

文章编号:1671-8860(2013)05-0599-05

文献标志码:A

利用浮动车数据提取停车场位置

李军¹ 秦其明¹ 游林² 解超¹

(1 北京大学遥感与地理信息系统研究所,北京市海淀区颐和园路5号,100871)

(2 宁波市规划与地理信息中心,宁波市江东区江澄北路575号,315042)

摘要:针对当前停车场信息采集手段周期长、成本高的问题,提出了利用浮动车数据自动提取停车场位置的方法。通过分析在停车场内采集的浮动车数据具有的典型特点后,使用DBSCAN算法检测出位于停车场内的点簇进而提取停车场的位置;同时针对DBSCAN算法具有高时间复杂度的缺点,结合定位点簇的空间尺寸限制条件构建了特定的空间索引,提高了聚类算法的效率。试验结果验证了该方法的有效性。

关键词:POI更新;停车场提取;浮动车数据;DBSCAN算法

中图法分类号:P208

随着国民经济的快速发展,我国的汽车保有量呈逐年递增的趋势,截至2011年末,全国民用汽车保有量达到10578万辆,比上年末增长16.4%,其中私人汽车保有量7872万辆,增长20.4%^[1]。汽车数量的增长导致各大中城市开始出现停车难问题,人们对停车场信息的需求越来越迫切,希望能通过手机或车载导航仪等设备随时随地获取当前地点附近的停车场分布。目前,包括停车场在内的感兴趣点(point of interest,POI)信息主要是依赖大量的人工调绘^[2]或专门的测量设备^[3-5]采集得到。在当前中国快速的城市建设进程中,停车场新建和改建工程频繁,位置变化快,而依靠人工调绘或专有设备测量周期长、成本高,远远满足不了实际需求^[6-7]。因此,POI信息快速采集方法逐渐成为国内外研究热点。

浮动车技术由于具有低投入、不受天气条件限制等优点,开始广泛应用于路况采集中^[8]。除了能获取路况信息外,浮动车数据也隐含着其他类型的交通信息,一些学者研究了浮动车数据的深层次挖掘应用,包括道路网提取^[9]、计算交叉口延误时间^[10]、检测交通事故^[11]、计算汽车尾气排放量^[12]等。以地理信息科学视角来看,一方面浮动车使用电子地图能够获取导航服务;另一方面可以将浮动车看作移动测量车,在行驶时以一定的时间间隔持续采集经过区域的位置信息,包括

经过的停车场,因此,浮动车数据可以为停车场信息的更新提供数据源。

本文以挖掘浮动车数据中的隐含信息为目的,通过分析停车场内采集的浮动车数据中存在的特定模式,提出了利用浮动车数据提取停车场位置的方法。该方法采用DBSCAN算法从浮动车数据中检测出位于停车场内的点簇进而提取停车场的位置;同时结合定位点簇的空间尺寸限制条件对DBSCAN算法进行改进,提高了停车场提取过程的效率。

1 方法原理与算法

从浮动车数据中提取停车场就是一个地理数据挖掘过程,具体来讲是从所有浮动车数据中识别出在停车场采集到的GPS定位点,并计算出用来表示停车场的点位置。其中,最基本的问题是研究停车场内采集的定位数据表现出的典型特点。

1.1 停车场定位数据的典型特点

浮动车在行驶的过程中经过的地物主要有两类:道路和停车场。在道路中有两处地方的车辆运动特点与停车场相似:拥堵频发地和交通信号灯路口。在本文中,将浮动车在道路上行驶时所采集的GPS点称为道路定位点,在停车场采集的GPS点称为停车场定位点,在拥堵频发地和交通

信号灯路口采集的 GPS 点称为道路瓶颈定位点。不同地物处采集的 GPS 定位点由于地物的功能不同具有对应的分布特点。如图 1 所示,在道路上,车辆通常以相对稳定的速度行驶,因此道路定位点一般沿着道路均匀分布,而且大部分速度不为零;在停车场内(区域 A),车辆通常长时间处于静止状态,且在一段时间内位于此区域的车辆数量较多,因此停车场定位点通常呈聚集分布,且大部分速度为零;在拥堵频发地和交通信号灯路口处(区域 B 和 C),由于车辆运动状态与在停车场里类似,则道路瓶颈定位点也呈聚集分布,但形成的点簇中心距离道路中线很近。

通过对比不同地物处定位点的空间分布与属性特点,可以总结出停车场定位数据具有与其他地物处定位数据所不同的典型特征:定位点分布密度大,形成明显的点簇;点簇中心离道路中线距离较远;大部分定位点的速度属性为零;单位面积内车辆数量多;单位面积内定位点数量多。

