

· 复杂性科学 ·

面向复杂城市道路网络的GPS轨迹匹配算法

刘 张^{1,2,3}, 王心迪^{3,4}, 闫小勇^{3,5}

(1. 电子科技大学能源科学与工程学院 成都 611731; 2. 成都师范学院物理与工程技术学院 成都 611130;
3. 电子科技大学大数据研究中心 成都 611731; 4. 电子科技大学英才实验学院 成都 611731;
5. 北京交通大学交通运输学院 北京 海淀区 100044)

【摘要】 车辆GPS轨迹的地图匹配是交通大数据挖掘中的一项重要的基础性工作, 可靠的轨迹匹配结果对于道路交通运行状态监测、实时交通信息发布、车辆定位与智能调度、出行路径选择行为分析等具有重要意义。由于城市道路网络中大量存在高架路、主辅路和立体交叉等复杂的道路场景, 传统的地图匹配算法在这些场景下难以对车辆轨迹进行准确匹配。针对这一问题, 该文提出一种基于道路网络拓扑结构的轨迹匹配算法, 将轨迹匹配问题转换为在加权道路网络中寻找最优路径的问题。利用成都市道路网络中上万辆出租车的实际运行轨迹数据对本文算法进行了验证, 结果表明在复杂的城市道路网络中应用该算法能够获得较高的匹配成功率和准确率。

关键词 GPS轨迹; 地图匹配; 道路网络; 出租车

中图分类号 TP301.6 文献标志码 A doi:10.3969/j.issn.1001-0548.2016.06.023

Map-Matching Algorithm for GPS Trajectories in Complex Urban Road Networks

LIU Zhang^{1,2,3}, WANG Xin-di^{3,4}, and YAN Xiao-yong^{3,5}

(1. School of Energy Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731;
2. College of Physics and Engineering, Chengdu Normal University Chengdu 611130;
3. Big Data Research Center, University of Electronic Science and Technology of China Chengdu 611731;
4. Yingcai Honors College, University of Electronic Science and Technology of China Chengdu 611731;
5. School of Traffic and Transportation, Beijing Jiaotong University Haidian Beijing 100044)

Abstract Map-matching for GPS trajectories is a key groundwork in mining transportation data. Reliable matching results are significant for monitoring traffic situation, publishing real-time transportation information, vehicle tracking, smart vehicle dispatching, and routing behavior analysis. In real urban road networks, there are numerous complicated road structures such as elevated roads, frontage roads, and interchange bridges. Traditional map-matching algorithms could not match trajectories on these structures accurately. In this paper, we propose a map-matching algorithm based on the topological structure of the road networks and transform the problem of matching GPS trajectories in road map into the problem of finding the shortest path in a weighted road network. We test the algorithm with the real data of GPS trajectories of tens of thousands of taxis in Chengdu. The results show that the presented algorithm can acquire a high success ratio and accuracy ratio in complicated urban road networks.

Key words GPS trajectories; map-matching algorithm; road networks; taxi

车辆GPS轨迹匹配是利用GPS轨迹点数据和数字地图信息, 确定车辆在数字地图上准确位置的一种定位技术。实际中由于各种因素的影响, 车载GPS设备得到的轨迹点和车辆真实位置之间总是存在误差, 这就需要利用轨迹匹配技术将获取的轨迹点数据准确匹配到数字地图上^[1]。GPS轨迹匹配技术被广泛应用于车辆定位^[2]、城市计算^[3-4]、人类移动行为

研究^[5]、交通预测及引导^[6-8]、搭乘服务^[9-10]、交通异常事件检测^[11-13]等领域, 准确的轨迹匹配结果对于为这些领域提供可靠的基础数据具有重要意义。

常见的GPS轨迹匹配算法主要有位置点匹配、轨迹相关算法、D-S证据理论算法、基于隐马尔科夫模型的算法等。位置点匹配算法^[14-15]是在多条候选路段中, 选择具有最小罚函数的路段作为匹配路段。

收稿日期: 2015-07-21; 修回日期: 2016-03-15

基金项目: 国家自然科学基金(61304177, 71671015); 四川省教育厅科研项目(15ZB0345)

