

DOI: 10.11830/ISSN.1000-5013.201701112



# 采用代表点插值的道路网提取方法

林杰<sup>1,2,3</sup>, 陈崇成<sup>1,3</sup>

- (1. 福州大学 空间数据挖掘与信息共享教育部重点实验室, 福建 福州 350002;
2. 福建工程学院 数理学院, 福建 福州 350108;
3. 福州大学 福建省空间信息工程研究中心, 福建 福州 350002)

**摘要:** 针对浮动车轨迹数据包含较多异常数据导致道路网提取效果不好的问题,提出一种渐近方式的代表点插值算法. 首先,采用 DBSCAN 聚类算法对转弯点聚类求出路口位置,计算经过两个路口间的轨迹,合并轨迹. 然后,采用改进的代表点算法提取代表点. 最后,利用最短路算法在由代表点建立的 Delaunay 三角网上插值,从而得到道路网. 实验结果表明:该方法能有效地从带有较多异常的轨迹数据中取出由复杂网络构建的道路网,具有较好的实用性.

**关键词:** 轨迹数据; 道路网; 海量数据挖掘; 聚类算法; 插值算法

**中图分类号:** TP 391      **文献标志码:** A      **文章编号:** 1000-5013(2018)01-0098-05

## Road Network Extraction Method Using Representative Points Interpolation

LIN Jie<sup>1,2,3</sup>, CHEN Chongcheng<sup>1,3</sup>

- (1. Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350002, China;
2. College of Mathematics and Physics, Fujian University of Technology, Fuzhou 350108, China;
3. Spatial Information Research Centre of Fujian, Fuzhou University, Fuzhou 350002, China)

**Abstract:** Aiming at the problem of the floating car trajectory data which contains many abnormal data leading to the bad road network extracting, the paper proposes a way of asymptotic representative point interpolation algorithm. Firstly, the method uses the DBSCAN clustering algorithm to find the intersections position, computes the trajectories between two intersections, and then merge the trajectories. Then an improved representative point algorithm is used for the extraction of representative points. Finally, using the shortest path algorithm to interpolate in Delaunay triangulation net which is established according to the representative points, road network is obtained. The experimental results show that the road network constructed by complex network can be extracted with the trajectory data which contains abnormal data, and has good practicability.

**Keywords:** trajectory data; road network; massive data mining; clustering algorithm; interpolation

随着全球定位系统和无线通信设备的广泛应用,收集车辆轨迹数据的技术正在逐步完善普及. 浮动车轨迹数据在城市研究中的一个重要应用是提取道路网. 道路网数据是一种非常重要的地理信息数据<sup>[1]</sup>,对方便人们交通出行、解决城市交通拥堵问题具有重要的意义. 车辆轨迹数据是车辆沿着道路行驶过程中利用车载定位设备获取的,是道路网络几何特征的直接反映<sup>[2]</sup>,这使从全球定位系统(GPS)轨

**收稿日期:** 2017-01-05

**通信作者:** 林杰(1980-),男,讲师,博士研究生,主要从事时空数据挖掘的研究. E-mail: fjljinjie@163.com.

**基金项目:** 国家自然科学基金资助项目(41471333)

迹数据中提取道路网数据成为可能. 目前, 一些学者已经用不同方法从浮动车数据中提取道路网<sup>[3-5]</sup>. Schroedl 等<sup>[6]</sup>提出从装有 GPS 接收器的车辆中提取高精度的道路网, 利用原有的数字地图把单独的轨迹划分成路段和路口, 并得到道路中心线. Brüntrup 等<sup>[7]</sup>提出一种基于 GPS 轨迹增量更新路网的方法, 成功地把每一轨迹融合到已推断出的路网中, 这种方法和 Li 等<sup>[8]</sup>提出的方法密切相关. Cao 等<sup>[9]</sup>把来自日常车辆的原始 GPS 轨迹转化为可路由的道路网, 该方法利用新建立的吸引力和排斥力模型分离不同道路的轨迹, 合并轨迹点生成路网. Karagiorgou 等<sup>[10]</sup>利用 1 s 或 5 s 这种时间间隔较小的轨迹数据, 经过路口识别、合并轨迹等步骤提取路网. 除了从 GPS 轨迹角度处理数据, 一些论文也利用 GPS 点作为基本处理单元提取路网信息. Worrall 等<sup>[11]</sup>把原始位置点聚类成有相同转向角的相似位置区域, 把聚类结果连接在一起, 形成连贯的链条, 最后, 利用非线性最小二乘拟合道路中心线. Guo 等<sup>[12]</sup>利用轨迹点数据提取代表点, 建立车辆轨迹的网络构造, 从而分析轨迹的分区模式. 此外, 一些学者利用图像处理方法从浮动车轨迹数据中提取道路网<sup>[2, 13-15]</sup>. 由于 GPS 终端被地面建筑物遮挡、车载 GPS 设备故障等客观原因, 接收到的轨迹数据会产生 GPS 数据漂移、车速值异常等情况<sup>[16]</sup>. 针对这种带异常较多的海量浮动车数据, 本文提出一种取用渐近的方式高效地提取代表点算法.