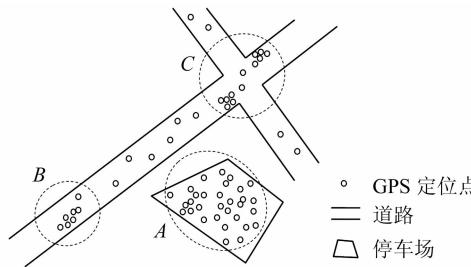


图 1 浮动车 GPS 定位点的空间分布特点

Fig. 1 Spatial Distribution of GPS Positioning Points Collected by Floating Cars

1.2 停车场提取方法

根据停车场定位数据的特点可知,停车场定位点通常聚集成一个高密度的点簇,识别浮动车数据中存在的高密度点簇常用的方法是空间聚类技术。空间聚类是指将一组对象的集合划分成为多个类或簇的过程,使簇内的对象之间具有较高相似度,而簇间对象具有较低相似度。常用的空间聚类方法有划分法、层次法、基于密度的方法和基于格网的方法^[13]等。相对停车场定位点来说,其他类型定位点如道路定位点就是噪声点,而且停车场没有固定的形状,因此,本文选择 DBSCAN 算法^[14]对定位点进行聚类,它是一种基于密度的方法,可以有效地处理噪声点,并能检测任意形状的聚类。

1.2.1 DBSCAN 算法

DBSCAN 算法的具体步骤如下。

1) 搜索数据集 D 中尚未处理过的对象,如果 p 未被处理过,则查询 p 的 Eps 邻域 $N_{\text{Eps}}(p)$,若包含的对象数不小于 MinPts ,则建立新簇 C ,将

$N_{\text{Eps}}(p)$ 中所有点添加到 C 中;

2) 依次搜索 C 中未被处理的对象 q ,查询 q 的 Eps 邻域 $N_{\text{Eps}}(p)$,若 $N_{\text{Eps}}(p)$ 包含至少 MinPts 个对象,则将 $N_{\text{Eps}}(p)$ 中未被划分到任何簇的对象添加到 C 中;

3) 重复执行步骤 2),继续搜索 C 中未被处理过的对象,直到 C 中所有对象都被处理过;

4) 重复执行步骤 1)~步骤 3),直到所有对象都被划分到某个簇中或被标记为噪声。

1.2.2 DBSCAN 算法的改进

从 DBSCAN 算法流程中可以看到,每一次计算都包含一次邻域查询,而邻域查询需要遍历整个数据集,当处理的数据量特别大时,邻域查询耗费的时间很长,因此,提高算法效率的关键是减少邻域查询的范围。本研究中的聚类分析被用于提取地理实体即停车场,聚类后得到的任一点簇即代表一个地理实体。在现实世界中地理实体的空间尺寸是有限的,所以对应地理实体的点簇也具有有限的空间尺寸,基于这个前提条件,本研究可以设定一个距离阈值限制邻域搜索空间。另外,分布密度小的点簇不能形成停车场,因此,可以忽视这些点,减少处理对象的数目。针对本文应用目的的特殊性,对 DBSCAN 算法做如下改进。

点簇的空间尺寸限制表达为:点簇中的任意两点间的距离不大于 d_{\max} 。在执行 DBSCAN 算法前,先计算包含所有对象的最小外包矩形,在该范围上构建格网(图 2),其中格网单元的边长为 d_{\max} ,然后确立每个对象与格网单元的一一对应关系,即格网索引。根据格网的最小点密度阈值 ρ_{\min} ,筛选出待处理的格网集合 $\{G_s \mid \text{Den}_s \geq \rho_{\min}\}$,将位于待处理格网集合中的对象组成搜索数据集合 D 。

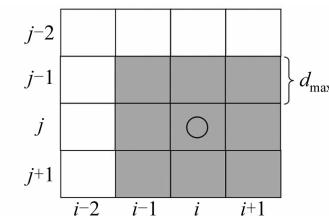


图 2 空间格网索引(标注有圆圈的格网为当前格网,灰色格网为邻域搜索格网)

Fig. 2 Spatial Grid Index

在查询某个对象的 Eps 邻域时,先根据格网索引得到此对象所在的格网单元 G_s ,设定以 G_s 为中心的 9 个格网作为邻域搜索格网(图 2),即 $\{G_{pq} \mid p=s-1, s, s+1; q=t-1, t, t+1\}$,只在邻域搜索格网内的对象中查询当前对象的 Eps 邻域。需要注意当一个对象位于边缘格网中时,搜

