

引用格式:王少帆,魏福豪,黄世雨,尹宝才.基于图密度峰值聚类算法的热点路段发现[J].中国传媒大学学报(自然科学版),2023,30(01):31-38.

文章编号:1673-4793(2023)01-0031-38

## 基于图密度峰值聚类算法的热点路段发现

王少帆,魏福豪,黄世雨,尹宝才\*

(北京工业大学信息学部,北京 100124)

**摘要:**传统的密度峰值聚类算法不仅具有较高的计算复杂度且未考虑路网固有的拓扑结构,无法衡量各路段之间的关联关系。针对这一问题,提出基于图密度峰值聚类算法的出行热点路段发现。该算法将交通路网用图模型结构,然后以各路段为基本单元计算局部密度及高局部密度距离并画出决策图找出聚类中心,最后结合实际区域的兴趣点分析该聚类簇成为热点路段的潜在可能。借助于图模型表达形式的优势,该算法不仅可以大幅度提升算法的计算复杂度,而且可以更加准确合理的找出热点路段。通过在滴滴-成都轨迹数据集上的实验表明,图密度峰值聚类算法具有更高的热点路段发现精度,并且在计算效率上有大幅度提升。

**关键词:**智能交通;出行热点;图密度峰值聚类;热点发现;滴滴数据集

**中图分类号:**U491 **文献标识码:**A

## Discovering hotspot road segments based on a graph density peak clustering algorithm

WANG Shaofan, WEI Fuhao, HUANG Shiyu, YIN Baoc\*

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

**Abstract:** The traditional density peak clustering algorithm not only has high computational complexity, but also does not consider the inherent topology of the road network. Hence, it cannot measure the intrinsic relationship between the various road segments. Aiming at this problem, this paper proposes a travel hotspot road segments discovery based on GDPC algorithm. The GDPC algorithm uses a graph model structure for the traffic road network, then uses each road segment as the basic unit to calculate the local density and the minimum high local density distance. Afterwards, the algorithm draws a decision diagram to find the cluster center, and finally combines the points of interest in the actual area to analyze the potential of the cluster to become a hot spot. With the advantage of the graph-based representation, the GDPC algorithm can not only greatly improve the computational complexity compared with traditional algorithms, but also find hot spots more accurately and reasonably. Experiments on the Chengdu Didi dataset show that the GDPC algorithm is more reasonable, and achieves a significant improvement in computational efficiency.

**Keywords:** intelligent transportation; travel hotspots; graph density peak clustering; hotspots discovery; DiDi dataset

基金项目:北京市自然科学基金项目(4222021);国家自然科学基金项目(U1811463)

作者简介(\*为通讯作者):王少帆(1982-),博士,副教授,主要从事模式识别、智能交通研究。Email:wangshaofan@bjut.edu.cn;尹宝才(1963-),博士,教授,主要从事数字多媒体、智能交通、多功能感知技术研究。Email:ybc@bjut.edu.cn

## 1 引言

随着大城市的汽车保有量迅猛增长,城市交通面临巨大的压力,道路拥堵已成为城市化进程中面临的严峻挑战。及时发现城市的出行热点,可以更好地了解城市的出行需求、优化交通资源配置,有利于实现对城市交通的智能管理与合理调度。Qu等<sup>[1]</sup>在研究中指出,传统的出租车车队可以超过50%的空闲时间,空置的出租车并不只是在固定的位置等待;相反,它们通常会四处行驶,燃烧燃料,拥堵有限的道路空间,造成公共资源浪费。随着网约出租车的迅速发展,这一问题得到了一定程度的解决;但是在一些高峰时段也会存在司机接单不及时导致乘客滞留时间较久的情况。为了能够寻找城市出租车分时段的热点区域,解决城市中心位置的拥堵问题,研究者们尝试采用了各种聚类算法。其中密度峰值聚类算法是Liu等<sup>[2]</sup>在2015年提出的一种典型聚类方法,该方法可以自动识别各种形状类簇,并且参数很容易确定。通过对历史数据的整合他们得到了与规划出租车等位点相对应的聚类结果,验证了算法的实用性,作为交通热点发现的经典方法,密度峰值聚类算法克服了其它聚类算法的类间密度差别大、邻域范围难以设定等问题,具有较高的鲁棒性。但是对于小范围的热点推荐应用不太现实,因为对于一个商业区如果有十几个热点,那么司机只要在商业区附近就容易接到客人不需要寻找热点<sup>[3]</sup>。在面对海量数据时,密度峰值聚类算法存在较高的计算复杂度以及处理效率低等问题。针对这一问题,何月等<sup>[4]</sup>针对出租车运营过程中缺少路径优化指导造成运营能力分布不均、空载率高的问题,提出了1种基于网格的出租车载客热点聚类算法,通过对出租车GPS轨迹数据进行处理和聚类分析,充分挖掘出租车热点区域,但该方法未考虑路网固有的拓扑结构,无法衡量各路段之间的关联关系。

