

基于向量识别的启发式路径推测算法

吕卫锋 吴东东 诸彤宇

(北京航空航天大学软件开发环境国家重点实验室 北京 100083)

摘 要 浮动车数据主要是由车辆的轨迹点数据组成,是一种重要的原始数据,可以广泛地用于各种交通应用,如交通管理和控制、路况计算等。但是原始的车辆 GPS 数据存在定位误差,必须经过路径推测的修正处理才可以应用。传统的路径推测算法主要采用两种方法:渐增式和全局式。两种方法各有优缺点,渐增式方法计算速度快但准确性差,全局式方法准确性好但计算速度慢。通过综合考虑两种传统算法,文中提出了一种基于向量识别的启发式路径推测算法,该算法采用了启发式图搜索方式,导入几何运算的约束条件,根据车辆轨迹点所形成的向量与路网模型比较来进行启发式搜索,并选择车辆所有可能行驶的候选路径。根据全局择优的方式从整体进行比较,确定车辆最有可能的行驶路径。实验结果表明,这种算法能够在复杂路网下,比较准确地推测距离间隔较大的车辆轨迹点,并且能够实时高效地处理大规模数据。

关键词 路径推测;浮动车数据;GPS;启发式搜索

中图法分类号 TP301 **DOI号**: 10.3724/SP.J.1016.2009.01443

A Heuristic Path-Estimating Algorithm by Using Vector-Based Recognition

LU Wei-Feng WU Dong-Dong ZHU Tong-Yu

(State Key Laboratory of Software Development Environment, Beihang University, Beijing 100083)

Abstract Floating Car Data (FCD) is an important material for a broad range of application such as traffic management and control, traffic conditions computation and so on. However, in the original data exists error, and the data must be handled by path-estimating and be related to the road. The traditional path-estimating algorithms mainly use two methods: the incremental method and the global method. Both of them have advantages and disadvantages of themselves; while the global map-matching algorithm produces better matching results, the incremental algorithm produces results of lower quality faster. All things considering the two traditional algorithms, this paper proposes a heuristic path-estimating algorithm by using vector-based recognition. Firstly, the algorithm uses the heuristic search method, and it makes use of geometric operation to form the restriction, and make the comparison between the vector formed with the vehicular GPS points and the special road network model to search and select the vehicular possible traveling routes. Secondly, it globally compares all the vehicular possible traveling routes, and then chooses the optimal one. The result of testing demonstrates the efficiency of the algorithm both at accuracy and computational speed when handling the large-scale data of GPS tracking data even under the complex road network conditions.

Keywords path-estimating; Floating Car Data (FCD); GPS; heuristic search

1 引言

在智能交通领域,实时和动态的交通信息能为车辆出行、交通运输等提供有效的交通诱导和出行规划信息,从而达到节省出行时间、减少尾气排放等目的.为了实现上述目标,一个重要研究课题是高效地处理实时、动态的交通信息.

浮动车(Float Car Data)技术,也被称作“探测车(Probe car)”,是近年来国际智能交通系统(ITS)中所采用的获取道路交通信息的先进技术手段之一.其基本原理是:根据装备车载全球定位系统(GPS)的浮动车在车辆行驶过程中定期记录的车辆位置、方向和速度信息,应用包括路径推测和道路交通拥堵信息计算等相关的计算模型和算法进行处理,从而使浮动车位置数据和城市道路在时间和空间上关联起来,最终得到浮动车所经过道路的车辆行驶速度以及道路的行车旅行时间等交通拥堵信息^[1].如果在城市中部署足够数量的浮动车,并将这些浮动车的位置数据通过无线通信系统定期、实时地传输到信息处理中心,经过综合处理,就可以获得整个城市动态、实时的交通拥堵信息^[2].

由于浮动车所采用的 GPS 设备一般会有 15m 以上的圆周误差^[3],一方面,在定位数据的准确性上,由于浮动车所采用的 GPS 设备一般会有 15m 以上的圆周误差^[3],在推测过程中存在将轨迹点匹配到多条道路上的可能性.另一方面,在计算效率上,浮动车采集 GPS 轨迹点数据的采样率较低,一般在 20~300s,造成两个连续轨迹点跨越了较长距离,两个轨迹点间有可能存在多条可行驶的路径.传统的路径搜索算法主要采用两种方式:渐增式^[4]和全局式^[5].渐增式方法分别对每个轨迹点进行路网搜索,计算轨迹点与其附近的道路的距离,选择与轨迹点距离最小的道路作为搜索结果.全局式方法则使用曲线匹配的方式,连接连续的多个轨迹点形成轨迹曲线,与路网中的路径做几何曲线匹配,通过计算 Fréchet 距离的方法选择相似的道路作为匹配结果.渐增式方法计算简单,因此处理速度快,但是准确性差;全局式方法从整体上进行考虑,因此准确性好,但是由于计算复杂,处理速度慢^[5].