索格网会少于9个。

1.2.3 停车场提取流程

1) 坐标系转换。将GPS定位数据从地理坐标系转换到投影坐标系,便于后面步骤中的距离计算;

2) 停车定位点筛选。从原始数据中筛选出速度零值点,为了剔除一些偶发因素(定位设备误差、临时停车)导致的速度零值点,添加距离、方向指标作为筛选条件。筛选方法为:对于一辆车的两个连续定位点 O_i 和 O_{i+1} ,若 $V_{O_i}=V_{O_{i+1}}=0$, $\text{dis}(O_i, O_{i+1}) \leq d_s$ 且 $\text{angledif}(O_i, O_{i+1}) \leq \text{ang}_s$,则 O_i 为停车定位点。

3) 停车定位点聚类。使用改进后的DBSCAN算法对步骤2)筛选出的停车定位点进行分类,将高密度的定位点集划分到同一类中,分类后得到若干个点簇 C_i ,并剔除噪声点。

4) 聚类中心计算。对任一个点簇 C_i ,提取簇心 $P_i(X_{P_i}, Y_{P_i})$,其中 $X_{P_i} = \frac{1}{|C_i|} \sum_{O_j \in C_i} X_{O_j}$, $Y_{P_i} = \frac{1}{|C_i|} \sum_{O_j \in C_i} Y_{O_j}$,该点簇中的定位点数量为 $ON_i = |C_i|$,并通过车牌号字段统计点簇包含的车辆数量 VN_i ,得到簇心集合 S 。

5) 停车场位置提取。将 S 中距离较近的簇心归并到同一类中,得到 S 的多个子集 S_k ,使 S_k 中任意两个簇心的距离不超过 d_c 。合并 S_k 中的簇心,得到新的簇心 (X_{S_k}, Y_{S_k}) ,其中 $X_{S_k} = \frac{\sum_{P_i \in S_k} ON_i \times X_{P_i}}{\sum_{P_i \in S_k} ON_i}$, $Y_{S_k} = \frac{\sum_{P_i \in S_k} ON_i \times Y_{P_i}}{\sum_{P_i \in S_k} ON_i}$,合并后簇心的定位点数量为 $ON_{S_k} = \sum_{P_i \in S_k} ON_i$,车辆数

量为 $VN_{S_k} = \sum_{P_i \in S_k} VN_i$ 。剔除车辆数量小于 N_{\min} 或与道路中线距离小于 d_r 的簇心,剩下的簇心即为停车场。

2 试验与分析

2.1 试验数据

本文使用全国重点营运车辆监控平台的浮动车数据,主要字段包括车牌号、经度、纬度、行驶速度、行驶方向、定位时间等。以北京市及周边地区作为试验区域(图3)。试验区内有多种类型的行政区域:大城市(北京市区)、中等城市(廊坊市)、小城市(大兴区、房山区等)和农村(北京郊区),便

于检验本文方法在多种类型地区的有效性。本文以监控中心接收到的来自试验区的浮动车的GPS定位数据作为试验数据。

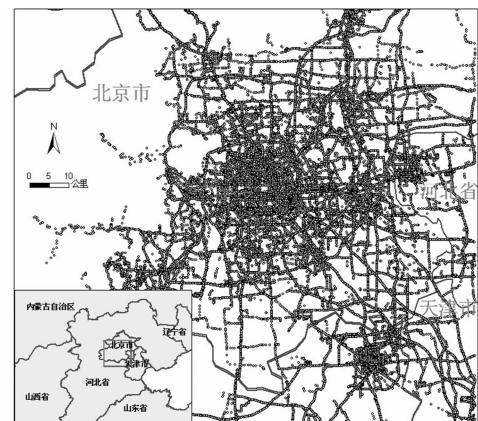


图3 试验区及试验数据的分布

Fig. 3 Experiment Area and Spatial Distribution of Test Data

2.2 试验结果

原始数据的坐标是在WGS-1984坐标系下,首先将其转换到UTM投影坐标系下。该监控平台中浮动车安装的GPS接收机定位误差为5~10 m,结合车辆的行驶特点,将距离参数 d_s 和角度 ang_s 分别设为10 m和60°。本文综合考虑具体应用目标和经验,将Eps和MinPts分别设定为30 m和30。在构建空间格网索引时, d_{\max} 和 ρ_{\min} 分别取300 m和20。

