确保轨迹间的高相似性，降低轨迹匿名时产生的信息损失，本文将起始时刻与结束时刻在给定时间间隔内的轨迹同步化，以构造轨迹等价类。

定义 8　轨迹等价类。若轨迹 $T_1,T_2,\cdots,T_n$(n>0)

满足：1) $\forall T_i \in \{T_1,T_2,\cdots,T_n\}$，轨迹 T_i 的起始时刻 $t_s^i \in [t_s-\Delta t, t_s+\Delta t]$，轨迹结束时刻 $t_e^i \in [t_e-\Delta t, t_e+\Delta t]$；2)轨迹 $T_1,T_2,\cdots,T_n$ 是同步的；则它们属于同一个轨迹等价类。

例如，设起始时刻 t_s 为 9:00，结束时刻 t_e 为 10:00，Δt 为 5 min，查找所有起始时刻在[8:55,9:05]，结束时刻在[9:55,10:05]的轨迹，并将它们同步化以构造等价类。起始时刻和结束时刻间的间隔取决于轨迹采样时刻的稀疏程度，如果轨迹的采样间隔很短，时间坐标密集，则可适当减少起始时刻和结束时刻的间隔；如果轨迹的采样间隔较长，时间坐标稀疏，则可适当增加起始时刻和结束时刻的间隔。

定义 9 轨迹等价类的距离矩阵。轨迹等价类 *TEC* 的距离矩阵 ***TDM*** 定义为对角线元素为 0 的 n 阶对称阵

$$\boldsymbol{TDM}=\begin{pmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & 0 \end{pmatrix}$$

其中，d_{ij} 为轨迹 T_i 和 T_j 间的距离。轨迹等价类的距离矩阵构造算法如算法 1 所示。

算法 1 轨迹等价类的距离矩阵构造算法(TDMCons)

输入：轨迹等价类 $TEC=T_1,T_2,\cdots,T_n$

输出：轨迹等价类的距离矩阵 ***TDM***，轨迹间最大距离 max，轨迹间最小距离 min

1) $TDM=\varnothing$；
2) for(i=1:n)
3) for(j=i+1:n)
4) 计算轨迹 T_i 和 T_j 间距离 $dis(T_i,T_j)$；
5) $TDM_{ii}=0$；
6) $TDM_{ij}=TDM_{ji}=dis(T_i,T_j)$；
7) 寻找 ***TDM*** 中的最大值 max 和最小值 min；
8) 返回 ***TDM***、max、min；

算法 1 的步骤 1)~6)为轨迹等价类的距离矩阵的构造，设轨迹由 m 个时空坐标构成，计算 2 条轨迹距离时需计算 k 次时空坐标间的距离，距离矩阵的构造共需计算 $n(n-1)/2$ 次轨迹间距离，此过程可在时间 $mO(n^2)$内完成；步骤 7)为查找距离矩阵中最大距离和最小距离，可在时间 $O(n\log n)$内完成。因此，算法 1 可在时间 $mO(n^2)$内完成。

3.2 个性化轨迹图构造

轨迹间相似性由轨迹的夹角和重合点的数目决定。轨迹间夹角越小，重合点越多，轨迹间相似性越高；轨迹间夹角越大，重合点越少，轨迹间相异性越大。根据用户对轨迹相似性的不同需求，本节利用轨迹间夹角和轨迹重合点的数目来控制轨迹图的规模变化，实现个性化的轨迹图构建。

定义 10 (s,λ)-覆盖。$T_p=(ID_p,(x_p^1,y_p^1,t_1),(x_p^2,y_p^2,t_2),\cdots,(x_p^n,y_p^n,t_n))$，$T_q=(ID_q,(x_q^1,y_q^1,t_1),(x_q^2,y_q^2,t_2),\cdots,(x_q^n,y_q^n,t_n))$是 2 条同步的轨迹，如果轨迹 T_p 和 $T_q(p\neq q)$是(s,λ)-覆盖的，则 T_p 和 T_q 满足：

1) $\lambda(0\leqslant\lambda\leqslant\pi/2)$为轨迹 T_p 和 T_q 间夹角的约束，T_p 和 T_q 间夹角 $\theta_{pq}\leqslant\lambda$，$\cos(T_p,T_q)\in[\cos\lambda,1]$；

2) $s(s>0)$为 T_p 的时空坐标重合数量的约束，存在不少于 s 个$(x_p^i,y_p^i)\in T_p(i=1,2,\cdots,n)$，$x_p^i\in[\min(x_q),\max(x_q)]$或 $y_p^i\in[\min(y_q),\max(y_q)]$，其中，$\min(x_q)$和 $\max(x_q)$分别为轨迹 T_q 在 x 轴上位置坐标的最小值和最大值，$\min(y_q)$和 $\max(y_q)$分别为轨迹 T_q 在 y 轴上位置坐标的最小值和最大值。

(s,λ)-覆盖描述了轨迹间的时空关系，可通过 s 和 λ 值来调整轨迹(s,λ)-覆盖的范围和轨迹图的规模变化，实现个性化轨迹图的构建。

定义 11 轨迹图[13]。轨迹等价类 $TEC=\{T_1,T_2,\cdots,T_n\}$，*TEC* 的轨迹图 $TG=(V,E,W)$是一个带权无向图，满足：

1) V 为 *TG* 节点的集合，$\forall v_i\in V$ 都对应等价类中的一条轨迹 $T_i(1\leqslant i\leqslant n)$；

2) E 为 *TG* 边的集合，如果轨迹 T_i 和 $T_j(1\leqslant i\leqslant n, 1\leqslant j\leqslant n, i\neq j)$是$(s,\lambda)$-覆盖的，则节点 v_i 和 v_j 间存在一条边 e_{ij}；

3) $\boldsymbol{W}$ 为 n 阶对称矩阵，其中，$w_{ij}(1\leqslant i,j\leqslant n, i\neq j)$为 e_{ij} 的边权。

