

路网匹配算法综述^{*}

高文超^{1,2}, 李国良², 塔娜³

¹(中国矿业大学(北京) 机电与信息工程学院, 北京 100083)

²(清华大学 计算机科学与技术系, 北京 100084)

³(中国人民大学 新闻学院, 北京 100872)

通讯作者: 李国良, E-mail: liguoliang@tsinghua.edu.cn



摘要: 路网匹配是基于位置服务中的关键预处理步骤, 它将 GPS 轨迹点匹配到实际路网上, 以此为基础对数据进行分析 and 挖掘, 能够辅助解决城市计算中相关问题, 例如建立智能交通系统、协助用户规划出行. 对国内外学者在该研究领域取得的成果进行了分类总结, 发现这些匹配算法可以较好地解决高采样率的路网匹配问题. 但是, 随着城市交通的快速发展, 获取和处理车辆位置信息的成本不断提高, 低频采样点越来越多, 现有算法匹配精确度大幅度下降. 于是, 近年来出现了基于隐马尔可夫模型(hidden Markov model, 简称 HMM)的路网匹配算法. 隐马尔可夫模型可以较为平滑地将噪声数据和路径约束进行整合, 从有许多可能状态的路径中选择一条最大似然路径. 重点总结了基于隐马尔可夫模型的路网匹配算法, 主要是从特点与实验结果的角度对其进行对比总结, 有些实验结果的正确率在一定条件下最高可达 90%, 这说明了基于隐马尔可夫模型的路网匹配算法在低采样率下的有效性. 最后, 对未来的研究可能采取的方法进行了展望.

关键词: 路网匹配; 隐马尔可夫模型; GPS 轨迹; 基于位置服务; 低采样率

中图法分类号: TP301

中文引用格式: 高文超, 李国良, 塔娜. 路网匹配算法综述. 软件学报, 2018, 29(2): 225–250. <http://www.jos.org.cn/1000-9825/5424.htm>

英文引用格式: Gao WC, Li GL, Ta N. Survey of map matching algorithms. Ruan Jian Xue Bao/Journal of Software, 2018, 29(2): 225–250 (in Chinese). <http://www.jos.org.cn/1000-9825/5424.htm>

Survey of Map Matching Algorithms

GAO Wen-Chao^{1,2}, LI Guo-Liang², TA Na³

¹(School of Mechanical Electronic and Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China)

²(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

³(School of Journalism and Communication, Renmin University of China, Beijing 100872, China)

Abstract: Map matching is a key preprocessing step in the location-based service to match GPS points into a digital road network. Data analysis on the map matched trajectory data can be used to facilitate many real city computing applications such as intelligent transportation system and trip planning. This survey provides a systematic summary of existing research achievements of map matching.

* 基金项目: 国家自然科学基金(61373024, 61632016, 61422205); 国家重点基础研究发展计划(973)(2015CB358700); 教育部人文社科基地重大项目(16JJD860008)

Foundation item: National Natural Science Foundation of China (61373024, 61632016, 61422205); National Program on Key Basic Research Project of China (973) (2015CB358700); Key Grant Project on Humanities and Social Sciences of the Ministry of Education of China (16JJD860008)

收稿时间: 2017-05-10; 修改时间: 2017-08-17; 采用时间: 2017-10-02; jos 在线出版时间: 2017-10-17

CNKI 网络优先出版: 2017-10-17 13:42:46, <http://kns.cnki.net/kcms/detail/11.2560.TP.20171017.1342.012.html>

With the rapid development of urban traffic, the cost of acquiring and processing vehicle location information is increasing, low-sampling-rate GPS tracking data is growing, and the accuracy of existing algorithms is not adequate. In recent years, map matching algorithm based on hidden Markov model (HMM) has been widely studies. HMM can smoothly assimilate noisy data with path constraints by choosing a maximum likelihood path. The accuracy of HMM-based algorithms can reach 90% under certain conditions, which confirms the validity of map matching algorithm based on HMM at low sampling rate. A perspective of future work in this research area is also discussed.

Key words: map matching; hidden Markov model; GPS trajectory; location-based service; low sampling rate

近几年来,中国在线出行服务行业发展迅速,服务模式从“线下重资产+线上服务”转为“互联网+共享经济/轻资产重服务”,服务场景从电脑端转向移动端.自 2012 年起,移动端出行服务模式大量涌现,包括出租车、拼车、专车(快车)、租车、代驾、定制巴士等.截止到 2015 年年底,移动端出行服务用户数量接近 4 亿,其中,出租车用户达到 2.5 亿,拼车用户达到 1.6 亿,专车(快车)用户达到 1.9 亿.2016 年,滴滴出行全平台订单数量达到 14.3 亿,这相当于美国 2015 全年订单量的近 2 倍(<http://www.aiweibang.com/yuedu/100330112.html>).

这些出行车辆装配的 GPS(global positioning system,全球定位系统)设备每天能够收集到大量的移动位置序列和车载状态信息,这些数据蕴含着丰富的交通信息和用户行为,可以实时地了解路网上乘客需求的时空变化以及乘客出行的有关信息.通过使用大数据处理技术,分析车辆轨迹数据和乘客需求数据,挖掘出对乘客打车和司机载客有帮助的价值信息,能够对乘客的出行提供有效的指导.同时,从一定程度上反映出城市交通状况和人群流动情况,便于了解交通流量规律,发现人群行为特征,协助改善交通状况等,具有极大的科研和应用价值.

另外,随着基于位置服务(location-based service)和移动社交网络(mobile social network)的发展和普及,越来越多的应用,如热路线查找^[1]、交通监管^[2]、城市规划^[3]、地理社会网络^[4],已经开始使用来自 GPS 的数据信息,以实现更好的服务质量.但是,由于 GPS 传感器的读数有定位误差和采样误差^[5],GPS 跟踪数据与实际轨迹之间的偏差是不可避免的.如果不进行路网匹配,那么交通工具可能不会正确显示到道路上,如图 1 所示.

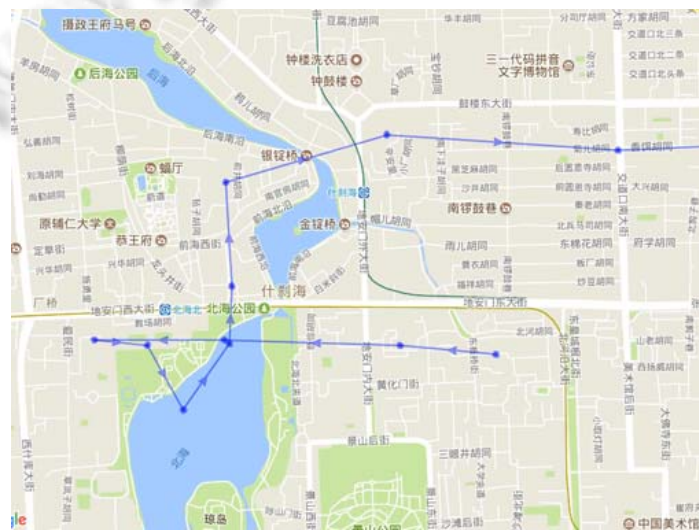


Fig.1 GPS device positioning illustration

图 1 GPS 设备定位示例

在图 1 中,实心点是 GPS 设备定位到的位置信息,可以看出,GPS 点离最近的路有一段距离,如果不做处理,这个点就直接定位到建筑、湖泊中,而不显示到路网上.最终得到的车辆行驶路线将不沿着正常道路,而会直接穿越地铁、房屋、草坪等,显然,这是不合理的.所以需要由 GPS 设备获得的定位信息进行一定的处理,使它能够为较为精确地匹配到路网上.

将所观察到的用户或者交通工具的 GPS 位置关联到给定电子地图的道路网络上的过程称为路网匹配。路网匹配对于许多基于位置的应用程序是一个基本的数据预处理步骤(如图 2 所示),只有处理好这一步,以后的研究才能有效地进行。其主要目的是识别用户(或车辆)正在/曾经行驶的真实路段,进行车辆导航^[6,7]、位置预测^[8,9]、路线或行程预测^[10,11]、交通状况预测与管理^[12-14]、轨迹深度理解^[15,16]、路径规划与推荐^[17-19]、活动识别^[20]和许多其他基于位置的服务^[21-23],同时,可以加入其他数据(如气象数据、环境数据等)来进行城市计算^[24,25],向社会提供广泛的服务。由于大多数的民用 GPS 设备和智能手机中的 GPS 模块价格便宜但精度不高,这就需要一种可靠且健壮的基于位置信息的路网匹配算法来提高 GPS 数据的准确性。

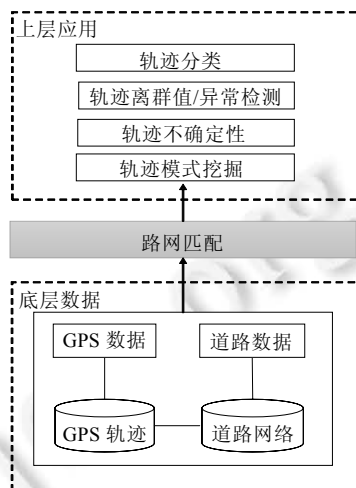


Fig.2 Paradigm of trajectory data mining

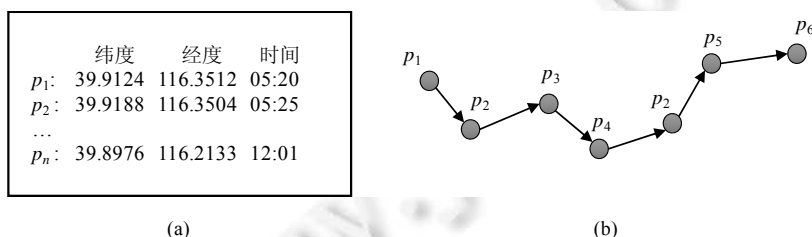
图 2 轨迹数据挖掘示意图

1 轨迹路网匹配问题描述

路网匹配是一个将原始的经纬度坐标转换成一个序列的处理过程,其涉及到的概念和定义如下。

定义 1(GPS 日志). 一个 GPS 日志是一系列 GPS 点的集合 $L=\{p_1, p_2, \dots, p_n\}$. 每个 GPS 点包含纬度、经度和时间戳信息,如图 3(a)所示。

定义 2(GPS 轨迹). 一条 GPS 轨迹 T 是一个 GPS 点的序列,任意连续的 GPS 点之间不超过一定的阈值 ΔT , $T: p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_n$, 其中, $p_i \in L, 0 < p_{i+1}.t - p_i.t < \Delta T (1 \leq i < n)$, ΔT 是采样间隔,如图 3(b)所示。



(a)

(b)

Fig.3 Illustration of GPS log and GPS trajectory

图 3 GPS 日志和轨迹的示意图

定义 3(路段). 一个路段 e 是一条有向边,具有编号 $e.id$ 、行驶速度 $e.v$ 、长度 $e.l$ 、开始点 $e.start$ 、结束点 $e.end$ 和中间点列表。一条道路可以包含多个路段,如图 4 所示。

定义 4(路网). 路网是一个有向图 $G(V, E)$, 其中, V 是顶点集合,代表路口或道路中的转折点; E 是一系列有向

边,代表路网中的路段.

定义 5(路径). 给定路网 G 中的两个顶点 v_i, v_j , 一条路径 P 是始于 v_i 止于 v_j 的一组连接的路段. $P: e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n$, 其中, $e_1.start = v_i, e_n.end = v_j, e_k.end = e_{k+1}.start, 1 \leq k < n$.



Fig.4 Illustration of a path in a road network

图 4 路网中的一条路径示意图

路网匹配问题可以描述为,给定未加工的 GPS 轨迹 T 和路网 $G(V,E)$,从 G 中寻找路径 P ,其实际路径匹配轨迹 T .路网匹配的一般流程如图 5 所示.

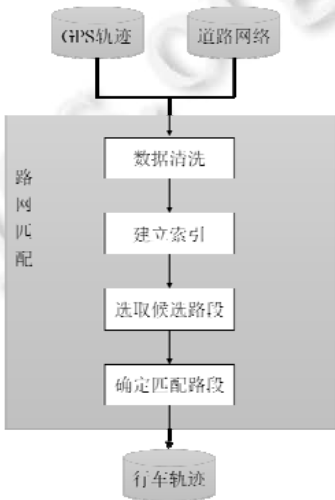


Fig.5 Paradigm of map matching

图 5 路网匹配流程图

直观来讲,解决路网匹配最简单的方法就是将 GPS 点关联到最近的路段上,但有些时候效果会非常差,如图 6 所示:假如有 3 个连续的点,第 1 个点和第 3 个点都在一条路上,而第 2 个点由于误差,会被定位到另外一条无关的路上,但却是最接近的.尤其是在采样频率较低、定位误差较大、信号易丢失的实际情况下,如何保持较高的路网匹配准确率,仍是当前研究的重点问题.目前,国内外学者提出了很多路网匹配算法来处理这个问题.



Fig.6 GPS point matched to the nearest road segment

图 6 GPS 点关联到最近的路段

2 轨迹路网匹配算法分类

根据算法的不同特性,现有的路网匹配算法可以有如下分类标准.

2.1 以数据涉及的信息来划分

根据输入数据所涉及的信息,现有的路网匹配算法可被分成 4 类,见表 1.

