



DOI:10.12404/j.issn.1671-1815.2401728

引用格式:刘华胜,杨莎,李津,等.基于出租车轨迹数据的需求响应式公交线路规划方法[J].科学技术与工程,2025,25(5):2135-2145.

Liu Huasheng, Yang Sha, Li Jin, et al. Demand responsive transit route planning based on mining taxi trajectory data[J]. Science Technology and Engineering, 2025, 25(5): 2135-2145.

基于出租车轨迹数据的需求响应式 公交线路规划方法

刘华胜,杨莎,李津*,邓浩然,赵玉琦,董旭

(吉林大学交通学院,长春 130025)

摘要 为了在缺乏公交准确需求信息的情况下能够估算需求响应式公交(demand responsive transit, DRT)的“潜在需求”,以便在试运营前为线路规划提供可行性方案,提出了一种基于出租车轨迹数据的需求响应式公交线路规划方法。首先,通过数据挖掘手段获取研究区域内的出租车轨迹数据并进行预处理,分析轨迹数据的空间分布特征,将其视为该区域内乘客出行的“潜在需求”。其次,利用K均值(K-means)聚类算法确定备选站点,将备选站点连接形成基准站点网络,并将边缘基准站点设为线路的起点和终点。根据线路长度的约束条件,采用K条最短路径(K-shortest paths, KSP)算法生成基准链条。最后,确定基准链条的子链条集合,并根据绕行临界值的约束条件搜索子链条的需求响应站点集合。结果表明:通过多次循环以上算法,可以生成时段内的所有备选线路,并根据各备选线路的综合评价指标选择该时段内初步的最优线路。同样的线路算法可确定不同时段内的最优线路初步方案。

关键词 需求响应式公交;站点选址;线路规划;数据挖掘;出租车轨迹数据

中图分类号 U491.1; **文献标志码** A

Demand Responsive Transit Route Planning Based on Mining Taxi Trajectory Data

LIU Hua-sheng, YANG Sha, LI Jin*, DENG Hao-ran, ZHAO Yu-qi, DONG Kui

(College of Transportation, Jilin University, Changchun 130025, China)

[Abstract] In the absence of accurate transit demand information, a demand responsive transit(DRT) route planning method based on taxi trajectory data was proposed to predict the “potential demand” of demand responsive transit and provide a feasible plan for route planning before transit operation. Firstly, taxi trajectory data in the study area was obtained through data mining, representing the “potential demand” for passenger travel in the area, and candidate station were determined using the K-means clustering algorithm. Secondly, a benchmark station network was established using these candidate station, with edge benchmark stations designated as the starting and ending points of routes. Utilizing the K-shortest paths(KSP) algorithm constrained by route length, benchmark chains were generated. Finally, after determining the sub-chain set of the benchmark chains, demand response stations within each sub-chain were searched based on circumferential critical value constraints. Using this algorithm, alternative routes were generated repeatedly within specific time periods, and an initial optimal route was selected based on comprehensive evaluation indices for each alternative route.

[Keywords] demand response transit; station siting; route planning; data mining; taxi trajectory data

20世纪80年代Flusberg^[1]首次提出了需求响应式公交(demand response transit, DRT)的概念。DRT通过相应乘客出行需求进行运营服务,其运营以乘客出行需求为导向,是一种不定点、不定线的公交运营模式。其灵活的运营方式使得运营企业能够根据乘客出行需求设计车辆运营路线,与常规公交固定站点、线路和班次的运行模式不同,DRT

的服务特性介于私家车和传统公交之间,具有低成本、个性化等特点^[2-3]。因此,研究DRT具有重要的应用实践场景和价值。

目前,对DRT的研究主要集中在车辆调度和路径优化等方面。在车辆调度方面,Liu等^[4]将DRT服务与固定线路运输网络联合优化,将服务区域划分为不相交的局部区域,固定线路网络为跨区域出

收稿日期:2024-03-12 修订日期:2024-11-19

基金项目:国家自然科学基金(71871103);吉林省教育厅科学研究项目(JJKH20231189KJ);吉林大学研究生创新研究计划(2023CX197)

第一作者:刘华胜(1987—),男,汉族,吉林长春人,博士,副教授。研究方向:交通运输系统规划。E-mail:jliuhs@jlu.edu.cn。

*通信作者:李津(1970—),女,汉族,吉林长春人,博士,副教授。研究方向:城市与区域交通规划。E-mail:li_jin@jlu.edu.cn。

行提供干线运输服务,而 DRT 系统则针对跨区域出行的需求响应服务;Gong 等^[5]提出了将定制公交网络与换乘运营和乘客路线分配联合设计,在优化乘客线路的同时,设计基于换乘的模块化车队定制公交网络;陆百川等^[6]同时考虑企业和乘客效益,以车辆运营成本 and 乘客出行成本最低构建多目标线路优化模型,该模型在节约企业运营成本的同时,最小化乘客的出行时间成本。在路径优化方面,Sun 等^[7]研究了基于共享 DRT 优化问题,建立了一种基于共享汽车的最佳 DRT 模型,并采用基于遗传算法的两阶段启发式算法求解;孙继洋等^[8]构建了需求响应式公交线路优化(vehicle routing problem, VRP)调度模型,采用启发式算法进行求解,并用案例验证了模型和算法的有效性和鲁棒性;王正武等^[9]研究了多换乘点之间的线路协调问题,建立了混合需求条件下的线路协调优化模型,设计了模拟退火算法求解。

由此可见,众多学者对 DRT 进行了较为深入的理论研究,构建了不同场景下、不同约束条件下、不同目标函数的公交线路和站点选址算法。然而实际线路运营中,由于缺乏准确的公交需求信息,如何有效地估算需求响应公交的“潜在需求”,并为其线路规划提供可行性方案成为亟待解决的问题。随着数据挖掘技术的发展,利用出租车轨迹数据进行 DRT 线路规划成为一种可行的研究方法。这种方法通过深入挖掘轨迹数据,可以更准确地理解城市出行的特征和规律,为 DRT 线路的规划提供了重要参考。

以往的出租车轨迹数据挖掘研究主要集中在利用出租车轨迹数据进行乘客出行热点区域识别和交通流量预测等方面。在出行热点区域识别方面,孙贵治^[10]利用 Python 编程语言,对出租车 GPS 轨迹数据进行清洗等预处理工作,并采用隐马尔可夫地图匹配算法进行匹配,进而提取上下车轨迹点以及出租车出行需求量,研究出行需求时空分布及出行热点区域分布。Lü 等^[11]通过分析出租车 GPS 轨迹数据挖掘区域内居民出行时空规律,确定了 DRT 站点选址;刘旭等^[12]基于武汉市出租车轨迹数据,通过改进的 Canopy-K-means 聚类算法预测了客源地的出行热点区域,以此构建公共交通站点预测模型,通过对比分析公交站点的合理性,利用数理统计方法研究了客流量的时空分布特征。在交通流量预测方面,汪孝之等^[13]利用出租车 GPS 轨迹数据进行短时流量预测,充分挖掘交通流量序列特征规律,减小序列非线性、非稳定性带来的影响,提出一种基于信号分解的预测模型(linear regression model empirical mode decomposition-random forest-long short-term memory network, LE-RL)。李欣然

等^[14]基于北京市出租车轨迹数据,采用深度重力模型(deep gravity)对不同空间尺度下的轨迹起讫点(origin-destination, OD)交通流进行预测。同时,引入 SHAP 值(Shapley additive explanations)揭示不同尺度下影响 OD 流预测建模的重要特征。李轶群等^[15]利用车载高精度 GPS 跟驰试验数据进行车辆跟驰建模研究,结合深度学习理论和数据驱动方法,构建了基于粒子群优化(particle swarm optimization, PSO)的长短期记忆(long short term memory, LSTM)车辆跟驰模型。