轨迹图的边权反映出相应轨迹间的相异度，边权越高，轨迹间差别越大。轨迹间的夹角和距离都能够反映出轨迹间的差别，用户对轨迹方向相同的需求越高，满足需求的轨迹间距离越大，产生的匿名域越大，轨迹的隐私保护程度越高；用户对轨迹距离相近的需求越高，产生的匿名域越小，轨迹数据的可用性越高。文献[14]首次提出使用轨迹间夹角和距离度量边权的方法，但其未标准化轨迹间距离，会出现距离对权重取值起决定性作用的情况，对参数的设置存在通用性和合理性的问题。因此，本文基于用户对轨迹方向和距离的需求，即隐私保

护程度和数据可用性的需求，提出一种基于标准化轨迹间距离，使用方向特征 α 和距离特征 β 调整轨迹间边权的个性化边权度量方法。

定义 12　轨迹图的边权。轨迹图 $TG=(V,E,W)$，轨迹 T_p、$T_q\in V$，如果边$(T_p,T_q)\in E$，则 T_p 和 T_q 间的边权 w_{pq} 定义为

$$w_{pq}=\begin{cases}\alpha(1-\cos\theta)+\beta\dfrac{dis(T_p,T_q)-\min}{\max-\min}, \min\neq\max\\ \alpha(1-\cos\theta), \min=\max\end{cases}$$

其中，$\alpha,\beta\in[0,1]$ 且 $\alpha+\beta=1$，θ 为轨迹 T_p 和 T_q 间的夹角，$dis(T_p,T_q)$为轨迹 T_p 和 T_q 间的距离，min 和 max 分别为等价类中轨迹间的最小距离和最大距离。

定理 2　等价类中任意 2 条(s,λ)-覆盖的轨迹 T_p 和 T_q 间边权 w_{pq} 的取值范围为$[0,\alpha(1-\cos\lambda)+\beta]$。

证明　根据定义 10，由于轨迹 T_p 和 T_q 是(s,λ)-覆盖的，设 θ 为轨迹 T_p 和 T_q 间的夹角，则 $\cos\theta\in[\cos\lambda,1]$，因此，$\alpha(1-\cos\theta)\in[0,\alpha(1-\cos\lambda)]$。由定理 2 条件，$T_p$ 和 T_q 是等价类中的 2 条轨迹，T_p 和 T_q 间的距离为 $dis(T_p,T_q)$，设 min 和 max 分别为等价类中 T_p 和 T_q 间的最小距离和最大距离，$dis=\dfrac{dis(T_p,T_q)-\min}{\max-\min}$。当 min=max 时，$dis=0$；当 min≠max 时，$0\leqslant dis\leqslant 1$。因此，$dis$ 的取值范围为[0,1]。

根据定义 12

$$w_{pq}=\begin{cases}\alpha(1-\cos\theta)+\beta\dfrac{dis(T_p,T_q)-\min}{\max-\min}, \min\neq\max\\ \alpha(1-\cos\theta), \min=\max\end{cases}$$

当 min≠max 时，$w_{pq}=\alpha(1-\cos\theta)+\beta dis$，此时 $w_{pq}\in[0,\alpha(1-\cos\lambda)+\beta]$；当 min=max 时，$w_{pq}=\alpha(1-\cos\theta)$，此时 $w_{pq}\in[0,\alpha(1-\cos\lambda)]$。

综上，$w_{pq}\in[0,\alpha(1-\cos\lambda)+\beta]$，特别地，当 $\lambda=\pi/2$ 时，$w_{pq}\in[0,1]$。

根据移动对象对发布轨迹数据的隐私保护程度和数据可用性的不同需求，本节提出一种使用方向特征 α 和距离特征 β 调整轨迹间边权的个性化边权度量方法，基本思想为：首先，分别计算 2 条轨迹间每个相应时间段内的轨迹向量间的夹角，并根据定义 4 计算轨迹间的夹角；然后，在轨迹等价类的距离矩阵中查找相应轨迹间距离，并根据定义 11 计算轨迹间权重。个性化轨迹间边权度量算法如算法 2 所示。

算法 2　个性化轨迹间权重度量算法(ITWCons)

输入：轨迹 T_p 和 $T_q(p\neq q)$，轨迹等价类的距离矩阵 ***TDM***，轨迹间最大距离 max，轨迹间最小距离 min，方向特征 α，距离特征 β

输出：T_p 和 T_q 间权重 w_{pq}

1) $s_{pq}=0$;
2) for(i=1:m−1)
3)　$s_{pq}=s_{pq}+\cos(T_{p[i,i+1]},T_{q[i,i+1]})$;
4) $s_{pq}=s_{pq}/m-1$;
5) if(max≠min)
6)　在 TDM 中查找 $dis(T_p,T_q)$;
7)　$w_{pq}=\alpha(1-s_{pq})+\beta\dfrac{dis(T_p,T_q)-\min}{\max-\min}$;
8) else $w_{pq}=\alpha(1-s_{pq})$;
9) 返回 w_{pq}。

设轨迹由 m 个时空坐标构成，算法 2 的步骤 2)~4)在计算 2 条轨迹间夹角时共需计算 $m-1$ 次相应区间的夹角，此过程可在时间$(m-1)O(1)$内完成；步骤 5)~9)为轨迹间权重的计算，可在时间 $O(n)$内完成；因此，算法 2 可在时间 $mO(1)+O(n)$内完成。综上，轨迹间权重的计算首先需时间 $mO(n^2)$构造轨迹等价类的距离矩阵，然后，每 2 条轨迹间的权重可在线性时间内完成。

