

厦门大数据安全开放创新应用大赛·交通专题

参赛编号：JT-CZYX-0053

赛道类型：算法赛

赛题名称（三选一）：厦门出租车运行特征分析研究

团队人数：4 人

1. 摘要

厦门作为东南沿海重要的旅游城市，具有门对门优势的出租车的运营情况直接影响到城市旅游业的发展。本方案依据信息共享服务平台提供的出租车订单数据、GPS 数据，综合使用统计分析方法结合信息熵理论深入挖掘厦门市出租车运营与居民出行特征。利用最大点密度圆及其半径衡量景区附近载客强度，并基于该指标结合出租车运营特征建立最大覆盖乘客出行需求的出租车泊位选择模型，提出了出租车泊位设置建议。

2. 问题需求

出租车作为城市客运交通系统的重要组成部分，以高效、便捷、灵活等优点深受出行者的青睐。出租车每天的运营中会产生大量的上下车点位相关信息，对这些数据进行科学合理的关联和挖掘，对比在工作日以及休息日、节假日的出租车数据的空间分布及其动态变化，对出租车候车泊位、管理调度和居民通勤特征的研究具有重要意义。厦门作为旅游的热点城市，出租车的智能调度在提升出行者旅行体验、提升运营效率等方面起到了重要的作用。

本方案将依据厦门市大数据交通信息共享服务平台提供的出租车（包括巡游

车和网约车) 订单数据、GPS 数据, 综合应用统计分析方法对出租车载客点的空间分布特征进行量化计算, 挖掘市民出行规律, 并给出出租车整体运行特征和载客热点区域分析, 对比分析工作日和非工作日、节假日和非节假日异同; 提取可视化指标, 从运算的结果中结合人文因素具体分析厦门出租车运行的特征。该结果可为政府制定出租车相关政策、出租车司机制定行驶路线、解决“打车难”问题提供数据支持。

同时根据景区周边出租车载客点的密度分布确定其周边载客点聚集分布的范围和强度, 分析其时变特性; 并结合整体分布特性以最大覆盖需求, 分析得出厦门市景区周边出租车候车泊位设置的建议, 拟解决当前出租车在景区周边载客无秩序混乱问题。

3. 数据应用

3.1. 数据清单

以下表格列出了本方案所要使用到的数据集及字段。

厦门巡游车订单数据（五一假期/普通工作日及双休日）

属性名	英文属性名	备注
车牌号	VEHICLE_PLATE_NUMBER	-
上车时间	GETON_DATE	YYYY/mm/dd HH:MM
上车经度	GETON_LONGITUDE	CGCS2000 坐标系
上车纬度	GETON_LATITUDE	CGCS2000 坐标系
下车时间	GETOFF_DATE	YYYY/mm/dd HH:MM
下车经度	GETOFF_LONGITUDE	CGCS2000 坐标系

下车纬度	GETOFF_LATITUDE	CGCS2000 坐标系
计程公里	PASS_MILE	单位：km
空驶公里	NOPASS_MILE	单位：km

厦门网约车订单数据（五一假期/普通工作日及双休日）

属性名	英文属性名	备注
车牌号码	VEHICLE_NO	-
车辆出发经度	DEP_LONGITUDE	CGCS2000 坐标系
车辆出发纬度	DEP_LATITUDE	CGCS2000 坐标系
上车地点	DEP_AREA	-
上车时间	DEP_TIME	YYYYmmddHHMMSS
车辆到达经度	DEST_LONGITUDE	CGCS2000 坐标系
车辆到达纬度	DEST_LATITUDE	CGCS2000 坐标系
下车地点	DEST_AREA	-
下车时间	DEST_TIME	YYYYmmddHHMMSS
载客里程	DRIVE_MILE	单位：km
载客时间	DRIVE_TIME	单位：秒
空驶里程	WAIT_MILE	单位：km

五一假期厦门巡游车辆 GPS 数据

属性名	英文属性名	备注
车牌号	VEHICLE_PLATE_NUMBER	-

GPS 速度	GPS_SPEED	单位: km/h
卫星定位时间	GPS_DATE	YYYY/mm/dd HH:MM
经度	LONGITUDE	WGS84 坐标系
纬度	LATITUDE	WGS84 坐标系
运营状态	RUNNING_STATUS	空车 (1)、载客 (2)、 电召 (4)、停运 (8)、 交班 (16)、包车 (32)

五一假期厦门网约车车辆 GPS 数据

属性名	英文属性名	备注
车牌号	VEHICLE_NO	-
定位时间	POSITION_TIME	YYYY/mm/dd HH:MM
经度	LONGITUDE	-
纬度	LATITUDE	-
坐标系	ENCRYPT	1: GCJ-02 标准、2: WGS84 GPS 标准、3: BD-09 百度标准 4: CGCS2000 北斗标准、 0: 其他
GPS 速度	SPEED	单位: km/h
运营状态	BIZ_STATUS	1: 载客、2: 接单、3: 空驶、4: 停运

3.2. 数据分析细节

3.2.1. 数据预处理

3.2.1.1. 对订单数据的清洗

现实中的数据大体上是不完善且存在异常的，因此要对订单的原始数据进行数据预处理，以提高数据质量，便于后期对数据进行挖掘，降低误差。以下展示了订单数据预处理的流程图：