综上所述,出租车轨迹数据挖掘比较全面地挖掘出行模式和行为特征以及交通流量预测,在一定程度上能够促进 DRT 的研究。因此,现提出一种基于出租车轨迹数据来规划 DRT 线路,旨在解决缺乏公交准确需求信息的情况下,运用出租车轨迹数据作为“潜在需求”数据,保证数据的易获取性,实现 DRT 线路规划的可行性。

1 DRT 线路规划设计方案

根据美国公共交通合作研究计划对 DRT 系统进行分类,按照服务模式将其分为 6 种主要类型,包括线路可偏移服务模式、站点可偏移服务模式、需求响应接驳服务模式、需求响应车站服务模式、区段灵活服务模式和区域灵活服务模式^[16]。其中站点可偏移服务模式是指在一定的运营区域内,同时考虑企业运营和乘客出行预约需求两方面,进而对原有固定线路进行调整。乘客在固定站点和部分需求相应站点进行上下车,通过灵活增加停靠站点的方式,规划运行的路径。站点可偏移服务模式示意图如图 1 所示。

站点可偏移公交服务区域通常较广,选取出租车服务覆盖区域作为研究范围,并依据出租车的运营特性将时间段分为平峰和高峰时段进行探讨。针对研究目的,设计方案分为 3 个步骤。

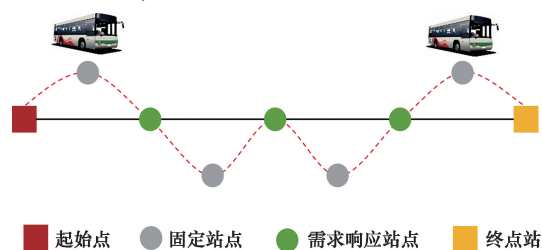


图 1 DRT 站点可偏移服务模式示意图

Fig. 1 Diagram of DRT station service model

1.1 数据的提取及预处理

利用出租车定位设备提取车辆在不同时间点的位置和载客状态等信息的轨迹数据,经过筛选、

清洗和格式调整,去除异常数据、填补缺失数据,并调整数据文件格式。采用小角度旋转的坐标转换方法将经纬度坐标系转换为平面直角坐标系。

通过 Python 分析工具对预处理后的原始数据进行分析,获取每辆出租车在 24 h 内的轨迹数据,并获取所有乘客上下车的时空分布信息。将此信息作为不同时段内站点可偏移的需求响应式公交服务模式的“潜在需求”。

1.2 生成基准站点链条

采用 K-means 聚类方法对“潜在需求”数据进行聚类,以获取乘客出行密集点。通过计算站点指标值确定备选站点的数量,根据站点客流量和时间离散程度将备选站点划分为基准站点和需求响应站点。连接基准站点形成基准网络,以线段的长度为约束条件,剔除长度过短或过长的线段。在网络中选择边缘基准站点作为基准站点链条的起始点和终点,利用 K 条最短路径 (K -shortest pathes, KSP) 算法生成站点链条集合。

1.3 确定最优的路线

从基准站点链条集合中,选择第 k 条链条,确定该链条的子链条,寻找子链条附近的需求响应站点集合,结合子链条集合和需求响应站点确定基准链条的备选线路。将基准站点和需求响应站点所承担的 OD 量,作为备选线路所承担的 OD 数量,计算子链条的评估指标,包括企业运营指标、乘客需求指标及综合评估指标。比较全部子链条的综合评价指标,得到第 k 条基准链条的最优线路。多次循环以上步骤,每次选择不同基准站点链条,找出每条链条的最佳线路,最终得到该时段在研究范围内最优线路的初步方案。

2 数据预处理及出行 OD 需求获取

2.1 数据来源

研究所使用的数据来自公开数据集,包含了 2012 年 10 月 11 日(周四)深圳市境内 14 728 辆出租车在 24 h 内的所有轨迹数据。数据以 CSV 格式存储,总计约有 4 700 万条记录,平均时间间隔为 26.57 s,文件大小约为 1.8 G。

2.2 数据预处理

利用车载卫星定位设备采集的轨迹数据可能受到信号干扰、设备故障等因素影响,导致数据缺失或异常情况的发生。为提高数据质量,需进行清洗和筛选等预处理步骤。预处理流程如图 2 所示。

2.2.1 数据缺失处理

数据缺失采用以下方法处理:首先,直接剔除位置信息和时间信息均为空的数据;其次,对于位置

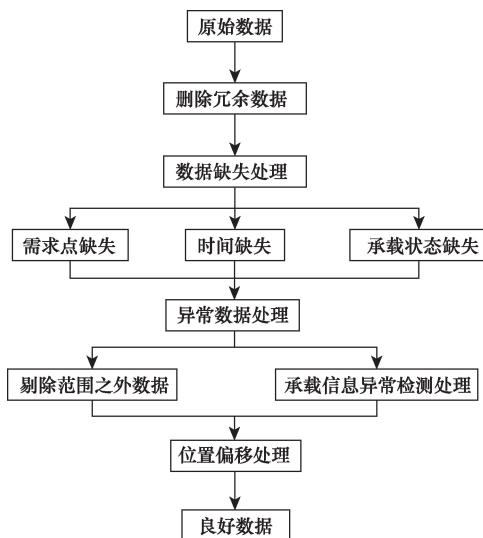


图2 数据预处理过程

Fig. 2 Data pre-processing process

信息缺失但时间信息完整的数据,可根据前后检测时间及位置进行估算;对于时间信息缺失但位置信息完整的数据,可利用前后检测时间的平均值进行估算;最后,对于承载状态缺失的数据,可根据前后 4 条数据的承载状态进行估算。

对于位置信息缺失但时间信息完整的第 n 条数据,设前时刻对应的位置为 $(x_{n-1}, y_{n-1}, t_{n-1})$,后时刻对应的位置为 $(x_{n+1}, y_{n+1}, t_{n+1})$,当前时刻对应的位置为 (x_n, y_n, t_n) ,则对当前位置的计算公式为

$$\begin{cases} \frac{x_n - x_{n-1}}{t_n - t_{n-1}} = \frac{x_{n+1} - x_n}{t_{n+1} - t_n} \\ \frac{y_n - y_{n-1}}{t_n - t_{n-1}} = \frac{y_{n+1} - y_n}{t_{n+1} - t_n} \end{cases} \quad (1)$$

对于承载状态信息缺失的第 n 条数据,根据前后 4 条数据的承载状态估算其状态值,估算原则为承载状态信息为连续信息。设前后状态值为 $S(t_{n-2})$ 、 $S(t_{n-1})$ 、 $S(t_n)$ 、 $S(t_{n+1})$ 和 $S(t_{n+2})$,且均为布尔变量, $S(t_n)$ 为承载状态缺失值,估算处理方式计算公式为

$$S(t_n) = \begin{cases} S(t_{n-1}) = S(t_{n+1}), \\ \quad S(t_{n-1}) = S(t_{n+1}) \\ S(t_{n-1}), \\ \quad S(t_{n-1}) \neq S(t_{n+1}) \text{ 或 } S(t_{n-1}) \neq S(t_{n-2}) \\ S(t_{n+1}), \\ \quad S(t_{n-1}) \neq S(t_{n+1}) \text{ 或 } S(t_{n+1}) \neq S(t_{n+2}) \end{cases} \quad (2)$$