本节给出一种个性化的轨迹图构造方法，其基本思想为：首先任取一条轨迹加入轨迹图 TG 的节点集合 V 中，剩余的轨迹加入节点集合 V'中；任取 V'中的一条轨迹 T_j，分别判断 T_j 与 V 中的每条轨迹 T_i是否为(s,λ)-覆盖的，如果 T_j 与 T_i 是(s,λ)-覆盖的，则将 T_j 加入集合 V 中，并计算 T_i 和 T_j 间的权重，将(T_i,T_j,w_{ij})加入边集合 E 中；重复此过程，直到集合 V'为空为止，并返回轨迹图 TG。具体的算法描述如算法 3 所示。

算法 3　个性化轨迹图构造算法(ITGCons)

输入：轨迹等价类 $TEC=T_1,T_2,\cdots,T_n$，夹角约束 λ、位置重合约束 s、方向特征 α、距离特征 β

输出：轨迹图 $TG=(V,E,W)$

1) TDMCons(TEC);
2) $V=\{T_1\},E=\varnothing,W=\varnothing$;
3) $V'=TEC-V$;
4) while($V'\neq\varnothing$)
5)　for(V 中的每个节点 T_i)
6)　$\forall T_j\in V'$;
7)　if(T_i 和 T_j 是(s,λ)-覆盖的)

8) w_{ij}= ITWCons(T_i,T_j, α, β);
9) E=E∪(T_i,T_j,w_{ij});
10) else w_{ij}=∞;
11) V=V∪T_j;
12) V'=V'−{T_j};
13) 返回 TG=(V,E,W)。

算法 3 的步骤 1 首先计算等价类的距离矩阵，设轨迹由 m 个时空坐标构成，步骤 1 需时间 $mO(n^2)$；步骤 2)~3)首先构造初始节点集合 V 和剩余节点集合 V'，可在 $O(n)$的时间内完成；步骤 5)~12)为轨迹图的构造，每执行一次循环需判断|V|*|V'|次(s, λ)-覆盖，并至多计算|V|*|V'|轨迹间权重，故步骤 5)~12)至多需时间 $O(n^2)$；最后，步骤 13)返回轨迹图 TG。综上，整个算法的个性化轨迹图构造过程可在 $mO(n^2)$的时间内完成。

3.3 轨迹匿名算法

定义 13 k-子图划分[13]。轨迹图 TG 的 k-子图划分 φ 将 TG 的节点集合划分为互不相交的多个子集，TG= S_1∪…∪S_n∪D_1∪…∪D_m，子集 S_i 和 D_i 的节点数目分别为|S_i|和|D_i|，满足 k≤|S_i|<2k，1≤|D_i|<k，且子集 S_i 的相应子图的节点间边权和较小。

定义 14 k-匿名集。φ 是轨迹图 TG 的一个 k-子图划分，TG= S_1∪…∪S_n∪D_1∪…∪D_m，若∀S_i∈TG，k≤|S_i|<2k，则 S_i 称作 TG 的一个 k-匿名集。

定义 15 合并代价。G_1 和 G_2 为轨迹图 TG 的 2 个子图，G_1∩G_2 =∅，若∃v∈G_1，v'∈G_2，v 和 v' 在 TG 中是相关联的且边(v,v')的边权为 $w_{vv'}$，则图 G_1 和 G_2 的合并代价为

$$\mathrm{Com}(G_1,G_2)=\min(w_{vv'})$$

本节给出一种基于轨迹 k-子图划分的个性化轨迹匿名方法，基本思想为：对于轨迹图中规模不小于 k 的连通分量，首先，将权重最小的边对应的定点加入集合 V'中；然后，寻找与 V'中节点相关联且不包含于 V'的权重最小的节点，加入 V'中，不断重复此过程，直至 V'中含有 k 个节点为止；最后，在轨迹图中删除 V'中的节点及其对应的边，按照此过程不断构造 k-匿名集。对于轨迹图中规模小于 k 的连通分量，将其加入合并代价最小的 k-匿名集中，若该子图与每个 k-匿名集都是不连通的，则隐匿该子图。最后匿名化每个 k-匿名集。具体如算法 4 所示。

算法 4 个性化轨迹匿名算法(ITAnony)
输入：轨迹等价类 TEC=T_1,T_2,…,T_n,k,s, λ, α, β
输出：匿名轨迹 T_1',T_2',…,T_n'

1) S=∅, V'=∅, W=∅;
2) ITGCons(TEC, α, β);
3) 隐匿 TG 中节点数目少于 k 的连通分量；
4) while(TG≠∅)
5) for(TG 中的每个连通分量 G)
6) if(|G|≥k)
7) 在轨迹图 TG 中寻找权重最小的边(v_e,v_s);
8) V'={v_e,v_s};
9) while(|V'|<k)
10) for(V'中的每个节点 v)
11) 寻找与 v 相关联的节点 v'及权重 w(v,v');
12) if(v'∉ V')
13) W=W∪ {w(v,v')};
14) 寻找 W 中最小权重 w(v,v');
15) V'=V'∪ {v'};
16) 在 TG 中删除 V'的节点及其关联的边；
17) S=S∪ V';
18) else 将 G 加入合并代价最小的匿名集中;
19) 轨迹 k-匿名化 S 中的每个集合;
20) 返回匿名轨迹 T_1',T_2',…,T_n'。

算法 4 的步骤 1)~3)为轨迹图的构建和预处理，设每条轨迹上都有 m 个时空坐标，步骤 1)~3)可在时间 $mO(n^2)$内完成；步骤 4)~18)为轨迹图的 k-子图划分，对于规模不小于 k 的连通分量，不断查找与 V'中节点相关联且边权最小的轨迹，以构造 k-匿名集，至多需时间 $kO(n^2)$，对于规模小于 k 的连通分量，将其加入合并代价最小的 k-匿名集中，至多需时间 $kO(n)$；步骤 19)为 k-匿名集的匿名化，每产生一个匿名域至多需时间 $O(k\log k)$，因此，每个轨迹 k-匿名集的匿名化可在时间 $mO(k\log k)$内完成，由于轨迹图中最多包含 n/k 个 k-匿名集，步骤 19)至多需时间 $mnO(k\log k)$。由于 k≪n，综上，算法 4 的轨迹匿名过程可在 $mO(n^2)$内完成。