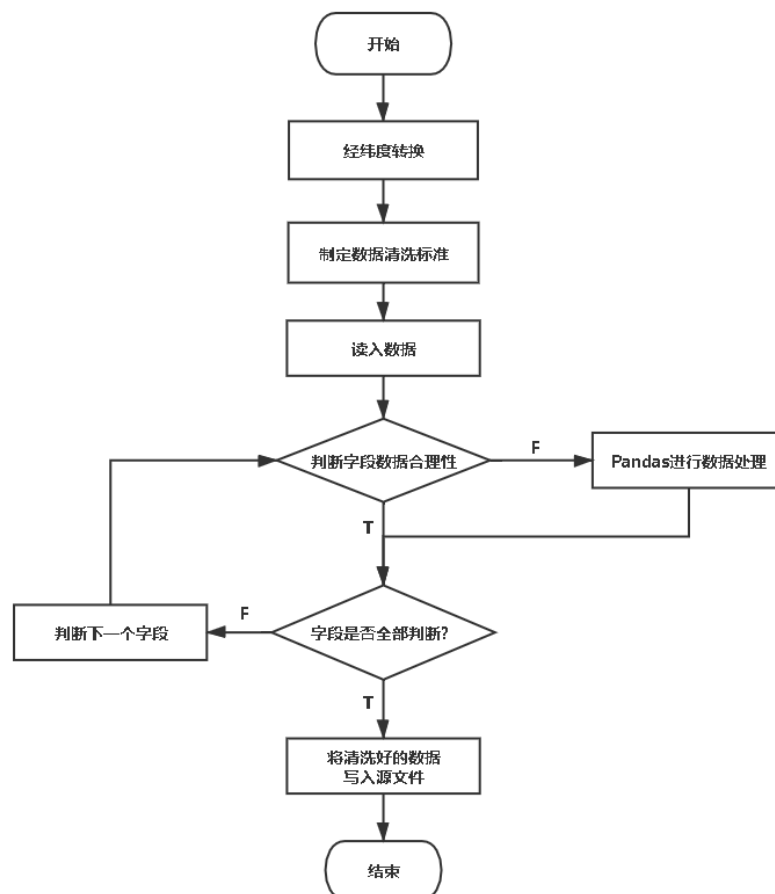


图 1 出租车订单数据清洗流程图

订单数据处理大致思路：

- (1) 经过调研，发现原始订单经纬度坐标为 CGCS2000 坐标系，为了更

方便地进行地图匹配和可视化，将原始经纬度坐标转换为 GCJ-02 坐标。

- (2) 根据厦门市的行政区划、一般经验、交规等定制出一份数据清洗标准，对行驶时间，平均速度等进行约束。
- (3) 利用 pandas 对不合理数据进行处理，如异常数据剔除、缺失值填充等。
- (4) 当数据中所有字段处理完毕后，将结果写入文件。

下面对具体清洗方法进行阐述：

- (1) 清洗异常数据。这里的异常包括上下车点的异常，速度和载客时间异常等。如对上下车点异常：如果二者直线距离小于 50 米的数据以及下车时间在上车时间之前的数据进行剔除。又如对速度和载客时间异常：首先根据上下车时间计算出该订单的行驶时间，接着根据行驶时间和行驶里程计算出该订单出租车的平均行驶速度。剔除载客时间超过 80min、载客里程超过 40km 和平均行驶速度超过 80km/h 的订单。
- (2) 清洗冗余数据。将车牌号和上车时间作为对比字段，当二者同时相同时，剔除多余的数据，留下该车牌第一次出现的订单数据。
- (3) 清洗无效属性信息数据。将字段信息不完整（如个别值为空）的数据进行缺失值补充，例如对于空驶公里字段，为了避免填充 0 所导致的混淆，于是使用-1 填充缺失值。并且对研究分析的内容无参考作用的字段予以剔除。

3.2.1.2. GPS 数据预处理

GPS 数据预处理流程如下图所示：

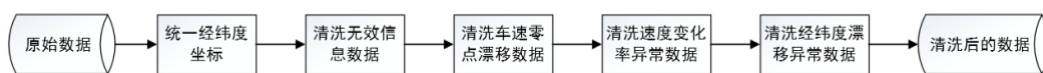


图 2 GPS 数据清洗流程

由于原始数据经纬度坐标是多样的，所以第一步要将所有数据的经纬度坐标统一，接着对厦门市出租车的 GPS 原始数据进行转换与分析，得到原始数据存在的几类错误：

- (1) 无效信息：轨迹点数据记录中存在部分属性字段出现无效值或重复值。
- (2) 车速零点漂移：时空轨迹数据中出现了当车辆静止时，车速非零值跳动现象。如：前一时刻车速 $v_{i-1} = 0$ ，后一时刻车速 $v_{i+1} = 0$ ，而当前时刻车速 $v_i \neq 0$ 。这可能是由于 GPS 终端设备受到外界信号干扰引起的。
- (3) 速度变化率异常：根据欧洲共和体(EEC)与联合国欧洲经济委员会(ECE)标准对制动效能检验要求以及评价参数的规定，不超过 8 座的小客车所能采取的最大制动减速度为 $5.8m/s^2$ 。同时出租车的驾驶行为特性决定了车辆在典型行驶工况下速度变化率一般不会超过 $2.5m/s^2$ ，因此，速度变化率的参考阈值可以设置为 $[-5.8m/s^2, 2.5m/s^2]$ ，数据中速度变化率超出这个区间外的为异常。
- (4) 经纬度漂移：漂移点即轨迹点数据不在实际行驶的道路上。

如果某次行程的轨迹被判定为不可用，则在订单数据中剔除相应的订单数据。下面对具体的清洗方法进行阐述：

- (1) 清洗无效信息数据。将属于字段信息采集不完整（如属性值全空），的数据进行填充，对研究分析的内容没有贡献的项或重复项予以剔除。
- (2) 清洗零点漂移数据。对于此类型数据，根据速度为 0 的时刻，以及前后时刻的经纬度的值是否相等，进行速度异常值修复处理，将当前点的速度置为 0。