作者简介: 刘张(1979-), 男, 博士生, 主要从事交通大数据挖掘方向的研究。

其算法简单, 反应速度快, 但罚函数鲁棒性差, 对城市道路中的并行路段(一条道路同时包含主路和辅路, 或高架道路和桥下道路水平位置重叠)容易造成误判。轨迹相关算法^[16]通过对比候选路段的航向增量, 选择具有最优值的路段作为合理的匹配路段。该算法能够矫正定位信息中的纵向误差分量, 但在GPS数据误差增加、道路状况变得复杂时, 地图匹配效果较差, 无法对城市道路中复杂立交等场景进行有效的处理。D-S证据理论算法^[17]以距离和航向为证据, 用信任函数计算某个证据对某条候选路段的支持程度, 选择支持度最大的候选路段作为当前路段。该算法能够处理不完备信息, 但当证据发生冲突时, 容易出错。基于隐马尔可夫模型(HMM)的地图匹配算法^[18-21]将车辆GPS轨迹点位置作为HMM中的观察变量, 将车辆实际所在位置作为HMM中的隐藏状态变量, 通过实际数据训练模型再进行真实位置的预测。尽管在相对简单的道路网络场景下HMM方法已经可以获得很高的精确度, 但对于复杂城市道路中大量存在并行路段、复杂立交等场景下该算法的适用性尚缺乏评估。

针对前述传统GPS轨迹地图匹配算法不能很好地处理复杂城市道路网络中并行路段、复杂立交等场景下的轨迹匹配这一问题, 本文提出了一种基于道路网络拓扑结构的GPS轨迹匹配算法, 将轨迹匹配问题转换为在加权的城市道路网络中寻找最优路径的问题, 并利用成都市道路网络中上万辆出租车的实际运行轨迹数据对算法进行了验证。

1 轨迹匹配算法

1.1 算法基本思路

在复杂的城市道路网络中, 对车辆的GPS轨迹进行地图匹配的难点在于并行路段和复杂立交场景的处理, 如图1所示。由于在这些场景下某些轨迹点距离多条路段的距离都非常接近, 单纯根据一个轨迹点的信息很难判定该轨迹点真正处于哪条路段上。但如果从一串轨迹点的角度观察, 则很容易发现车辆能够运行的路径是非常有限的。如在图1中, 从该图左下角的道路经过立交驶入左上角的道路, 必须要通过立交右上方的左转匝道才可能完成左转弯过程。因此, 处理并行路段和复杂立交场景下轨迹匹配的关键在于利用道路网络的拓扑结构信息, 找出轨迹首末点之间的可行路径。在有多条可行路径的前提下, 则可以通过判断轨迹点到不同路径的距离, 找到轨迹点最可能通过的路径。

以上过程可以通过在加权道路网络中搜索轨迹首末点之间的加权最短路径来实现。基本思路是:

- 1) 将路段长度作为道路网络上连边的基本权值;
- 2) 将每个轨迹点投影给周边路段, 用轨迹点到路段的距离来更新路段权值, 使得投影轨迹点越多, 距离轨迹点越近的路段获得更低的权值;
- 3) 以权值最小为目标在轨迹首末点之间搜索一条路径。该方法可以保证得到的路径在拓扑上是正确的, 同时由于用轨迹点到路段的距离更新路段权值, 可以使所得路径距离轨迹的整体距离最小, 因此最终得到的路径最有可能是轨迹点实际所处的路径。