综上所述,传统的交通热点发现方法存在应用范围受限、计算复杂度高等问题,无法准确刻画客观存在的交通路网上路段与路口的交通热点状态。针对上述问题,本文基于城市路网的拓扑结构构建图模型,提出图密度峰值聚类算法,寻找在特定时段下的热点路段,帮助出租车更精准的寻找订单位置,大大提高出租车的运营效率。

## 2 相关研究

近年来,基于网约车数据的出租车热点发现研究

越来越多<sup>[5,6]</sup>。Qu等<sup>[7]</sup>利用自动ARIMA模型进行时间序列分析,但是对于小量需求时,分析结果并不可靠。Li等<sup>[8]</sup>通过对ARIMA模型改进对杭州市区热点乘客数量预测并做热点推荐,然而该模型由于仅考虑了前一时间段数据对预测的影响,造成推荐热点不准确导致司机浪费时间。Liu等<sup>[9]</sup>分析武汉7200辆出租车两个月的GPS数据数据样本,发现高收入司机总是能及时抓住乘客热点位置,并在相应时段巡航于热点区域,因此如果能使出租车司机及时了解区域热点的实时热度和区域概率信息,将有利于优化出租车运输系统。Dow等<sup>[10][11]</sup>分析台湾出租车公司的历史出租车GPS记录数据,并提取在真实环境中拥有大量载客记录的区域,对台湾地区热点进行分析,在此基础上建立自适应模型对热点区域进行推荐,结果表明该方案在减少空车巡航时间上优于其他方案。Xie等<sup>[12]</sup>对Arcgis软件获取的位置信息进行分析,获得旅行需求的分布特征。Verma<sup>[13]</sup>对Uber纽约数据进行基于PAM(Partitioning Around Medoids)的聚类分析,将获得的数据分为工作日上午、工作日晚上和周末高峰时段,同时将区域分成住宅场所、办公场所和聚会场所。

盛宇裕等<sup>[14]</sup>通过构建循环神经网络模型对城市热点区域进行预测,采用参数服务器与工作架构解决大规模参数分布式存储和更新一致性问题,但由于交通信息存在于城市空间的复杂开放系统中,受地理要素相互依存、协同演化的影响,从宏观角度来看,通过简单的数据归纳和逻辑推理来对不确定性结果进行解释是不充分的。陈锐等<sup>[15]</sup>利用导航大数据,提出基于谱聚类和蚁群算法的城市热点区域关联性挖掘方法,将关联性分析转化为图的最优划分问题以及最优路径求解问题;但该方法未考虑时间维度信息,无法研究热点区域关联性的时空演化特征。刘尧等<sup>[16]</sup>采用神经网络与统计分析方法,分别从行政区划以及道路网络上进行交通事故的时空特征分析;然而,该方法局限于分析交通事故的时空分布,无法应用其他属性信息,定量分析影响交通事故严重程度的因素。彭大芹等<sup>[17]</sup>提出了将手机信令数据融合兴趣点数据来挖掘城市的热点区域并进行功能类型标定,通过基于密度的聚类算法判别热点区域。徐伟<sup>[18]</sup>提出基于机器学习方法的热点区域识别和需求预测框架。借助聚类算法实现对热点区域的识别和划分,借助集成学习实现对出行需求的预测;但该方法存在两个问题:停车规范化问题、调度优化问题。梁卓灵等<sup>[19]</sup>提出了

基于改进谱聚类的热点区域挖掘算法,该算法对比传统方法,其对孤立点不敏感,适用于任意形状的样本空间,能够准确地获得聚类结果;但该方法仍然存在着处理和计算效率低下等问题。毕硕本等<sup>[20]</sup>提出基于多因素排序的出租车乘客候车热点区域挖掘方法,通过OPTICS(Ordering Points To Identify the Clustering Structure)算法对轨迹数据中的停留点提取聚类结果,得到乘客候车热点序列。曲昭伟等<sup>[21]</sup>分析了出行需求的时空分布特性,提出轨迹线密度方法在热点区域内进一步确定热点路段的位置;但该方法未考虑城市路网中各路段之间的关联关系。刘家良等<sup>[22]</sup>针对交通拥堵问题提出了城市出行热点区域出租车调度点配置方法,该方法能够得到大连市出租车调度点的最优配置方案,且具有一定的实时性;但不足之处在于数据源单一,适用场景有限。周博等<sup>[23]</sup>提出了复杂网络视角下的城市热点区域空间交互分析,发现了城市热点区域在空间联系方面表现的一系列规律性信息,如抱团现象、小世界效应和无标度特征。此外,杨文亮等<sup>[24]</sup>结合数据挖掘和可视化技术,研究兰州市城市居民出行规律和城市空间交互特征,发现出行量聚类结构总体呈现与兰州市河谷型地形相匹配的“哑铃”状分布形状。全威等<sup>[25]</sup>提出了基于出租车轨迹数据挖掘的城市区域公交可达性评估优化方法,利用核

