

# 一种路网拓扑约束下的增量型地图匹配算法

朱 递<sup>1</sup> 刘 瑜<sup>1</sup>

1 北京大学地球与空间科学学院遥感与地理信息系统研究所,北京,100871

**摘 要:**着眼于低频浮动车轨迹数据,对地图匹配问题进行了抽象,并分析了影响匹配结果的几何约束与拓扑约束。针对 GPS 采样的低频性和城市路网的复杂性,提出了一种路网拓扑约束下的增量型地图匹配算法(topology-constrained incremental matching algorithm, TIM)。选取北京市浮动车的 GPS 样例轨迹数据进行匹配,结果表明,该匹配算法在不同复杂程度的城市路网下均表现较好。

**关键词:**地图匹配;低频 GPS 轨迹;路网拓扑;增量

**中图法分类号:**P208; P283.1      **文献标志码:**A

在大数据时代,浮动车技术广泛应用,带来了海量的城市车辆轨迹数据。这些数据一方面为研究城市空间结构提供了新的视角,另一方面可以为各类智能交通系统应用提供研究基础<sup>[1-3]</sup>。其中,智能交通系统为了得到每条路段的交通流量、车速等信息,需要进行 GPS 轨迹的地图匹配。地图匹配是一种基于软件技术的定位修正手段,其基本思路是将定位装置所获得的车辆定位轨迹与数字地图数据库中的路网数据相联系,从而确定车辆的路网参考位置,或正确识别出车辆在特定时刻所处的路段<sup>[4-5]</sup>。由于获得的 GPS 数据质量存在不一致,路网数据精度也存在差异,国内外学者对于地图匹配算法的研究也多种多样。目前,对于高频采样的 GPS 数据和相对简化的路网数据,已有很多较为成熟的匹配算法<sup>[6-7]</sup>。然而,针对低频 GPS 采样数据和高复杂度的城市路网,成熟并实用的地图匹配算法并不多,算法的匹配效率和匹配精度也普遍难以达到研究和应用的要求。

本文针对低频 GPS 采样数据在复杂城市路网下的地图匹配问题,综合考虑影响地图匹配的几何约束与拓扑约束,提出了一种路网拓扑约束下的增量型地图匹配算法(topology-constrained incremental matching algorithm, TIM)。该算法能够较好地解决 GPS 低频性质带来的匹配不稳定性和复杂城市路网带来的计算复杂性,为相关研究提供思路。

## 1 地图匹配问题抽象与影响因素分析

### 1.1 地图匹配问题抽象

街道系统,即真实的路网,可以表示为  $\bar{N}$ ,移动对象的真实轨迹为  $\bar{T}$ ,移动对象在  $t$  时刻的真实位置为  $\bar{P}^t$ ,GPS 定位的  $t$  时刻位置为  $P^t$ 。低频 GPS 采样数据是指  $P^t$  的采样时间间隔大于 1 min。获得真实街道系统  $\bar{N}$  往往是不可行的,通常研究基于的都是路网的数字表达形式,记作  $N$ 。 $N$  由有限数量的弧段组成,每一个弧段  $A \in N$  是由一系列点  $(A^0, A^1, A^2, \dots, A^n)$  构成,其中  $A^0$  与  $A^n$  是弧段起始点和终止点,而  $A^1, A^2, \dots, A^{n-1}$  通常被称作形状点。

地图匹配问题就是将 GPS 定位位置  $P^t$  匹配到数字路网  $N$  中与之对应的弧段  $A$  上,从而确定真实街道系统中对应真实位置  $\bar{P}^t$  的道路  $\bar{A} \in \bar{N}$ 。针对不同的研究目的,有的研究还会进一步确定  $A$  弧段上  $\bar{P}^t$  所对应的具体位置,如图 1 所示。

