

• 复杂性科学专栏 •



## 地铁乘客站点的选择行为分析及预测

王璞<sup>1,2\*</sup>, 肖健和<sup>1,2</sup>, 李明伦<sup>1,2</sup>, 郭宝<sup>1,2</sup>

(1. 中南大学交通运输工程学院 长沙 410075; 2. 轨道交通大数据湖南省重点实验室 长沙 410075)

**【摘要】**采用数据互补的方法,融合智能卡数据和公交车 GPS 轨迹数据,获取了公交和地铁乘客在地铁新线路开通前后的乘客个体出行信息。利用居民空间行为指标——“出行质心”对地铁乘客的公交出行质心定位,并分析了地铁站点与乘客公交出行质心之间的距离对乘客地铁站点选择的影响。结果表明有 86.15% 的乘客选择距离他们公交出行质心更近的地铁站点,地铁站点与乘客公交出行质心之间的距离是影响乘客出行站点选择的重要因素。基于这一发现,建立 Logit 模型对乘客是否会选择新地铁站点进行预测,预测准确率、精确率、召回率和特异性分别为 83.87%、84.23%、83.66% 和 84.09%,预测结果表现良好。研究结果可用于评估计划新建的地铁站点对邻近既有站点的影响,有助于地铁运营计划的设计。

**关键词** 数据融合; 出行特征; 地铁新开线; 智能卡数据

中图分类号 TU984; U121

文献标志码 A

doi:10.12178/1001-0548.2022031

## Analyzing and Predicting Station Choice Behavior of Subway Passengers

WANG Pu<sup>1,2\*</sup>, XIAO Jianhe<sup>1,2</sup>, LI Minglun<sup>1,2</sup>, and GUO Bao<sup>1,2</sup>

(1. School of Traffic and Transportation Engineering, Central South University Changsha 410075;

2. Rail Data Research and Application Key Laboratory of Hunan Province Changsha 410075)

**Abstract** Data fusion method is used to obtain the individual travel information of bus and subway passengers before and after the opening of the new subway line. The important human mobility index, center of mass, is used to locate the center of mass of passengers' bus boarding locations, and the impact of the distance between a subway station and the center of mass on passengers' choice of subway stations is analyzed. Results show that 86.15% of the passengers choose subway stations closer to their center of mass, and the distance is a key factor affecting passengers' station choice behavior. Based on the findings above, we develop a Logit model to predict whether a passenger will choose to use the new subway station. The prediction accuracy, precision, recall and specificity are 83.87%, 84.23%, 83.66% and 84.09%, respectively, indicating that the model performs well. The results of this study can be used to evaluate the impact of planned new subway stations on the adjacent existing subway stations and contribute to the design of subway operation plan.

**Key words** data fusion; mobility patterns; new subway lines; smart card data

地铁是城市公共交通的骨干,具有速度快、运力大等优点<sup>[1]</sup>。发展地铁被广泛认为是缓解大城市交通拥堵的有效方法<sup>[2]</sup>。但新建或扩建现有地铁线路需要高额投资,因此在规划阶段需要详细评估所规划的线路能否满足居民的实际出行需求。研究地铁网络扩建中乘客的站点选择行为,对于地铁新线路的规划和选址,以及提高地铁新线路的运营管理水平具有十分重要的理论和实际意义。

步行是乘客到达地铁站点的主要方式<sup>[3]</sup>,乘客从出行起点到地铁站点间的步行距离是影响乘客出行选择的重要因素。但在过去的研究中,由于难以获取高空间分辨率的乘客出行起点信息,研究人员通常利用集计模型研究步行距离与站点选择之间的关系。文献 [4] 基于乘客购买火车票的邮政编码数据,以邮政编码区中心点为乘客出行起点,分析了荷兰铁路出行用户对火车站点的出行选择。文献 [5]

收稿日期: 2022-01-24; 修回日期: 2022-05-18

基金项目: 国家自然科学基金(71871224); 湖南省交通运输厅科技进步与创新项目(212102)

作者简介: 王璞(1983-),男,博士,教授,主要从事交通大数据、智能交通等方面的研究。

\*通信作者: 王璞, E-mail: wangpu@csu.edu.cn

基于日本东京交通小区至地铁站点的客流数据,提取交通小区到地铁站点的道路网络距离,构建乘客出行站点选择模型。文献 [6] 采用将个人层面数据汇总的方法,利用上海市人口栅格数据,以栅格中心点和地铁站点之间的距离作为乘客步行距离进行乘客出行站点的选择分析。集计模型以交通小区为研究对象,缺乏乘客个体特征,对模型预测准确性会造成一定的影响。