- (3) 清洗速度变化率异常数据。对所有运营车辆的速度变化率进行计算，计算公式为：

$$a_i = \frac{v_i - v_{i-1}}{t_i - t_{i-1}} \quad (1)$$

这里将第 i 时刻与第 $i-1$ 时刻之间计算的加速度作为第 i 时刻的加速度。根据计算结果剔除速度变化率超出阈值的轨迹数据点。

- (4) 经纬度漂移：按照车辆行驶的速度和时间差可以估算出两点间的最大位移量 X ，而按照车辆行驶轨迹中的经纬度数据也可以求出两点间的距离 L 。据地球半径 R 和弧度 c 得出球面上 A 、 B 两点间的距离 L 。计算公式如下：

$$\cos \theta = \sin(lat_A) \sin(lat_B) + \cos(lat_A) \cos(lat_B) \cos(lat_A - lat_B) \quad (2)$$

$$c(\text{弧度}) = \frac{\arccos \theta}{180} * \pi \quad (3)$$

$$L = R * c(\text{弧度}) \quad (4)$$

$$X = \max\{v_i, v_{i+1}\} * (t_{i+1} - t_i) \quad (5)$$

参数名称	参数说明	备注
lat	纬度	北纬
R	地球半径	6371km
c	弧度	起终点与地心所成弧的弧度
θ	方位角	起终点与地心所成角的度数
v	速度	单位:km/h
t	采集时刻	-

表 1 经纬度漂移计算公式参数说明

例如，依据其行驶速度和时长，可以按公式(两点中的速度最大值与时间差的乘积)求出汽车速度在 93km/h 时，其单位时间内最大位移量 X 为 25.83333 米，令 $K=L/X$ ，若 K 超过一定的阈值范围，则视为异常。

3.2.2. 出行起终点识别技术构建

利用 3.2.1 节中介绍的数据预处理方法对原始数据进行清洗，然后将网约车和巡游车清洗后的订单数据进行融合，最后利用地图 API 进行起终点识别，并输出结果到文件。

出行起终点识别技术构建流程如下图所示：

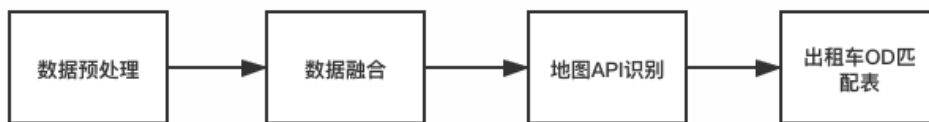


图 3 出行起终点识别流程

3.2.3. 出租车运行特性分析

厦门市出租车运行特征分析流程如下图所示：

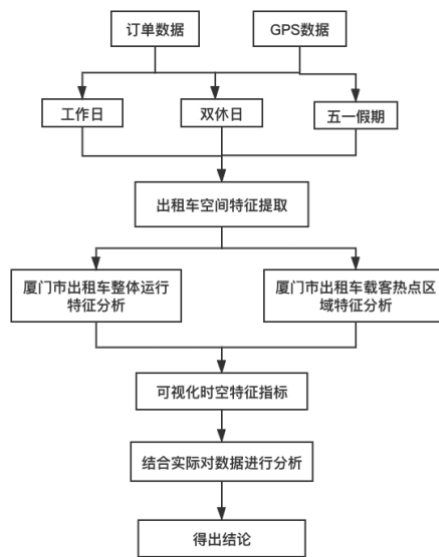


图 4 厦门市出租车运行特征分析流程

首先，对经过预处理后的订单数据、GPS 数据从时间维度上（工作日/双休日/五一假期）进行分割；接着，对每一个时间维度分别提取该时间范围内的出租车空间运行特征；然后，计算厦门市出租车整体运行特征指标和载客热点区域的出租车运行指标，并对二者加以对比，以更好地反映载客热点区域的时空特征；最后，将计算好的各项出租车时空指标进行可视化，并结合厦门市的政策、重要设施分布等实际情况进行分析，总结得出厦门市的出租车运行特征。

在技术上，使用了空间统计分析技术、密度分析技术、信息熵分析技术、数据可视化技术等分析方法。其中，空间统计分析技术包括利用地理信息系统软件 ArcGIS 进行中位数中心、标准差椭圆、凸壳等空间指标的计算；密度分析技术包括使用机器学习中的密度聚类算法 DBSCAN，以及核函数分析等方法对载客点空间分布进行分析；信息熵分析技术将传统的信息熵理论应用到对交通领域数据的分析中，使其具有更好的可解释性；数据可视化技术包括使用 Python 的可视化工具 Matplotlib 和 GIS 可视化工具 Folium 等对各时空指标的计算结果进行可视化展示。

3.2.4. 景区周边出租车运行特性分析

景区周边出租车运行特征分析流程如下图所示：

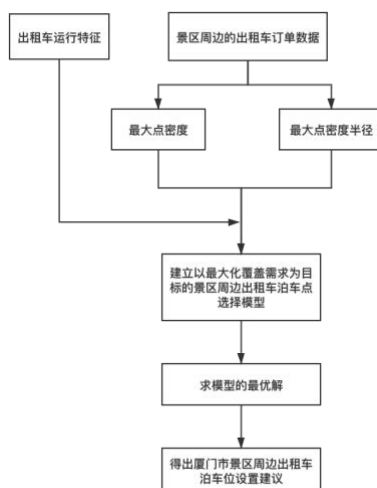


图 5 景区周边出租车运行特性分析流程

首先，筛选在景区附近的载客订单，并以景区为中心，尝试寻找载客点密度最大的圆，这个圆即最大点密度，它反映了景区周边载客点的聚集程度，最大点密度圆的半径即最大点密度半径，它从数值上量化了乘客的聚集强度；接着，以最大点密度半径为基准，结合在 3.2.3 节中得出的厦门市出租车运行特征，建立以最大化覆盖乘车需求的景区周边出租车泊车点选择模型；然后，对该模型求最优解，最后得出景区周边出租车泊车点的设置建议。