# 1 路网提取算法

文中路网提取算法流程, 如图 1 所示. 算法流程主要分为预处理、提取代表点和轨迹插值, 其步骤如下: 1) 利用轨迹数据计算转弯点, 采用 DBSCAN 聚类算法对转弯点聚类求出路口位置, 计算经过两个路口间的轨迹, 合并轨迹; 2) 经过第 1 步处理后得到的轨迹数据, 已经去除大量的异常数据, 再采用改进代表点算法提取代表点; 3) 先利用代表点建立 Delaunay 三角网结构, 再利用步骤 2) 合并后轨迹的每条线段的起点和终点查找最近的代表点, 用最短路算法求出起点和终点在三角网上的最短路径并保存, 最后, 计算复杂网络图边的权值.

## 1.1 预处理

预处理的第 1 个步骤是计算转弯点. 当浮动车轨迹的相邻 3 个轨迹点构成的夹角大于  $50^\circ$  且小于  $150^\circ$  时, 判断中间的轨迹点的转弯点. 绝大多数转弯点都发生在路口附近, 从而根据转弯点聚集的位置与聚类算法找到路口位置、细节<sup>[10, 17]</sup>. 由于对数据的概率分布没有任何先验知识, 不知道目标区域有多少个交叉口, 这就要求所选择的聚类算法能自动确定聚类的个数. 对此, 文中采用基于密度的聚类算法 DBSCAN 算法, 聚类数量为路口的数量, 聚类中心为路口位置. DBSCAN 算法包括领域半径值 Eps 和领域密度阈值 MinPts 两个参数. 当减少 MinPts 值时, 路口数量增加, 但有可能提取出伪路口<sup>[17]</sup>; 当增加 MinPts 值时, 路口数量减少, 这时会忽略真实的路口位置. 经过综合考虑, 将 Eps 和 MinPts 分别设为 40 m 和 10, 提取出 209 个路口位置. 根据浮动车的轨迹数据, 当采样点和路口的距离小于 45 m 时, 认为这辆车经过这个路口, 保存连接两个路口间的轨迹. 当两个路口的共有轨迹量超过一定阈值(大于 5 条)时, 即可认为这两个路口之间有道路相连.

预处理的最后 1 个步骤是多条轨迹合并. 这个处理步骤可去除一些异常的轨迹, 并简化轨迹点. 多条轨迹合并算法的主要步骤为: 两路口的轨迹中任意找 1 条, 每隔 30 m 取 1 个点  $P_i$ . 以点  $P_i$  为圆心, 以 30 m 为半径做圆, 找出落在圆内的轨迹点(当轨迹点数量小于 5 时, 换 1 条轨迹, 去除一些异常的轨迹), 并求出圆内轨迹点的重心  $c_i$ . 把所有的  $c_i$  按顺序连起来, 即可构成两路口的多条轨迹合并后的一条轨迹. 举例说明, 如图 2 所示. 路口 A, D 间有 A, B, C, D 这 4 个轨迹点, 间隔 30 m 取点, 得到的点为  $P_1 \sim P_6$ , 即  $P_1P_2$  的长度为 30 m,  $P_2B$  与  $BP_3$  的长度和为 30 m.