非集计模型以实际交通出行的个人为单位,研究结果可以更好地反映个体选择行为,因而在新地铁线路的出行需求评估中得到广泛应用。在地铁网络扩建情景相关研究中,国内外研究人员通常基于调查数据获取乘客出行特征和个体特征,使用非集计模型进行乘客个体选择地铁新线路的行为分析。文献 [7] 对希腊雅典的居民进行意向 (stated preference, SP) 调查,包括乘客的出行时间、出行成本、出行目的等出行特征,构建层次极值 Logit 模型探究新地铁线路开通后乘客选择不同交通方式的驱动因素。文献 [8] 利用 SP 调查方法,调查了不同性别、职业、收入、出行目的的乘客在西安新地铁线路开通前后出行方式的选择情况,通过逻辑回归模型分析了更倾向于使用新地铁线路的乘客的个体特征。文献 [9] 借助乌鲁木齐市居民的出行方式选择行为 SP 调查数据,不仅调查了出行者的出行特征和个体特征,还调查了居民对交通信息的获取和采纳情况以及乘客的出行方式选择习惯,构建巢式 Logit 模型,预测了新地铁线路开通后各出行方式的出行分担比例。利用调查数据研究乘客个体选择使用新地铁线路的驱动因素,方法简单易行。但这通常需要耗费巨大的人力和物力资源,并且受样本代表性的影响较大。

近年来,数据驱动的方法被广泛用于研究各类交通问题,如交通流量的估计<sup>[10]</sup>、交通速度分布估计<sup>[11]</sup>、出行需求预测<sup>[12]</sup>等。公交数据的空间分辨率较高,被广泛用于研究公交网络瓶颈路段甄别<sup>[13]</sup>、公交乘客的移动模式<sup>[14]</sup>、通勤模式<sup>[15]</sup>以及来源信息<sup>[16]</sup>。大数据技术和双层交通网络融合方法<sup>[17-18]</sup>的不断成熟,使得大范围研究地铁乘客的站点选择行为成为可能。因此本文通过融合公交、地铁智能卡数据及公交车 GPS 轨迹数据,采用大数据驱动的方法在更精细的空间尺度上分析了乘客公交出行质心与地铁站点之间的距离对乘客选择新地铁站点的影响,并进一步建立 Logit 模型预测乘客是否选择使用新地铁站点。

## 1 数据

### 1.1 深圳市公共交通地理信息数据

深圳市地铁地理信息系统 (geographic information system, GIS) 数据由深圳市交通运输委员会提供。2016 年 10 月 28 日前,深圳地铁共有 6 条线路 (1~5 号线、11 号线), 132 个站点。2016 年 10 月 28 日,深圳地铁 7 号线、9 号线开通运营,站点数量增加到 166 个。7 号线和 9 号线与 16 个换乘站相连。

深圳市公交站点 GIS 数据也由深圳市交通运输委员会提供,深圳市共有公交站点 9 114 个。公交站点密度远高于地铁站点密度,这意味着利用公交站点能够以更高的空间分辨率记录乘客的出行起点位置信息。而且,在地铁新线路投入运营之前,公交站点就已经存在。因此在新地铁站点投入运营之前,其周边的公交乘客出行信息可以用于预测乘客在新地铁站点开通后的出行行为。

### 1.2 乘客智能卡数据

本文所使用的地铁智能卡数据和公交智能卡数据均由深圳市交通运输委员会提供。在两组智能卡数据中,乘客拥有唯一的匿名 ID。因此,可以同时研究一个乘客的公交出行和地铁出行。这为从乘客历史公交出行中推断出该乘客未来的地铁出行起点创造了条件。

地铁智能卡数据的收集时间为 2016 年 8 月-2016 年 12 月,共有 10 775 905 名乘客产生了 599 786 003 条地铁智能卡记录。其中有 12 天数据缺失,本研究仅使用剩余的 141 天地铁智能卡数据。每条地铁智能卡记录包含乘客 ID、记录时间、交易状态和设备编号。根据设备编号可以得到乘客进站或出站的站点 ID。

公交智能卡数据的收集时间为 2016 年 8 月-2016 年 12 月,共有 10 112 676 名乘客产生了 451 814 608 条公交智能卡记录。每条公交智能卡记录包含乘客 ID、公交车牌号和记录时间。

