

# 基于路网的GPS轨迹在线压缩方法<sup>①</sup>

程倩<sup>1,2</sup>, 丁云峰<sup>2</sup>

<sup>1</sup>(中国科学院大学, 北京 100049)

<sup>2</sup>(中国科学院沈阳计算技术研究所, 沈阳 110168)

**摘要:** 在传统的GPS轨迹压缩中, 其压缩的结果与原始轨迹差别较大, 在压缩过程中不同程度的忽略了轨迹点的速度信息、方向信息以及轨迹的形状, 在本文中, 在保证压缩率的前提下将保存速度、方向、轨迹形状等GPS轨迹特征信息, 作为重点研究的问题. 本文算法基于路网信息、OW(Opening Window, 开放窗口)算法、关键点法以及停留点法之上提出了一种能够保存GPS轨迹时空特性的在线压缩算法. 实验结果表明, 该压缩算法较现有的压缩算法在保证压缩率的前提下, 在保留轨迹时空特性的基础上有所改进.

**关键词:** 路网; GPS 轨迹; 在线压缩; 地图匹配; 时空特性

## Online Compression Method for GPS Trajectory Based on Road Network

CHENG Qian<sup>1,2</sup>, DING Yun-Feng<sup>2</sup>

<sup>1</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>2</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

**Abstract:** The compression of the GPS trajectories in a conventional way is greatly different with the original one for ignoring the speed information, the direction of GPS points and the shape of tracks. In this paper, we focus on saving the speed, the direction information and shape features of tracks on the premise of keeping the compression ratio. The Algorithm in this article is based on the road network information, OW(Opening Window) algorithm, the core point algorithm and the stay point algorithm, then presents a online compression algorithm which is capable of saving the temporal characteristics of GPS tracks. Experimental results show that the compression algorithm compareing to the existing compression algorithms, ensures the compression ratio while improving to retain the temporal characteristics of the track.

**Key words:** road network; GPS trajectories; online compression; map-matching; temporal characteristics

近年来, GPS 定位功能在智能移动终端已得到广泛普及, GPS 定位模块成为移动终端不可或缺的功能. 由移动设备采集的 GPS 定位信息呈爆炸式增长. 如何减少从移动终端到服务器的数据传输量以及压缩存储服务器端海量的 GPS 轨迹信息成为当前的研究热点<sup>[1]</sup>. 通过移动终端可以很容易获得用户的位置, 记录用户的行为, 分析 GPS 轨迹信息, 具有巨大的商业应用前景. 智能终端获取的 GPS 轨迹数据, 需要通过 GPRS 传送到服务器端进行存储, GPRS 传输数据是通过计算流量的方式来进行收费的. 通过在线压缩的 GPS 轨迹算法, 可以有选择的向服务器传送轨迹点, 即减少

了数据的传输量, 缓解了服务器的存储压力, 也降低了数据的通信费用.

GPS 轨迹压缩的目的就是为了简化轨迹数据, 尽可能在保证轨迹原型的基础上存储轨迹. 基于路网的 GPS 轨迹压缩<sup>[2,3]</sup>, 将轨迹匹配到路网上, 不仅可以用路网信息表示较长的 GPS 轨迹信息, 大大提高 GPS 轨迹压缩率, 也在一定程度上, 减少了 GPS 采样导致的误差, 保证了轨迹压缩的质量.

在现有的 GPS 轨迹压缩算法中, 大部分压缩算法很难保留原 GPS 轨迹中的移动速度、轨迹形状和方向等特征信息, 压缩过程中的损失率较大. 由 Douglas 和

① 收稿时间:2015-10-14;收到修改稿时间:2015-12-03 [doi:10.15888/j.cnki.csa.005212]

Peucker<sup>[4]</sup>于 1973 年提出的道格拉斯-普克算法 (Douglas-Peucker Algorithm, 简称 DP 算法) 和 Meratnia<sup>[5]</sup>等人在 2004 年提出的基于速度和时间比例的自顶向下轨迹压缩算法 (Top-Down Time-Ratio Algorithm, 简称 TD-TR 算法) 等算法都是集中对 GPS 轨迹的形状进行压缩. Schmid<sup>[6]</sup>等人提出了一种利用路网信息的 GPS 轨迹压缩方法, 首先将移动用户的轨迹进行地图匹配, 指出用户正在移动的道路信息, 用该路段表示压缩轨迹, 大大提高了轨迹数据的存储量. 本文结合停留点法、地图匹配、开放窗口法以及关键点法, 提出一种能够保存 GPS 轨迹时空特性的在线轨迹压缩算法.