4 轨迹数据的信息损失

轨迹数据的可用性由轨迹匿名域的大小决定，匿名域越大，数据可用性越低，轨迹数据的信息损失越大。现有研究中的轨迹信息损失度量方法存在

部分影响因素无法体现影响力和匿名域的规模度量不准确的问题。因此，本文提出一种使用所有等价类的轨迹匿名域之和，与整个轨迹空间的面积比例的均值来度量轨迹数据的信息损失的方法，该方法能够更准确地反应出匿名轨迹的信息损失。形式化描述为

$$IL=\frac{1}{n_{TEC}Area}\sum_{i=1}^{n_{TEC}}S(t_i)$$

其中，n_{TEC}为等价类的数量，$S(t_i)$为 t_i时刻轨迹集产生的匿名域的面积，$Area$ 为整个轨迹空间的面积。

为满足个性化的轨迹匿名，在轨迹数据发布过程中会隐匿一些轨迹。隐匿率是关系数据隐私保护中常用的信息损失度量方法，由于关系数据和轨迹数据具有相似性：1) 一条信息对应一个特定个体；2) 多维性，关系数据中一条元组包含多个属性值，轨迹数据中一条轨迹包含多个时空坐标。因此，本文使用轨迹隐匿率作为衡量算法信息损失的标准之一，即被隐匿的轨迹数目占等价类中轨迹总数的比例。定义为

$$TSR=\frac{1}{n_{TEC}}\sum_{i=1}^{n_{TEC}}\frac{n_s}{|TEC|}$$

其中，n_{TEC}为等价类的数量，n_s为隐匿的轨迹数目，显然，轨迹隐匿率越小，被隐匿的轨迹数目越少，轨迹数据的信息损失越小，发布数据的可用性越高。

5　实验分析

5.1　实验数据和环境

实验数据集OLDEN由Network-Based Generator of Moving Objects[15]模拟器来模拟用户在德国奥登伯格市交通网络图（区域面积为 23.57 km×26.92 km）上的 10 000 条移动轨迹。表 1 给出了 OLDEN 数据集的统计信息，其中，$|D|$为数据集中轨迹的数目，$|D_{pre}|$为数据预处理后轨迹的数目，$|D_{TEC}|$为数据集中等价类的数目，Max 为规模最大的等价类包含轨迹的数量，Min 为规模最小的等价类包含轨迹的数量。

表 1　　数据集统计信息

数据集	$\|D\|$	$\|D_{pre}\|$	$\|D_{TEC}\|$	Max	Min
OLDEN	10 000	27 844	141	392	53

本实验从匿名数据质量和算法执行效率两方面进行分析，并将本文所提 ITAnony 算法与文献[13]所提算法（简称 GrePar 算法）、文献[14]所提算法（简称 CMPT 算法）进行对比实验。实验环境为 Intel Core 2 Duo CPU T8300@2.40 GHz；2 GB 内存；Microsoft Windows XP 操作系统；算法在 Matlab 7.0 和 Visual C++ 6.0 混合编程环境下实现。

5.2　实验结果分析

1) 实验 1 信息损失分析

图 3 给出了随匿名约束 k 值增大，s=1，$\lambda=\pi/2$，方向特征α、距离特征β分别为(0,1)、(0.3,0.7)、(0.7,0.3)、(1,0)时 ITAnony 算法的信息损失比较。由图 3 可知，随方向特征α的增加和距离特征β的降低，ITAnony 算法的信息损失增加，且信息损失随 k 值增大而增加。由于α增加、β降低时，用户对轨迹间方向相似性需求增加，轨迹间距离需求降低，即用户对数据隐私保护需求逐渐增加，对数据可用性需求逐渐降低，因此，ITAnony 算法产生的匿名域逐渐增加，算法的信息损失逐渐增加。同时，随 k 值的增加，ITAnony 算法构建的匿名域面积增加，因此 ITAnony 算法的信息损失随 k 值增大而增加。

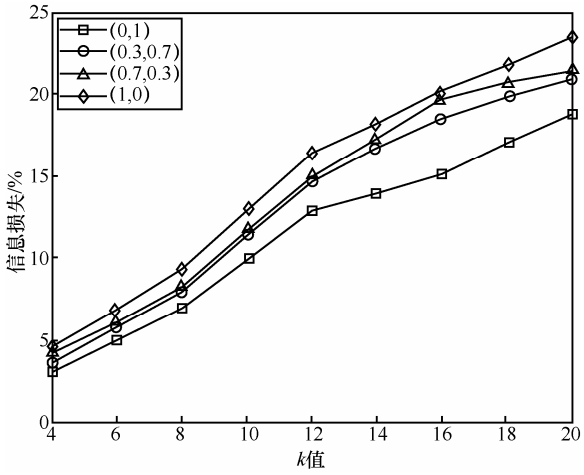


图 3　不同(α,β)、k 值下算法信息损失比较

图 4 给出了位置关联 s 值增大，匿名约束 k=12，方向特征α、距离特征β分别为(0.3,0.7)，轨迹间夹角约束λ分别为$\pi/6$、$\pi/4$、$\pi/3$、$\pi/2$时 ITAnony 算法的信息损失比较。由图 4 可知，ITAnony 算法的信息损失随 s 值增大而降低，且随λ值增大而增加。s 值增大和λ值减小时，ITAnony 算法对轨迹间关联的约束增加，轨迹间方向相似度增加且距离减少，轨迹匿名域减小，因此ITAnony算法的信息损失随s值增大而降低，且随λ值增大而增加。