### 1.3 公交车 GPS 轨迹数据

为了推断地铁新线路开通前乘客的公交上车站点,本研究使用了 2016 年 8 月-2016 年 10 月的公交车 GPS 轨迹数据。每条数据记录包含公交车牌号、记录时间、公交车经纬度,在数据记录期内共有 16 192 辆公交车产生了 3 632 007 303 条公交车 GPS 轨迹记录。具体信息如表 1 所示。

表1 深圳市公共交通数据信息表

数据名称	数据信息	数据量/条
地铁地理信息数据	经度、纬度、站点名称、地铁线路名称	166
公交地理信息数据	经度、纬度、站点名称、公交线路编号	9114
地铁闸机数据	站点名称、地铁线路名称、设备编号(前六位)	194
地铁智能卡数据	乘客ID、交易状态、设备编号、记录时间	599786003
公交智能卡数据	乘客ID、交易状态、车牌号、记录时间	451814608
公交车GPS轨迹数据	经度、纬度、线路编号、车牌号、记录时间	3632007303

## 2 方法

为了利用乘客公交出行数据充分探究乘客出行起点与地铁站点间的步行距离对乘客使用新地铁站点的影响, 本文首先分析了地铁站点吸引区域及竞争地铁站点, 并提出了识别被新地铁站点吸引的乘客及未被新地铁站点吸引的乘客的方法。在此基础上, 利用居民空间行为指标——出行质心, 计算乘客的公交出行质心, 并将乘客的公交出行质心估计为乘客的出行起点, 计算乘客出行起点与地铁站点间的步行距离。

### 2.1 识别地铁站点吸引区域和竞争地铁站点

为了确定可能使用新地铁站点的乘客, 首先分析了地铁站点的吸引区域。如图1所示, 以地铁站点为圆心, 半径800 m内的区域被估计为地铁站点的吸引区域<sup>[19-20]</sup>。地铁站点吸引区域内的乘客更偏向乘坐地铁出行。当新地铁站点的吸引区域与既有地铁站点的吸引区域重叠时, 部分乘客可能会由在既有地铁站点乘车转变为在新地铁站点乘车。其中, 将吸引区域与新地铁站点有重叠区域的既有地铁站点定义为竞争地铁站点。

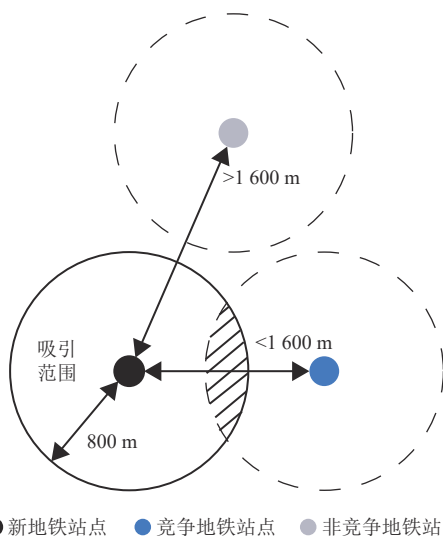


图1 新地铁站点的竞争地铁站点与非竞争地铁站点示意图

### 2.2 识别地铁新线路潜在影响的乘客

研究乘客的站点选择行为需要获取该乘客的历史出行数据。在新地铁站点及其竞争地铁站点的吸引区域内至少有3次公交出行且在新线路开通前平均每周使用地铁大于等于1次的乘客被定义为潜在受影响的乘客。新线路开通后, 平均每周使用新地铁站点大于等于1次, 且使用新地铁站点次数大于使用竞争地铁站点次数的潜在受影响乘客定义为被吸引乘客 $p_a$ ; 新线路开通后, 平均每周使用竞争地铁站点大于等于1次, 且使用竞争地铁站点次数大于使用新地铁站点次数的潜在受影响乘客定义为未被吸引乘客 $p_{na}$ 。

### 2.3 推断乘客公交上车站点

乘客历史公交出行的上车站点数据是计算乘客公交出行质心的基础。采用以下方法获取乘客的公交上车站点。首先将公交车的GPS记录点按时间排序, 将公交轨迹根据公交线路的起终点分为多段公交行程。然后, 计算每个公交车的GPS记录点与该线路中每个公交站点 $k$ 之间的距离, 将每段行程中距离 $k$ 站点最近的GPS点的记录时刻视为公交车 $b$ 在 $k$ 站点的停靠时刻 $t_k^b$ , 依此可以得到所有公交车 $b$ 在各个公交站点的停靠时刻。最后, 对于每个乘客乘车记录 $p$ , 以车辆到达各站点的时刻 $t_k^p$ 作为聚类中心, 以最小时间差为标准, 将乘客的乘车记录时刻 $t_k^p$ 聚类到各个类别中, 各聚类中心的站点 $k$ 为该类别中乘客的上车站点 $k$ <sup>[16, 21-22]</sup>。

