

基于OD识别和客流预测的城市公共交通 多源数据分析关键技术研究

毕宇航¹, 许玲², 叶晓飞³, 白桦⁴

¹宁波市民卡运营管理有限公司, 浙江 宁波

²宁波市轨道交通集团有限公司, 浙江 宁波

³宁波大学海运学院, 浙江 宁波

⁴华设设计集团股份有限公司大数据及交通数字化规划研究中心, 江苏 南京

收稿日期: 2023年10月20日; 录用日期: 2023年11月10日; 发布日期: 2023年11月20日

摘要

随着城市公共交通体系的逐步完善, 通过大数据等新兴技术手段, 深入研判和预测公共交通客流情况至关重要。目前公交OD (origin-destination) 识别效率及预测效率都比较有限, 客流分配不够合理, 无法有效支撑公交出行特征分析、公交专项规划管理等, 亟需从理论和实践出发, 开展进一步的深入研究和探索, 因此需要对现有技术进行改良。本文结合大数据及机器学习等, 对公共交通数据进行分析, 首先结合Transcad最短路径方法对公交OD识别算法进行优化, 其次分别采用五种模型构建公交短时客流预测算法并进行对比分析。结果表明本文在优化公交OD识别算法后, 算法效率上有较大提升, 同时公交短时客流预测中Tensor + ARIMA方法在效率和精度方面表现较好, 能够更好地适用于该领域的实践应用。

关键词

公共交通, OD识别, 客流预测

Research on Key Technologies of Multi-Source Data Analysis of Urban Public Transport Based on OD Identification and Passenger Flow Prediction

Yuhang Bi¹, Ling Xu², Xiaofei Ye³, Hua Bai⁴

¹Ningbo Citizen Card Operation Management Co., Ltd., Ningbo Zhejiang

²Ningbo Rail Transit Group Co., LTD., Ningbo Zhejiang

³Faculty of Maritime and Transportation, Ningbo University, Ningbo Zhejiang

Abstract

With the gradual improvement of the urban public transportation system, it is crucial to study and predict the public transportation passenger flow in depth through big data and other emerging technical means. The current public transportation OD (origin-destination) identification efficiency and prediction efficiency are relatively limited, and the passenger flow distribution is not reasonable enough to effectively support the analysis of public transportation travel characteristics and public transportation special planning management, etc. Further in-depth research and exploration are urgently needed from theory and practice, so the existing technology needs to be improved. This article combines big data and machine learning techniques to analyze public transportation data. Firstly, it optimizes the public transit OD identification algorithm by integrating Transcad's shortest path method. Secondly, it constructs five different models to develop short-term public transportation passenger flow prediction algorithms and compares them. The results show that after optimizing the public transit OD identification algorithm, there is a significant improvement in algorithm efficiency. Additionally, the Tensor + ARIMA method performs well in terms of efficiency and accuracy for short-term public transportation passenger flow forecasting, making it more suitable for practical applications in this field.

Keywords

Public Transport, OD Identification, Passenger Flow Forecasting

Copyright © 2023 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

随着城市公共交通体系的逐步完善,通过大数据等新兴技术手段,深入研判和预测枢纽出行特征至关重要。目前在公共交通出行方面,受多模式换乘、单次售检票、多渠道支付等因素影响,尚存在公交OD识别率低、预测精度有限、客流分配结果与实际相差甚远等诸多问题,无法有效支撑公交出行特征分析,公交专项规划管理等,亟需从理论和实践出发,开展进一步的深入研究和探索。

在公交OD识别方面,国内主要围绕IC卡刷卡记录、公交GPS(全球定位系统)等相关数据进行研究,主要涉及乘客公交出行的上下站点识别[1][2][3]和公交出行OD矩阵的推导[4][5][6][7]等。国外主要围绕挖掘IC卡数据以及利用其推导公交OD矩阵等方面进行研究[8][9][10][11][12]。

在短时交通流预测方面,国内外通常采用线性回归[13]、卡尔曼滤波法[14]、神经网络[15][16]等模型进行研究。因为短时客流预测对样本信息和预测精度的要求较高,近些年国内外非常注重对组合预测方法的研究[17][18][19][20],并研究出了一系列成果。

本文应用多源大数据融合分析、人工智能机器学习等新兴技术,突破城市综合交通枢纽出行定量分析中的公交OD识别、多方式短时预测两大技术瓶颈,并且本文将扬州市公交数据作为案例对模型进行了精度校核和有效性验证。