2.2.2 数据异常处理

研究将数据承载信息不准确和车辆经纬度不在研究区域范围内的情况视为数据异常。采用 Py-

thon 中 Shapely 模块的 intersect 几何交集函数工具对所有数据点进行范围检查,以剔除异常数据点。

2.2.3 数据清洗、筛选及坐标转换

车载设备在信息收集过程中可能会出现数据缺失、漂移、时间错乱等问题,因此需要对原始数据进行筛选和清洗,以确保数据质量,并节省程序处理数据的时间。采用 Python 中的 GeoDataFrame 模块对所有数据点进行检查,筛选出在行政区域范围内的 OD 出行数据。基于小旋转角度的变换方法,将预处理之后的轨迹数据,由 WGS84 坐标系数据转换为平面直角坐标系数据。

2.3 出行 OD 需求的获取

预处理后得到的数据为出租车轨迹点数据,按照车牌号和时间顺序进行二次排列,以观察每辆出租车全天的载客状态变化情况,并提取乘客出行的 OD 需求数据。图 3 为某辆出租车全天的载客状态变化情况,发生一次上车则记为 1,发生一次下车则记为 0。因此,可见,该出租车一天内发生 23 次“上车-下车”事件,其中上车时刻对应出行的起点,下车时刻对应出行的终点。利用 Python 分析工具生成了 23 条出行 OD 数据。按照相同的处理方法,可以获取 24 h 内所有出租车的出行 OD 数据。

3 DRT 线路规划算法

3.1 评估指标

选取企业运营指标和乘客需求指标,作为研究区域内站点预选址和线路初步设置的评估指标。其中,企业运营指标包括客流强度 $\rho_{k,h}$ 、基准站点时间离散程度 $\delta_{k,h}^b$ 、需求响应站点时间离散程度 $\delta_{k,h}^d$; 乘客需求指标包括平均松弛系数 $\varphi_{k,h}^d$ 、步行距离系数 $\varphi_{k,h}^b$ 。由企业运营指标和乘客需求指标计算综合评估指标 $E(\hat{C}_k)$,用于确定线路规划设计方案中最优线路的评估指标。

3.1.1 企业运营指标

(1) 客流强度 $\rho_{k,h}$ 。客流强度为线路承担 OD 数量 $Q_{k,h}$ 与线路长度 $L_{k,h}$ 之比,如式(3)所示,代表

线路平均每公里的承载能力,是反映线路运营效率和经济效率的重要指标,为正向指标。

$$\rho_{k,h} = \frac{Q_{k,h}}{L_{k,h}} \quad (3)$$

(2) 基准站点时间离散系数 $\delta_{k,h}^b$ 。基准站点时间离散系数是某时段内服务线路 $k-h$ 基准站点的 OD 数量的标准差与均值的比值,如式(4)所示。该指标反映在基准站点处需求密度是否均匀分布,为负向指标,值越高表明某时段内乘客出行越集中在某几个基准站点处,需求响应线路效率越低。

$$\delta_{k,h}^b = \frac{\text{std}(Q_{k,h}^b)}{\bar{Q}_{k,h}^b} \quad (4)$$

式(4)中: $Q_{k,h}^b$ 为时段内线路 $k-h$ 所有基准站点的需求密度; $\bar{Q}_{k,h}^b$ 为时段内线路 $k-h$ 所有基准站点需求密度的平均值。

(3) 需求响应站点时间离散系数 $\delta_{k,h}^d$ 。需求响应站点时间离散系数是某时段内服务线路 $k-h$ 需求响应站点的 OD 数量的标准差与均值的比值,如式(5)所示。该指标为正向指标,值越高表明某时段内乘客出行越集中在某几个需求响应站点处,需求响应线路效率越高。

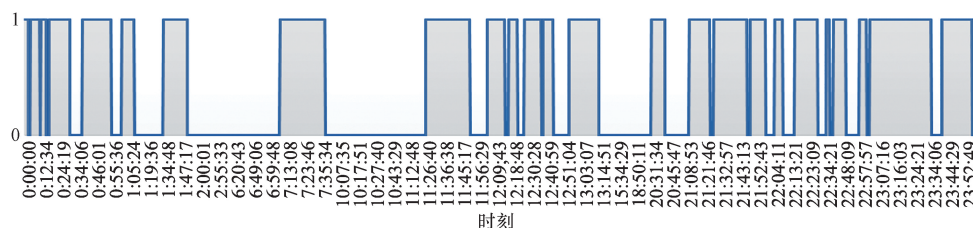
$$\delta_{k,h}^d = \frac{\text{std}(Q_{k,h}^d)}{\bar{Q}_{k,h}^d} \quad (5)$$

式(5)中: $Q_{k,h}^d$ 为时段内线路 $k-h$ 所有需求响应站点的需求密度; $\bar{Q}_{k,h}^d$ 为时段内线路 $k-h$ 所有需求响应站点需求密度的平均值。

3.1.2 乘客需求指标

(1) 平均松弛系数 $\varphi_{k,h}^d$ 。平均松弛系数定义为线路 $k-h$ 运营所需的全部松弛时间与所有需求响应站点的客流量之比,如式(6)~式(8)所示,是指乘客对出行方案的接受程度,即愿意接受的额外等待时间。该指标为负向指标,可以评估乘客满意度和出行体验。

$$\varphi_{k,h}^d = \frac{\sum (l_{i,j} + l_{j,j+1} - 1) P_j T_j}{\sum Q_j} \quad (6)$$



1 为发生一次上车;0 为发生一次下车

图3 某出租车全天载客状态变化情况

Fig. 3 Changes in the 24-hour passenger status of a taxi

$$P_j = \frac{Q_j}{\sum Q_j} \quad (7)$$

$$T_j = \frac{l_{i,j} + l_{j,j+1}}{\bar{v}_{k,h}} + t_j \quad (8)$$

式中: $i \in S_{k,h}^b$; $j \in S_{k,h}^d$; $l_{i,j} + l_{j,i+1}$ 为线路 $k-h$ 需求响应站点 j 与相邻基准站点 i 与 $i+1$ 之间的距离之和; $l_{i,i+1}$ 为线路 $k-h$ 上相邻基准站点 $i, i+1$ 之间的距离; Q_j 为需求响应站点 j 的 OD 数量; P_j 为需求响应站点 j 出现的概率; T_j 为线路 $k-h$ 在需求响应站点 j 的松弛时间; $\bar{v}_{k,h}$ 为线路 $k-h$ 公交车辆平均行驶速度; t_j 为车辆在需求响应站点 j 的停靠时间。

(2) 步行距离系数 $\varphi_{k,h}^b$ 。步行距离系数为乘客平均步行距离与基准站点服务半径之比, 如式(9)所示。平均步行距离是指每次出行到达以及离开线路 $k-h$ 基准站点所需要步行距离的平均值, 如式(10)所示。该指标为负向指标, 反映乘客对步行距离的接受程度, 能够评估乘客满意度和线路可达性。

$$l_{k,h} = \frac{\sum [\text{dist}(o,i) + \text{dist}(i,d)]}{\sum c_i}, i \in S_{k,h}^b \quad (9)$$

$$\varphi_{k,h}^b = \frac{\bar{l}_{k,h}}{r} \quad (10)$$

式(10)中: $\bar{l}_{k,h}$ 为出行到达或离开基准站点的平均步行距离; $\text{dist}(o,i)$ 为出发点到上车时基准站点 i 的步行距离; $\text{dist}(o,i)$ 为下车时基准站点 i 到目的地的步行距离; c_i 为达到或离开线路 $k-h$ 基准站点 i 的次数; r 为基准站点服务半径。