### 1.2 几何约束影响

影响地图匹配算法设计和匹配结果的几何约束主要包括距离因素、航向因素与方位因素。国内外已有相关研究构建起相对完善的权重匹配模型来实现地图匹配。

随着采样点和弧段垂直距离  $D(>0)$  的减小,点和弧段的邻近性增加,因此,距离因素对匹配结

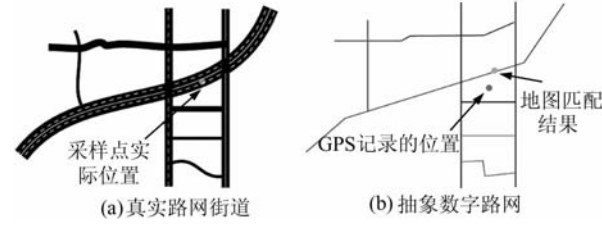


图 1 地图匹配问题抽象化图示  
Fig. 1 Abstraction of Map-Matching

果的权重  $W_D$  可以定义为:

$$W_D = \lambda_D / D \quad (1)$$

其中,  $\lambda_D (> 0)$  为可调节的权重参数。

同时,配置有航位推算系统 (dead reckoning, DR) 的 GPS 能够提供带有移动个体行驶航向的原始数据, GPS/DR 数据的航向参数非常精确, 通常能够用于辅助匹配和纠正匹配的结果。

方位因素对于地图匹配结果也有明显的影响, 常见因素有: (1) 相邻 GPS 采样点连线与路段相交程度  $W_i = \lambda_i \cos \theta$ ,  $\lambda_i$  为可调节的参数,  $\theta$  为相邻 GPS 采样点连线与路段的夹角。 (2) GPS 采样点相对于路段端点的位置  $W_r = \lambda_r \cos \alpha$ ,  $\lambda_r$  为可调节的参数,  $\alpha$  为采样点与路段端点连线和交叉口各路段的夹角。

### 1.3 拓扑约束影响

空间拓扑关系能够反映出实体和地理间的相互关系, 为了保证地图匹配结果的合理性, 构建拓扑约束至关重要。本文主要考虑了两方面的拓扑约束:

1) GPS 误差缓冲区与路网的叠加拓扑约束。GPS 定位存在误差, 基于概率准则定义误差椭圆, 可以有效缩小匹配搜索时的状态空间。从 GPS 接收机的输出数据中可以直接读出系统正向与北向测量误差的标准差  $\sigma_x$  和  $\sigma_y$ , 设  $\sigma_{xy}$  和  $\sigma_{yx}$  为协方差, 定位系统的方差、协方差矩阵  $\mathbf{M}$  便可以模型化为:

$$\mathbf{M} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix} \quad (2)$$

由此定义的定位误差椭圆半长轴  $a$  和半短轴  $b$  以及椭圆半长轴取向与正北方向的夹角  $\varphi$  为:

$$a = \sigma_0^2 \cdot$$

$$\sqrt{\frac{1}{2} \cdot (\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2) + 4\sigma_{xy}^2})} \quad (3)$$

$$b = \sigma_0^2 \cdot$$

$$\sqrt{\frac{1}{2} \cdot (\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2) + 4\sigma_{xy}^2})} \quad (4)$$

$$\varphi = \frac{\pi}{2} - \frac{1}{2} \arctan\left(\frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}\right) \quad (5)$$

在式 (3) ~ (5) 中,  $\sigma_0^2$  为单位权值的后验方差, 可以通过改变  $\sigma_0$  的值来调整误差区域的大小, 从而产生不同的误差椭圆置信度。

2) 路网拓扑连通约束。本文处理的是低频 GPS 采样数据, 而路网又是精度极高的复杂城市路网, 在匹配结果中, 跳路是很常见的 (即相邻两次匹配结果所对应的路段满足拓扑邻接的概率是很低的)。于是, 本文提出的路网拓扑连通约束是指相邻两个 GPS 采样点所匹配到的路段之间, 在路网中应当存在一条连通路径, 即匹配结果在路网表达上的连通性。

## 2 路网拓扑约束下的增量型地图匹配算法

本文旨在解决低频 GPS 采样数据在复杂城市路网下的地图匹配问题, 处理 GPS 低频性质带来的匹配不稳定性和复杂城市路网带来的计算复杂性显得格外重要。前人也有过将路网拓扑约束和最短路径思想同时考虑并设计算法的相关研究<sup>[8]</sup>。但是, 该研究只是在已知轨迹起始、终止位置的前提下, 进行最短路径的检索, 忽略了轨迹中段 GPS 采样点所带来拓扑约束影响以及轨迹之间的差异性。

