

· 数字经济 ·

基于天气因素的共享单车骑行量预测



□ 晏鹏宇^{1,2} 张 华¹ 王 雪³ 黎 鹏¹ 刘雨轩⁴ 杨 东⁴

[1. 电子科技大学 成都 611731; 2. 四川省国际科技合作 (以色列) 研究院 成都 611731;
3. 深圳职业技术学院 深圳 518055; 4. 成都市气象局 成都 610072]

【摘 要】 【目的/意义】 气温、降雨量和风力等天气因素对共享单车骑行量构成了非线性影响, 而传统的多元线性回归模型以及基于时序的预测模型, 由于其对非线性关系表达能力不足, 预测结果往往存在较大误差。【设计/方法】 在单车价格等经济因素以及市场竞争态势等相对稳定情况下, 结合用户潮汐骑行规律和地理空间兴趣点数据构建了基于BP神经网络的天气因素对共享单车骑行量的预测模型, 并利用成都市某区域内摩拜单车骑行与天气数据, 分别训练了“逐日”和“逐时”为单位的BP预测模型。【结论/发现】 对比实验结果表明, 基于BP神经网络的预测模型具有更高的预测准确度, 实验结果还表明不同兴趣点在不同时刻的骑行量受温度和降雨量的影响程度差异较大。研究结果为共享单车公司实现基于天气因素的单车骑行量预测提供了科学方法, 并且为单车数量的精准投放与回收提供了依据。

【关键词】 共享单车; 骑行量; BP神经网络; 天气因素

【中图分类号】 U491.1

【文献标识码】 A

【DOI】 10.14071/j.1008-8105(2021)-3014

Predicting the Riding Volume of Shared Bikes Based on Weather Factors

YAN Peng-yu^{1,2} ZHANG Hua¹ WANG Xue³ LI Peng¹ LIU Yu-xuan⁴ YANG Dong⁴

(1. University of Electronic Science and Technology of China Chengdu 611731 China;

2. Sichuan Institute of International Science and Technology Collaboration (Israel) Chengdu 611731 China;

3. Shenzhen Polytechnic Shenzhen 518055 China;

4. Chengdu Meteorological Office Chengdu 610072 China)

Abstract [Purpose/Significance] Weather factors such as temperature, rainfall and wind constitute the nonlinear influence on the riding volume of shared bikes, but the traditional multiple linear regression model and the prediction model based on time sequence often have large errors due to their poor expression ability of the nonlinear relationship. [Design/Methodology] In this paper, the economic factors such as price and the marketing competition situations are supposed relatively stable. A BP neural network model combined with the user tidal riding rule and geographical spatial Point of Interest data to predict the riding volume of shared bikes affected by weather factors. The two prediction models based on per “day” and per “hour” were respectively trained by the real riding data and weather data in Chengdu city, China. [Findings/Conclusions] The comparative experimental results indicate that the BP neural network model has higher prediction accuracy. The results also show that the riding volume of shared bikes in different POIs at different time is greatly affected by temperature and rainfall. This study provides a scientific approach to estimate the riding volume and also a basis for the precise delivery and collection of bikes for shared bike enterprises.

Key words shared bike; riding volume; BP neural network; weather factor

[收稿日期] 2021-05-08

[基金项目] 国家自然科学基金项目 (71971044); 国家自然科学基金重大项目 (72091213); 四川省科技计划应用基础研究项目 (2020YJ0026)。

[作者简介] 晏鹏宇 (1982-) 男, 博士, 电子科技大学经济与管理学院教授; 四川省国际科技合作 (以色列) 研究院研究员; 张华 (1997-) 女, 电子科技大学经济与管理学院硕士研究生; 王雪 (1968-) 女, 深圳职业技术学院管理学院副教授。

引言

为了丰富大中城市市民短途出行方式,全球各大城市推出了有桩和无桩共享单车租赁服务。世界上最早的共享单车概念起源于1965年荷兰的阿姆斯特丹^[1]。中国共享单车服务始于2014年,目前已经覆盖60多个大中城市。截至2016年底总计约43万辆公共自行车,全球排第一位,是第二名法国的9.5倍^[2]。共享单车有效解决了人们短途出行需求,并节约了出行成本。此外,从城市交通系统角度,其还有效提升了城市道路的利用率,丰富了城市交通工具结构,对构建绿色低碳交通体系起到了积极作用。