本文的贡献总结如下：1) 基于多源数据的公交出行特征分析技术；2) 基于机器学习的分方式客流短时预测技术。

2. 数据库构建

为了便于数据输入输出和后续项目研究，设计了一套公交数据库标准，包含原始数据表、中间数据表、指标数据表三类。数据从扬州市公交数据中选取。

原始数据表包括公交设施类、公交需求类以及公交运行类。公交设施类包括分方向公交线路表、公交站点表、道路路段表和道路节点表。公交需求类包括巴士交易数据表记录乘客刷卡信息。公交运行类包括巴士 GPS 表存储轨迹信息。

中间数据表包括公交设施类、公交需求类以及公交运行类。公交设施类包括同线路公交站间成本表、公交站间路径表和公交站间最优路径表。公交需求类包括同线路巴士出行表配对和公交出行表。公交运行类包括巴士班线表用于存储巴士到达、离开各站点的时间。

指标数据表包括公交需求类以及公交运行类。公交需求类包括公交站间出行客流表和公交小区出行客流表。公交运行类包括公交站点客流指标表、公交线路客流指标表、公交网络客流指标表、公交站点换乘客流表和道路运行指标表。

3. 公交 OD 识别算法

为方便管理各类数据表，基础数据存储采用 SQL Server、算法处理采用 Python 语言编程(Pandas、Geopandas、Numpy 数据模块等)、交通 OTD (出发地 - 出行方式 - 目的地)识别采用 Transcad 最短路径方法。本节输入数据为公交线路走向、物理站点、GPS 及交易原始数据；输出数据为公交 OD、线路、站点、断面等多维度客流指标。

3.1. 建立地图数据

本小节利用公交站点 GIS (地理信息系统)输入物理站点表、线路位表以及线路属性表，将物理站点转换为线路站点(不分方向)，并将垂足坐标加入线路轨迹列表以更新公交站点地理信息数据表，最终的公交站点地理信息数据表如表 1 所示。并利用垂足坐标和站点间直线距离矩阵，输出站间距离表以及步行换乘表。得出的结果利用高德地图进行数据校核，站点基本无缺漏，匹配率较好。

Table 1. Example of a transit stop geospatial data structure

表 1. 公交站点地理信息数据结构示例

站点名称	站点经度	站点纬度	线路编号	站点序号	站点方位角
扬农集团	119.426691	32.372699	69	1	71
文峰塔	119.42825	32.37296	69	2	60
制药厂	119.432293	32.373743	69	3	88
教院附中	119.434716	32.37554	69	4	350

3.2. 乘客上车站点识别

首先对车辆 GPS 数据进行清洗。接下来将 GPS 数据与公交站点 GIS 数据进行关联匹配，根据经纬度来确定公交车与公交站点之间的空间距离[8]：

$$d = R * \arccos \left[\sin(\text{Lat}_{\text{GPS}}) * \sin(\text{Lat}_{\text{GIS}}) + \cos(\text{Lat}_{\text{GPS}}) * \cos(\text{Lat}_{\text{GIS}}) * \cos(\text{Lon}_{\text{GPS}} - \text{Lon}_{\text{GIS}}) \right] \frac{\pi}{180} \quad (1)$$

式中, R 为地球半径, 取 6,371,004 m, d 为站点与车辆的空间距离, Lon_{GPS} 为公交 GPS 轨迹信息表中 Lng 的值, Lat_{GPS} 为公交 GPS 轨迹信息表中 Lat 的值, Lon_{GIS} 为 GIS 中的公交线路站点的经度字段的值, Lat_{GIS} 为 GIS 中的公交线路站点的纬度字段的值。

GPS 数据的记录的時刻并不一定是真实的公交进站時刻, 因此需要设置一个空间误差范围, 即以公交站点经纬度位置为圆心, 设置误差半径 LZ 的圆形识别区域(简称缓冲区), 一般可设为公交站点沿道路方向的最大长度(一般可设为 30 m), 若存在落在缓冲区的 GPS 数据记录点, 且该记录点的速度值小于设定的速度阈值 VZ (一般可设为 3 km/h), 同时满足车辆行驶方向角与站点方向角差值的绝对值 ≤ 60 时, 可确定该记录点为特征点, 即为有效进站点。