本文提出一种拓扑约束下的增量型地图匹配算法。该算法首先结合 GPS 采样点的几何约束构建误差缓冲区, 通过将 GPS 误差缓冲区与路网缓冲区进行叠加, 得到 GPS 采样点的候选匹配边集。接着, 基于候选匹配边集的属性进行轨迹增量化处理, 并构建匹配时的拓扑约束。最后, 利用拓扑邻接约束下的  $A^*$  算法实现轨迹增量的地图匹配, 将 GPS 采样点和对应的路段实现匹配。TIM 算法流程如图 2 所示。

### 2.1 拓扑邻接字典的建立

TIM 算法在 GPS 轨迹预处理和路网简单拓扑构建完成之后, 首先需要完善并抽象路网拓扑关系, 使之成为可用的路网邻接模型。所谓路网拓扑邻接字典 (topological adjacency dictionary, TAD), 是指利用简单拓扑构建后的边表的 Source 和 Target 字段, 检索与每一条边相邻的所有边, 并将这个巨大的邻接关系存储为字典的形式, 其示意图见图 3。

将 TAD 转存为 JSON 文档格式, 计算机只需要读取该 JSON 文档便能够为后续的匹配计算提供路网信息。一个典型的 TAD 记录在 JSON 文档中表示为:

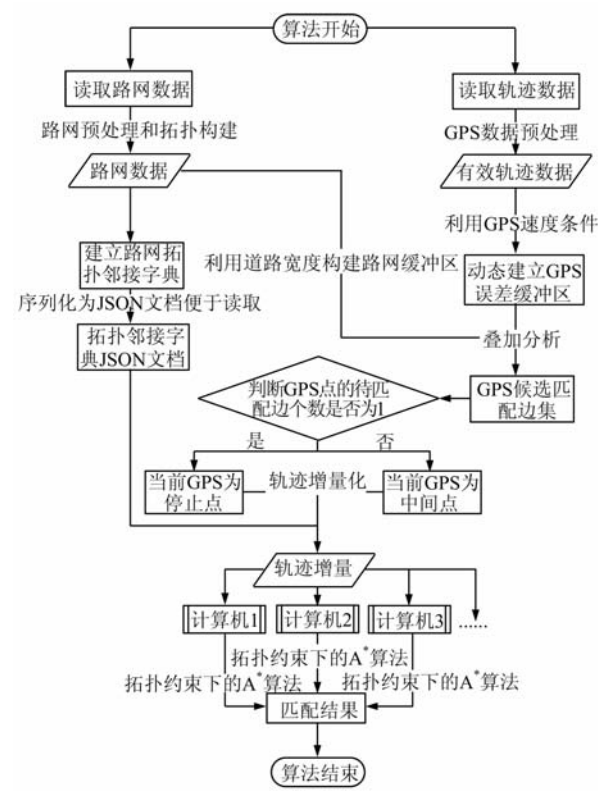


图 2 TIM 算法流程图  
Fig. 2 Flowchart of TIM Algorithm

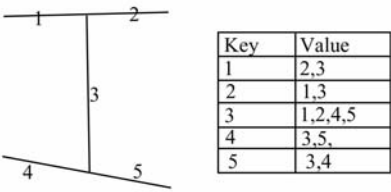


图 3 路网拓扑邻接字典构建示意图  
Fig. 3 Diagram of Topological Adjacency Dictionary

```
{"Key":1204,"Value":[1205,1160,1162,1206]}
```

Key 对应当前边的 ID, Value 对应与该边相邻的边的 ID。经测试,读取北京市路网数据的 JSON 文档并转换为字典格式,用时为 1 200 ~ 1 500 ms,是完全能够接受的。

## 2.2 GPS 数据误差缓冲区动态生成

地图匹配的关键假设是 GPS 定位存在误差,且误差是在一定范围之内的。于是,确定误差区域以便提取候选匹配道路便成为了非常重要的一环。TIM 算法同样运用构建 GPS 误差缓冲区的方式来确定 GPS 的待匹配边集。由于对于误差椭圆的包含和相交运算需要累计执行大量乘积与开方,前人在构建 GPS 误差缓冲区时常选取误差圆的方法。但是,多数研究在构建误差缓冲区时没有考虑不同 GPS 采样点之间的差异性。由于