在技术上，使用了空间建模技术。空间建模技术包括根据载客点与景区的相对位置，量化计算乘客的乘车需求，以及根据这些需求建立景区周边出租车泊车点选择模型。

3.3. 数据安全保障

针对大赛所提供的数据，将采用以下方法来保障使用过程中所使用的数据的安全：

- (1) 订单数据所涉及的订单号或车牌号等将截取其中的一部分作为代表展示，不会显示完整车牌号和订单号。本次针对数据的分析仅涉及大赛所提出的问题，不对数据所涉及的网约车公司的运营策略等进行分析。通过以上的方法保障企业的合法商业机密不外泄。
- (2) 巡游车和网约车所涉及到的 GPS 信息，将完全隐匿出行轨迹的车牌号信息。数据只分析乘客的起终点数据和部分行车轨迹等数据，不会对司机的行车方式和路径选择等进行过度的解读。个人隐私的保护措施将在所有的模型结果中得到体现。
- (3) 在进行数据可视化时，仅展示计算后的数据，不会展示车牌号，订单号等数据，保障用户隐私的安全。

4. 技术架构

下图展示了项目的技术路线图：

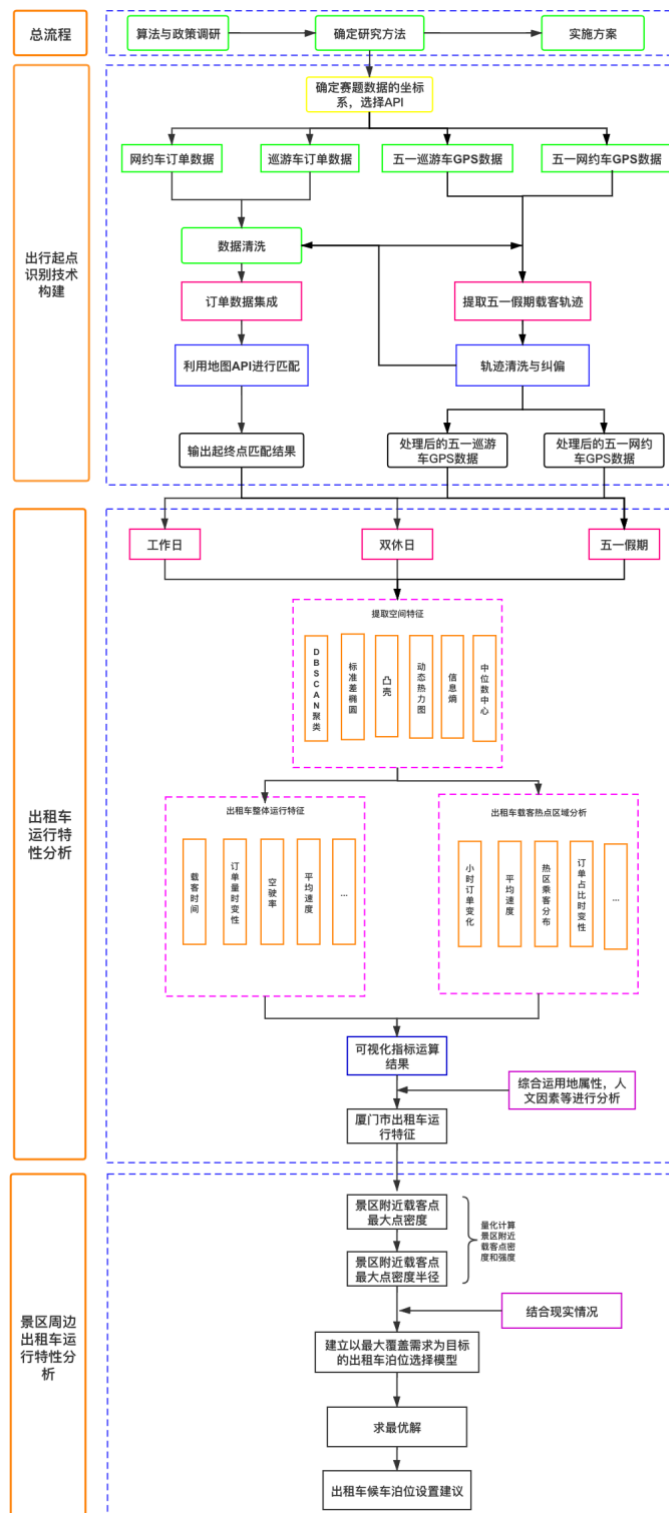


图 6 厦门出租车运行特性分析研究技术路线图

从宏观上，项目分为三个阶段：前期算法与相关政策的调研、制定解决方案和实施解决方案。在制定解决方案阶段，又将其细分为三个子阶段：出行起终点

识别技术构建、出租车运行特性分析和景区周边出租车运行特性分析。三个子阶段相辅相成，共同构成了解决方案的整体。下面，将对这三个子阶段的技术架构进行详细介绍。

4.1. 出行起终点识别技术构建技术架构

在本阶段，需要对原始的巡游车、网约车订单数据进行数据清洗，然后对清洗后有效订单数据中的经纬度，利用地图 API 进行地点匹配，最终生成出租车行程起终点匹配表，如下表所示：

字段名	中文解释	备注
VEHICLE_NO	车牌号	-
DEP_TIME	上车时间	-
DEP_LONGITUDE	上车经度	GCJ-02 坐标系
DEP_LATITUDE	上车纬度	GCJ-02 坐标系
DEP_AREA	上车地点	-
DEST_TIME	下车时间	-
DEST_LONGITUDE	下车经度	GCJ-02 坐标系
DEST_LATITUDE	下车纬度	GCJ-02 坐标系
DEST_AREA	下车地点	-

表 2 出租车起终点匹配表

首先，用 3.2.1 节中所提到的方法对原始数据进行预处理，包括无效数据的剔除、空值的填充等。由于本阶段更关心的是乘客的出行起终点，与运营车辆类型无关，所以将巡游车和网约车经过预处理后的订单数据进行融合，统一生成一

张表。

然后，为了通过订单数据中的上下车经纬度信息得到上下车地点的详细地址，故提出起终点识别技术。该技术以地图服务 API 为基础，每个订单起终点的经纬度数据为输入，具体的字符串地理位置信息为输出。