3.1.3 综合评估指标

综合评估指标是评价备选线路的最终决策指标, 由以上五个指标构成。对以上 5 个指标进行 Z-score 标准化处理之后得到均在 $(-1, 1)$ 的指标值, 处理方式如式(11)所示。

$$x^* = (x - \mu) / \sigma \quad (11)$$

式(11)中: x 、 μ 、 σ 分别为某一评估指标所有线路的指标实际值、总体数值均值、总体数值标准差。

综合评估指标计算方法如式(12)所示。

$$E(\hat{C}_k) = \theta_{11}\rho_{k,h} + \theta_{12}(1 - \delta_{k,h}^b) + \theta_{13}\delta_{k,h}^d + \theta_{21}(1 - \varphi_{k,h}^d) + \theta_{22}(1 - \varphi_{k,h}^b) \quad (12)$$

式(12)中: θ_{11} 、 θ_{12} 、 θ_{13} 、 θ_{21} 、 θ_{22} 分别为根据城市性质、发展水平、出行方式的选择、线路运营、乘客的预约等情况对各项评估指标赋予的权重。

3.2 基准链条和最优线路算法

根据以上模式分析, 进行 t 时段内基准链条和

最优线路算法设计, 算法流程如图 4 所示。

3.2.1 生成基准链条

步骤 1 将“潜在出行”作为 t 时段的出行 OD 矩阵 Q_t , 并输入 Q_t 。

步骤 2 采用 K-mean 聚类算法对 OD 矩阵 Q_t 进行聚类, 设定聚类半径为 1 km, 得到聚类站点 i , 将其纳入备选站点集合 S 。

步骤 3 提取聚类站点 i 乘客上下车数量 q_i , 计算离散程度指标 δ_i , 根据乘客上下车数量 q_i 和离散程度指标 δ_i 计算站点指标 S_i , 如式(13)所示。根据站点指标 S_i 确定备选站点, 设定 S^* , 如果 $S_i \geq S^*$, 站点 i 作为备选站点。

$$S_i = q_i + \alpha\delta_i \quad (13)$$

步骤 4 设定临界 q^* 和 δ^* , 对备选站点进行分类, 规则为

$$\begin{cases} i \in S^b, q_i > q^*, \forall \delta_i \\ i \in S^d, 0 < q_i < q^*, \delta_i > \delta^* \end{cases} \quad (14)$$

式(14)中: S^b 为备选基准站点集合; S^d 为备选需求响应站点集合。

步骤 5 由备选基准站点形成基准网络 W , 站间距约束条件为

$$l_{\min} \leq l_{i,j} \leq l_{\max}, i, j \in S^b \quad (15)$$

式(15)中: i 和 j 为连接站点; $l_{i,j}$ 为基准站点 i, j 的直线距离。

步骤 6 网络 W 中的边缘基准站点集合为 S^{bb} , 将其作为线路的两端点, 对任意端点 i, j 利用 KSP 算法生成基准站点链条, 如果线路总长度满足式(16), 则将该基准站点链条纳入基准链条集合 C 。

$$L_{\min} \leq L_{i,j} \leq L_{\max}, i, j \in S^{bb} \quad (16)$$

式(16)中: $L_{i,j}$ 为基准链条长度。

步骤 7 输出该时段所有基准站点链条方案 C , 链条数量为 n 。

3.2.2 确定最优线路

步骤 1 $k=1$ 。

步骤 2 确定基准链条 k 的子链条集合 \hat{C}_k 。对链条 k 的 m 个基准站点进行编号, 对所有 $i, j \in S_k^b$ ($0 < i < j < m$) 形成以 i, j 为起终点的子链条, 并纳入集合 \hat{C}_k , 基准链条 k 的子链条数量为 e 。

步骤 3 确定子链条 h 的需求响应站点。对于任意站点 i, j ($i \in S_{k,h}^b, j \in S^d$), 绕行约束条件为

$$\frac{l_{i,j} + l_{j,j+1}}{l_{i,i+1}} \leq l^* \quad (17)$$

式(17)中: $j \in S_{k,h}^d$; $S_{k,h}^b$ 、 $S_{k,h}^d$ 分别为子链条 h 的基准站点和需求响应站点集合; l^* 为绕行约束的临

界值。

步骤4 计算备选线路 $k-h$ 的 OD 需求 $Q_{k,h}$ 。设基准站点 $i (i \in S_{k,h}^b)$ 和需求响应站点 $j (j \in S_{k,h}^d)$ 服务半径为 r , 某乘客出行起终点为 o 和 d , 若 $\text{dist}(o, i/j) \leq r$ 且 $\text{dist}(i/j, d) \leq r$, 将该需求服务纳入 $Q_{k,h}$, 其中 $\text{dist}(o, i/j)$ 表示起点 o 至站点 i 或 j 的距离, $\text{dist}(i/j, d)$ 表示站点 i 或 j 至终点 d 的距离。

步骤5 计算备选线路 $k-h$ 的评估指标, 包括运营指标和乘客需求指标以及综合评估指标 $E(C_k)$ 。

步骤6 如果 $k < n, k = k + 1$, 则转至步骤2。

步骤7 比较所有备选线路的综合评估指标 $E_{k,h}$, 选择指标最大值对应的线路作为优选方案。

步骤8 将步骤7 优选的线路所涵盖的站点及对应的 OD 需求, 从 S^b 、 S^d 、 Q_t 中剔除, 并更新 S^b 、 S^d 、 Q_t , 转至 3.2.1 节的步骤5。

步骤9 输出 t 时段最优线路方案集合 R_t 。

4 算例验证

利用出租车运营数据获取乘客出行的分布特征, 根据出行特征将出行时段分为第一高峰时段、第二高峰时段和平峰时段, 如图5所示。

针对第一高峰时段第七条基准链条, 通过实证分析, 详细描述了算法分析过程以及获得最优线路方案的步骤。采用相同的算法求解了第二高峰时段和平峰时段的最优线路。

使用 Python 获取第一高峰时段内研究区域范围内的 OD 数据, 采用 K-means 聚类方法确定备选站点, 并进一步将其划分为基准站点和备选需求响应站点。在基准链条的基础上形成备选线路并识别其 OD 客流, 计算线路指标并输出最优线路。通过循环算法, 并最终确定了研究范围内的最优线路集合。