## 1 GPS 轨迹压缩方法

### 1.1 停留点法

在车辆行驶过程中, 会出现车辆在一段时间内停止在某个地点的情况. 通过停留点法判断可以有效的减少车辆临时停车或交通拥堵情况下导致的相邻地点的重复定位, 去除冗余点.

停留点的判断过程如下所示. 已知  $P_i$  是 GPS 轨迹中的某点,  $Dis(P_s, P_e)$  表示  $P_s$  与  $P_e$  之间的距离.  $Dr$  表示距离的阈值, 它是一个定值.  $T(P_s, P_e)$  表示  $P_s$  与  $P_e$  之间的时间差.  $Tr$  表示时间的阈值, 满足以下公式即可求解出车辆行驶过程中的停留点.

$$s < i < e \quad (1)$$

$$Dis(P_s, P_i) \leq Dr \quad (2)$$

$$Dis(P_s, P_{e+1}) > Dr \quad (3)$$

$$T(P_s, P_e) \geq Tr \quad (4)$$

在轨迹段中, 判断出用户的停止点, 并对停止点进行压缩. 假设停留点的定位分布如图 1 所示.  $S(x, y, t_s, t_e)$  表示停留点压缩后有效的轨迹信息.

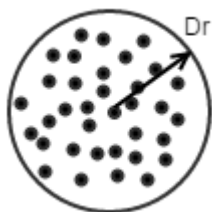


图 1 停留点定位分布图

压缩后的  $S$  的轨迹信息的计算方式如下所示,  $PNum$  表示停留点的总个数.

$$S.x = \sum_{i=s}^e P_i.x / PNum \quad (5)$$

$$S.y = \sum_{i=s}^e P_i.y / PNum \quad (6)$$

$$S.t_s = P_s.t_s \quad (7)$$

$$S.t_e = P_e.t_e \quad (8)$$

### 1.2 地图匹配

地图匹配<sup>[7]</sup>是利用算法将车辆定位的信息修正到数字地图的路段上, 从而获得车辆相对于数字地图的位置. 具体的匹配流程如图 2 所示:

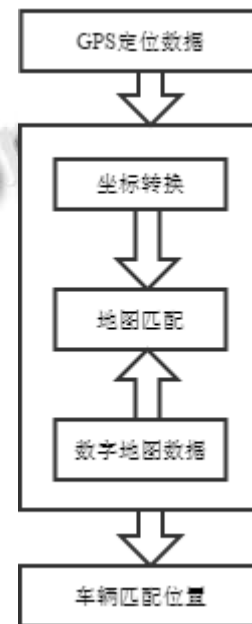


图 2 地图匹配流程

地图匹配算法主要分为基于几何信息的地图匹配, 基于拓扑关系的地图匹配, 基于概念论的地图匹配以及其他基于模糊逻辑的地图匹配算法, 本篇文章中首先从最基本的匹配算法入手, 结合实际问题的需要来进行分析. 在 GPS 定位过程中, 可能会出现定位误差的问题. 将定位信息进行地图匹配, 可以将轨迹点匹配到相应的道路上, 从而减小了定位误差. 如图 3 所示,  $C_1$ 、 $C_2$ 、 $C_3$  分别表示起始节点到道路端节点的道路向量,  $P_{i-1}P_i$  是 GPS 的定位点,  $l_i$  表示  $P_{i-1}P_i$  之间的向量. 道路在  $C_3$  之后出现交叉口, 要将  $P_i$  点匹配到对应的道路上, 前提是  $P_{i-1}$  已匹配到确定的道路上.