图4(a)和4(b)分别展示了两个局部区域的聚类结果,图中相同颜色的点组成一个点簇,代表一个具有高分布密度的点集,白色点为噪声点,不属于任何一个点簇。由于停车场中的车辆多且停止时间长,因此这些点簇位于停车场内的可能性比噪声点大。

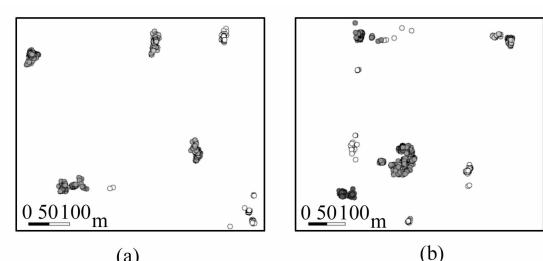


图4 GPS定位点聚类结果

Fig. 4 Clustering Results of GPS Positioning Points

在计算出聚类中心坐标后,设置 d_c 为50 m, N_{\min} 为2, d_r 为15 m,融合距离较近的聚类中心并筛选出可能为停车场的簇心。再使用道路缓冲区剔除掉由堵车、路口停车等因素产生的簇心。最终从定位数据中提取出了463个浮动车常用的

型停车场,如图5所示。

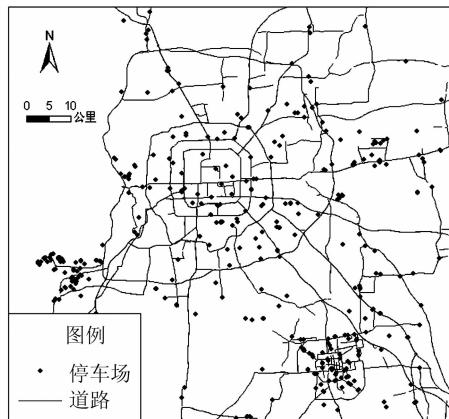


图5 停车场提取结果

Fig. 5 Parking Lot Extraction Result

2.3 效率验证

本文从实验数据中依次选取包含不同数量浮动车定位点的子集,分别使用经典的DBSCAN算法和改进后的算法对各子集进行聚类处理,图6给出了两种算法的执行时间和处理结果。可以看到,改进的DBSCAN算法大大提高了算法效率,并且得到的聚类数量和经典算法基本一致。

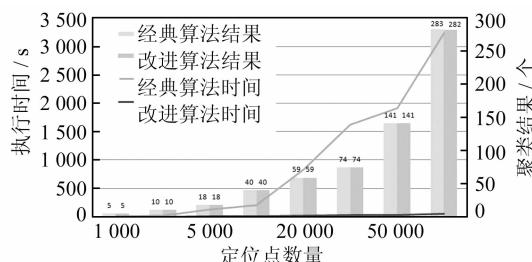


图6 经典和改进DBSCAN算法的效率对比

Fig. 6 Contrast in Efficiency Between the Classical and Modified DBSCAN Algorithms

2.4 精度验证

为了验证提取出的停车场位置的准确度,将停车场点图层叠加到GoogleEarth卫星影像图上。经过比较,在通过自动提取得到的463个停车场中,395个为真实停车场,其正确提取率为85.3%。通过对停车场点的位置与卫星影像中的停车场区域,发现点数据大多位于停车场的中心区域,与卫星影像匹配得非常好,图7展示了部分停车场的匹配效果。

3 结语

目前,电子地图中的停车场信息采集主要依赖大量的人工调绘或专门的测量设备,更新周期长且成本高,难以满足实际要求。本文以探索从

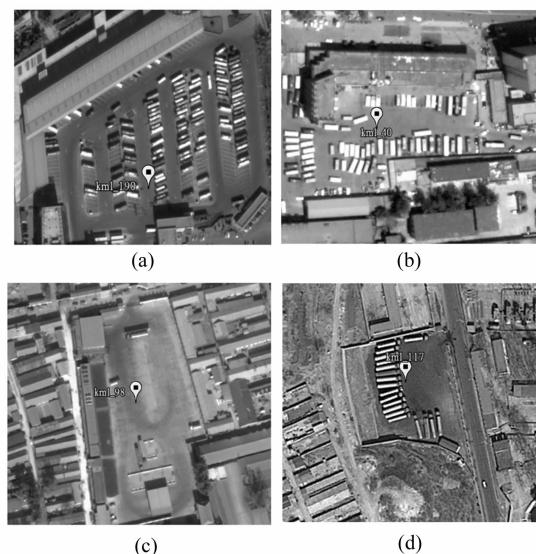


图7 部分停车场的点数据

Fig. 7 Points Representing Parking Lots

浮动车数据中提取停车场位置为目的,在总结停车场内浮动车数据的典型特点的基础上,使用DBSCAN算法从浮动车数据中提取代表停车场的点簇,从而计算停车场位置,并结合点簇的空间尺寸限制条件对算法进行改进,提高了算法效率。最后,将该方法应用于全国重点营运车辆监控平台的浮动车数据中,试验结果表明该方法能准确提取出代表停车场区域的标记点位置,可以为小比例尺电子地图提供数据源。