密度分析和热点探测,识别出租车用户出行敏感区,并利用出行热点表征区域用户需求,研究敏感区域公交可达性。

### 3 研究方法

#### 3.1 模型框架

基于图模型的热点路段发现研究主要由3个部分构成:基于轨迹数据统计路段流量,根据路段之间的拓扑关系计算每个路段局部密度 $\rho_i$ 和高局部密度点距离 $\delta_i$ ,以及建立决策树以找出聚类中心。本文研究内容的整体框架见图1。首先,基于路段划分及当天的轨迹数据,统计每小时各路段的流入流出数据;其次,以每个路段为单位,根据路段之间拓扑关系,统计每个路段及其一邻域的轨迹点数量,计算每个路段的局部密度 $\rho_i$ ;再计算任意两路段之间的距离并建立距离矩阵,然后计算每个路段的高局部密度距离 $\delta_i$ ;最后根据得到的局部密度 $\rho_i$ 和高局部密度距离 $\delta_i$ 建立以 $\delta$ 为纵轴, $\rho$ 为横轴的决策图,找出 $\delta, \rho$ 均为大值的数据点对应的路段作为聚类中心,同时将每个剩余路段分配到它最近邻且密度比其大的路段所在的簇中。得到聚类中心后,本文根据成都市实际的各个功能区分析其作为热点的可行性。

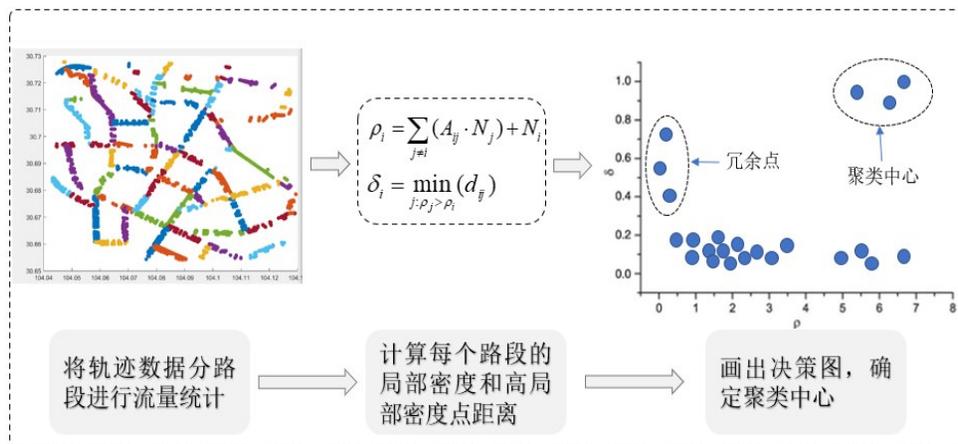


图1 基于图密度峰值聚类算法的出行热点路段发现

#### 3.2 数据预处理

##### 3.2.1 数据集介绍

实验使用的轨迹数据集来自2016年11月成都市区滴滴盖亚数据开放计划的脱敏数据。滴滴出行作为世界领先的公共出行平台,其日均订单量已超过3000万,平台数据库每天存储的轨迹数据已超过

106TB,每日数据处理量达到4875TB,每日出租司机发出的路径规划请求有400亿多次。在保障用户数据的安全和隐私前提下,滴滴平台把一部分出行数据审慎地开放。

原始轨迹数据信息见表1,数据集轨迹点包括司机ID、订单ID、经度、纬度和时间戳,其中纬度范围为

[30.65, 30.73], 经度范围为[104.04, 104.13], 约覆盖了成都市的大部分城区, 时间戳间隔10秒左右。

表1 原始出租车轨迹数据记录详情

Tab.1 Original taxi track data record details

字段	类型	实例	备注
司机ID	String	dkjx.jrrdDo8ntdd	已脱敏
订单ID	String	Jsdj3kKlvst@qnwz	已脱敏
时间戳	String	1521584670	Unix时间戳
经度	String	104.04282	GCS-02坐标
纬度	String	30.71984	GCS-02坐标

