Statistics and Application 统计学与应用, 2025, 14(8), 427-440
Published Online August 2025 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2025.148247

Hans汉斯

基于聚类算法的城市交通热点挖掘及居民出行时空需求分析

——以成都市出租车轨迹数据为例

王伟鑫[1]，康　硕[2]

[1]四川外国语大学国际工商管理学院，重庆
[2]河北工业大学经济管理学院，天津

收稿日期：2025年7月26日；录用日期：2025年8月16日；发布日期：2025年8月28日

摘　要

为精准感知城市交通运行状态并深度洞察居民出行规律，本文以成都市出租车GPS轨迹数据为例，提出了一套基于聚类算法的城市交通“热点路段”挖掘、量化及应用分析框架。研究首先对原始轨迹数据进行预处理，提取高质量的OD (起讫点)数据集；随后，分别采用DBSCAN与K-means算法进行交通热点路段挖掘，并通过轮廓系数、离群点率等多个指标对比验证了DBSCAN算法在识别路网形态热点上的优越性；在此基础上，构建了综合出行规模与空间密度的热度量化模型，对识别出的热点路段进行了定量评估。研究结果表明：1) 成功识别出人民南路三段等多个关键交通热点，并发现其热度与周边功能属性高度相关；2) 居民出行在工作日与休息日呈现显著不同的高峰时段特征，且工作日高峰时段的OD点在空间上与通勤规律高度吻合。本研究不仅为利用交通大数据进行精细化城市分析提供了方法论参考，也为交通资源优化配置、拥堵疏导及城市规划提供了数据驱动的决策支持。

关键词

轨迹数据，热点聚类，交通拥堵，居民出行时空需求

Mining Urban Traffic Hotspots and Analyzing Spatiotemporal Travel Demand of Residents Based on Clustering Algorithms

—A Case Study of Taxi Trajectory Data in Chengdu

Weixin Wang[1], Shuo Kang[2]

[1]School of International Business and Management, Sichuan International Studies University, Chongqing
[2]School of Economics and Management, Hebei University of Technology, Tianjin

文章引用: 王伟鑫, 康硕. 基于聚类算法的城市交通热点挖掘及居民出行时空需求分析[J]. 统计学与应用, 2025, 14(8): 427-440. DOI: 10.12677/sa.2025.148247

Received: Jul. 26th, 2025; accepted: Aug. 16th, 2025; published: Aug. 28th, 2025

Abstract

To accurately perceive the operational status of urban traffic and gain deep insights into residents' travel patterns, this paper proposes a framework for mining, quantifying, and analyzing urban traffic "hotspot road segments" based on clustering algorithms, using GPS trajectory data from taxis in Chengdu as an example. The study first preprocesses the raw trajectory data to extract a high-quality OD (origin-destination) dataset. Subsequently, DBSCAN and K-means algorithms are employed to mine traffic hotspot road segments, and the superiority of the DBSCAN algorithm in identifying hotspots in road network patterns is verified through multiple indicators such as the contour coefficient and outlier rate. Based on this, a heat quantification model that integrates travel scale and spatial density is constructed to quantitatively evaluate the identified hotspot road segments. The research results show that: 1) Several key traffic hotspots, such as the third section of Renmin South Road, have been successfully identified, and their heat is highly correlated with surrounding functional attributes; 2) Residents' travel patterns exhibit significantly different peak period characteristics on weekdays and weekends, and the OD points during peak periods on weekdays are highly spatially consistent with commuting patterns. This study not only provides a methodological reference for utilizing traffic big data for refined urban analysis but also offers data-driven decision support for optimizing traffic resource allocation, easing congestion, and urban planning.

Keywords

Trajectory Data, Hotspot Clustering, Traffic Congestion, Residents' Spatiotemporal Travel Demand

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

城市化进程的加速与机动化水平的提升，使得交通拥堵、出行效率低下已成为全球各大城市面临的共同挑战，严重制约了城市的可持续发展与居民生活品质的提升[1] [2]。在此背景下，精准感知城市交通运行状态，深入洞察居民出行规律，对于优化交通资源配置、制定科学的交通管理策略至关重要[3]。智能交通系统的发展，特别是基于位置服务技术的普及，催生了海量的城市交通大数据，其中，出租车 GPS 轨迹数据因其覆盖广、时效性强、蕴含丰富的出行信息，为城市交通研究提供了宝贵的数据源[4] [5]。

挖掘城市交通中的“热点”，即交通活动高度集中的区域或路段，是理解城市交通供需关系、识别交通瓶颈、优化交通网络布局的基础[6]。传统的热点分析多集中于静态的兴趣点或区域，而对于动态交通流作用下形成的“热点路段”及其“热度”的精确识别与量化，是当前研究的重点与难点[7]。识别这些热点路段，不仅能够揭示交通流在路网上的集聚效应，还能为交通信号优化、拥堵疏导、警力部署等提供直接依据。进一步地，基于这些热点路段分析居民的出行时空需求，即探究居民在不同时间往来于这些热点路段的行为模式，对于理解城市功能区互动、优化公共交通线路、提升出租车运营效率具有深远意义[8] [9]。

聚类算法作为无监督学习的核心技术，能够有效地从大规模数据中发现内在结构和模式，已被广泛应用于交通数据挖掘领域[10]。例如，DBSCAN 算法因其能发现任意形状的簇且对噪声不敏感，常用于

识别上下客热点区域[11]；K-means 算法则因其简洁高效，也常被用于初步的交通模式划分[12]。然而，现有研究在以下方面仍有提升空间：首先，针对“热点路段”的精细化挖掘方法尚不成熟，多数研究侧重于点状或面状热区，对线性路段特征的关注不足；其次，对热点路段“热度”的量化方法有待深化，如何综合考虑交通流量、时空聚集性等因素进行科学评估是一个挑战[13] [14]；再次，将热点路段挖掘与居民出行时空需求进行深度耦合分析，特别是探究围绕这些特定路段的出行生成与吸引规律的研究相对较少；最后，不同聚类算法在交通热点路段挖掘任务中的适用性与表现对比[15] [16]，及其与核密度估计等方法结合进行热度量化的研究也需进一步探索。