我国早期多个共享单车平台为快速占领市场份额,进行了多轮疯狂“圈地运动”,使得许多大中城市的共享单车投放规模远超于实际需求^[3]。为鼓励和规范共享单车的可持续性健康发展,进一步提升共享单车平台的运维管理水平,我国中央和地方政府部门出台了多项规范共享单车发展和车辆动态投放与回收的相关制度与措施^[4-5]。与此同时,积极探索利用卫星导航定位、地理信息系统和蓝牙通信等技术,建立共享单车政府监管与服务平台,规范共享自行车的管理。例如,成都市成华区城管局建立了第三方共享单车智能管理平台,通过蓝牙技术实时监控该区域内地铁站口、高校、商圈和景点等地点的单车使用和空闲堆积情况,并要求共享单车平台对区域内的单车实现实时动态投放和回收,提升车辆的有效利用率,避免多余车辆堆积。同时,还将每日单车闲置率和废弃率等作为区域内共享单车平台考核评价的指标和下一阶段单车投放总量的重要参考。然而,共享单车平台在实际运营中难以准确预测用户骑行量,时常造成单车投放与需求在时空维度上的不平衡^[6],出现人流多的地点无车可借,人流量少的地点车辆堆积,堵塞人行道和机动车道,甚至影响市民的正常出行,给城市交通秩序造成不良影响。

共享单车的需求不仅受到通勤用户潮汐出行规律等因素影响^[7],还受到气温、降雨量和风力等天气因素以及地理空间等多维复杂因素的影响。共享单车作为一种暴露在交通环境中的非机动出行方式,用户骑行需求受天气等自然因素的影响程度更为显著。天气状态不仅会改变出行者的出行环境,还会影响出行者的生理和心理状态,进而影响交通的有序运行。纵观目前大多研究内容主要集中于机动车交通工具,对于共享单车此类新型交通工具,特别是天气因素对共享单车骑行量的影响,研究尚

存在不足。研究天气因素是否影响、如何影响共享单车作为交通工具出行的选择以及各天气因素之间的相互关系,有利于指导单车企业与相关管理部门更加科学合理地投放和管理。

上述现实运营压力和政府管制要求,迫使共享单车平台的运营管理方式必须从粗放型向智能和精细化方式转变。平台需要根据实时天气数据、用户骑行规律以及时空特征因素,准确预测用户骑行(需求)量,为日常运营中车辆投放与回收决策提供科学依据。本文在当前单车骑行价格等经济因素以及市场竞争态势等相对稳定情况下,构建了基于BP神经网络的天气因素对共享单车骑行量的预测模型,利用2018年8~10月成都市成华区电子科技大学(沙河校区)及其周边的摩拜单车骑行数据和成都市气象局提供的该区域实时天气数据,并结合用户的潮汐骑行规律和地理空间的兴趣点(Point of Interest, POI)数据^[8],分别训练“每日”和“每时”的用户骑行量预测模型,为管理决策者提供每天单车总投放数量决策和一天内单车在兴趣点之间的动态回收和调度决策。本文进一步利用该预测模型,分析了不同天气条件下共享单车骑行量的变化特征,为下一步共享单车平台或者第三方监管平台实现基于天气因素的实时精准投放与回收提供了科学方法与依据。