若公交站点识别区域内没有特征点, 则寻找同一线路上距离站点最近的两个记录点, 可以假设公交在两点间匀速运行, 通过记录的车辆位置坐标得到这两点间的公交线路长度 d_z 和它们分别与该站点间的线路长度 d_1 和 d_2 , 与两点的记录時刻 t_1 和 t_2 , 可知两点间假定的匀速运行速度 \bar{V} 为:

$$\bar{V} = \frac{d_z}{|t_1 - t_2|} \quad (2)$$

从而得到公交进站的近似時刻 \bar{t}_z 为:

$$\bar{t}_z = t_1 + \frac{d_1}{\bar{V}} \quad (3)$$

或反方向为:

$$\bar{t}_z = t_2 + \frac{d_2}{\bar{V}} \quad (4)$$

由于特征点的速度值最小, 所以可认为特征点记录的時刻极其接近进/出站時刻。在无法确定车辆进、出站時刻的情况下, 以特征点记录時刻为区间中心, 车辆的最大驻站时间为区间半径, 构造出一个驻站识别区间(缓冲区)。当特征点记录時刻为 t_z , 车辆最大驻站时间为 $t_{s,\text{max}}$ (一般设为 30 s), 则驻站识别区间为 $(t_z - t_{s,\text{max}}, t_z + t_{s,\text{max}})$ 。同理, 跳站情形下, 被跳站点的驻站识别区间为 $(\bar{t}_z - t_{s,\text{max}}, \bar{t}_z + t_{s,\text{max}})$ 。

对公交 IC 卡消费记录进行清洗。关联 GPS 数据与公交站点 GIS 信息, 得到 GPS 数据相对应的公交站点, 然后匹配 GPS 数据和 IC 卡消费记录, 得出乘客的上车站点。

3.3. 乘客下车站点识别

首先针对城市公交线路设置一个空间判断阈值, 取两个相邻公交站点间的最大距离, 用 l_{om} (一般为 200 m) 表示。根据前后两次乘车过程的公交 IC 卡刷卡记录来判断这两次公交出行是否构成空间衔接, 判断依据为: 若以后一次乘车过程的上车站点为圆心, l_{om} 为半径的圆形区域内存在至少一个属于乘客前一次乘车过程乘坐的公交线路的上车站点在其线路运行方向上的后继站点, 则认为这两个公交乘车过程可以在空间上构成衔接; 反之则不构成衔接。

假设乘客一天中所有公交乘车过程所对应的车辆运行情况如图 1 所示。根据判断逻辑, 可认为在站点 A 上车的乘车过程与在站点 B 的乘车过程可能构成衔接, 并且站点 S1 和 S2 可能为第一次刷卡记录的乘车过程的下车站点。依次类推, 站点 C、D 刷卡记录的乘车行程同样可以判断是否会在空间上有衔接。

在设置空间判断阈值的基础上, 还需要添加时间判断阈值。只有两者都满足的情况下, 才能判断两次乘车行为构成时空上的衔接。具体来说, 乘客在站点 A 上车的乘车过程与在站点 B 的乘车过程在判断是否形成空间衔接后, 还要判断乘客在站点 S1 或站点 S2 下车的不同情况下, 分别到达站点 B 的時刻。

假设乘客在站点 B 的上车刷卡记录時刻为 t , 乘客在站点 A 乘坐的车辆到达站点 S1 与站点 S2 的特征点记录時刻分别为 t_1 和 t_2 , 站点 S1、S2 与站点 B 的距离分别为 d_1 和 d_2 , 另设 v_{max} 为人的最大步行