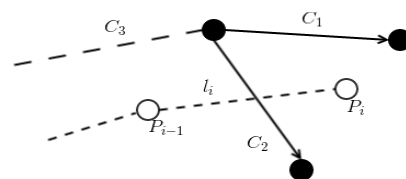


图 3 地图匹配

$$s_d(p_i, c_j) = u_d \cdot \alpha \cdot d(p_i, c_j)^{n_d} \quad (9)$$

$$s_a(p_i, c_j) = u_a - \cos(a_{i,j})^{n_a} \quad (10)$$

$$s = s_d + s_a \quad (11)$$

在以上公式中,  $u_d, n_d, u_a, n_a$  是常量值,  $d(p_i, c_j)$  表示  $p_i$  到  $c_j$  的距离.  $a_{i,j}$  表示向量  $l_i$  与  $c_j$  之间的角度差.  $s_d$  是  $p_i$  到  $c_j$  距离的评估指标,  $s_a$  是  $p_i$  到  $c_j$  之间角度的评估指标, 当公式(11)中  $s$  的值越大时, 则表示该道路的匹配是最佳的匹配结果.

### 1.3 开放窗口压缩方法

为了支持在线轨迹压缩, Meratnia 等人提出了 Opening Window(开放窗口压缩方法). OW 方法利用 SED( Synchronized Euclidean Distance, 同步欧氏距离) 来进行轨迹点到轨迹段距离的计算. 其中 SED 的计算如图 4 所示.

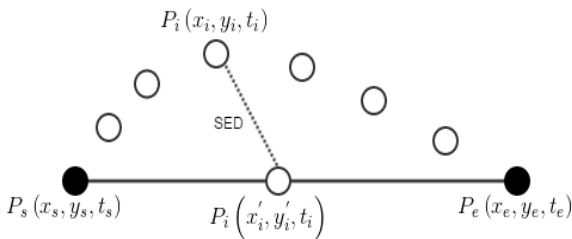


图 4 SED 计算

在上图中  $P_s, P_e, P_i$  是原始的轨迹定位点,  $P_s$  到  $P_e$  的轨迹被压缩为  $P_sP_e$  轨迹段, 原轨迹段  $P_sP_e$  之间的  $P_i$  根据公式(12)(13)在  $P_sP_e$  的投影点  $P_i'$ .

$$x_i' = x_s + \frac{t_i - t_s}{t_e - t_s} (x_e - x_s) \quad (12)$$

$$y_i' = y_s + \frac{t_i - t_s}{t_e - t_s} (y_e - y_s) \quad (13)$$

在 OW 算法中, 算法首先选取一个采样点作为起始点, 后续的轨迹点逐个加入窗口中. 窗口每加入一个轨迹点, 窗口值自动加 1, 并检查初始点到浮动点和窗口中新加入的轨迹点组成的轨迹段的 SED 是否大于阈值. 将第三个点作为浮动点, 即起始点后的第二点作为浮动点. 如上图中所示  $P_s$  为起始点,  $P_e$  为浮动点. 计算起始点到浮动点之间最新来的轨迹点的 SED 是否大于阈值, 例如上图中的  $P_i$  点, 如果 SED 小于等于阈值, 浮动点向序列方向后移一个; 若 SED 大于阈值, 则停止该窗口. 选取新的起始点的方式有两种策略, 分别是普通开放窗口方法和向前开放窗口方法. 在阈值相同的情况下, 向前开放窗口的执行速度要大于普

通开放窗口方法. 在向前开放窗口算法中选择上个窗口中倒数第二个轨迹点作为新的起始点.

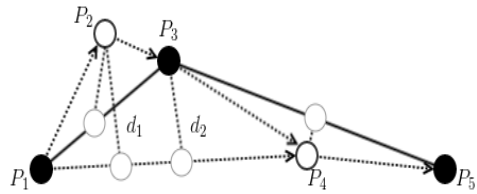


图 5 开放窗口法

如图 5 所示,  $P_1, P_2, P_3, P_4, P_5$  为原生轨迹点, 在  $P_1, P_2, P_3$  进入窗口时, 选取  $P_1P_3$  点作为压缩轨迹段, 计算  $P_2$  到  $P_1P_3$  的 SED 的值.  $P_2$  点到  $P_1P_3$  上的 SED 小于阈值, 继续向窗口中加入下一个轨迹点, 当  $P_4$  加入窗口后,  $P_2, P_3$  到  $P_1P_4$  的 SED(即  $d_1$  和  $d_2$ )都超过了阈值. 根据向前开放窗口算法, 应该选取引起误差  $P_4$  的前一点  $P_3$  作为新的起始点, 循环直到某条路段的轨迹点结束. 当用户移动的道路发生改变时, 起始点和浮动点也相应变成该道路的起点和终点.