最后将经纬度匹配后的地理位置数据添加进融合后的数据表，或对其原有信息进行地理位置修正。

4.2. 出租车运行特性分析技术架构

根据 4.1 节所构建的乘客出行起终点数据和 GPS 数据预处理结果，将订单数据和 GPS 轨迹数据分成不同的时间区间（工作日/双休日/五一假期），方便后续对比分析；接着，先提取载客点分布的空间特征，以获取载客热点区域，再计算出出租车整体和热点区域的运行指标；然后将时空指标进行可视化；最后结合实际情况进行对比分析，并得出结论。

4.2.1. 载客点空间特征分析

首先，在地图上对乘客的上/下车点进行地理标注，可以简单统计载客点的空间计数，对每一个时间段内的情况进行宏观把握。如下图所示：



图 7 载客点分区域计数

从图中可以看出，基于样本数据，岛内的打车热度整体上要远大于岛外的打车热度。

为了更具体地了解载客点的时空分布和聚集性特征，将信息论中的信息熵理论引入到了对载客点数据的时空分析中。信息熵是信息论之父香农于 1948 年提出的用于衡量数据中信息量的量化指标。

由于信息熵可以衡量随机变量的不确定性程度，而出租车载客点的分布就是一个随机性很强的事件。故可将信息熵理论引入到对出租车载客点的空间分析中。将厦门市按照其行政区划分为 r 个区域，设随机变量 μ_i 表示载客点随机落入第 i 个区域，其概率为 $p(\mu_i)$ ， i 的取值为 $0, 1, 2, \dots, r-1$ 。则定义出租车载客点的信息熵 H 为：

$$H = - \sum_{i=0}^{r-1} p(\mu_i) \log_b p(\mu_i) \quad (6)$$

其中， b 是对数底，取 2 时，信息熵的单位是比特，取自然对数 e 时，信息熵的单位为奈特；取 10 时，信息熵的单位为 dit 。为了方便后续的计算和可视化，将信息熵的计算结果进行归一化处理，即将信息熵的取值映射到 $[0, 1]$ 区间，由于信息熵的取值在 $[0, H_m]$ 之间，定义归一化的公式如下：

$$J = \frac{H}{H_m} \quad (7)$$

其中， H_m 表示当前计算时间区间中最大的信息熵， J 表示载客点在某个时间点内的分布均衡程度， J 越大，表示载客点的空间分布约均衡， J 越小，表示载客点的空间分布越不均衡。

由信息熵得到整体的载客点时空分布均衡性特征后，需要了解其分布细节，故采用了基于密度的 DBSCAN 聚类算法。该算法优势是不需要事先确定聚类的个数，同时还能识别数据中的噪声点，让聚类结果更加清晰明确，是交通领域中常用的聚类算法之一。DBSCAN 算法需要确定两个参数，分别是数据点的 ϵ 领域和包含在 ϵ 领域中最少点的个数 MinPts。这两个参数的选择是 DBSCAN 算法的难点。在本阶段，采用了类似 K-Means 算法中的“肘部法则”和 k 近邻距离来确定这两个参数。

在确定 MinPts 参数的过程中，通过将 ϵ 值保持不变，观察不同的 MinPts 所聚类数和噪声点数，可以得到较确切的 MinPts 取值。例如对于样本数据，将 MinPts 的取值范围设置为[0,10]， ϵ 领域取值设置为[0.002,0.007]，可得以下图表：

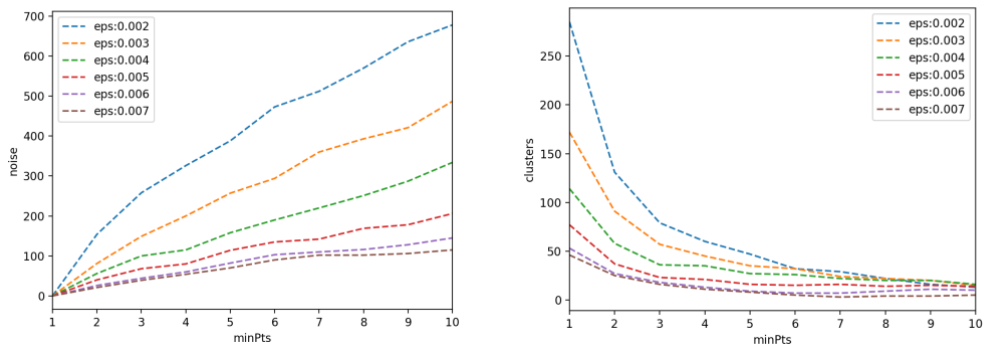


图 8 MinPts 与噪声关系图(左)，MinPts 与聚类个数关系图(右)

从图中可以发现，当 MinPts 取值大于等于 5 时，在大部分 ϵ 取值下，噪声点的增长趋势放缓，聚类的簇数也趋于稳定，所以，将样本数据 DBSCAN 算法的 MinPts 定为 5。

为了确定 ϵ 领域的具体取值，使用了 k 近邻距离算法，算法的流程图如下图所示：

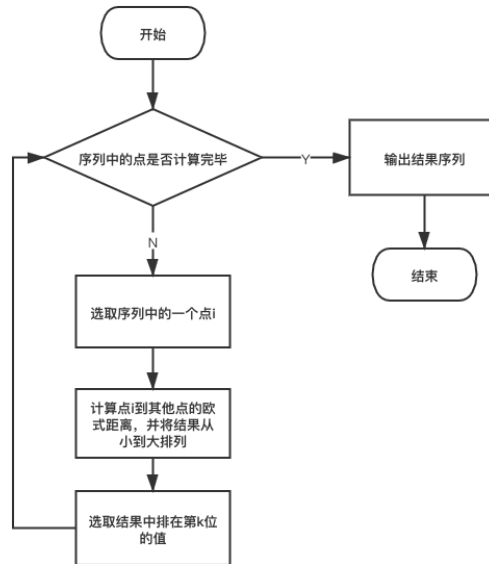


图 9 k 近邻距离算法流程图

需要说明的是，当选定 MinPts 的值后，k 值即为 MinPts 的值。该算法的目标是找到一个 ϵ 领域取值范围，使得载客点在这个取值下，图像线性增长且较平缓，然后通过可视化的方法来确定最终取值，即存在某一个 ϵ ，能保留较小的类簇，又不会将中等大小的簇合并成一个大簇。可以根据 k 近邻距离算法作出载客点数与 ϵ 领域关系图，如下图所示：

