

建成环境对共享单车时间集聚模式的非线性影响

戢晓峰, 邓若凡, 乔 新, 关昊天

(昆明理工大学 交通工程学院, 昆明 650500)

摘 要: 针对共享单车潮汐式供需失衡问题, 从社会人口、交通设计与土地利用 3 个维度构建了建成环境指标体系, 构建基于凝聚层次聚类模型提炼共享单车典型时间集聚模式, 使用随机森林与沙普利加法解释模型揭示两者之间的非线性关系与阈值效应, 并以云南省昆明市为例进行验证。结果表明: 共享单车时间集聚模式主要为先入后出模式与先出后入模式。建成环境与共享单车时间集聚模式间存在显著的非线性关系, 共享单车时间集聚模式主要受公交线网密度、公园景点邻近度、地铁站点邻近度和公司企业 POI 密度影响。地铁站点邻近度 < 0.5 km 时, 对共享单车时间集聚表现为先入后出模式概率具有抬升作用; 公司企业密度 < 25 个/栅格时, 对其具有压降作用。

关键词: 交通运输系统工程; 共享单车; 时间集聚模式; 随机森林分类器; 非线性影响; 建成环境

中图分类号: U491.225 **文献标志码:** A **文章编号:** 1671-5497(2025)07-2233-10

DOI: 10.13229/j.cnki.jdxbgxb.20231050

Nonlinear influence of built environment on temporal aggregation modes of shared bicycles

Ji Xiao-feng, DENG Ruo-fan, QIAO Xin, GUAN Hao-tian

(College of Transportation Engineering, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Aiming at the problem of tidal supply-demand imbalance of shared bicycles, a built environment indicator system was constructed from the three dimensions of socio-demographic, transportation design, and land use. Based on the Agglomerative Hierarchical Clustering model, typical time-aggregation patterns of shared bicycles were obtained from clustering. The nonlinear relationship and the threshold effect between them were revealed by using the Random Forest and SHapley Additive exPlanations models. Ultimately, Kunming was an example to prove that. The results show that the temporal aggregation modes of shared bicycles mainly show the first-in-last-out mode and the first-out-last-in mode, accounting for 91.4% of the sum of all modes. There is a significant nonlinear relationship between the built environment and the temporal aggregation modes of shared bicycles. The temporal aggregation modes of shared bicycles are primarily influenced by the density of the bus line network, proximity of parks and attractions, proximity of metro stations, and POI density of companies. When the proximity of metro stations is less

收稿日期: 2023-10-05.

基金项目: 云南省交通运输厅科技创新及示范项目(2022-27(二)).

作者简介: 戢晓峰(1982-), 男, 教授, 博士. 研究方向: 交通大数据. E-mail: yiluxinshi@sina.com

than 0.5 km, there is an elevating effect on the probability of showing the first-in-last-out mode. When the POI density of companies is less than 25 per cell, there is a depressing effect on the probability of showing the first-in-last-out mode.

Key words: engineering of communication and transportation system; shared bicycle; temporal aggregation modes; random forest classifier; nonlinear influence; built environment

0 引言

共享单车作为解决城市出行“最后 1 km”问题的出行方式^[1],在我国各大城市得到了广泛的应用。然而,共享单车大量堆积、潮汐式供需失衡现象愈发普遍,极大降低了共享单车与城市公共交通的接驳效率。该问题的本质在于区域共享单车需求量与供给量在时间维度上不匹配,表现为区域共享单车集聚数量在时间维度上的变化。已有研究表明,土地利用类型是影响共享单车时空均衡性的主要因素^[2],因此,为揭示土地利用与潮汐式供需失衡现象之间的关系,亟需明晰不同区域共享单车集聚数量在时间维度上的变化模式,研究建成环境对共享单车时间集聚模式的影响机理,为城市共享单车管理与交通规划提供理论依据。

目前,建成环境对共享单车时空出行特征的影响逐步成为研究热点。现有研究主要集中于建成环境与共享单车出行量或到达量的关系,如 Faghih 等^[3]考察了加拿大蒙特利尔市共享单车使用量的影响因素,发现离中央商务区越近,共享单车的出发量越大。黄颢昊等^[4]以北京市路段作为研究单元,发现娱乐设施数量与共享单车流量之间存在负相关关系。马新卫等^[5]以南京市为案例,发现娱乐型兴趣点(Point of interest, POI)在郊区和主城区对共享单车需求的影响存在异质性。上述研究多采用线性研究方法,多元线性回归模型被广泛用于研究建成环境对共享单车的使用影响重要程度^[6];广义线性回归模型常被用于预测不同建成环境区域共享单车出行量的大小^[7];空间自回归模型则被用于研究建成环境对共享单车出行的空间异质性影响^[8]。然而, Ding 等^[9]的研究表明非线性效应将使建成环境的作用机理发生变化;高楹等^[10]发现建成环境与共享单车集聚间存在非线性关系,而现有研究普遍忽略了建成环境对共享单车使用的非线性作用。同时,现有研究局限于分别研究建成环境与共享单

车出行量或到达量的关系,难以精确刻画土地利用对共享单车时间集聚模式的影响,无法为解决共享单车潮汐式供需失衡问题提供有效的理论支撑。

因此,本文统计各研究单元共享单车净流入量时间序列,基于凝聚层次聚类模型(Agglomerative hierarchical clustering, AHC)对时间集聚变化相似的曲线进行聚类,得到典型时间集聚模式,构建随机森林(Random forest, RF)和沙普利加法解释(Shapley additive explanations, SHAP)模型,探究城市建成环境对共享单车时间集聚模式的非线性影响,为共享单车停放管理与调度提供理论依据。