针对上述问题，本文以快速发展的特大城市——成都市为例，利用其海量的出租车 GPS 轨迹数据，旨在构建一套基于聚类算法的城市交通热点路段挖掘与热度量化分析框架，并在此基础上深入剖析与这些热点路段相关的居民出行时空需求特征。基于上述分析，本文首先对出租车轨迹数据进行精细化预处理，提取准确的上下客事件及行驶路径信息；其次，分别采用 DBSCAN 和 K-means 聚类算法，结合路网信息，进行交通热点路段的挖掘与识别，并对比分析两种算法的优劣；再次，引入核密度估计等方法对识别出的热点路段进行热度量化，评估其交通承载压力与活跃程度；最后，基于挖掘出的热点路段，系统分析居民出行的空间分布特征、时间分布规律以及热点区域的功能属性。本研究不仅期望为成都市的交通规划与管理提供数据驱动的决策支持，也试图为其他城市利用轨迹大数据进行精细化交通分析提供方法论参考。

2. 交通热点路段挖掘及热度量化分析

2.1. 问题描述

随着城市化进程的深化，交通拥堵已成为制约城市发展与居民生活品质的瓶颈。传统交通数据采集方法因其局限性，难以满足现代城市精细化管理的需求。以出租车 GPS 轨迹为代表的城市感知大数据，凭借其高时空分辨率和丰富的出行语义，为深度洞察城市交通规律提供了全新视角。然而，将海量原始轨迹数据转化为有价值的决策依据，仍面临若干挑战。现有研究虽已取得进展，但在以下方面亟待深化：首先，多数研究侧重于识别点状或面状的“热点区域”，而对承载实际交通流、更具路网特征的线状“热点路段”的精确挖掘尚不充分，这限制了对交通瓶颈的精准定位。其次，针对已识别热点的“热度”评估，缺乏一套能综合 OD 点数量与空间集聚特征的科学量化方法，导致难以对不同路段的重要性进行客观比较。此外，不同聚类算法在挖掘“热点路段”这一特定任务上的适用性与性能差异，也需要更系统的实证对比与分析。为应对上述挑战，本研究旨在聚焦于城市交通“热点路段”的挖掘与量化分析。

2.2. 数据来源与描述

出租车 GPS 轨迹数据来源于 DataCastle 交通线路通达时间预测算法竞赛，共包含成都市出组车轨迹数据约 20,000 条。采样日期为 2014 年 8 月 3 日和 8 月 4 日，分别代表休息日和工作日。采样的时间点为每天早上 6:00 到凌晨 0:00。虽然该数据距今已有十余年，但是这些年成都市各区的城市格局并未发生较大的改变。因此，该数据在今天仍有一定的研究意义，为挖掘交通热点和居民出行的规律提供了数据支持。

轨迹数据的具体信息如表 1 所示，包括 id (出租车标识)、lat (采样时出租车所处的纬度)、lon (采样时出租车所处的经度)、status (载客状态)，其中“0”表示采样时出租车处于空载状态，“1”表示采样时出租车已搭载乘客、day (采样时间)。

2.3. 数据预处理

本文中使用到的数据包括出租车完成订单的所有位置。为实现热点路段挖掘的目的，我们需要将

Table 1. Structure of taxi GPS trajectory data table
表 1. 出租车 GPS 轨迹数据表结构

字段	含义	例子
id	出租车标识	13,920
lat	纬度	30.5707
lon	经度	103.957
status	载客状态	0
day	采样时间	2014-08-04 21:47:13

其中的起始点和结束点提取出来，合成 OD 数据集。该数据集忽略了出租车的行驶路径，简化了研究过程。以 OD 数据集为研究对象进行时间和空间上的研究，有助于我们了解居民出行时间规律以及城市热点。如图 1 所示，当载客状态由 0 变成 1 时，说明此刻该出租车的状态由空载变成了载客，我们称这个状态改变的点为 O 点。如：轨迹段 a 向轨迹段 b 转换的轨迹点；当载客状态由 1 变成 0 时，说明此刻该出租车的状态由载客变成了空载，我们称这个状态改变的点为 D 点。如：轨迹段 b 向轨迹段 c 转换的轨迹点。对 2014 年 8 月 3 日和 4 日成都市出租车轨迹数据进行分析，共提取到 O 点 794 个，D 点 792 个。

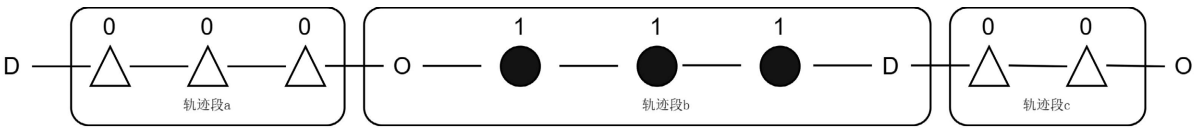


Figure 1. Explanation of OD points
图 1. OD 点说明

2.4. 热点路段挖掘与量化模型

2.4.1. 基于 DBSCAN 的交通热点挖掘

1) 确定参数(eps, MinPts)的范围

决定 DBSCAN 聚类结果优劣的参数主要有搜索半径 eps 和以距离阈值为半径的样本个数阈值 MinPts。为找到二者的最优组合，本文参考学者们的研究方法并结合实际情况设定了 eps 为浮点数，在 1 到 5 之间，MinPts 为整数，在 2 到 10 之间。图 2 表示在确定不同的样本个数阈值的情况下，调节距离阈值所造成的聚类热点个数变化。由图可得出：聚类热点的个数基本上随着 eps 的增大而减少。当 eps 小于 1.5 时，可能导致过拟合，将噪声点视为簇；当 eps 大于 3 时，聚类热点的个数偏少，导致丢失部分重要信息，出现欠拟合现象。因此，eps 的取值在 1.5~3 之间更有效。图 3 展示了在确定距离阈值处于 1.5~3 之间的情况下，调节样本个数阈值所造成的聚类热点个数改变。由图中我们可以看出：聚类热点的个数随着 MinPts 的增大而减少。当 MinPts 大于 4 时，聚类热点个数较少，簇划分的较为粗糙，可能导致结果欠拟合。此外，由于研究的是交通热点问题，所以本文认为一个路段在一定时间内至少产生 2 个及以上的点才能视为热点。因此，为使聚类结果更有效，MinPts 的取值范围应该在 2~4 之间，既能避免过拟合，又能避免欠拟合。