Table 1 Algorithm classification by data involved in the information
表 1 以数据涉及信息进行算法分类

算法	代表性工作	算法描述	优缺点分析
几何匹配算法	点到点匹配方法 ^[26]	首先计算 GPS 点与道路网络上每个节点的距离,然后将 GPS 点匹配到距离最近的节点上	算法简单,容易实现,计算速度快,但受道路网络上节点密度及分布影响,容易造成错误匹配,实用性不强
	点到线匹配方法 ^[27]	将道路网络中的所有路段都看作候选路段,首先计算 GPS 点到每个候选路段的投影距离,然后选择距 GPS 点最近的路段为匹配路段,相应的投影点就是匹配节点	准确性提高,但由于没有考虑道路的拓扑结构和车辆的历史信息,所以稳定性不好,在复杂路况下容易发生错误匹配
	线到线匹配方法 ^[28]	首先获得车辆的历史轨迹曲线,然后与候选路段进行相似性比较,最终选择相似度最高的路段	考虑车辆的历史信息,提高了匹配效率,但计算量大,受异常值影响较大,并且某个 GPS 点的匹配错误可能会引起一系列的错误匹配
拓扑关系算法	简单拓扑关系匹配方法 ^[29]	运用车辆历史数据、轨迹之间关系和道路拓扑特征等信息来限制采样点的候选匹配	利用历史采集点和空间道路网的数据,提高匹配效率,但容易受到采集噪声和数据稀疏性的影响,在复杂路况下容易发生错误匹配
	加权拓扑匹配方法 ^[30]	首先对轨迹方向、GPS 点到道路的距离以及相关性等进行加权计算得到该路段的总权重,然后选择权重最大的路段为匹配路段	根据 GPS 信息和道路几何形状,使用较少的输入,引入额外的参数和准则来提高匹配效率,但易受采集噪声和数据稀疏性的影响,在复杂路况下容易发生错误匹配
概率统计算法	置信区间匹配方法 ^[31,32]	基于多假设思想,首先在置信区域内选出多条路段加入到候选路段集中,然后对每条候选路段都计算出一个得分,最终选择得分最高的路段作为匹配路段	即使某个 GPS 点匹配错误也不会对后续匹配产生太大影响,但同时也降低了路网匹配程序的执行速度
	新型概率匹配方法 ^[33,34]	在道路交叉口处设置一个椭圆或矩形的置信区域,通过一些基于经验性研究准则来检测车辆在道路交叉口的转弯动作	计算速度较快,稳定性较好,对识别车辆转弯很有效,即使车辆低速运行时也能正确匹配
先进匹配算法	模糊逻辑方法 ^[23,35-38]	采用模糊逻辑的方法解决路网匹配问题,通过隶属函数描述以及候选道路来定义误差模型	实时性好,鲁棒性较好,匹配效率较高,但计算开销较大,缺乏理论依据,对输入要素的利用率较低,实用性较差
	证据推理方法 ^[39-42]	基于 D-S 证据理论,首先利用车辆位置与方向信息为证据分别构造基本概率分配函数,再将其融合为一个联合支持度函数,然后比较融合后的各候选路段支持度大小,最终选定支持度最高的路段作为匹配路段	通过证据推理可以得到唯一正确的判断结果,但是只考虑 GPS 点的两种信息:位置和方向,准确度不高,而且算法实现复杂、计算开销大
	卡尔曼滤波方法 ^[43-46]	考虑 GPS 系统误差特点,基于卡尔曼滤波中的白噪声假设,通过分析卡尔曼滤波后的模型误差特性是否满足高斯白噪声分布来完成地图匹配	稳定性较高,特别是在交叉路口,但是由于没有充分考虑车辆信息和路网信息,匹配最初阶段效果并不理想
	贝叶斯推理 ^[47,48]	使用多假设技术(multiple hypothesis technique,简称 MHT),为置信区域内所有路段产生伪测量值,路网拓扑分析(连接、方向和道路设计参数)与伪测量值一起被用来为每个车辆推导出一组假设及概率值	缺少初始匹配方法,在很大程度上影响了随后匹配的性能,因此存在错匹配可能性

2.1.1 基于几何信息的匹配算法

基于几何信息的匹配算法是由 Bemstein 等人在 20 世纪 90 年代提出来的,它仅利用时空道路网络的几何信息进行匹配,如距离、角度、形状等,而不考虑路段间的连通性^[26].

基于几何信息的匹配算法相对简单,易于实现,在定位系统没有较大误差的情况下匹配度较高,但是它们没有使用 GPS 采集点之间的联系、交通规则信息和道路的拓扑信息等,导致准确率下降;而且一旦出现错误匹配无法及时修正,稳定性较差.虽然它们可以为密集的 GPS 轨迹提供良好的匹配,但如果是针对有噪声且稀疏的数据就不再适用.

2.1.2 基于拓扑信息的匹配算法

拓扑信息是指实体(点、线、多边形)之间的相连(线)、邻接(多边形)、包围(多边形与点)等关系.基于拓扑信息的匹配算法^[49,50]关注路网的连通性,运用历史数据、车辆速度和道路拓扑特征等额外信息来限制采样点的候选匹配.代表性的算法^[51]使用弗雷歇距离(Fréchet distance)来衡量 GPS 序列和候选路段序列的匹配度,处理具有噪声和低采样率的轨迹.算法^[30]使用道路网络的几何形状和 GPS 导航数据之间的各种相似性标准(如车辆速度、封闭性)将用户轨迹映射到道路网络上.

基于拓扑信息的路网匹配算法不仅要考虑 GPS 采集点与道路的距离,而且要考虑路网的拓扑连通性以及历史采集点,从而使匹配效率和准确性在一定程度上有所提高.但它们仍然容易受到采集噪声和数据稀疏性的影响,还不能完全解决复杂的城市道路问题.

2.1.3 基于概率信息的匹配算法

Honey 等人在 1989 年首先将概率统计方法运用在路网匹配算法中^[52].该算法在汽车动态运动过程中,利用计算机来估计车辆的行驶方向以及位置距离,然后使用航位推测法来推测汽车所在的路段.基于概率信息的匹配算法^[23,33,53,54]给每个 GPS 样本点设定一个椭圆或矩形的置信区域,并根据 GPS 的点和在置信区域内的位置之间的距离获得概率值,根据概率值的大小来决定最佳匹配路径.

基于概率信息算法不要求车辆总是在道路上.如果车辆接收到的 GPS 数据偏离了已知的路网,算法将会不断地比较接收到的 GPS 坐标与偏离路段的坐标,并识别与车辆匹配的路段^[55].但是大多数基于概率信息算法都涉及公式推导,不容易理解,实现起来比较困难;而且算法计算开销较大,匹配准确性很低,满足不了实时性和大众对 GPS 系统的要求.

2.1.4 高级路网匹配算法

高级路网匹配算法往往综合考虑全面信息,包括最近出现的道路网络的拓扑结构和轨迹数据中的噪音;使用更精致的概念,如卡尔曼滤波^[43]、模糊逻辑模型^[23]、隐式马尔可夫模型^[53,56]等.由于采用了先进的技术手段,这些高级路网匹配算法的准确率普遍较高.尤其是在空旷的道路,比如郊区、农村等.在城市道路上效果也不错,但受复杂路网、采样率等因素影响较大.当采样率较低时,这些算法仍然表现不佳,当采样间隔超过 5min 时,错误率超过 50%^[56].

2.2 以考虑采样点的范围来划分

根据匹配轨迹时考虑采样点的范围,路网匹配算法可以分为两类:局部/增量和全局方法.具体见表 2.

Table 2 Algorithm classification by range of sampling points
表 2 以考虑采样点的范围进行算法分类

算法	代表性工作	算法描述	复杂度	优缺点分析
局部 / 增量方法	文献[57]	使用距离相似度和方向相似度组合来评价候选边	$O(n)$, n 是采样点个数	除了位置样本外,该方法假设预期行驶路线没有任何进一步知识,而且也不使用任何速度信息
	文献[58]	“自适应裁剪”,将轨迹分段,采用 Dijkstra 算法在两个采样点之间构建最短路径	$O(mn\log m)$, m 是路网的边数, n 是采样点个数	增量算法快,质量低;全局算法慢,质量好.达到二者的平衡,提高计算速度
	文献[51]	使用一个恒定深度的基于局部匹配的几何形状的递归前瞻策略	$O(mn\log mn)$, m 是路网的边数, n 是采样点个数	文章提出 2 类 3 种算法:局部/增量 1 种;全局算法 2 种.实验结果表明,全局算法的匹配效果更好
	文献[59]	引入基于分段的匹配算法,给不同采样点分配置信度值.先匹配高置信度分段,然后使用之前匹配的边来匹配低置信度分段	$O(m\log m + m\log G + m C S)$, S 是调用相似度函数所需要的时间, m 是采样点个数, C 是候选结果数, $ G $ 表示地图的大小	采样率很高时,执行速度快且效果好.如果是低采样率的话,两个采样点之间断了联系则很难匹配,精度显著下降
	文献[17]	追求一小部分轨迹的局部匹配.当匹配一个新 GPS 点时,要考虑其前一个位置和最后匹配的边	—	对于只考虑当前位置的方法,结果受到测量误差影响很大.因此,局部/增量算法的准确性普遍较低,因为相邻点的相关性被完全忽略

Table 2 Algorithm classification by range of sampling points (Continued)

表 2 以考虑采样点的范围进行算法分类(续)

算法	代表性工作	算法描述	复杂度	优缺点分析
全局算法	文献[57]	该算法在概念上从左到右(沿着轨迹的方向)在所有自由空间上扫描一条线,同时保持在扫描线上的点通过左下角开始的自由空间中的一些单调路径可达.然后在扫描线推进的同时更新该 Dijkstra 风格的可达性信息	$O(mn\log_2 mn)$, m 是路网的边数, n 是采样点个数	该算法对所有临界值应用参数搜索.然后它通过在从左下角到右上角的自由空间中找到单调路径来解决决策问题
	文献[51]	提出两种全局算法:一种使用弗雷歇距离来衡量 GPS 轨迹和候选路段序列的匹配度,旨在最小化弗雷歇距离;另一种扩展上述算法,用弱弗雷歇距离(weak-Fréchet-distance)来减少异常值的影响	$O(mn\log^2 mn)$, m 是路网的边数, n 是采样点个数	采用一种前瞻技术来减少遗漏路段的情况,进一步提高准确率.然而,当前方法没有尝试使用真实路径来评估实际匹配精度.同时,它忽略了轨迹中的时间/速度信息
	文献[49]	采用离线捕捉方法,提出了一种基于道路网络的加权图表示的算法,其中,每个边的权重表示轨迹到地图上的弧序列的距离.通过用于加权图的 Dijkstra 最短路径算法来找到道路网络中的匹配轨迹,即与初始轨迹具有最小距离的路线		基于类似于平均弗雷歇距离(average-Fréchet-distance,简称 AFD)的测量,然而在匹配曲线上没有给出整体质量保证.同时,数据集上的一些细节,如类型和大小缺失
	文献[60]	通过将相邻时刻轨迹点的候选匹配路段构建成候选匹配图,并从该图中求解并搜寻一条最佳匹配路径,该路径即对应着最终的匹配结果	$O(nk^2m\log m+nk^2)$, n 是轨迹点个数, m 是路网路段数, k 是每采样点候选点个数	文献[56,61,62]与文献[60]的思路类似,只是在最优化模型的求解上分别采用了投票机制、隐马尔可夫模型及 A^* 算法等
	文献[63]	采用最优化模型来求解匹配结果	$O(knm^2)$, m 是候选路段数, n 是查询长度	其最可能行驶路径是通过历史轨迹数据的查询来获得的

2.2.1 局部/增量路网匹配算法

局部/增量算法利用当前 GPS 轨迹点的局部特征来进行路网匹配.算法使用贪心策略,从已经匹配好的解决方案中依次延伸得到最终的全局解.文献[57]根据 GPS 轨迹点与路段的几何距离、行车方向以及前后时刻路径可达性等因素来综合判断匹配结果.文献[53,64]则进一步考虑了在拓扑规则约束和交通规则约束下的路段连通性,匹配结果更为可靠.

局部/增量算法计算速度快、实时性强,所以经常用于有在线需求的应用.当采样速率较高时(比如 2s~5s),能够获得良好的匹配效果.但随着采样率的降低,局部/增量算法将呈现“弧跳”现象^[57],造成精度的显著降低.由于只考虑当前位置,相邻点的相关性被完全忽略,结果受到测量误差影响很大.因此,局部/增量算法的准确性普遍较低.

2.2.2 全局路网匹配算法

全局算法考虑整个采样轨迹,从路网中找到一条与其最接近的匹配路径.大多数全局路网匹配算法^[49,51]使用“弗雷歇距离”或其变体来测量采样轨迹和匹配路径的相似度.“弗雷歇距离”考虑了曲线的连续性,因此适用于比较轨迹.两曲线之间的“弗雷歇距离”由文献[59]提出,文献[29]给出了一种计算方法.文献[56,60,62,63]则同时考虑多个相邻时刻 GPS 轨迹点的位置信息,将路网匹配问题转化为给定目标函数的优化求解问题,其最优解即对应路网匹配的结果.

与局部/增量算法相比,全局算法更准确,对采样率降低表现出更大的健壮性,能够较好地处理长采样间隔情况下的路网匹配问题,但计算成本较高.由于缺乏真实的车辆和道路实验,使得它难以评估实际的匹配精度.此外,目前的全局匹配算法只采用空间分析,而忽略了轨迹的时间/速度约束,这使它们也容易受到采样率下降的影响.由于全局算法需要整个行驶轨迹的信息,所以通常适用于离线任务(例如在所有轨迹产生后,挖掘频繁的轨迹模式).

2.3 以采样的频率来划分

如表 3 所示,现有的路网匹配算法根据 GPS 数据的采样密度可分为高频采样算法、低频采样算法和更低频率采样算法.

Table 3 Algorithm classification by sampling rate
表 3 以采样的频率进行算法分类

算法	采样设备	采样频率	代表
高频采样算法	浮动车 GPS+航位推算罗盘/高度计	1s~10s	所有局部算法、部分全局算法
低频采样算法	浮动车 GPS+航位推算罗盘/高度计	30s 以上	HMMM ^[56] ,ST-Matching ^[60] ,IVVM ^[61] 等
更低频率采样算法	手机 GPS/手机基站+WIFI/惯性传感器/蜂窝指纹	2min 以上	VTrack ^[65] ,CTrack ^[66] ,WheelLoc ^[67] ,AutoWitness ^[68] ,semMatch ^[69] ,SnapNet ^[70,71]

2.3.1 高频采样算法

GPS 样本点的采样频率影响着算法的匹配精确度,当 GPS 采样频率较高时(一般是 1s~10s),即使仅将轨迹点匹配至最近的道路片段,也可得到较高的正确率.高采样率轨迹的路网匹配技术已经被商业化为个人导航设备.几乎上述所有的路网匹配算法都是基于高样本率数据集的.