DBSCAN算法需要事先确定Eps和MinPts两个参数的值,参数的选择对聚类效果有一定的影响,今后将进一步研究处理不同时间周期的浮动车数据时的参数选取优化方法。此外,浮动车数据中也包含了收费站、加油站等地物的空间位置信息,今后也将探索从浮动车数据中提取其他类型POI的方法。

参 考 文 献

- [1] 中华人民共和国国家统计局. 2011年国民经济和社会发展统计公报 [EB/OL]. http://www.stats.gov.cn/tjgb/ndtjgb/qgndtjgb/t20120222_402786440.htm, 2012
- [2] 崔健,魏金鹏,仲伟政,等. 基于栅格基本图的POI数据更新方法研究 [J]. 测绘通报, 2011(3): 66-68
- [3] 李德仁. 移动测量技术及其应用 [J]. 地理空间信息, 2006, 4(4): 1-5
- [4] 王龙波. 基于GPS与实景影像的POI快速采集技术 [J]. 中国科技信息, 2007, 12: 121-122
- [5] 徐敬海,李清泉. 基于时态GIS的导航电子地图增量更新研究 [J]. 武汉大学学报·信息科学版, 2008, 33(11): 1182-1185

- [6] 崔铁军,李宏利,郭鹏,等.导航电子地图增量更新服务及关键技术研究[J].地理信息世界,2011(4):36-39
- [7] 李连营,许小兰,吴中恒,等.多源导航电子地图数据融合更新方法研究[J].武汉大学学报·信息科学版,2008,33(4):409-412
- [8] 李宇光,熊普选,乐阳.基于大样本浮动车数据的武汉市车辆行驶速度获取与分析[J].交通信息与安全,2009,27(150):26-29
- [9] 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: 238-247
- [10] 于泉,孙玲,荣建.基于浮动车数据调查方法的交叉口延误计算[J].重庆交通大学学报(自然科学版),2009,28(2):283-286
- [11] 张存保,杨晓光,严新平.基于浮动车的高速公路交通事故事件自动判别方法研究[J].武汉理工大学学报
- [12] 朱倩茹,刘永红,曾伟良,等.基于GPS浮动车法的机动车尾气排放量分布特征[J].环境科学研究,2011,24(10):1 097-1 103
- [13] Han Jiawei, Kamber M, Tung A K H. Geographic Data Mining and Knowledge Discovery[C]. Chapter Spatial Clustering Methods in Data Mining: a Survey, Taylor & Francis, 2001
- [14] Ester M, Kriegel H P, Sander J, et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise[C]. The Second International Conference on Knowledge Discovery and Data Mining, Portland, USA, 1996
-
- 第一作者简介:**李军,博士生。现主要从事地理数据挖掘和交通地理信息系统方面的研究。
E-mail: newton2069@126.com

Parking Lot Extraction Method Based on Floating Car Data

LI Jun¹ QIN Qiming¹ YOU Lin² XIE Chao¹

(1) Institute of Remote Sensing and Geographic Information Systems, Peking University, 5 Yiheyuan Road, Haidian District, Beijing 100871, China

(2) Ningbo Surveying, Mapping and Geoinformation Bureau, 575 North Jiangcheng Road, Jiangdong District, Ningbo 315042, China)

Abstract: The phenomena that parking is difficult in large and medium-sized cities makes people's desire for parking lot information become stronger and stronger, and the level of detail and accuracy of parking lot information in electronic maps directly impact the service quality of maps. In view of the problems of current surveying methods of parking lots, this paper proposes an approach to automatically extract the locations of parking lots from floating car data. In this paper, the DBSCAN algorithm is used to detect point clusters located within parking areas, and a special spatial grid index is established to decrease the time complexity of the clustering algorithm. Experiment results validate the proposed method.

Key words: POI update; parking lot extraction; floating car data; DBSCAN algorithm

About the first author: LI Jun, Ph.D candidate. His main research fields are geographic data mining and geographic information system for transportation.

E-mail: newton2069@126.com