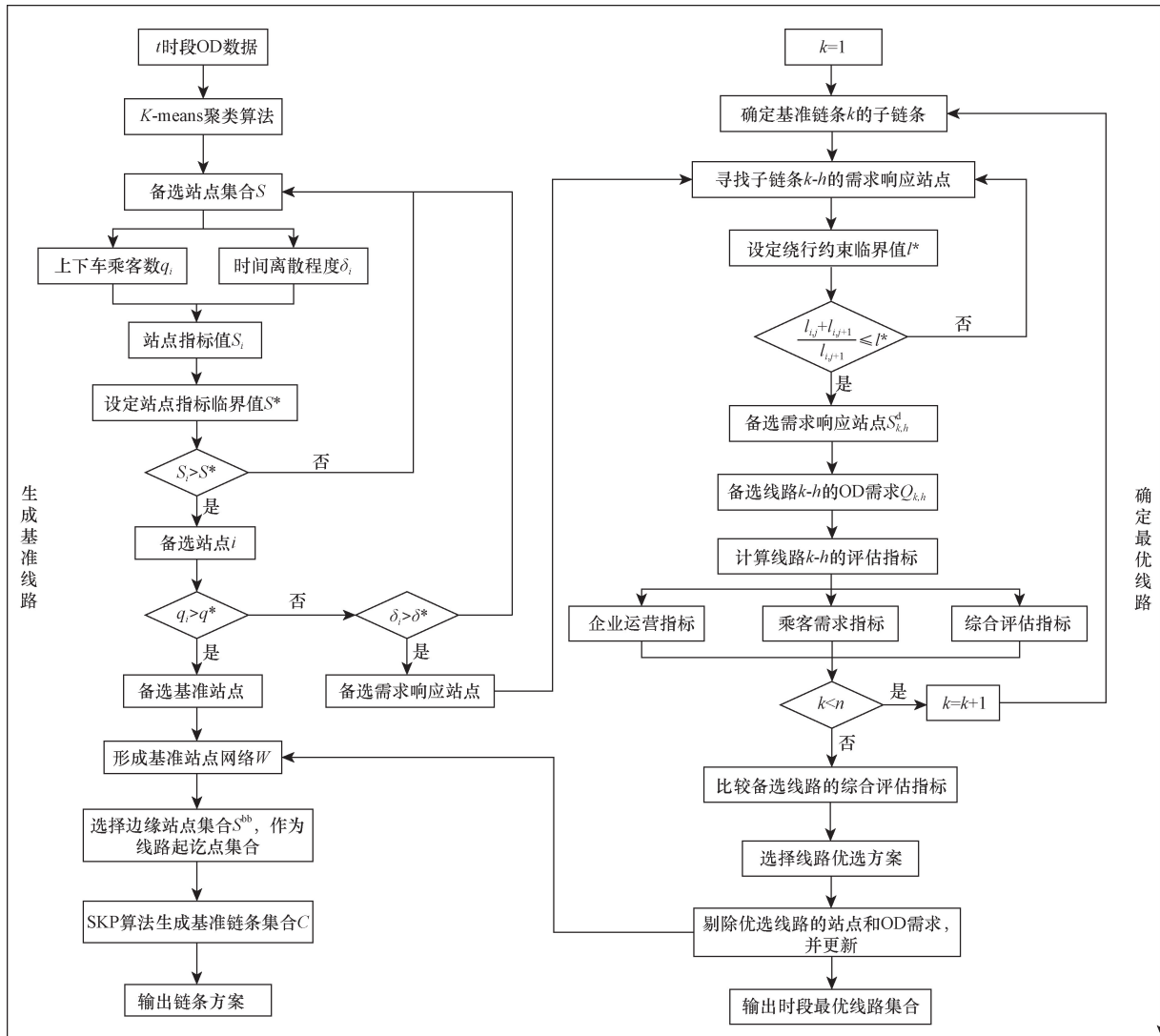


图4 算法流程图

Fig. 4 Algorithm flowchart

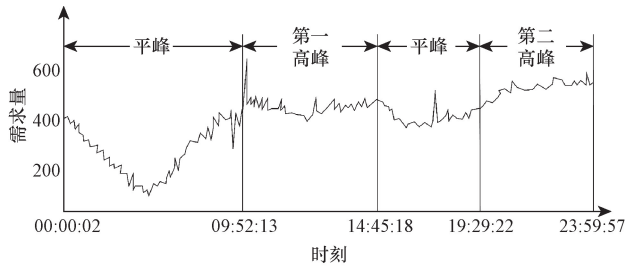


图5 出行时段划分示意图

Fig. 5 Diagram of travel time division

4.1 基准站点链条的生成

第一高峰时段提取到 103 021 条有效 OD 数据,对数据进行预处理。运用 K-means 算法聚类出行密集点,选取 500 个聚类点结果,在 Open StreetMap 上做可视化分析,如图 6 所示。

计算 500 个潜在备选站点的站点指标值 S_i ,设站点指标临界值 $S^* = 1 000$,选出 119 个备选站点。参照图 7 备选站点分类标准示意图,将备选站点划分为基准站点和需求响应站点。

设站点客流量临界值 $q^* = 2 500$ 、时间离散系数临界值 $\delta^* = 0.5$ 。最终选出 119 个备选站点,其中基准站点 16 个,需求响应站点 103 个,备选站点分类结果如图 8 所示。

选出的 16 个基准站点形成全互通网络,站间距约束条件为 $1 < l_{i,j} < 10 (i,j \in S^b)$ (单位为 km),

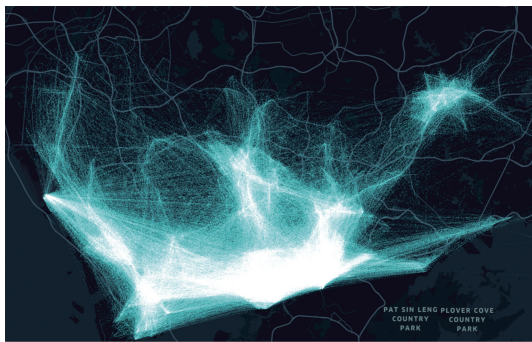


图6 第一高峰时段内有效 OD 分布情况

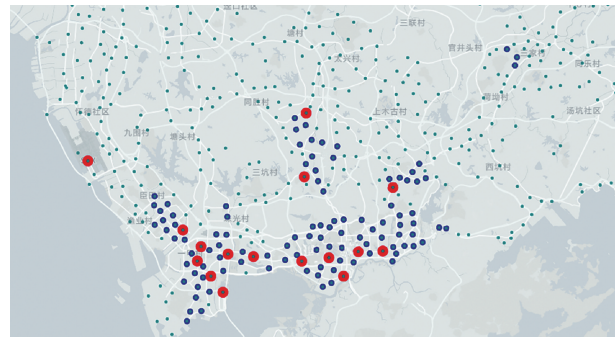
Fig. 6 Effective OD distribution during the first peak hour



图7 备选站点分类标准

Fig. 7 Alternative station classification criteria

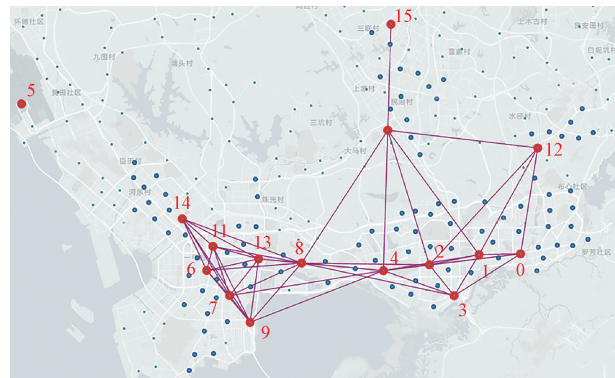
剔除不符合站间距约束条件的基准站点,最终构建的基准站点网络结果如图 9 所示,网络中部分站点信息如表 1 所示,其中 A、B 为两个基准站点。将网络中边缘的基准站点 15、12、0、3、9、14 作为链条的起终点,最终构建 10 条备选基准站点链条,基准链条备选线路方案如表 2 所示。



红色为备选基准站点,蓝色为备选需求响应站点

图8 备选站点分类结果

Fig. 8 Alternative station classification results



红色为备选基准站点,蓝色为备选需求响应站点

图9 第一高峰时段基准站点网络

Fig. 9 The first peak period base station network

表2 基准链条备选线路链条方案

Table 2 Benchmark chain alternative route scheme

编号	基准链条路径
1	15-10-2-1-0
2	15-10-4-2-1-0
3	15-10-4-2-3-1-0
4	15-10-4-8-13-11-14
5	15-10-4-8-13-7-9
6	12-0-1-2-4-8-13-11-14
7	12-0-1-2-4-8-13-7-9
8	12-0-1-3-2-4-8-13-11-14
9	12-0-1-3-2-4-8-13-7-9
10	12-0-1-3-4-8-13-11-14