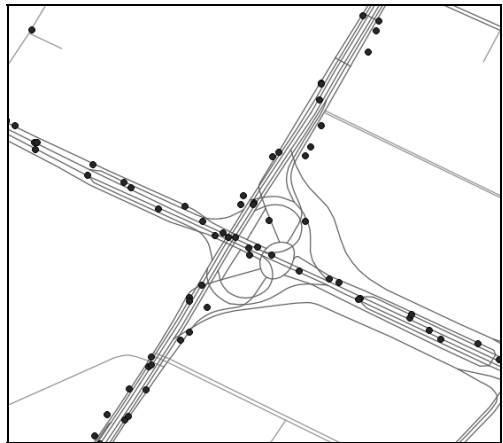


图1 并行路段和复杂立交场景下的GPS轨迹

1.2 算法步骤描述

下面给出该方法的具体实现步骤, 算法流程如图2所示。

初始化: 根据道路地图建立有向含权网络模型, 将道路路段作为网络中的有向边, 路段长度作为边的初始权值, 路段相交处作为网络节点。

- 1) 从车辆运行轨迹中截取时长为 T 的轨迹片段, 将轨迹片段中的第一个点(记为 s)连接到以 s 点为圆心、 D 为半径的圆形所覆盖的所有路段的起点上, 并将 s 点到这些路段起点的权值设为0; 同时将轨迹片段中的最后一个点(记为 t)连接到以 t 点为圆心、 D 为半径的圆形所覆盖的所有路段的终点上, 并将 t 点到这些路段终点的权值设为0。

- 2) 对于轨迹片段中除 s 、 t 点之外的每个轨迹点 o , 找出以 o 为圆心、 D 为半径的圆形范围内所覆盖的路段 l , 将 o 到 l 的垂直距离 D_o 除以 D 作为权值更新系数, 乘在 l 的权值上, 得到新的路段权值, 同时将点 o 的时间戳以及点 o 到路段 l 的距离 D_{ol} 记录在路段 l 上。

- 3) 用最短路径算法搜索 s 和 t 点之间权值最小的路径。

4) 从路径起点到终点依次检查路径中包含的每条路段，以轨迹点到路段的距离为标准去除重复的轨迹点(即对于同一个轨迹点判断它到哪条路段的距离更小就将它归属到哪条路段上)，得到轨迹点和路段的一一对应关系，该段轨迹匹配工作完成。

5) 返回步骤1)，进行下一个轨迹片段的匹配。

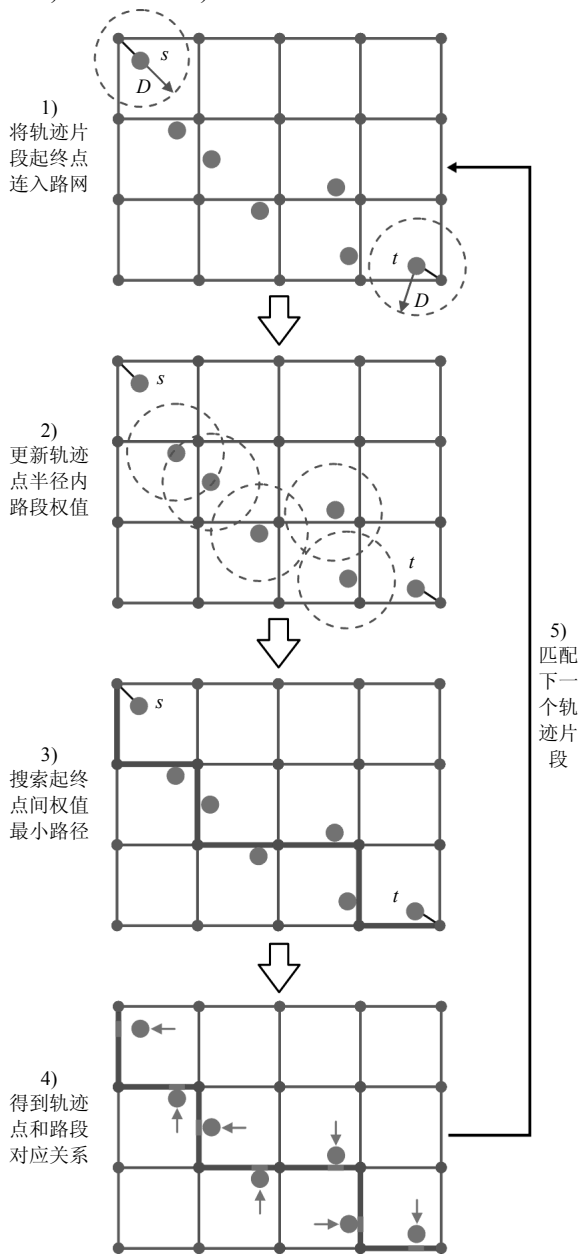


图2 算法流程示意图

1.3 算法实现要点

1) 具体参数的选取

算法中包含两个待定参数，一是轨迹片段的截取时长 T ，二是空间距离阈值 D 。对于轨迹片段的截取时长 T ，过长和过短都会影响轨迹匹配的成功率和准确率，后续算法应用实例部分会结合具体的应用