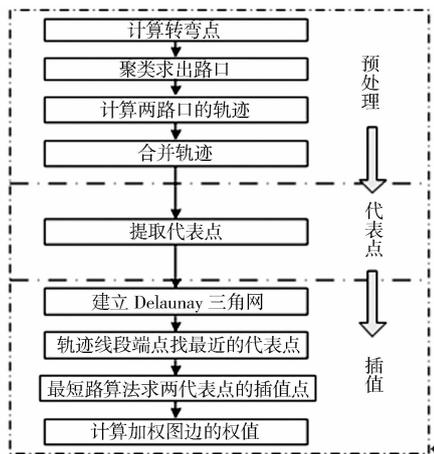


图 1 算法流程图

Fig. 1 Flow chart of proposed approach

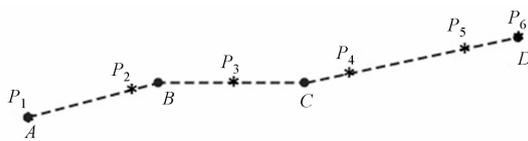


图 2 间隔 30 m 取点示意图

Fig. 2 Sketch map of interval 30 m taking points

## 1.2 提取代表点

经过预处理步骤的轨迹已去除大量的异常轨迹,但还不具备拓扑连接结构.采用改进的提取代表点算法自动提取出具有代表性的轨迹点.算法思路为:任意选择1个不是代表点的轨迹点,以特定半径作圆,计算出落在圆内的所有轨迹点的重心坐标作为代表点.这时得到的代表点靠近道路中心.然后,以这个代表点作圆,落在这个圆内的所有的点都设置这个点作为代表点,也可理解为删除圆内的其他点,只保留该代表点.提取代表点可以减少轨迹点的数量和冗余,便于插值.如果设置圆的半径太小,道路网的一些地方会不平整.根据测试结果,圆的半径选择为55 m.查找在半径为55 m的圆内的所有点时,建立kd树索引可提高查找效率.

文中提取代表点算法包括以下5个关键步骤.

**步骤1** 从任意的GPS点 $s$ 开始,代表点集 $C=\emptyset$ .

**步骤2** 找出点 $s$ 的55 m内的全部GPS点(包含点 $s$ ),计算这些GPS点的重心 $c$ .

**步骤3** 找出离点 $c$  55 m的GPS点.对每个点 $p_i$ ,当 $p_i$ 还不是代表点,指派 $p_i$ 到 $c$ (即 $p_i$ 用 $c$ 代表);当 $p_i$ 已经指派到另外的代表点 $q$ ,则重新指派 $p_i$ 到点 $c$ .

**步骤4** 从剩余的并不是代表点中选择下一个点 $s$ .

**步骤5** 重复步骤2~4,直到所有GPS点都有代表点.

文中改进提取代表点算法<sup>[12]</sup>,其中,与原算法比较,步骤3,4在效率上均有提高.在步骤3中,文献[12]在 $p_i$ 已经指派到另外的代表点 $q$ 这种情况时,需要分别计算 $p_i$ 到点 $q$ 和点 $c$ 的距离,当 $p_i$ 靠近 $c$ 时,重新指派 $p_i$ 到 $c$ .因为点 $p_i$ 指派到哪个代表点不会影响代表点的提取,从而点 $p_i$ 可任意指派到点 $q$ 或点 $c$ .在步骤4中,文献[12]在选择下一个点 $s$ 时,要求它是任一 $\{p_i\}$ 的邻居且没有代表过,这要求下一个代表点在 $\{p_i\}$ 的附近.而文中是随机从剩余的并不是代表点中选择点 $s$ ,其优点是在执行步骤4时均可减少1次最近邻查询操作.

## 1.3 插值代表点

利用步骤2提取的代表点建立Delaunay三角网.根据预处理的第3个步骤中得到的两路口间的轨迹,对于每两个轨迹点间的线段,设点 $A, B$ 是轨迹任意一条线段的起点和终点.分别找到离 $A, B$ 最近的代表点 $A', B'$ .根据修正的距离度量和最短路算法,找到 $A', B'$ 间在Delaunay三角网上的最短路径,即 $A, B$ 间的插值路径.对每段轨迹都做类似的插值处理,计算Delaunay三角网各条边的权值,从而得到由复杂网络构建的道路网.

设Delaunay三角网1条边的两个端点为 $u=(x_u, y_u), v=(x_v, y_v)$ .边长定义为欧氏距离的 $\alpha$ 次方<sup>[12]</sup>,即

$$d(u, v) = (\sqrt{(x_u - x_v)^2 + (y_u - y_v)^2})^\alpha. \quad (1)$$

式(1)中:当 $\alpha$ 大于1时,应用最短路算法会插值更多的代表点.改变 $\alpha$ 的取值可以控制1条直线路径和1条有更多代表点的曲线路径之间做平衡.经过实验,文中采用参数 $\alpha=1.5$ .

## 2 实验评价

为了对上述算法进行验证,实验1采用连续一周的上海市6 000辆浮动车轨迹数据.区域为 $[(121.424\ 9, 31.248\ 0), (121.487\ 9, 31.202\ 0)]$ ,位于上海市中心.涉及的轨迹数据共有3 640万个轨迹点,主要包括经度、纬度、时间等属性.原始数据的坐标是WGS-1984坐标系,先将其转换到UTM投影坐标系.