1 研究区域与变量选取

1.1 研究单元划分

本文选取城市共享单车出行热点区域(所有共享单车骑行区域中骑行量约前 30% 的区域^[11])作为研究范围,借助 ArcGIS 软件数据管理工具中的生成渔网功能进行栅格划分,选取边长为 500 m^[12]的完整正方形栅格作为研究单元,使用叠加分析工具中的空间连接功能计算区域内路网密度、交叉口密度以及 POI 密度等指标;并使用邻域分析工具中的生成近邻表功能计算公共交通站点及各类 POI 可达性指标。

1.2 自变量选取

基于建成环境“5D”模型(即密度、多样性、街区设计、目的地可达性与交通临近度)^[13],本文围绕社会人口、交通设计和土地利用 3 个维度对自变量进行选取。已有研究证明,共享单车的时空集聚性易受到公司、商业区、公园等土地利用类型的影响^[2],因此,本文选取公司企业、购物、餐饮、居住、体育休闲与公园 6 类 POI 密度对密度进行表征;选取公交站点、交叉口、路网以及公交线网密度指标对街区设计进行表征;选取商圈以及公园景点邻近度对目的地可达性进行表征;选取地铁站点以及公交站点邻近度对交通邻近度进行表

征;此外,选取现有研究中常用的土地利用混合度指标来表征多样性^[14],进而构成本文的自变量指标体系,共计 15 个自变量,自变量指标体系具体描述见表 1。

表 1 建成环境指标信息
Table 1 Index information on built environment indicators

维度	指标名称	指标释义	单位
社会人口	居住密度	栅格内居住 POI 数量	个/栅格
	公交站点密度	栅格内公交站点数量	个/栅格
	交叉口密度	栅格内交叉口数量	个/栅格
	路网密度	栅格内道路总长度	m/栅格
交通设计	公交线网密度	栅格内公交线路总长度	km/栅格
	地铁站点邻近度	栅格中心到最近地铁站的直线距离	km
		栅格中心到最近公交站的直线距离	m
	公交站点邻近度		
土地利用	体育休闲 POI 密度	栅格内体育休闲 POI 数量	个/栅格
	公司企业 POI 密度	栅格内公司企业 POI 数量	个/栅格
	公园景点 POI 密度	栅格内公园景点 POI 数量	个/栅格
	购物服务 POI 密度	栅格内购物服务 POI 数量	个/栅格
	餐饮服务 POI 密度	栅格内餐饮服务 POI 数量	个/栅格
	商圈邻近度	栅格中心到最近商场或超级市场距离	m
	公园景点邻近度	栅格中心到最近公园景点距离	m
	土地利用混合度	栅格内 5 类 POI 的熵指数	—

土地利用混合度通过各个栅格内体育休闲、公司企业、公园景点等 5 类 POI 计算得出的熵指数进行表征,其计算公式为:

$$E_i = \frac{-\sum p_{ij} \cdot \ln p_{ij}}{\ln A_i} \tag{1}$$

式中: E_i 为第 i 个栅格的土地利用混合度; p_{ij} 为第 i 个栅格内第 j 类 POI 在所有类别 POI 总数中的占比; A_i 为第 i 个栅格内 POI 类别数。

1.3 因变量构建

为描述共享单车时间集聚模式,需要计算不同时间各栅格内共享单车的集聚数量,本文选择 1 h 作为研究时间粒度,用共享单车净流入量 I_{it} 对其进行表征,其计算公式为:

$$I_{it} = \begin{cases} I_{i(t-1)} + D_{it} - O_{it}, & t \neq 0 \\ D_{it} - O_{it}, & t = 0 \end{cases} \tag{2}$$

式中: I_{it} 为第 i 个栅格内第 t h 结束后该栅格的共享单车净流入量; D_{it} 为第 i 个栅格内第 $t \sim t + 1$ h 之间共享单车的到达量; O_{it} 为第 i 个栅格内第 $t \sim$

$t + 1$ h 之间共享单车的出发量。

通过计算各栅格每小时的共享单车净流入量,可构成每日 24 h 共享单车净流入量时间序列,以净流入量随时间的变化表征共享单车的时间集聚变化。通过时间序列聚类方法集聚变化特征相似的时间序列,得到典型时间集聚模式,最终以共享单车典型时间集聚模式作为本文的因变量。

2 模型与方法

由于共享单车的净流入量时间序列曲线的局部拐点出现时间、峰值出现时间和总体增减趋势等时变特征呈现出明显的分类性,不同集群之间存在明显的差异和分离性,因此需提炼典型的共享单车时间集聚变化呈现模式,进而研究建成环境对共享单车时间集聚模式的非线性影响。具体地,在对数据进行预处理之后,统计各栅格内共享单车净流入量时间序列,提炼典型共享单车时间集聚模式,并输入不同模型进行训练,探究各建成环境因素对共享单车时间集聚模式的非线性影响;最终,对模型结果进行解释,明确其作用机制与阈值效应。

2.1 凝聚层次聚类算法

凝聚层次聚类算法是一种自下而上的层次聚类方法,它从每个样本开始,将每个样本看作 1 簇,逐渐将 2 个距离最近的簇合并成为 1 个更大的簇,直到所有样本都被合并。该算法具有可以设置不同的相似度度量、便于可视化与解释、无需设置聚类数量和可以处理不同属性维度之间数据等特点。

相似度度量方法层面,由于欧氏距离易受到时序曲线振幅的影响,难以反映其形状和趋势动态变化的相似性,因此并不适用;同时,由于时间序列的长度相同,而动态时间规整(Dynamic time warping, DTW)方法是一种非对称的相似度度量方法,其只考虑时间序列之间的对齐,而不考虑它们之间的顺序关系,早晚高峰变化趋势相反的时间序列易归为一类,因此也不适用。本文采用布雷柯蒂斯相异度(Bray-curtis dissimilarity, BCD)来衡量样本之间的距离^[15]。布雷柯蒂斯相异度是一种常用于衡量时间序列相似性的指标,其具有能衡量时间序列如何随时间共同变化、可解释性强、计算高效等优势,能够较好地反映共享单车