然而,高采样率的数据集的采集和存储需要消耗更多的存储空间、GPS 模块的操作和计算时间.基于节约能源以及浮动车本身特性等原因,在实际应用中,返回给调度中心的数据往往是低采样率的(1min~2min 采样一次,甚至更低).

2.3.2 低频采样算法

在实际中,存在大量的低采样率 GPS 跟踪数据(采样间隔超过 1min).例如,为了节省通信和能源成本,出租车通常以较低的频率向调度中心报告 GPS 位置.文献[61]给出了 10000+北京出租车 1 周产生的 GPS 轨迹的采样间隔的统计分布,如图 7 所示.可以指出,只有 34%的数据是高采样率(采样间隔小于 1min)的数据,超过 60%的数据是低采样率数据.数据集的平均时间间隔为 3.27min.

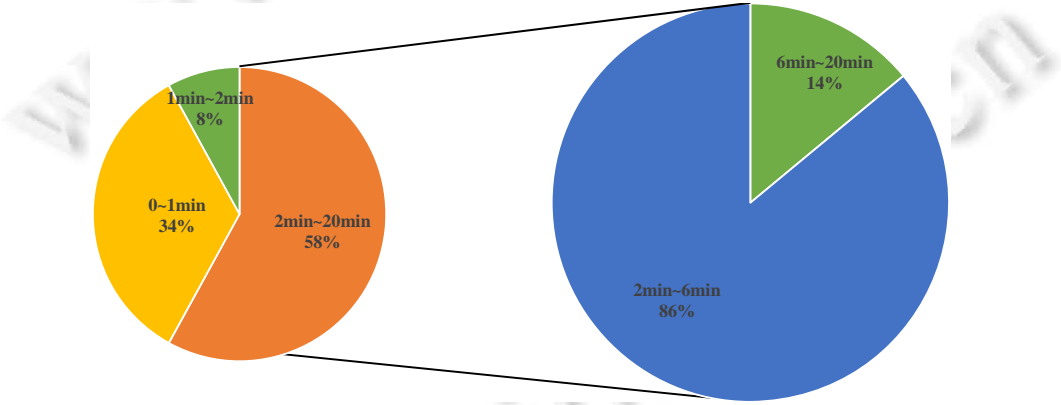


Fig.7 Distribution of the sampling interval ^[61]
图 7 采样间隔的分布^[61]

在这种情况下,基于高样本率的路网匹配算法无法适用于低样本率的数据集.针对此问题,研究人员已经提出了一些基于低采样率的数据集的路网匹配算法,如 ST-Matching^[60],IVVM^[61]等,但仍存在着一些误计算.由 IVVM^[61]的实验结果可见:给定一个轨迹的采样速率约为 2min,路网匹配算法的最高精度约为 70%.低采样率的轨迹处理技术仍然非常具有挑战性.

2.3.3 更低频采样算法

上述的路网匹配技术主要用于 GPS 定位数据,GPS 具有全球可用性和较高的精度.然而在许多形况下,基于

粗粒度网络的定位信息是唯一可行的解决方案,例如从蜂窝提供商处定位,在需要高效节能定位的情况下^[72-76],应用涉及低端手机没有 GPS 或其他传感器,以及人群的遥感应用中 GPS 通常关闭^[77]。

因此,一些新的路网匹配技术开始出现,利用手机上的其他传感器(如 WiFi^[65],具有周围手机信号塔信息的蜂窝指纹^[66]和惯性传感器^[67,68])来提高手机定位精度,加入一些基于道路语义的判断^[69],从而提供更好的路网匹配精度。然而,它们所使用的传感器仍然具有噪声,并且功率消耗较高,导致定位不准确。

现在,手机使用率非常高,运营网络也是现成的。如果能够发动手机用户积极参与交通流预测,则会极大地提高预测的实时性、准确性和覆盖率,而且投入和维护的成本都比较低。因此,这种方法非常易于推广普及。但在这种情况下,传统路网匹配算法的典型假设不成立。手机位置不再接近实际路段,位置误差是以 km 计,改变手机塔关联会有来回转换现象,且输入位置样本高度稀疏。所以,亟需一些能够处理更低采样率和定位精度的路网匹配算法。

2.4 以计算的实时性来划分

路网匹配算法可以分为实时和离线算法:实时算法将记录过程中的位置与道路网络联系起来;离线算法在数据被记录后使用,将数据匹配到道路网络。

在实时应用中,如交通感知,希望进行路网匹配,而未来的数据尚不可知,这就要求算法必须在线。实时应用程序只能基于给定时间前的点计算(而不是一个完整的旅程),意在现实生活环境使用。上文介绍的增量算法^[30,78,79]采用本地化的战略,将输入轨迹划分成较小的部分按顺序处理,适合在线匹配,但是有时会导致一个次优的解决方案。这就需要在性能和精度之间取得平衡。一个可靠的在线路网匹配算法要充分考虑到这些问题,并保证输出的准确性和及时性。

而离线应用可以考虑所有的点,所以可以容忍较慢的性能而有利于提高精度。上文介绍的全局算法^[49,80]批量输入轨迹从而生成整个解决方案,需要整个行驶轨迹的信息,所以通常适用于离线任务。

目前,因为要节省能源成本和通信成本以及车辆本身的特性,输入 GPS 数据的采样率普遍较低,两个采样点之间的许多细节很容易被丢失。例如,当采样间隔为 2min 时,即使车速为 30km/h,两个采样点之间的距离也可达到 1km。特别是在复杂的城市道路网络中,车速较高、街区较短、误差较大的情况下,大部分 GPS 点之间的关联信息将会丢失^[81],从而导致上述绝大多数传统路网匹配方法失效。因为它们只能处理高采样率的 GPS 数据,对低采样率的点有效性下降严重。如文献[82]中所述,当 GPS 采样间隔大于 120s 时,大多数全局算法表现不佳,平均正确率小于 60%。

在近几年的研究中出现了一些基于低采样率的数据集的路网匹配算法,可以较好地处理这个问题,比较著名的方法就是基于隐马尔可夫模型(hidden Markov model,简称 HMM)的路网匹配算法,其结果的正确率在一定条件下最高可以达到 90%。下一节将重点介绍基于 HMM 的路网匹配算法。

3 基于 HMM 的路网匹配算法

3.1 隐马尔可夫模型

20 世纪 60 年代,Baum 等人和其他作者在一系列的统计学论文^[83-85]中描述了隐马尔可夫模型。它最初的应用之一是语音识别,80 年代成为信号处理的研究重点,现已成功地用于故障诊断、行为识别、文字识别、自然语言处理以及生物信息等领域。

隐马尔可夫模型主要包含 5 个要素——2 个状态集合和 3 个概率矩阵(<http://baike.baidu.com/>)。

- 隐含状态 S : 马尔可夫模型中实际所隐含的状态,通常无法通过直接观测得到。这些状态之间满足马尔可夫性质。令 N 代表隐含状态数目。
- 可观测状态 O : 可以通过直接观测而得到的状态,在隐马尔可夫模型中与隐含状态相关联。令 M 代表可观测状态数目。
- 状态转移概率矩阵 A : 描述隐马尔可夫模型中各个状态之间的转移概率 $A=[a_{ij}]_{N \times N}$, 其中, $a_{ij}=P(q_{t+1}=S_j|$

$q_i=S_i), 1 \leq i, j \leq N$. 表示在 t 时刻状态是 S_i 的条件下, 在 $t+1$ 时刻状态为 S_j 的概率.

- 观测状态概率矩阵 $B: B=[b_j(k)]_{N \times M}$, 其中, $b_j(k)=P(o_t=O_k|q_t=S_j), 1 \leq j \leq N, 1 \leq k \leq M$. 表示在 t 时刻隐含状态是 S_j 条件下, 其可观测状态为 O_k 的概率.
- 初始状态概率矩阵 $\pi: \pi=(\pi_i), \pi_i=P(q_1=s_i), 1 \leq i \leq N$. 表示隐含状态在初始时刻 $t=1$ 的概率矩阵.

3.2 基于隐马尔可夫模型的路网匹配

隐马尔可夫模型通过对道路的连通性进行明确建模, 并同时考虑许多不同的路径假设来解决路网匹配问题. 最早利用 HMM 来进行路网匹配的应用来自 Lamb 等人^[86]的工作. 他们使用的是卡尔曼滤波器和 HMM 的组合. 几个卡尔曼滤波器沿着不同的假设路径跟踪车辆, HMM 在这中间进行选择. 来自 Hummel^[87]和 Krumm^[88]的著作使用隐马尔可夫模型来平衡测量噪声和路径概率.

隐马尔可夫模型可以解决 3 类问题.

- 评价: 将最可能的系统与观测序列匹配, 使用前向算法(forward algorithm)求解.
- 解码: 确定最可能产生观察序列的隐藏序列, 使用维特比算法(Viterbi algorithm)求解.
- 学习: 确定模型参数最可能产生的一系列观察, 使用前向-后向算法(forward-backward algorithm)求解.

路网匹配问题实际上是一个解码问题. 假设移动对象在路网的交叉点间移动, 且移动的过程符合马尔可夫模型. 虽然无法直接观察到移动对象所经过的道路轨迹(可以把这些轨迹看作隐含状态), 但是可以获得隐含状态的输出, 即 GPS 样本点(假设观测值仅与移动对象的隐含状态相关). 于是, 基于 HMM 的路网匹配算法就变成在给定一系列观察的前提下, 寻找最有可能产生这个观察序列的隐含状态序列, 也就是实际的移动轨迹. 这里有两个状态集合.

- 可观测状态集合: 从 GPS 设备中得到的位置信息(经度、纬度).
- 隐含状态集合: 拥有 GPS 设备的物体(车、人等)实际所在的位置.

这样, 就可以把路网匹配问题与隐马尔可夫模型结合起来. 然后, 用一定的规则去拟合历史的真实数据得到状态转移概率矩阵 A 、观测状态转移概率矩阵 B 以及初始状态概率矩阵 π . 接下来将详细介绍基于 HMM 的几种算法, 其中, 定义的规则要满足人的先验知识, 主要有以下 3 种.

- 观测状态概率. 观测的 GPS 样本点离候选路段越近, 这个样本点在这个路段上的概率就越大.
- 状态转移概率. 这里有两种解决思路:
 - 一是前后两个 GPS 样本点的距离越近, 状态转移的概率就越大;
 - 二是真实路段上的前后两个点之间距离与 GPS 观测的前后两个样本点之间距离越接近, 状态转移概率就越大.
- 初始状态概率. 代表移动对象初始时位于某个路段上的概率, 使用相应 GPS 样本点的观测状态概率来表示.

在基于 HMM 的路网匹配算法中, 对于每一个 GPS 轨迹点, 首先确定一组候选路段. 每个候选路段被表示为马尔可夫链中的隐藏状态(顶点), 并具有观测状态概率, 这是观察 GPS 点是否与候选路段匹配的可能性. 如果发现 GPS 点非常接近某个路段, 就给这个路段指定一个较高的概率值. 然后, 为马尔可夫链中连接每一对相邻顶点的边计算权重, 即状态转移概率. 最终, 在马尔可夫链上找到具有最高观测状态概率和状态转移概率的最大似然路径. 一般使用维特比算法(Viterbi algorithm)进行求解, 实际上是用一个动态规划(dynamic programming, 简称 DP)算法解隐马尔可夫模型预测问题, 即用动态规划在道路网络中快速找到使观测概率和转移概率的乘积最大化的最优路径. 这个过程如图 8 所示.

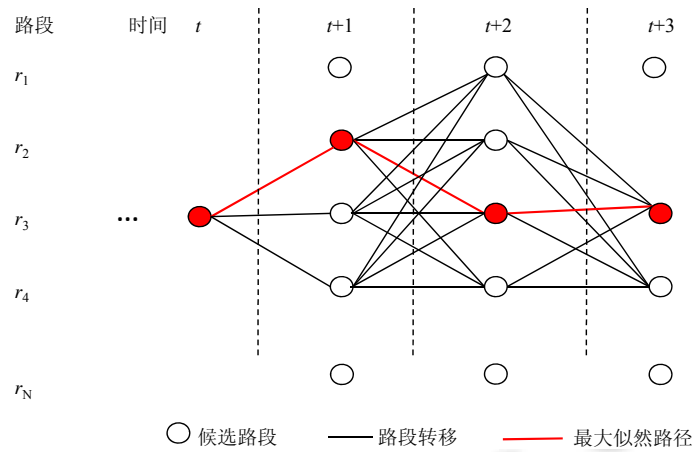


Fig.8 HMM-Based map matching process

图 8 基于隐马尔可夫模型的路网匹配过程

图 9 用一个简单的例子来说明基于隐马尔可夫模型的路网匹配过程,观测状态概率和状态转移概率都在图中列出.

