

基于最小边界扇形的移动对象轨迹实时化简算法

王欣然*, 杨智应

(上海海事大学 信息工程学院, 上海 201306)

(* 通信作者电子邮箱 wxr0905@gmail.com)

摘要: 为了对全球定位系统(GPS)设备采集到的移动对象原始轨迹数据进行简化, 提高轨迹数据的使用效率, 降低移动终端的通信代价和计算开销, 提出了一种基于最小边界扇形(MBS)的移动对象轨迹实时化简算法。该算法不同于用一条折线来近似原始轨迹的方法, 它利用扇形预测范围来估计、简化原始轨迹, 在角度和距离两个层面上对简化误差进行控制。提出了新的误差度量方法——基于等极径的误差度量方法, 并讨论了 GPS 定位误差对简化算法的影响。实验结果表明, 所提算法的简化轨迹高效、稳定, 所得到的简化轨迹与原始轨迹之间误差较小(不超过误差阈值的 20%), 对 GPS 定位误差有较好的容错能力。

关键词: 移动对象; 轨迹化简; 数据压缩; 最小边界扇形; 实时化简; 全球定位系统; 定位误差

中图分类号: TP391.412 **文献标志码:** A

Real-time trajectory simplification algorithm of moving objects based on minimum bounding sector

WANG Xinran*, YANG Zhiying

(College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract: To improve the efficiency of the application of trajectory data, reduce communication cost and computational overhead of mobile terminal, the raw trajectory data of moving objects which were collected by Global Positioning System (GPS) equipment must be simplified. A method based on Minimum Bounding Sector (MBS) for real-time trajectory simplification of moving objects was proposed. The algorithm is different from those which approximated the original trajectory with a polygonal line. It adopted sector to predict the moving range, which could estimate and simplify the original trajectory. In order to control simplification error efficiently, the identical polar radius error metric method was proposed based on the characteristics of sector angle and distance. In addition, the affect of GPS positioning error on the simplified algorithm was discussed. The experimental results show that, the simplified trajectory of the proposed algorithm is efficient and stable, it has smaller error (no more than 20% of the error threshold) in comparison with the original trajectory and has good fault tolerant ability on GPS positioning error.

Key words: moving object; trajectory simplification; data reduction; Minimum Bounding Sector (MBS); real-time simplification; Global Positioning System (GPS); positioning error

0 引言

随着内置定位功能的移动设备不断发展以及对位置服务功能需求的进一步扩大, 轨迹数据变得日益重要和无处不在。然而, 由全球定位系统(Global Positioning System, GPS)设备产生的未经任何处理的原始轨迹数据的数据量十分庞大, 这直接导致在对这些数据进行传输、存储和查询的过程中效率低下, 成本代价高昂。例如, 如果每 10 s 收集一次位置数据, 那么以文献[1]中的计算可以看出 100 MB 的内存只够记录 500 个移动对象 1 d 的轨迹数据。更为重要的是, 这些轨迹数据中存在着大量的冗余数据, 这会导致网络传输负担沉重, 特别是对于移动终端用户来说, 这意味着长的上传时间以及大量的电量消耗。无论是对于位置服务提供商还是用户来说, 一方面, 希望提高位置信息的收集频率以更好地感知移动对象的位置变化; 另一方面又希望能够缩小轨迹数据的规模以提高传输和处理效率, 降低电量消耗。

基于以上的问题需求, 用户希望对得到的原始轨迹数据

进行压缩化简, 将不能够反映出移动对象轨迹变化的位置信息进行过滤, 从而在提高感知频率的同时降低数据规模。事实上, 文献[2]展开了对于移动对象原始轨迹的简化压缩的研究工作, 比较典型的几种算法还有: Douglas 等^[3]提出的算法、Bellman^[4]提出的算法以及 STTrace (Sampling Trajectory Trace) 算法^[5]等。这些算法从响应方式上, 可以分为实时算法和非实时算法。从压缩的策略上又可以分为无损压缩算法和有损压缩算法。利用无损压缩算法可以准确地重建原始轨迹, 然而在实际情况中, 由于 GPS 本身在定位中就存在着一定的误差, 因而在允许存在一定误差的环境下, 利用有损压缩算法可以大幅度地简化轨迹数据, 并使得到的简化轨迹与原始轨迹的误差在可以接受的范围以内。考虑到大多数的简化轨迹算法都没有考虑到 GPS 的定位误差, 本文提出了一种新的移动对象轨迹简化算法, 其特点如下:

1) 对终端设备的性能要求低, 只需要其收集到移动对象的位置信息。

2) 是一种简单、高效的实时轨迹化简算法, 并对 GPS 的

收稿日期: 2014-03-03; 修回日期: 2014-04-17。 基金项目: 上海海事大学研究生创新基金资助项目(CK2013080)。

作者简介: 王欣然(1989-), 男, 河南沈丘人, 硕士研究生, 主要研究方向: 移动对象数据库; 杨智应(1964-), 男, 广西钟山人, 教授, 博士, 主要研究方向: 算法与复杂性、移动计算、分布式计算、软件工程。