场景给出建议的 T 取值范围。

对于空间距离阈值 D ，根据我国《城市道路交通规划设计规范》^[22]规定，城市道路中等级最高的快速路宽度为40~45 m，因此本文取道路的最大宽度为45 m，再考虑到车载GPS的定位误差为±15 m^[23]，本文取 $D=60$ m。这样取值能够保证如果某时刻车辆在某条道路上行驶，那么这条道路一定处在以该时刻车辆GPS轨迹点为圆心、60 m为半径的圆形范围内。如果 D 取值小于60 m，可能有部分轨迹点无法找到匹配道路；而 D 取值大于60 m也不会增加算法精确度，只会增加空间查询操作的工作量。

2) 空间查询的效率问题

前述算法中涉及到大量空间查询操作(查找距离轨迹点 D 范围内的路段)，如果直接在全图中进行搜索，效率无疑是非常低的。因此，必须通过建立空间索引的方式，将全网中的路段按照一定的空间层次进行预排列，以提高空间查询的效率。本文采取了经典的 R 树索引结构^[24]为路段建立空间索引，可以有效提高路段空间查询的效率。

3) 路径搜索的效率问题

常用的网络最短路径算法(如Dijkstra算法、Floyd算法等)的复杂度是 $O(n^3)$ ^[25]，其中 n 是道路网络中节点的个数。一般大城市的道路网络所包含的节点往往多达数万个，在如此大规模的路网中反复进行最短路径搜索是非常耗时的。为提高路径搜索效率，本文采用缩减网络规模的办法，即只在轨迹片段周边一定范围内的路段所构成的子网中搜索路径。具体方法是：首先在轨迹片段中相邻的两个轨迹点之间连线，然后建立以连线为中心、半径为 D 的缓冲区，最后将此缓冲区与整个路网进行空间交集运算，抽取一个路网子集，如图3所示。在此子网上进行最短路径搜索，可以有效缩短搜索时间，提高整个算法的运行效率。

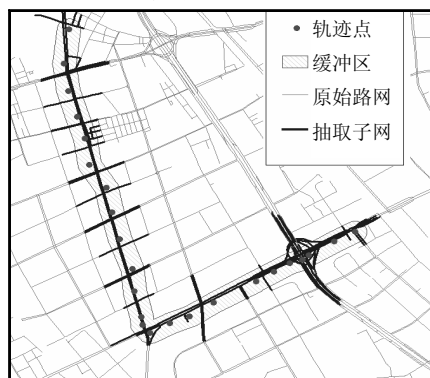


图3 根据轨迹点抽取子网示意图

2 算法应用实例

2.1 数据集描述

本文用成都市出租车车载GPS所记录的轨迹数据对算法进行测试。该数据集中包含成都市13 933辆出租车在一天之内的行驶记录, 每辆出租车的车载GPS设备以每10 s为间隔记载该车当前所处的时间、经纬度以及空重载状态, 共有数据76 071 067条。同时, 本文还使用了图4所示的成都市道路网络数据, 该数据集中记录了成都市绕城高速以内全部道路的拓扑结构和长度等信息, 网络中共包含路段30 658条, 节点21 921个。



图4 成都市道路网络地图

2.2 算法匹配结果分析

图5是本文算法的一个典型匹配结果。从图中可以看到, 不论车辆运行轨迹是通过两座复杂立交的转向匝道, 还是在两座立交之间的辅路上进行切换, 本文算法都给出了准确的匹配结果。这一结果在某种程度上验证了本文算法能够很好地处理复杂道路网络条件下的车辆轨迹匹配问题。