随着 GPS 行驶速度的增大,定位误差会在一定程度上减少<sup>[6]</sup>。在 TIM 算法中,通过利用 GPS/DR 数据中所带有的速度字段,构建一个动态的误差缓冲区,从而进一步提高待匹配边集的有效性和准确性。

鉴于目前没有成熟的研究将 GPS 误差圆半径随速度的变化关系量化表示,在 GPS 轨迹数据预处理之后(包括纠偏、去除飘点和冗余数据等),综合考虑算法效率和匹配结果,TIM 算法采用简单的线性关系对其进行刻画,其关系为:

$$R = \begin{cases} 60 - \frac{1}{3}v\Delta t, & 0 \leq v \leq 90 \\ 30, & v \leq 90 \end{cases} \quad (6)$$

式中, $R$  为误差圆半径(m); $v$  为行驶速度(m/s); $\Delta t$  为单位时间(s)。由于轨迹预处理时已经将纠偏后偏离路网距离过大的采样点去除,根据实验结果,将  $R$  范围设置在[30 m, 60 m]能够较好地处理本文的需求。

生成的 GPS 误差缓冲区如图 4 所示。图 4 中紫色的连线为按照时间顺序的 GPS 采样点连线,考虑到速度因素,GPS 误差缓冲区大小不恒定,动态误差缓冲区的思想有非常重要的意义。



图 4 GPS 误差缓冲区动态生成效果图  
Fig. 4 Dynamically Building GPS Error Buffer

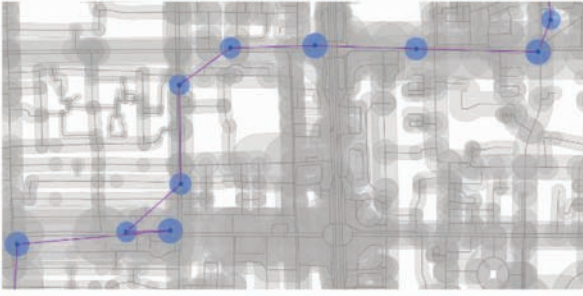
## 2.3 GPS 数据候选匹配边集的建立

在将 GPS 误差缓冲区与路网进行叠加分析时,利用原始数据中较完善的路段宽度字段,构建路网的双侧缓冲区,再将误差缓冲区与路网缓冲区求交,从而得到每一个 GPS 采样数据  $P^i$  的候选匹配边集  $E_{P^i}$ ,其效果如图 5 所示。

与建立路网 TAD 时的思路一样,本文也将 GPS 数据的  $E_{P^i}$  存储为 JSON 文档,以便于匹配时的快速检索与查询。一个典型的  $E_{P^i}$  记录表示为:

```
{"Key": "20121130073700.000000000", "Value": [1452, 1316, 2128, 1317, 1457, 1461, 1458, 2127]}
```

Key 表示 GPS 点的采样时间,与 GPS 数据

图 5 GPS 数据候选匹配边集  $E_{P^i}$  的建立Fig. 5 Generation of  $E_{P^i}$ 

一一对应, Value 对应该 GPS 可能匹配到的边的集合, 集合中的元素对应边的 ID, 与 TAD 中的 Key 相对应。

$E_{P^i}$  在 TIM 算法中具有非常重要的作用。首先, 对于每一个 GPS 采样点,  $E_{P^i}$  反应了该点对于轨迹匹配几何约束的贡献, 每一个采样点匹配的结果只能够在对应  $E_{P^i}$  中进行选择。其次,  $E_{P^i}$  有效地将 GPS 点和路网拓扑关系连接了起来, 通过检索  $E_{P^i}$  中边集的相邻边, 可以在路径搜索中进行有效的拓扑约束。另外,  $E_{P^i}$  为 GPS 轨迹提供了一种分类标准, 从而使得轨迹能够增量化,  $E_{P^i}$  的大小会直接影响到算法的效率。

## 2.4 轨迹增量化

在进行轨迹增量化处理之前, 需要对轨迹上的 GPS 点进行分类, 利用分类后 GPS 点的不同属性, 将轨迹划分为一系列增量。这种将轨迹增量化的思想参考了分治思想, 能够降低匹配的复杂度, 显著提高算法的效率。