定位误差具有一定的容错能力。

3) 相比基于最小边界矩形 (Minimum Bounding Rectangle, MBR) 的轨迹简化算法来说, 避免丢失较多的轨迹运动特征。

4) 定义一种新的度量原始轨迹与简化轨迹之间偏离大小的误差度量方法。

1 相关工作

首先给出移动对象轨迹简化的有关定义: 移动对象在本质上是随时间而变化的几何实体^[6], 其在运动过程中获得的以时间标注的位置信息流即称为轨迹。

1.1 GPS 的误差分析

一般来讲, 在 GPS 定位中, 影响测量的偏差可分为 3 类: 1) 与 GPS 卫星有关的偏差; 2) 与观测有关的偏差; 3) 与观测站有关的偏差^[7]。其中: 第 1) 类误差是对于每一个用户接收机所公有的, 即卫星钟误差、星历误差、电离层误差和对流层误差, 利用差分技术, 第 1) 类误差可以完全消除; 第 2) 类误差是不能由用户测量或由校正模型来计算的传播延时误差, 其大部分可以被消除, 主要取决于基准接收机和用户接收机的距离; 第 3) 类误差是各用户接收机所固有的误差, 即内部噪声、通道延迟和多路径效应, 这类误差是无法消除的。GPS 定位误差^[8]由式(1)近似表示:

$$\sigma_p = GDOP \cdot \sigma_{URE} \quad (1)$$

其中: σ_p 为 GPS 定位误差的标准偏差, σ_{URE} 为伪距测量误差的标准偏差, $GDOP$ 为几何精度衰减因子 (Geometric Dilution Of Precision)。即在移动对象轨迹简化问题中, 本文所考虑的主要是 GPS 定位偏差的第 2) 类。这里伪距测量误差就是用户等效距离误差 (User Equivalent Range Error, UERE)。

1.2 轨迹简化算法研究进展

在最近的研究中, 文献[9]根据 GPS 中存在定位误差的实际情况, 提出了一种基于最小边界矩形的实时轨迹简化方法以及一种基于封闭面积的误差度量方法。该方法通过一套针对最小边界矩形的分割/合并操作以及选择策略来对空间轨迹数据进行简化, 同时通过实验论证了基于封闭面积的误差度量方法对于 GPS 定位误差的健壮性。更为重要的是本文提供了一种设计轨迹简化算法的新思路, 即利用标准图形 (这里的标准图形是指具有相似的面积或形状的图形, 文献[9]中利用的标准图形是标准矩形) 去逐段匹配整个轨迹。但同时其也存在着如下的不足: 首先, 并不能以单位 MBR 中的信息量的多少来作为值得保存点数多少的依据。MBR 中出现较多的点时既可能意味着需要进行保留的点较多, 也可能意味着冗余的点较多。其次, 由于以矩形的面积较大, 且无

明显的特征, 因而掩盖了原始轨迹较多的细节, 简化的结果变成了一系列矩形的集合。另外, 需要用户设定参数过多。在实际应用中, 对用户不够友好。最后, 在实时轨迹简化中, 由于需要进行分割和合并因而效率比较低。进一步分析可以发现, 导致以上问题的一个重要原因是由于只关注于位置信息而忽略了对于方向 (角度) 信息的利用。文献[10]着重论述了方向信息的重要性, 并论证了利用方向信息的轨迹简化算法所得到的简化轨迹在位置误差上的有界性以及利用位置信息的轨迹简化算法会导致所得到的简化轨迹在方向误差上的无界性。

本文所讨论的前提假设是针对内置 GPS 的设备 (例如智能手机、车载终端) 只具有基本的定位功能, 即只能提供基本的移动对象的位置信息而无法获得移动对象的速度、方向等信息, 并且考虑 GPS 定位误差。其最终目标是希望通过设计出合适的轨迹简化算法在终端对定位信息进行实时的分析简化, 进而达到降低定位数据规模、提高处理效率的目的。

2 实时简化框架

本文以在移动终端进行实时轨迹简化为应用背景, 考虑 GPS 的定位误差。以时间复杂度和空间复杂度为基础, 着重考察简化算法的简化率与误差大小。下面形式化给出框架描述。

2.1 问题定义

定义 1 原始轨迹。在给定的欧氏空间中, 移动对象所有的位置信息所构成的轨迹序列, 它可以用 $O = \{o_1, o_2, \dots, o_n\}$ 表示, 在图形中它可以用一条连接所有点的折线 LO 表示。

定义 2 简化轨迹。在给定的欧氏空间中, 对移动对象的原始序列进行简化后的序列, 它可以用 $S = \{s_1, s_2, \dots, s_m\}$ 表示, 其中 $m \leq n$ 。在图形中它也可以用一条连接所有点的折线 LS 表示。其中: 记 $O(S)$ 为 $n(m)$ 个以时间序列组成的三元组集, 并且 $o_i(s_i)$ ($1 \leq i \leq n(m)$) 为三元组集 $O(S)$ 中的第 i 个三元组, 即 $o_i(s_i) = (t_i, x_i, y_i)$, 这里 (x_i, y_i) 为轨迹在二维空间中第 t_i 时刻的位置, 记为 $p_i = (x_i, y_i)$ 。S 是 O 的子集。当 $S \subset O$ 时, 称 S 为 O 的强简化^[11], 这是本文所讨论的。