4.2 最优线路的确定

从基准站点网络中提取连接边缘站点的 K 最短路线作为备选基准站点链条,只对其中第一条~第七

表 1 基准站点网络部分站点情况

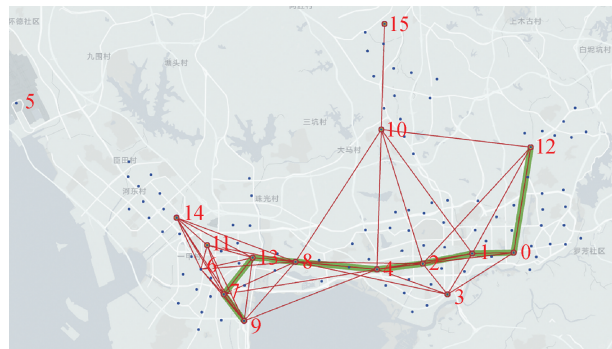
Table 1 Benchmark station network part of the station situation

A 基准站点				B 基准站点			
站点	经度/(°)	纬度/(°)	站点指标	站点	经度/(°)	纬度/(°)	站点指标
0	22.669 60	114.027 4	1 105.390	5	22.611 70	114.025 5	1 388.923
1	22.563 32	113.904 0	1 114.494		222.541 38	113.949 1	1 233.588
1	22.563 22	113.904 0	1 114.494	4	22.548 25	113.922 1	1 347.335
1	22.563 22	113.904 0	1 114.494	8	22.506 72	113.944 1	1 641.552
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
11	22.534 95	114.023 0	2 816.098	15	22.544 16	114.104 1	5 393.093
12	22.521 25	114.064 8	4 270.933	15	22.538 11	114.050 2	5 359.093
13	22.538 11	114.050 2	4 326.331	14	22.544 16	114.079 5	5 157.528
13	22.538 11	114.050 2	4 326.331	15	22.544 16	114.104 1	5 359.093
14	22.543 70	114.079 5	5 157.528	15	22.544 16	114.104 1	5 359.093

条基准链条描述分析过程, $k = 7$, 链条 $C_7 = \{S_{71}^b, S_{72}^b, \dots, S_{7i}^b, \dots, S_{79}^b\}$, $i \in S_7^b, 0 < i < 9$, $S_{71}^b, S_{72}^b, \dots, S_{7i}^b, \dots, S_{79}^b$ 为基准站点。

不同的起终点 i, j 组合构成不同的子链条, 9 个基准站点形成 36 种不同的组合。图 10 为计算得到的第七条链条, 站点顺序为 ‘12-0-1-2-4-8-13-7-9’, 站点信息如表 3 所示。

DRT 公交线路在基准站点之间提供一定程度上的需求响应服务, 但由于受松弛系数的约束, 需求响应站点不能太偏离基准站点连线, 即经过需求



绿线为线路; 红点为备选基准站点; 蓝点为备选需求响应站点

图 10 第一高峰时段第七条基准链条

Fig. 10 Article 7 benchmark chain in the first peak period

表 3 第七条基准链条站点信息

Table 3 Article 7 benchmark chain station information

站点	X/m	Y/m	经度/(°)	纬度/(°)
12	52 765.44	22 463.51	114.114 2	22.602 02
0	51 728.30	16 029.79	114.104 1	22.544 16
1	49 204.53	15 978.32	114.079 5	22.543 70
2	46 194.98	15 357.11	114.050 2	22.538 11
4	43 403.15	15 005.93	114.023 0	22.534 95
8	38 440.33	15 470.78	113.974 6	22.539 13
13	35 824.40	15 720.56	113.949 1	22.541 38
7	34 065.81	13 484.74	113.932 0	22.521 27
9	35 308.78	11 866.28	113.944 1	22.506 72

注: X、Y 为经过将 WGS84 坐标系数据转换为平面直角坐标系下的坐标位置。

响应站点的折线距离与相邻基准站点之间直线距离的比值不能超过容忍松弛系数, 设最大容忍松弛系数 $l^* = 1.3$ 。得到第七条基准站点链条的子链条和需求响应站点分布情况, 如图 11 所示。结合第一高峰时段的 OD 出行数据, 提取起讫点在子链条站点服务区内的 OD 出行数据, 站点服务半径设为 300m, 线路承担的 OD 需求为 1123, 其中 125 个 OD 起讫点站点均为基准站点, 382 个 OD 起讫点均为需求响应站点, 如图 12 所示。



绿色线路是以 S_{72}^b 和 S_{78}^b 为起终点的第七条链条的子链条, 即线路 $C_{7,2} = \{S_{71}^b, S_{72}^b, \dots, S_{78}^b\}$; 红色为基准站点; 蓝色为需求响应站点

图 11 子链条线路 $C_{7,2}$

Fig. 11 Sub-chain route $C_{7,2}$

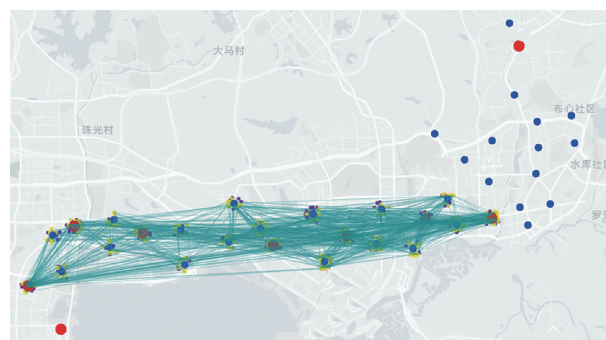


图 12 线路 $C_{7,2}$ 承担的 OD 数量

Fig. 12 Number of OD borne by route $C_{7,2}$

4.3 指标计算

在第一高峰时段有 10 条备选链条线路, 以第七条链条的子链条 $C_{7,2}$ 为例阐述指标计算过程。

4.3.1 企业运营指标

(1) 客流强度 $\rho_{7,2}$ 。线路 $C_{7,2}$ 承担的 OD 数量 $Q_{7,2}$ 为 1 123 个,线路 L 长度为 18.87 km,因此线路 $C_{7,2}$ 的客流强度 $\rho_{7,2}$ 为 129.97。

(2) 基准站点时间离散系数 δ_{72}^b 。从线路 OD 出行中提取到达离开基准站点的时间,部分 OD 信息如表 4 所示。以 10 min 间隔进行计数,其中到达基准站点的 OD 数据有 741 条,离开基准站点的 OD 数据有 125 条,根据式(4)得到基准站点时间离散系数为 0.51。

(3) 需求响应站点离散系数 δ_{72}^d 。从 OD 中提取上车站点为需求响应站点的出发时间以及下车站点为需求响应站点的到达时间。以 10 min 为间隔,根据式(5)得到出行 OD 的离散系数, δ_{72}^d 为 0.57。

4.3.2 乘客需求指标

(1) 平均松弛系数 φ_{72}^d 。 $C_{7,2}$ 包含 16 个需求响应站点,具体信息如表 5 所示。根据表 5 的数据,利用式(6)~式(8)得到 $C_{7,2}$ 的平均松弛系数为 0.097。

(2) 步行距离系数 φ_{72}^b 。从 OD 出发点或者到达点位置和服务半径可以确定乘客上车站点和下车站点,进而可以计算乘客到达上下车站点的步行距离。表 6 为乘客上下车的步行距离情况。根据式(9),平均上车步行距离为 186.8 m,平均下车步行距离为 186.3 m,整体平均步行距离为 186.52 m。设平均服务半径为 300 m,根据式(10),线路 $C_{7,2}$ 的步行距离系数为 0.622。

