

城市轨道交通客流与站点特征分析

李敏祥¹ 谢佳茵^{1*} 金辉² 李永贞¹

1. 辽宁工业大学汽车与交通工程学院, 辽宁 锦州 121001

2. 辽宁理工职业大学物流学院, 辽宁 锦州 121001

[摘要] 交通客流与站点特征的精准分析, 对缓解城市交通拥堵, 促进轨道交通的良好运营起着至关重要的作用。文中以成都市轨道交通为例, 基于 AFC 数据完成客流特征分析, 并采用改进的 k-means 算法对成都地铁 1、2、3 号线 104 个站点进行聚类, 最终分为居住导向型、就业导向型、职住交错型、职住交错偏居住型、职住交错偏就业型和综合型 6 类站点。望为后续车站发展规划及周边土地利用等提供参考依据。

[关键词] 轨道交通; AFC 数据; 客流特征; k-means 算法

DOI: 10.33142/sca.v7i7.12793

中图分类号: U491

文献标识码: A

Analysis of Passenger Flow and Station Characteristics in Urban Rail Transit

LI Minxiang¹, XIE Jiayin^{1*}, JIN Hui², LI Yongzhen¹

1. School of Automobile and Traffic Engineering, Liaoning University of Technology, Jinzhou, Liaoning, 121001, China

2. School of Logistics, Liaoning Vocational University of Technology, Jinzhou, Liaoning, 121001, China

Abstract: The precise analysis of passenger flow and station characteristics plays a crucial role in alleviating urban traffic congestion and promoting the good operation of rail transit. Taking Chengdu rail transit as an example, this article completes passenger flow characteristic analysis based on AFC data, and uses an improved k-means algorithm to cluster 104 stations on Chengdu Metro Lines 1, 2 and 3. Finally, it is divided into six types of stations: residential oriented, employment oriented, occupational residential staggered, occupational residential staggered, occupational residential staggered, and employment oriented, and comprehensive, which will provide reference for the development planning of subsequent stations and the utilization of surrounding land.

Keywords: rail transit; AFC data; passenger flow characteristics; K-means algorithm

1 研究背景

城市轨道交通因其环保、准时、运量大等优点, 正逐渐成为日常居民出行的主要交通工具之一。然而有限的运输资源、站点客流时空不均等成为急需解决的难题。为此科学分析交通客流特征, 准确识别站点类别, 对于城市轨道交通线网规划起着至关重要的作用。

聚类算法作为数据挖掘的重要技术, 已被广泛应用在人工智能、城市交通等领域^[1]。其中 K-means 聚类算法因其简单高效、可扩展性强, 且对于噪声数据具有一定的容忍性, 获得了广泛应用。大批学者加入了对算法的研究和改进。如陈玉如, 王才雪等基于 Hadoop 平台对深圳通刷卡数据进行客流特征分析, 并采用聚类算法进行站点聚类, 并将站点分为居住区车站、居住办公组团车站、一般车站、交通枢纽区车站以及商业区车站 5 类^[2]。王晨, 石俊刚等结合 AFC 客流数据, 提出一种基于 K-means 算法的双层规划聚类方法对全线所有车站进行聚类并划分车站类型^[3]。但部分学者在完成交通站点聚类时, 对聚类因素及算法考虑不够全面。为此, 本文以成都市为例, 选取 2020 年 1 月 6 日至 1 月 31 日成都地铁 1、2、3、4、5、7 和 10 号线的客流数据为研究对象, 每日客流时段为 5:00~24:00。

基于 AFC 数据分析城市轨道交通客流特征, 运用 K-means 聚类分析法进行站点分类, 以对不同属性的站点合理分配各类资源, 为优化车站运营管理提供参考依据。

2 成都市轨道交通客流特征分析

2.1 AFC 数据及处理

成都地铁于 2010 年开始通车运营, 截至 2023 年末, 已开通线路 13 条, 累计运送旅客 115 亿人/次, 运营里程达 601.7 公里, 已成为成都居民出行的主要方式。本部分依据成都市轨道交通自动收费系统 (Automatic Fare Collection System, 简称 AFC) 记录的历史刷卡数据完成客流分析。笔者首先对刷卡数据进行清洗, 研究数据已无重复和缺失现象。其次, 选取 2020 年 1 月某一工作日的客流数据, 从线路、站点以及时间维度对成都地铁客流数据进行分类统计, 完成客流特征分析。