定义 3 线段的角度 (方向)。对于轨迹折线 T 中的某一线段 $\overline{p_i p_{i+1}}$, 该线段的方向定义为以 p_i 为极点的从极轴 x 到该线段的逆时针角度, 记为 $\theta(\overline{p_i p_{i+1}})$, 其取值范围为 $[0, 2\pi)$ 。另外定义两个角度 θ_1 与 θ_2 的角度差为 $\Delta(\theta_1, \theta_2) = \min\{|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|\}$ 。如图 1 所示: 图 1(a) 展示了角度定义的两个实例, 在图 1(b) 与 1(c) 中, 分别展示了角度差 $\Delta(\theta_1, \theta_2)$ 为 $|\theta_1 - \theta_2|$ 和 $2\pi - |\theta_1 - \theta_2|$ 的两种情况, 其中角度差的范围为 $[0, \pi]$ 。

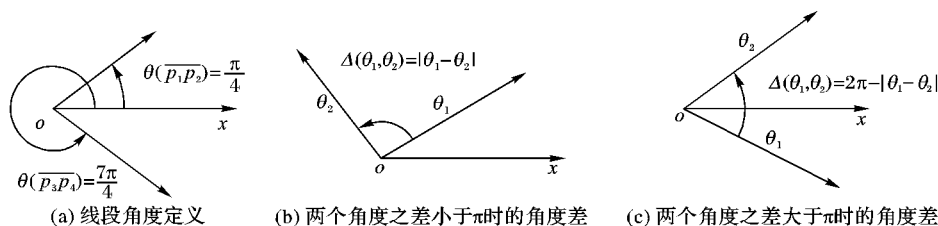


图1 有关角度以及角度差的定义示例

另一个需要注意的问题是有关实时轨迹简化的定义, 文献[12]中给出了有关定义: 当移动终端获取到新的位置信息数据时, 其自身能够及时地根据用户的要求来判断是否将该

位置信息数据发送到服务器端。其中, 用户的要求可以通过多种方式表达, 如对轨迹误差的要求、对数据量大小的要求等。其公式形式可以表示为:

$\forall o_i \in [o_1, o_c]$: 可以判断 $o_i \in S$ 是否成立 (2)

其中 o_c 为当前位置信息数据。因此, 由于移动终端无法知道整个轨迹中所有位置信息(特别是未来的轨迹数据), 实时轨迹化简是一种贪心算法。

2.2 基于等极径的误差度量方式

在移动对象轨迹化简问题中, 通常应用的误差度量方式主要有3种: 垂直欧氏距离、同步欧氏距离和 Fréchet 距离^[13]。此外, 文献[9]针对其简化算法的特性提出了一种新的误差度量方式: 基于有界区域面积(Enclosed Area)的误差度量方式。为了适应构造最小边界扇形算法的实现思路, 本文设计了基于等极径的误差度量方式, 如图2所示, 其中: 原始轨迹 $O = \{o_1, o_2, o_3, o_4, o_5, o_6\}$, $\overline{o_1 e}$ 为该扇形中心角的平分线。其具体误差的度量方法是: 首先在已得到的边界扇形内, 将扇形顶点 o_1 作为极坐标的极点, 计算中心角平分线的角度(记为 β) 以及扇形内原始轨迹点 o_i 到极点的距离(记为 ρ), 然后根据式(3) 计算边界扇形中各个原始轨迹点对应的估计点 s_i 的坐标, 最后该估计点到原轨迹点之间的距离即为简化后的轨迹误差:

$$\begin{cases} s_i, x = o_1, x + \rho \times \cos \beta \\ s_i, y = o_1, y + \rho \times \sin \beta \end{cases} \quad (3)$$

例如, 在图2中 s_5 为 o_5 在 $\overline{o_1 e}$ 上的对应估计位置, 其到极点 o_1 的极径与 o_5 到极点 o_1 的极径相同, 这两点之间的距离即为误差 d_1 。从另一方面来说, 通过这种误差度量方式得到的最大误差即为腰长为扇形半径的等腰三角形(在图2中即为 $\triangle o_1 a e$, a 为边界扇形的一个边界点) 顶角所对边的长度, 在图2中即为 \overline{ae} 的长度 d 。换句话说, 只要能够控制该线段的长度, 就能够对整个简化过程进行误差控制。下面本文将讨论如何在原始轨迹中逐段构造最小边界扇形。