4.3.3 综合评价指标 $E(C_{7,2})$

对以上 5 个指标进行 Z-score 标准化处理之后评价指标值均在 (-1,1) 范围内。考虑到各种指标的重要程度,设 $\theta_{11} = \theta_{12} = 0.3, \theta_{13} = 0.2, \theta_{21} = \theta_{22} = 0.1$,得到 $C_{7,2}$ 线路的综合评价指标 $E(C_{7,2}) = 0.195$ 。

表 5 线路 $C_{7,2}$ 松弛系数

Table 5 Relaxation coefficient of the route $C_{7,2}$

站点	X/m	Y/m	Quantity_in_300 m/个	φ_{72}^d
57	50 048.08	16 736.06	966	0.171 231
219	50 389.49	15 785.73	1 845	0.014 692
127	47 527.64	16 355.79	1 119	0.101 267
201	48 729.52	14 884.92	1 563	0.226 905
302	47 338.10	15 064.67	3 207	0.060 201
12	44 931.14	16 154.58	1 485	0.210 446
9	45 357.86	14 396.97	2 048	0.180 312
37	40 032.71	14 280.92	894	0.090 443
150	39 879.11	15 559.98	1 285	0.004 881
143	41 905.69	16 569.07	1 470	0.163 579
236	41 700.44	15 128.89	1 631	0.000 118
146	42 908.92	15 627.89	2 485	0.056 419
4	37 183.74	14 933.05	942	0.117 954
183	37 329.09	15 959.63	1 230	0.041 764
373	35 338.38	14 018.87	495	0.107 305
286	34 997.40	15 380.93	544	0.056 999

注: X、Y 为站点坐标; Quantity_in_300 m 为需求响应站点 300 m 半径以内的预约出行数量。

表 6 乘客上下车步行距离长度

Table 6 length of walking distance for passengers to get on and off the bus

ID	上车步行距离/m	下车步行距离/m
70 832	223.912 796	251.628 305
92 231	271.040 583	272.639 097
201 552	297.974 743	252.477 826
237 636	135.473 966	59.196 022
237 852	266.623 061	290.863 097
⋮	⋮	⋮
189 281	157.897 998	66.205 365
246 418	177.995 018	43.563 146
272 568	34.702 092	196.855 805
161 721	188.034 310	204.617 829
357 343	204.273 967	249.307 115

表 4 到达基准站点的 OD 数据信息

Table 4 OD data information arriving at the fiducial stations

车牌号	起点时刻	起点横坐标/m	起点纵坐标/m	终点时刻	终点横坐标/m	终点纵坐标/m
24 343	10:52:49	48 983.346	15 943.472	11:10:57	51 523.064	15 884.205
24 975	11:38:35	51 505.824	15 943.976	11:50:46	49 125.781	15 717.302
28 240	14:10:33	49 494.594	15 910.114	14:15:59	51 969.781	15 954.703
29 420	10:40:14	51 788.539	15 908.446	10:50:24	49 219.779	15 921.122
29 424	14:23:01	51 977.974	15 936.245	14:37:38	48 922.185	15 908.446
32 870	14:19:54	34 105.434	13 499.073	14:28:57	35 251.688	15 487.684
33 950	14:31:26	34 948.654	15 574.861	14:37:28	34 092.093	13 513.750
34 060	13:52:27	34 955.735	15 483.792	13:58:11	34 125.034	13 576.909
⋮	⋮	⋮	⋮	⋮	⋮	⋮
33 950	12:25:39	34 146.174	13 597.369	12:32:04	35 972.689	15 774.789
34 628	11:40:48	35 786.333	15 765.449	11:49:54	34 083.576	13 406.447
35 770	13:36:32	33 933.967	13 256.334	13:47:16	36 008.576	15 769.341
36 384	11:56:53	35 651.697	15 726.642	12:06:15	33 969.977	13 426.796
36 628	14:02:47	33 982.496	13 502.742	14:12:48	35 715.118	15 886.207

不同的起终点 i, j 组合形成不同的子链条 $C_{i,j}$, 计算不同子链条综合评价指标 $E(C_{ij})$, 综合评价指数结果如图 13 所示, 其中横轴和纵轴表示起终点编号。由图 14 可知, $i = 1, j = 8$ 的子链条综合评价指标最大, 为 $E(C_{1,8}) = 0.476$ 。因此子链条 $C_{1,8} = \{S_{71}^b, S_{72}^b, \dots, S_{78}^b\}$ 作为第七条基准链条的最优子链条。

循环多次, 对不同的 k 值进行分析, 比较所有线路的 $E(C_k)$, 结果如表 7 所示, 选择综合评价指标值 $E(C_k)$ 作为第一高峰时段最优线路方案。由于第一、第三、第七条基准线路的综合评价指标值很突出且很接近, 因此三条都可以作为线路的初步方案, 其中第一条和第三条方案站点数量太少, 不适合运营, 被排除。后期线路正式运营考虑站点覆盖率等因素, 可以选择第七条基准线路作为第一高峰时段内最优线路, 线路方案如图 15 所示。

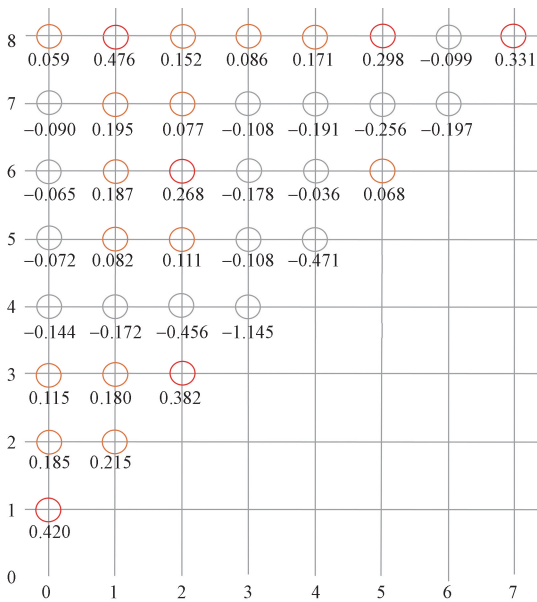


图 13 不同线路的综合评价指标值

Fig. 13 Comprehensive evaluation index value of different lines

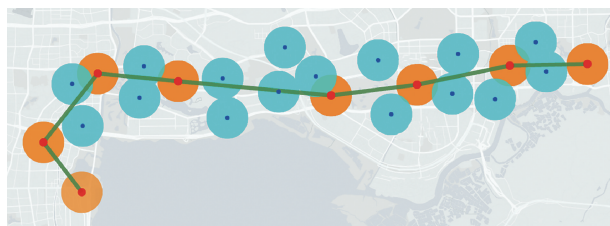


图 14 第一高峰时段最优线路方案

Fig. 14 The optimal route scheme in the first peak period

4.4 不同时段最终线路方案

以上站点可偏移服务模式最优线路是在第一高峰时段根据综合评价指标所确定的最优方案。在其他不同时段的最优线路选取方案与第一高峰

时段的设计思路相同。图 15 和图 16 分别为第二高峰时段和平峰时段的最优线路方案。

表 7 所有基准线路的最优线路

Table 7 The optimal lines of all reference lines

k	最优线路	$E(C_k)$
1	10-2-1-0	0.562
2	10-4-2-1-0	0.422
3	10-4-2-3-1-0	0.490
4	10-4-8-13-11	0.276
5	4-8-13-7-9	0.298
6	2-4-8-13-11	0.251
7	0-1-2-4-8-13-7-9	0.476
8	4-8-13-11-14	0.417
9	3-2-4-8-13-7-9	0.241
10	4-8-13-11-14	0.417