2) 参考轮廓系数缩小参数范围

轮廓系数是判断 DBSCAN 聚类结果优劣的一个重要指标。分别计算上一步骤中不同参数组合下的轮廓系数并将其作为确定参数大小的重要依据，有助于达到最优的聚类效果。计算轮廓系数 $S(i)$的公式为：

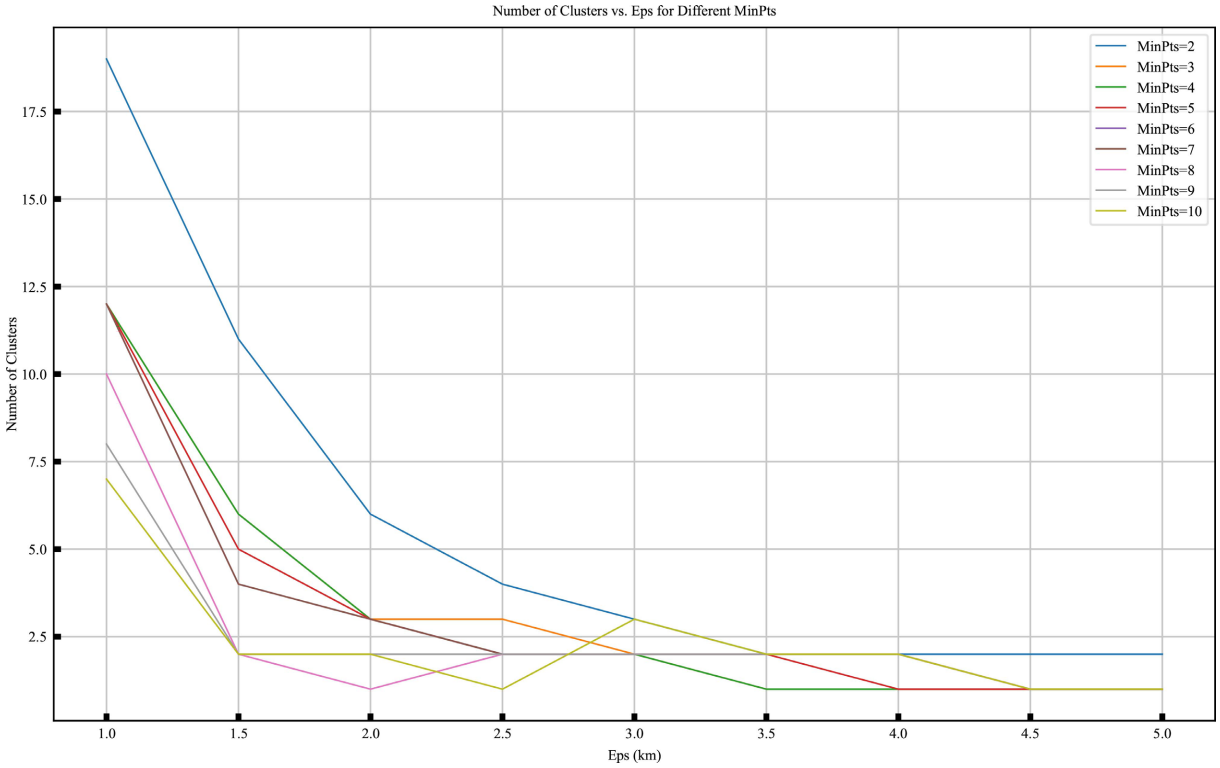


Figure 2. Change in the number of clustering hotspots caused by adjusting eps
图 2. 调节 eps 造成的聚类热点个数变化

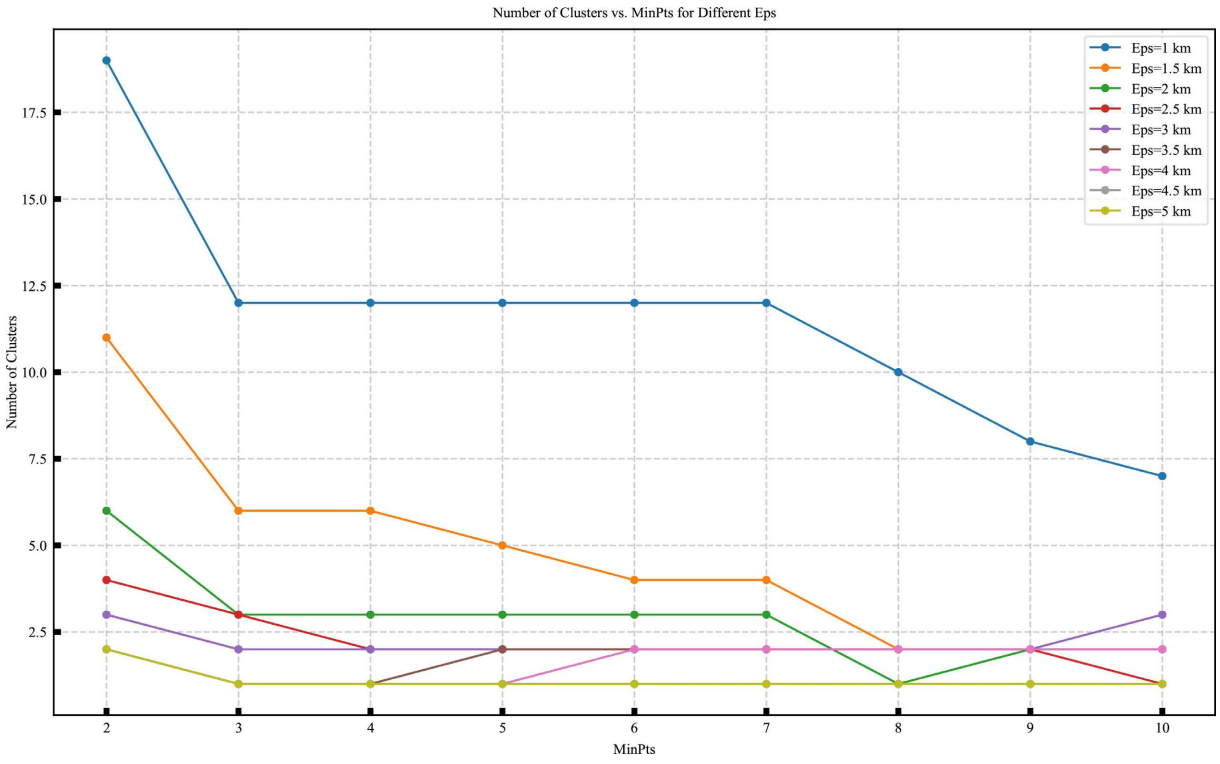


Figure 3. Change in the number of clustering hotspots caused by adjusting MinPts
图 3. 调节 MinPts 造成的聚类热点个数变化

