

# 基于时空特性的 GPS 轨迹数据压缩算法\*

张达夫<sup>1</sup> 张昕明<sup>2</sup>

(1. 辽宁大学信息学院 沈阳 110036; 2. 哈尔滨工业大学交通科学与工程学院 哈尔滨 150001)

**摘 要** 车辆 GPS 数据中蕴含的车辆轨迹信息具有重要的理论和应用价值。为减少轨迹数据存储空间,提高数据分析及传送速度,提出了一种基于时空特性的轨迹数据压缩算法,根据 GPS 数据点的时空三维特征,计算轨迹特征点判断的距离标准,更准确的提取轨迹信息。算例分析结果表明,基于时空特性的 GPS 轨迹数据压缩算法大幅度减少轨迹压缩误差,提高压缩效率。

**关键词** GPS 轨迹数据;数据压缩;时空特性

中图分类号:U491.1+16 文献标志码:A doi:10.3963/j.issn.1674-4861.2013.03.002

## 0 引 言

随着技术的成熟和经济的发展,全球卫星定位系统(GPS)在车辆上的应用已经非常普遍。无论是用于导航还是定位,GPS 都存储着海量的数据信息。据统计,如果每 10 s 采集一次 GPS 数据,则 100 Mb 的存储空间只够容纳 400 辆车 1 d 的数据<sup>[1]</sup>。随着车辆数和时间的增多,数据量的扩张难以置信。庞大的数据量为数据的存储、查询、分析及传送造成很大的困难。所以,对 GPS 数据的压缩算法的研究成为这一领域的热点问题之一。

GPS 数据中蕴含的车辆行驶轨迹信息,对于交通科学的理论研究及车辆营运管理的实践工作,具有重要价值。在对车辆轨迹进行分析研究时,其实只需要存储准确描述车辆轨迹的信息特征点,其他数据点可以简化处理,这是对 GPS 轨迹数据压缩的基本思想。GPS 轨迹数据压缩的目的是在保留数据所包含的信息的前提下,尽可能的减少数据量,缩小数据所占用的存储空间。即在保证车辆轨迹准确性的前提下,去除冗余定位点,从而减少数据量。但在简化数据点的同时,必然会丢失一定量的信息,所以目前出现的轨迹数据压缩算法,是在数据信息的准确性和数据存储空间两者之间进行权衡。

## 1 轨迹数据压缩算法

轨迹数据压缩算法一般分为 2 大类,一是将运动轨迹进行分段线性化,由于其算法形式简单,计算复杂度低,是最常用的算法;另一种是非线性的轨迹拟合,如 Bezier 曲线<sup>[2]</sup>,非线性拟合更接近真实轨迹,但是其算法复杂,计算量大。车辆的运动轨迹受到实际交通路网的限制,所以线性化的表示方法能够很好的描述现实中车辆的运动状态。因此,线性化方法更适合于车辆轨迹数据的压缩。Top-Down 算法及 Opening Window 算法是具有代表性的 2 种线性化压缩算法。

### 1.1 Top-Down 数据压缩算法

Top-Down 算法根据首末点确定轨迹的大致方向,然后计算所有系列点在该方向的偏离值,偏离值大于给定阈值的即认定为轨迹特征点,将所有轨迹特征点存储、完成数据压缩<sup>[3]</sup>。

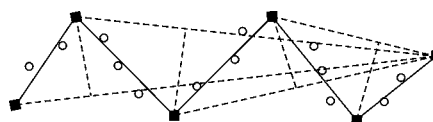


图 1 Top-Down 轨迹压缩算法

Fig. 1 Top-Down trajectory compression algorithm

如图 1 所示,Top-Down 算法将数据序列的第一个点定义为起始点,将最后一个点设为浮动点。①计算其他所有数据点到起始点与浮动点之

收稿日期:2013-01-05 修回日期:2013-05-10

\* 国家高技术研究发展计划(863 计划)项目(批准号:2012AA112310)资助

第一作者简介:张昕明(1987—),博士研究生,研究方向:智能交通系统. E-mail:12b332001@hit.edu.cn

间连线的垂直距离,找出距离最大的点;②如果最大距离大于给定阈值,则原线段分割为 2 部分,最大距离点变为前一线段的浮动点,同时为后一段线段的起始点,储存该最大距离点;③返回步骤 1。如此循环,直到所有距离均小于阈值,则所存储的点构成的轨迹,即为压缩后的轨迹。