因此,应用于浮动车信息系统的路径推测算法既要有高效的计算速度,又要有较高的准确性.传统的方法无法很好地适用于这种应用.基于此,本文设计了采用启发式搜索路径的路径推测算法.首先在

算法的计算速度上,为了满足实时性的要求,采用了渐增式方法.一方面,重新设计了路网模型和路网拓扑,将其与算法进行整合^[6];另一方面,利用启发式搜索方式,算法在特殊的路网模型支持下,利用车辆轨迹点所形成的向量与路网相比较,导入三角形的几何理论作为约束,采用启发式的图搜索方式快速地搜索浮动车可能行驶的候选道路.两方面相结合有效地提高了算法的计算速度.其次,在算法的准确性上,采用全局择优的方法,通过树结构保存与车辆轨迹相符合的所有候选路径,然后比较每条路径上轨迹点的匹配权值,从整体上选择最接近车辆轨迹的一条路径作为推测结果.

本文第 2 节给出相关研究;第 3 节介绍算法使用的路网模型和路网拓扑;第 4 节对问题进行描述与分析;第 5 节描述算法的设计;第 6 节分析实验测试的结果;最后对本文进行总结,并给出下一步研究工作.

2 相关研究

路径推测是浮动车信息系统的关键技术,是正确计算道路交通状况的基础.路径推测是通过匹配车辆较大间隔的行驶轨迹点数据,搜索车辆正确行驶路径的技术,是地图匹配的一种处理方法.

传统的方法中,渐增式方法基于已完成点匹配的两点进行搜索,需要进行两次独立的地图匹配运算,然后基于路网拓扑进行道路的广度搜索找出车辆可能的行驶路径,搜索范围大、效率较低、准确性偏低;全局式方法则使用曲线匹配的方式,将整体轨迹点形成轨迹曲线,通过与路网道路之间计算 Fréchet 距离或弱 Fréchet 距离的方法选择候选道路作为匹配结果,计算量大、效率低.

3 路网模型和路网拓扑

路网数据是路径推测的基础,同时高效的路径推测算法应该整合相适应的路网模型和拓扑.常见的电子地图路网模型一般是由基于桌面 GIS(地理信息系统)的路网模型改进得来.这种模型的缺陷在于难以表达复杂道路的拓扑关系,对路网的描述比较单一,无法满足针对大规模浮动车数据处理的需求.因此,必须对原有路网模型进行改进,以适应高效路径推测的要求.基于此,本文通过对数据的分层处理和抽象,建立了新型的路网模型和路网拓扑.

3.1 路网模型

路网模型用于描述区域内的道路网,是由车辆轨迹点进行路径推测的基础.图 1 显示了路网模型的基本元素,包括节点和路段.整个路网模型可以表述为

$$\begin{cases} R_n=(N,S) \\ N=\{n|n=(x,y),x,y\in Coordinates\} \\ S=\{\langle m,n\rangle|m,n\in N\} \end{cases} \quad (1)$$

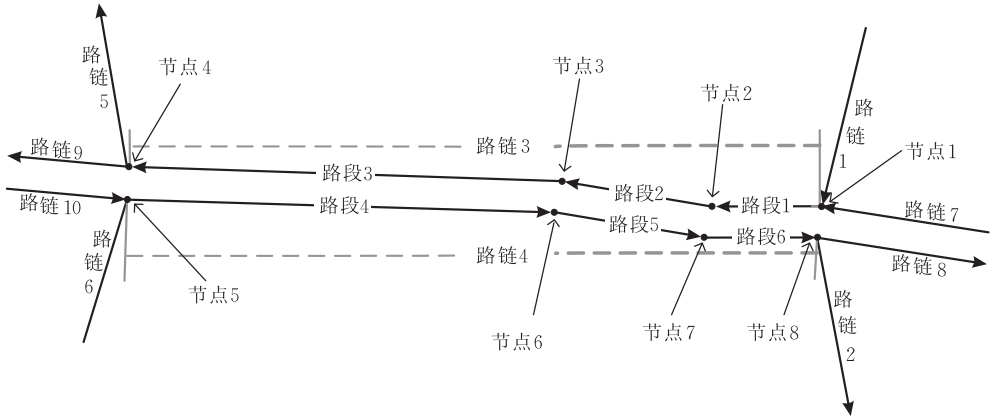


图 1 路网模型中的基本元素

若在路网 R_n 中,存在一个路段的序列 $\langle s_i, s_j, \dots, s_k \rangle$,其中每条路段的终止节点是下一个路段的起始节点,则此序列代表路网上的一条有向通路,如图 1 中,路链 3 和路链 4 分别是由 3 条路段构成的通路.