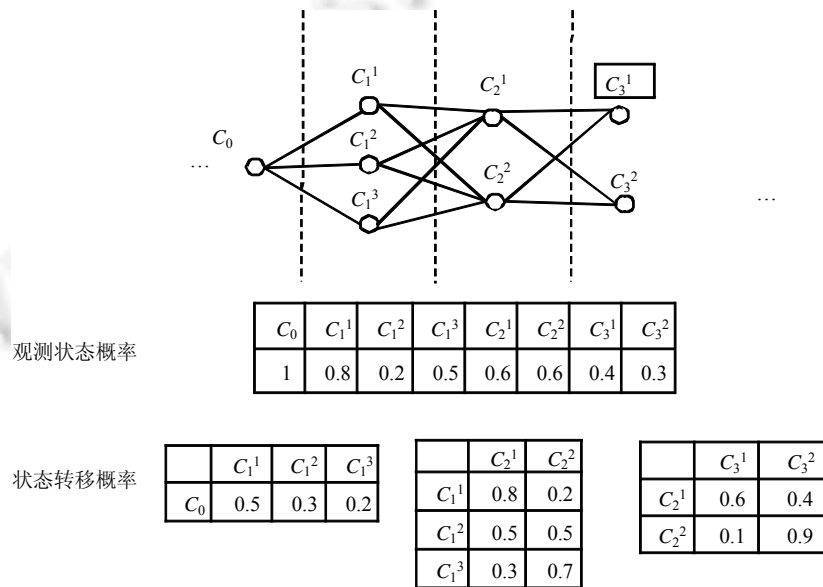


Fig.9 An example of HMM-based map matching process

图 9 基于隐马尔可夫模型的路网匹配过程示例

从 C_0 点出发,有 3 个候选点 C_1^1, C_1^2, C_1^3 , 它们的得分分别是 0.4, 0.06, 0.1, 最大值是 C_1^1 , 于是, 选择 C_1^1 作为 C_0 匹配的结果. 接下来考虑 C_1^1 , 先计算路径 $C_1^1 \rightarrow C_2^1 = 0.6 \times 0.8 = 0.48$, $C_1^1 \rightarrow C_2^2$, $C_1^2 \rightarrow C_3^1$ 类似. 于是, C_2^1 的得分等于 $0.4 + 0.6 \times 0.8, 0.06 + 0.6 \times 0.5, 0.1 + 0.6 \times 0.3$ 中的最大值, 即 0.88. 然后, 针对每个候选点重复上述过程. 基于之前存储的计算信息, 算法最终得到匹配结果是 $C_0 \rightarrow C_1^1 \rightarrow C_2^1 \rightarrow C_3^1$, 整个过程如图 10 所示.

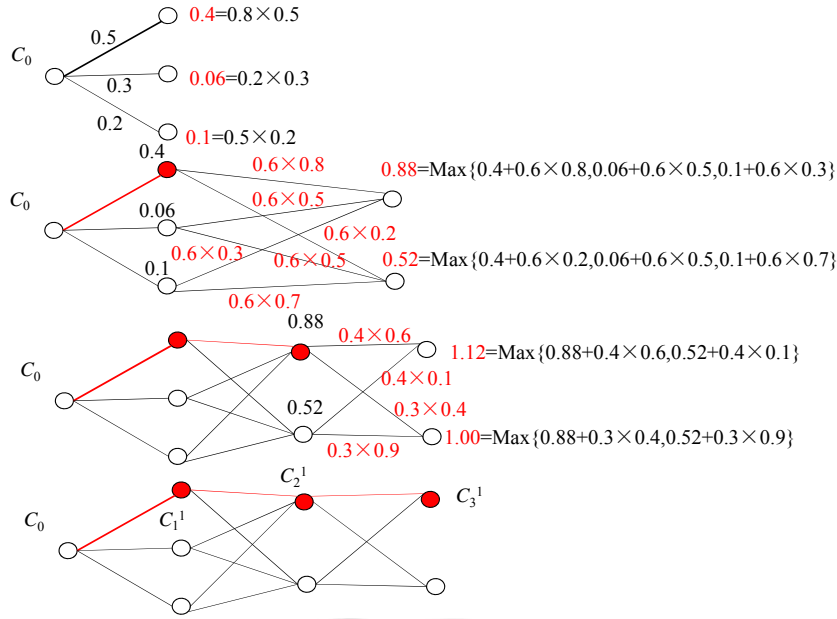


Fig.10 Calculation process of HMM-based map matching algorithm

图 10 基于隐马尔可夫模型的路网匹配算法计算过程

3.3 基于隐马尔可夫模型的路网匹配算法

3.3.1 ST-Matching 算法(spatial temporal matching algorithm)^[60]

ST-Matching 算法结合时间和空间特征,包括几何拓扑结构和速度约束.文献[60]使用高斯分布来模拟观测概率转移,而状态转移概率则是综合了速率信息、GPS 观测点和真实点的信息计算出来,具体公式如下.

- 观测概率矩阵 B

$$b_j(k) = P(o_t = o_k | q_t = r_j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\|o_t - r_j\|^2}{2\sigma^2}} \quad (1)$$

一般来说,GPS 观测点的测量误差可以被合理地描述为一个 GPS 观测点和实际路段之间距离的高斯分布.这里采用的是一个均值 $\mu=0$,标准差 $\sigma=20$ 米的高斯分布. $\|o_t - r_j\| = \text{dist}(c_i^j, p_i)$,即观测点 p_i 和其候选点 c_i^j (c_i^j 代表观测点 p_i 的第 j 个候选点)之间的距离.它表明一个 GPS 观测点 p_i 可否匹配到真正道路上的一个候选点 c_i^j ,而不考虑其邻近点.时刻 t 的观测点与候选点之间的距离越小,这个候选点是真正的实际点的概率就越大.这充分考虑了道路网络的几何性质和局部特性.

- 状态转移矩阵 A

$$\left. \begin{aligned} a_{ij} &= P(q_{t+1} = r_j | q_t = r_i) = F_i(r_i \rightarrow r_j) \cdot V(r_i \rightarrow r_j) \\ F_i(r_i \rightarrow r_j) &= \frac{d_{t \rightarrow t+1}}{w_{t \rightarrow t+1}} = \frac{\|o_t - o_{t+1}\|_{\text{Euclidean_Distance}}}{\|r_i - r_j\|_{\text{route}}} \\ V(r_i \rightarrow r_j) &= \frac{\sum_{u=1}^k (e'_u \cdot v_r \times \bar{v}_{ij})}{\sqrt{\sum_{u=1}^k (e'_u \cdot v_r)^2} \times \sqrt{\sum_{u=1}^k \bar{v}_{ij}^2}} \end{aligned} \right\} \quad (2)$$

$F_i(r_i \rightarrow r_j)$ 是空间分析函数,考虑了道路网络的拓扑结构.根据前后两个时间点 t 和 $t+1$ 之间的观测点 p_{t-1}, p_t 及其候选点 c_{t-1}^i, c_t^s 的信息,推测从 p_{t-1} 到 p_t 的真实路径是 c_{t-1}^i 到 c_t^s 最短路径的可能性.其中, $d_{t \rightarrow t+1} = d_{i1 \rightarrow i} = \text{dist}(p_i, p_{i-1})$ 是相邻观测点 p_i 和 p_{i-1} 之间的欧氏距离(Euclidean distance). $w_{t \rightarrow t+1} = w_{(i1, t) \rightarrow (i, s)}$ 为通过候选节点 c_{t-1}^i 和

c_i^s 的最短路径的长度.这样匹配出的结果可以避免出现绕路或者道路不连通等情况.

$V(r_i \rightarrow r_j)$ 是时间分析函数,考虑前后两个时间点之间的速率信息.用余弦距离(cosine distance)来衡量从 c_{i-1}^t 到 c_i^s 平均速度跟路段速率约束的相似度. c_{i-1}^t 到 c_i^s 的最短路径定义为一个路段列表 $[e'_1, e'_2, \dots, e'_k]$, 每个路段 e'_u 还与其限制的最大速率 $e'_u.v_r$ 相关联. \bar{v}_{ij} 是这个最短路径的平均速率.

算法根据路段投影过程为每个 GPS 采样点检索一组候选路段和候选点.然后构造一个候选图,图中的节点是每个 GPS 观测的候选点集合,而边是任意两个相邻的候选点之间的最短路径集.节点和边都基于空间/时间分析结果分配了权重值.结合观测和转移概率,匹配算法在候选图中寻找一条具有最高得分的路径,最大限度地提高全局匹配概率.

3.3.2 IVMM 算法(interactive-voting based map matching algorithm)^[61]

ST-Matching 在匹配路网时,会把 GPS 点对应到距离较近的路段上,并不考虑其相邻点的关系,所以会出现锯齿形的线路.IVMM 的提出,就是为了“拉直”不匹配的路段,使其更适合地面真实路径.它在 ST-Matching 的基础上加入了“互动投票”,就是说,每个候选点会对在同一条单点最优的轨迹上的其他候选点进行加强,即投票,最后打分最高的点连起来形成匹配路径.

观测概率矩阵(同公式(1))和状态转移矩阵(同公式(2))的计算方法和 ST-Matching 相同,只是公式(1)中的取值不同,这里 $\mu=5\text{m}$, $\sigma=10\text{m}$.先选取相邻观测点 p_{i-1}, p_i 和相应的候选节点 c_{i-1}^t, c_i^s , 经过匹配函数计算出初始静态匹配矩阵.然后定义每个观测点的距离加权函数,全局计算得到动态匹配矩阵.在每个观测点对应的动态匹配矩阵中,并行进行单点投票.最后对全局投票结果进行汇总.

IVMM 算法从全局角度考虑观测点之间的相关性.所有观测点之间进行相互投票,完全并行,这样就不需要进行大量的计算得到可能的匹配结果,而只需从候选集合中投票产生正确匹配结果,算法复杂度较小.

3.3.3 HMMM(hidden markov model map matching through noise and sparseness)算法^[56]

与 ST-Matching 算法相近,但没有考虑速率信息.它是用前后两个相邻 GPS 观测点的距离与两个候选点的距离之差的绝对值来表示这两段距离的接近程度,越接近,概率就越大,且最后的概率是用指数函数来拟合.具体公式如下.

- 状态转移矩阵 A

$$a_{ij} = P(q_{t+1} = r_j | q_t = r_i) = \frac{1}{\beta} e^{\frac{-d_t}{\beta}} \quad \left\{ \begin{array}{l} d_t = ||o_t - o_{t+1}||_{\text{great_circle}} - ||r_i - r_j||_{\text{route}} \end{array} \right. \quad (3)$$

其中, β 选用 Gather 等人^[89]建议的鲁棒性较强的估计值,见公式(4).

$$\beta = \frac{1}{\ln(2)} \text{median}_t (||o_t - o_{t+1}||_{\text{great_circle}} - ||r_i - r_j||_{\text{route}}) \quad (4)$$

β 较大、大圆距离和候选点之间最短距离越大,说明对非直线路径越容忍.

- 观测概率矩阵 B

与 ST-Matching 算法相同(见公式(1)),都是使用均值 $\mu=0$ 高斯分布.但标准差 σ 是使用绝对中位差(median absolute deviation,简称 MAD)进行鲁棒估计,见公式(5).

$$\sigma = 1.4826 \text{median}_t (||o_t - r_j||_{\text{great_circle}}) \quad (5)$$

其中, $||o_t - r_j||$ 使用的是观测点和其候选点之间的地球表面上的大圆距离.对于使用的测试数据,该值为 $\sigma=4.07\text{m}$, 这是 GPS 噪声的合理值. σ 值越大,噪声就越大,表示位置测量越不可信. σ 和 β 两个参数需要确定合适值才能使算法性能最高.

- 初始状态概率 π

$\pi_i = P(p_1 | r_i), 1 \leq i \leq N$, 在初始时刻 $t=1$ 时的观测概率矩阵,有 $\pi_i = P(p_1 | r_i)$, 即使用第 1 个 GPS 观测点 p_1 得到 r_i 是所有道路中第 1 条匹配道路的概率.

状态转移概率依赖于当前状态以及上一个观察状态,所以不能严格满足维特比算法中对状态转移概率的要求,从而产生两个问题:一是利用公式(1)得到的近似概率并不一定是其最大似然结果;二是公式(1)将观测点在道路上的投影作为隐性状态的一部分,但这个投影并不唯一,从而增大了最大似然解的寻找难度.文献[81]中的算法简化状态转换概率以适应理想的 HMM 框架,避免了这些缺点.

3.3.4 Simplified HMM 算法(simplification of the more-complex HMM-based method)^[81]

上述论文中使用的模型并不是完美的原始的 HMM,因为在计算状态转移概率时,既要考虑候选点的信息,也要考虑观测点的信息,可以说是 HMM 的一个变种.本文将之前的 HMM 模型简化,在计算状态转移概率时不考虑观测点之间的信息,所以可以看作简化的 HMM 方法.图 11 给出了 Simplified HMM 算法所采用的模型与 HMMM 算法模型的对比示意图.

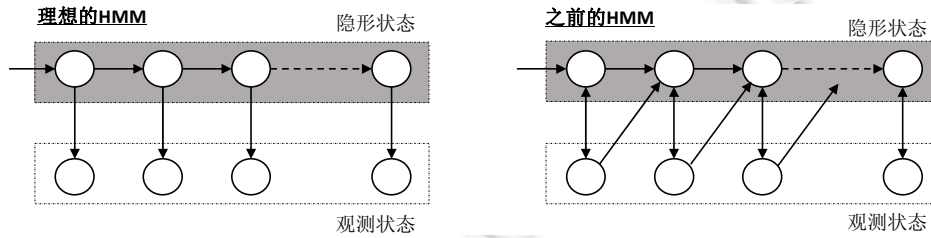


Fig.11 Comparison of HMMs

图 11 HMM 模型对比

隐性马尔可夫模型包含的 3 种概率如下.

- 观测概率矩阵 B

和之前的算法一样(见公式(1)),用高斯分布模拟(本文是 $\mu=0, \sigma=1.0$ 的高斯分布).

- 状态转移矩阵 A

车辆从一个道路交叉点转移到另一个交叉点的概率,对应于图 9 中的理想隐性马尔可夫模型中的状态转移.文献[81]中假设状态转移概率跟两个交叉点间的距离成正比,即

$$a_{ij} = P(q_{t+1} = r_j | q_t = r_i) \propto e^{-\beta d_{ij}} \quad (6)$$

其中, β 用于调节两个交叉点间最短距离 d_{ij} 对结果的影响.距离 d_{ij} 可以考虑行驶时间、成本等距离之外的其他因素.

- 初始状态概率 π

代表车辆在初始时刻处于某个道路交叉点的概率,文献[81]中使用相应交叉点的观测概率来表示车辆的初始状态概率.

之前的算法的状态转移概率公式中都有 $\|O_t - O_{t+1}\|$,说明状态转移概率依赖于当前及上一次观察状态,不能严格满足 Viterbi 算法中对状态转移概率的要求.Simplified HMM 算法按照理想模型设定了状态转移概率公式,错误率比 HMMM 算法略大,说明变种 HMM 更接近现实路网匹配.但 Simplified HMM 算法的匹配效率提高了很多,约为前者的 2 倍.该算法实现更为简单,且在处理低频采样的 GPS 轨迹时仍有较好的性能.