### 2.4 计算乘客公交出行质心与新地铁站点、竞争地铁站点间的距离

如图2所示, 新地铁站点与竞争地铁站点吸引范围的并集构成了研究乘客地铁站点选择行为的区域。经统计, 深圳地铁站点800 m吸引范围内平均有26个公交站点, 公交站点的分布密度远高于地铁站点的分布密度。文献[23]研究发现, 公共交通密度更高的区域, 乘客的平均步行距离更短。乘客乘坐公交的平均步行距离要远低于乘坐地铁的平均步行距离, 这意味着公交站点能够以更高的空间分辨率记录乘客的出行起点位置信息。综合考虑每个公交站点的数量和乘客在公交站点的上车次数对乘客的出行起点进行估计。

近年来, 居民空间行为分析与建模领域发展迅速<sup>[24-28]</sup>, 本文利用居民空间行为指标——出行质心<sup>[29]</sup>以及乘客的公交上车站点来计算乘客的公交出行质心, 并将该位置估算为乘客的出行起点。其中, 对



于乘客的公交上车站点, 不考虑乘客从地铁换乘公交时的公交上车记录, 即乘客从地铁出站后 30 分钟内的公交上车记录<sup>[17]</sup>。

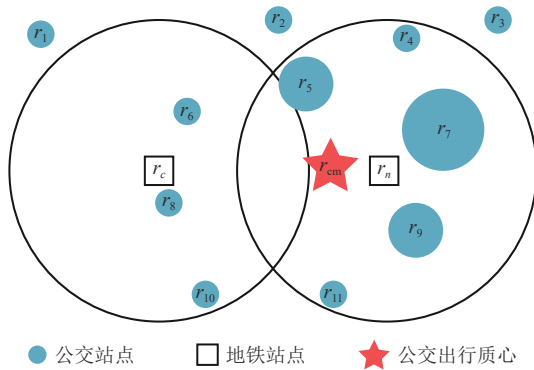


图 2 乘客公交出行质心计算示意图

乘客公交出行质心的计算方法如图 2 所示, 对于每个潜在受影响乘客, 将研究区域内的公交站点 (如  $r_4, r_5, r_6, r_7, r_8, r_9, r_{10}, r_{11}$ ) 视为质点, 不考虑研究区域外的公交站点 (如  $r_1, r_2, r_3$ )。然后将乘客在每个公交站点上的上车次数作为每个质点的权重, 反映在图 2 中为圆圈的大小。最后, 加权平均各个质点的位置得到乘客公交出行质心:

$$r_{cm} = \frac{\sum_i n_i r_i}{N} \quad (1)$$

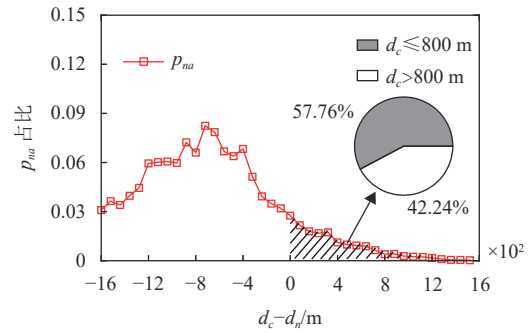
式中,  $i$  为公交站点序号;  $n_i$  为乘客在公交站点  $i$  的上车次数;  $r_i$  为公交站点  $i$  的位置坐标;  $N$  为乘客的公交出行总次数。

在计算了每个乘客的公交出行质心  $r_{cm}$  后, 分别计算乘客公交出行质心  $r_{cm}$  与新地铁站点  $r_n$  和竞争地铁站点  $r_c$  之间的距离, 分别用  $d_n$  和  $d_c$  表示, 用于评估乘客从出行起点前往新地铁站点和相应竞争地铁站点的便利程度。