3.2 路网拓扑结构

假设 n 是路网 R_n 中的一个节点.出度 $o(n) = \#\{\langle n,z \rangle | \langle n,z \rangle \in S\}$ 表示以 n 为起始节点的路段的个数;入度 $i(n) = \#\{\langle m,n \rangle | \langle m,n \rangle \in S\}$ 表示以 n 为终止节点的路段的个数,如图 1 中节点 1 的入度为 1,出度为 2.

路网 R_n 中的节点 n ,当 $o(n) \neq 1$ 或 $i(n) \neq 1$ 时,

式(1)中 R_n 代表道路网络; N 代表节点集,表示路网中构成道路的坐标点集合,其元素为经纬度值对 (x,y) ; S 代表路网的路段集合,其元素是有序对 $\langle m,n \rangle$,表示存在一条有向直线道路 s ,其起始节点 $m=begin(s)$,终止节点 $n=end(s)$,且 s 上不存在其它节点,如图 1 中,节点 1 和节点 2 之间存在着路段 1.

n 称为连通性节点.在路网 R_n 中,存在一条通路 $L\langle s_1 \dots s_n \rangle$,若 S_1 的起始节点和 S_n 的终止节点均为连通性节点,且其它路段的节点的入度和出度均为 1,则此路段的序列 L 称为一条路链,路链的集合为

$$\begin{aligned} L(R_n) = \{l | l = \langle s_0 \dots s_k \rangle, i(begin(s_0)) \neq 1, \\ o(end(s_k)) \neq 1 \text{ 且 } i(begin(s_i)) = 1, \\ o(end(s_i)) = 1, 0 < i < k\} \end{aligned} \quad (2)$$

在路网模型的基础上,根据道路网络的连通性和路链结构,建立路链间的拓扑结构图 $G(R_n) = (V,E)$,其中 $V=L(R_n)$,表示在路网 R_n 上的路链集合, E 则是根据道路的连通性建立的路链之间的连通关系.

按照路链之间的连接性,路链的驶入路链称

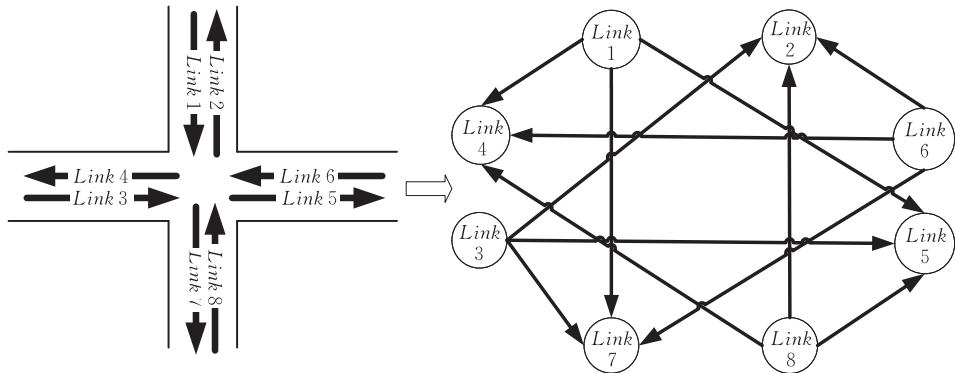


图 2 十字交叉路口拓扑结构示意图

为此路链的前继路链,路链的驶出路链称为此路链的后继路链, $link \rightarrow link'$ 表示 $link$ 是 $link'$ 的前继路链,而 $link'$ 是 $link$ 的后继路链^[7]. 如图 2 中所示, $Link1 \rightarrow Link5$, $Link6 \rightarrow Link4$.

3.3 向 量

在地图上取一坐标点 $c_i(x_i, y_i)$ (x_i 为 c_i 点的经度, y_i 为 c_i 点的纬度), 从 c_i 出发向某点 $c_j(x_j, y_j)$ (x_j 为 c_j 点的经度, y_j 为 c_j 点的纬度) 引一条线段, 有方向且有长度, 这种有向线段称为向量, 记作 $c_1 c_2$, 它的长度记作 $\phi(c_1 c_2)$, 表示 c_i 与 c_j 之间的距离, 它的方向记作 $\phi(c_1 c_2)$, 表示为 $c_1 c_2$ 与正北方向的夹角 (顺时针方向, 方向值取值 $0 \sim 360$).

对于 $L(R_n)$ 中的任一路链 l , 其向量长度记为 $\phi(l)$ 表示路链起点与路链终点所构成有向线段的距离, 其向量方向记作 $\phi(l)$, 表示路链起点与路链终点所构成有向线段的方向.