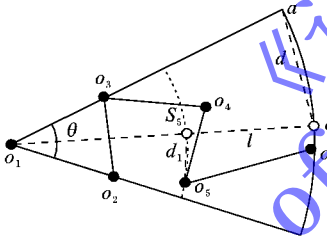


图2 基于等极径的误差度量方式

3 本文方法

为了在轨迹简化过程中避免丢失过多的移动对象运动特征, 同时保留利用标准图形去逐段匹配整个轨迹的设计思路, 本文在距离和角度(此处的角度信息是通过位置信息相对变化计算出来的, 而不是直接通过移动终端获取的) 两个层面对移动对象轨迹简化过程进行控制, 从而设计出了基于最小边界扇形的移动对象轨迹简化算法。

在移动目标运动的过程中, 本文采用最小边界扇形(Minimum Bounding Sector, MBS)来近似简化其移动轨迹。如图3所示, 连接 o_1, o_2, \dots, o_6 的实线表示移动对象的实际轨迹, 其构造最小边界扇形的方法根据扇形顶点选择方式的不同可以分为两种: 第1种是以当前获得的位置点为顶点向后构造最小边界扇形(back Minimum Bounding Sector, back_MBS), 该扇形顶点的选择随着位置信息的更新不断变化, 如图3(a)所示。当移动端获得位置信息点 $o_3(o_4, o_5)$ 时, 以该点为扇形的顶点, 以当前时刻之前所有轨迹点到该点的最远距离为半径, 形成了圆心角为 $\theta_1(\theta_2, \theta_3)$ 的最小边界扇形。第2

种是以上一次简化更新后的下一个位置点或是感知定位刚开始的最初位置点为顶点向前构造最小边界扇形(forward Minimum Bounding Sector, forward_MBS)。如图3(b)所示, 对于最初位置点 o_1 向前构造包围点集 o_2, o_3, \dots, o_6 的最小边界扇形。对于移动对象轨迹 $O = \{o_1, o_2, \dots\}$, 在某个时间段内, 以某一起始点 o_k 为顶点, 同时包含 $P = \{o_k, o_{k+1}, \dots, o_n\} (P \subset O)$ 中所有点的最小扇形是唯一的。考虑到移动对象轨迹点的坐标实际上是包含了时间信息的3维时空坐标点, 本文算法需要先将 P 中的每个轨迹点投影到平面直角坐标系中, 再将直角坐标转换为极坐标, 然后将每个顶点对应的极角、极半径排序, 求出最小角、最大角和最大半径。这样就求出了包围该移动轨迹的最小边界扇形最小角、最大角和最大半径。

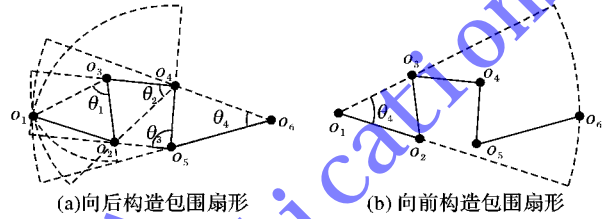


图3 两种构造边界扇形的方法

在完成了最小边界扇形的构造方法以后, 需要一定的选择策略来近似估计边界扇形内的各个轨迹点, 在这里, 本文算法以过边界扇形顶角的平分线上的点作为边界扇形中各点的对应点, 并且原始点与其对应点到扇形顶点的距离相同。通过上文可知, 在构造边界扇形的同时需要对误差进行控制, 当目前的误差超过设定的误差阈值时, 就需要对外更新移动对象的位置信息。从其选择策略可以得到边界扇形中最大的误差产生在其中的某一个轨迹点位于边界扇形最边界位置上的情况, 即如图2中点 a 的位置, 其对应的最大误差为 d 。该值与该边界扇形的中心角大小和扇形的半径有关, 在等腰三角形 $\triangle o_1 a e$ 中, 利用余弦定理可得:

$$\begin{cases} d^2 = 2l^2 - 2l^2 \cos(\theta/2) \\ 2l^2 - 2l^2 \cos(\theta/2) \leq st_error^2 \end{cases} \quad (4)$$

其中: l 为扇形的半径长度, θ 为扇形的中心角度值, d 为该边界扇形的最大误差值, st_error 为根据用户要求所设定的误差阈值。当设定了简化过程中的误差阈值后, 每当获得新的位置信息后, 只要判断新得到的边界扇形是否满足式(4) 就可以知道是否需要向服务端更新该移动对象的位置信息数据。其具体的算法如算法1~2所示, 其中两个算法的输入参数分别为设定误差阈值 st_error 以及原始轨迹数据集 $Traj$ 。函数 $CalcMBS(Buf, point)$ 的第一个参数 Buf 表示存储历史轨迹点的集合, 第二个参数 $point$ 为最新得到的位置信息, 该函数用于计算当前的最小边界扇形; 函数 $f_4(newMBS)$ 的参数 $newMBS$ 表示当前的最小边界扇形的信息, 该函数用于计算当前的最小边界扇形是否满足式(4); 函数 $BoundInfor(newMBS)$ 为获取当前最小边界扇形的边界信息, 包括扇形中心点坐标以及距离该坐标点的最大值和到中心点与 X 轴所成最大角、最小角四个数据。