$$S(i)=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{1}$$

$a(i)$表示点 i 到他所属的簇中其他所有样本点的平均距离，$b(i)$表示点 i 到与它相邻最近的簇中所有样本点的平均距离。

假设带聚类的数据集中共有 N 个点，那么所有样本轮廓系数的平均值 SM 就称为聚类结果的总轮廓系数。即

$$SM=\frac{1}{N}\sum_{1}^{N}S(i) \tag{2}$$

由公式(1)可以看出，轮廓系数的取值在−1 到 1 之间，这个值越接近 1，证明不同簇的样本点之间的距离远大于同一簇内样本点之间的距离，即聚类效果更优。通常情况下，当轮廓系数小于 0.3 时，我们认为聚类效果不太理想，代表性不强，可能需要重新调整参数范围；轮廓系数在 0.3~0.7 之间，表明聚类结果虽然可能存在噪声，但结果较好且基本有效，可以用于下一步研究；若轮廓系数大于 0.7，则可以认为聚类效果非常好，同一簇内非常紧密且不同簇间有着明显的分离。

3) 依据离群点率缩小参数范围

离群点率指的是离群点数量占总样本量的百分比，这些点也被称为噪声。离群点越多，则说明原始数据的使用率越低。因此我们优先选择离群率低的聚类参数组合。

4) 依据戴维森堡丁指数缩小参数范围

davies-bouldin 属于一种内部评估指标，这个指数通过判断同一簇内的紧密程度和不同簇间的分离程度来评估聚类算法的效果。DBI 越小，则说明同一簇内紧密且不同簇间有着明显的分离，即聚类效果越好。

各项聚类效果评价指标计算过程如下：

```
#计算轮廓系数(仅在有多个簇时计算)
if len(set(labels)) > 1：#至少有两个簇
   silhouette_avg = silhouette_score(points, labels, metric = haversine_distance)
else:
   silhouette_avg = np.nan
#计算离群点率
n_noise = list(labels).count(-1)#离群点的数量
outlier_ratio = n_noise/len(labels)
#计算戴维森堡丁指数（仅在有多个簇时计算）
if len(set(labels)) > 1：#至少有两个簇
   dbi = davies_bouldin_score(points, labels)
else:
   dbi = np.nan
```

表 2 为聚类效果的各项评价指标结果。由该表我们可以看出：综合考虑各个评估指标，当 eps 取 2，MinPts 取 2 时，轮廓系数为 0.3512，离群点率为 0.32%，DBI 值为 1.4734，可以认为聚类结果基本有效。

2.4.2. 基于 K-Means 的交通热点挖掘

使用 K-means 聚类算法分析时，首先应确定聚类数 k 的大小。本文参考学者们的研究方法并且结合实际情况，采用肘部法则确定 k 值。该法则的核心思想是通过绘制出不同聚类数 k 值所对应的误差平方和曲线，观察这个曲线的“肘部位置”。所谓“肘部位置”，即指：随着 k 值的不断增大，误差平方和

Table 2. Results of various indicators
表 2. 各项指标结果

搜索半径	样本个数阈值	轮廓系数	离群点率	戴维森堡丁指数
1.5	2	0.1866	0.0057	1.9013
1.5	3	0.3154	0.0120	1.3066
1.5	4	0.3158	0.0132	1.3235
2	2	0.3512	0.0032	1.4734
2	3	0.4436	0.0069	1.4440
2	4	0.4436	0.0076	1.3743
2.5	2	0.3883	0.0013	2.1734
2.5	3	0.3915	0.0025	3.9111
2.5	4	0.5684	0.0050	1.9335
3	2	0.5741	0.0013	2.4406
3	3	0.5920	0.0025	4.9904
3	4	0.5920	0.0025	4.9904

会逐渐减小，但是在某一点处，误差平方和的下降幅度会由急剧下降转化为缓慢下降。我们称这个转变的点为该曲线的“肘部”。肘部法则适用于各种数据集且方便直观，但是主观性较强。此外，当曲线图没有明显的“肘部”时，我们还需要结合轮廓系数法、交叉验证法等方法来确定最优的 k 值。图 4 为 OD 数据集的 SSE 图，横轴表示聚类簇数 k，纵轴表示所有聚类簇中的样本与聚类中心点之间的距离平方和的总和 TWSS。当最佳聚类簇数 k 接近肘部时，TWSS 下降的幅度会明显减小。