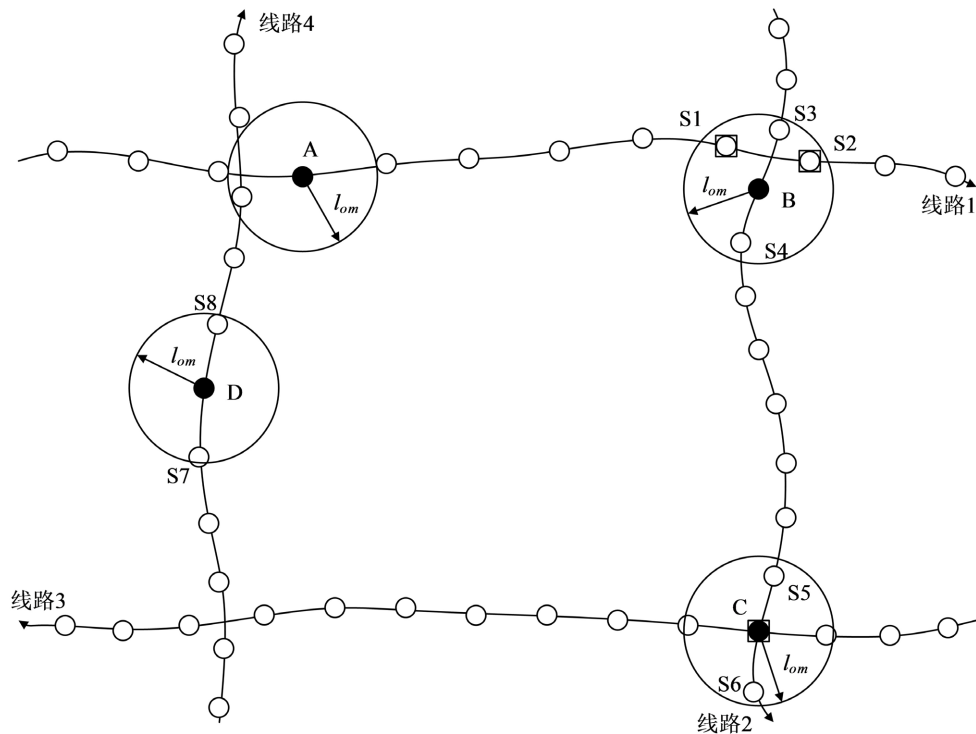


Figure 1. Spatial relationship between front and rear passes
图 1. 前后两次乘车空间关系图

速度，一般设为 1.5 m/s ， t_{cm} 为同一站点经过的相邻的两辆同一线路不同班次的公交车的最大间隔时间，一般设为 10 min ，作为公交出行链上的时间判断阈值。定义 $(t-t_{cm}, t)$ 为换乘行为的判断区间，如处于早晚高峰期，该区间可以扩大 $(t-2t_{cm}, t)$ 。通过上述定义，可做出如下分析，乘客从站点 S1、S2 下车后前往站点 B 的到达的具体时刻分别为 $t_1 + d_1/v_{\max}$ 和 $t_2 + d_2/v_{\max}$ 。通过判断到达时刻与区间的关系，可分辨出下车行为和换乘行为。

因此结合具体情况，可得换乘与下车站点的识别过程如下：

- (1) 若之前没有匹配出上车站点，则下车站点无法确定。
- (2) 若之前匹配出上车站点，且乘客当天只有一次乘车刷卡记录，则无法确定该记录对应的下车站点。
- (3) 若乘客在同一个站点连续刷卡，可能存在代替他人刷卡行为，则无法确定被代替刷卡的乘客下车站点。
- (4) 若乘客当日最后一次刷卡记录所搭乘的线路，就是当日最早搭乘的线路。则分如下几种情况：① 如果两次乘车方向相反或者为循环线，那么该记录对应的下车站点是当前线路上距离当日最早乘车站点最近的站点。② 如果两次乘车方向相同或者为非循环线，那么该记录对应的下车站点是当前线路上距离次日最早上车站点最近且小于 l_{om} 的站点。
- (5) 若乘客当日最后一次刷卡记录中所搭乘的线路与当日最早搭乘的线路不同，那么该记录对应的下车站点是当前线路的当前方向上与当日最早乘车站点距离最近且小于 l_{om} 的站点，如果没有这样的站点则无法确定该记录对应的下车站点。
- (6) 若乘客当前乘车的站点与下一次刷卡记录中乘车的站点不同，且非前五种情况，那么当前乘车过程的下车站点是当前线路的当前方向上距离下一次刷卡的站点距离最近且小于 l_{om} 的站点。
- (7) 在(6)的基础上，区分下车行为与换乘行为：① 当乘客到达下次乘车站点的时刻在换乘行为定义

区间内时,可认为当前下车站点为换乘站点。② 当乘客到达下次乘车站点的时刻在换乘行为定义区间的左侧时,可认为当前下车站点为普通的下车站点。③ 当乘客到达下次乘车站点的时刻在换乘行为定义区间的右侧时,可认为这两次公交乘车过程在时间上不衔接,无法形成出行链。乘客到达下一个乘车站点的时刻如下:

$$t_d = t_u + \frac{d}{v_{\max}} \quad (5)$$

其中 t_u 为车辆到达当前下车站点的时刻, t_d 为乘客到达下一个乘车站点的时刻, d 为从当前下车站点到下一次乘车站点的距离, v_{\max} 为人步行的最大速度。

在判断出下车及换乘行为后,需要在已匹配的上/下车站点中,去除那些已被确定为换乘行为所对应的站点,以明确乘客公交出行的真实目的地/出发地。

整体算法流程下所示:

STEP 01: 对乘客刷卡数据进行出行时间聚类, 30 天内出行天数 ≤ 9 的乘客定义为低频乘客, 出行次数 > 9 的乘客定义为高频乘客。

STEP 02: 对乘客上车站点进行匹配, 统计乘客当天的出行距离和出行时间, 是否满足出行链在空间和时间的阈值范围。

STEP 03: 若乘客当天存在出行链, 通过乘客刷卡时间差或换乘距离作为判别换乘的约束参数, 来判断真实出行起点、换乘、终点; 若乘客当天不存在出行链, 高频乘客下车站点根据该乘客在该线路下游站点中多天下车量确定, 低频乘客下车站点根据所有乘客在该线路下游站点当天下客量确定。

STEP 04: 将 24 小时按照 15 分钟一个时间片(共 96 个时间片)进行聚类, 对每个时间片内的各条线路上客量、下客量、换乘量进行统计分析。

3.4. 运行结果

本文通过上述方法进行 OD 识别, 以扬州城市对比同类规模城市, 算法效率上有较大的优势: 扬州有 273 条线路, 1 个月数据需要算 4.7 小时; 南昌有 752 条线路, 1 个月数据需要算 1 周。两者效率差约 13 倍。

4. 公交短时客流预测算法

本文采用间接法 SARIMA-RF/XGBoost 方法与 CNN-LSTM (或 LSTM-CNN)方法, 和直接法 Tensor-ARIMA (融合张量分解与差分自回归滑动平均法)方法进行公交短时客流预测, 如图 2 所示。

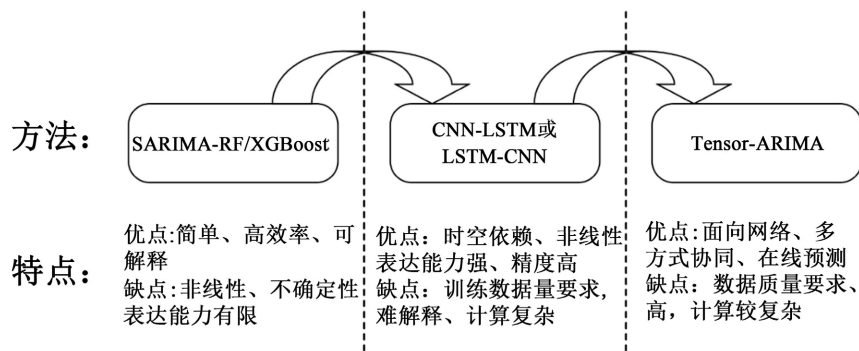


Figure 2. Different methods for short-term passenger flow prediction

图 2. 客流短时预测不同方法

4.1. SARIMA-RF/XGBoost 方法

基于 SARIMA-RF/XGBoost 的公交客流 OD 短时预测基本思想是首先利用 SARIMA-RF/XGBoost 模型对各站点的上/下客流进行预测, 然后以历史公交 OD 作为种子 OD, 利用交通分布模型对未来时间段内的 OD 进行预测。模型结构如下图 3 所示。