当  $E_{P^i}$  的构建完成后, TIM 算法顺序遍历所轨迹  $T(P^1, P^2, P^3, \dots, P^i, \dots, P^t)$  中的每一个 GPS 采样点  $P^i$ , 并对  $P^i$  进行分类, 分类标准为:

- 1) 若当前  $P^i$  的  $E_{P^i}$  为空, 从  $T(P^1, P^2, P^3, \dots, P^i, \dots, P^t)$  中删除  $P^i$ 。
- 2) 如果当前  $P^i$  的  $E_{P^i}$  只包含一条边, 将  $P^i$  标记为停止点。
- 3) 如果当前  $P^i$  的  $E_{P^i}$  包含多条边, 将  $P^i$  标记为中间点。

通过 GPS 采样点分类后, 轨迹中的点被分成了停止点和中间点两类。停止点在轨迹中的匹配可靠性是非常强的, 即轨迹一定会通过该点对应  $E_{P^i}$  中的那条边, 中间点则能够有效地约束轨迹匹配时的抖动幅度。利用停止点的特殊性, 原始轨迹的匹配问题便可以进行分治, 划分为一个个增量地图匹配子问题, 同时很好地保证了匹配的准确性。

本文定义增量为一个停止点和下一个停止点

之间的轨迹片段, 抽象为  $I(P^{i_s}, P^{i_1}, P^{i_2}, \dots, P^{i_j}, \dots, P^{i_t})$ , 其中,  $P^{i_s}$  与  $P^{i_t}$  为停止点, 而  $P^{i_1}, P^{i_2}, \dots, P^{i_j}, \dots, P^{i_{t-1}}$  均为中间点。

在增量匹配阶段, 只保留那些起始点和终止点均为停止点的轨迹片段作为增量, 若  $P^{i_s}$  与  $P^{i_t}$  不能满足均为停止点, 则舍弃该轨迹片段。

## 2.5 轨迹约束下的 T-A\* 算法搜索

假设出租车司机是具有经验的路径选择者, 在目的地确定的条件下, 出租车司机会选择行驶时间最短的路线行驶。由于本文无法获得路网实时速度数据集, 假设全路网行驶速度恒定, 那么出租车司机会选择距离最短的路线来行驶。

经过了轨迹增量化后, 本文的研究目标转换为一个增量的匹配问题。轨迹增量具有确定的起始路段和确定的终止路段, 而中间点的候选匹配边集约束并减小了中间点可能匹配路段的搜索状态空间, TIM 算法采用基于轨迹约束的启发式 A\* 算法, 能够实现轨迹增量的快速地图匹配。区别于最常见的状态空间搜索, 如深度优先搜索和广度优先搜索, 启发式 A\* 算法的优点在于它并不是在完整路网所对应的状态空间下进行穷举, 而是对每一个搜索的位置进行评估, 并对所有评估过的位置进行评估值的排序, 将评估值最小的位置作为最优并继续搜索, 由此缩小了搜索的状态空间。

按照传统 A\* 算法, 节点  $n$  的估价函数  $f'(n)$  可以表示为:

$$f'(n) = g'(n) + h'(n) \quad (7)$$

$$g'(n) \leq g(n), h'(n) \geq h(n) \quad (8)$$

式中,  $f'(n)$  为节点  $n$  的 A\* 估价函数值;  $g(n)$  为状态空间中从初始节点搜索至节点  $n$  的实际代价;  $h(n)$  为节点  $n$  到目标节点的最佳路径的估计代价; 要求  $g'(n)$  恒小于  $g(n)$ , 而  $h'(n)$  恒大于  $h(n)$ 。计算时取  $g'(n)$  为起点到节点  $n$  的最短路径, 取  $h'(n)$  为节点  $n$  到目标节点的欧氏直线距离。可以证明, 这样的估价函数能够保证找到从起始位置到目标位置的最短路径。

受到 A\* 算法的启发, 本文针对路网的拓扑结构特点提出 T-A\* 算法。T-A\* 算法的状态空间中处理的对象是轨迹增量的中间点, 每一个位置的搜索结果是路段。