4 问题描述和问题分析

4.1 问题描述

在路网 R_n 中, 假设由其形成的路网拓扑为 G , 运动车辆在一个时间段 T 内的连续 GPS 轨迹点集合 P 表示为 $\{p_0, \dots, p_n\}$, $p_i(z_i, v_i)$ (z_i 为车辆定位坐标点, v_i 为车行方向), $0 \leq i \leq n$.

假设车辆在 T 内所行驶过的通路为 κ , 如何将轨迹点集合 P 映射为 κ , 就是路径推测所要解决的最主要问题, 即

$$Mm(P, R_n) \rightarrow \kappa \tag{3}$$

通过构建匹配函数 $Mm(P, R_n)$, 完成 P 与 R_n 的推测计算, 得到车辆的正确行驶路径.

4.2 车辆行驶规律分析

4.2.1 车辆行驶状态分析

车辆在短时间内的行驶轨迹具有一定的相似

性^[8], 通过对车辆的行驶状态分析归纳, 可以掌握车辆的行驶规律, 再与路网模型相结合, 进一步得到了路径搜索的启发式条件. 下面通过实例的方式对车辆在路网上的行驶状态进行分析, 并归纳路径推测算法的启发式搜索条件.

(1) 车辆在单一路链行驶.

车辆的连续轨迹点位于同一路链上. 如图 3 中, p_1 在 $link1$ 上的匹配点 m_1 与 p_2 构成的轨迹向量其长度小于 m_1 与 $link1$ 的终点构成的向量长度, 即 $\phi(m_1 p_2) < \phi(m_1 e_1)$ ($e_1 = end(link1)$) 且 $|\phi(m_1 p_2) - \phi(m_1 e_1)| < 90^\circ$.

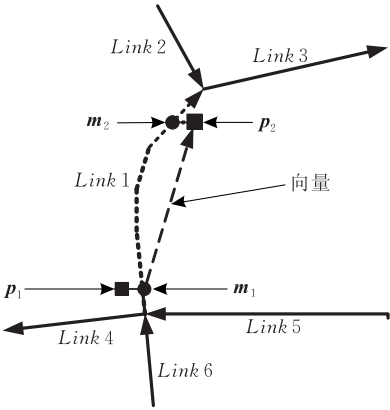


图 3 车辆在单一路链上行驶

(2) 车辆跨路链行驶

车辆的连续轨迹点跨越了两条或两条以上路链, 车辆基本存在了两种行车状态: 一种是保持直行, 如图 4 所示; 另一种是转弯行驶, 如图 5(a) 和图 5(b) 分别所示.

在示意车辆跨路链直行的图 4 中 $\phi(m_1 p_2) < \phi(m_1 e_1) + \phi(Link2) + \phi(s_3 p_2) < 2 \times \phi(m_1 p_2)$ ($e_1 = end(link1)$, $s_3 = begin(link3)$) 且 $|\phi(m_1 p_2) - \phi(m_1 e_1)| < 90^\circ$, $|\phi(m_1 p_2) - \phi(Link2)| < 90^\circ$, $|\phi(m_1 p_2) - \phi(Link3)| < 90^\circ$.