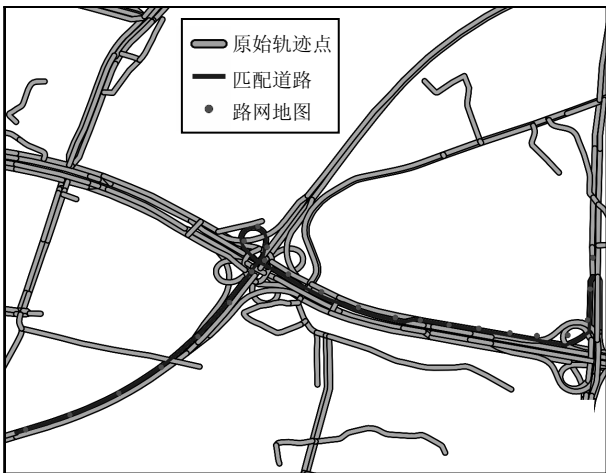


图5 算法匹配结果示意图

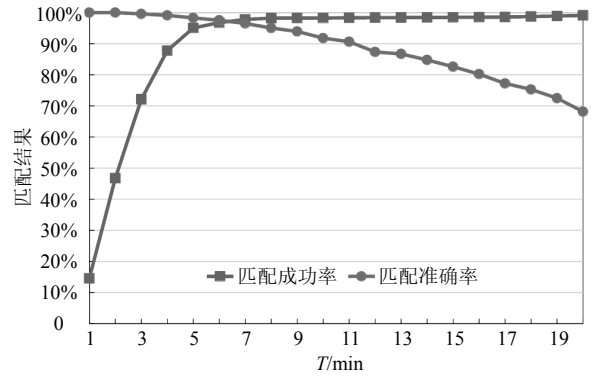
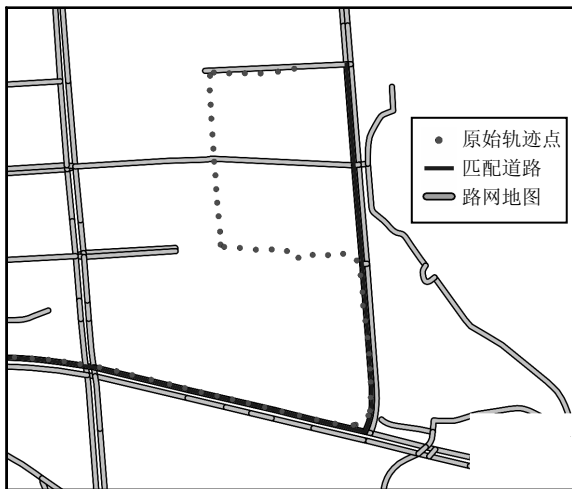


图6 轨迹长度T对匹配结果的影响

为对本文算法进行更全面的测试, 本文分析了在轨迹片段长度 T 取值不同的情况下, 算法匹配成功率和准确率的变化规律, 结果如图6所示。这里的成功率定义为: 在所有被匹配的轨迹片段中, 成功找到一条有效路径的轨迹片段所占的比例。找不到一条有效路径的情况主要是由于在 T 过小时, 有较大比例的车辆可能停驶不动或只移动了很短的距离, 使得算法无法搜索到一条权值大于0的路径。因此, 单纯从提高算法匹配成功率的角度来看, T 取值越大, 出现车辆停驶的可能性就越低, 匹配成功率就越高。但过大的 T 会降低算法匹配的准确率。本文采用以下方法定义准确率: 对于成功匹配的轨迹片段, 如果其中存在一个轨迹点距离所匹配路段的最大距离超过了60 m, 本文认为匹配结果是不可信的。这是由于在1.3节中已经提到, 综合考虑GPS定位的误差和城市道路最大宽度, 一个轨迹点和它所匹配的道路距离超过60 m是不可能的。因此如果算法结果中出现了超过60 m的匹配距离则判定结果是错误的, 反之就认为结果是准确的。当然这种准确性也是相对的, 因为除非真正知道车辆的实际运行路段, 否则无法客观地判定匹配结果的准确性。根据准确匹配轨迹片段在全部成功匹配轨迹片段中所占的比例, 可以计算出算法匹配结果的准确率。

从图6的结果可以看到, 随着 T 的增大, 算法匹配的准确率逐渐下降。通过分析未能准确匹配的轨迹片段特征, 发现造成匹配错误的原因主要有两方面: 一是真实路网中存在某些路段, 但在地图数据中这些路段却是缺失的, 如图7a所示。在这种情况下, 轨迹只能匹配到相对更近的道路上去, 造成结果的偏差。二是车辆存在绕行的情况, 即在一个轨迹片段中车辆两次或多次经过了同一个节点, 导致路径搜索算法忽略了绕行部分的路段(因为这样总权值更小), 如图7b所示。绕行的原因可能是出租车