算法1 back_MBS method。

```
function calcMinBoundingSector(st_error, Traj) {
    Buf = null;           //保存接收到的轨迹点位置信息集合
    newMBS = null;        //保存一个最小边界扇形信息的集合
    WHILE( Traj.hasNext() ) DO
        point = Traj.next();
        //前两个轨迹点直接保存到集合中
```

```

IF (Buf.Count >= 2) THEN
    //计算加入新位置点后的最小边界扇形
    newMBS = CalcMBS(Buf, point);
    //如果误差在给定阈值以内不作其他处理
    IF (f4(newMBS) <= st_error) THEN
        Buf.add(point);
    //否则将当前的最小边界扇形信息向服务器端更新
    ELSE
        output(BoundInfor(newMBS));
        //将超出误差阈值的点加入下一个 MBS 信息集合中
        Buf = null; Buf.add(point);
    END IF
ELSE
    Buff.add(point);
END IF
END WHILE
}
算法 2 forward_MBS method。
function calcMinBoundingSector(st_error, Traj) {
    //保存一个最小边界扇形信息的集合
    newMBS = null;
    WHILE( Traj.hasNext() ) DO
        point = Traj.next();
        //计算加入新位置点后的最小边界扇形
        newMBS = CalcMBS(newMBS, point);
        IF (f4(newMBS) >= st_error) THEN
            //否则将当前的最小边界扇形信息向服务器端更新
            output(BoundInfor(newMBS));
        END IF
    END WHILE
}

```

对算法 1 ~ 2 进行分析可知, 由于其都可以在某一个常数时间 C 内, 完成对当前轨迹点信息的简化判断, 因而其时间复杂度都为 $O(1)$, 而就空间复杂度而言, 对于算法 1, 其最坏的情况可以达到 $O(n)$ (例如, 移动对象做直线运动时, 其需要将历时轨迹信息逐一记录); 对于算法 2, 由于其只需要保存一个 MBS 的边界信息的集合, 因而其空间复杂度为 $O(1)$ 。这种基于最小边界扇形的移动对象轨迹简化方法根据移动目标在某一范围内, 距离起始位置 (当前位置) 的最大距离, 以及和极轴形成的最大角和最小角所确定的扇形区域来近似地表示其在该段时间范围内运动轨迹, 其优点在于其对移动对象本身的需求的信息量较少 (只需要其位置信息, 无需瞬时速度等信息), 且相对于最小边界矩形的方法来说近似的精度更高, 包含的信息量更多, 如该移动对象大概的运动趋势。以 forward_MBS 算法为例, 应用该算法后的简化结果如图 4 所示, 从中可以看出轨迹的大致运行趋势, 并且覆盖范围比基于 MBR 轨迹化简算法更为精确。

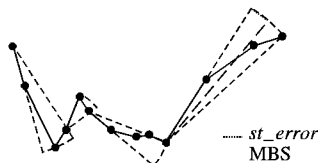


图 4 应用 forward_MBS 的实时轨迹化简算法后的结果

4 实验及分析

为了分析与评估算法的性能, 实验选取台式电脑作为测试硬件平台, 具体配置为: Pentium Dual-Core CPU E5500

2.80 GHz, 内存为 2 GB; 实验环境为 Windows 7 操作系统和 Eclipse 开发系统。实验数据是从项目 INFATI (Intelligent Farttilpasing)^[14] 得到, INFATI 的位置数据是由 20 个 GPS 数据集构成。

4.1 GPS 定位误差分析

正如上文所述, GPS 中总会存在几米到十几米的固有误差。对于一个特定地域来说, 这种定位误差并不是固定的, 该误差可以看作符合标准正态分布, 即对于大多数定位信息来说其误差是在一个很小的范围内 (在商业 GPS 中这些微小的误差可以通过某些平滑技术来消除), 而只有很少的特殊点会产生较大的误差^[9]。如图 5 所示, 轨迹点 o_5 存在着较大的误差, 其实际位置为 o_5' , 而其余位置点的误差忽略不计。

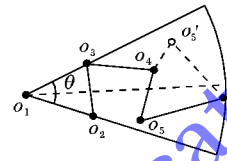


图 5 轨迹简化问题中的误差分析

为了在轨迹简化问题中分析 GPS 定位误差对于简化算法的影响, 可以将原始轨迹中的各个位置点信息分成两类: 一种是可以被简化的位置点, 即运动变化微小可以在限定误差范围内被预测; 另一种是不可被简化的位置点, 即运动变化显著即简化轨迹 S 中的点集。GPS 不确定性对于这两类点集的影响结果是不一样的: 首先对于不可被简化的位置点 (如图 5 中的 o_1) 来说, 当其产生较大误差时, 其很大程度上将直接导致简化结果集的改变。换句话说, 任何简化算法对于不可简化点产生较大误差这种情况的容错能力都是无法分析, 其既可能导致简化结果完全改变, 也可能对简化结果没有影响。另一方面, 对于可以被简化的位置点 (如图 5 中的 o_2, o_3, o_4, o_5, o_6) 来说, 简化算法的不同将直接影响到其简化结果对于 GPS 定位误差的健壮性。由于本文改变了以往用一条折线去近似原始轨迹折线的方法, 而是通过一系列的预测范围 (通常以某种相似的图形) 来估计并简化原始轨迹, 因而对于 GPS 定位误差有了较强的容错能力。如图 5 所示点 o_5 的实际位置点为 o_5' , 然而最后的简化结果集并没有因 GPS 定位误差而产生影响。事实上, 只要 o_5 的实际位置在该边界扇形的范围内, 该点的 GPS 定位误差都不会对最后的简化结果产生影响。