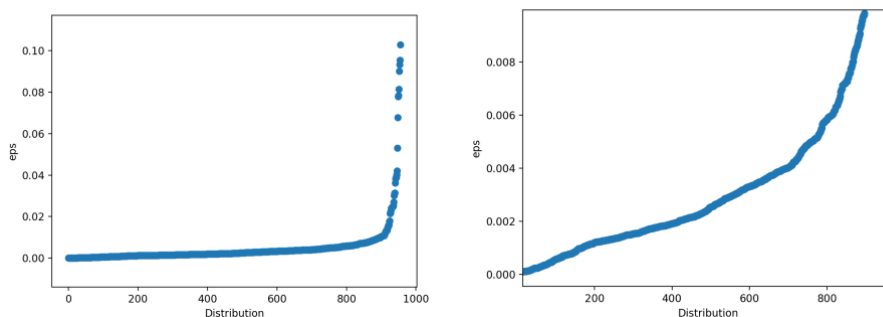


图 10 ϵ 领域取值与载客点分布图

从图中可以发现，基于样本数据，当 ϵ 在 $[0.001, 0.004]$ 附近时，曲线呈现出线

性增长，且比较平缓。通过作图比对，确定样本数据的 ϵ 最终取值为 0.003。此时，DBSCAN 算法对样本数据共聚出 35 类，可视化效果如下图所示：



图 11 DBSCAN 聚类可视化

分析出载客点的聚集分布后，需要深入分析载客点聚集区域的内部情况，即在指定聚类簇中，哪些位置载客点分布更为密集，哪些更为稀疏。本项目使用了核密度分析方法。

核密度分析是空间统计分析中密度分析的一个重要组成部分，核函数是一个明确、光滑且无界的函数，核函数的作用是，可以根据一定的带宽，将单位面积的量值拟合光滑的锥状表面。在本方案中，核函数选取的是 Epanechnikov 核函数。可以利用 ArcGIS 中的核密度分析工具或者其他核密度分析库函数进行计算，最终得出的核密度图如下：



图 12 核密度分析图

通过核密度分析，可以勾画出载客热点区域，为了更好研究载客热区在空间中随时间变化情况，引入了动态热力图加以可视化展示：



图 13 动态热力图

在空间指标分析中，同时也计算了其他一些常用空间指标，用来反映载客点的空间分布特征，如下表所示：

指标名称	指标含义	效果图
------	------	-----

标准差椭圆	分析载客点空间分布 异向性	
凸壳	载客点空间分布最大 范围	
中位数中心	由于中位数中心总是 偏向密度密集的一 侧，故可用它来代表 每日载客密度中心	

表 3 其他空间统计指标

4.2.2. 厦门市出租车整体运行特征分析

为了更好分析和挖掘厦门市出租车运行特征，了解厦门市民的整体出行规律，同时为了与 4.2.1 节中的空间特征结合分析以及将在 4.2.3 节中讨论的载客热区特征做对比，计算了一些指标用于反映厦门市的出租车整体运行特征，如下表所示。这些特征将会分不同的日期属性进行对比分析。

指标名称	计算公式（如果有）	参数说明（如果有）	指标含义
------	-----------	-----------	------

日均载客时间	$T = \frac{\sum t_i}{m}$	t_i : 第 i 辆车的单日总载客时间。 $i = 0,1,2, \dots, m$ m : 单日所有运营车辆的数量	反映出租车司机单日平均劳动强度
日均载客里程	$L = \frac{\sum l_i}{m}$	l_i : 第 i 辆车的单日总载客里程。 $i = 0,1,2, \dots, m$ m : 单日所有运营车辆的数量	反映运营车辆的道路占用情况
日均载客次数	$C = \frac{\sum N_i^{on}}{m}$	N_i^{on} : 第 i 辆车的单日载客次数。 $i = 0,1,2, \dots, m$ m : 单日所有运营车辆的数量	反映司机的劳动强度和城市居民出行需求
日均小时订单变化	-	-	反映居民在不同时间范围内的出行需求
行程平均速度	$v = \frac{\sum s_t}{T}$	s_t : 在 T 时间范围内订单的总行驶里程。 $t \in T$ T : 观察时间尺度	反映不同时间段内的道路整体拥堵情况
空驶里程率	$E = \frac{s_{car}}{s_{all}}$	s_{car} : 车均空驶里程 s_{all} : 车均总行驶里程	反映城市出租车运行效率

表 4 厦门市出租车整体运行指标

4.2.3. 厦门市出租车载客热点区域分析

载客热点是本阶段关注的重点。经过 4.2.1 节的出租车载客点空间特征分析后，可以勾画出出租车的载客热点区域大致范围，现针对载客热点区域，做下表的指标计算，并与厦门市出租车整体运行水平进行对比，进一步挖掘载客热点区域的出租车运行特征和居民出行特征。

指标名称	指标含义
热区订单小时变化	反映热点区域居民出行需求随时间变化特征
热区订单占总订单比	反映热区居民整体出行需求
热区订单 OD 分布	反映往来热区居民的出行特征
热区载客出行平均时间/里程	反映以热区为中心的居民活动范围
行程平均速度	反映热区附近路面拥堵情况

表 5 厦门市出租车载客热点区域计算指标

对于上述表格中的平均速度的计算，需要做如下的补充说明：由于工作日/双休日出租车 GPS 数据的缺失，故在计算该指标时，将订单 OD 均在热区范围内或者起点在热区，终点在热区附近的订单数据纳入计算，将其作为近似值；对于五一假期，可以根据 GPS 数据和热区范围进行计算。