净流入量时间序列的动态变化趋势相似性。布雷柯蒂斯相异度的取值在 $[0,1]$ 之间,取值越接近1表示两时间序列间的相异程度越高。布雷柯蒂斯相异度计算公式为:

$$b_{(X,Y)} = \frac{\sum_{q=1}^m (X_q - Y_q)}{\sum_{q=1}^m (X_q + Y_q)} \quad (3)$$

式中: $b_{(X,Y)}$ 为样本 X 与样本 Y 之间的布雷柯蒂斯相异度; X_q 为样本 X 中第 q 个数据点; Y_q 为样本 Y 中第 q 个数据点; m 为样本容量,在本文中为每日的小时数,取值为24。

对于簇间距离,为降低极端值的影响,增强模型的鲁棒性,本文使用平均连接聚类规则进行计算,其公式为:

$$d_{(C_l, C_p)} = \frac{\sum_{X \in C_l} \sum_{Y \in C_p} r_{(X,Y)}}{\|C_l\| \|C_p\|} \quad (4)$$

式中: $d_{(C_l, C_p)}$ 为簇 C_l 与簇 C_p 之间的距离; l, p 为簇的编号。

2.2 随机森林分类器

随机森林模型是一种 Bagging 集成模型,其个体学习器为决策树,通过决策树投票的方式得出最终的分类结果,其具体步骤为:

步骤1:从训练集的所有样本中,使用 Bootstrap Sampling 采样方法提取出新的训练集(此采样方法为放回采样),并基于该新训练集训练决策树。

步骤2:从训练集的所有特征中随机选取一部分特征,基于该部分特征寻找决策树的最优划分点,将样本划分到2个子节点中。

步骤3:反复执行步骤2,直到所有的节点都无法继续向下划分为止。

步骤4:检查训练的决策树数量是否足够,若数量足够则转入步骤5;若数量不足则转入步骤1。

步骤5:基于所有的决策树分类器结果进行投票,最终的投票结果即为随机森林分类器输出的结果。

随机森林分类器对于噪声值和异常值具有很高的鲁棒性,可以减少特征之间相关性的影响,提高模型的泛化能力;同时,由于随机森林算法中的决策树都相互独立,因此可以并行运算,使随机森林模型处理高维或大规模数据时更加高效,适用于本文中特征较多的情况。

2.3 模型评价方法

针对随机森林分类器模型,采取常用于分类模型评价的精确率(Precision)、召回率(Recall)和F1分数(F1-score)3个指标对模型分类效果进行度量^[16],并使用 $P-R$ 图对模型性能进行可视化以便于与其他模型进行比较。3个指标的取值区间均为 $[0,1]$,取值越接近1则模型表现越好。模型性能度量指标计算公式为:

$$P = \frac{n_{TP}}{n_{TP} + n_{FP}} \quad (5)$$

$$R = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (6)$$

$$F = \frac{2PR}{P + R} \quad (7)$$

式中: P 为模型精确率; R 为模型召回率; F 为模型F1分数; n_{TP} 为将正类识别为正类的数量; n_{TN} 为将负类识别为负类的数量; n_{FP} 为将负类识别为正类的数量; n_{FN} 为将正类识别为负类的数量。

2.4 SHAP 模型

虽然机器学习模型相较于传统的线性模型在精确度等指标上有所提升,但由于模型内部结构复杂,其在可解释性方面有明显不足,因此,需要使用可解释模型来帮助理解模型的决策过程并对其进行可视化。SHAP 模型是一种基于 Shapley value 的可加性解释模型,是合作博弈论中用于解决贡献问题的一种方法。有别于粒子相关性图(Partial dependence plot, PDP)和个体条件期望(Individual conditional expectation, ICE)等方法,该方法通过计算各特征的 SHAP 值来衡量各个特征对最终输出结果的影响,揭示特征值的异质效应,SHAP 模型能够较好地适用于随机森林等树模型^[17],因此选取 SHAP 模型为本文中随机森林分类器提供一个可解释的方案。各特征的 SHAP 值计算公式为:

$$\phi_k = \sum_{S \subset N \setminus \{k\}} \left\{ \frac{|S|!(M-|S|-1)!}{M!} \cdot [f_x(S \cup \{k\}) - f_x(S)] \right\} \quad (8)$$

式中: ϕ_k 为特征 k 的 SHAP 值; N 为训练集中所有特征的集合; M 为 N 的维度; S 为 N 的一个子集; $|S|$ 为 S 的维度; $N \setminus \{k\}$ 为集合 N 去掉特征 k 后组成的集合; $S \cup \{k\}$ 为集合 S 加上特征 k 后组成的集合; $f_x(S)$ 为只利用特征集合 S 根据随机森林模型中树的结构等计算出的样本的平均值。

SHAP 模型不仅能够表示各建成环境指标对

共享单车时间集聚模式影响的重要程度,还能够清楚地表达各特征指标对于分类结果的正向或负向影响,从而揭示建成环境对共享单车时间集聚特征的非线性影响。

3 实证分析

3.1 研究数据选取与描述性统计

本文以云南省昆明市作为研究案例城市,选取2022年3月14~20日(一周内天气皆为晴天)共享单车订单数据,于2022年3~7月,使用高德开放平台API爬取昆明市各类POI数据、交叉口数据和地铁公交站点数据,使用Open Street Map平台提取路网数据。在剔除异常数据后,保留有效共享单车订单数据96 593条。使用ArcGIS软件将主城区划分为边长500 m栅格,统计所有存在共享单车骑行数据的栅格中的热点区域(7 d内共享单车出发量与到达量之和 >120 次,即所有共享单车出行区域的前31.0%),有效区域数量为546个。本文研究区域如图1所示。