一、文献综述

目前国内外针对共享单车在社会影响、经济效益、法律条例以及运营管理等各方面均有较深入的研究^[9-18],部分研究集中在某区域内兴趣地点和时段对骑行需求量的影响。例如,文献^[19]从空间和时间角度分析了共享单车出行需求存在时空不平衡现象,并利用长短期记忆神经网络方法(LSTM NNs)建立共享单车动态需求预测模型,预测了不同时间间隔(10 min、15 min、20 min和30 min)的共享单车出行量和吸引力。文献^[20]提出基于数据驱动图滤波器(GCNN-DDGF)的新型图卷积神经网络模型,该模型能够学习站点间隐藏的异构配对关联,从而预测大型共享单车网络中站点级每小时的需求。文献^[21]将人工智能系统与人工神经网络预测技术相结合,开发了一种用于数值预测的新型预测框架(AIS-ANN)进行单车需求预测。在提出的AIS-ANN预测框架中,建立预测系统的机制主要有三种:人工神经网络、克隆选择和相似性度量。文献^[22]采用支持向量机构建了基于SVM的共享单车需求预测模型。仿真实验的复杂度方面利用主成分

分析法进行了降低,对于数据噪声的扰动,采用小波消噪的方式进行了减弱和消除。而文献[23~24]则针对时间出行分布的非严格周期性,分别采用引入注意力机制的AM-LSTM和以随机森林作基准的GRU-LSTM双重方法结合的模型进行短期需求预测。文献[25]基于Tanh, Logistic, Relu和Identity四种不同激活函数构建了BP神经网络需求预测模型,并利用模型评估指标筛选了最优单车需求预测模型。文献[26]针对共享单车动态的时空特性,对系统状态的短期时空分布预测进一步扩大。并采用卷积长短期记忆网络(conv-LSTM)方法,解决了空间依赖和时间依赖。通过以上文献回顾可发现,大部分文献从区域人口和经济总量等社会经济宏观因素评估共享单车的投放量,但未从日常运作的微观角度考虑天气条件对单车骑行量的影响。在具体预测方法方面,通过BP神经网络模型同时结合天气因素来研究骑行需求的文献尚较缺乏,且大多主要利用传统的多元线性回归模型以及基于时序的预测模型,由于这类模型自身对非线性关系表达能力较差,预测结果往往存在较大误差。

目前仅少部分文献考虑了温度、降雨量等自然因素对单车骑行量的影响。文献[27]基于历史数据、天气数据和时间数据,采用两层深度的LSTM模型实时预测未来一段时间内城市不同区域的单车租赁和归还情况。文献[28]则综合考虑了时间、天气、建筑环境、基础设施和空间等因素,开发了广义极值(GEV)计数模型,预测各个站点每小时的单车开关锁数量。文献[29]则借助天气数据(温度、湿度、风速、能见度、太阳辐射、降雪、降雨),每小时租用单车的数量和日期信息,运用数据挖掘技术建立了每小时单车需求预测模型,提出了一种特征滤波方法以消除不可预测的参数。文献[30]基于套索回归、随机森林和迭代决策树等机器学习方法研究了共享单车短期(基于小时)内需求预测的主要影响因素,研究结果发现,影响共享单车短期需求的主要因素包括:特定的位置因素,时间因素以及天气条件因素。文献[31]则采用长短期记忆(LSTM)神经网络模型预测了共享单车的短期需求量,并与传统的循环神经网络(RNN)和BP神经网络模型预测结果进行了对比分析,验证了LSTM鲁棒性更高,泛化能力更强,且预测结果曲线与真实结果曲线相吻合。此外,预测实验结果表明:影响单车需求量的主要因素包括温度、节假日、季节和早晚高峰时间段等。

本文将利用摩拜单车平台提供的成都市电子科

技大学(沙河校区)及其周边的历史骑行数据和成都市气象站提供的天气数据,分析影响单车骑行量的天气因素及其相关性的强弱,并结合主成分分析法选取其中主要的影响因素作为输入变量,在此基础上构建基于BP神经网络的共享单车骑行量预测模型,并与传统的多元线性回归模型以及基于时序的预测模型进行对比分析。最后基于真实的数据集验证所建预测模型的准确性、有效性和可靠性。以此帮助共享单车企业更加智能和精细地预测用户需求的特征与规律,为共享单车的投放和调运提供更加科学的解决方案,从而提升企业的运营管理水平。