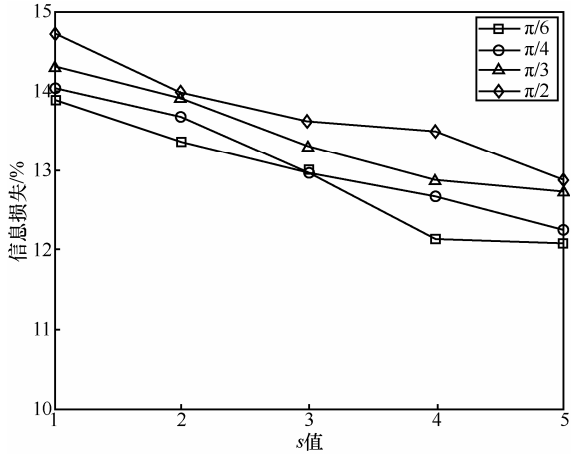


图4 不同 λ、s 参数下算法信息损失比较

图5给出了随匿名约束 k 值增大，s=1，$\lambda=\pi/2$，方向特征 α、距离特征 β 分别为(0.3,0.7)时ITAnony算法、GrePar算法和CMPT算法的信息损失比较。由图5可知，3种算法的信息损失都随 k 值的增加而增加，(α,β)=(0.3,0.7)时ITAnony算法的信息损失低于CMPT算法且略高于GrePar算法。随 k 值增加，3种算法构建的匿名域面积增加，算法的信息损失都随 k 值增加而增加。由于GrePar算法在度量轨迹间边权时仅考虑轨迹间距离，更注重匿名数据的效用，GrePar算法的信息损失低于另2种算法；而ITAnony算法构建轨迹间关联时的约束条件多于CMPT算法，构建出的轨迹匿名域面积较小，ITAnony算法的信息损失低于CMPT算法。

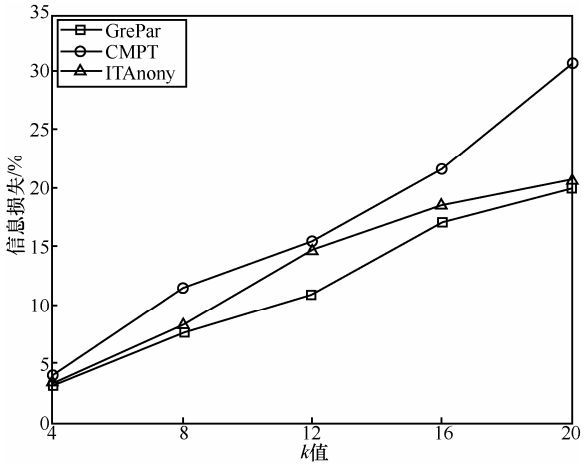


图5 不同 k 值下算法信息损失比较

图6给出了随匿名约束 k 值增大，s=1，$\lambda=\pi/2$，方向特征 α、距离特征 β 分别为(0.3,0.7)时ITAnony算法、GrePar算法和CMPT算法的在应用文献[13]的信息损失度量标准（简称GrePar-IL）时的信息损失比较。由图6可知，3种算法的GrePar-IL都随 k 值的增加而增加，GrePar算法的GrePar-IL最高，CMPT算法的GrePar-IL最低。随 k 值的增加，2种算法构建的匿名域面积增加，隐匿的轨迹数目增加，2种算法的GrePar-IL也随 k 值的增加而增加。虽然GrePar算法产生的匿名域小于ITAnony算法，但其隐匿的轨迹数目远多于ITAnony算法；且CMPT算法无轨迹隐匿步骤，在计算GrePar-IL时不需考虑隐匿轨迹的数目。因此，GrePar算法的GrePar-IL最高，CMPT算法的GrePar-IL最低。

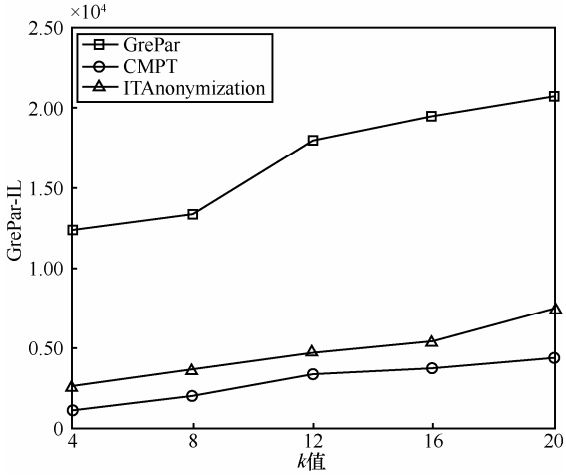


图6 不同 k 值下算法GrePar-IL比较

图7给出了随匿名约束 k 值增大，s=1，$\lambda=\pi/2$，方向特征 α、距离特征 β 分别为(0.3,0.7)时ITAnony算法、GrePar算法和CMPT算法的在应用文献[14]的信息损失度量标准（简称CMPT-IL）时的信息损失比较。由图7可知，3种算法的CMPT-IL都随 k 值的增加而增加，CMPT算法的CMPT-IL最高，ITAnony算法的CMPT-IL略高于GrePar算法。随 k 值的增加，3种算法构建的匿名域面积增加，3种算法的CMPT-IL也随之增加。由于GrePar算法注重匿名数据的可用性，产生的匿名域面积最小；ITAnony算法的轨迹间关联约束条件多于CMPT算法，轨迹间的相似度较高，产生的匿名域的面积较小。因此，CMPT算法的CMPT-IL最高，ITAnony算法的CMPT-IL略高于GrePar算法。

2) 轨迹隐匿率分析

图8给出了随匿名约束 k 值增大，s=1，$\lambda=\pi/2$，方向特征 α、距离特征 β 分别为(0,1)、(0.3,0.7)、(0.7,0.3)、(1,0)时ITAnony算法的轨迹隐匿率比较。