浮动车轨迹点原始数据,如图3所示.由图3可知:原始数据带有大量异常数据点;在一些地方,由于轨迹点过于密集,导致分辨不出道路形状.经过预处理后,得到55 150个轨迹点,这时已经去除大量的异常轨迹点,可以清晰分辨出道路,如图4所示.利用文中的代表点提取算法,共提取2 246个代表点,如图5所示.

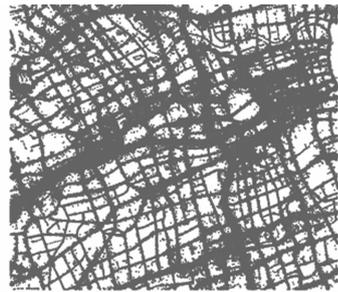


图3 原始的浮动车轨迹点  
Fig. 3 Original points of floating car trajectory

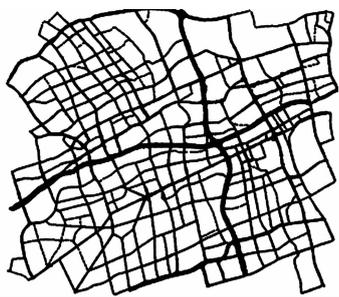


图 4 预处理后的轨迹点

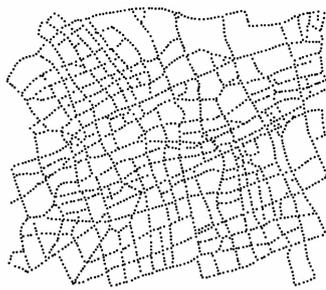


图 5 文中算法提取的代表点

Fig. 4 Trajectory points after pretreatment

Fig. 5 Extracted representative points with our algorithm

这些代表点均匀分布,是道路网的特征点,经过插值操作可获取道路网.分别从代表点提取质量和运行时间两方面进行测试比较.1)针对部分的原始轨迹点直接提取代表点(全部的原始轨迹点提取代表点的运行时间太长,故只用部分原始轨迹点测试),将得到杂乱分布的代表点,从而无法提取出道路网.2)采用相同的 55 150 个轨迹点测试提取代表点的运行时间.文中算法执行时间为 113 s,原算法的执行时间为 191 s,文中提取代表点算法运行时间仅为原算法的 59%.需要指出的是,原算法是直接对原始数据执行提取代表点算法,而原始数据点是 3 000 多万万个轨迹点,需要耗费大量时间提取代表点,远多于 191 s.而文中是对预处理后的轨迹点执行代表点算法,预处理后的轨迹点仅有 55 150 个.因此,文中方法的执行时间远少于原方法提取代表点算法执行时间.

将提取的道路网图层叠加到百度地图进行比较,如图 6 所示.由图 6 可知:利用浮动车轨迹数据提取的道路网与现有的路网结构基本一致.但从细节上看有些差异,在实验区中部提取的道路与百度地图的道路有偏移,在实际应用时需对提取的道路网进行修正.



图 6 实验结果与百度地图叠加比较  
Fig. 6 Comparison of experimental result with Baidu map

实验 2 的轨迹数据是由文献[5]提供的美国芝加哥市轨迹数据,覆盖范围是 7.0 km×4.5 km,如图 7 所示.将文中方法与文献[10]算法的提取结果相比较,如图 8 所示.图 8 中:两处椭圆框标注处为两种方法提取结果的差异.由图 8 可知:左下方标注内应该有道路,文献[10]没有提取出;右上方标注内没有道路,文献[10]将其误提取,而文中方法均得到正确的结果.

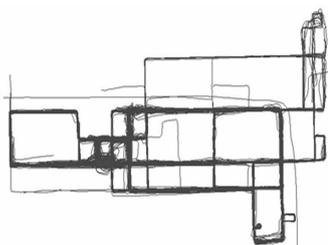
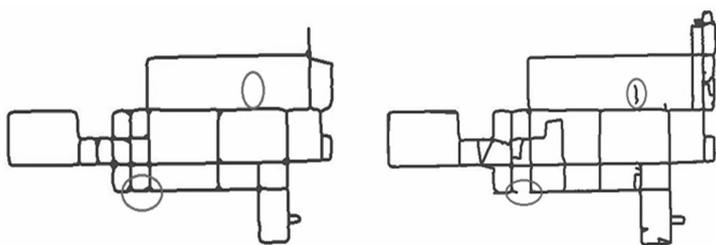


图 7 芝加哥市轨迹数据

Fig. 7 Chicago trajectory data



(a) 文中的提取结果

(b) 文献[10]的提取结果

图 8 两种方法提取结果比较

Fig. 8 Comparison of two extracting results

### 3 结论

针对浮动车轨迹数据存在较大异常数据的实际情况,文中提出预处理、提取代表点和轨迹插值 3 大步骤的方法流程.文中方法具有以下 4 个特点.