3.3.5 OHMM(online map-matching based on hidden Markov model)算法^[71]

算法使用在线的维特比算法来实现实时的路网匹配.计算观测概率矩阵时考虑了路段的宽度信息和速度信息.在计算状态转移矩阵时使用了机器学习的方法.

- 观测概率矩阵 B

$$b_j(k) = P(o_t = o_k | q_t = r_j) = S(v_t, v_r) P(\text{observation}) \quad (7)$$

$$P(\text{observation}) = \frac{1}{2w} \int_{-w}^w \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(l-d)^2}{2\sigma^2}} dl \quad (8)$$

$P(\text{observation})$ 与之前的算法相似,也是使用 $\mu=0$ 的高斯分布拟合.但这里也考虑了路段的宽度信息,为了能够更好地区分路段,特别是在路口.其中, $2w$ 是路段宽度, d 是GPS样本点 t 和道路路段 r 之间点到曲线的大圆距离.GPS误差的估计标准偏差 σ 是通过分析地面真实数据的扰动来估计的.对于每个轨迹点,计算该点到离其最近路段中心的大圆距离.然后,用与文献[56,90]相同的基于距离的绝对中位差计算标准偏差.基于整个数据集最终获得 $\sigma=6.86\text{m}$.

$$S(v_t, v_r) = \frac{v_r}{\max(0, v_t - v_r) + v_r} \quad (9)$$

$S(v_t, v_r)$ 是引入的一个超速处罚因子,它是基于这样的假设:驾驶员不太可能在很大程度上超过速度限制.目的是帮助区分从同一路口分叉出的可能具有不同速度限制的紧密间隔的平行道路.因为在这种情况下,单独的位置测量不足以区分路段,因为记录的轨迹点可能落在它们之间.其中, v_t 是 t 的记录速度, v_r 是路段 r 的速度限制.如果遵守了速度限制,那么 $\max(0, v_t - v_r)=0$,不会产生代价. $S(v_t, v_r)$ 的作用和算法^[60]中的速率变化功能是一样的.

• 状态转移矩阵 A

使用支持向量机(support vector machine,简称 SVM)来探悉状态转移概率,而不是像之前的算法一样先验地选择模型,然后估计其参数.这种方法的主要优点是提供了一个数据驱动的框架,用于集成多个转移评分函数.文献[71]中设计了两个评分函数.

$$\text{距离差异函数: } T(d_{i \rightarrow j}, D_{i \rightarrow j}) = \frac{|d_{i \rightarrow j} - D_{i \rightarrow j}|}{D_{i \rightarrow j}} \quad (10)$$

$$\text{动量变化函数: } M(\vec{v}_0, \vec{v}_1, l_1, \dots, \vec{v}_k, l_k) = \frac{\sum_i^K l_i |\vec{v}_i - \vec{v}_{i-1}|}{\vec{v}_{i \rightarrow j} \sum_i^K l_i} \quad (11)$$

距离差异函数测量传感器推导的行驶距离和最短路径长度之间的差异. $P_{i \rightarrow j}$ 定义为从 i 行驶到 j 时车辆最大可能采取的最短路径. $d_{i \rightarrow j}$ 表示前后两个时刻之间以平均速率 $\vec{v}_{i \rightarrow j}$ 行驶的距离, $D_{i \rightarrow j}$ 是 $P_{i \rightarrow j}$ 的路径长度.公式通过比较候选路径 $P_{i \rightarrow j}$ 的长度与导航推算(deduced reckoning,简称 DR)估计来评估候选路径的可行性.如果 $P_{i \rightarrow j}$ 是真实路径,则距离差异接近于0.

动量变化函数测量车辆在 $P_{i \rightarrow j}$ 中采取每个路段上所引起的平均动量变化.这里, $(\vec{v}_1, \dots, \vec{v}_k)$ 是 $P_{i \rightarrow j}$ 上各个路段车辆的速度矢量,符合线性分布. (l_1, \dots, l_k) 是对应路段的长度.动量变化函数可以描述为平滑因子来惩罚一些不可行的转换,比如有很多急转弯.

这两个公式计算出来的都只是一个特征,然后使用被标记为正确或不正确转换的实例训练支持向量机分类器,其中,特征向量由上述两个评分函数给出的分数组成.使用这种分类方法, $p(i \rightarrow j)$ 就是输入的分数组属于“正确转换”类的概率,也就是从 i 到 j 的转移概率.

为了利用增量方法实现在线的全局路网匹配算法,需要在不知道未来输入的情况下沿着马尔可夫链进行不可逆的在线决策,同时确保部分解在组合时产生全局最优解.

文献[71]使用在线维特比算法来解决 $V_n = b_j(k) \max \{a_{ij} V_{n-1}\}$ 中的递推关系.当前幸存路径在马尔可夫链中的某点收敛时,所有未来的幸存路径将包含直到收敛点的相同子路径^[91,92].

3.3.6 QMM(quick map matching using multi-core CPUs)算法^[93]

QMM是第1个强调运行时间的路网匹配算法.算法设计在多核CPU上运行,因为路段的处理可以在建立索引时彼此分离,并且每个样本轨迹点在匹配时相互独立.算法应用多线程技术极大地减少了运行时间,同时在HMMM算法基础上作了一些改进.

- 观测概率矩阵 B :算法考虑了侧路问题,基于车辆更可能在具有较高速度约束的主要道路上行驶而不是侧路行驶的假设,在计算观测概率时加入路段的速度限制 V_r :

$$b_j(k) = P(o_i = o_k | q_i = r_j) \cdot V_r \quad (12)$$

由于存在不可避免的GPS测量误差,在主路上行驶的车辆记录的GPS采样点位置可能更接近侧路,原始算法可能会错把这些点与侧路匹配.

此外,算法也考虑了类似于 HMMM 算法^[56]的 HMM 中断问题(详细说明见第 3.5.2 节),提出适当地放大参数 β ,可以有效地减少中断情况,对匹配精度具有不可估量的影响.并通过实验,给出不同采样周期内 β 的最优值.

上述基于隐马尔可夫模型的路网匹配算法的对比见表 4 和表 5.隐马尔可夫模型被理论和实践证明适用于低采样率的路网匹配问题,为了进一步提高匹配精度,在研究中引入了一些优化算法、启发式算法等.如吴刚等人^[94]提出的一种将 HMM 和遗传算法相结合的新型地图匹配算法.在使用传统的 HMM 地图匹配算法预测出一组概率较大的路径序列,然后把这组路段序列作为种群,通过遗传算法寻找最优的路段序列,提高了 HMM 算法的匹配精确度.

Table 4 Comparison of HMM-based map matching algorithm setup
表 4 基于隐马尔可夫模型的路网算法设置比较

算法	观测概率矩阵	状态转移矩阵	初始状态概率
ST-Matching	均值 $\mu=0$,标准差 $\sigma=20\text{m}$ 的高斯分布.考虑道路网络的几何性质和局部特性	空间分析函数和时间分析函数的乘积.考虑了道路网络的拓扑结构和前后两个时间点之间的速率信息	—
IVMM	$\mu=5\text{m},\sigma=10\text{m}$ 的高斯分布	空间分析函数和时间分析函数的乘积.考虑了道路网络的拓扑结构和前后两个时间点之间的速率信息	—
HMMM	$\mu=0,\sigma=4.07\text{m}$ 的高斯分布. σ 使用绝对中位差进行鲁棒估计	用指数函数来拟合前后两个相邻 GPS 观测点的距离与两个候选点的距离之差的绝对值	在初始时刻时的观测概率矩阵
Simplified HMM	$\mu=0,\sigma=1.0$ 的高斯分布	假设状态转移概率与两个交叉点间的距离正相关,也考虑行驶时间、成本等距离之外的其他因素	使用相应交叉点的观测概率来表示移动
OHMM	$\mu=0,\sigma=6.86\text{m}$ 的高斯分布. σ 使用绝对中位差进行鲁棒估计.考虑了路段的宽度信息,引入一个超速处罚因子	使用支持向量机来探悉状态转移概率.两个特征为距离差异函数和动量变化函数	—
QMM	$\mu=0,\sigma=6.86\text{m}$ 的高斯分布.加入路段的速度限制	用指数函数来拟合前后两个相邻 GPS 观测点的距离与两个候选点的距离之差的绝对值	—

Table 5 Comparison of HMM-based map matching algorithm characteristic
表 5 基于隐马尔可夫模型的路网算法性质比较

算法	数据信息	考虑范围	采样频率	计算实时性	特点
ST-Matching	几何信息+拓扑信息	全局	低采样率	线下计算	较早的算法,结合时间和空间特征.与增量算法和基于 AFD 的全局匹配算法相比,具有更高的匹配精度
IVMM	拓扑信息	全局	低采样率	线下计算	在 ST-Matching 的基础上加入了“互动投票”环节,比 ST-Matching 有更高的匹配精度和更好的鲁棒性
HMMM	几何信息+拓扑信息	全局	低采样率	线下计算	首次提出基于隐马尔可夫模型路网匹配的概念,并公开实验数据,是之后许多研究的基础
Simplified HMM	几何信息	全局	低采样率	线下计算	简化的隐马尔可夫模型,实现虽然简单,但仍有较好的性能.而且匹配速度快
OHMM	拓扑信息	局部/增量	低采样率	在线计算	在线的路网匹配算法.以较低的输出延迟实现了较优的匹配结果
QMM	几何信息	全局	低采样率	线下计算	第 1 种强调运行时间的路网匹配算法.采用多线程技术极大地减少了运行时间

3.3.7 基于 HMM 的手机轨迹路网匹配算法

随着手机的普及,移动大数据成为重要的资源,提供了丰富的信息,而且几乎没有成本.近几年新出现的一个研究方向和热点是将 HMM 用在采样周期低、噪声误差大的手机定位.高频率采样的智能手机定位精确但无法实时定位,因为它消耗过多的智能手机的带宽和电池电源.因此,手机数据一般非常稀疏且噪声较大.有必要在确保精确和及时的情况下,为手机轨迹路网匹配开发鲁棒算法.

SnapNet 系统^[70,71]使用增量 HMM 算法为具有噪声和稀疏性特征的粗粒度蜂窝数据提供实时地图匹配,运用一系列的过滤器和一些启发式方法来减少噪声,利用插补解决数据稀疏问题.George 等人^[95]结合了 HMM 的方法和驾驶员路线选择的概念,使用 HMM 以在线方式为噪声和稀疏数据生成部分地图匹配路径,然后利用一个路线选择模型,从真实驾驶数据估计,重新评估每个 HMM 生成的部分路径以及一组可行的替代路径.Essam

等人^[96]总结并比较了一些利用手机数据进行路网匹配的方法,并提出了一种基于自适应 HMM 的模型,使用大量移动数据来学习模型参数,并利用数据的稀疏性来提供实时快速维特比处理,将各个手机轨迹映射到道路段.该模型首次单独使用手机大数据进行细粒度路网匹配.

隐马尔可夫模型对于低采样率路网匹配问题的适用性,使得它在手机数据匹配方面也得到较好的应用.由于数据的来源不同,接下来的实验比较中没有加入手机轨迹路网算法.

3.4 实验设置与测试

上述基于 HMM 的算法的实验环境与评价标准各有千秋(见表 6),详细分析如下.

Table 6 Experiment results comparison of HMM-based map matching algorithm
表 6 基于隐马尔可夫模型的路网算法实验结果对比

算法	评价标准	测试环境	比较对象	实验结果
ST-Matching	匹配质量, 运行时间	合成数据	增量算法, 基于 AFD 的全局匹配算法	当采样间隔为 2.91min 时, A_L 达到 95%, A_N 约为 92%. 当采样间隔降低到 5.77min 时, A_L 仍有 86%, A_N 为 80%
		城市: 北京, Geolife 中 28 条轨迹	增量算法, 基于 AFD 的全局匹配算法	当采样间隔为 30s 时, A_L 约为 87%, A_N 为 86%. 当采样间隔下降到 5min, A_L 约为 70%, A_N 为 68%. 采样间隔为 2min, 可以维持在 80% 的准确率
IVMM	匹配质量, 运行时间	城市: 北京, Geolife 中 28 条轨迹	ST-Matching 算法	当采样间隔为 1.5min~6.5min 时, IVMM 的精度维持在 70% 左右, 比 ST-Matching 算法有 10% 的改进. 当采样间隔超过 6.5min 时, IVMM 和 ST-Matching 算法的性能变得接近, 但仍具有 5% 的优势
HMMM	匹配质量,	城市: 西雅图, 7 531 个采样点	—	当采样间隔为 30s 时, RMF 为 0.11%. 随着采样率间隔不断增加, RMF 上升, 采样间隔 2min 的 RMF 为 0.1, 当到达 5min 以上时, RMF 超过 0.5
Simplified HMM	匹配质量, 运行时间	城市: 西雅图, 7 531 个采样点	HMMM 算法	当采样间隔为 90s 时, RMF 为 0.2. 采样间隔为 5min 时, RMF 为 0.6. 随着采样间隔的增加, RMF 逐渐上升. 当采样增加至 8min 时, RMF 为 1.1, 之后出现少许下降. 与 HMMM 算法相比, RMF 略大, 但效率较高, 约为前者的 2 倍
OHMM	匹配质量, 运行时间	城市、郊区: 新加坡, 4 条公共线路	城市路线和农村路线比较	除了大于 4min 的采样间隔外, 农村路线的准确性比城市路线的准确度高出大约 5%. 在小于 1min 的采样间隔, 两条路线的精度都高于 0.9. 在这两种情况下, 精度都随着采样间隔的增加而下降
QMM	匹配质量, 运行时间	ACM SIGSPATIAL Cup 2012 竞赛数据	不同线程下的比较	当采样间隔为 1s~30s 时, 不匹配率保持在 2% 以内, 波动不大

3.4.1 数据集

算法测试实验使用的数据源有 3 种类型: 合成数据、真实数据和竞赛数据.