由图 8 可知，4 种(α,β)参数下 ITAnony 算法的轨迹隐匿率基本相同，且轨迹隐匿率随 k 值增大而增加。这是因为 s、λ 值一定时，4 种参数下 ITAnony 算法构造的轨迹图虽然边权不同，但规模相同，因此 ITAnony 算法在 4 种(α,β)参数下的轨迹隐匿率基本相同。同时，随 k 值的增加，ITAnony 算法构建的轨迹图中规模小于 k 的连通分量数目增加，隐匿的轨迹数量增多，因此 ITAnony 算法的轨迹隐匿率随 k 值增大而增加。

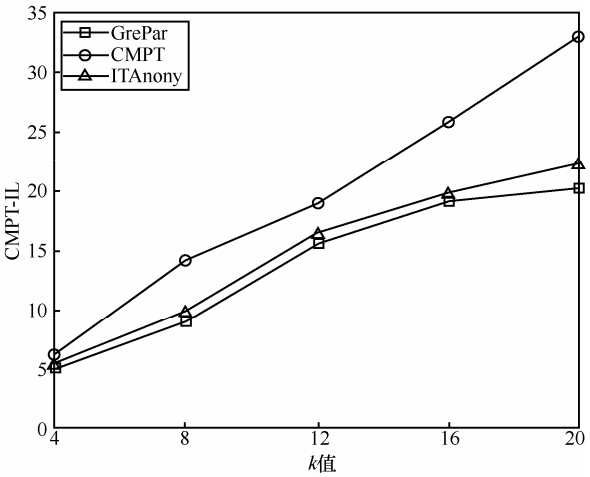


图 7　不同 k 值下算法 CMPT-IL 比较

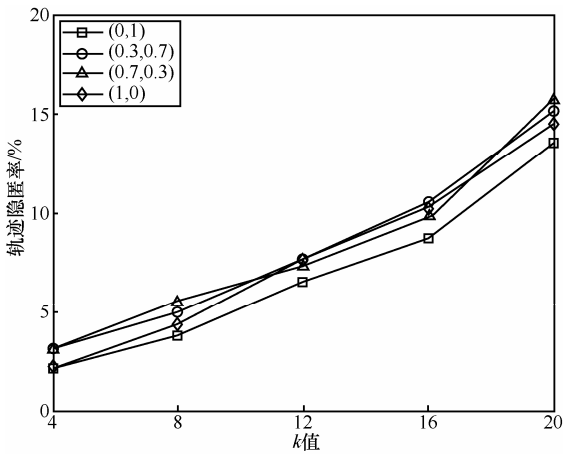


图 8　不同(α,β)、k 值下算法轨迹隐匿率比较

图 9 给出了随位置关联 s 值增大，匿名约束 k=12，方向特征α、距离特征β分别为(0.3,0.7)，轨迹间夹角约束λ分别为$\pi/6$、$\pi/4$、$\pi/3$、$\pi/2$时 ITAnony 算法的轨迹隐匿率比较。由图 9 可知，ITAnony 算法的轨迹隐匿率随 s 值的增大而增加，且随λ值的增大而降低。这是因为随 s 值的增大和λ值的减小，ITAnony 算法对轨迹间的关联约束增加，构造的轨迹图中规模小于 k 的联通分量数目增加，隐匿的轨迹数量增加，因此，ITAnony 算法的轨迹隐匿率随 s 值的增大和λ值的减小而增加。

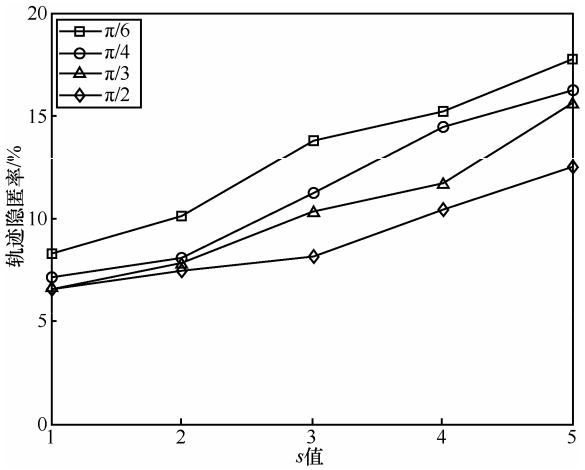


图 9　不同 λ、s 参数下算法轨迹隐匿率比较

图 10 给出了随匿名约束 k 值增大，s=1, $\lambda=\pi/2$，方向特征α、距离特征β分别为(0.3,0.7)时 ITAnony 算法、GrePar 算法的轨迹隐匿率比较。由图 10 可知，2 种算法的轨迹隐匿率都随 k 值的增加而增加，且 ITAnony 算法的轨迹隐匿率低于 GrePar 算法。这是因为随 k 值的增加，2 种算法构建的轨迹图中规模小于 k 的连通分量数目增加，因此 2 种算法的轨迹隐匿率随 k 值的增加而增加。由于 ITAnony 算法较 GrePar 算法多出匿名集的合并过程，因此 ITAnony 算法的轨迹隐匿率低于 GrePar 算法。由于 CMPT 算法将所有轨迹都划分到相应轨迹匿名集中，无轨迹隐匿步骤，相应的信息损失体现在轨迹匿名域的面积的增加上，因此 CMPT 算法不参与隐匿率对比实验。