### 3.2.2 路段流量统计

由于密度峰值聚类算法要求计算每个数据点的局部密度, 因此首先要计算任意两点之间的距离, 这必然

导致传统的密度峰值聚类算法在面对庞大数据集时的效率十分低下, 正因为这样传统的密度峰值聚类算法面对较大数据时容易出现困难。因此本文以划分路段为基本单元, 根据各路段的拓扑关系建立图模型。一方面, 在计算每个路段的局部密度时能够降低算法的计算复杂度; 另一方面, 由于考虑了各路段之间的关联关系, 能够更准确合理地找出居民的出行热点。

为了能够清晰地探索城市居民出行在各路段的热点, 首先需要对成都市路网进行路段划分, 并基于这种划分对原始的轨迹数据做路段的流量统计。图2(a)为成都市某一天的原始轨迹热图, 研究区域为8 km\*8 km的正方形区域, 将该区域的路网进行路段划分, 划分方法见图2(b), 统计每小时各路段的轨迹点数量。

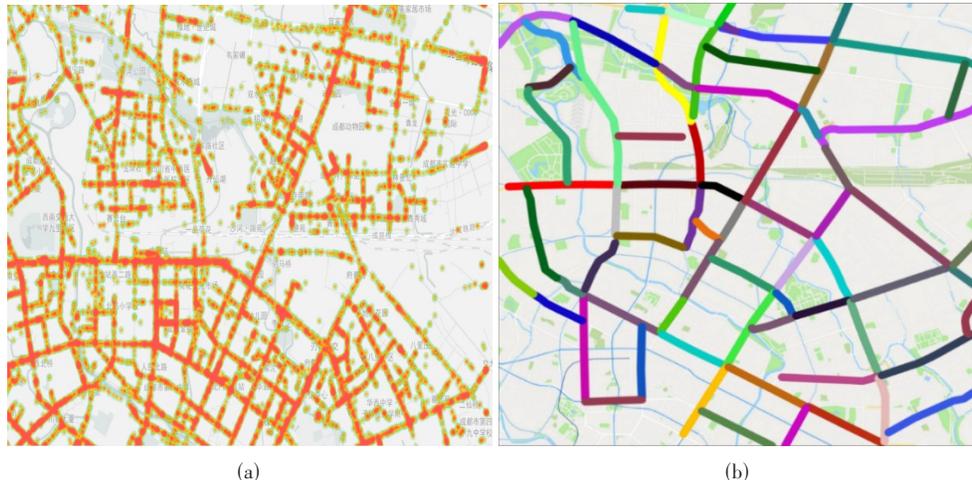


图2 原始轨迹分布热图与成都市区路网的路段划分示意图

### 3.2.3 图密度峰值聚类算法

统计各路段的轨迹点密度数据后, 需计算各路段的局部密度和高局部密度距离。在计算局部密度时, 首先根据路网的拓扑结构构建图模型如下:

$$G = (V, \mathcal{E}) \quad (1)$$

其中  $V$  表示以路段为节点构成的节点集,  $\mathcal{E}$  表示路口连接的路段之间构成的边集。

定义  $A$  为图  $G$  的邻接矩阵, 则矩阵的第  $i$  行、第  $j$  列元素  $A_{ij}$  可以表示为:

$$A_{ij} = \begin{cases} 1, & \text{若 } (v_i, v_j) \text{ 有边} \\ 0, & \text{若 } (v_i, v_j) \text{ 无边} \end{cases} \quad (2)$$

其中  $v_i, v_j$  表示第  $i$  个节点与第  $j$  个节点。

传统的密度峰值聚类算法需要先确定截断距离  $d_c$ , 再统计在截断距离阈值范围内节点  $i$  周围数据点的数量。而图密度峰值聚类算法中, 本文根据图的邻

接矩阵可以使路段的局部密度为该路段的数据点与其邻域路段的数据点之和。具体地, 图密度峰值聚类算法的局部密度计算见式(3)。

$$\rho_i = \sum_{j \neq i} (A_{ij} \cdot N_j) + N_i \quad (3)$$

其中,  $A_{ij}$  为邻接矩阵的第  $i$  行、第  $j$  列元素;  $N_j$  表示路段  $v_j$  中打车数据量;  $N_i$  表示  $v_i$  路段自身中打车数据量。

计算符合邻域的路段中数据量总和加上自身路段内数据量即为第  $i$  个路段的局部密度  $\rho_i$ 。具体示例见图3, 在计算路段 a 的局部密度时, 仅考虑与路段 a 相连的路段, 路段 a 的局部密度为与其相连的路段 1 至 6 的数据量总和加上路段 a 自身的数据量。

当计算所有路段的局部密度  $\rho$  后, 需要寻找每个路段对应的高于自身局部密度的路段的最小距离, 即

高局部密度距离 $\delta$ 。在计算高局部密度距离 $\delta$ 之前,本文计算所有路段的距离矩阵 $D$ ,其中矩阵 $D$ 的第 $i$ 行、第 $j$ 列元素 $d_{ij}$ 表示路段 $i,j$ 之间的距离。那么第 $i$ 路段的高局部密度 $\delta_i$ 计算公式为:

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}) \quad (4)$$

其中 $\rho_i, \rho_j$ 分别为路段 $i, j$ 的局部密度。



图3 局部密度计算方式示意图

#### 算法1:图密度峰值聚类算法

输入:轨迹数据(包含轨迹经纬度与时间戳);成都市地图

输出:聚类结果

1. 将成都市路网按路段划分,并以路段为节点,路口连接的路段构成边建立图,并计算图的邻接矩阵 $A$
2. 以每个路段为基本单元,统计各路段、各时段的流入流出流量,剔除掉数据点远远低于设置阈值的路段
3. 剩余路段中,根据邻接矩阵 $A$ 计算出每个路段的局部密度 $\rho_i$
4. 计算剩余路段中任意两个路段之间的距离,建立距离矩阵并计算出每个路段的 $\delta_i$
5. 以 $\rho$ 为横轴,以 $\delta$ 为纵轴画出 $\rho$ 和 $\delta$ 的决策图,在决策图找出 $\rho_i$ 和 $\delta_i$ 都相对较高的路段,将其设为簇中心,最后将剩余的路段分配到它的最近邻且密度比其大的数据所在的簇

在计算完每个路段的局部密度 $\rho$ 和高局部密度距离 $\delta$ 后,对每个路段进行分析并找出聚类中心,即热点路段。为了更加直观的展现数据的分布,本文以局部密度 $\rho$ 为横轴,以高局部密度距离 $\delta$ 为纵轴做决策图,根据局部密度与高局部密度距离的定义,以及成为聚类中心的两个基本假设:1)聚类中心与其阈值范围内的相邻点相比,局部密度要最大;2)不同类中心之间的距离要相对很远。因此必然从决策图中选用局部密度 $\rho$ 和高局部密度距离 $\delta$ 值都较大的点作为聚类中心。该算法流程见算法1。

表2 最高密度与次高密度聚类中心的局部密度和高局部密度距离

时段	局部密度	高局部密度距离
8:00-9:00	39166	744.8469
8:00-9:00	37144	152.0690
9:00-10:00	44293	744.8469
9:00-10:00	43526	148.0135
10:00-11:00	38923	744.8469
10:00-11:00	38872	98.7927
11:00-12:00	37852	133.9328
11:00-12:00	36066	34.8855
12:00-13:00	37038	37.5898
12:00-13:00	33432	27.6586
13:00-14:00	37545	133.2704
13:00-14:00	31567	152.0690
14:00-15:00	36188	133.9328
14:00-15:00	31809	94.1541
15:00-16:00	30610	744.8469
15:00-16:00	27633	135.4843
16:00-17:00	31413	12.3693
16:00-17:00	26634	175.8664
17:00-18:00	27647	71.0633
17:00-18:00	26646	133.2704
18:00-19:00	21448	53.2353
18:00-19:00	18473	118.8780
19:00-20:00	11874	75.21303
19:00-20:00	9937	74.1080

针对传统的密度峰值聚类算法的缺陷,图密度峰值聚类算法以划分路段为基本单元,不仅不影响聚类结果,还在计算每个路段的局部密度时能够降低算法的计算复杂度;其次,根据各路段的拓扑关系建立图模型,由于考虑了各路段之间的关联关系,能够更准确地合理地找出居民的出行热点。

## 4 实验结果与分析

### 4.1 实验参数设定

实验中使用滴滴-成都轨迹数据集2016年11月1日数据,共包含202503条轨迹数据,数据所覆盖的经纬度范围为:[30.652828, 30.727818], [104.043102, 104.129591]。在统计各路段的流入流出流量时,将一天中的08:00—20:00区间进行统计,参数设置见表3。

表3 基于滴滴-成都轨迹数据集的图密度峰值聚类算法参数

轨迹数据	202 503 条
数据覆盖范围	8 km*8 km
划分路段数	69
划分路口数	51
统计时间间隔	1h

### 4.2 局部密度及高局部密度距离计算

实验中根据式(3)和(4)对局部密度和高局部密

度距离进行计算。如表2所示,按照不同时段最大的两个局部密度,列出了高局部密度距离。

### 4.3 确定聚类中心

为了决定聚类中心,本文以 $\rho$ 为横轴,以 $\delta$ 为纵轴画出 $\rho$ 和 $\delta$ 的决策图。图4为2016年11月1日08:00—09:00的散点图,从图中可以看出右上侧几个散点的 $\rho$ 值和 $\delta$ 值都较大,这几个点对应的路段可以看作是聚类中心。由于其他散点无法直接进行判断,本文设计了结合局部密度 $\rho$ 和高局部密度距离 $\delta$ 的综合考量值 $\gamma$ ,见式(4)。

$$\gamma_i = \rho_i \delta_i \quad (4)$$

其中 $\gamma_i$ 越大,路段*i*越有成为聚类中心的可能。实验中,将所有 $\gamma$ 的平均值181523.345作为阈值,在不同时段成为聚类中心的路段数约为6个。