4.2.4. 指标结果分析与得出结论

将出租车载客点空间分布特征利用 ArcGIS 等工具出图可视化，同时在计算好出租车整体运行特征指标与出租车载客热点区域运行特征指标后，将这些指标通过绘图库(如 Matplotlib)进行可视化，将二者进行对比分析，同时结合厦门市当地政策、人文因素以及土地用途等进行综合分析，最终得出厦门市出租车的运行特征。

4.3. 景区周边出租车运行特性分析技术架构

这一阶段的技术架构是：先计算以景区为中心的载客点最大点密度和最大点密度半径，接着以此为基准建立渔网，然后建立以最大覆盖乘客乘车需求的出租车泊车点选择模型，最后通过计算模型最优解得出泊车点设置方案。

4.3.1. 景区周边载客点聚集范围和强度的计算

为了衡量景区周边载客点聚集范围和强度，本阶段引入了两项指标：景区周边载客点最大点密度和景区周边载客点最大点密度半径。最大点密度是指以景区为圆心，寻找载客点密度最大的圆。它反映了景区周边的载客点密度范围，其强度用这个最大密度圆的半径来衡量，这就是最大点密度半径。

最大点密度的计算方法流程图如下图所示：

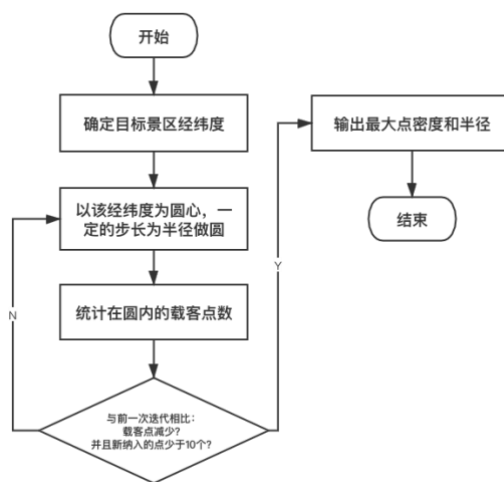


图 14 最大点密度及其半径计算流程图

在计算载客点密度时，参考了人口密度的计算方法，即：

$$\rho = \frac{n}{S} \quad (8)$$

其中， ρ 表示载客点密度， n 表示落入圆内的载客点数， S 表示圆面积。在算法迭代过程中，每次都将在当前迭代时，新纳入的载客点数，与上一次迭代时纳入的载客点数进行对比，如果纳入的载客点数目下降且少于 10 个，表示当前密度圆的边界所扫过的区域载客点较为稀疏，可以停止半径扩张了。然后根据式 (3)，即可计算以景区为中心的最大点密度和最大点密度半径。

4.3.2. 出租车候车泊位设置模型的建立

确立了以景区为中心的最大点密度范围和最大点半径后，基于计算结果，勾画出拟建立出租车候车泊位点的范围，建立渔网，同时建立最大覆盖需求点的出租车泊位设置模型。模型和参数约束如下：

$$\max Z = \sum_{i \in I} \sum_{j \in J} h_i y_{ij} \quad (9)$$

s.t

$$\sum_{j \in J} c_j k_j x_j \geq \sum_{i \in I} h_i y_{ij}, \forall j \in J \quad (10)$$

$$\sum_{j \in J} c_j k_j x_j \leq \sum_{i \in I} h_i \quad (11)$$

$$\sum_{j \in J} x_j = P \quad (12)$$

$$x_j = \{0,1\}, \forall j \in J \quad (13)$$

$$y_{ij} = \begin{cases} 1, d_{ij} < S_{min} \forall i \in I, j \in J \\ \frac{S_{max} - d_{ij}}{S_{max} - S_{min}}, S_{min} \leq d_{ij} \leq S_{max} \forall i \in I, j \in J \\ 0, d_{ij} > S_{max} \forall i \in I, j \in J \end{cases} \quad (14)$$

$$y_{ij} \leq x_j, \forall i \in I, j \in J \quad (15)$$

$$\sum_{j \in J} y_{ij} \leq 1, \forall i \in I \quad (16)$$

各参数的含义如下表所示：

参数名	含义
-----	----

I	出行需求点的集合，需求点 $i \in I$ （拟计划将渔网中心作为需求点代表）
J	出租车泊车位候选点集， $j \in J$ （拟根据 4.2 节结论进行选择）
x_j	是否在 j 处设立泊车点（0-为建设；1-已建设）
k_j	泊车点 j 在高峰小时的泊位周转率
h_i	需求点 i 在高峰小时的出行需求
y_{ij}	对于需求点 i 中乘客前往泊车点 j 乘车比例
P	计划建设的泊车点数
d_{ij}	需求点 i 距泊车点 j 的距离
S_{min}	泊车点 j 的最小覆盖范围
S_{max}	泊车点 j 的最大覆盖范围
c_j	停靠站 j 的泊车位数量

表 6 模型参数说明

约束条件说明如下：式（9）表示停靠站覆盖的出行需求最大。式（10）、（11）表示各停靠站所提供的运力能够满足高峰小时乘客的乘车需求，且不会超出乘客在高峰小时的总需求，以免造成运力浪费。式（12）表示建设的停靠站数目等于预期数目。式（13）表示 x_j 为决策变量。式（14）表示某个需求点 i 前往泊车点 j 乘车的比例跟泊车点 j 的覆盖范围有关，需求点离泊车点越远，前往比例越低。式（15）表示出行者若处在泊车点的服务范围内，则考虑前往泊车点乘车，乘客的选择受到泊车点设置约束。式（16）表示若处于泊车点的服务范围内，每一个需求点的部分或者全部乘客考虑前往泊车点乘车。