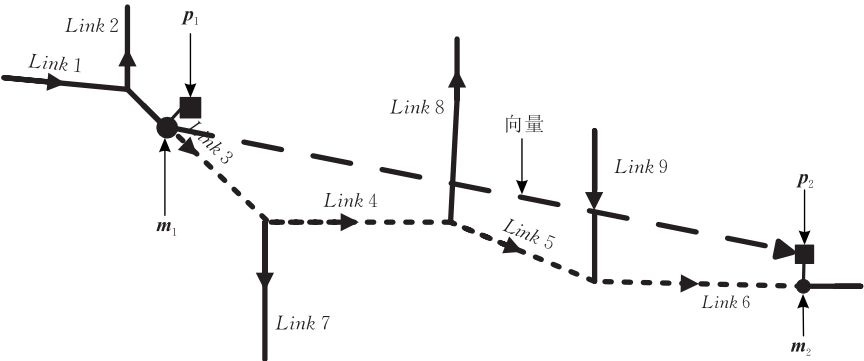


图 4 车辆跨路链直行

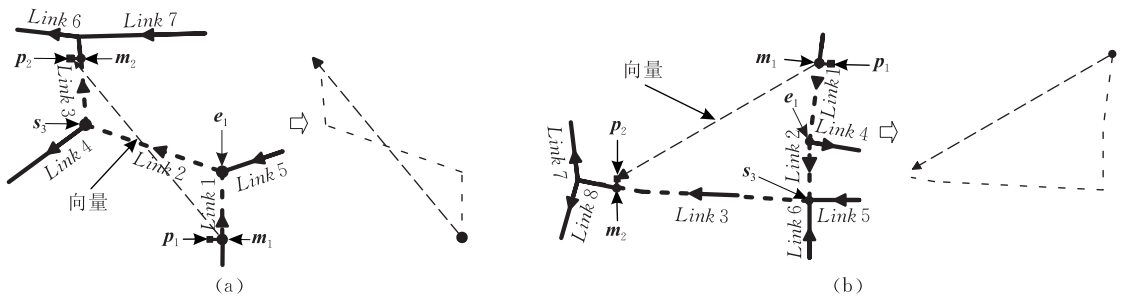


图 5 车辆跨路链转变行驶

转弯的情况比较复杂,如图 5 所示,轨迹向量和车辆所行驶过的轨迹路链可以形成类三角形。

在普通的路网中,交叉道路的夹角幅度一般会在 $60^\circ \sim 180^\circ$ 之间. 假设三角形的三边分别为 a 、 b 和 c , 其中 a 、 b 两边的夹角 C 大于 60° , 根据三角形余弦定理和极限定理有下式:

$$a+b > 2c \Rightarrow \cos C > \frac{3(a^2+b^2)}{2ab} - \frac{1}{4} \Rightarrow \cos C > \frac{1}{2} \\ \Rightarrow C < 60^\circ,$$

则有

$$C \geq 60^\circ \Rightarrow a+b \leq 2c \quad (4)$$

即当三角形中两边的夹角大于 60° 时, 此两边之和小于第三边的 2 倍。

由此可得, 在图 5 中, 车辆的轨迹点 p_1 的匹配点 m_1 到 link1 的终点距离 $\phi(m_1 e_1)$, link2 的向量长度 $\phi(link2)$ 与 link3 的起点 s_3 到轨迹点 p_2 的距离 $\phi(s_3 p_2)$ 之和小于 $2 \times \phi(m_1 p_2)$, 即

$$2 \times \phi(m_1 p_2) > \phi(m_1 e_1) + \phi(link2) + \phi(s_3 p_2).$$

另外, 在方向上, 车辆行驶路链的方向与向量的方向值必定会小于 90° . 因此, 车辆行驶时满足的约束条件可以归纳如下:

(1) 路链的方向与匹配点和车辆轨迹点形成向量的方向值之差小于 90° , 即

$$|\phi(m_i p_{i+1}) - \phi(l_j)| < 90^\circ \quad (5)$$

(2) 车辆行驶经过路链的累计向量长度之和小于匹配点与车辆轨迹点形成向量的长度的 2 倍, 即

$$2 \times \phi(m_i p_{i+1}) > \phi(m_i e_j) + \phi(l_{j+1}) + \dots + \phi(l_k) + \phi(s_k p_{i+1}) \quad (6)$$

5 算法设计

5.1 路径搜索树

为了提高搜索效率, 采用启发式搜索思想, 其中路径搜索树定义为 $Ps-T$, $Ps-T = (U, A)$, 其中 U

是路网拓扑 $G(R_n) = (V, E)$ 中 V 的非空子集, A 为 U 集合中边的关系。

从路径搜索树中得到算法的输出, 是车辆记录的 GPS 定位数据的时间范围内所行驶的路径, 表现为若干条相连的路链。

5.2 车辆行驶路径搜索

5.2.1 候选匹配点

为了提高地图匹配的准确性, 本文提取了点匹配过程中两个特征: 定义 dis 为 GPS 轨迹点到路段的距离, $dirDiff$ 为行车方向与路链的角度差, 帮助判断定位点 M 的所在. 两个特征对应的后验概率为 $p(match|dis)$ 和 $p(match|dirDiff)$, 然后用 SVM 分类器进行分类. SVM 的分类结果表示为

$$\begin{cases} match = \{true | \partial_1 \times p(match|dis) + \\ \partial_2 \times p(match|dirDiff) \geq threshold\} \\ match = \{false | \partial_1 \times p(match|dis) + \\ \partial_2 \times p(match|dirDiff) > threshold\} \\ \partial_1 + \partial_2 = 1 \end{cases} \quad (7)$$

利用 GPS 匹配点的两个特征值定义匹配概率 η 如下:

$$\eta(\psi, \phi) = \partial_1 \times \psi + \partial_2 \times \phi \quad (8)$$

对于给定的任一 GPS 轨迹点 p , 它在路链 l 上的候选匹配点被定义为匹配概率大于匹配阈值 λ 的点, 记作 $\tau(p, l) = z$ 且 $\eta > \lambda$, 其中 z 为 p 在 l 上的匹配点。

每个 GPS 轨迹点均可以通过点匹配 $Pmm(p, R_n)$ 得到其在路网上 R_n 的匹配路链集合, 表述如下:

$$Pmm(p, R_n) = \{l | \tau(p, l) = z \text{ 且 } \eta(\psi, \phi) > \lambda\} \quad (9)$$