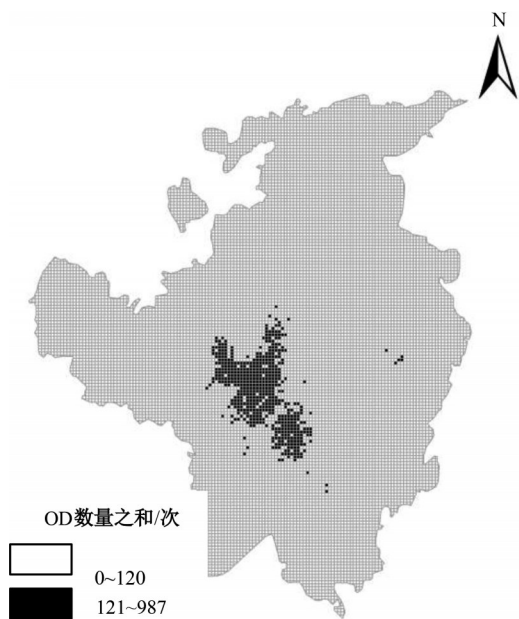


图1 研究区域示意图

Fig. 1 Schematic diagram of research scope

为避免自变量之间高度相关对分类过程造成影响,将自变量与因变量导入SPSS中对自变量进行多重共线性检验,结果显示所有自变量的方差膨胀因子(Variance inflation factor, VIF)值均 <5 ,不存在高度相关的自变量,满足研究对多重共线性的要求。本文构建的自变量指标体系描述性统计见表2。

表2 自变量描述性统计

Table 2 Descriptive statistics of the independent variables

自变量	VIF	均值	标准差
交叉口密度/(个·栅格 $^{-1}$)	2.25	4.24	4.72
公交站点密度/(个·栅格 $^{-1}$)	1.19	1.80	1.66
体育休闲POI密度/(个·栅格 $^{-1}$)	2.21	6.62	8.95
公司企业POI密度/(个·栅格 $^{-1}$)	1.55	28.09	30.62
公园景点POI密度/(个·栅格 $^{-1}$)	1.16	0.95	2.76
居住密度/(个·栅格 $^{-1}$)	2.45	6.98	6.29
购物服务POI密度/(个·栅格 $^{-1}$)	2.28	111.14	150.64
餐饮服务POI密度/(个·栅格 $^{-1}$)	2.55	56.69	52.70
公交站点邻近度/m	2.60	201.81	199.92
公园景点邻近度/m	2.71	393.68	436.48
公交线网密度/(km·栅格 $^{-1}$)	1.32	4.20	3.52
路网密度/(km·栅格 $^{-1}$)	1.49	1 661.04	1 155.09
商圈邻近度/m	1.12	189.53	125.54
地铁站点邻近度/km	1.22	0.85	0.57
土地利用混合度	1.96	0.64	0.12

3.2 时序聚类结果与分析

基于研究热点区域,集计得到各栅格每小时共享单车平均净流入量并构成时间变化序列曲线,使用AHC聚类方法对曲线聚类,聚类树形图如图2所示。

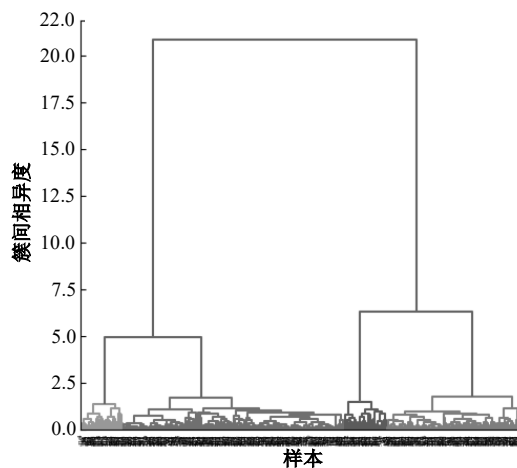


图2 凝聚层次聚类树形图

Fig. 2 Hierarchical clustering tree diagram

由图2可知,当簇数量 >4 时,随着簇的数量每次提升,簇间距离的提升量并不大;当簇的数量为4时,簇间的布雷柯蒂斯相异度差距较大,聚类结果能够同时较好地反映同类时间序列曲线之间的动态相似性与不同类时间序列曲线之间的差异性,因此选择4类作为本文时间序列曲线聚类数量。时序曲线聚类结果如图3所示。