需要注意的是,本文从日常微观运作角度出发,研究关于电子科技大学(沙河校区)及其周边区域内,摩拜单车平台的骑行量在“每日”和“每时”受天气因素的影响。在研究问题中,假设对影响市场中整体骑行量的经济因素,如共享单车骑行的价格和成本,以及该区域内其他共享单车平台的竞争策略与行为等均保持不变或者相对稳定。其次,由于本文研究区域——电子科技大学(沙河校区)及其周边的共享单车总体呈现供大于求的情况,并且在城市管理部门的实时监管下该平台能够及时地进行单车调度,总体上实现了单车供需的基本匹配。因此,本文采用摩拜单车平台记录的骑行量来反映需求量具有一定的可靠性和准确性保证。

二、数据描述与相关性分析

本文从摩拜单车公司(现为美团共享单车)获取了成都市成华区电子科技大学(沙河校区)及其周边2018年8月1日~10月31日的骑行数据,具体信息包括:骑行日期、单车ID、单车开锁和闭锁的时间(精确到秒)与开关锁的经纬度,共25.4万条记录;以及成都市气象局提供的该区域气象观测站同期以小时为单位的气温和降雨量数据,共2929条记录。其中部分共享单车骑行数据的原始记录如表1所示。

(一)影响因素分析

因缺少底护装置和控温设备,共享单车与地铁、公交车、汽车等其他交通工具相比,其用户的骑行意愿受温度、降雨量等外部天气因素影响较大。此外,天气因素还包括:风力、降雪量、空气质量等。由于成都地区常年风力较弱,基本无降雪,空气质量良好,故本文以气温、降雨量以及降雨时长作为主要指标进行考量。每小时整点时刻的温度记录为时刻温度,精确到0.1℃,每小时的总

表 1 部分单车骑行数据原始记录

日期	bikeid	开锁时间	关锁时间	开锁经度	开锁纬度	关锁经度	关锁纬度
20181018	0286660990	18:15:10	18:21:26	104.03453	30.72214	104.03211	30.72559
20181018	0286660990	08:33:32	08:38:45	104.03115	30.72972	104.03213	30.72437
20181018	0286660990	08:39:21	08:44:31	104.03213	30.72437	104.03306	30.71884
20180831	0286560539	17:44:56	17:47:52	104.05316	30.68159	104.05263	30.68481
20180831	0286560539	19:32:58	19:39:08	104.05279	30.68476	104.04766	30.68233
20181018	0286560539	16:01:55	16:07:53	104.05168	30.70209	104.05794	30.69477
20181011	0286560539	09:37:28	09:42:21	104.04162	30.69643	104.04229	30.69254
20181018	0286560539	09:18:35	09:22:19	104.04988	30.70552	104.05182	30.70192
20180919	0286560539	12:18:12	12:22:12	104.07219	30.65385	104.07115	30.65258
20180831	0286560539	00:05:46	00:12:04	104.04563	30.66294	104.04854	30.66776

降雨量记录为时刻降雨量，精确到0.1 mm，并以小时为单位统计了各个地点在每个时段内的骑行量。同时，为了避免不同地点和潮汐现象等微观因素对预测效果的影响，本文在所研究区域内选取了8个具有代表性的兴趣地点，主要包括有地铁口、公交站口、高校和商场入口等。

（二）相关性分析和主成分分析

基于所获取的单车骑行数据和气象数据，本文首先以“日”为单位进行影响因素的相关性分析。从表2可以看出，日均温度和日下雨时长之间的相关性为-0.125，呈弱负相关关系；日均温度和日总雨量之间的相关性为0.122，呈弱正相关关系；而日总雨量与日下雨时长为强相关关系，其相关性为0.709。其中，日均温度、日总雨量和日下雨时长与日骑行量均呈强负相关关系。以上数据说明研究天气因素对日骑行量的影响具有一定的必要性，因日总雨量与日下雨时长关于骑行量的解释具有较大的信息重叠性，故考虑通过主成分分析法进一步筛选变量。

表 2 天气指标相关性

	日均温度	日下雨时长	日总雨量	日骑行量
日均温度	1	-0.125	0.122	-0.367**
日下雨时长	-0.125	1	0.709**	-0.486**
日总雨量	0.122	0.709**	1	-0.454**
日骑行量	-0.367**	-0.486**	-0.454**	1