### 4.4 热点路段发现

实际的成都市区中公园、商业街、学校、住宅区、办公区、车站等区域的分布见图5。本文将不同时段的聚类中心,即热点路段绘制成热图,见图6。

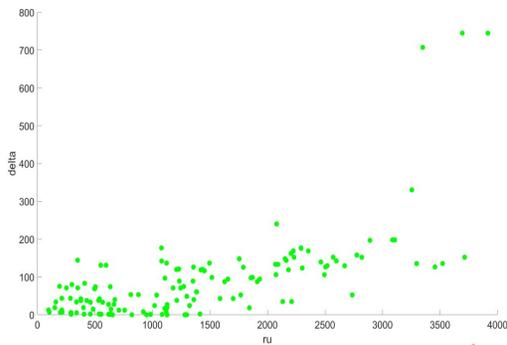


图4  $\rho$ 和 $\delta$ 的决策图

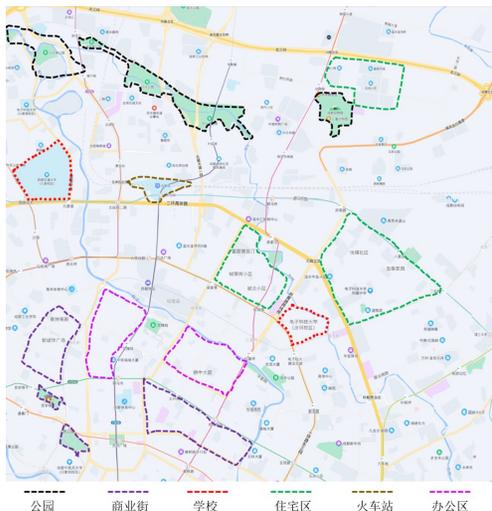


图5 成都市区各功能区分布示意图

从时序的角度来看,在08:00—09:00时段内,热点路段主要集中在北三环附近的住宅区及电子科技大学(沙河校区)附近,09:00—10:00热点路段则集中于解放路二段及北大街,为一些办公楼集中地区及城隍庙地铁站,因此可以判断上午的这段时间,乘客的出行目的可能包括上午通勤和部分的旅游群体。在10:00—11:00内,热点路段仍集中于解放路二段及北大街,但在成都动物园、沙河源公园、西南交大附近路段也逐渐出现峰值,11:00—12:00热点路段基本集中在西南交大和宽窄巷子附近的一些商业街,因此在临近中午乘客出行的热点路段基本往商业街区域移动,少部分则往公园等景点移动。而在下午,12:00—13:00,热点路段又分布在北三环附近,导致该段成为热点或许是因为下午外出旅游人群。在13:00—15:00,热点路段基本分布在一些学校、公园和商业街,包括成都动物园、西南交大、电子科大、奥塔莱斯、宽窄巷子附近路段,而在15:00—20:00办公区域附近路段逐渐出现峰值,导致该热点的出现可能的原因为乘客的晚通勤。

按功能区的角度来看,位于学校附近的路段基本在大部分时段都处于热点,考虑到大学生的活动较其他群体更为自由,且大学生的出行受工作日、非工作日的影 响较小,因此该功能区附近路段的打车需求量大。部分住宅区,如电子科技大学(沙河校区)附近住宅区,则会受到学校的影响大部分时段都处于热点,位于北三环附近的住宅区则在早晚时段会出现较大峰值,主要原因可能源于早晚通勤群体。此外位于西南交大、宽窄巷子以及城隍庙等附近路段则主要在下午和晚上时段出现峰值,由于旅游群体一般会选择下午出行,且考虑到晚上下班后群体以及学生都会偏向于商业区,导致商业区在该时段人口密度较大,乘客打车需求量增大。最后,位于成都站附近路段,几乎全天都出现峰值,显然火车站附近的打车需求较别的功能区而言,受时间影响较小。

### 4.5 计算效率分析

本文将图密度峰值聚类算法与传统的密度峰值聚类算法及k-means算法在滴滴-成都轨迹数据集上进行了计算效率对比实验,结果见表4。可以看出,本文提出的图密度峰值聚类算法在计算效率上有大幅度提升,这主要归因于本文算法避免了传统算法计算任意两轨迹点之间的距离带来的时间损耗,而基于路段划分的图密度峰值聚类算法可以将位于同一路段的轨迹点看作一个基本单元,大大简化了算法的时间复杂度,适用于很多有实时需求的场景。