5.2.2 $Ps-T$ 的生成

路径搜索树将根据匹配路链集合的不同, 进行不同的初始化: (1) 如果候选匹配路链集合的元素只有一个, 则以匹配路链建立根节点, 即 $Ps-T = (Pmm(p_0, R_n), \{\})$; (2) 如果候选匹配路链集合的

元素大于 1,则建立一个空节点作为根节点,以每条候选匹配点路链分别建立根节点的子节点,即

$$P_{S-T}=(l_{\text{null}}\cup Pmm(p_0,R_n),\{a|a=(l_{\text{null}}\rightarrow l),l\in Pmm(p_0,R_n)\}).$$

依据 GPS 定位的车辆轨迹,车辆在一段时间内的所有可能行驶路径都保存在路径搜索树中. 利用路网拓扑和车辆的定位数据,按照车辆行驶时满足的约束条件(5)和(6),对车辆的行驶路径进行搜索时,这棵树的深度和广度就会不断地增加,树节点里保存着车辆可能行驶过的路链.

从路径搜索树中得到算法的输出,是车辆记录的 GPS 定位数据的时间范围内所行驶的路径,表现为若干条相连的路链.

5.3 车辆行驶路径选择

生成最终的 P_{S-T} 后,如果有多个叶节点中保存着轨迹点集合 P 的最后一个点的匹配路链,则表明车辆可能存在多条行驶路径,每个包含匹配路链的叶节点向上回溯到根节点对应一条车辆的候选行驶路径. 累计每条路径上的候选匹配点的匹配概率,选择累计值最小的一条路径作为最终的路径推测结果.

5.4 算法时间复杂度分析

基于路网拓扑 G 进行浮动车行驶路径的搜索通常是一个 NP 问题,每 2 个轨迹点之间进行图遍历搜索.

在本文介绍的方法中,设车辆在一个时间段 T 内共采集到 n 个 GPS 轨迹点,2 个轨迹点之间进行一轮路径搜索. 假设前一匹配点和后续的轨迹点所建立的向量平均长度为 d ,路链的平均长度为 l ,路链的后续路链数平均为 3,通过约束条件(4)的限制,加入 P_{S-T} 的路链数平均为 2. 两点之间 P_{S-T} 的深度可达到 $2\times d/l$,起始两点完成搜索后,路径搜索树的叶节点为 $2^{\lfloor 2\times d/l \rfloor - 1}$,假设叶节点中有 1/2 保存了匹配到的路链,因此,算法的时间复杂度为 $O((2^{\lfloor 2\times d/l \rfloor - 2})^n)$,为指数时间算法. 实际上,经过约束条件(5)和(6)的限制,其搜索空间范围非常有限,远远小于 $(2^{\lfloor 2\times d/l \rfloor - 2})^n$.

6 实验分析

为了验证本文提出的算法,作者使用了北京市的导航电子地图(包括 86000 条路链)和 13000 辆出租车发回的以 60~120s 为间隔的 GPS 数据作为浮动车数据进行了实验.

6.1 提高地图匹配精度

在利用样本数据对匹配距离和角度差两个匹配特征值使用 SVM 分类器进行分类后,得到匹配距离 dis 和角度差 $dirDiff$ 的后验概率线性拟合情况,如图 6 所示.

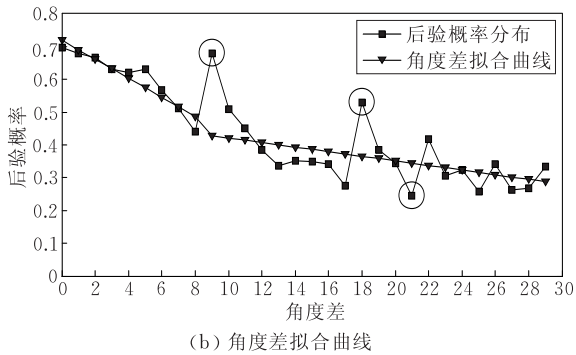
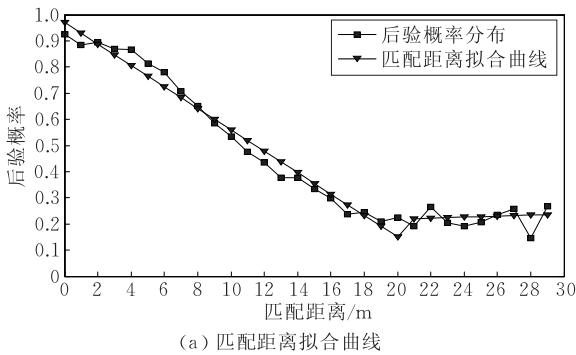


图 6 匹配特征参数的后验概率线性拟合

SVM 分类得 $[\partial_1=0.64,\partial_1=0.36]$, 阈值 $\lambda=0.592$.