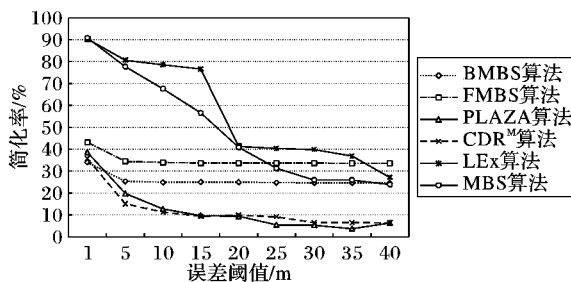
4.2 MBS 算法性能分析

为了进一步分析 MBS 算法的特性, 本文首先引入几种典型的实时轨迹简化算法: 线性外推 (Linear Extrapolation, LEx) 算法^[15] 以前两个轨迹点定义运动矢量, 然后以该运动矢量对后续的轨迹进行预测, 最后根据预测结果来对实际轨迹进行简化。这种算法实现简单适用于轨迹变化平缓的情况。连接-保持推测算法 (Connection-preserving Dead Reckoning with a predefined parameter M , CDR^M)^[16] 利用一个具有固定大小的存储空间来存储部分历史轨迹点信息, 并利用这些历史轨迹点来对当前时刻的轨迹点作出去留判断, 同时根据以上判断来更新这个历史轨迹点集合。基于分区角度的分段线性近似 (Piecewise Linear Approximation with Zoning Angle, PLAZA) 算法是首先是由 Soroush 等^[17] 针对在条件有限的传感器网络中为保证高效数据流传输所设计出来的, 他们提出了一个叫做“分区角度” (zoning angle) 的概念, 并利用该概念进行轨迹简化。首先, 本文对以上各种算法进行复杂度分析, 结果如表 1 所示。其中: n 为轨迹点的数量, m 为给定的存储空间的大小。

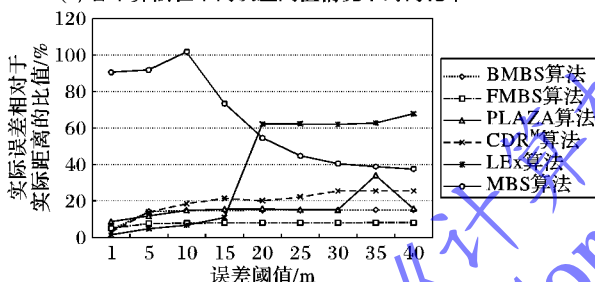
表1 几种算法的复杂度分析

算法	时间复杂度	空间复杂度	算法	时间复杂度	空间复杂度
LEx 算法	$O(n)$	$O(1)$	MBR 算法	$O(n \log m)$	$O(m)$
BMBS 算法	$O(n)$	$O(n)$	PLAZA 算法	$O(n)$	$O(1)$
FMBS 算法	$O(n)$	$O(1)$	CDR ^M 算法	$O(m \log m)$	$O(m)$

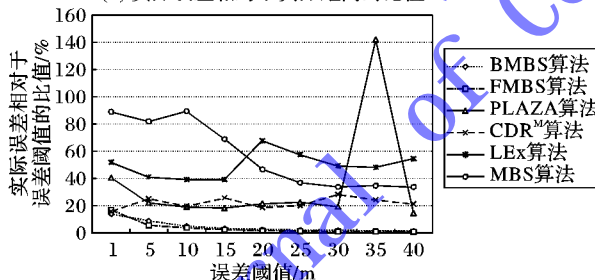
接着,针对以上算法从简化率、运行时间以及误差大小这几个方面来进行性能分析与比较。轨迹简化率 rr 为简化轨迹中轨迹点的数量与原始轨迹中轨迹点的数量的比值,当这个比值越小时,反映了该算法的简化程度越好。换句话说存储空间节省率即为 $1 - rr$ 。通过改变误差阈值 ε 的值来得到几种算法的简化效果,结果如图 6(a) 所示。



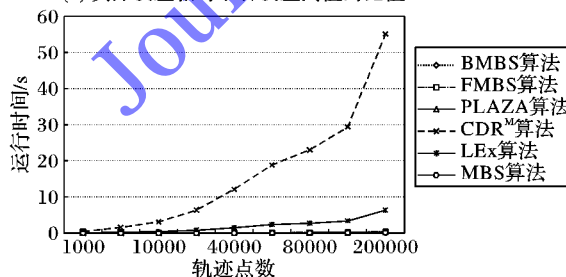
(a) 各个算法在不同误差阈值情况下的简化率



(b) 实际误差相对于实际距离的比值



(c) 实际误差相对于各误差阈值的比值



(d) 几种算法的运算时间对比

图6 几种算法的性能比较

为了更好地分析各个算法,本文定义了两种误差度量函数来评价所得简化轨迹与原始轨迹之间误差程度:

1) 在时刻 i 时,简化轨迹点 S_i 与原始轨迹点 P_i 的距离记为 $error$,设定的误差阈值为 $seterror$ 。该误差度量方法记为

$errormeasurement1 = error / seterror$ 。

2) 在时刻 i 时,简化轨迹点 S_i 与原始轨迹点 P_i 的距离记为 $error$,原始轨迹点 P_i 与前一个时刻原始轨迹点 P_{i-1} 的距离记为 $distantp$,则该误差度量方法记为 $errormeasurement2 = error / (distantp + error)$ 。

从误差度量方法 $errormeasurement1$ 中,可以得到实际的误差结果与事先设定的误差阈值的比值,如果比值越接近 1,反映出算法对于误差控制方面能起到很好的效果,使得到的简化轨迹与原始轨迹之间的误差能够在给定的误差域值之内;另一方面,为了得出误差相对于总体距离的比值,则需要另一种误差度量方法,考虑到实际误差与实际距离的值都可能取 0,因而得到了上述定义,其反映了误差对于实际距离的影响程度,比值越小则反映出其对实际距离影响的程度越低,当实际距离与误差同时为 0 时,则将该比值定义为 0。其结果分别如图 6(b)、(c) 所示。

通过以上的结果可以看出,本文所提出的两种算法在各方面表现出了良好的性能特点。总的来说,从简化率的角度来看这两种算法在不同的误差精度要求下都表现出了较强的稳定性,这说明本文所提出的算法能够适应不同的应用环境要求;从简化误差来看,这两种算法均优于目前较为优秀的简化算法,在保证高精度的前提下依然保持着很好的稳定性;从简化算法的运行时间来说,本文算法耗时时间很低,有效地保证了算法的实时性。具体来说,CDR^M 算法在简化率以及简化精度方面均表现优秀,但由于该算法计算复杂,在实际应用中耗时较长(如图 6(d) 所示),这对于实时轨迹简化场合来说可能会导致较为严重的数据积压;PLAZA 算法的简化率以及算法复杂性都较为优秀,但其获得的简化轨迹忽略了角度的轨迹细节,误差较大;LEx 算法作为最为简单的简化算法,在简化率以及简化精度等方面均不够理想。

5 结语

文献[9]中所讨论的是 GPS 不确定性因素对度量方式的影响,而本文所讨论的是 GPS 不确定性因素对简化算法本身的影响。本文改变了以往用一条折线去近似原始轨迹折线的方法,而是通过一系列的预测范围(通常以某种相似的图形)来估计并简化原始轨迹。基于最小边界扇形的移动对象轨迹简化信息中只需要记录扇形中心点坐标,以及距离该坐标点的最大值和到中心点与 X 轴所成最大角、最小角这 4 个数据。通过分析可以得到,这种以预测范围来简化原始轨迹的方法对于 GPS 不确定性因素的影响具有更好的健壮性。此外,该简化算法以一种更具实际意义(即可以判断出其运动的方向)的边界扇形对原始轨迹进行简化,避免了一些轨迹特征的丢失。在实际应用中对于用户也十分友好。更为重要的是,文献[18]中已经完成了对于最小边界扇形的空间索引方法的研究,这意味着在实际应用中,可以以一套完整的应用体系(先对移动位置数据信息进行简化,再对其运用索引方法存储于数据库中)来对移动对象的位置信息数据进行处理。当然这种以预测范围来简化原始轨迹的方法,在对轨迹重建时也会产生相应的问题,即重建后的轨迹不够细致。能否找到更为合适的预测范围(某种图形)以及如何利用该算法得到的简化轨迹来对原始轨迹进行更好的重建将是今后所需研究的问题。

参考文献:

- [1] MERATNIA N, ROLF A. Spatiotemporal compression techniques for moving point objects [C]// EDBT 2004: Proceedings of the 9th International Conference on Extending Database Technology. Berlin: Springer, 2004: 765–782.
- [2] CAO H, WOLFSON O, TRAJCEVSKI G. Spatio-temporal data reduction with deterministic error bounds [J]. The VLDB Journal—the International Journal on Very Large Data Bases, 2006, 15(3): 211–228.
- [3] DOUGLAS D H, PEUCKER T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature [J]. Cartographica: the International Journal for Geographic Information and Geovisualization, 1973, 10(2): 112–122.
- [4] BELLMAN R. On the approximation of curves by line segments using dynamic programming [J]. Communications of the ACM, 1961, 4(6): 284.
- [5] POTAMIAS M, PATROUMPAS K, SELLIS T. Sampling trajectory streams with spatiotemporal criteria [C]// SSDBM'06: Proceedings of the 18th International Conference on Scientific and Statistical Database Management. Piscataway: IEEE Press, 2006: 275–284.
- [6] GUTING R H, SCHNEIDER M. Moving objects databases [M]. Amsterdam: Elsevier, 2005: 25–30.
- [7] YUAN J, LUO J, YUE X. Principle and application of satellite navigation [M]. Beijing: China Astronautic Publishing House, 2003: 191–218. (袁建平, 罗建军, 岳晓奎. 卫星导航原理与应用[M]. 北京: 中国宇航出版社, 2003: 191–218.)
- [8] KAPLAN E D, HEGARTY C J. Understanding GPS: principles and applications [M]. Boston: Artech House, 2006: 379–390.
- [9] LIU G, IWAI M, SEZAKI K. An online method for trajectory simplification under uncertainty of GPS [J]. Information and Media Technologies, 2013, 8(3): 665–674.
- [10] LONG C, WONG R C W, JAGADISH H V. Direction-preserving trajectory simplification [J]. Proceedings of the VLDB Endowment, 2013, 6(10): 949–960.
- [11] TRAJCEVSKI G, CAO H, SCHEUERMANN P, *et al.* On-line data reduction and the quality of history in moving objects databases [C]// Proceedings of the 5th ACM International Workshop on Data Engineering for Wireless and Mobile Access. New York: ACM Press, 2006: 19–26.
- [12] PARK H, LEE Y J, CHAE J, *et al.* Online approach for spatio-temporal trajectory data reduction for portable devices [J]. Journal of Computer Science and Technology, 2013, 28(4): 597–604.
- [13] FAN C. The study of geometric algorithms: color-spanning set, Voronoi diagram and Fréchet distance [D]. Changsha: Central South University, 2011. (范成林. 基于颜色支撑点集 Voronoi 图和 Fréchet 距离的几何算法研究[D]. 长沙: 中南大学, 2011.)
- [14] JENSEN C S, LAHRMANN H, PAKALNIS S, *et al.* The INFATI data [EB/OL]. [2014-01-11]. <http://arxiv.org/abs/cs/0410001>.
- [15] HÖNLE N, GROSSMANN M, REIMANN S, *et al.* Usability analysis of compression algorithms for position data streams [C]// Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2010: 240–249.
- [16] LANGE R, DÜRR F, ROTHERMEL K. Online trajectory data reduction using connection-preserving dead reckoning [C]// Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services. Brussels: Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2008: 52.
- [17] SOROUSH E, WU K, PEI J. Fast and quality-guaranteed data streaming in resource-constrained sensor networks [C]// Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM Press, 2008: 391–400.
- [18] LI Y. The spatial index method based on minimum bounding circle and minimum encircling sector [D]. Harbin: Harbin University of Science and Technology, 2009. (李杨. 基于最小边界圆和最小包围扇形的空间索引方法[D]. 哈尔滨: 哈尔滨理工大学, 2009.)
- [11] SHI C, ZHANG M. Analysis of logistic regression models [J]. Computer Aided Engineering, 2005, 14(3): 74–78. (施朝健, 张明铭. Logistic 回归模型分析[J]. 计算机辅助工程, 2005, 14(3): 74–78.)
- [12] CHA M, HADDADI H, BENEVENUTO F, *et al.* Measuring user influence in twitter: the million follower fallacy [C]// ICWSM 2012: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media. Menlo Park: AAAI Press, 2010: 10–17.
- [13] ZHANG Y, ZHANG H, ZHANG W. Quick ranking algorithm for network user based on power law distribution [J]. Journal of Chinese Information Processing, 2012, 26(4): 122–128. (张玥, 张宏莉, 张伟哲. 基于幂律分布的网络用户快速排序算法[J]. 中文信息学报, 2012, 26(4): 122–128.)
- [14] RICHARDSON M, DOMINGOS P. Combining link and content information in Web search [M]// Web dynamics: adapting to change in content, size, topology and use. Berlin: Springer, 2004: 179–193.
- [15] LIU Q, ZHANG H, BAI S. An open resource platform for Chinese NLP [J]. Applied Linguistics, 2002(4): 50–56. (刘群, 张浩, 白硕. 自然语言处理开放资源平台[J]. 语言文字应用, 2002(4): 50–56.)
- [16] LIN X, WANG W. Set and string similarity queries: a survey [J]. Chinese Journal of Computers, 2011, 34(10): 1853–1862. (林学民, 王伟. 集合和字符串的相似度查询[J]. 计算机学报, 2011, 34(10): 1853–1862.)
- [17] LIAN J, ZHOU X, CAO W, *et al.* SINA microblog data retrieval [J]. Journal of Tsinghua University: Science and Technology, 2011, 51(10): 1300–1305. (廉捷, 周欣, 曹伟, 等. 新浪微博数据挖掘方案[J]. 清华大学学报: 自然科学版, 2011, 51(10): 1300–1305.)
- [18] LI Y, YU H, LIU L. Predict algorithm of micro-blog retweet scale based on SVM [J]. Application Research of Computers, 2013, 30(9): 2594–2597. (李英乐, 于洪涛, 刘力雄. 基于 SVM 的微博转发规模预测方法[J]. 计算机应用研究, 2013, 30(9): 2594–2597.)

(上接第2408页)