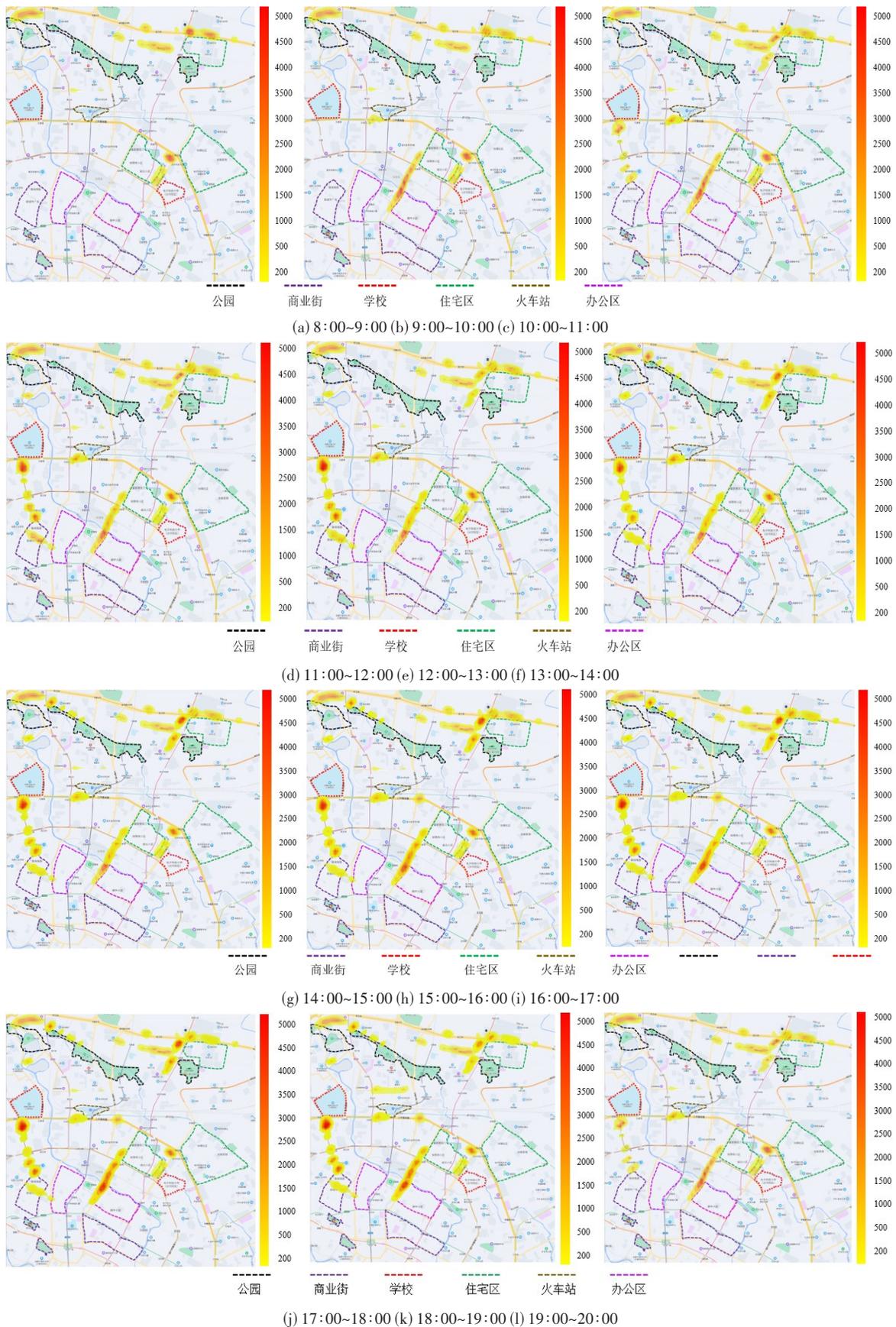


图6 成都市区2016年11月1日08:00—20:00热点路段

表4 图密度峰值聚类算法与传统聚类算法的计算效率对比

算法名称	计算时间/s
密度峰值聚类算法	593.2911
k-means算法	409.2416
图密度峰值聚类算法	32.80

## 5 结论

传统的密度峰值聚类算法具有较高的计算复杂度,且传统算法未考虑路网固有的拓扑结构,无法衡量各路段之间的关联关系。针对上述问题,本文提出基于图密度峰值聚类算法的出行热点路段发现方法,将路网以图结构引入该算法,图中以路段为基本单元不仅可以大幅度提升算法的计算复杂度,而且可以更加准确合理的找出热点路段。通过在滴滴-成都轨迹数据集的实验表明,图密度峰值聚类算法具有更高的热点路段发现精度,并且在计算效率上有大幅度提升。