建立模型后，对模型求最优解，最终得出景区周边出租车泊车点设置方案。

5. 预期成果

本项目的成果可以用于城市交通治理、交通规划、绿色出行等多个领域。

对乘客起终点数据、出租车 GPS 数据的挖掘结果，例如乘客的上下车热点，载客点分布等，最终可利用可交互式地图进行展示，这种方式的优点是简单直观，正符合数据可视化的核心观念，进而帮助出租车运营公司或出租车司机合理地制定不同时段的行程路线，提高城市出租车的运营效率，降低运营成本。对政府决策者来说，可以利用挖掘结果针对厦门市出租车运营环境，量身制定交通政策，从政府层面推动城市出租车的发展，进一步巩固出租车作为城市公共交通的重要补充的作用。

景区周边往往是容易发生交通拥堵与“打车难”问题的地方，本文建立以最大化覆盖乘车需求的景区周边出租车泊车点选择模型，提出了景区周边出租车候车泊位的设置方案，望通过该方案能够帮助改善景区周边的交通状况，同时满足乘客的乘车需求。

6. 作品价值

6.1. 社会效益

6.1.1. 对居民

(1) 更大限度的满足居民日益增长的出行需求

随着社会的不断进步，人们的出行需求日益增长并朝着多样化发展。越来越多的城市存在着“打车难”的问题。合理安排出租车的时空分布可以提高空载出

租车和用户的匹配率，从而达到满足更多用户出行需求的效果。

(2) 给用户出行带来便利

本方案分析了出租车载客点的分布特征，深入挖掘了出租车的空间分布及其动态变化，根据休息日、节假日的异同，分别得出其载客点的空间分布特征，给出租车运营公司合理投放运力和巡游路径规划提供一定的数据支撑，如何时何地投放运力载客率较高，在给乘客出行带来了便利的同时，一定程度上降低了出租车运营成本。

6.1.2. 对城市

出租车行业是公共交通的重要组成部分，其功能是完成用户点到点的空间位移，以高效、便捷、舒适、“个性化”等作为其产品质量的要素。是满足社会需求，对社会有一定的贡献力，是社会效益的反映。根据本赛题提出的解决方案，总结出以下几点：

(1) 优化出租车资源配置

根据资源优化配置理论：在可替代的几种资源使用方案中选择较优的一种，以达到社会的最高效率和社会利益的最大满足。因此，在出租车无法替代的前提下，优化其资源配置可以在满足少部分人的出行需求的同时，达到城市交通的利益最大化。

(2) 降低出租车占用城市道路交通资源的机会成本

出租车乘客人均占用的道路交通资源大，且占用相同城市道路资源的产出效率低，这就导致了其占用城市道路交通的机会成本过大，影响城市的交通资源的使用。此解决方案分析了载客点的时空分布，同时为出租车的候车泊位的设置提出了建议。在城市交通道路占有率相同的前提下提高其载客率，保证有限的资源得到最大限度的利用。

(3) 提升城市治理水平

对出租车进行合理的调度，实时了解出租车的总体运营状况有利于优化城市交通结构，提高城市交通的效率。而出租车行业的统一管理，包括规划、经营、管理会得到一定程度上的提高。最终达到提升城市治理水平的效果。

6.2. 经济效益

(1) 提高出租车的运行效益

出租车司机可以根据工作日和休息日的载客时空特征对其载客策略进行适当的调整，以达到在相同的时间成本下得到更高的收益。避免由于驾驶经验不足、路况不熟、用户出行需求不了解而引起的盲目寻找乘客等情况的发生。

(2) 降低空驶率，提高能源利用率

本方案对后期分析用户通勤特征有着重要的意义，了解了用户的分布特征能够使得出租车更快地找到客源，有效的减少空驶里程，从而提高出租车的载客率、达到提高能源利用率的效果。

(3) 推动城市经济发展

推动高质量发展是推动城市经济发展的必然要求，而高质量发展需要生产要素投入少、资源环境成本低、经济效益好。当下，通过对出租车的运行特征分析得到的解决方案能够有效帮助城市功能的合理布局，提升城市的交通承载力，从而对城市的经济起到一定的推动作用。

7. 附件

7.1. 数据空间可视化演示网址列表

以下所有演示的数据源为初赛提供的样本数据。

建议使用 Chrome、Firefox、Safari 等现代浏览器获得更好的体验效果。

[厦门出租车乘客分区域计数](#)

[厦门出租车载客点 DBSCAN 聚类可视化](#)

[厦门出租车乘客上车点分布](#)

[厦门出租车乘客下车点分布](#)

[厦门出租车载客点核密度分析](#)

[厦门出租车载客点动态热力图演示](#)

7.2. 主要参考文献

[1] 叶臻, 王海洋, 贺明光, 等. 基于出行需求时空分布的出租车停靠站选址方法[J]. 华东交通大学学报, 2017, 34(06): 97–103.

[2] 祁文田. 基于 GPS 数据的出租车载客点空间特征分析[D]. 吉林大学, 2013.

[3] 林鹏飞, 翁剑成, 刘文韬, 等. 基于多源数据的网络约租车与出租车运营特征分析[J]. 交通工程, 2020, 20(01): 26–33.

[4] 周静一, 李晔. 基于浮动车技术的杭州市出租汽车运行特征分析[J]. 科技信息, 2011(03): 120–122.

[5] 韩勇, 樊顺, 周林, 等. 基于聚类算法的出租载客点时空分布特征研究[J]. 中国海洋大学学报(自然科学版), 2019(S1 vo 49): 155–162.

[6] 金雷, 谢秉磊. 基于上落客时空特征的出租车停靠站选址模型[J]. 交通运输系统工程与信息, 2015, 15(02): 182-188+194.

[7] 胡兰兰. 基于 GPS 出租车高收益热点区域推荐[D]. 温州大学, 2019.

[8] 高静文, 蔡永香, 甘艺垚. 运输车辆轨迹点数据清洗方法研究[J]. 电脑知识与技术, 2019, 15(36): 189-192+194.

[9] 侯聪, 王贺武, 欧阳明高. 中美 PHEV 能耗评价方法对比研究[J]. 汽车工程, 2015, 37(01): 1-8.