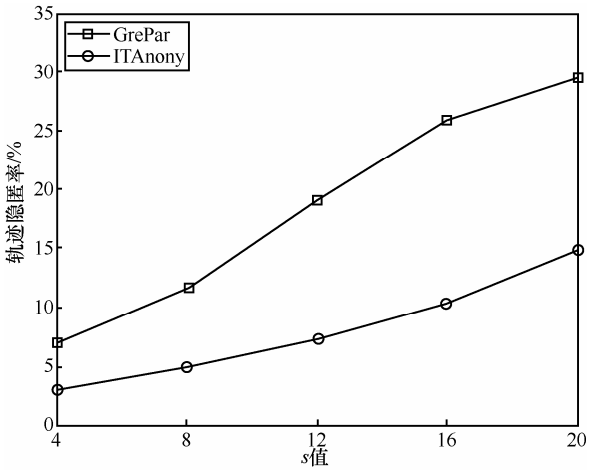


图 10　不同 k 值下算法轨迹隐匿率比较

3) 实验3执行时间分析

图 11 给出了随匿名约束 k 值增大，s=1，$\lambda=\pi/2$，方向特征 α、距离特征 β 分别为(0,1)、(0.3,0.7)、(0.7,0.3)、(1,0)时 ITAnony 算法的执行时间比较。由图 11 可知，4 种(α, β)参数下 ITAnony 算法的执行时间基本相同，且执行时间随 k 值增大而增加。这是因为 s、λ 值一定时，4 种参数下 ITAnony 算法构造的轨迹图虽然边权不同，但规模相同，因此 ITAnony 算法在 4 种(α, β)参数下所需的时间基本相同。同时，随 k 值的增加，ITAnony 算法在寻找近似最优的 k 条轨迹构成匿名集时需要更多的时间开销，因此 ITAnony 算法的执行时间随 k 值增大而增加。

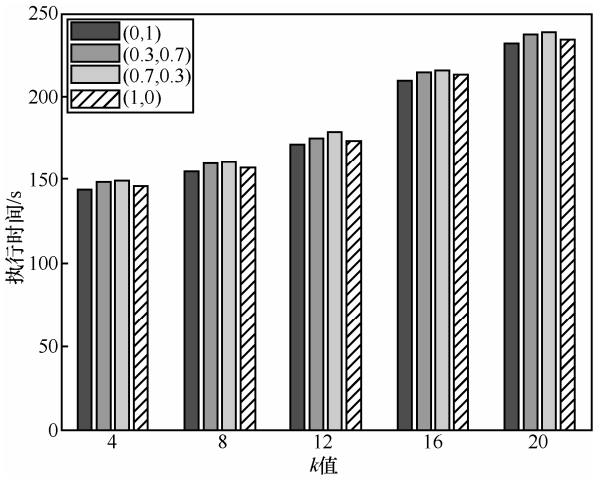


图 11 不同(α, β)、k 值下算法执行时间比较

图 12 给出了随位置关联 s 值增大，匿名约束 k=12，方向特征 α、距离特征 β 分别为(0.3,0.7)，轨迹间夹角约束 λ 分别为 $\pi/6$、$\pi/4$、$\pi/3$、$\pi/2$ 时 ITAnony 算法的执行时间比较。由图 12 可知，ITAnony 算法的执行时间随 s 值的增大而降低，且随 λ 值的增大而增加。这是因为随 s 值的增大，ITAnony 算法对轨迹间的关联约束增加，构造的轨迹图中规模小于 k 的联通分量数目增加，隐匿的轨迹数量增加，算法计算合并代价的次数减少，因此随 s 值的增大 ITAnony 算法的执行时间降低。同理，随 λ 值的增加，ITAnony 算法对轨迹间的关联约束降低，算法的执行时间增加。

图 13 给出了随匿名约束 k 值增大，s=1, $\lambda=\pi/2$，方向特征 α、距离特征 β 分别为(0.3,0.7)时 ITAnony 算法、GrePar 算法和 CMPT 算法的执行时间比较。由图 13 可知，3 种算法的执行时间都随 k 值的增加而增加，ITAnony 算法的执行时间略高于另外 2 种算法。随 k 值的增加，3 种算法在寻找 k 条轨迹构建匿名集时需要更多的时间开销，因此 3 种算法的执行时间都随 k 值的增加而增加。由于 ITAnony 算法在构建轨迹间关联时的约束条件更多，因此 ITAnony 算法的时间开销略高于另外 2 种算法。但使用 ITAnony 算法轨迹匿名后的数据可用性更高，隐私保护效果更好，略高的时间开销是可以被接受的。

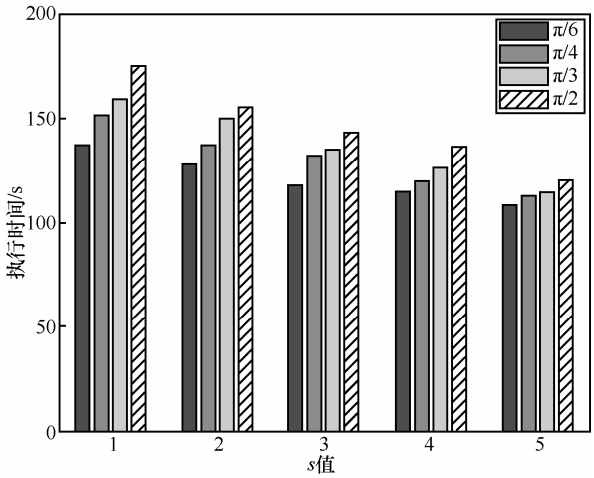


图 12 不同 λ 和 s 值下算法执行时间比较

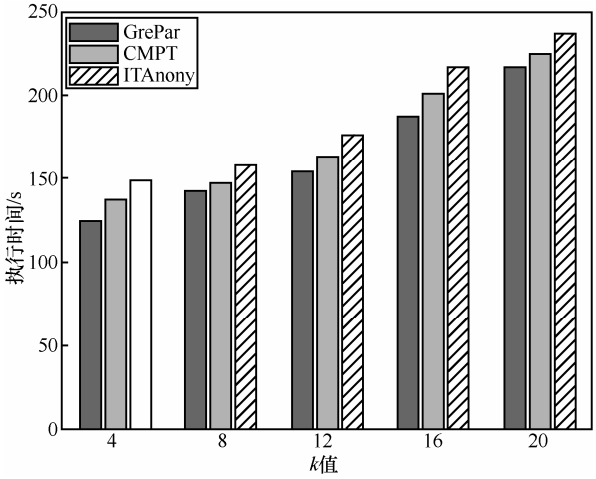


图 13 不同 k 值下算法执行时间比较

6 结束语

随着移动终端和无线定位技术的发展，基于位置的服务已深入人们的生活，但随之而来的轨迹隐私泄露也逐渐被人们所关注。本文针对用户对轨迹 k-匿名集的个性化需求，提出了一种基于轨迹间时空关系的(s, λ)-覆盖个性化轨迹关联构建方法，通过调整轨迹夹角约束 λ 和位置重合约束 s 控制轨迹