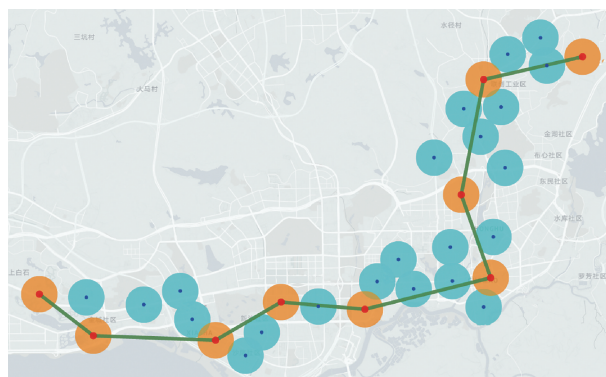


图 15 第二高峰时段最优线路方案

Fig. 15 The optimal route scheme for the second peak period

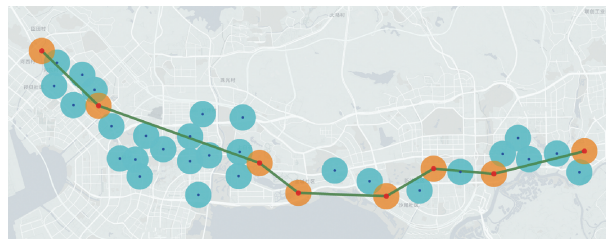


图 16 平峰时段最优线路方案

Fig. 16 The optimal route scheme in flat peak period

5 结论

(1) 在数据挖掘方面, 基于出租车原始轨迹数据进行预处理, 保证了数据分析的准确性。预处理后的出租车轨迹数据集从时间和空间两方面分析乘客的出行偏好、行为模式以及出行特征分布, 为公交线路规划提供时段性的参考依据, 同时也为规划不同时段公交线路提供基础数据。

(2) 在公交线路规划方面, 基于需求响应式公交服务模式与出租车服务模式具有相似性, 可利用预处理后的出租车轨迹数据作为公交出行的“潜在需求”。利用 Python 编程语言实现基准线路和需求

响应站点的算法设计过程。该算法不仅考虑了线路的企业运营效率,还充分考虑乘客的出行需求,以确保公交线路的规划和优化更加贴近实际情况,解决站点选择和线路规划问题。

参 考 文 献

- [1] Flusberg M. An innovative public transportation system for a small city: the Merrill, Wisconsin, case study [J]. *Transportation Research Record*, 1976, 606: 54-59.
- [2] Franco P, Johnston R, McCormick E. Demand responsive transport: generation of activity patterns from mobile phone network data to support the operation of new mobility services[J]. *Transportation Research Part A: Policy and Practice*, 2020, 131: 244-266.
- [3] Giuffrida N, LePira M, Inturri G, et al. On-demand flexible transit in fast-growing cities: the case of Dubai[J]. *Sustainability*, 2020, 12(11): 4455-4469.
- [4] Liu Y N, Ouyang Y F. Mobility service design via joint optimization of transit networks and demand responsive services[J]. *Transportation Research Part B: Methodological*, 2021, 151: 22-41.
- [5] Gong M L, Hu Y C, Chen Z W, et al. Transfer based customized modular bus system design with passenger-route assignment optimization[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2021, 153: 102422.
- [6] 陆百川, 黄镜轶, 张冬梅, 等. 站点需求响应式公交调度系统优化研究[J]. *重庆理工大学学报(自然科学)* 2021, 35(2): 34-43.
Lu Baichuan, Huang Jingyi, Zhang Dongmei, et al. Research on optimization of station demand-responsive public transit dispatching system[J]. *Journal of Chongqing Institute of Technology (Natural Science)*, 2021, 35(2): 34-43.
- [7] Sun B, Wei M, Wu W. An optimization model for demand-responsive feeder transit services based on ride-sharing car[J]. *Information*, 2019, 10(12): 370.
- [8] 孙继洋, 黄建玲, 陈艳艳, 等. 面向多目标站的灵活型公交线路优化调度模型[J]. *交通运输系统工程与信息*, 2019, 19(6): 105-111.
Sun Jiyang, Huang Jianling, Chen Yanyan, et al. Flexible bus route optimization scheduling model for multi-target stations [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2019, 19(6): 105-111.
- [9] 王正武, 宋名群. 多换乘点响应型接驳公交运行线路的协调优化[J]. *中国公路学报*, 2019, 32(9): 164-174.
Wang Zhengwu, Song Mingqun. Coordinated optimization of operation routes for responsive feeder transit systems with multiple transfer points [J]. *China Journal of Highway and Transport*, 2019, 32(9): 164-174.
- [10] 孙贵治. 基于出租车 GPS 轨迹数据的热点区域出行需求预测 [D]. 北京: 北京交通大学, 2019.
Sun Guizhi. Prediction of travel demand in hot spots based on taxi GPS trajectory data [D]. Beijing: Beijing Jiaotong University, 2019.
- [11] Lü Y, Chow C Y, Lee V C S, et al. T2CBS: mining taxi trajectories for customized bus systems [C]//2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). San Francisco: IEEE, 2016: 441-446.
- [12] 刘旭, 陈云波, 施昆, 等. 结合 Canopy-K-means 算法和出租车轨迹数据的公交车站预测方法[J]. *测绘通报*, 2018(11): 63-68.
Lin Xu, Chen Yunbo, Shi Kun, et al. Bus station prediction method combining Canopy-K-means algorithm and taxi trajectory data [J]. *Mapping Bulletin*, 2018(11): 63-68.
- [13] 汪孝之, 牟凤云, 张用川, 等. 利用出租车 GPS 轨迹数据进行短时交通流量预测: 以重庆市解放碑街区为例[J]. *科学技术与工程*, 2023, 23(28): 12265-12274.
Wang Xiaozhi, Mu Fengyun, Zhang Yongchuan, et al. Short-term traffic flow forecasting using GPS track data of cabs: take the Jiefangbei neighborhood of Chongqing City as example [J]. *Science Technology and Engineering*, 2023, 23(28): 12265-12274.
- [14] 李欣然, 贺日兴, 姜超, 等. 北京市出租车 OD 流多尺度空间预测深度重力模型及其可解释性研究[J]. *地球信息科学学报*, 2024, 26(6): 1390-1406.
Li Xinran, He Rixing, Jiang Chao, et al. A multiscale spatial prediction model for taxi OD flow based on deep gravity and its interpretability research in Beijing [J]. *Journal of Geo-information Science*, 2024, 26(6): 1390-1406.
- [15] 李轶群, 石家继, 陈森, 等. GPS 数据驱动的车辆跟驰行为建模与分析[J]. *科学技术与工程*, 2024, 24(1): 408-415.
Li Yiqun, Shi Jiaji, Chen Miao, et al. Modeling and analysis of car following behavior based on GPS data [J]. *Science Technology and Engineering*, 2024, 24(1): 408-415.
- [16] 潘述亮, 俞洁, 卢小林, 等. 灵活型公交服务系统及其研究进展综述[J]. *城市交通*, 2014, 12(2): 62-68, 58.
Pan Shuliang, Yu Jie, Lu Xiaolin, et al. Overview of flexible transit service system and its research progress [J]. *Urban Traffic*, 2014, 12(2): 62-68, 58.