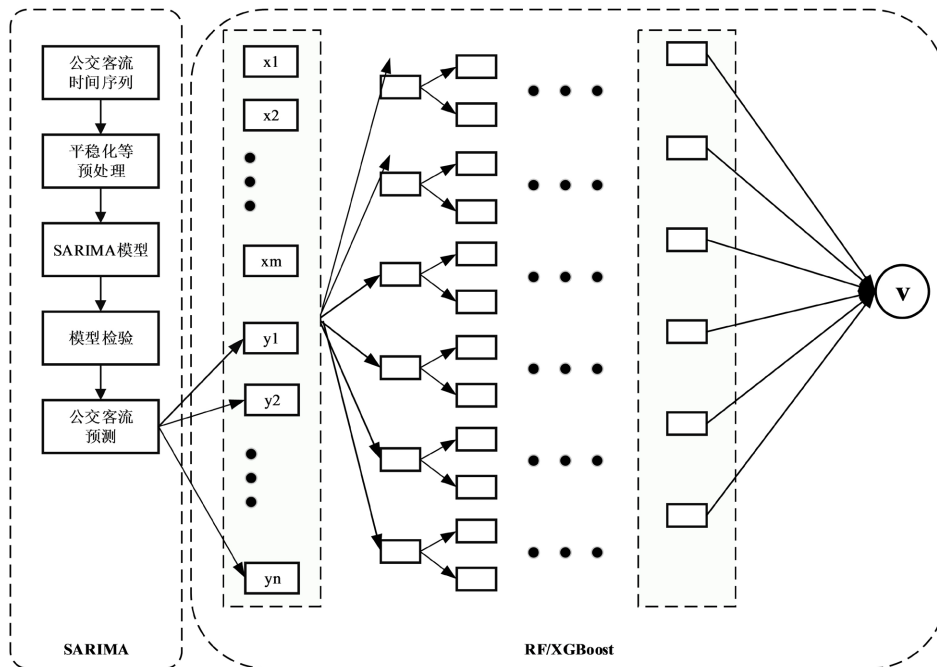


Figure 3. The combined model structure diagram

图 3. 组合模型结构图

首先构建季节性差分自回归滑动平均模型 SARIMA, 利用历史公交客流时间序列数据对未来 n 个时间段内的客流进行预测, 获得 y_1, y_2, \dots, y_n ; 其次提取历史客流、用地业态、天气、接驳设施等客流影响特征, 并与 y_1, y_2, \dots, y_n 融合作为随机森林 RF 与梯度提升机 XGBoost 的输入特征; 然后构建并训练随机森林 RF 或者梯度提升机 XGBoost 模型; 最后利用训练得到的 SARIMA-RF/XGBoost 模型对下一时刻客流量进行预测, 得到 $V(T + \Delta t)$ 。

在各公交站点上客量已知的前提下, 采用分布模型中常用的增长率法估算站间 OD。增长率法的基本原理就是在现状年出行分布矩阵的基础上, 乘以相应的增长函数, 得到规划年的出行分布矩阵。

4.2. CNN-LSTM 或 LSTM-CNN 方法

基于深度学习的公交客流 OD 短时预测首先利用 LSTM-CNN 或 CNN-LSTM 深度学习模型对各站点的上/下客流进行预测, 然后以历史公交 OD 作为种子 OD, 利用交通分布模型对未来时间段内的 OD 进行预测。模型结构如图 4 所示。

LSTM-CNN 深度学习组合模型如图 4(a)所示。LSTM-CNN 模型由一个初始 LSTM 层构成, 它将接收公交站点上/下客流序列数据中每一个时间段内的上/下客流作为输入。然后, 将 LSTM 层的输出输入到可以提取局部特征的卷积层中, 提取空间依赖特征。最后, 卷积层的输出通过池化将被汇集到一个较小的纬度, 并串联为特征向量输入至全连接层。通过对全连接层进行 dropout 等操作后, 向后逐层学习输

入的抽象特征，以此用于预测站点上/下客流。

CNN-LSTM 深度学习组合模型如图 4(b)所示。与 LSTM-CNN 模型不同，CNN-LSTM 模型的初始层由卷积层构成，以网络中不同站点在各个历史时间段内的上/下客流作为输入，提取各站点之间的空间相关特征。然后，将卷积层的输出通过池化运算汇集到一个较小的尺寸，通过 dropout 等操作后，联接成为向量输入至 LSTM 层，通过 LSTM 循环组件提取历史上不同时间段内上/下客流的时间相关特征。最后，将 LSTM 的输出作为全连接层的输入向量，向后逐层学习模型输入的抽象特征表达，并基于该特征表达对公交站点未来时间段内的上/下客流进行预测。

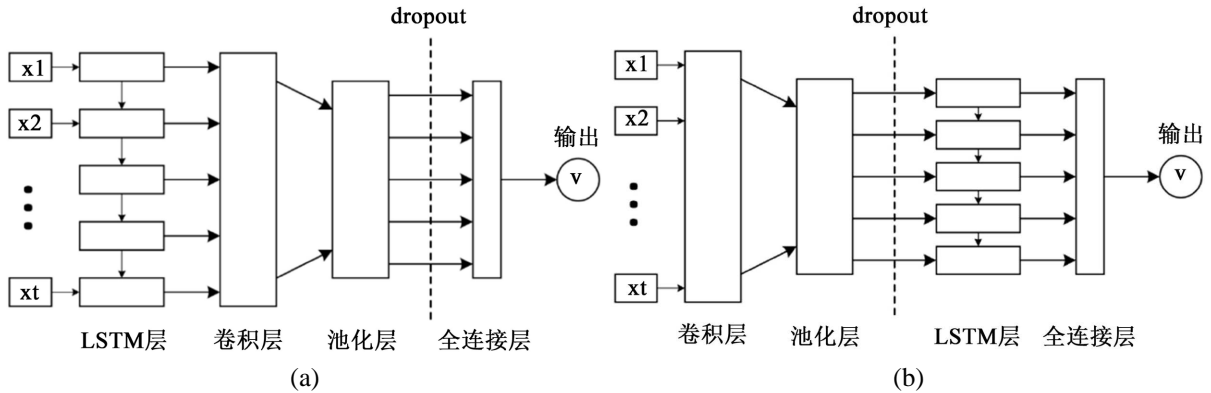


Figure 4. Deep learning model for OD prediction of bus passenger flow

图 4. 公交客流 OD 短时预测深度学习模型

后续仍用交通分布模型中的增长率法进行公交 OD 短时预测。

4.3. Tensor-ARIMA 方法

基于 Tensor-ARIMA 的公交客流 OD 短时预测直接面向客流 OD 矩阵，以历史时间段内的客流 OD 矩阵数据作为输入，通过结合张量构建、张量分解、ARIMA 序列预测、张量重构等步骤，直接对未来时间段内的公交客流 OD 进行预测，如图 5 所示。