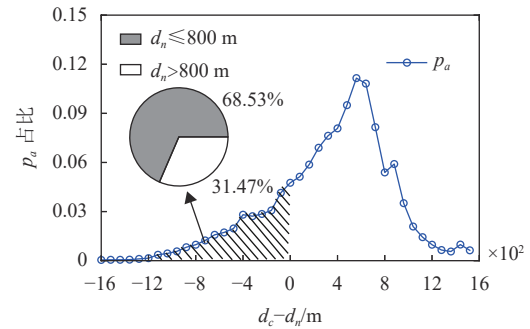
### 3 结果

#### 3.1 乘客地铁出行站点选择行为分析

分别对被吸引乘客  $p_a$  和未被吸引乘客  $p_{na}$  的公交出行质心  $r_{cm}$  与新地铁站点  $r_n$ 、竞争地铁站点  $r_c$  间的距离  $d_n$  和  $d_c$  进行分析, 如图 3 所示。结果表明, 有 79.94% 的被吸引乘客的公交出行质心更靠近新地铁站点 ( $d_c - d_n > 0$ ), 而 86.37% 的未被吸引乘客的公交出行质心更靠近竞争地铁站点 ( $d_c - d_n < 0$ )。结果表明, 大多数乘客 (86.15%) 使用地铁出行时会选择距离他们公交出行质心更近的地铁站点。



a. 选择竞争地铁站点的乘客占比



b. 选择新地铁站点的乘客占比

图 3 不同  $d_c - d_n$  下的乘客地铁出行站点选择概率

少数乘客在使用地铁出行时会选择距离其公交出行质心较远的地铁站点, 这可能是因为乘客的公交出行质心在两个站点吸引区域的重叠区域内, 距离因素的影响有所降低。如图 3 所示, 对于被新地铁站点吸引但质心离竞争地铁站点更近的乘客, 有 68.53% 的乘客质心在新地铁站点 800 m 的吸引范围内; 而未被新地铁站点吸引但质心离新地铁站点更近的乘客, 有 57.76% 的乘客质心在竞争地铁站点 800 m 的吸引范围内。

#### 3.2 乘客地铁出行站点选择预测

Logit 模型是研究出行选择行为时常用的离散选择模型。Logit 模型假设出行者会选择随机效用最高的交通方式, 被广泛应用于交通方式划分问题。本文使用 Logit 模型预测乘客是否会选择乘坐新地铁站点。

本文利用乘客公交出行质心与地铁站点间的距离 ( $d_n$  和  $d_c$ ) 建立 Logit 模型, 选择使用新地铁站点与其相应竞争地铁站点的概率关系为:

$$\frac{p_i^c}{p_i^n} = \frac{e^{V_i^c}}{e^{V_i^n}} \quad (2)$$

其中,

$$V_i^c = \alpha d_i^c + \beta \quad (3)$$

$$V_i^n = \alpha d_i^n + \beta \quad (4)$$

因此,

$$\ln\left(\frac{p_i^c}{p_i^n}\right) = \alpha(d_i^c - d_i^n) + \beta \quad (5)$$

式中,  $p_i^c$ 表示乘客*i*继续选择竞争地铁站点的概率;  $p_i^n$ 表示乘客*i*选择新地铁站点的概率;  $V_i^c$ 表示乘客*i*继续选择竞争地铁站点的效用;  $V_i^n$ 表示乘客*i*选择新地铁站点的效用;  $\alpha$ 和 $\beta$ 为待拟合系数。

本研究随机抽取 50% 的乘客数据集作为模型的训练集, 剩余 50% 的数据集作为模型的测试集。由于数据集中被吸引乘客  $p_a$  和未被吸引乘客  $p_{na}$  样本数之比约为 1:30, 属于典型的类别不平衡问题。本文通过欠采样<sup>[30]</sup>来调整数据的不平衡, 即随机抽取数据集中未被吸引乘客  $p_{na}$ , 使得模型的训练集和测试集中的被吸引乘客  $p_a$  和未被吸引乘客  $p_{na}$  的样本数保持相同。然后借助极大似然估计方法(式(6)), 求得  $\alpha = 1.003$ ,  $\beta = 1.2972$ 。Logit 模型预测结果的混淆矩阵如图 4 所示。

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta) \quad (6)$$

其中,

$$L(\theta) = \prod_{i=1}^n P(X_i = x_i; \theta)$$

式中,  $x_i$ 为样本  $X_i$  的观测值,  $X_i \in X$ ;  $P(X_i = x_i; \theta)$  为总体  $X$  的分布律;  $\theta$  是未知参数,  $\theta \in \Theta$ ,  $\Theta$  是参数空间。

	实际 $P_a$	实际 $P_{na}$
预测 $P_a$	4 668	874
预测 $P_{na}$	912	4 620

图 4 Logit 模型预测结果的混淆矩阵

准确率 (accuracy)、精确率 (precision)、召回率 (recall)、特异性 (specificity) 是确定分类模型性能的常用度量指标<sup>[31]</sup>, 分别为:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad (8)$$