### 1.4 关键点法

在车辆行驶过程中, 有一类定位点在轨迹中具有很重要的作用, 即速度发生变化. 例如, 车辆在转弯时会发生车速以及方向的变化. 而这一类点在压缩过程中是必须要保留的点. 在 GPS 轨迹中的两点进行连线, 我们进行角度以及速度的判断. 当速度和角度之差满足公式(14)(15)时, 则该点是必须要保留的点. 通过关键点的判断可以有效的保持轨迹点的形状和时空特性.

$$S(P_i, P_{i-1}) > Sr \quad (14)$$

$$A(P_i, P_{i-1}) > Ar \quad (15)$$

在给定的轨迹中, 根据公式判断出是否为关键点, 例如图 6 所示.  $P_1, P_2, P_3$  为轨迹的点, 假如  $P_3$  为当前点, 计算出  $P_1P_2$  和  $P_2P_3$  的角度差, 判断是否大于  $Ar$ (阈值为常量), 并且计算出  $P_1P_2$  与  $P_2P_3$  之间的速度差, 判断是否大于  $Sr$ (阈值为常量). 两者同时满足条件, 则  $P_3$  点位关键点.

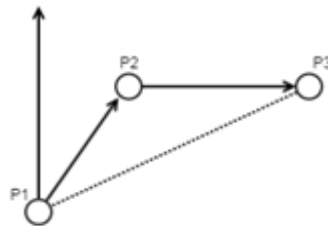


图 6 关键点判断

## 2 轨迹数据压缩实验及分析

### 2.1 实验步骤

步骤一. 首先利用停留点法对定位数据进行第一轮筛选, 利用公式(1)、(2)、(3)、(4)进行停留点判断, 判断为停留点之后, 使用临时变量进行保存, 然后根据(5)、(6)、(7)、(8)对临时变量中的点进行计算, 计算出该停留点的平均位置, 这样就将停留点收敛在一个位置, 也减少后期轨迹压缩的计算. 在  $D_r$  以及  $T_r$  阈值选取上, 应当根据实验中提供的数据进行设置. 例如本文中  $T_r$  的值为间隔定位的时间 5s. 根据速度以及定位误差分析  $D_r$  设置的范围则是最好在 5~25 米之内.

步骤二. 使用 Brakatsoulas 的匹配方法进行地图匹配, 使用(9)、(10)对候选路径进行评估, 由(11)来确定匹配路径. 将定位点匹配到道路上, 也是减少定位误差的过程. 实验中的参数设置分别为  $u_d=10$ ,  $n_d=1.4$ ,  $u_a=10$ ,  $n_a=4$ ,  $\alpha=0.17$ .

步骤三. 使用开放窗口法和关键点相结合的方法对匹配后的轨迹进行压缩. 首先根据(14)、(15)判断该点是否在角度和速度方面都发生变化且超过阈值, 如果是, 即为关键点, 直接进行保留, 如果不是, 利用开放窗口方法根据公式(12)、(13)计算该点的 SED 是否大于阈值, 大于则保留, 小于则舍弃. 在本次实验中, 阈值  $S_r$  设置为上个轨迹点速度的 1/4, 阈值  $A_r$  设置为 30. 重复步骤一、步骤二、步骤三直到轨迹结束.

### 2.2 实验结果及分析

实验数据来自某物资运输公司 50 辆车的 276 条轨迹. 定位数据为每 5s 钟定位一次, 定位信息包括经度、纬度、定位时间、速度等信息. 为了进行不同的 SED 阈值, 本文中 SED 阈值的变化值从 10 到 100, 并对 SED 误差、速度误差以及压缩率进行分析.