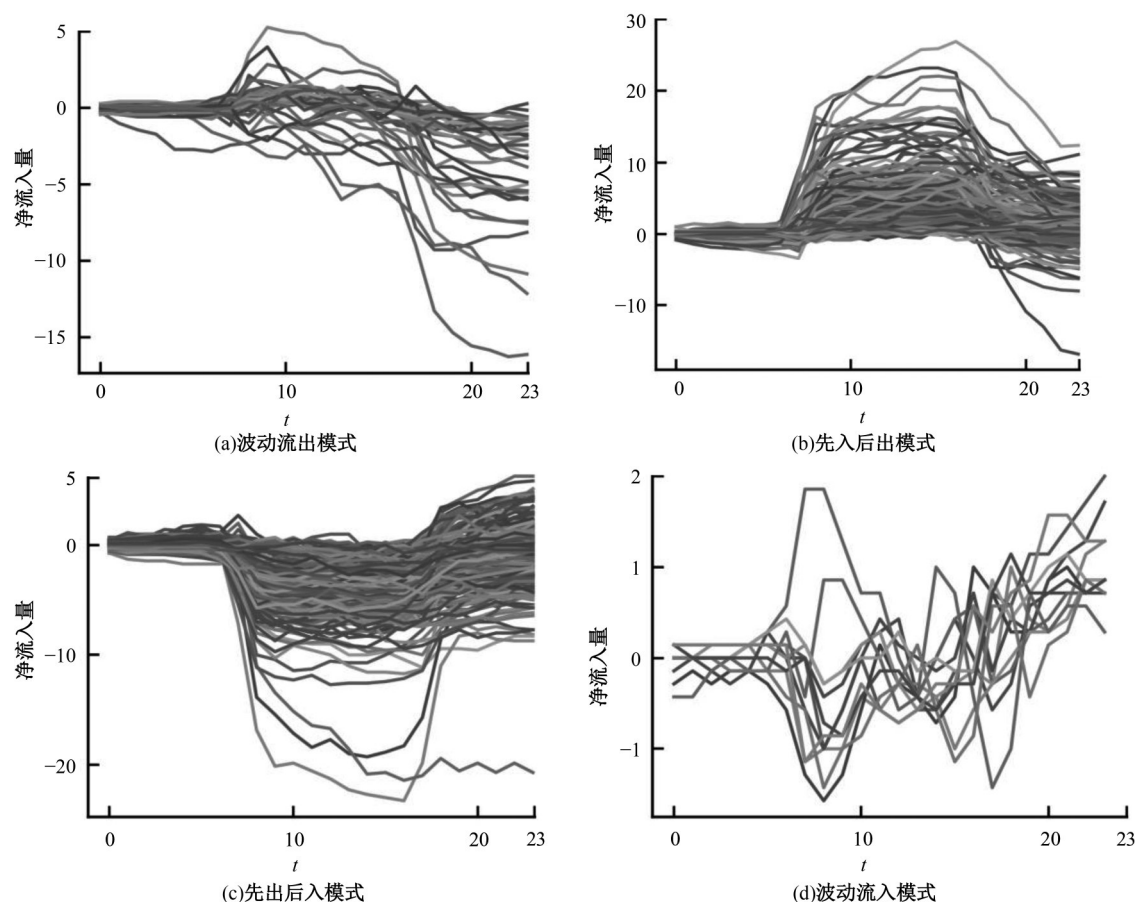


图 3 时间序列曲线聚类结果

Fig. 3 Time series curve clustering results

图 3 中纵坐标为共享单车净流入量,横坐标为时间。为了便于描述,根据各类时间序列曲线时变特征呈现模式将其记为波动流出模式、先入后出模式、先出后入模式和波动流入模式,各模式曲线的具体特征如下:

(1)波动流出模式:时序曲线总数为 35 条,时变特征表现为 0 至早高峰开始时净流入量稳定在 0 附近,从早高峰开始时至一天结束呈现出波动特征,增减变化复杂,总体净流入量呈现下降趋势。

(2)先入后出模式:时序曲线总数为 203 条,其时变特征较为简明,表现为 0 至早高峰开始时净流入量稳定在 0 附近,早高峰开始后逐渐攀升,早高峰结束后恢复平稳状态至晚高峰开始,从晚高峰开始至一天结束净流入量逐渐下降。

(3)先出后入模式:时序曲线总数为 296 条,其时变特征与先入后出模式曲线相反,表现为 0 至早高峰开始时净流入量稳定在 0 附近,早高峰开始后逐渐下降,早高峰结束后恢复平稳状态至晚高峰开始,从晚高峰开始至一天结束净流入量

逐渐攀升。

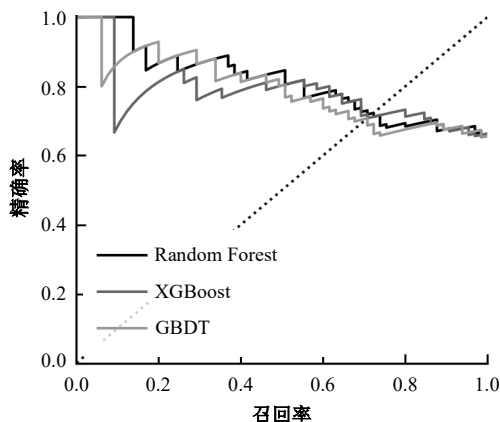
(4)波动流入模式:时序曲线总数为 12 条,其时变特征表现为 0 至早高峰开始时净流入量稳定在 0 附近,从早高峰开始时至一天结束呈现出波动特征,增减变化复杂,总体净流入量呈现增长趋势。

先入后出模式与先出后入模式曲线为共享单车时间序列曲线的主要类型,其数量占总数的 91.4%。

3.3 分类结果

由于波动流入模式与波动流出模式时间序列数量过少,为更精确地表达建成环境对共享单车时间集聚模式的非线性影响,选择先入后出模式与先出后入模式 2 类典型时序曲线模式作为因变量输入分类器。由于 2 种模式时序曲线数量并不均衡,为了使分类结果可信度更高,因此在划分训练集与测试集时使各自 2 种模式曲线数量比例与原数据集中 2 种模式时间序列曲线数量比例保持相同,同时按照 8:2 的比例划分训练集与测试集。将数据分别输入随机森林分类器、极度梯度提升

树(Extreme gradient boosting, XGBoost)分类器与梯度提升决策树(Gradient boosting decision tree, GBDT)分类器中,使用随机搜索法调校模型的超参数,使用最优超参数对数据进行拟合并进行5折交叉检验,以模型的精确率为纵轴,召回率为横轴,绘制得到3种模型的 $P-R$ 图如图4所示。