式中, TP 表示实际是  $p_a$  且预测为  $p_a$  的数量; FN 表

示实际是  $p_a$  但预测为  $p_{na}$  的数量; FP 表示实际是  $p_{na}$  但预测为  $p_a$  的数量; TN 表示实际是  $p_{na}$  且预测为  $p_{na}$  的数量。

Logit 模型的准确率为 83.87%, 精确率为 84.23%, 召回率为 83.66%, 特异性为 84.09%。结果表明, 通过引入出行质心度量, Logit 模型能够有效预测潜在受影响乘客出行时是否会使用其公交出行质心附近的新地铁站点或继续使用竞争地铁站点。

Logit 模型是乘客出行选择研究的传统模型。地铁乘客的站点选择问题属于二分类问题, BP (back propagation) 神经网络和支持向量机 (support vector machine, SVM) 是机器学习中解决二分类问题的常用方法, 部分研究人员也通过机器学习中 BP 神经网络<sup>[32]</sup>和支持向量机<sup>[33]</sup>对乘客的出行选择进行预测。

BP 神经网络是一种误差逆向传播的多层前馈神经网络, 该网络包含了输入层、隐藏层和输出层, 本文以乘客公交出行质心与地铁站点间的距离 ( $d_n$ 和 $d_c$ ) 作为输入, 乘客是否选择新地铁站点作为输出。本研究隐藏层取 7 个节点数, 随机数生成器选取种子数为 1, 惩罚参数为 0.00001、采用默认的激活函数 (relu 函数)、最大迭代次数 200、优化容忍度 0.0001。本文首先通过式 (9) 对输入特征进行归一化, 并根据式 (10) 得到隐藏层节点个数  $K$ 。最后以最小化误差为目标, 通过拟牛顿法优化器进行优化求解, 取误差最小的隐藏层节点数的结果作为预测结果:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{f_{\text{std}}(x_j)} \quad (9)$$

式中,  $x_{ij}$ 为样本  $i$  的第  $j$  个特征;  $x_{ij}^*$ 为归一化后的特征;  $\bar{x}_j$ 为第  $j$  个特征的均值;  $f_{\text{std}}(\cdot)$ 为标准差函数。

$$K = \sqrt{m+n} + a \quad (10)$$

式中,  $K$ 是隐藏层的节点数;  $m$ 是输入层的节点数;  $n$ 是输出层的节点数;  $a$ 是 0~10 之间的整数。

SVM 是一种监督学习的分类器, 它通过输入特征构建的特征空间中的超平面, 将待学习样本进行分类。本文基于乘客公交出行质心与地铁站点间的距离 ( $d_n$ 和 $d_c$ ) 作为输入特征, 通过式 (9) 进行归一化, 利用多项式核函数, 对乘客的出行站点选择进行分类学习和预测。本研究使用的错误项惩罚系数为 1, 核函数阶数为 3, 核函数系数为样本特征数的倒数, 核函数独立项为 0, 采用启发式收缩方

式, 取 0.001 为停止训练的误差精度。

3 类模型的乘客站点选择分类预测结果如表 2 所示。

表 2 3 类模型的乘客站点选择预测结果

模型	准确率/%	精确率/%	召回率/%	特异性/%
Logit模型	83.87	84.23	83.66	84.09
BP神经网络	84.31	84.20	84.58	84.05
支持向量机	84.14	84.07	84.34	83.94

3 类模型的预测结果差异不大且都表现良好, 这表明基于乘客出行质心的方法能够有效地对乘客出行站点选择进行预测。Logit 模型在 3 类模型中不仅有更高的精确率 (84.23%) 和特异性 (84.09%), 而且 Logit 模型基于随机效用理论进行选择预测, 相较于机器学习模型有更好的可解释性。

## 4 结束语

1) 通过利用数据互补、跨交通方式数据融合技术, 在更精细的空间尺度上分析了地铁乘客出行站点的选择行为, 弥补了先前研究通过调查获取数据的不足, 提出了研究地铁乘客站点选择行为的新方法。

2) 引入居民空间行为指标——出行质心, 发现乘客公交出行质心与地铁站点间的距离是影响乘客使用新地铁站点的重要因素, 并建立了相应的站点选择 Logit 预测模型, 为探索影响乘客使用新地铁站点的因素提供了新思路。

3) 研究发现乘客通常会选择距离自身公交出行质心更近的地铁站点, 研究有助于在地铁新线路开通前识别被新地铁站点吸引的乘客, 为地铁新线路的选址规划及运营管理提供关键信息。