假设当前处理的 GPS 数据采样点为中间点  $P^i$ , 其候选匹配边集表示为  $E_{P^i} \{E_1, E_2, E_3, \dots, E_n\}$ , 其中,  $E_i$  的起始节点为  $N_{is}$ , 终止节点为  $N_{it}$ 。那么, 定义 T-A\* 算法的估价函数为:

$$f'(P^i) = \min_i \{g'(N_{is}) + l(E_i) + h'(N_{it})\} \quad (9)$$



式中,  $g'(N_{is})$  为增量起始采样点对应匹配边的终止节点到  $E_i$  起始点  $N_{is}$  的  $A^*$  最短路径长度;  $l(E_i)$  为求候选边  $E_i$  的长度的函数;  $h'(N_{it})$  为  $E_i$  终止节点  $N_{it}$  到增量终止采样点对应匹配边的起始节点的欧氏直线距离。

T-A\* 算法执行时, 首先遍历增量的全部 GPS 采样点, 根据 T-A\* 估价函数确定所有  $P'$  点的匹配边, 降低轨迹匹配的抖动性。接着, 可以利用传统  $A^*$  算法得到相邻两 GPS 采样点之间的路径则。根据上述流程便可以给出一个增量的完整匹配结果, 并按照一定格式输出。

### 3  算法应用与讨论

#### 3.1  数据预处理

从北京市 2009 年 11~12 月的 6 000 余辆出租车行车轨迹中选择了较具有代表性的出租车在 2009-11-30T06:32~2009-12-01T00:06 的轨迹作为原始样例, 经过火星坐标纠偏处理、重采样和去除飘点及冗余数据后, 样例轨迹的有效采样点从 1 624 点减少为 853 点。另外, 选取北京市城区内的路网数据, 并将其与编号为该出租车的行驶轨迹进行匹配, 城区路网数据如图 6 所示, 共计边 102 575 条。

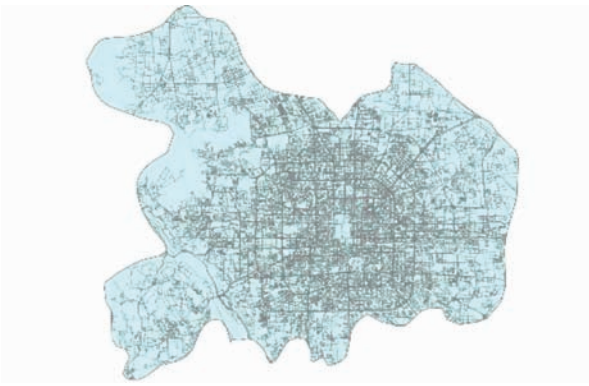


图 6  北京市城区路网示意图

Fig. 6  Road Network of Beijing Urban Area

统计分析发现, 样例轨迹平均采样间隔为 1.19 min, 满足低频数据的特点, 采样间隔最大为 55 min, 绝大部分为 1~3 min, 少量间隔小于 1 min 或大于 5 min。低频采样的特征且样例数据的速度起伏较大, 采样密度的空间差异非常明显 (见图 7)。仅从笔者了解到的情况来看, 处理如此低频的轨迹数据的相关研究非常少。