2.2 客流特征分析

2.2.1 线路客流特征分析

成都市轨道交通不同线路客流量存在明显差异, 但同一线路的进、出站客流量大体相等。无论进站还是出站时间段, 各线路总运量排名为: 2 号线>3 号线>1 号线>4 号线>7 号线>5 号线>10 号线, 如图 1 所示。

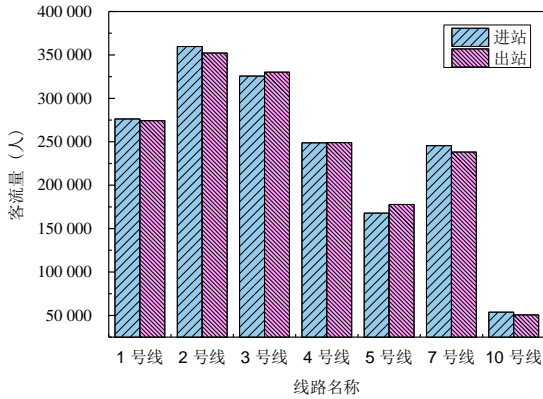


图1 分线路客流量进出站对比图

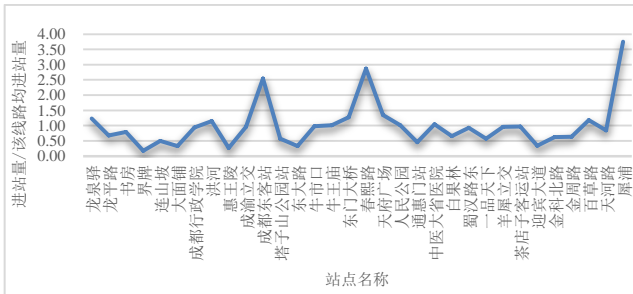


图2 地铁2号线车站进站量均衡度

2.2.2 站点客流特征分析

成都市轨道交通线网共有 386 个站点,由于每个站点所在的区域位置不同,存在各站点客流分布不均衡、少数站点承载着绝大部分客流的现象。以 2 号线为例,计算车站客流的均衡度,如图 2 所示。本文规定以均衡程度介于 0.8 到 1.2 之间作为正常范围。由图 2 可知,成都东客站、春熙路和犀浦站的均衡度分别为 2.55、2.87 和 3.75,远高于正常范围,说明这 3 个站点所在地对乘客的吸引力较强。相反,界牌、大面铺、惠王陵以及迎宾大道车站等站点的均衡度远低于正常范围,可见,这些站点的承载水平尚未达到其应承担的客流运量水平。

2.2.3 出行客流时间特征分析

针对各个分线路客流进行全天 1 小时粒度的客流量分布趋势对比,结果如图 3 所示。

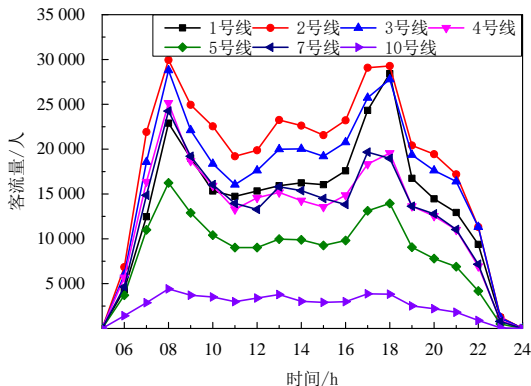


图3 各线路客流量时间分布图

可以看出,全天时间段,各线路的客流变化趋势基本一致,均在 7:00~9:00 时段以及 17:00~19:00 时段客流量有明显的高峰值,为乘客通勤的早晚高峰区间。

3 基于 k-means 算法的轨道交通站点聚类研究

轨道交通站点功能的演变受到区域经济发展、交通区位、站区功能结构等诸多因素的综合影响。科学地针对交通站点分类,可以更好地实现车站与轨道线网的有效配合^[4]。为此本部分以成都地铁 1、2、3 号线连续 20 个工作日的客流数据为研究对象,将客流数据进行主成分分析,并提取主要特征。时间区间为每天 5:00~24:00,以 1 小时为间隔,进、出站客流一共有 38 个小时区间段,38 个成分的方差贡献率达到 100%,且特征值大于 1 的成分有 5 个。为此,这 5 个成分即是提取的主成分,累计贡献率达 94%。采用主成分分析方法识别了车站客流的差异后,将车站的客流特性作为划分变量,利用 k-means 聚类法对车站进行聚类分析。