• 合成数据

ST-Matching 算法测试使用的合成数据是由模拟器随机生成的轨迹. 它首先随机选择两个顶点的道路网络, 并计算它们之间最短的 k 条路径. 然后, 随机从 k 条路径中选择一个轨迹作为地面实况, 记为 $G:e_1,e_2,e_3,\dots,e_n$. 其背后的动机是移动物体通常遵循从源到目的地的方向, 但不一定严格遵循最短路径. 注意到任何两个相邻点之间的时间间隔不均匀. 为了检索具有所需采样间隔的轨迹, 模拟器从 G 中每个 k' 段选择一个路段, 比如选择 $e_1,e_{1+k'},\dots,e_{1+mk'}$, 其中, $m=\left\lfloor\frac{n}{i}\right\rfloor$. 因此, 可以通过改变 k' 的值来实现抽样率的调整. 模拟器生成一个具有每个选定路段的估计时间戳信息的 GPS 点. 论文研究集中在低采样率 GPS 轨迹上, 因此, 合成数据的采样间隔范围为 2min 至近 6min.

• 真实数据

ST-Matching 算法和 IVVM 算法实验中, 以人们标记的真实路径数据作为地面实况. 路网数据采用北京的城市道路网络图, 包含 58 624 个顶点和 130 714 个路段. GPS 轨迹数据都是从 Geolife 系统^[97,98]收集的真实数据. ST-Matching 算法从中选择了 28 条轨迹. 它们在空间范围、长度、时间间隔和道路段数上有所不同, 采样间隔

范围为 30s~5min. IVVM 算法数据集包含 26 个具有不同采样点和平均速度的轨迹. 30 多个小时的轨迹覆盖约 643km, 采样间隔范围为 30s~10.5min.

HMMM 和 Simplified HMM 算法也是在真实数据集上进行测试. 使用的是美国华盛顿西雅图的 GPS 数据 (<http://research.microsoft.com/en-us/um/people/jckrumm/MapMatchingData/data.htm>), 行驶距离为 80km, 包含 7 531 个 GPS 采样点, 采样间隔为 1s. 为了体现采样点的稀疏性, 还模拟了 90s, 120s, 180s, 240s, 300s, 360s, 420s, 480s, 540s 和 600s 等的采样周期. 所有测试数据、地面实况数据和道路网络都已公开, 供其他研究人员开发、测试和比较自己的路网匹配算法.

OHMM 算法使用在新加坡的 4 条公交线路收集的地面真实数据来评估算法的性能. 为了比较不同环境下的表现, 选择了 4 条涵盖新加坡农村和城市地区的路线. 农村路线的转弯次数较少, 主要由直通高速公路等直线路线组成. 城市路线除了大量分岔路外, 还覆盖着高层建筑密集的城市街区. 长度分别为 36.3km, 11.3km, 27.3km, 32.5km. 此外, 为了模拟不同采样频率的轨迹, 以 10s~5min 的采样间隔对原始数据(每 1s~3s 记录 1 次)进行二次采样.

• 竞赛数据

QMM 算法采用 2012 年 ACM SIGSPATIAL Cup (ACM SIGSPATIAL Cup 2012. Available from: <http://depts.washington.edu/giscup/home>) 的训练数据, 这是由 ACM SIGSPATIAL 主办的全球范围内的关于地图匹配算法的科技竞赛, 这也是此国际会议举办的第 1 次竞赛. 这次科技竞赛影响力吸引了来自全球 31 支专业级的参赛队伍. 所有算法当中匹配准确率最高的两个都是基于 HMM 的匹配算法. SIGSPATIAL GIS 编写的数据集可以帮助各研究所的研究人员掌握地面真实数据. 比赛是美国华盛顿州路网信息的简化版本, 原始地图数据从 Open Street Map (OSM) (OpenStreetMap. Available from: <http://www.openstreetmap.org/>) 获得, 采样间隔为 1s~30s.

3.4.2 评价标准

ST-Matching 算法在运行时间和匹配质量上进行了广泛的评价. 实验中使用实际的程序执行时间来衡量运行时间. 通过两个精度指标“准确的数量 A_N ”和“准确的长度 A_L ”来评估匹配质量, 其中,

$$A_N = \frac{\text{#正确匹配的路段数}}{\text{#轨迹中所有的路段数}},$$

$$A_L = \frac{\sum \text{匹配路段的长度}}{\text{轨迹的总长度}}.$$

IVMM 算法在匹配质量和效率方面与 ST-Matching 算法进行比较. 为了评估算法的效率, 实验比较了运行时间. 为了评估匹配质量, 使用下式计算的正确匹配百分比 (correct matching percentage, 简称 CMP).

$$CMP = \frac{\text{#正确匹配的样本点数}}{\text{#需要匹配的样本点数}} \times 100\%.$$

HMMM 算法、ST-Matching 算法、Simplified HMM 算法评价标准都采用道路错误匹配指数 (route mismatch fraction, 简称 RMF), 是对从正确路径错误添加和错误减去的不正确道路的长度求和, 然后除以正确路径的长度来计算不正确路径所占比例, 也就是结果记录的误差值, 如图 12 所示. RMF 指标由匹配正确和匹配错误的点数量共同决定.

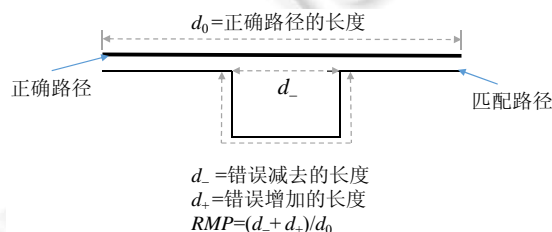


Fig.12 Evaluation norm

图 12 评价指标

OHMM 算法评估使用两个性能指标:匹配精度和输出延迟.精度定义为地面真实路径中正确匹配的轨迹点的比例 CMP,某个轨迹点映射到地面真实路径中的任何道路段就是一个正确的匹配;输出延迟是对于每个轨迹点算法所产生的平均输出延迟,它通过在获得匹配结果之前经过的秒数来量化.

QMM 算法使用不匹配率(mismatching percentage,简称 MP)作为评价标准,采样间隔为 1s~30s,不匹配率保持在 2%以内.

$$MP = \frac{\text{\#不匹配样本点数}}{\text{\#所有样本点数}} \times 100\%.$$

3.4.3 实验结果

• ST-Matching 算法

算法比较对象为基于点的先前匹配结果执行匹配的增量算法和旨在最小化平均弗雷歇距离(average Fréchet distance,简称 AFD)的全局匹配算法.合成数据的测试结果表明 ST-Matching 算法显著优于增量式和基于 AFD 的算法.同时我们注意到,当采样率降低时,增量算法的性能急剧退化,而全局算法对采样率的变化更为鲁棒.当采样间隔为 2.91min 时,ST-Matching 算法匹配质量最好, A_L 达到 95%, A_N 约为 92%;当采样间隔降低到 5.77min 时, A_L 仍有 86%, A_N 为 80%.

ST-Matching 算法在 GeoLife 收集的实际数据上的测试结果也要优于增量式和基于 AFD 的算法.随着采样率的降低,增量算法的精度急剧下降,而两种全局算法再次表现出更一致的性能.当采样间隔为 30s 时,ST-Matching 算法的 A_L 约为 87%, A_N 为 86%;当采样间隔下降到 5min 时, A_L 约为 70%, A_N 为 68%;当采样间隔为 2min 时,可以维持在 80%的准确率.

为了评估算法的效率,将算法的运行时间与基于 AFD 的全局方法进行比较.当采样点数目较少时,两种方法表现出相似的运行时间;然而随着采样点的增加,基于 ADF 的方法的运行时间急剧增加.

实验结果表明:ST-Matching 算法在低采样轨迹的匹配精度方面显著优于增量算法;而且,与基于 AFD 的全局匹配算法相比,ST-Matching 还提高了匹配精度和运行时间.

• IVMM 算法

IVMM 算法在匹配质量和效率方面与 ST-Matching 算法进行比较.为了评估算法的效率,实验比较了运行时间,结果表明:对于低采样率和高采样率数据,IVMM 算法与 ST-Matching 算法一样快,IVMM 甚至更好一些;而且随着采样间隔的增加,要匹配的采样点数量相应地减少,使得两种算法的时间成本都快速降低.

匹配结果显示,IVMM 算法在所有采样间隔中都显著优于 ST-Matching 算法.回顾北京实际出租车 GPS 数据的统计结果(如图 6 所示),大多数低采样率 GSP 轨迹的采样间隔为 2min~6min(在所有低采样率数据中约为 86%).当采样间隔为 1.5min~6.5min 时,IVMM 的精度维持在 70%左右,比 ST-Matching 算法有 10%的改进.当采样间隔超过 6.5min 时,IVMM 和 ST-Matching 算法的性能变得接近,但仍具有 5%的优势.验证了 IVMM 算法代表采样点之间的有效交互影响,比 ST-Matching 算法具有更好的鲁棒性,在低采样率的情况下匹配效果好.

• HMMM 算法

实验结果表明:当采样间隔在 10s 以内时,几乎不会产生错误;当采样间隔增长到 30s 时,误差只有 0.11%.随着采样率间隔的不断增加,错误率上升,采样间隔 2min 的误差为 10%;当采样间隔达 5min 以上时,误差超过 50%.

• Simplified HMM 算法

从实验结果可以看出,Simplified HMM 算法与 HMMM 性能相似.当采样间隔为 90s 时,RMF 为 0.2;当采样间隔上升至 180s 时,RMF 为 0.6;当采样增加至 480s 时,RMF 为 1.1,然后出现少许下降.采样间隔增加,采样点的数量也随之降低,算法的正确率逐渐下降,RMF 逐渐上升.与 HMMM 算法相比,RMF 略大,但效率较高,约为前者的 2 倍.

Simplified HMM 算法在采样率较低的情况下,准确率较高,实现较为简单,而且匹配速度较快.这是因为 Simplified HMM 算法中没有搜索候选路段的过程,所以不需要计算 GPS 点到道路上的投影,而且状态转移概率公式较为简单.

- OHMM 算法

除了大于 4min 的采样间隔外,农村路线的准确性比城市路线的准确度高出大约 5%.采样间隔不超过 1min 时,农村路线和城市路线的精度均在 0.9 之上.在这两种情况下,精度都随着采样间隔的增加而下降.研究结果表明:当操作实时输入流时,OHMM 算法以较低的输出延迟实现了最优或接近最优的匹配结果.

- QMM 算法

应用 ACM SIGSPATIAL Cup 竞赛数据测试,不匹配率保持在 2%以内,波动不大.利用美国华盛顿州道路网的 1 283 539 个路段来测试多线程技术的应用.假设把使用一个线程进行匹配需要的时间设置为 100%,则 2 个线程匹配时间为 65%,4 个线程匹配时间为 43%.而当使用 4 个线程时,程序仅需要 0.37s 来构建网格索引.结果表明:对原始隐马尔可夫模型映射匹配算法的修改,纠正了一些特殊的不匹配情况.当多核 CPU 的资源被充分利用时,索引构建时间和路网匹配时间都大为减少.

3.5 算法参数设置

3.5.1 窗 口

大多数现有的增量/在线算法、轨迹具有太多点的全局算法,往往使用滑动窗口策略.滑动窗口方法简单地将轨迹划分为固定大小的输入序列,并且独立地处理它们.

文献[63]中的算法使用固定滑动窗口(fixed sliding window,简称 FSW),通过在轨迹的滑动窗上构造局部候选图来实现全局路网匹配.每个局部候选图由轨迹的子集构成,找到这个子集的最佳匹配路径序列,然后滑动窗口,重复该过程.将轨迹分割为窗口可以减少平均延迟并节省在线处理的存储空间,但不一定加快总体处理时间,因为算法中花费最大的部分是最短路径计算.为了降低计算复杂度,算法试探性地保留具有顶部 l 个分数的候选点,从而减少下一个采样点需要计算的最短路径数量.当 $l=1$ 时,算法退化成增量算法.从真实轨迹上的实验结果可以看出:只要选择适当的窗口大小,基于滑动窗口的全局匹配算法就可以实际使用.随着窗口尺寸的增大,算法可以得到更好的匹配,但输出延迟增加.而且保留顶部节点可以显著降低匹配准确度,特别是当窗口较大时.滑动窗口方法虽然易于实现,但可能导致次优解决方案和较长的输出延迟.

文献[71]中的算法使用可变滑动窗口(variable sliding window,简称 VSW)将输入分成更小的子问题,并证明算法得到的是全局最优解.可以针对实时输入流很快产生输出.当在由窗口覆盖的马尔可夫链中的任何地方发现收敛点时,窗口随着新的轨迹点被处理而从后面收缩,向前扩展.注意,滑动窗口的大小可以根据状态空间的结构而变化.在概念上类似于在线维特比算法^[91,92].然而,VSW 的一个缺点是不能保证最坏情况的窗口大小,因此在极端情况下,输出延迟可以是任意大的.通过设置窗口大小的上限来修改 VSW,使得当达到阈值时,算法将输出,直到当前阶段的最大似然解,即有界滑动窗(bounded variable sliding window,简称 BVSF)方法.但不同于 VSW,这种方法可能导致次优解决方案.

3.5.2 HMM 中断

基于隐马尔可夫模型路网匹配算法在实际应用中可能会出现中断的情况,从一个时间点到下一个时间点的所有转移概率为 0.导致中断的原因如下.

- 路线本地化.在选择某个 GPS 点的候选路段时,为了避免不合理的计算量,一般会忽略距离 GPS 点超过一定距离的路段,而不将它们添加到 HMM 里面,这可能会导致在特定时间中没有匹配候选.当车辆在未标记在地图上的道路上行驶时,或者经历特别高的噪声时(比如当车辆进入隧道或城市峡谷时),可能出现这种情况.
- 低概率路线.当路线变得迂回时,对于该路线上的任何两个采样点之间的最短路径距离可能远大于它们之间的欧式距离.在这种情况下,对应于该路线的转移概率将变得非常小,这将触发 HMM 中断.
- GPS 离群值.作为 HMM 状态之间转换计算的任何路线都要做合理性测试:如果计算出的路线需要车辆以超过 50m/s 的速度,或者以超过 3 倍限制的速度行驶,就认为路线是不合理的,并将其概率设置为 0.