图4 分类器 $P-R$ 曲线图Fig. 4 Classifier $P-R$ curve

由图4可知,3种分类器并未有任何一种 $P-R$ 曲线完全位于其他某种之下的情况,因此比较平衡点(即图4中虚线与分类器交点)的位置,随机森林曲线的平衡点略高于GBDT与XGBoost曲线,因此,随机森林分类器略优于GBDT与XGBoost分类器。随机森林分类器的模型精确率 P 为65.34%,召回率 R 为80.26%, $F1$ 分数 F 为71.52%。

3.4 SHAP模型结果与非线性影响分析

将随机森林分类器模型与自变量数据输入SHAP模型中,得到在分类过程中各自变量的重要程度排序,绘制重要程度排序图如图5所示。

由图5可知,共享单车时间集聚模式主要受公交线网密度、公园景点邻近度和地铁站点邻近度等因素的影响,其中公交线网密度对分类结果的影响重要度最大,而公园景点POI密度、公交站点密度和交叉口密度等因素对共享单车时间集聚模式影响相对较小。

通过SHAP归因分析得到自变量对共享单车时间集聚模式分类结果的影响范围和影响方向,SHAP值为负表示具有负向贡献,SHAP值为正表示具有正向贡献。以先入后出模式为例绘制自变量正负向影响蜂群图,如图6所示,先出后入

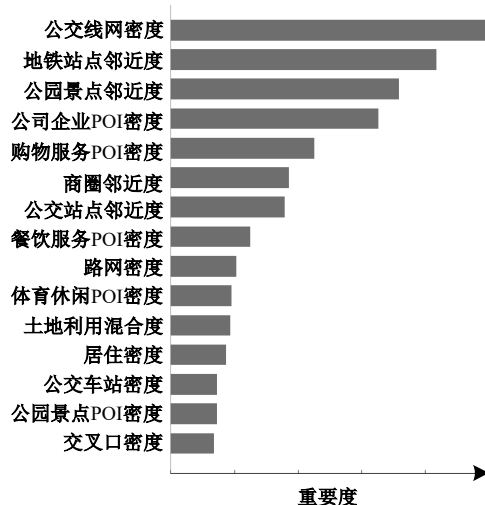


图5 自变量重要度排序

Fig. 5 Importance ranking of independent variables

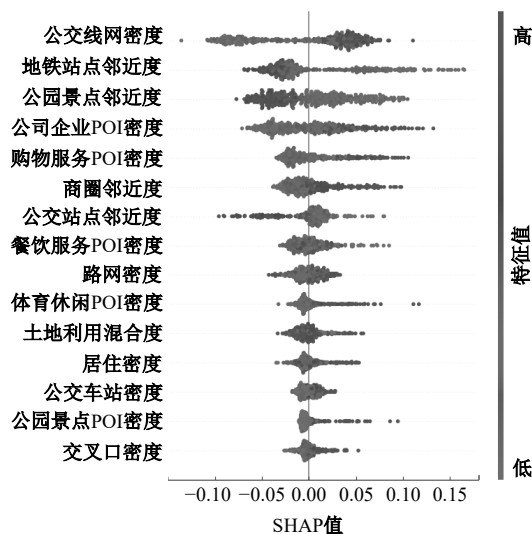


图6 自变量正负向影响

Fig. 6 Positive and negative effects of independent variables

模式与先入后出模式的结果相反。

由图6可知:①公交线网密度、公司企业POI密度和商圈邻近度等指标对共享单车时间集聚呈现为先入后出模式具有正向贡献。公交线网密度、公司企业POI密度越高,栅格中心离商圈越远,则该地区共享单车时间集聚呈现为先入后出模式的概率越高;②公园景点邻近度、地铁站点邻近度与公交站点邻近度等指标对共享单车时间集聚表现为先入后出模式具有负向贡献,即栅格中心离地铁站、公交站或公园景点越远,则该栅格内共享单车时间集聚越可能表现为与先入后出模式相反的模式,即先出后入模式。