4) 由于缺乏步行数据, 本文估算的乘客出行起点与实际出行起点存在一定的误差。未来在获取步行数据的情况下, 可在现有研究基础上研究地铁站点的实际吸引区域, 预测新地铁站点开通后从竞争地铁站点到新地铁站点的客流转移。

## 参 考 文 献

- [1] BAO X. Urban rail transit present situation and future development trends in China: Overall analysis based on national policies and strategic plans in 2016–2020[J]. *Urban Rail Transit*, 2018, 4(1): 1-12.
- [2] PAN K. Current situation and development trend of urban traffic engineering management[J]. *BCP Business & Management*, 2021, 13: 366-371.
- [3] DANUELS R, MULLEY C. Explaining walking distance to public transport: The dominance of public transport supply[J]. *Journal of Transport & Land Use*, 2011, 6(2): 5-20.
- [4] DEBREZION G, PELS E, RIETVELD P. Modelling the joint access mode and railway station choice[J]. *SSRN Electronic Journal*, 2007, 45: 270-283.
- [5] 李俊芳, 杜慎旭, 钱卫力. 城市轨道交通车站客流吸引范围重叠区域划分模型[J]. *城市交通*, 2015, 13(6): 4.  
LI J F, DU S X, QIAN W L. Overlapping region partition model for urban rail transit station passenger attraction[J]. *Urban Transport*, 2015, 13(6): 4.
- [6] 刘晓臻. 地铁站可达性和便利性对通勤选择的影响研究——以上海市中心城区地铁站为例[J]. *上海管理科学*, 2018, 40(2): 7.  
LIU X C. Study on the impact of accessibility and convenience of metro stations on the commuters' choice—a case study of metro station in downtown Shanghai[J]. *Shanghai Management Science*, 2018, 40(2): 7.
- [7] GOLIAS J C. Analysis of traffic corridor impacts from the introduction of the new Athens Metro system[J]. *Journal of Transport Geography*, 2002, 10(2): 91-97.
- [8] WANG Y, LI L, WANG Z, et al. Mode shift behavior impacts from the introduction of metro service: Case study of Xi'an, China[J]. *Journal of Urban Planning and Development*, 2013, 139(3): 216-225.
- [9] 曹建青, 王立晓, 孙小慧, 等. 基于 Nested Logit 模型的城市轨道交通客流转移研究[J]. *中国科技论文*, 2017, 12(7): 749-753.  
CAO J Q, WANG L X, SUN X H, et al. Study of passenger flow shift of urban rail transit based on Nested Logit model[J]. *China Sciencepaper*, 2017, 12(7): 749-753.
- [10] WANG P, LAI J Y, HUANG Z R, et al. Estimating traffic flow in large road networks based on multi-source traffic data[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, PP(99): 1-12.
- [11] WANG P, HUANG Z R, LAI J Y, et al. Traffic speed estimation based on multi-source GPS data and mixture model[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, DOI: 10.1109/TITS.2021.3095408.
- [12] WANG P, HUNTER T, BAYEN A M, et al. Understanding road usage patterns in urban areas[J]. *Scientific Reports*, 2012, 2(1): 1-6.
- [13] 王璞, 谭倩, 徐仲之, 等. 城市公交网络运行演化分析与瓶颈甄别[J]. *电子科技大学学报*, 2018, 47(3): 455-461.  
WANG P, TAN Q, XU Z Z, et al. Evolution analysis and bottleneck identification of urban bus network[J]. *Journal of University of Electronic Science and Technology of China*, 2018, 47(3): 455-461.
- [14] 王明生, 黄琳, 闫小勇. 探索城市公交客流移动模式[J]. *电子科技大学学报*, 2012, 41(1): 2-7.  
WANG M S, HUANG L, YAN X Y. Exploring the mobility patterns of public transport passengers[J]. *Journal of University of Electronic Science and Technology of China*, 2012, 41(1): 2-7.
- [15] MA X L, LIU C, WEN H, et al. Understanding commuting patterns using transit smart card data[J].