由于上述的简化限制了将被考虑的匹配候选路线的数量,可能会找不到通过 HMM 的完整路径,这就是

HMM 中断.HMM 中断发生的频率越高,匹配算法的效率就越低.

文献[56]中的算法通过以下方式处理 HMM 中断:当在时间 t 和 $t+1$ 之间检测到中断时,从模型中移除测量点 z_t 和 z_{t+1} ,并检查中断是否已经愈合.如果 $t-1$ 和 $t+2$ 的测量点在 HMM 中重新连接,则中断愈合;否则,继续去除中断任一侧上的点,直到中断愈合或中断超过 180s.如果中断超过 180s,就将数据拆分为前后两个单独的行程,并分别对每个行程进行路网匹配.文献[92]中的算法使用的方法是适当地扩大参数 β ,可以有效地减少中断情况,对匹配精度具有不可估量的影响.同时,通过实验,给出不同采样周期内 β 的最优值.

4 研究挑战与趋势

GPS 轨迹进行路网匹配是轨迹挖掘的基础,而 GPS 数据规模和数据质量都对路网匹配算法提出了挑战.下一步考虑在这两个方面进行的研究工作.

4.1 使用大数据技术提高算法性能

配有 GPS 设备的车辆会产生海量数据.比如北京市具有约 6 万余辆出租车,一天的时间就能产生上亿条 GPS 数据,更不用提产生更多数据的车牌识别、交通视频监控等.当前,与交通有关的数据跃升到 PB 级别.这种情况下,再使用传统交通流方法来对数据进行处理与分析,效果不甚理想.

另外,交通轨迹数据在空间和时间上呈现某种特性,如果能够充分利用时空特点,可以关联更多的异构数据,从而实现更加准确的分析,但是相应地也增大了数据分析的维度,对数据查询和处理提出了更高的要求.尤其在轨迹数据规模较大的情况下,计算效率变得更加重要.

日趋成熟的 Hadoop,Spark 等分布式计算平台能够高效地支持大规模数据的分析与建模问题.合理的分布式策略可以线性地提高交通计算的速度.如,Map-Reduce 就是一个性能很高的批处理分布式计算框架.它能够并行处理海量轨迹数据,可以利用集群的计算能力来提高轨迹数据处理的效率.应海量数据处理需求发展起来的云计算等新兴综合应用,可以帮助路网匹配算法降低海量数据处理的难度、节约成本、提高处理效率.

此外,对时空数据做索引,支持快速存取,可以缩小定位范围,减少候选路段数量,从而极大地缩短了 GPS 数据匹配时间,可以更高效、更精确地确定候选匹配路段,提高路网匹配的实时性和准确性.时空数据索引依赖于空间数据库索引技术,如网格索引、R-tree、K-D Tree、Quadtree 等.

4.2 利用道路语义提高匹配准确性

GPS 设备精度和测量误差会严重影响路网匹配的准确率.可以考虑使用 GPS 信息对车辆所处的道路情况进行预判,从而根据不同的道路场景选择相应的路网匹配算法来提高路网匹配的实时性和准确性.

- 首先,可以使用机器学习和计算机视觉的方法来提取路边的街道、街景、交通标志等细节信息,从而建立 GPS 信息之外的新数据层,这样能够在很大程度上帮助路网匹配算法提高准确率.比如,路牌信息可以更好地起到指针作用,如果驾驶者在听到导航指示的时候能够匹配所看到的街景,比如车辆所在街道旁边的企业、单位或是商铺的招牌等信息也可以帮助车辆更快、更准地定位;路边的停止标记、地上的车道或者转弯限制等交通标志信息虽然很难捕捉,但是可以帮助提高匹配的效率和质量.
- 其次,智能手机传感器能感受到不同的道路特性^[77],如减速道或隧道.例如,驶过减速带的车辆将会小范围地上下移动,这将引起手机测量的重力加速度发生变化;U 形转弯的车辆将产生大约 180°的方向转变.因此,可以利用手机传感器的这种效应来监测道路条件的变化.文献[99–101]使用加速度计来检测坑洞和交通条件.使用智能手机的传感器实时识别大范围的道路语义,可以自动丰富当前的数字地图,使基于这些道路语义丰富的数字地图上的路网匹配变得更加精确.
- 最后,普通用户也可以参与到标记道路语义信息的工作中来,比如获取公园、步行道以及其他街景车无法进入的地方的信息,完善更多精确路线.随着信息不断被开发、路网匹配精度不断提高,最终呈现出来的将是更为丰富、灵活的交通生活地图.

5 结 论

本文详细调研了国内外轨迹路网匹配技术,包括全局算法和局部算法.很多算法只适合于高采样率的轨迹,不适合于低采样率的轨迹,而基于隐马尔可夫的算法成为主流,因此,本文详细综述了基于隐马尔可夫模型的路网匹配算法,分析了各种不同算法的优势和不足,最后给出了研究挑战和趋势.

References:

- [1] Li X, Han J, Lee J, Gonzalez H. Traffic densitybased discovery of hot routes in road networks. In: Proc. of the 10th Int'l Symp. on Spatial and Temporal Databases, Vol.4605. 2007. 441–459. [doi: 10.1007/978-3-540-73540-3_25]
- [2] Yuan YF, Van LH, Van WF, Hoogendoorn S. Network-Wide traffic state estimation using loop detector and floating car data. Journal of Intelligent Transportation Systems, 2014,18(1):41–50. [doi: 10.1080/15472450.2013.773225]
- [3] Yuan NJ, Zheng Y, Xie X, Wang YZ, Zheng K, Xiong H. Discovering urban functional zones using latent activity trajectories. IEEE Trans. on Knowledge and Data, 2015,27(3):712–725. [doi: 10.1109/TKDE.2014.2345405]
- [4] Zheng, Y, Wang L, Xie X, Ma W. GeoLife: Managing and understanding your past life over maps. In: Proc. of the 9th Int'l Conf. on Mobile Data Management. 2008. 211–212. [doi: 10.1109/MDM.2008.20]
- [5] Pfoer D, Jensen CS. Capturing the uncertainty of moving-object representations. Lecture Notes in Computer Science, 1999,1651: 111–131. [doi: 10.1007/3-540-48482-5_9]
- [6] Joshi RR. A new approach to map matching for in-vehicle navigation systems: The rotational variation metric. In: Proc. of the ITS. 2001. 33–38. [doi: 10.1109/ITSC.2001.948625]
- [7] Feijoo C, Ramos J, Perez F. A system for fleet management using differential GPS and VHF data transmission mobile networks. In: Proc. of the VNIS. 1993. 445–448. [doi: 10.1109/VNIS.1993.585667]
- [8] Chen M, Liu Y, Yu X. Predicting next locations with object clustering and trajectory clustering. In: Proc. of the PAKDD. 2015. 344–356. [doi: 10.1007/978-3-319-18032-8_27]
- [9] Alvarez-Garcia JA, Ortega JA, Gonzalez-Abril L, Velasco F. Trip destination prediction based on past GPS log using a hidden Markov model. Expert Systems with Applications, 2010,37(12):8166–8171. [doi: 10.1016/j.eswa.2010.05.070]
- [10] Mori U, Mendiburu A, Álvarez M, Lozano JA. A review of travel time estimation and forecasting for advanced traveler information systems. Transportmetrica A: Transport Science, 2015,11(2):119–157. [doi: 10.1080/23249935.2014.932469]
- [11] Krumm J. A Markov model for driver turn prediction. SAE World Congress, 2008,22(1):1–25.
- [12] Kuehne RRP, Schiifert J, Mikat KV, Tbiessenbusent V, Lorkowski BS. New approaches for traffic management in metropolitan areas. In: Proc. of the 10th IFAC Symp. on Control in Transportation Systems. 2003. 209–214.
- [13] Lü L, Chen M, Liu Y, Yu XH. A plane moving average algorithm for short-term traffic flow prediction. In: Proc. of the Advances in Knowledge Discovery and Data Mining. 2015. 357–369.
- [14] Pang LX, Chawla S, Liu W, Zheng Y. On detection of emerging anomalous traffic patterns using GPS data. Data and Knowledge Engineering, 2013,87(9):357–373. [doi: 10.1016/j.datak.2013.05.002]
- [15] Long X, Jin L, Joshi J. Exploring trajectory-driven local geographic topics in foursquare. In: Proc. of the Int'l Workshop on Location-based Social Networks. 2012. 927–934. [doi: 10.1145/2370216.2370423]
- [16] Yuan Q, Cong G, Ma Z, Sun A, Magnenat-Thalmann N. Who, where, when and what: Discover spatio-temporal topics for Twitter users. In: Proc. of the SIGKDD. 2013. 605–613. [doi: 10.1145/2487575.2487576]
- [17] Gonzalez H, Han J, Li X, Myslinska M, Sondag JP. Adaptive fastest path computation on a road network: A traffic mining approach. In: Proc. of the VLDB. 2007. 794–805.
- [18] Wei LY, Zheng Y, Peng WC. Constructing popular routes from uncertain trajectories. In: Proc. of the SIGKDD. 2012. 195–203.
- [19] Yuan J, Zheng Y, Xie X, Sun GZ. T-Drive: Enhancing driving directions with taxi drivers' intelligence. IEEE Trans. on Knowledge and Data Engineering, 2013,26(1):220–232. [doi: 10.1109/TKDE.2011.200]
- [20] Patterson DJ, Liao L, Fox D, Kautzy H. Inferring high-level behavior from low-level sensors. In: Proc. of the UbiComp. Springer-Verlag, 2003. 73–89. [doi: 10.1007/978-3-540-39653-6_6]

- [21] Rauschert I, Agrawal P, Sharma R, Fuhrmann S, Brewer I, MacEachren A, Wang HM, Cai G. Designing a human-centered, multimodal GIS interface to support emergency management. In: Proc. of the Int'l Symp. on Advances in Geographic Information Systems. 2003. 119–124. [doi: 10.1145/585147.585172]
- [22] Li M, Zhang Y, Wang W. Analysis of congestion points based on probe car data. In: Proc. of the ITS. 2009. 1–5. [doi: 10.1109/ITSC.2009.5309869]
- [23] Mohammed AQ, Robert BN, Washington YO. A high accuracy fuzzy logic based map matching algorithm for road transport. Journal of Intelligent Transportation Systems, 2006,10(3):103–115. [doi: 10.1080/15472450600793560]
- [24] Yuan NJ, Zheng Y, Xie X, Wang YZ, Zheng K, Xiong H. Discovering urban functional zones using latent activity trajectories. IEEE Trans. on Knowledge and Data Engineering, 2015,27(3):712–725. [doi: 10.1109/TKDE.2014.2345405]
- [25] Shang J, Zheng Y, Tong W, Chang E, Yu Y. Inferring GAS consumption and pollution emission of vehicles throughout a city. In: Proc. of the SIGKDD. 2014. 1027–1036. [doi: 10.1145/2623330.2623653]
- [26] Bernstein D, Kornhauser A. An introduction to map matching for personal navigation assistants. Geometric Distributions, 1996, 122(7):1082–1083.
- [27] White CE, Bernstein D, Kornhauser AL. Some map matching algorithms for personal navigation assistants. Transportation Research Part C: Emerging Technologies, 2000,8(1):91–108. [doi: 10.1016/S0968-090X(00)00026-7]
- [28] Taylor G, Blewitt G, Steup D, Corbett S, Car A. Road reduction filtering for GPS-GIS navigation. Trans. in GIS, 2001,5(3): 193–207. [doi: 10.1111/1467-9671.00077]
- [29] Alt H, Efrat A, Rote G, Wenk C. Matching planar maps. Journal of Algorithms, 2003,49(2):262–283. [doi: 10.1016/S0196-6774(03)00085-3]
- [30] Quddus MA, Ochieng WY, Zhao L, Noland RB. A general map matching algorithm for transport telematics applications. GPS Solutions, 2003,7(3):157–167. [doi: 10.1007/s10291-003-0069-z]
- [31] Abbour M, Bonnifait P, Cherfaoui V. Map matching integrity using multi-sensor fusion and multi-hypothesis road tracking. Journal of Intelligent Transportation Systems Technology Planning and Operations, 2008,6(4):189–201.
- [32] Nadine S, Kay WA. Map matching of GPS traces on high-resolution navigation networks using the multiple hypothesis technique (MHT). Working Paper Transport and Spatial Planning, 2009,(10):568–588.
- [33] Ochieng WY, Quddus M, Noland RB. Map-Matching in complex urban road networks. Revista Brasileira De Cartografia, 2003, 55(2):1–14.
- [34] Quddus MA, Noland RB, Ochieng WY. Validation of map matching algorithm using high precision positioning with GPS. Journal of Navigation, 2004,58(2):257–271. [doi: 10.1017/S0373463305003231]
- [35] Kim S, Kim J. Adaptive fuzzy-network based C-measure map-matching algorithm for car navigation system. IEEE Trans. on Industrial Electronics, 2001,48(2):432–440. [doi: 10.1109/41.915423]
- [36] Syed S, Cannon ME. Fuzzy logic-based map-matching algorithm for vehicle navigation system in urban canyons. In: Proc. of the Ion National Technical Meeting. 2004. 982–993.
- [37] Su H, Chen J, Xu J. A adaptive map matching algorithm based on fuzzy-neural-network for vehicle navigation system. In: Proc. of the WCICA. 2008. 4448–4452. [doi: 10.1109/WCICA.2008.4593639]
- [38] Haibin S, Jiansheng T, Chaozhen H. A integrated map matching algorithm based on fuzzy theory for vehicle navigation system. In: Proc. of the ICCIAS. 2006. 916–919. [doi: 10.1109/ICCIAS.2006.294272]
- [39] Zhang L, Liu JW, Wang RC, Wang HY. Trust evaluation model based on improved D-S evidence theory. Journal on Communications, 2013,34(7):167–173 (in Chinese with English abstract).
- [40] El Najjar ME, Bonnifait P. A road-matching method for precise vehicle localization using belief theory and kalman filtering. Autonomous Robots, 2005,19(2):173–191. [doi: 10.1007/s10514-005-0609-1]
- [41] Nassreddine G, Abdallah F, Denoeux T. Map matching algorithm using belief function theory. In: Proc. of the IF. 2008. 1–8.
- [42] Yang D, Cai B, Yuan Y. An improved map-matching algorithm used in vehicle navigation system. Intelligent Transportation Systems, 2003,11(2):1246–1250. [doi: 10.1109/ITSC.2003.1252683]
- [43] Obradovic D, Lenz H, Schupfner M. Fusion of map and sensor data in a modern car navigation system. Journal of Signal Processing Systems, 2006,45(1):111–122. [doi: 10.1007/s11265-006-9775-4]