## 参考文献(References):

- [1] QU Z W, WANG X. Location Optimization for Urban Taxi Stands Based on Taxi GPS Trajectory big data [J]. IEEE Access, 2019, 6: 62273-62283.
- [2] LIU D, CHENG S F, YANG Y. Density peaks clustering approach for discovering demand hot spots in city-scale taxi fleet dataset [C]. IEEE 18<sup>th</sup> International Conference on Intelligent Transportation Systems, 2015: 1831-1836.
- [3] SUN C, QUAN W. Evaluation of bus accessibility based on hotspot detection and matter-element analysis [J]. IEEE Access, 2020, 8: 138800-138809.
- [4] 何月,王崇倡.基于时空聚类的出租车载客热点区域挖掘研究[J].测绘与空间地理信息,2020,43(1):99-102.
- [5] YUAN J, ZHENG Y, ZHANG L, et al. Where to find my next passenger [C]. Proceedings of the 13<sup>th</sup> International Conference on Ubiquitous Computing, 2011: 109-118.
- [6] SHEN Y, ZHAO L, FAN J. Analysis and visualization for hot spot based route recommendation using short-dated taxi GPS traces [J]. Information, 2016, 6(2): 134-151.
- [7] QU Z, WANG X, SONG X, et al. Location optimization for urban taxi stands based on taxi GPS trajectory big data [J]. IEEE Access, 2019, 7: 62273-62283.
- [8] LI X, PAN G, WU Z, et al. Prediction of urban human mobility using large-scale taxi traces and its applications [J]. Frontiers of Computer Science, 2012, 6(1): 111-121.
- [9] LIU L Q, WU C Z, ZHANG H, et al. Research on taxi drivers' passenger hotspot selecting patterns based on GPS data: A case study in Wuhan [C]. 4<sup>th</sup> International Conference on Transportation Information and Safety (ICTIS), 2017-432-441.
- [10] DOW C R, CHEN H C, HWANG S F. A hotspot aware taxi zone queuing system [C]//2018 International Conference and Workshop on Computing and Communication (IEMCON). IEEE, 2016: 1-4.
- [11] DOW C R, NGUYEN D B, CHEN H C, et al. An adaptive and hotspot aware taxi zone queuing system on internet of vehicles [J]. International Journal on Semantic Web and Information Systems, 2017, 13(3): 89-106.
- [12] XIE X, YU Y, JIANG Y. Passenger hotspot mining based on taxi GPS data-taking Chengdu as an example [M]. ICTE 2019. Reston, VA: American Society of Civil Engineers, 2020: 883-892.
- [13] VERMA N, BALIYAN N. PAM clustering based taxi hotspot detection for informed driving [C]. 8th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2017: 1-7.
- [14] 盛宇裕,毕硕本,王国杰,等.运用循环神经网络模型预测城市热点区域[J].武汉大学学报(信息科学版),2021.
- [15] 陈锐,陈明剑,姚翔,等.利用导航大数据挖掘城市热点区域关联性[J].地球信息科学学报,2019,21(6):826-835.
- [16] 刘尧,王颖志,王立君,等.交通事故的时空热点分析[J].浙江大学学报(力学版),2020,47(1):52-59.
- [17] 彭大芹,罗裕枫,江德潮,等.基于移动信令数据的城市热点识别方法[J].重庆邮电大学学报(自然科学版),2019,31(1):95-102.
- [18] 徐伟.基于机器学习的共享单车热点区域识别及需求预测[J].交通发展改革专题,2019,41(5):1-6.
- [19] 梁卓灵,元昌安,覃晓,等.基于改进谱聚类的热点区域挖掘方法[J].重庆理工大学学报(自然科学版),2021,35(1):129-137.
- [20] 毕硕本,徐瑞壮,万蕾,等.基于多因素排序的南京市出租车乘客候车热点区域挖掘方法[J].中国科技论文,2020,15(1):23-30.
- [21] 曲昭伟,王鑫,宋现敏,等.基于出租车GPS大数据的城市热点出行路段识别方法[J].交通运输工程与信息,2019,19(2):238-246.
- [22] 刘家良,孙立双.城市出行热点区域的出租车调度点配置[J].中国科技论文,2018,13(9):1012-1017.
- [23] 周博,马林兵.复杂网络视角下的城市热点区域空间交互分析[J].地理空间信息,2021,19(6):115-120.
- [24] 杨文亮,冯慧芳.基于出租车GPS轨迹的城市区域时空交互特征分析[J].计算机与现代化,2021,305:87-93.
- [25] 全威,孙超.基于热点探测的城市区域公交可达性研究[J].交通运输系统工程与信息,2020,20(2):231-236.