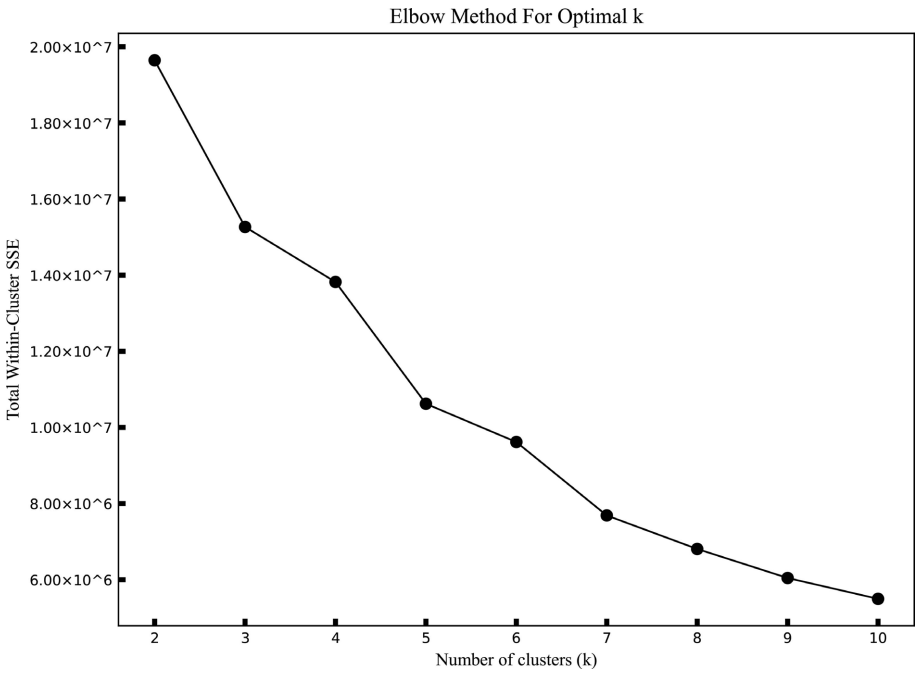


Figure 4. SSE diagram
图 4. SSE 图

结合图 4 和 DBSCAN 参数设置，本文选择 6 作为聚类个数 k 值。经过 Python 运行后得出当迭代 30 次时，聚类效果达到最优。K-means 聚类效果评价指标计算过程如下：此时轮廓系数为 0.3408，离群点率为 7.44%。

```
            #计算轮廓系数
            average_silhouette_score = silhouette_score(distance_matrix,kmeans.labels_)
            print(“平均轮廓系数:",average_ silhouette_score)
            #计算离群点率
            distances = [ ]
for i in range (len (original _ data)):
      cluster_index = kmeans. labels_[i]
      cluster center = original data[np.where (kmeans.labels_== cluster index) [0][0]]#获取该聚类的中心点
      distance = haversine (original _ data [i] [0], original _ data [i] [1], cluster _ center [0], cluster _ center [1])
      distances.append (distance)
      threshold = np.mean (distances) + np.std (distances) #设定阈值为距离的平均值加一个标准差
       outliers = [d for d in distances if d > threshold]
       outlier_rate = len (outliers) / len (data
       print (“离群点率：", outlier_rate)
```

将 K-means 聚类结果与地理数据相结合，得到如图 5 所示。

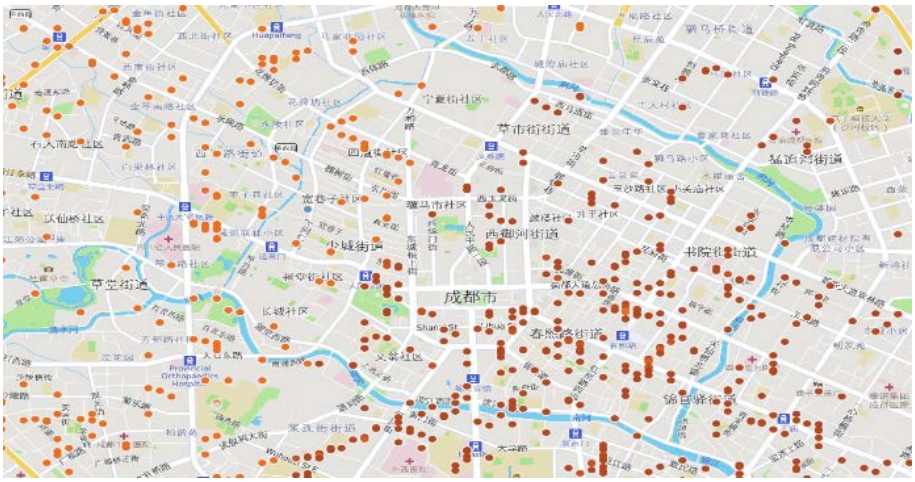


Figure 5. K-means clustering results combined with geographic data
图 5. K-means 聚类结果结合地理数据

为客观评估两种算法的性能，本节对 DBSCAN 与 K-means 的聚类结果进行综合对比分析。从量化指标上看，K-means 聚类结果的轮廓系数(0.3408)显著低于 DBSCAN (0.3512)，表明其在簇内紧密性和簇间分离度上的表现均较差。究其原因，这与两种算法的内在机制密切相关。如图 5 所示，K-means 将所有样本点(包括噪声)强制分配至某一簇，导致了簇内样本分散、边界模糊的现象。相比之下，DBSCAN 算法表现出显著的优越性：首先，其基于密度的特性能够有效识别并剔除噪声点，从而得到更为纯净、凝聚的热点区域；其次，它能够发现并拟合沿道路网络分布的、任意形状的簇结构，这与交通热点的实际空间形态高度契合。

综上所述，无论是在聚类质量的量化评估，还是对交通热点空间形态的拟合能力上，DBSCAN 算法均是更适用于本研究场景的选择。因此，后续的热点热度量化及居民出行需求分析，将全部基于 DBSCAN

的聚类结果展开。

2.4.3. 基于核密度估计的热点热度量化

1) DBSCAN 聚类中心结合街道

按照参数设置进行 DBSCAN 聚类，并得到聚类中心的坐标如下表 3 所示。通过 Tableau 结合开源定位软件将上述地理坐标映射到街道地图中，得到的结果如下表 4 所示。

Table 3. Geographical coordinates of clustering centers
表 3. 聚类中心地理坐标

簇编号	中心纬度	中心经度
0	30.64504593	104.0624625
1	30.7323	104.1461429
2	30.69695	104.1645
3	30.8039	103.888
4	30.7483	103.9575
5	30.765	103.9348571