- [44] Xu H, Liu HC, Tan CW, Bao Y. Development and application of an enhanced kalman filter and global positioning system error-correction approach for improved map matching. *Journal of Intelligent Transportation Systems: Technology, Planning and Operations*, 2010,14(1):27–36. [doi: 10.1080/15472450903386013]
- [45] Kim W, Jee GI, Lee JG. Efficient use of digital road map in various positioning for ITS. In: *Proc. of the Position Location and Navigation Symp.* 2000. 170–176. [doi: 10.1109/PLANS.2000.838299]
- [46] Li Z, Chen W. A new approach to map-matching and parameter correcting for vehicle navigation system in the area of shadow of GPS signal. In: *Proc. of the ITSC*, Vol.5. 2005. 425–430. [doi: 10.1109/ITSC.2005.1520086]
- [47] Pyo JS, Shin DH, Sung TK. Development of a map matching method using the multiple hypothesis technique. In: *Proc. of the Intelligent Transportation Systems*. 2001. 23–27. [doi: 10.1109/ITSC.2001.948623]
- [48] Cherif S, El-Najjar ME, François C. A road matching method for precise vehicle localization using hybrid Bayesian network. *Intelligent Transportation Systems*, 2008,12(4):176–188. [doi: 10.1080/15472450802448153]
- [49] Yin H, Wolfson O. A weight-based map matching method in moving objects databases. In: *Proc. of the SSDBM*, Vol.16. 2004. 437–438. [doi: 10.1109/SSDM.2004.1311248]
- [50] Blazquez C, Vonderohe A. Simple map-matching algorithm applied to intelligent winter maintenance vehicle data. *Journal of the Transportation Research Board*, 2005,1935(1):68–76. [doi: 10.3141/1935-08]
- [51] Brakatsoulas S, Pfoser D, Salas R, Wenk C. On map-matching vehicle tracking data. In: *Proc. of the VLDB*. 2005. 853–864.
- [52] Honey SK, Zavoli WB, Milnes KA, Phillips AC, Jr White MS, Jr Loughmiller GE. *Vehicle navigational system and method*: US. US4796191, 1989.
- [53] Pink O, Hummel B. A statistical approach to map matching using road network geometry, topology and vehicular motion constraints. In: *Proc. of the ITS*. 2008. 862–867. [doi: 10.1109/ITSC.2008.4732697]
- [54] Bierlaire M, Chen J, Newman J. A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, 2013,26(1):78–98. [doi: 10.1016/j.trc.2012.08.001]
- [55] Zhang ZH, Cui TJ, Yao HM. New map matching calculation method in vehicle navigation system. *Hydrographic Surveying and Charting*, 2006,26(2):55–58 (in Chinese with English abstract).
- [56] Newson P, Krumm J. Hidden Markov map matching through noise and sparseness. In: *Proc. of the ACM-GIS*. 2009. 336–343. [doi: 10.1145/1653771.1653818]
- [57] Greenfield JS. Matching GPS observations to locations on a digital map. In: *Proc. of the TRB*. 2002. <https://trid.trb.org/Results?q=&serial=%22Transportation%20Research%20Board%2081st%20Annual%20Meeting%22>
- [58] Wenk C, Salas R, Pfoser D. Addressing the need for map-matching speed: Localizing global curve-matching algorithms. In: *Proc. of the SSDBM*. 2006. 379–388. [doi: 10.1109/SSDBM.2006.11]
- [59] Chawathe SS. Segment-Based map matching. In: *Proc. of the Intelligent Vehicles Symp.* 2007. 1190–1197. [doi: 10.1109/IVS.2007.4290280]
- [60] Lou Y, Zhang C, Zheng Y, Wang W, Huang Y. Map-Matching for low-sampling-rate GPS trajectories. In: *Proc. of the ACM-GIS*. 2009. 352–361.
- [61] Yuan J, Zheng Y, Zhang C, Xie X, Sun JZ. An interactive-voting based map matching algorithm. In: *Proc. of the MDM*. 2010. 43–52. [doi: 10.1109/MDM.2010.14]
- [62] Giovannini L. A novel map-matching procedure for low-sampling GPS data with applications to traffic flow analysis [Ph.D. Thesis]. Università Di Bologna, 2011.
- [63] Zheng K, Zheng Y, Xie X, Zhou XF. Reducing uncertainty of low sampling rate trajectories. In: *Proc. of the ICDE*. 2012. 1144–1155.
- [64] Civilis A, Jensen CS, Pakalnis S. Techniques for efficient road-network based tracking of moving objects. *IEEE Trans. on Knowledge and Data Engineering*, 2005,17(5):698–712. [doi: 10.1109/TKDE.2005.80]
- [65] Thiagarajan A, Ravindranath L, Lacurts K, Madden S, Balakrishnan H, Toledo S, Eriksson J. VTrack: Accurate, energy-aware road traffic delay estimation using mobile phones. In: *Proc. of the Sensys*. 2009. 85–98.
- [66] Thiagarajan A, Ravindranath L, Balakrishnan H, Madden S, Girod L. Accurate, low-energy trajectory mapping for mobile devices. In: *Proc. of the USENIX*. 2011. 267–280.

- [67] Wang H, Wang Z, Shen G, Li F, Han S, Zhao F. WheelLoc: Enabling continuous location service on mobile phone for outdoor scenarios. In: Proc. of the INFOCOM. 2013. 2733–2741. [doi: 10.1109/INFOCOM.2013.6567082]
- [68] Guha S, Plarre K, Lissner D, Mitra S, Krishna B, Dutta P, Kumar S. AutoWitness: Locating and tracking stolen property while tolerating GPS and radio outages. In: Proc. of the TOSN. 2010. 29–42. [doi: 10.1145/1869983.1869988]
- [69] Aly H, Youssef M. semMatch: Road semantics-based accurate map matching for challenging positioning data. In: Proc. of the 23rd SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems. 2015. Article No.5. [doi: 10.1145/2820783.2820824]
- [70] Mohamed R, Aly H, Youssef M. Accurate and efficient map matching for challenging environments. In: Proc. of the SIGSPATIAL. 2014. 401–404. [doi: 10.1145/2666310.2666429]
- [71] Mohamed R, Aly H, Youssef M. Accurate real-time map matching for challenging environments. IEEE Trans. on Intelligent Transportation Systems, 2017,18(4):847–857. [doi: 10.1109/TITS.2016.2591958]
- [72] Aly H, Youssef M. Dejavu: An accurate energy-efficient outdoor localization system. In: Proc. of the ACM-GIS. 2013. 154–163. [doi: 10.1145/2525314.2525338]
- [73] Alzantot M, Youssef M. UPTIME: Ubiquitous pedestrian tracking using mobile phones. In: Proc. of the WCNC. 2012. 3204–3209. [doi: 10.1109/WCNC.2012.6214359]
- [74] Ibrahim M, Youssef M. CellSense: A probabilistic RSSI-based GSM positioning system. In: Proc. of the GLOBECOM. 2010. 1–5. [doi: 10.1109/GLOCOM.2010.5683779]
- [75] Ibrahim M, Youssef M. A hidden Markov model for localization using low-end GSM cell phones. Proc. of the ICC, 2011,47(10): 1–5. [doi: 10.1109/icc.2011.5962993]
- [76] Ibrahim M, Youssef M. CellSense: An accurate energy-efficient GSM positioning system. IEEE Trans. on Vehicular Technology, 2011,61(1):286–296. [doi: 10.1109/TVT.2011.2173771]
- [77] Aly H, Basalamah A, Youssef M. Map++: A crowd-sensing system for automatic map semantics identification. In: Proc. of the SECON. 2014. 546–554. [doi: 10.1109/SAHCN.2014.6990394]
- [78] Goh CY, Dauwels J, Mitrovic N, Asif MT, Oran A, Jaillet P. Online map-matching based on hidden Markov model for real-time traffic sensing applications. In: Proc. of the ITSC. 2012. 776–781. [doi: 10.1109/ITSC.2012.6338627]
- [79] Rohani M, Gingras D, Gruyer D. A novel approach for improved vehicular positioning using cooperative map matching and dynamic base station DGPS concept. IEEE Trans. on Intelligent Transportation Systems, 2015,17(1):1–10. [doi: 10.1109/TITS.2015.2465141]
- [80] Pereira FC, Costa H, Pereira NM. An off-line map-matching algorithm for incomplete map databases. European Transport Research Review, 2009,1(3):107–124. [doi: 10.1007/s12544-009-0013-6]
- [81] Miwa T, Kiuchi D, Yamamoto T, Morikawa T. Development of map matching algorithm for low frequency probe data. Transportation Research Part C: Emerging Technologies, 2012,22(5):132–145. [doi: 10.1016/j.trc.2012.01.005]
- [82] Raymond R, Morimura T, Osogami T, Hirose N. Map matching with hidden Markov model on sampled road network. In: Proc. of the ICPR. 2012. 2242–2245.
- [83] Baum LE. Statistical inference for probabilistic functions of finite state Markov chains. Annals of Mathematical Statistics, 1966, 37(6):1554–1563. [doi: 10.1214/aoms/1177699147]
- [84] Baum LE, Eagon JA. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. Bulletin of the American Mathematical Society, 1967,73(3):360–363. [doi: 10.1090/S0002-9904-1967-11751-8]
- [85] Baum LE, Sell GR. Growth transformation for functions on manifolds. Pacific Journal of Mathematics, 1968,27(2):211–227. [doi: 10.2140/pjm.1968.27.211]
- [86] Lamb P, Thiebaux S. Avoiding explicit map-matching in vehicle location. In: Proc. of the ITS. 1999. 8–12.
- [87] Hummel B. Map matching for vehicle guidance. In: Dynamic and Mobile GIS: Investigating Changes in Space and Time. Boca Raton: Taylor & Francis, 2006.
- [88] Krumm J, Letchner J, Horvitz E. Map matching with travel time constraints. SAE Technical Paper, 2007-01-1102, 2007. [doi: 10.4271/2007-01-1102]

- [89] Gather U, Schultze V. Robust estimation of scale of an exponential distribution. *Statistica Neerlandica*, 1999,53(53): 327–341. [doi: 10.1111/1467-9574.00115]
- [90] Wang Y, Zhu Y, He Z, Yue Y, Li Q. Challenges and opportunities in exploiting large-scale GPS probe data. HP Laboratories, Technical Report, HPL-2011-109, 2011.
- [91] Bloit J, Rodet X. Short-Time viterbi for online HMM decoding: Evaluation on a real-time phone recognition task. In: *Proc. of the ICASSP*. 2008. 2121–2124. [doi: 10.1109/ICASSP.2008.4518061]
- [92] Šrámek R, Brejová B, Vinař T. On-Line viterbi algorithm and its relationship to random walks. arXiv:0704.0062v1, 2007.
- [93] Song R, Lu W, Sun W, Huang W, Chen C. Quick map matching using multi-core CPUs. In: *Proc. of the ACM-GIS*. 2012. 605–608. [doi: 10.1145/2424321.2424428]
- [94] Wu G, Qiu YJ, Wang GR. Map matching algorithm based on hidden Markov model and genetic algorithm. *Journal of Northeastern University (Natural Science)*, 2017,38(4):472–475 (in Chinese with English abstract).
- [95] George RJ, Thambipillai S. Online map-matching of noisy and sparse location data with hidden Markov and route choice models. *IEEE Trans. on Intelligent Transportation Systems*, 2017,18(9):2423–2434. [doi: 10.1109/TITS.2017.2647967]
- [96] Essam A, Tetsuji O, Ahmed E. Real-Time large-scale map matching using mobile phone data. *ACM Trans. on Knowledge Discovery from Data*, 2017,11(4):1–38. [doi: 10.1145/3046945]
- [97] Zheng Y, Chen Y, Xie X, Ma W. GeoLife2.0: A location-based social networking service. In: *Proc. of the MDM*. 2009. 357–358.
- [98] Zheng Y, Liu L, Wang L, Xie X. Learning transportation mode from raw GPS data for geographic applications on the Web. In: *Proc. of the WWW*. 2008. 247–256.
- [99] Eriksson J, Girod L, Hull B, Newton Y, Madden S, Balakrishnan H. The pothole patrol: Using a mobile sensor network for road surface monitoring. In: *Proc. of the MobiSys*. 2008. 29–39. [doi: 10.1145/1378600.1378605]
- [100] Mednis A, Strazdiņš G, Zviedris R, Kanonirs G, Selavo L. Real time pothole detection using Android smartphones with accelerometers. In: *Proc. of the DCOSS*. 2011. 1–6. [doi: 10.1109/DCOSS.2011.5982206]
- [101] Mohan P, Padmanabhan VN, Ramjee R. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In: *Proc. of the SenSys*. 2008. 323–336. [doi: 10.1145/1460412.1460444]

附中文参考文献:

- [39] 张琳,刘婧文,王汝传,王海艳.基于改进 D-S 证据理论的信任评估模型.通信学报,2013,34(7):167–173.
- [55] 张振辉,崔铁军,姚慧敏.车辆导航系统中地图匹配新算法.海洋测绘,2006,26(2):55–58.
- [94] 吴刚,邱煜晶,王国仁.基于隐马尔可夫模型和遗传算法的地图匹配算法.东北大学学报(自然科学版),2017,38(4):472–475.



高文超(1986—),女,山东泰安人,博士,讲师, CCF 专业会员,主要研究领域为时空大数据,计算机辅助设计.



塔娜(1982—),女,博士,讲师,主要研究领域为城市计算,计算传播学.



李国良(1980—),男,博士,教授,CCF 高级会员,主要研究领域为大数据挖掘与分析,群智计算,时空大数据分析.