图的规模变化，有效地实现了个性化轨迹图的构建。同时，将轨迹 *k*-匿名集的构建转化为轨迹图划分问题，提出了一种基于贪心策略寻找近似最优的 *k* 条轨迹的轨迹 *k*-匿名集构建算法，有效地提高了轨迹数据的可用性，并满足了移动轨迹的个性化隐私保护。实验结果表明，虽然该方法的时间开销略高于已有方法，但其能够满足用户对轨迹 *k*-匿名集的个性化需求，且具有较低的数据隐匿率和信息损失。下一步的研究工作是进一步优化轨迹 *k*-匿名集的构建策略，在减少算法时间开销的同时实现轨迹数据的个性化隐私保护。

参考文献：

[1] MOKBEL M F. Privacy in location-based services: start-of-the-art and research directions[A]. The Int Conf on Mobile Data Management (MDM'07)[C]. Mannheim: IEEE, German, 2007. 228-235.

[2] 吴英杰, 唐庆明, 倪巍伟. 基于聚类杂交的隐私保护轨迹数据发布算法[J]. 计算机研究与发展, 2013, 50(3): 578-593.
WU Y J, TANG Q M, NI W W. A clustering hybrid based algorithm for privacy preserving trajectory data publishing[J]. Journal of Computer Research and Development, 2013, 50(3): 578-593.

[3] 张建明, 赵玉娟, 姜浩斌等. 车辆自组网的位置隐私保护技术研究[J]. 通信学报, 2012, 33(8): 180-189.
ZHANG J M, ZHAO Y J, JIANG H B, *et al*. Research on protection technology for location privacy in VANET[J]. Journal on Communications, 2012, 33(8): 180-189.

[4] CHOW C, MOKBEL M F, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based services[A]. Proc of the ACM Symposium on Advances in Geographic Information Systems(ACM GIS'06)[C]. Arlington: ACM, USA, 2006. 171-178.

[5] GHINITA G, KALNIS P, KHOSHGOZARAN A. Private queries in location based services: anonymizers are not necessary[A]. ACM SIGMOD Int Conf on Management of data (SIGMOD'08)[C]. New York: ACM, USA, 2008. 121-132.

[6] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[A]. Proc of the Int Conf on Mobile Systems, Applications, and Services (MobiSys'03)[C]. San Francisco: USA, 2003.163-168.

[7] SWEENEY L. *k*-Anonymity: a model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Konwledge Based Systems, 2002, 10(5): 557-570.

[8] ARIS G D, PANOS K, VASSILIOS S. Providing *k*-anonymity in location based services[J] ACM SIGKDD Explorations Newsletter, 2010, 1(12): 3-10.

[9] YAROVOY R, BONCHI F, LAK SHMANAN L V S. Anonymizing moving objects: how to hide a MOB in a crowd[A]. Proc of the 12th Int Conf on Extending Database Technology: Advances in Database Technology(EDBT'09)[C]. New York: ACM, USA, 2009. 72-83.

[10] CHOW C Y, MOHAMED F M. Trajectory privacy in location-based services and data publication[J]. ACM SIGKDD Explorations Newsletter, 2011, 1(13):19-29.

[11] NERGIZ M E, ATZORI M, SAYGIN Y. Towards trajectory anonymization: a generalization-based approach[A]. Proc of the SIGSPATIAL ACM GIS 2008 Int Workshop on Security and Privacy in GIS and LBS(SPRINGL'08)[C]. New York: ACM, USA, 2008. 52-61.

[12] ABUL O, BONCHI F, NANNI M. Never walk alone: uncertainty for anonymity in moving objects databases[A]. Proc of the IEEE 24th Int Conf on Data Engineering(ICDE'08)[C]. Cancun: IEEE, Mexico, 2008. 376-385.

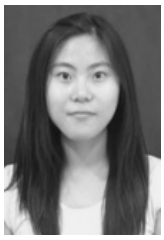
[13] HUO Z, HU H, MENG X. History trajectory privacy-preserving through graph partition[A]. Proc of the 1st Int Conf on Mobile Location-based service(MLBS'11)[C]. Beijing: ACM, China, 2011. 88-97.

[14] GAO S, MA J, SUN C. Balancing trajectory privacy and data utility using a personalized anonymization model[J]. Journal of Network and Computer Applications, 2014, 38: 125-134.

[15] BRINKHO T. Generating traffic data[J]. Bulletin of the Technical Committee on Data Engineering, 2003, 26(2):19-25.

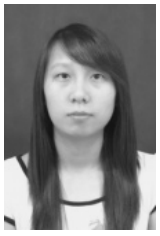
作者简介：

杨静（1962-），女，黑龙江哈尔滨人，哈尔滨工程大学教授、博士生导师，主要研究方向为数据挖掘、隐私保护等。



张冰[通信作者]（1986-），女，黑龙江哈尔滨人，哈尔滨工程大学博士生，主要研究方向为隐私保护、人工智能。E-mail:zhangbing006@hrbeu.edu.cn。

张健沛（1956-），男，黑龙江哈尔滨人，哈尔滨工程大学教授、博士生导师，主要研究方向为社会网络、数据挖掘。



谢静（1986-），女，湖北随州人，哈尔滨工程大学博士生，主要研究方向为数据挖掘、隐私保护。