（**代表相关性显著）

将以上三种因子解释骑行量的占比由高到低进行排序，观察各因子累计解释占比。统计结果如表3所示：

表 3 总方差解释

成分	总计	方差百分比	累计
日均温度	1.709	56.964	56.964
日总雨量	1.041	34.685	91.649
日下雨时长	0.251	8.351	100

从表3可以看到，日均温度、日总雨量和日下雨时长的解释占比分别为56.964%、34.685%和8.351%。日均温度和日总雨量两个因子的累计解释

占比为91.649%，这表明绝大部分的信息可以被日均温度和日下雨总量所表达，即这两个因子基本可以解释日骑行量，故在以“日”为单位的BP神经网络预测模型中，日总雨量和日下雨时长选择其中之一作为输入变量即可，本文选取日总雨量进行进一步的研究。

三、改进的BP神经网络预测模型

以往使用的ARIMA等模型均为基于时序的预测应用，随着时间推移进行顺序处理，不仅会耗时长，而且对计算资源要求极高。而BP神经网络模型无以上基于时序的预测模型弊端，且其理论发展已较为成熟，作为一种前向传输网络，其还具有强大的并行处理能力、非线性映射能力、自适应性和泛化能力等优点。

（一）输入数据处理

1. “日”和“时”为单位的数据处理

通过统计分析发现在每天00:00~5:59时段内的单车骑行量不足总数据的1%，故本文仅统计每天06:00~23:59这个时段共计18个小时的骑行数据。为排除其他干扰因素，统计数据中剔除周末和节假日的单车骑行数据。在以“日”为单位的统计中，日下雨时长在[0, 18]小时范围内取整数，经统计得到以“日”为单位的“骑行-天气”数据共62个。在以“时”为单位的统计中，由于统计骑行时间段为06:00~23:59，将观察到的“时刻”以小时为单位依次编号：6，7，…，23，经过统计整理共得到8928个“骑行-天气”数据。

2. 兴趣地点的数据处理

本文在所研究的电子科技大学（沙河校区）及其周边区域内选取了8个具有代表性的兴趣地点，其中包括有：前锋路地铁站、电子科大食堂、图书馆等，将其依次编号为：1，2，…，8，并根据历史骑行数据，绘制了如图1所示的8个兴趣地点的热力图。这些兴趣点的日平均开锁次数在10次以上，代表了该区域内用户密集借还单车的位置，这也是单车运营管理者对单车回收和投放的重要地点。

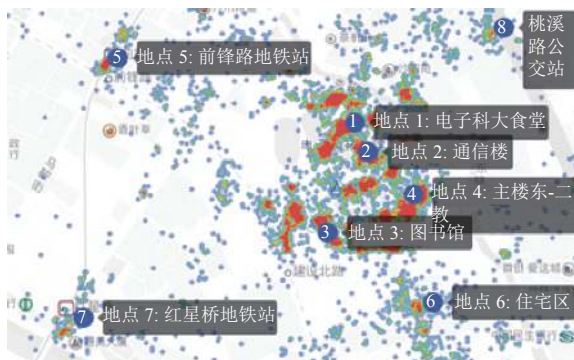


图1 兴趣地点1-8骑行热力图

(二) 模型构建

1. 输入层和输出层设计

由于输入变量会直接影响到网络的输出结果,故输入变量和节点数的选取特别重要,根据上述相关性和主成分分析,初步确定日均温度和日总雨量作为输入变量。本文研究内容为天气因素对骑行量的影响,故输出变量确定为“日骑行量”或“时刻骑行量”。

2. 网络参数的确定

本文根据以往普遍经验进行各个参数初始值的选取,初始值在 $(-3/\sqrt{m}, 3/\sqrt{m})$ 范围内取值,其中 m 为输入变量个数。初步误差设定为0.0002,学习率初步取值为0.1。