该算法只能针对静态数据计算,是一种批处理算法。且该算法只适用于明确轨迹的起点和终点的情况下。

## 1.2 Opening Window 数据压缩算法

Opening Window 算法与 Top-Down 算法的特征点判断标准类似,同样认为与原方向偏离最大的点为特征点,但 Opening Window 算法的起始点和浮动点的选取方法不同。

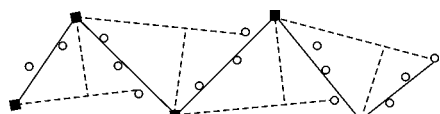


图 2 Opening Window 轨迹压缩算法

Fig. 2 Opening Window trajectory compression algorithm

如图 2 所示,Opening Window 算法选取数据系列中第一个点作为起始点,第三个点设为浮动点。①计算所有位于起始点和浮动点之间数据点到起始点与浮动点之间连线的垂直距离;②如果所有距离都小于给定阈值,则浮动点沿序列方向向后推移一个,返回步骤 1;③若出现偏移距离大于阈值的点,则存储该点,且该点设为起始点,其后的第三个点作为浮动点,返回步骤 1。循环结束后,所存储的点构成的轨迹,即为压缩后的轨迹。

Opening Window 算法可以对动态添加的数据进行计算,是一种实时的压缩算法。该算法只需知道轨迹起点,不需确定轨迹终点。

## 2 基于时空特性的轨迹数据压缩算法

无论是 Top-Down 算法还是 Opening Window 算法,都是以数据点偏离轨迹方向的垂直距离作为轨迹特征点的选取标准,仅考虑了 GPS 数据的空间特性。而 GPS 轨迹数据是具有时间序列特征的空间数据,车辆 GPS 数据不仅包含车辆的空间位置——经纬度  $(x, y)$ ,同时也记录了时间信息  $t$ ,这 3 个数据准确的记录了车辆的时空位置  $(x, y, t)$ 。此外,GPS 数据还包含车辆的瞬时

速度及车头朝向 2 个信息。即每条数据记录了车辆在某一时空位置在哪、该时刻的行驶速度如何及车辆正朝哪个方向行驶。

GPS 数据所记录的不仅具有线性的特征,同时也是具有时间序列特性的轨迹特征。所有仅利用位置上的偏移距离作为轨迹特征点的提取并不全面,对于其时间特性的考虑同样重要。

如图 3 所示,轨迹起点为  $P_s$ ,轨迹终点为  $P_e$ , $P_i$  为原始轨迹数据点,各点的时空三维坐标如图所示。则:

整个轨迹所经历的时间  $\Delta e = t_e - t_s$ ,由起点  $P_s$  到  $P_i$  所经历的时间  $\Delta i = t_i - t_s$ ,综合考虑数据点的时空特性,点  $P_i$  相对于轨迹方向的偏移距离应为线段  $P_i P'_i$  的长度,点  $P'_i$  的坐标计算方法如下:

$$x'_i = x_s + \frac{\Delta i}{\Delta e}(x_e - x_s) \quad (1)$$

$$y'_i = y_s + \frac{\Delta i}{\Delta e}(y_e - y_s) \quad (2)$$

数据压缩过程中,将线段  $P_i P'_i$  的长度作为点  $P_i$  的偏移距离,与给定阈值进行比较判断。这个距离考虑了 GPS 数据点的时空三维特性,更符合车辆运行轨迹实际。

基于时空偏移距离,本文提出 GPS 轨迹数据压缩算法的通用代码表达如下:

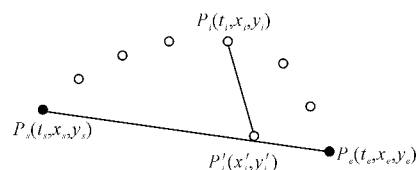


图 3 基于时空特性的轨迹偏差距离

Fig. 3 Spatiotemporal deviation distance

```

Comp( tra, max_dis)
If( len(tra)<=2 )
Then return tra
Else
{ store=false
e=2
While (e<= len(tra) and not store) do
{ i=2
While(i<=e and not store) do
{
Δe=tra[e]t - tra[1]t
Δi=tra[i]t - tra[1]t
(xi', yi') = tra[1]loc + (tra[e]loc - tra[1]loc) Δi/Δe
If ( dis(tra[i]loc, (xi', yi') ) >= max_dis )

```

```
Then store=true
Else i=i+1
}
If( store )
Then return
[tra[1]]++Comp(tra[i, len(tra)], max_dis)
e=e+1
}
If( not store)
Then return [tra[1], tra[len(tra)]]
}
```

变量解释表见表 1。

表 1 变量解释表  
Tab. 1 Variables table

变量名	涵义
<i>tra</i>	轨迹 <i>tra</i>
<i>len(tra)</i>	<i>tra</i> 中所包含的数据点数量
<i>tra[i]</i>	<i>tra</i> 中的第 <i>i</i> 个点
<i>tra[k, m]</i>	<i>tra</i> 中从第 <i>k</i> 点到第 <i>m</i> 点的子轨迹
<i>P<sub>t</sub>, P<sub>loc</sub></i>	数据点 <i>P</i> 的时间属性及空间位置
<i>dis(P, Q)</i>	返回数据点 <i>P</i> 和点 <i>Q</i> 的空间距离
<i>tra++s</i>	合并轨迹 <i>tra</i> 和轨迹 <i>s</i>

本文提出的基于时空特性的轨迹压缩算法,只需要明确轨迹数据的起始点,就能够根据偏移距离的判断标准,综合考虑数据的时空属性,准确提取轨迹数据的特征点,在精确保留轨迹信息的前提下,对数据进行压缩,减少数据量。

### 3 算例分析

为检验算法的应用性和性能,利用哈尔滨市出租车 GPS 数据作为实际算例数据,选取一辆出租车在上午 10:00:52 到 11:53:14 时之间的 GPS 数据,数据每 30 s 采集 1 次,即共有 200 个数据点。

由于本文提出的基于时空特性的压缩算法在对数据的处理方式上与 Opening Window 算法相似,即只需要轨迹起点就可以进行数据压缩。所以分别用 Opening Window 算法和本文提出的基于时空特性的轨迹数据压缩算法对算例数据进行压缩。分别对压缩误差及压缩率 2 个参数进行比较。压缩误差定义为原始数据点与压缩后轨迹上对应的拟合数据点之间的距离,同样考虑数据点的时间特征及空间位置<sup>[5]</sup>,即

$$\delta = \sum_{i=1}^n \sqrt{(x_{ii} - x'_{ii})^2 + (y_{ii} - y'_{ii})^2} \quad (3)$$

为进行深入的比较,在偏差距离阈值的选取上,本文从 30~100 m 之间选取了 15 个阈值。压

缩率计算结果见图 4。

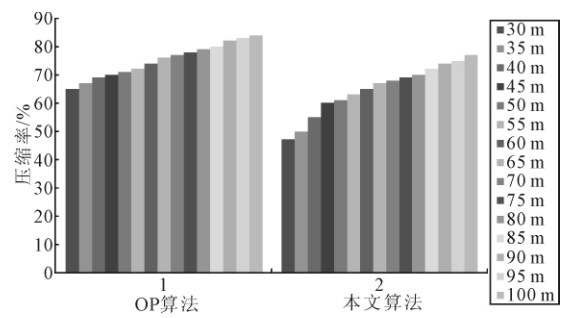


图 4 压缩率比较

Fig. 4 Comparison of compression rate

由图 4 可知,本文提出的算法在数据压缩率上较 OP 算法有小幅度的改进。同时可以得出,无论哪种算法,数据压缩率都随着阈值的增大而增加,说明选取的阈值越大,能够保留的数据点就越少,压缩率越大。随着阈值的增加,也带来了信息丢失的风险。