- 1) 预处理去除大量异常轨迹点,使代表点提取质量上优于直接在原始轨迹点中提取代表点.这十分重要,因为高质量的代表点才能够提取出高质量的道路网.
- 2) 预处理简化了原始轨迹点的数量,简化效率为 0.15%,提取效率优于直接对原始轨迹点提取代

表点.

3) 改进了原有的提取代表点算法,提高了运行效率.

4) 针对包含大量异常轨迹数据提取出的道路网质量较高,提取结果具有复杂网络的拓扑结构.利用文中提出的方法,可以避免耗费大量的人力物力,同时为道路网更新和变化发现提供关键技术支撑.

#### 参考文献:

- [1] 邹利林,胡学东,梁发超,等.武汉市道路网空间结构与住宅地价的协同性分析[J].华侨大学学报(自然科学版),2015,36(5):587-592. DOI:10.11830/ISSN.1000-5013.2015.05.0587.
- [2] 蒋益娟,李响,李小杰,等.利用车辆轨迹数据提取道路网络的几何特征与精度分析[J].地球信息科学学报,2012,14(2):165-170. DOI:10.3724/SP.J.1047.2012.00165.
- [3] BIAGIONI J,ERIKSSON J. Inferring road maps from global positioning system traces: Survey and comparative evaluation[J]. Transportation Research Record: Journal of the Transportation Research Board,2012,2291(1):61-71. DOI:10.3141/2291-08.
- [4] FATHI A,KRUMM J. Detecting road intersections from GPS traces[J]. Geographic Information Science,2010,6292:56-69.
- [5] AHMED M,KARAGIORGOU S,PFOSER D,*et al.* A comparison and evaluation of map construction algorithms using vehicle tracking data[J]. Geoinformatica,2015,19(3):601-632.
- [6] SCHROEDL S,WAGSTAFF K,ROGERS S,*et al.* Mining GPS traces for map refinement[J]. Data Mining and Knowledge Discovery,2004,9(1):59-87.
- [7] BRÜNTRUP R,EDELKAMP S,JABBAR S,*et al.* Incremental map generation with GPS traces[C]//Proceedings of Intelligent Transportation Systems, Vienna;IEEE Press,2005:574-579. DOI:10.1109/ITSC.2005.1520084.
- [8] LI Jun,QIN Qiming,XIE Chao,*et al.* Integrated use of spatial and semantic relationships for extracting road networks from floating car data[J]. International Journal of Applied Earth Observation and Geoinformation,2012,19(1):238-247. DOI:10.1016/j.jag.2012.05.013.
- [9] CAO L,KRUMM J. From GPS traces to a routable road map[C]//Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York;ACM,2009:3-12. DOI:10.1145/1653771.1653776.
- [10] KARAGIORGOU S,PFOSER D. On vehicle tracking data-based road network generation[C]//Proceedings of the 20th International Conference on Advances in Geographic Information Systems, New York;ACM,2012:89-98.
- [11] WORRALL S,NEBOT E. Automated process for generating digitised maps through GPS data compression[C]// Australasian Conference on Robotics and Automation, Brisbane:[s. n.],2007:1-6.
- [12] GUO Diansheng,LIU Shufan,JIN Hai. A graph-based approach to vehicle trajectory analysis[J]. Journal of Location Based Services,2010,4(3/4):183-199. DOI:10.1080/17489725.2010.537449.
- [13] 张莉婷,陈云波,左小清,等.出租车轨迹数据快速提取道路骨架线方法[J].测绘科学,2015,40(1):110-112. DOI:10.16251/j.cnki.1009-2307.2015.01.024.
- [14] WU C,AYERS P D,ANDERSON A B. Validating a GIS-based multi-criteria method for potential road identification[J]. Journal of Terramechanics,2007,44(3):255-263. DOI:10.1016/j.jterra.2007.04.001.
- [15] ZHAO Yue,LIU Jian,CHEN Runqiang,*et al.* A new method of road network updating based on floating car data [C]// Geoscience and Remote Sensing Symposium, Vancouver; IEEE Press, 2011: 1878-1881. DOI: 10.1109/IGARSS.2011.6049490.
- [16] 辛飞飞,陈小鸿,林航飞.浮动车数据路网时空分布特征研究[J].中国公路学报,2008,21(4):105-110.
- [17] 朱云龙,刘微.基于轨迹聚类的路网构造[J].吉林师范大学学报(自然科学版),2015,36(1):11-17.

(责任编辑:钱筠 英文审校:吴逢铁)