#### 3.2  匹配结果图

实验结果表明, TIM 算法在处理北京城市复杂路网下的低频采样数据时表现出较好的稳定

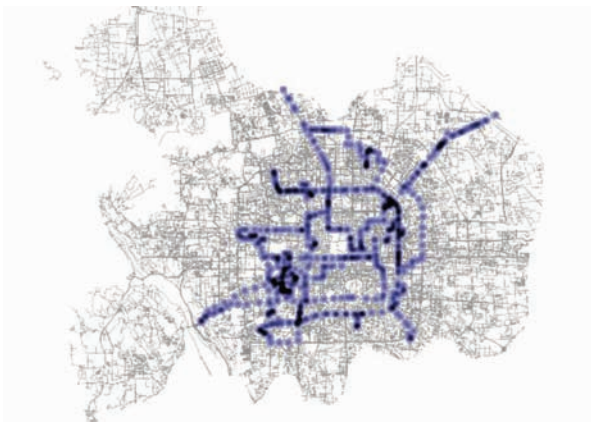


图 7  样例轨迹点密度图

Fig. 7  Dot Density Map of Sample Trajectory

性, 本文选择具有代表性的区域进行匹配结果的展示。

图 8~10 分别显示了 TIM 算法在带有立交的高速路段、特色明显的北京胡同路段以及采样距离较大的直线路段均能给出相对令人满意的结果, 图 8~10 中, 红色点为增量采样点。这是由于 T-A\* 算法充分考虑了几何约束与拓扑约束, 同时加入一些如路网等级、宽度等因素对匹配结果进行了修正。由于 TIM 对于路网拓扑结构的要求较高, 路段拓扑缺失可能会触发 T-A\* 算法在检索时的终止条件, 从而对增量片段再分段出现异常, 如图 11 所示。



图 8  07:00~07:59 机场高速路段匹配结果

Fig. 8  Match Results for Airport Expressway  
at 07:00—07:59

#### 3.3  算法效率与准确性分析

本文对于测试样例轨迹进行匹配的测试环境为 Intel(R) Core(TM) i5-3230M CPU @2.60 GHz、4 G RAM、WIN8 x64、VS2012 . NET Framework 4.5。在 TAD、 $E_{P'}$  和轨迹增量的构建完成后, 对于 72 个增量的匹配共计用时 116.64 s, 平均每个增量用时 1.62 s。由于平均每个增量含 11.03 个采样点, 每一个采样点的匹配以及相



图 9 09:58~11:08 北二环胡同  
Fig. 9 Match Results for Beierhuan Alleyway  
at 09:58-11:08



图 10 13:33~13:46 南二环护城河段  
Fig. 10 Match Results for Nanerhuan Moat  
at 13:33-13:46



图 11 15:59~16:08 玉渊潭南侧(拓扑异常)  
Fig. 11 Error Match at the South of Yuyuantan  
at 15:59-16:08

邻采样点间最短路径的检索需耗时 0.147 s, 总体效率可以满足普通实时 ITS 应用的需求。

另外, 本文的样例出租车轨迹真实数据不可知, 在进行  $A_N$  数量准确性分析时无法逐个检查。考虑到 GPS 采样点的数量并不是很大, 本文通过逐个采样点目视观察的方法判断每一个 GPS 点匹配是否正确(见式(10))。853 个 GPS 采样点得到正确匹配的采样点数量为 785,  $A_N = 92.03\%$ 。

$$A_N = \frac{\text{正确匹配的路段}}{\text{轨迹对应的全部路段}} \quad (10)$$

由于目视时有可能忽略路网拓扑因素, 直观结果和算法结果可能会出现不一致性, 这也有可能导致  $A_N$  的值不准确。

3.4 问题探讨

通过分析 TIM 算法的流程, 可以总结出两个会影响结果准确性的重要因素或参数。

1) 路网现势性问题以及拓扑邻接字典的准确性。如果基础路网边表的 Source 和 Target 字段和真实路网情况不一致, 或者由于路网导入空间数据库时拓扑容差的阈值过大, 则会导致构建的拓扑邻接字典出现拓扑错误或者拓扑缺失, 从而导致异常匹配。本文中并没有对基础路网的拓扑和准确性进行检查, 是有待商榷的。

2) GPS 数据候选匹配边集的大小。候选集平均大小会明显影响 T-A\* 算法的效率, 将  $E_{pt}$  的平均大小按照合理规则控制在 10 以内将会有效地保证匹配效率, 如图 12 所示。本文中, TIM 算法在误差圆半径与速度之间建立起简单的线性关系, 从而获得候选匹配边集, 进一步构建增量进行地图匹配, 可以较好地满足算法效率和准确性的要求。

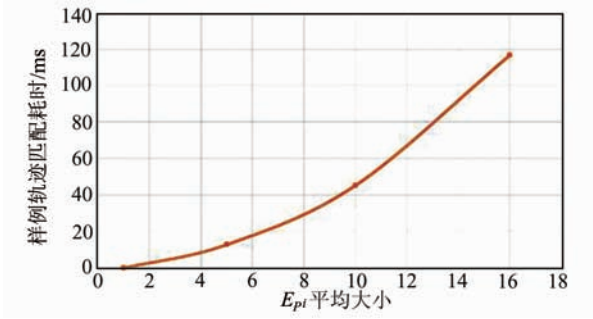


图 12 候选匹配边集大小与匹配效率的关系  
Fig. 12 Influence of  $E_{pt}$  Set on Matching Efficiency