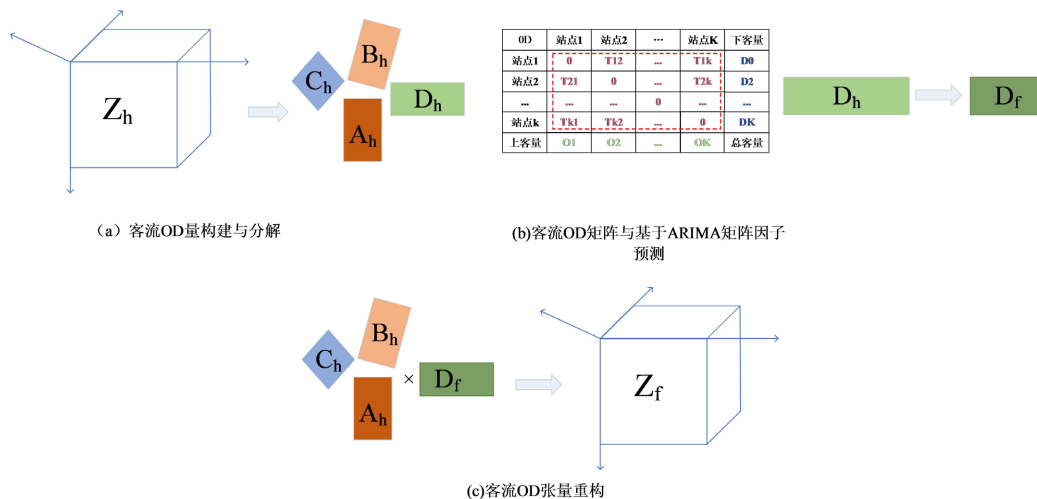


Figure 5. Short-time prediction of OD of bus passenger flow based on Tensor-ARIMA

图 5. 基于 Tensor-ARIMA 的公交客流 OD 短时预测

基于 Tensor-ARIMA 的公交客流 OD 短时预测模型算法流程可以大致划分为以下步骤:

输入: T 个历史时间段内的 OD 矩阵 $T(h)$;

输出: $T + n$ 时间段内的 OD 矩阵 $T(f)$;

STEP 01: 模型建立。基于历史 OD 数据, 通过数据预处理, 构建当前时间段 OD 张量 $T(h)$ 。

STEP 02: 参数标定。利用 CP 或者 Tucker 方法对 OD 张量进行分解, 得到因子矩阵。

STEP 03: 因子预测。利用 ARIMA 等时间序列预测方法对时间因子矩阵进行预测, 获得未来时间段内的时间因子矩阵。

STEP 04: OD 预测。利用 CP 或者 Tucker 方法对 OD 张量进行重构, 得到未来时间段内的 OD 预测值 $T(f)$ 。

4.4. 运行结果

根据上述三类五种模型, 本文对其精度和效率进行计算, 其中精度由平均绝对百分比误差和均方误差表示, 效率由训练时间和预测时间表示。运算结果如下表 2 所示。

Table 2. Operation results of three kinds of five models
表 2. 三类五种模型运算结果

模型	平均绝对百分比误差 MAPE (%)	均方误差 MSE	训练时间(s)	预测时间(s)
SARIMA + RF	14.76	4.72	76030	122 + 分布
SARIMA + XGBoost	14.59	4.49	78492	135 + 分布
CNN + LSTM	14.91	4.96	60437	53 + 分布
LSTM + CNN	15.24	5.15	56484	45 + 分布
Tensor + ARIMA	14.26	4.11	816	62

从上表可以得出, 从平均绝对百分比误差和均方误差来看, Tensor + ARIMA 的精度较好, SARIMA + XGBoost 其次。从训练时间和预测时间来看, Tensor + ARIMA 的效率最快, LSTM + CNN 和 CNN + LSTM 其次。

5. 结论

本文对公交 OD 识别和短时客流预测算法进行了优化设计, 以扬州市公交数据对模型进行精度校核和有效性验证。运行结果表明: 通过优化算法进行 OD 识别, 以扬州城市对比同类规模城市, 算法效率上有较大的优势; 五种短时客流预测算法中 Tensor + ARIMA 在精度和效率上表现最佳。