选取自变量中关键影响因素,使用SHAP模

型对其进行偏依赖分析,探究各关键因素对共享单车时间聚集模式的非线性影响机制。以各关键

因素为横坐标,其 SHAP 值为纵坐标,绘制关键因素部分依赖图如图 7 所示。

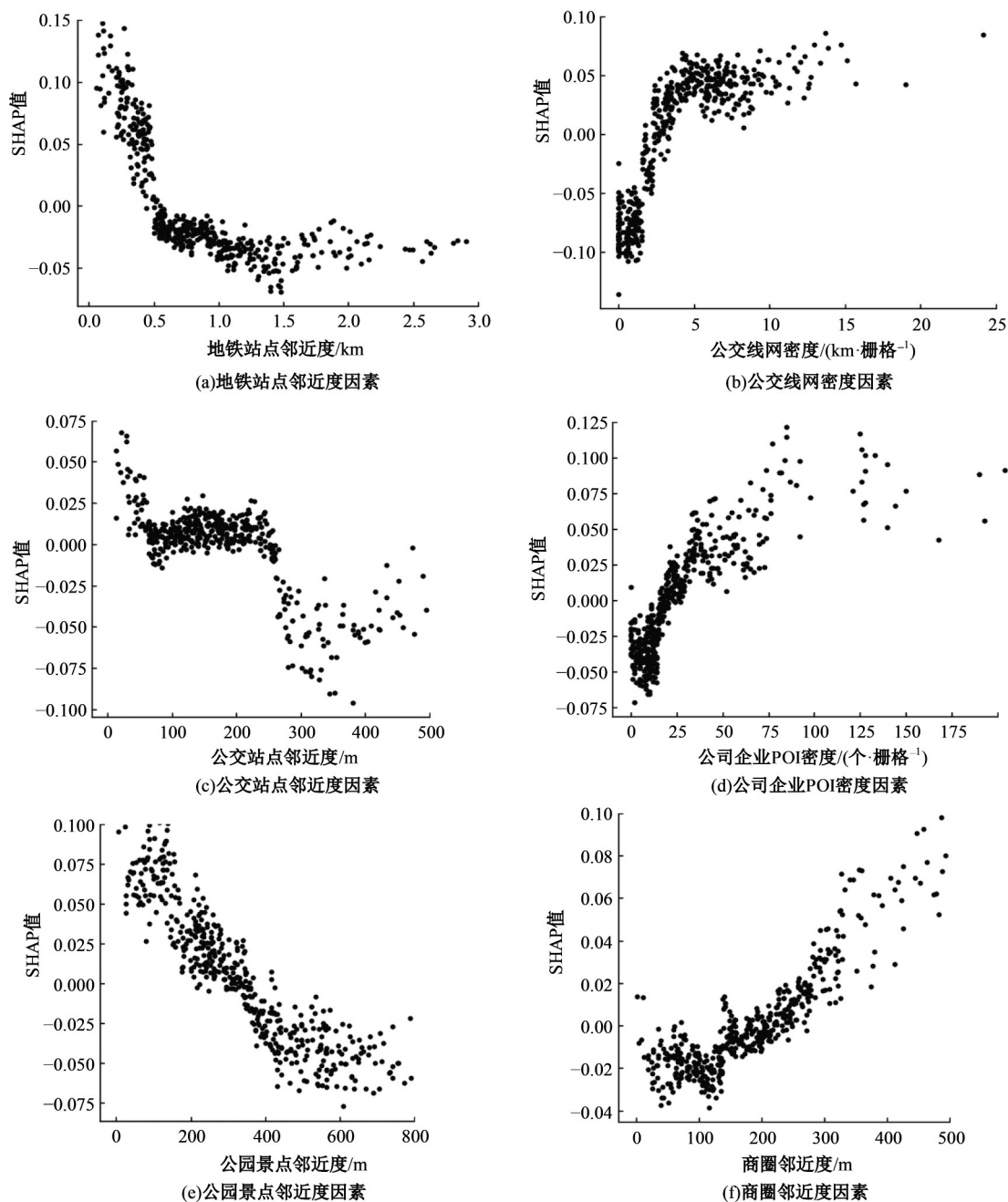


图 7 关键因素部分依赖图

Fig. 7 Key factors partial dependency

由图 7 可知,城市建成环境关键因素对共享单车时间集聚模式的非线性影响如下:

(1) 地铁站点邻近度在 $[0, 0.5]$ km 区间内时,对共享单车时间集聚呈现为先入后出模式概率具有抬升作用;且在 $[0, 0.6]$ km 区间内时,地铁站离栅格中心距离越近,呈现为先入后出模式的概率越大;超出 0.6 km 范围地铁站邻近度对共享单车时间集聚模式的影响程度基本保持不变。

地铁站附近早晚高峰客流以通勤客流为主,早高峰时通勤群体从居住地乘共享单车,接驳地铁前往目的地,而晚高峰则乘地铁,接驳共享单车前往居住地。当邻近度超过 0.6 km 时,距离地铁站已经较远,地铁站对共享单车聚集产生的影响较小。

(2) 公交线网密度在 $[0, 2.7]$ km/栅格区间内时,对共享单车时间集聚呈现为先入后出模式概率具有压降作用。公交线网密度在 $[0, 5]$ km/

栅格区间时,公交线网密度越大,共享单车时间集聚呈现为先入后出模式的概率越大,且在密度超过5 km/栅格时,公交线网密度对共享单车时间集聚模式的影响程度基本不再变化。公交线网密度越大表明区域内通过的公交线路越多,因此该区域选择“共享单车-公交”接驳出行的通勤者越多,越会致使早高峰共享单车流入而晚高峰时期流出,净流入量在其余时间段基本保持平稳。

(3)公交站点邻近度对共享单车时间集聚模式表现出非常显著的非线性影响。当公交站点邻近度取值在 $[0, 80]$ m区间内时,对共享单车时间集聚呈现为先入后出模式概率具有抬升作用。在 $[80, 250]$ m区间内时,对共享单车时间集聚模式的影响程度较小,且随着邻近度的增大影响程度不发生变化。公交站点邻近度超过250 m时,公交站点对共享单车时间集聚模式的影响不足以作用到该区域,因此该区域共享单车时间集聚呈现为先入后入模式概率较大。

(4)公司企业POI密度与先入后出集聚模式之间具有显著正相关关系,在 $[0, 25]$ 个/栅格区间内时,公司企业POI密度对共享单车时间集聚表现为先入后出模式具有压降作用,随着公司企业POI密度的升高,区域共享单车时间集聚越倾向于表现为先入后出模式。公司企业越密集,早高峰时通勤人群越多,早高峰共享单车流入量越大,晚高峰共享单车流出量越大,使得该区域共享单车时间集聚表现为先入后出模式。

(5)公园景点邻近度取值在 $[0, 350]$ m区间内时,对共享单车时间集聚呈现为先入后出模式概率具有抬升作用。这是由于公园景点区域基本不具备居住属性,且多数景点及部分公园夜间存在关闭时间限制,旅游群体需离开景点区域。邻近度超过450 m时,公园景点对共享单车时间集聚模式的影响不足以作用到该区域。

(6)商圈邻近度在 $[0, 200]$ m区间内,对共享单车时间集聚模式表现为先入后出模式具有压降作用,邻近度值越大,共享单车时间集聚表现为先入后出模式的可能性越小;超过200 m时,对共享单车时间集聚模式表现为先入后出模式具有抬升作用,邻近度值越大,共享单车时间集聚表现为先入后出模式的可能性越大。

4 结 论

(1)共享单车时间集聚主要呈现为先入后出模式与先出后入模式两大模式,其数量之和占所有模式的91.4%。

(2)相较于XGBoost分类器与GBDT分类器,随机森林分类器对于共享单车时间序列曲线具有较好的分类效果。模型精确率 P 为65.34%,召回率 R 为80.26%,F1分数 F 为71.52%。