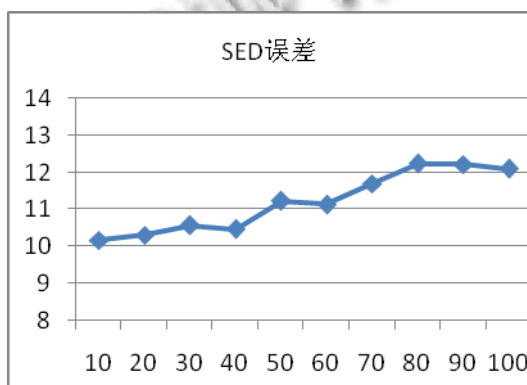


图7 SED 误差

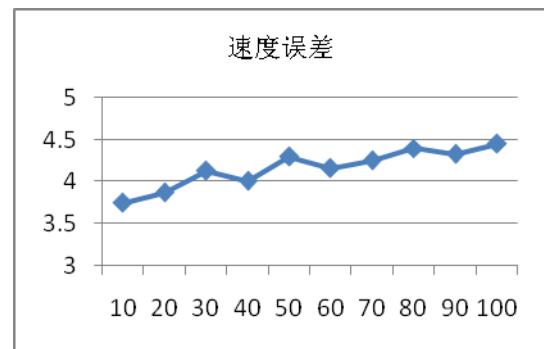


图8 速度误差

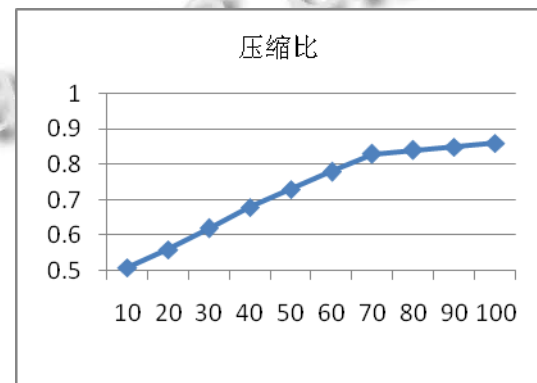


图9 压缩比

实验结果表明, 随着 SED 阈值增大, 压缩率增大, SED 误差和速度误差的变动的范围较小, 比较好的保证了轨迹的时空特性. 从图 7、图 8 得知, 速度的误差受 SED 误差的影响. 从图 9 得知随着 SED 阈值的增大, 压缩率在不断增加, 最高能达到 87% 左右. 结论得出在保证较高压缩率的情况下, 在一定程度上保证了轨迹的速度、时间等时空特性.

本文针对物流车辆监控结合目前已有的几种算法提出了以上的轨迹压缩算法. 旨在减少数据传输成本和数据存储空间. 本文研究中还存在一些不足之处, 如十字交叉路口地图匹配的精确度、在线压缩算法效率等问题. 另外, 在分析轨迹的同时, 还发现有很多轨迹重合的情况, 即匹配到相同的道路上. 在实际物流运输过程中, 如果是相同的起始点和终点, 就会存在选择相同路径的情况. 在已有的方法中, 通常使用车辆在道路的入点和出点代表车辆在该道路上的轨迹, 压缩率达 95% 以上, 但是这种方式只是大致描述了车辆的行车路径, 失真率较大. 如何重用相似轨迹, 并保存当前轨迹时空特性, 将需要进一步的研究.

## 参考文献

- 1 罗丽云.基于 WebGIS 的车辆监控系统的设计与实现[学位论文].南昌:南昌大学,2008.
- 2 李清泉,黄练,谭文霞.基于道路特征的海量 GPS 监控数据压缩方法.武汉大学学报:信息科学版,2008,33(4):337-340.
- 3 冯神柱.路网轨迹数据的压缩存储技术研究[学位论文].杭州:杭州电子科技大学,2014.
- 4 Douglas DH, Peucker TK. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, 1973, 10(2): 112-122.
- 5 Meratnia N, Rolf A. Spatiotemporal compression techniques for moving point objects, Advances in Database Technology-EDBT 2004. Springer. 2004. 765-782.
- 6 Schmid F, Richter KF, Laube P. Semantic trajectory compression. Proc. 11th International Symposium, SSTD. 2009, 5644. 411-416.
- 7 Brakatsoulas S, Pfoser D, Salas R, Wenk C. On map-matching vehicle tracking data. Proc. of the 31st Very Large Data Bases International Conference. 2005. 853-864.

WWW.C-S-A.ORG.CN

WWW.C-S-A.ORG.CN