4 结 语

本文提出了一种路网拓扑约束下的增量型地图匹配算法, 实现了不同城市路段的低频 GPS 采样数据地图匹配, 分析了拓扑邻接字典、GPS 误差缓冲区、候选匹配边集对于算法匹配结果的影响机制, 为今后有关地图匹配的研究提供思路 and 依据。

本文研究中尚未考虑航向与路段行驶方向的约束, 也未论证数据的代表性。后续研究可以从动态缓冲区大小选择的角度来探讨影响误差缓冲区的因素及相应关系。

参 考 文 献

[1]   Liu Y, Wang F, Xiao Y, et al. Urban Land Uses and Traffic “Source-Sink Areas”: Evidence from GPS-Enabled Taxi Data in Shanghai[J]. *Landscape and Urban Planning*, 2012, 106(1): 73-87

[2]   Liu X, Gong L, Gong Y, et al. Revealing Travel Patterns and City Structure with Taxi Trip Data [J]. *Journal of Transport Geography*, 2015, 43 (1): 78-90

[3]   Zhou Y, Fang Z, Thill J C, et al. Functionally Critical Locations in an Urban Transportation Network: Identification and Space-Time Analysis Using Taxi Trajectories[J]. *Computers, Environment and Urban Systems*, 2015, 52(1): 34-47

[4]   Quddus M A, Ochieng W Y, Zhao L, et al. A General Map-Matching Algorithm for Transport Telematics Applications[J]. *GPS Solutions*, 2003, 7(3):157-167

[5]   Greenfeld J S. Matching GPS Observations to Locations on a Digital Map[C]. The 81th Annual Meeting of the Transportation Research Board, Washington D C, 2002

[6]   Quddus M A, Noland R B, Ochieng W Y. Current Map-Matching Algorithms for Transport Applications: State of the Art and Future Research Directions[J]. *Transportation Research Part C*, 2007, 8 (15):312-328

[7]   White C E, Bernstein D, Kornhauser A L. Some Map-Matching Algorithms for Personal Navigation Assistants[J]. *Transportation Research Part C*, 2000, 8(1):91-108

[8]   Li Qingquan, Hu Bo, Yue Yang. Flowing Car Data Map-Matching Based on Constrained Shortest Path Algorithm[J]. *Geomatics and Information Science of Wuhan University*, 2013, 38(7):805-808 (李清泉, 胡波, 乐阳. 一种基于约束的最短路径低频浮动车数据地图匹配算法[J]. 武汉大学学报·信息科学版, 2013, 38(7):805-808)

An Incremental Map-Matching Method Based on Road Network Topology

ZHU Di<sup>1</sup>   LIU Yu<sup>1</sup>

1   School of Earth and Space Sciences, Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China

**Abstract:** The emergence of big spatio-temporal data brings brand new perspectives as well as challenges for us to investigate and understand urban space. Due to existence of GPS position error, it is inevitable to adopt the map-matching methods to map the spatio-temporal trajectories onto geographic space. This research focuses on the low-sampling trajectories of floating cars in urban road networks by formalizing the map-matching process and exploring the influence of both the geometric and topology constraints on matching results. To solve the problem of matching low-sampling GPS data in the context of complex urban road networks, we propose a topology-constrained incremental matching algorithm (TIM). Utilizing a sample GPS trajectory of Beijing float car as an example, the TIM algorithm is verified to be efficient and accurate give various road network complexity. Our study is valuable for the pre-processing of massive spatio-temporal data, and has the potential to benefit trajectory data mining and related urban informatics research in the future.

**Key words:** map-matching; low-sampling GPS trajectory; road network topology; increment

**First author:** ZHU Di, postgraduate, specializes in GIS, spatio-temporal big data and visualization. E-mail: zhudi-001@163.com

**Foundation support:** The National Natural Science Foundation of China, Nos. 41271386, 41428102.