司机在绕行路段部分的某个地点放下乘客后又搭载了其他乘客,或者空车返回到其他地点载客等。



a. 路段缺失的情况



b. 车辆绕行的情况

图7 轨迹匹配不成功的两种典型情况

在以上两种情况中,第一种情况的改善依赖于道路网络地图完整度与准确度的提高,与算法本身无关。理论上如果有一个完全准确的路网地图,这种情况不可能发生。但第二种情况则与轨迹片段长度 T 的取值有关,因为 T 越长,出租车将以更大的可能性返回之前走过的一条路段。综合图6中成功率和准确率两方面的结果,本文建议 T 取值在5~10 min之间是一个比较合理的范围,这也与一般在线交通状态发布的时间间隔要求相符合。

3 结束语

本文针对传统的地图匹配算法不能有效处理复杂城市道路网络中并行路段和复杂立交场景下的车辆轨迹匹配的问题,提出了一种基于道路网络拓扑结构的轨迹匹配算法。该算法充分利用了网络拓扑结构信息,可确保找到的路径一定是拓扑上可行的

路径。同时,也充分考虑了轨迹点到路段的投影距离信息,使得最终匹配的路径以很大的可能性是车辆真实运行的路径。通过在成都市大规模路网中上万辆出租车的运行轨迹数据对本文算法进行了实际验证,结果表明在对轨迹片段长度进行合理取值的情况下,本文算法能获得较高的匹配成功率和准确率,能够满足实际应用的需求。这一研究结果为道路交通运行状态识别、在线交通信息发布、车辆定位与智能调度、出行路径选择行为分析等应用领域中的车辆GPS轨迹匹配提供了一种全新的方法。

相对于传统的地图匹配算法,本文算法的主要缺点是计算效率较低,主要原因是本文算法需要反复进行网络路径搜索和空间对象查询等耗时较高的运算。在实际应用中,可以从以下3方面着手提高该算法的计算效率:1)采用更为高效的最短路径搜索算法,例如基于堆结构和桶结构优先级队列的最短路径搜索算法^[25];2)采用更为高级的空间索引结构对路段进行索引,进一步提高空间查询效率;3)由于本文算法对不同轨迹片段的匹配是完全分开处理的,因此非常适合采用并行计算的方法,可从根本上提升算法的计算效率。

本文的研究工作得到成都师范学院科研项目(CS142B02)以及成都市委政研室提供的基础数据的资助,在此表示感谢。

参 考 文 献

- [1] 王建辉, 李爱光. 顾及多要素影响的路网匹配改进算法[J]. 测绘与空间地理信息, 2015, 38(3): 109-113.
WANG Jian-hui, LI Ai-guang. Improved algorithm of network matching base on multi factors influence[J]. Geomatics & Spatial Information Technology, 2015, 38(3): 109-113.
- [2] YUAN J, ZHENG Y, ZHANG L, et al. Where to find my next passenger[C]//Proceedings of the 13th International Conference on Ubiquitous Computing. New York: ACM, 2011: 109-118.
- [3] ZHENG Y, LIU Y, YUAN J, et al. Urban computing with taxicabs[C]//Proceedings of the 13th International Conference on Ubiquitous Computing. [S.l.]: ACM, 2011: 89-98.
- [4] YUAN J, ZHENG Y, XIE X. Discovering regions of different functions in a city using human mobility and POIs[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]: ACM, 2012: 186-194.
- [5] 周涛, 韩筱璞, 闫小勇, 等. 人类行为时空特性的统计力学[J]. 电子科技大学学报, 2013, 42(4): 481-540.