6.2 准确性

为了验证算法推测车辆行驶路线的准确性,在实验中作者抽取了 4 辆作为浮动车的出租车,在正常运营时间范围内选取了北京市的主要道路作为规定的行车路线,分 4 组进行了 11 次路侧. 通过算法

计算出的车辆行驶旅行与实际线路对比的正确率作为评测的标准.

在推测结果的效果图 7 上,粗体路线代表推测正确的道路,虚的路线代表车辆正确行驶而算法没有进行正确推测的道路. 通过图 8 对测试结果的统计,可发现算法的准确率达到 93% 以上.

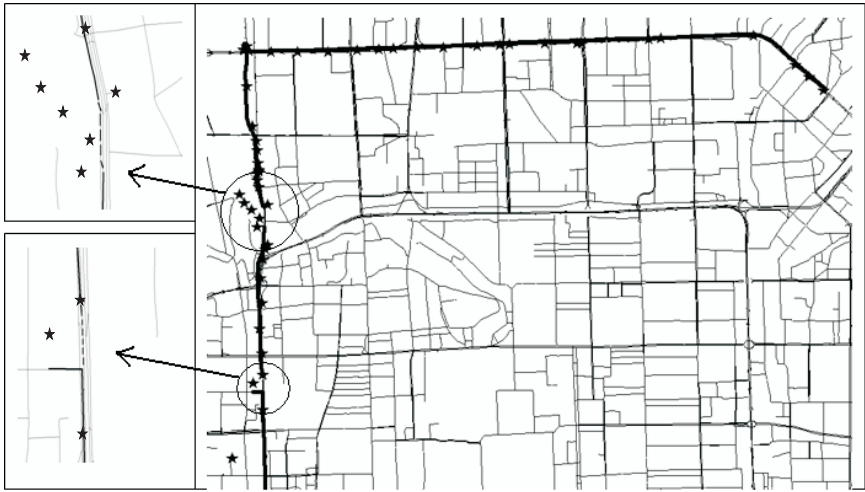


图 7 测试路线的路径推测结果

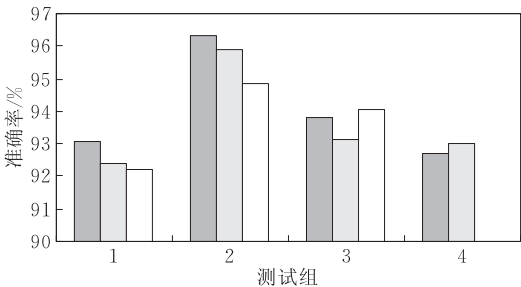


图 8 不同测试路线的匹配准确率统计

6.3 计算速度

由于浮动车系统是对最近一段时间范围内(一般采用 5~15min)采集的全部数据进行批量处理,因此本文选取早 8:00 至晚 8:00 的不同时间段,采用缓存的 5~15min 时间范围内的浮动车数据对算法的计算速度进行测试,并对不同处理周期下的平均计算时间做了统计,如图 9 所示. 测试环境为 Pentium(R) 4 CPU 3.20GHz, 2GB 内存.

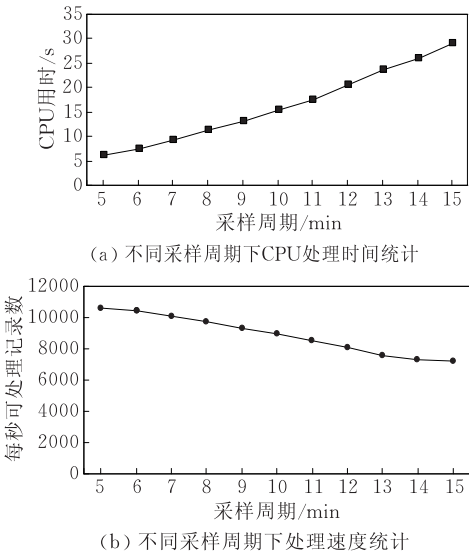


图 9 算法的计算效率测试统计

对算法的运行效率上的测试可以看出算法具有很高的运行效率,但是随着采样周期的延长,计算速度有所下降.从系统发布路况的要求来看,处理的时间应该控制在 30~40s 以内,因此,算法可以充分满足浮动车系统的实时性要求.

6.4 对比测试

使用同样的测试数据,本文对渐进式和全局式的匹配算法进行了测试,并在准确性和计算效率上与本文介绍的启发式算法进行了对比.从表 1 对不同算法的比较可以看出,启发式算法在准确性和计算效率比渐进式和全局式算法具有明显优势,满足了实际 ITS 系统信息的需求,快速准确地提供了道路交通实时信息.