压缩误差计算结果见图 5。

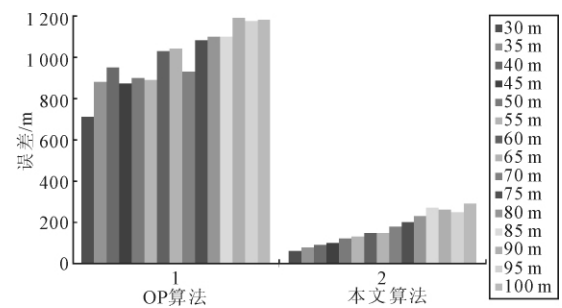


图 5 压缩误差比较

Fig. 5 Comparison of compression error

由图 5 可见,由于考虑了数据点的时间特性,所以基于时空特性的压缩算法较 OP 算法在压缩精度上有明显的改进。此外,对于 OP 算法,阈值选取的不同对于轨迹压缩精度有着较大影响,而本文提出的基于时空特性算法则对阈值的敏感性不高,因此可以选取较大的阈值以提高压缩率,而不会影响轨迹的描述。

### 4 结束语

对于车辆的 GPS 轨迹数据压缩,线性化的压缩算法因其计算简单,拟合度好而普遍被采用。但是现有的压缩算法没有完全利用 GPS 数据信息,只考虑 GPS 数据中的空间位置信息,造成压缩效率不高且信息丢失严重。基于时空特性的 GPS 轨迹数据压缩算法,考虑到了 GPS 数据点的时空三维信息,建立了新的偏差距离计算方法,实践证明,该算法能够大幅减少信息丢失,减少轨迹

提取的误差,并能够提高数据压缩效率。

GPS 除记录了车辆的时空三维信息,同时还记录了车辆的即时速度和方位角等信息,这些数据对于轨迹数据压缩同样具有重要作用。如轨迹特征点前后的数据一般减速和加速转换的特征等。如何利用速度及方位角数据,进行更为准确高效的轨迹数据压缩,是今后研究的重要方向。

### 参考文献

- [1] Meratnia N, de By R A. Spatio-temporal compression techniques for moving point objects[J]. Computer Science. 2004,2992:765-782.
- [2] Foley J D, van Dam A, Feiner S K, et al. Computer graphics: principles and practice[M]. Tokyo: Ltd. Ohmsha,2001.
- [3] Muckell J, Hwang J H, Lawson C T, et al. Algorithms for compressing GPS trajectory data: an empirical evaluation [C] // Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, New York: ACM, 2010:402-405.
- [4] Keogh E J, Chu S, Hart D, et al. An online algorithm for segmenting time series[C]//Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on. San Jose, CA: IEEE, 2001: 289-296.
- [5] Potamias M, Patroumpas K, Sellis T. Sampling trajectory streams with spatiotemporal criteria[C]//Scientific and Statistical Database Management, 2006. 18th International Conference on. Washington: IEEE, 2006:275-284.

## A Spatiotemporal Compression Algorithm for GPS Trajectory Data

ZHANG Dafu<sup>1</sup> ZHANG Xinming<sup>2</sup>

(1. College of Information, Liaoning University, Shenyang 110036, China;

2. School of Transportation Science and Engineering,  
Harbin Institute of Technology, Harbin 150001, China)

**Abstract:** GPS trajectory data of vehicles is valuable for important theoretical work and practice. Compressing trajectory data are conducive to storage, computation and transmission. In this paper, a spatiotemporal compression algorithm for GPS trajectory data is proposed. According to the spatiotemporal three-dimensional characteristics of GPS data points, a new method for calculating the deviation distance is utilized to extract the trajectory data more accurately. Results of a case study show that the algorithm is able to significantly reduce the errors and improve compression efficiency.

**Key words:** GPS trajectory data; data compression; spatiotemporal characteristics