Table 4. Names of hot travel areas
表 4. 热点出行区域名称

簇编号	所属路段
0	人民南路三段
1	熊猫大道
2	华泰路
3	蜀信西路
4	迪康大道
5	百草路

2) 交通热点区域热度量化

本文旨在构建一个量化模型以评估各热点的相对“热度”。一个科学的热度指标需综合反映区域交通的两个核心维度：1) 出行需求的规模，即 OD 点的绝对数量；2) 出行需求的集中程度，即 OD 点的空间分布密度。为有效融合上述两个维度，本研究构建了如式(1)所示的热度量化模型。该模型的设计逻辑如下：首先，采用簇内各样本点到其质心距离的倒数之和来表征空间密度，这种方法能有效增强近距离点的影响力并具备对异常值的鲁棒性。其次，将密度指标与簇内样本总数 n 相乘，从而将规模因素纳入考量。最后，对结果取平方根，以调整热力值的量纲，避免因数值过大或过小而影响可比性。综上所述，我们可以设置如下化简后的热力值计算公式：

$$p_m = \sqrt{\left\{ n \Big/ \sum_{i=1}^{n} d\left(i, c_m\right) \right\} \times n} \tag{3}$$

其中，p_m 表示第 m 簇的热力值；n 表示该簇所包含的样本点总数；$d\left(i, c_m\right)$ 表示该簇中第 i 个样本点到簇中心 c_m 的距离。

将以上六个热点区域根据该公式进行热力值计算，得到的结果如下表 5 所示。其中，在编号为 3 的簇中同一辆出租车的 O 点与间隔 30 秒后的 D 点经纬度完全相同，我们认为这种点属于刷单行为。即司机接到自己发出的订单，然后快速完成且进行好评，以此来提升自己的业绩和好评率。因此，在统计热力值时，我们忽略编号为 3 的簇。

Table 5. Thermal value table of hot spot areas
表 5. 热点区域热力值表

簇编号	所属路段	中心纬度	中心经度	热力值
0	人民南路三段	30.64504593	104.0624625	63.0709249
1	熊猫大道	30.7323	104.1461429	0.365256605
2	华泰路	30.69695	104.1645	0.743761378
4	迪康大道	30.7483	103.9575	1.199277829
5	百草路	30.765	103.9348571	0.537022116

2.4.4. 热度模型有效性验证

为检验本文所构建热度量化模型(公式(3))的科学性与有效性，需验证其是否能准确反映交通热点的真实“热度”。本节选取了两个公认的、物理意义明确的基准指标，通过对比分析来验证模型的性能。基准指标选择如下：

基准一：簇内 OD 点总数：这是衡量交通活动规模最直接的指标，即每个热点簇内的 OD 点数量 n。一个有效的热度模型应与 OD 点总数呈现强正相关。

基准二：核密度估计峰值：该指标用于衡量空间集聚的峰值强度。对每个热点簇内的 OD 点进行独立的核密度估计(Kernel Density Estimation, KDE)，并提取其密度最大值。该值代表了该区域最核心位置的 OD 事件密集程度。一个有效的热度模型也应与 KDE 峰值正相关。

计算每个有效热点簇的上述两个基准指标，并将其与本文提出的原始热度值 p 进行皮尔逊相关性分析(Pearson Correlation Analysis)。结果如表 6 和表 7 所示。

Table 6. Heat values and benchmark indicator values of each hotspot cluster
表 6. 各热点簇的热度值与基准指标值

簇编号	所属路段	原始热力值(p)	OD 点总数(n)	KDE 峰值
0	人民南路三段	63.07	112	15.82
1	熊猫大道	0.37	7	1.15
2	华泰路	0.74	4	1.89
4	迪康大道	1.20	5	2.53
5	百草路	0.54	7	1.46

Table 7. Correlation analysis between heat model and benchmark indicators
表 7. 热度模型与基准指标的相关性分析

对比项	皮尔逊相关系数(r)	P 值
原始热力值(p) vs. OD 点总数(n)	0.998	<0.001
原始热力值(p) vs. KDE 峰值	0.989	<0.001

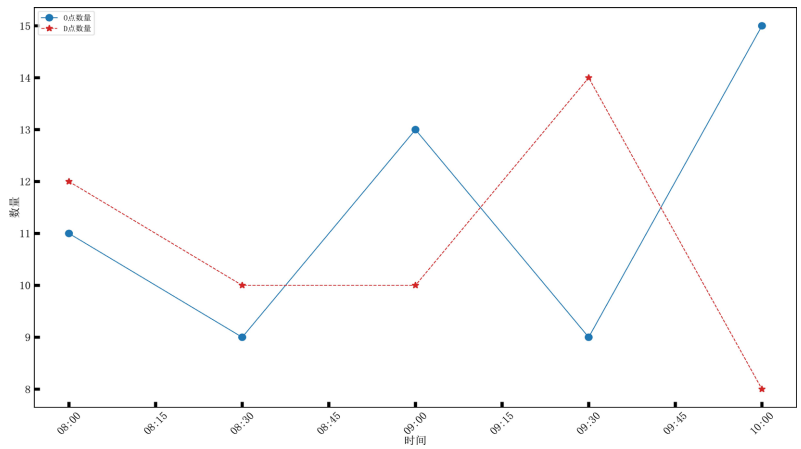
相关性分析结果显示，本文提出的热度量化模型计算出的原始热力值(p)与 OD 点总数(n)的皮尔逊相关系数高达 0.998，与 KDE 峰值的相关系数也达到了 0.989，且 P 值均远小于 0.01，表明存在极强的统计显著性正相关关系。

3. 基于交通热点的居民出行时空需求及区域分析

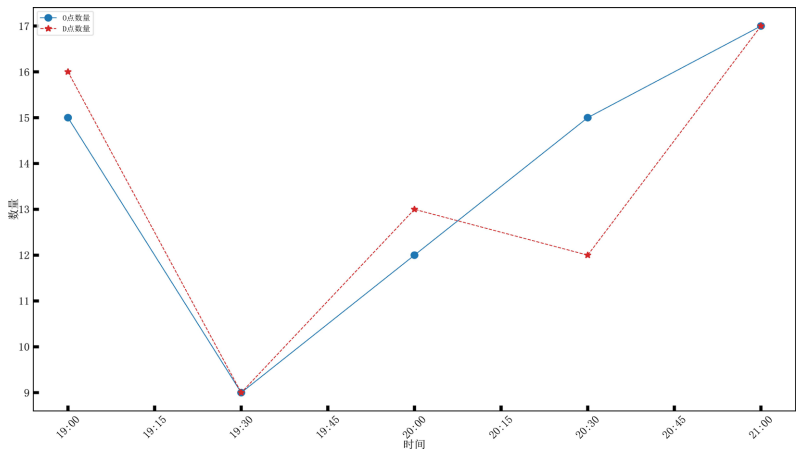
深入了解居民出行的时空需求不仅有利于政府部门进行更加完善的城市规划建设，能缓解城市交通压力、便利居民出行并提升市民生活幸福感。从空间、时间和出行区域三个方面进一步分析居民出行需求。