表 1 算法的对比测试结果

	准确性/%	计算效率/GPS 记录数 · s ⁻¹
渐进式	86	4547
全局式	93	3235
启发式	93.75	8885

7 结 论

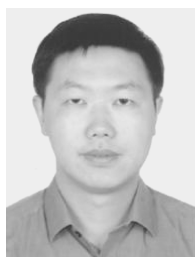
综上所述,本文提供了一种适用于实时处理大规模浮动车 GPS 数据的路径推算算法.它具有启发式搜索的特征,算法的效率不受路网的复杂性影响,将路径搜索的效率进行了有效的提高.通过实际数据的测试可以看出,其特点是实时性好,效率高,并且具有较好的准确性.下一步,我们将在候选路径的选择方法上进行深入的研究和实验,进一步提高算法的准确性.

参 考 文 献

- [1] Dai Xiao-Wen, Ferman M A. A simulation evaluation of a real-time traffic information system using probe vehicles// Proceedings of the IEEE ITSC. Shanghai, China, 2003: 12-15
- [2] Kuehne R, Schaefer R-P, Mikat J, Thiessenhusen K-U, Boettger U, Lorkowski S. New approaches for traffic management in metropolitan areas//Proceedings of the IFAC CTS Symposium. Tokyo, Japan, 2003: 42-45
- [3] Brakatsoulas S, Pfoer D, Tryfona N. Practical data management techniques for vehicle tracking data//Proceedings of the 21st International Conference on Data Engineering, 2005: 324-325
- [4] Marchal F, Hackney J K, Axhausen K W. Efficient map-matching of large GPS data sets—Test on a speed monitoring experiment in Zurich. Transportation Research Record,

2004, Monte Verità, 2004: 23-25

- [5] Brakatsoulas S, Pfoer D, Salas R, Wenk C. On map-matching vehicle tracking data//Proceedings of the International Conference on Very Large Data Bases. Trondheim, Norway, 2005: 853-864
- [6] Lu W F, Zhu T Y, Wu D D. A heuristic path-estimating algorithm for large-scale real-time traffic information calculating. Science in China Series E: Technological Sciences, 2008, 51(Supplement I): 165-174
- [7] Zhang Shan-Qi, Wu Jin-Pei, Yang Dong-Kai. Intelligent Vehicle Positioning Navigation System and Application. Beijing: Science Press, 2002(in Chinese)
(张其善, 吴今培, 杨东凯. 智能车辆定位导航系统及应用. 北京: 科学出版社, 2002)
- [8] Wu Dong-Dong, Zhu Tong-Yu, Lu Wei-Feng. A heuristic map-matching algorithm by using vector-based recognition// Proceedings of the IEEE ICCGI. Guadeloupe, French Caribbean, 2007: 18-18



LU Wei-Feng, born in 1972, Ph. D., professor. His research interests focus on ITS, network management.

WU Dong-Dong, born in 1982, master, engineer. His research interests focus on ITS.

ZHU Tong-Yu, born in 1969, Ph. D., professor. His research interests focus on ITS.

Background

How to acquire real-time and dynamic transportation information is becoming prevailing research focus in the community of Intelligent Transport Systems (ITS) nowadays. This information can be applied in the transportation area like vehicle navigation, traffic control and so on. Floating car data (FCD) refers to assess the overall traffic conditions. Using technologies like orientation, wireless communication and information processing to achieve the collection and disposal of the raw data like instantaneous speed, position and travel time of the floating cars. A pre-processing step that matches FCD to the road network is needed. This technique is commonly referred to as path-estimating. Large-scale, real-time and discrete distribution is the main characteristics of FCD, and the core processing algorithms must provide good performance of high efficiency and accuracy in order to meet the needs of real traffic information system. The traditional path-estimating algorithm adopts two mechanisms: incremental and global. The incremental method is simple and fast, while

the accuracy of the results is poor; the global method is complicated and the accuracy of the result is pretty good, while the processing speed is slow. For this reason, the paper studies and designs a heuristic route-estimating algorithm by using vector-based recognition. The proposed algorithm is efficient both at accuracy and computational speed when handling the large-scale data of GPS tracking data even under the complex road network conditions. This subject belongs to the Next Generation Internet Demonstrated Project—"Beijing Internet ITS Demonstration and Application Project" and the National High-Tech Research and Development Plan of China "The research to key technologies of high-quality dynamic traffic information services for vehicles' navigation". With the result of the projects, it provides real-time picture of traffic conditions of the whole road network of Beijing through handling large volume of GPS data recorded by 13000 taxis. The results of this paper provide the core computations which make the "raw" FCD to be related which the city road.