基金项目

基于 GPS 和 IC 卡数据的公交出行 OD 识别技术研究(HS2022000100), 宁波市交通运输科技项目(202214), 国家自然科学基金(52272334)。

参考文献

- [1] 陈绍辉, 陈艳艳, 赖见辉. 基于 GPS 与 IC 卡数据的公交站点匹配方法[J]. 公路交通科技, 2012, 29(5): 102-108.
- [2] 吴祥国. 基于公交 IC 卡和 GPS 数据的居民公交出行 OD 矩阵推导与应用[D]: [硕士学位论文]. 济南: 山东大学, 2011: 63-66.

- [3] 贾奕宁. 基于下车和换乘行为一体化识别的公交客流 OD 推算研究[D]: [硕士学位论文]. 武汉: 华中科技大学, 2019.
- [4] 吴子啸, 任西锋, 胡静宇. 基于公交 GPS 和 IC 卡数据的综合交通建模新思路[J]. 城市交通, 2011, 9(1): 47-51.
- [5] 章威, 徐建闽. 基于 GPS 与 IC 卡的公交 OD 量采集方法[J]. 交通与计算机, 2006, 24(2): 21-23.
- [6] 徐建闽, 熊文华, 游峰. 基于 GPS 和 IC 卡的单线公交 OD 生成方法[J]. 微计算机信息, 2008, 24(22): 221-223.
- [7] 章玉. 基于数据挖掘的动态公交客流 OD 获取方法研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2010.
- [8] Farzin, J.M. (2008) Constructing an Automated Bus Origin-Destination Matrix Using Farecard and Global Positioning System Data in São Paulo, Brazil. *Transportation Research Record Journal of the Transportation Research Board*, **2072**, 30-37. <https://doi.org/10.3141/2072-04>
- [9] Ma, X., Wu, Y.J., Wang, Y., et al. (2013) Mining Smart Card Data for Transit Riders' Travel Patterns. *Transportation Research Part C Emerging Technologies*, **36**, 1-12. <https://doi.org/10.1016/j.trc.2013.07.010>
- [10] Hofmann, M. and O'Mahony, M. (2005) Transfer Journey Identification and Analyses from Electronic Fare Collection Data. *Proceedings of the 8th International IEEE Conference on Intelligent Transportation System*, Vienna, 13-16 September 2005, 34-39.
- [11] Hofmann, M., Wilson, S.P. and White, P. (2009) Automated Identification of Linked Trips at Trip Level Using Electronic Fare Collection Date. *Transportation Research Board 88th Annual Meeting*, Washington DC, 11-15 January 2009, 18 p.
- [12] Kusakabe, T. and Asakura, Y. (2014) Behavioural Data Mining of Transit Smart Card Data: A Data Fusion Approach. *Transportation Research Part C Emerging Technologies*, **46**, 179-191. <https://doi.org/10.1016/j.trc.2014.05.012>
- [13] 李小红, 王申令. 带线性约束的多元线性回归模型参数估计[J]. 统计研究, 2016, 33(11): 86-92.
- [14] 张春辉, 宋瑞, 孙杨. 基于卡尔曼滤波的公交站点短时客流预测[J]. 交通运输系统工程与信息, 2011, 11(4): 154-159.
- [15] 杨信丰, 刘兰芬. 基于 AP 聚类的支持向量机公交站点短时客流预测[J]. 武汉理工大学学报(交通科学与工程版), 2016, 40(1): 36-40.
- [16] Costarelli, D. and Vinti, G. (2016) Pointwise and Uniform Approximation by Multivariate Neural Network Operators of the Max-Product Type. *Neural Networks*, **81**, 81-90. <https://doi.org/10.1016/j.neunet.2016.06.002>
- [17] 施原. 线性回归在世博会散客流量预测中的实验[J]. 淮阴工学院学报, 2010, 19(4): 49-56.
- [18] 李得伟, 颜艺星, 曾险峰. 城市轨道交通进站客流量短时组合预测模型[J]. 都市快轨交通, 2017, 30(1): 54-58, 64.
- [19] 韩媛媛. 成长生时序数据挖掘在系统运营管理中的分析与应用计算机应用与软件[J]. 2010, 27(11): 160-162.
- [20] 李晓俊, 吕晓艳, 刘军. 基于径向神经网络的铁路短期客流预测[J]. 铁道运输与经济, 2011, 33(6): 86-89.