(3)公交线网密度、公园景点邻近度、地铁站点邻近度和公司企业POI密度对于共享单车时间集聚模式影响较大,而公园景点POI密度、体育休闲POI密度和交叉口密度等因素对共享单车时间集聚模式影响相对较小。

(4)建成环境与共享单车时间集聚模式存在显著的非线性关系。当地铁站点邻近度在 $[0, 0.5]$ km,公交站点邻近度取值在 $[0, 80]$ m,公园景点邻近度取值在 $[0, 350]$ m区间内时,对共享单车时间集聚表现为先入后出模式具有抬升作用;公交线网密度在 $[0, 2.7]$ km/栅格,公司企业密度在 $[0, 25]$ 个/栅格,商圈邻近度在 $[0, 200]$ m区间内时,对共享单车时间集聚表现为先入后出模式具有压降作用。

参考文献:

- [1] Shaheen S, Guzman S. Worldwide bikesharing[J]. ACCESS Magazine, 2011, 39(1): 22-27.
- [2] 邝嘉恒, 邱群勇. 接驳地铁站的共享单车时空均衡性分析与吸引区域优化[J]. 地球信息科学学报, 2022, 24(7): 1337-1348.
Kuang Jia-heng, Wu Qun-yong. Spatial-temporal equilibrium analysis and attraction area optimization of dockless sharing bicycles connected to subway stations [J]. Journal of Geo-information Science, 2022, 24(7): 1337-1348.
- [3] Faghih Imani A, Eluru N, El-Geneidy A M, et al. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system(BIXI) in montreal[J]. Journal of Transport Geography, 2014, 41: 306-314.
- [4] 黄颀昊, 杨新苗, 岳锦涛. 基于多尺度地理加权回归模型的城市道路骑行流量分析[J]. 清华大学学报: 自然科学版, 2022, 62(7): 1132-1141.
Huang Yong-hao, Yang Xin-miao, Yue Jin-tao. Urban street bicycle flow analysis based on multi-scale

- geographically weighted regression model[J]. Journal of Tsinghua University(Science and Technology), 2022, 62(7): 1132-1141.
- [5] 马新卫, 季彦婕, 金雨川, 等. 基于时空地理加权回归的共享单车需求影响因素分析[J]. 吉林大学学报: 工学版, 2020, 50(4): 1344-1354.
Ma Xin-wei, Ji Yan-jie, Jin Yu-chuan, et al. Geographically and temporally weighted regression for modeling spatio-temporal variation in dockless bike-share usage demand[J]. Journal of Jilin University (Engineering and Technology Edition), 2020, 50(4): 1344-1354.
- [6] Ji Y, Ma X, Yang M, et al. Exploring spatially varying influences on metro-bikeshare transfer: A geographically weighted poisson regression approach[J]. Sustainability, 2018, 10(5): No. 1526.
- [7] Wang K, Chen Y J. Joint Analysis of the impacts of built environment on bikeshare station capacity and trip attractions[J]. Journal of Transport Geography, 2020, 82: No. 102603.
- [8] Shen Y, Zhang X, Zhao J. Understanding the usage of dockless bike sharing in Singapore[J]. International Journal of Sustainable Transportation, 2018, 12(9): 686-700.
- [9] Ding C, Chen P, Jiao J. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: a machine learning approach[J]. Accident Analysis & Prevention, 2018, 112: 116-126.
- [10] 高楹, 宋辞, 郭思慧, 等. 接驳地铁站的共享单车源汇时空特征及其影响因素[J]. 地球信息科学学报, 2021, 23(1): 155-170.
Gao Ying, Song Ci, Guo Si-hui, et al. Spatial-temporal characteristics and influencing factors of source and sink of dockless sharing bicycles connected to subway stations[J]. Journal of Geo-information Science, 2021, 23(1): 155-170.
- [11] Cao M, Cai B, Ma S, et al. Analysis of the cycling flow between origin and destination for dockless shared bicycles based on singular value decomposition [J]. ISPRS International Journal of Geo-Information, 2019, 8(12): 573.
- [12] 李武, 戢晓峰, 陈方, 等. 建成环境对常规公交空间公平性的非线性影响[J]. 公路交通科技, 2023, 40(8): 207-213.
Li Wu, Ji Xiao-feng, Chen Fang, et al. Nonlinear influence of built environment on spatial fairness of conventional transit[J]. Journal of Highway and Transportation Research and Development, 2023, 40(8): 207-213.
- [13] Ewing R, Cervero R. Travel and the built environment: A synthesis[J]. Transportation Research Record, 2001, 1780(1): 87-114.
- [14] 王建军, 王赛, 宋明洋, 等. 大数据背景下城市建成环境对出行行为影响研究综述[J]. 长安大学学报: 自然科学版, 2022, 42(1): 61-78.
Wang Jian-jun, Wang Sai, Song Ming-yang, et al. Review of research on impact of urban built environment on travel behavior in big data context[J]. Journal of Chang'an University(Natural Science Edition), 2022, 42(1): 61-78.
- [15] Rodionov I V, Sozontov A N. 2020. On confidence estimation based on quantitative similarity coefficients [J]. Automation and Remote Control, 81(2): 320-332.
- [16] 李冬梅, 杨宇, 孟湘皓, 等. 多标签分类综述[J]. 计算机科学与探索, 2023, 17(11): 2529-2542.
Li Dong-mei, Yang Yu, Meng Xiang-hao, et al. Review on multi-label classification[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(11): 2529-2542.
- [17] Aas K, Jullum M, Loland A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values[J]. Artificial Intelligence, 2021, 298: No. 103502.