- ZHOU Tao, HAN Xiao-pu, YAN Xiao-yong, et al. Statistical mechanics on temporal and spatial activities of human[J]. Journal of University of Electronic Science and Technology of China, 2013, 42(4): 481-540.
- [6] YUAN J, ZHENG Y, ZHANG C, et al. T-drive: Driving directions based on taxi trajectories[C]//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2010: 99-108.
- [7] JENELIUS E, KOUTSOPOULOS H N. Travel time estimation for urban road networks using low frequency probe vehicle data[J]. Transportation Research Part B: Methodological, 2013, 53: 64-81.
- [8] YUAN J, ZHENG Y, XIE X, et al. Driving with knowledge from the physical world[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California: ACM, 2011: 316-324.
- [9] MA S, ZHENG Y, WOLFSON O. T-share: a large-scale dynamic taxi ridesharing service[C]//2013 IEEE 29th International Conference on Data Engineering (ICDE). [S.l.]: IEEE, 2013: 410-421.
- [10] HE W, LI D, ZHANG T, et al. Mining regular routes from GPS data for ridesharing recommendations[C]//Proceedings of the ACM SIGKDD International Workshop on Urban Computing. New York: ACM, 2012: 79-86.
- [11] CHAWLA S, ZHENG Y, HU J. Inferring the root cause in road traffic anomalies[C]//12th IEEE International Conference on Data Mining. Brussels: ICDM, 2012: 141-150.
- [12] ZHANG J T. Smarter outlier detection and deeper understanding of large-scale taxi trip records: a case study of NYC[C]//Proceedings of the ACM SIGKDD International Workshop on Urban Computing. New York: ACM, 2012: 157-162.
- [13] CHEN C, ZHANG D, CASTRO P S, et al. iBOAT: Isolation-based online anomalous trajectory detection[J]. IEEE Transaction on Intelligent Transportation Systems, 2013, 14(2): 806-818.
- [14] ERIC C A. Land-vehicle navigation system: an examination of the influence of individual navigation aids on system performance[D]. California: Stanford University, 1997.
- [15] SINN K, JONG H K. Adaptive fuzzy-network-based c-measure map-matching algorithm for car navigation system[J]. IEEE Transaction on Industrial Electronics, 2001, 48(2): 432-441.
- [16] BISHNU P P. Adaptive trajectory segmentation method and its application in in-car navigation system[D]. Columbus, Ohio: The Ohio State University, 2001.
- [17] CHEN Z W, SUN Y R, YUAN X. Development of an algorithm for car navigation system based on Dempster-Shafer evidence reasoning[C]//5th International IEEE Conference on Intelligent Transportation Systems. Singapore: IEEE, 2002: 534-537.
- [18] LOU Y, ZHANG C Y, ZHENG Y, et al. Map-matching for low-sampling-rate GPS trajectories[C]//Proceedings of the 17th ACM SIGSPATIAL. Seattle, WA: ACM, 2009: 352-361.
- [19] NEWSON P, KRUMM J. Hidden Markov map matching through noise and sparseness[C]//Proceedings of the 17th ACM SIGSPATIAL. Redmond, WA: ACM, 2009: 336-343.
- [20] RAYMOND R, MORIMURA T, OSOGAMI T, et al. Map matching with hidden Markov model on sampled road network[C]//21st International Conference on Pattern Recognition. Tsukuba: [s.n.], 2012: 2242-2245.
- [21] GOH C Y, DAUWELS J, MITROVIC N, et al. Online map-matching based on hidden Markov model for real-time traffic sensing applications[C]//15th International IEEE Conference on ITSC. Anchorage. Alaska: [s.n.], 2012: 776-781.
- [22] 中华人民共和国住房和城乡建设部. 城市道路交通规划设计规范: GB50220-1995[S]. 北京: 建设部标准定额研究所, 1995.
- MOHURD. Code for transport planning and design on urban road: GB50220-1995[S]. Beijing: Research Institute of Standard Quota of Ministry of Construction, 1995.
- [23] ZAIDI A S, SUDDLE M R. Global navigation satellite systems: a survey[C]//International Conference on Advances in Space Technologies. [S.l.]: IEEE, 2006: 84-87.
- [24] GUTTMAN A. R-Trees: a dynamic index structure for spatial searching[C]//Proceedings of the 1984 ACM SIGMOD. Boston: ACM, 1984: 47-57.
- [25] ZHAN F B. Three fastest shortest path algorithms on real road networks[J]. Journal of Geographic Information and Decision Analysis, 1997, 1(1): 70-82.