- Journal of Transport Geography*, 2017, 58: 135-145.
- [16] MA X L, WANG Y H, CEHN F, et al. Transit smart card data mining for passenger origin information extraction[J]. *Journal of Zhejiang University Science C*, 2012, 13(10): 750-760.
- [17] ZHENG Z, HUANG Z, FAN Z, et al. Understanding coupling dynamics of public transportation networks[J]. *EPJ Data Science*, 2018, 7: 1-16.
- [18] GU C G, ZOU S R, XU X L, et al. Onset of cooperation between layered networks[J]. *Physical Review E*, 2011, 84(2): 026101.
- [19] DANIELS R, MULLEY C. Explaining walking distance to public transport: The dominance of public transport supply.[J]. *Journal of Transport and Land Use*, 2013, 6(2): 5-20.
- [20] WEI J, LONG K, GU J, et al. Optimizing bus line based on metro-bus integration[J]. *Sustainability*, 2020, 12(4): 1493.
- [21] 阎逸飞, 于琛, 李彬, 等. 基于上海公交 IC 卡的公交客流走廊 OD 获取方法[C]//2019 年中国城市交通规划年会. 成都: 中国建筑工业出版社, 2019: 3791-3801.  
YAN Y F, YU C, LI B, et al. OD acquisition method of bus passenger flow corridor based on Shanghai bus IC card[C]//2019 China Urban Transport Planning Annual Conference. Chengdu: China Architecture & Building Press, 2019: 3791-3801.
- [22] 王磊. 公共交通数据的融合研究与上海市实践[C]//2017 年中国城市交通规划年会. 上海: 中国建筑工业出版社, 2017: 2595-2605.  
WANG L. Fusion research of public transport data and practice in Shanghai [C]//Proceedings of the 2017 China Urban Transport Planning Annual Conference. Shanghai: China Architecture & Building Press, 2017: 2595-2605.
- [23] ALSHALALFAH B W, SHALABY A S. Case study: Relationship of walk access distance to transit with service, travel and personal characteristics[J]. *Journal of Urban Planning and Development*, 2007, 133(2): 114-118.
- [24] YAN X Y, ZHAO C, FAN Y, et al. Universal predictability of mobility patterns in cities[J]. *Journal of the Royal Society Interface*, 2014, 11(100): 20140834.
- [25] YAN X Y, WANG W X, GAO Z Y, et al. Universal model of individual and population mobility on diverse spatial scales[J]. *Nature Communications*, 2017, 8(1): 1639.
- [26] LIU E J, YAN X Y. A universal opportunity model for human mobility[J]. *Scientific Reports*, 2020, 10(1): 1-10.
- [27] XU F, LI Y, JIN D, et al. Emergence of urban growth patterns from human mobility behavior[J]. *Nature Computational Science*, 2021, 1(12): 791-800.
- [28] YAN X Y, HAN X P, WANG B H, et al. Diversity of individual mobility patterns and emergence of aggregated scaling laws[J]. *Scientific Reports*, 2013, 3: 2678.
- [29] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns[J]. *Nature*, 2008, 453(7196): 779-782.
- [30] DRUMMOND C, HOLTE R C. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling[C]//Workshop on Learning from Imbalanced Datasets II. Washington DC: Citeseer, 2003, 11: 1-8.
- [31] FARRUGIA S, ELLUL J, AZZOPARDI G. Detection of illicit accounts over the Ethereum blockchain[J]. *Expert Systems with Applications*, 2020, 150: 113318.
- [32] GAO J, ZHAO P, ZHUGE C, et al. Impact of transit network layout on resident mode choice[J]. *Mathematical Problems in Engineering*, 2013, 2013(6): 927-940.
- [33] WENG J, TU Q, YUAN R, et al. Modeling mode choice behaviors for public transport commuters in Beijing[J]. *Journal of Urban Planning and Development*, 2018, 144(3): 05018013.

编辑 叶芳

(上接第 622 页)

- [16] DUHAYYIM M A, ALSHAHRANI H M, AL-WESABI F N, et al. Deep learning empowered cybersecurity spam bot detection for online social networks[J]. *CMC-Computers, Materials & Continua*, 2022, 70(3): 6257-6270.
- [17] KIPF T N, WELING M. Semi-Supervised classification with graph convolutional networks[EB/OL]. [2021-09-01]. <http://arxiv.org/pdf/1609.02907.pdf>.
- [18] CHEN M, WEI Z, HUANG Z, et al. Simple and deep graph convolutional networks[C]//International Conference on Machine Learning. Vienna: [s.n.], 2020: 1725-1735.
- [19] LIU L, JIA K. Detecting spam in chinese microblogs - A study on sina Weibo[C]//Proceedings of the 2012 8th International Conference on Computational Intelligence and Security. Guangzhou: [s.n.], 2012: 578-581.
- [20] LIAN Y, DONG X, CHI Y, et al. An internet water army detection supernetwork model[J]. *IEEE Access*, 2019, 7: 55108-55120.

编辑 税红