3.1 改进的 k-means 聚类算法

k 均值聚类算法 (k-means clustering algorithm) 是一种迭代求解的聚类分析算法,传统的 k-means 算法通过随机方式选择初始聚类中心,为提升聚类效果,本文对此进行了改进,采用初始聚类中心之间的距离尽可能远的原则。思路为:首先从初始数据集中随机选择一个对象作为第一个聚类中心,接着计算数据集中每个对象与其的距离,距离越大的对象,被选为下一个初始中心的概率也越大。其算法步骤,见表 1。

表1 改进的 k 均值聚类算法步骤

序号	具体步骤
1	从整个样本集中选取一个数据对象作为第一个聚类中心 C_1
2	计算样本集中的每个对象 x_i 与 C_1 之间的最短距离,记为 $D(x)$ 表示
3	确定 $D(x)$ 值最大的对象作为下一个聚类中心
4	重复操作 2 和 3,直至 k 个初始聚类中心全部选出
5	按照欧式距离最短原则,将数据对象分到距离最近聚类簇中
6	针对聚类结果,计算每个聚类簇中对象的均值,更新聚类中心
7	重复迭代,聚类中心不变时停止,并输出结果

3.2 基于改进的 K-means 算法对轨道交通站点进行分类

3.2.1 确认聚类指标

依据主成分分析结果,为准确描述站点客流特征,本文确定 6 类聚类指标,包括偏度、峰度、高峰小时系数、早晚高峰比、分布均匀系数和客流高峰数。

3.2.2 k-means 算法聚类结果

结合手肘法和轮廓系数法,本文最终确定成都市地铁 1、2、3 号线的 104 个站点最佳聚类数 k 为 6。并按照表 1 的步骤完成 k-means 聚类运算,聚类结果如表 2 所示。

表 2 k-means 算法聚类结果

类别	站点名称	总计
1	韦家碾、四河、广都、五根松、华阳、海昌路、广福、塔子山公园	8
2	锦江宾馆、金融城、孵化园、锦城广场、世纪城、天府三街、天府五街、红石公园、武汉路、兴隆湖、天府广场2、东门大桥、磨子桥	13
3	人民北路、省体育馆1、倪家桥、百草路、牛市口、龙平路、龙泉驿、双流广场、东升、龙桥路、太平园、红牌楼、高升桥、新南门、马超西路、钟楼、石油大学、成都医学院	20
4	升仙湖、华府大道、天河路、羊犀立交、一品天下、蜀汉路东、白果林、东大路、成渝立交、惠王陵、洪河、成都行政学院、大面铺、连山坡、界牌、书房、双流西站、三里坝、迎春桥、航都大街、双凤桥、前锋路、李家沱、驷马桥、昭觉寺南路、动物园、植物园、金华寺东路、三河场、锦水河、团结新区	31
5	文殊院、骡马市、天府广场1、桐梓林、高新、科学城、金科北路、人民公园、牛王庙、武侯立交、衣冠庙、省体育馆3、市二医院、红星桥、天府公园、金周路、迎宾大道、武清南路、川藏立交	19
6	火车北站、华西坝、火车南站、麓湖、西博城、广州路、犀浦、中医大省医院、春熙路2、成都东客运站、春熙路3、熊猫大道、军区总医院	13

4 结论

(1) 基于 AFC 数据分析可知，成都轨道交通存在客

流时空分布不均衡等问题。

(2) 与传统 k-means 算法相比，改进后聚类效果更佳。对成都市轨道交通 1 号线、2 号线和 3 号线站点完成聚类分析，其地铁站点可分为 6 类，分别为居住导向型、就业导向型、职住交错型、职住交错偏居住型、职住交错偏就业型和综合型。

(3) 在后续的地铁运营中，成都市可结合站点聚类结果，采取差异化的行车组织及营销管理方式。同时，改进的 k-means 聚类算法可推广用于其他城市的轨道交通管理中。

【参考文献】

- [1]朱倩. 基于大数据的城市轨道交通客流预测方法研究[D]. 四川: 西南交通大学, 2019.
- [2]陈玉如, 王才雪. 基于深圳通刷卡数据的轨道交通客流与站点特征分析[J]. 交通世界, 2023(26): 4-6.
- [3]王晨, 石俊刚. 基于 AFC 数据的南昌轨道交通车站精细化分类[J]. 都市快轨交通, 2023, 36(6): 49-56.
- [4]夏雪, 盖靖元. 基于 K-Means 聚类算法的城市轨道交通站点分类及客流特征分析[J]. 现代城市轨道交通, 2021(4): 112-118.

作者简介: 李敏祥 (2001—), 男, 硕士, 学生, 车辆工程; 谢佳茵 (1981—), 女, 硕士; 辽宁工业大学汽车与交通工程学院讲师, 载运工具运用工程。