3.1. 居民出行时间需求分析

在本节我们将对以上这些热点时间段进行细分如图 6 和图 7 所示。对工作日高峰时段的细粒度剖析揭示了其内部更为复杂的出行需求结构。早高峰呈现双阶段特征：9:30 前集中的到达需求(D 点)与 10:00 左右的出发需求(O 点)分别对应了传统通勤与商业服务业等不同职业群体的出行节律。午后 16:00~16:30 的 O 点小高峰则与学生放学时间高度吻合，而晚高峰则清晰地展现了通勤返程链条，即 19:00 后的出发潮与 20:00 左右的到达潮。这种对出行需求在时间维度上的精准刻画，为交通管理部门实施动态优化策略提供了实证依据。

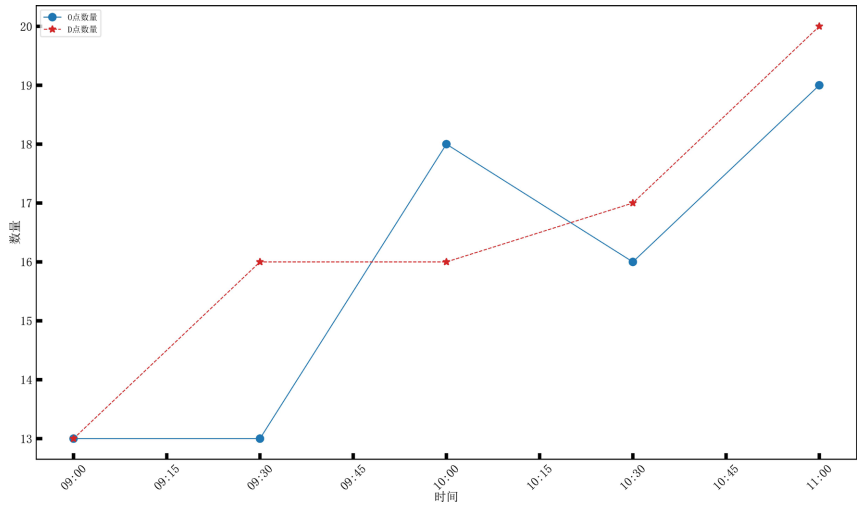


(a)

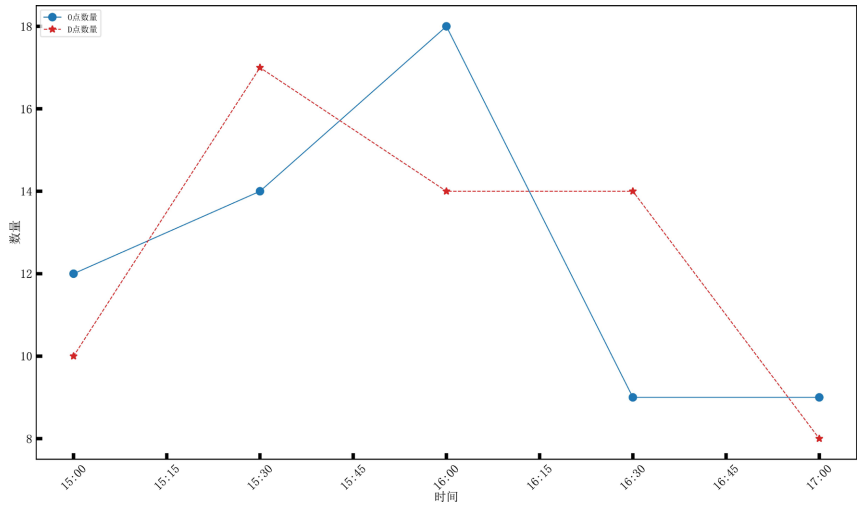


(b)

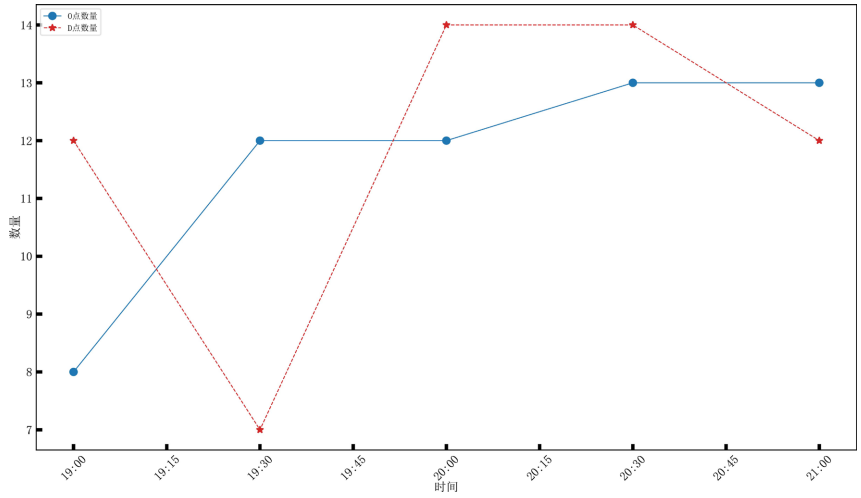
Figure 6. Division of hot time periods on rest days
图 6. 休息日热点时间段划分



(a)



(b)



(c)

Figure 7. Division of hot time periods on working days
图 7. 工作日热点时间段划分

3.2. 交通热点区域分析

在排除虚假订单的基础上，对热点区域进行了热度量化。包括以下五个热点路段：人民南路三段、熊猫大道、华泰路、迪康大道和百草路。接下来我们将对热点路段进行具体的分析，以人民南路三段为例，结合百度地图得到人民南路三段附近的各类现有基础设施如图 8 所示。

人民南路三段附近分布着很多文化教育设施，如成都石室联合中学、成都石室锦城外国语学校、四川大学以及西华大学等。同时，这个区域还包含了很多商业中心，例如成都来福士广场、皇城商业广场等。除此之外，附近还有四川省计算机研究院等科研机构。这些场所的存在可能导致工作日特定时段和休息日出现出租车运行的高峰期。因此，出租车公司可以考虑在高峰期来临前完成车辆调度，为市民出行提供一定的便利。同时也可以在一定程度上提升公司的收益和口碑。

(a) (b)

Figure 8. Infrastructure near section 3 of renmin south road
图 8. 人民南路三段附近基础设施

4. 结论

本文基于成都市出租车 GPS 轨迹数据，通过构建一套集热点挖掘、量化与时空需求分析于一体的研究框架，得出以下核心结论：

1) 本研究证实了 DBSCAN 算法在挖掘城市交通“热点路段”任务上的优越性。相较于 K-means，DBSCAN 无需预设聚类数目，且在轮廓系数等关键评估指标上表现更佳，能够更精确地识别出沿道路网络分布、符合实际交通流特征的任意形状热点区域。

2) 识别出人民南路三段等多个关键交通热点路段，并揭示了其重要性的差异。通过构建的热度量化模型发现，各热点区域的交通吸引力存在显著差异，其中人民南路三段的热度值最高，表明其在城市交通网络中扮演着核心节点的角色。

3) 居民出行的时空需求特征与区域功能属性是热点形成的根本原因。工作日呈现的通勤(9:00~11:00, 19:00~21:00)与特定活动(15:00~17:00，如学生放学)构成了三大出行高峰。对热度最高的人民南路三段的深入分析表明，其周边密集的商业、教育与科研等多功能设施是吸引海量、持续出行需求的直接动因，这为其高热度现象提供了合理解释。

基金项目

重庆市社科规划基金资助项目(2022NDYB72)；重庆市教委人文社科重点项目(25SKGH144)。

参考文献

[1] Shi, Y., Lin, Y., Lim, M.K., Tseng, M., Tan, C. and Li, Y. (2022) An Intelligent Green Scheduling System for Sustainable Cold Chain Logistics. *Expert Systems with Applications*, **209**, Article ID: 118378. https://doi.org/10.1016/j.eswa.2022.118378

[2] Guo, X., Zhang, W. and Liu, B. (2022) Low-Carbon Routing for Cold-Chain Logistics Considering the Time-Dependent Effects of Traffic Congestion. *Transportation Research Part D*: *Transport and Environment*, **113**, Article ID: 103502. https://doi.org/10.1016/j.trd.2022.103502

[3] 余海燕, 唐婉倩, 吴腾宇. 带硬时间窗的 O2O 生鲜外卖即时配送路径优化[J]. 系统管理学报, 2021, 30(3): 584-591.

[4] 李倩, 蒋丽, 梁昌勇. 基于模糊时间窗的多目标冷链配送优化[J]. 计算机工程与应用, 2021, 57(23): 255-262.

[5] 庄峻, 杨东. 面向生鲜电商的前置仓选址及订单履约决策优化研究[J]. 中国管理科学, 2024, 32(2): 188-198.

[6] 刘明, 曹杰, 章定. 数据驱动的疫情应急物流网络动态调整优化[J]. 系统工程理论与实践, 2020, 40(2): 437-448.

[7] Wang, Y., Li, Q., Guan, X., Fan, J., Xu, M. and Wang, H. (2021) Collaborative Multi-Depot Pickup and Delivery Vehicle Routing Problem with Split Loads and Time Windows. *Knowledge-Based Systems*, **231**, Article ID: 107412. https://doi.org/10.1016/j.knosys.2021.107412

[8] Zhen, L., Ma, C., Wang, K., Xiao, L. and Zhang, W. (2020) Multi-Depot Multi-Trip Vehicle Routing Problem with Time Windows and Release Dates. *Transportation Research Part E*: *Logistics and Transportation Review*, **135**, Article ID: 101866. https://doi.org/10.1016/j.tre.2020.101866

[9] Schmidt, C.E., Silva, A.C.L., Darvish, M. and Coelho, L.C. (2023) Time-Dependent Fleet Size and Mix Multi-Depot Vehicle Routing Problem. *International Journal of Production Economics*, **255**, Article ID: 108653. https://doi.org/10.1016/j.ijpe.2022.108653

[10] Fan, H.M., Zhang, Y.G., Tian, P.J., *et al*. (2021) Time-Dependent Multi-Depot Green Vehicle Routing Problem with Time Windows Considering Temporal-Spatial Distance. *Computers & Operations Research*, **129**, Article ID: 105211. https://doi.org/10.1016/j.cor.2021.105211

[11] 辜勇, 袁源乙, 张列, 等. 带时间窗的多中心半开放式车辆路径问题[J]. 中国机械工程, 2020, 31(14): 1733-1740.

[12] 戴树贵, 陈文兰, 潘荫荣, 等. 多配送中心车辆路径安排问题混合蚁群算法[J]. 四川大学学报(工程科学版), 2008(6): 154-158.

[13] 王新杰, 陈淮莉. 多中心联合取送货车辆路径优化与利润分配[J]. 计算机工程与应用, 2023, 59(3): 300-307.

[14] 王万良, 陈浩立, 李国庆, 等. 基于深度强化学习的多配送中心车辆路径规划[J]. 控制与决策, 2022, 37(8): 2101-2109.

[15] Xue, G., Wang, Y., Guan, X. and Wang, Z. (2022) A Combined GA-TS Algorithm for Two-Echelon Dynamic Vehicle Routing with Proactive Satellite Stations. *Computers & Industrial Engineering*, **164**, Article ID: 107899. https://doi.org/10.1016/j.cie.2021.107899

[16] Moazzeni, S., Tavana, M. and Mostafayi Darmian, S. (2022) A Dynamic Location-Arc Routing Optimization Model for Electric Waste Collection Vehicles. *Journal of Cleaner Production*, **364**, Article ID: 132571. https://doi.org/10.1016/j.jclepro.2022.132571