3. 隐含层设计

隐含层选取Sigmoid函数,输出层采用线性函数。本文参考公式 $N = \sqrt{m+n} + b$ 来确定隐含层节点数,其中, N 为隐含层节点数, m 为输入变量个数, n 为输出变量个数, b 为常数。初步设定隐含层节点数为3,后续通过不断增加节点数进行多次试验,对比分析误差和学习性能,从而确定最优的隐含层节点数。根据BP神经网络理论可知,隐含层层数和节点数是影响训练结果的重要指标。故为了进一步研究神经元节点数和验证所确定的隐含层层数的最优性,进行神经元个数和隐含层层数的组合测试,实验结果证实了最优的网络结构为3层神经网络层。而在相关节点选取方面,经研究发现结果最优的神经网络模型为14个节点。

此外,传递函数的选择也需要进一步优化,本文通过数值实验发现:tansig相比传递函数logsig输出的平均公差更小,所对应的网络结构更优。此外,影响BP神经网络预测性能的还有学习训练的函数类型,经过研究发现系统公差最小的为trainlm参数,故本文选取的训练函数为trainlm。采用随机抽样的方法,从总体样本中选取了12天作为试验样

本,根据以上确定的参数和相关函数所构建的基于天气因素的BP神经网络模型,以“日”为单位进行骑行量的预测,此时预测输出值和实际值(即期望值)的对比如图2所示,误差结果如图3所示。

从图2的曲线走势可以发现,11个走势中,除第8个预测结果稍有些异常导致7→8和8→9的两个趋势预测不够精准之外,其余9个趋势预测均较为准确,且准确率约为81.82%。进而从图3关于BP神经网络模型预测误差曲线图中,可看出第2、3、8个预测样本误差偏大,但相对误差不超过15%,故此模型具有较高的可靠性与准确性。为进一步检验其准确性,下面将进行模型误差对比分析。

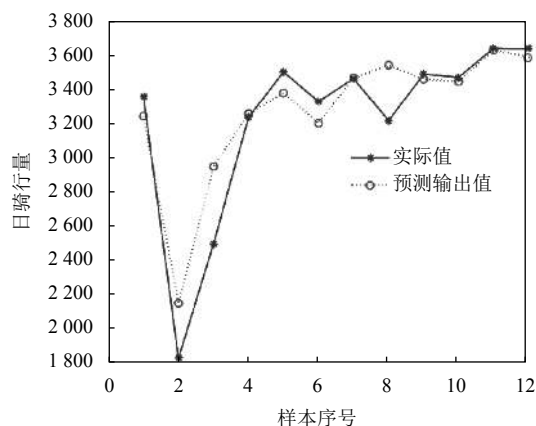


图2 实际值-预测值对比(日骑行量)

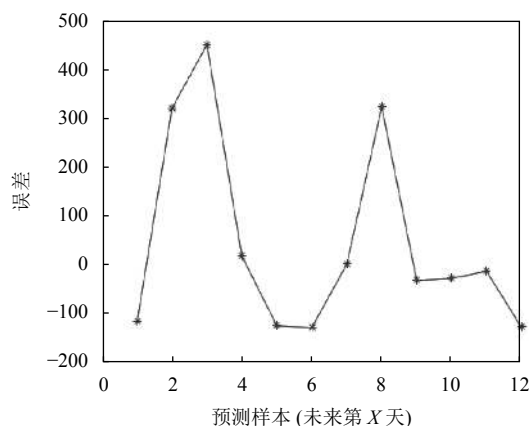


图3 BP神经网络预测误差(日骑行量)

(三) 模型准确性检验

K折交叉验证法是神经网络模型普遍采用的检验方法,K值通常由其样本总数来确定。本文选取五折交叉验证法进行检验,即将总样本数均分为5份,轮流选取其中1份作为测试样本,其余4份作为训练样本,共进行5轮验证,最后将5次的测试样本误差和训练误差求平均。本文以RMSE作为误差指标,并将其结果与ARIMA模型、多元线性回归的误差进行对比,结果如表4所示。

表 4 模型验证对比 (日骑行量)

模型类型	训练样本	测试样本	总样本
BP神经网络模型	RMSE: 125.1	RMSE: 289.1	RMSE: 163.12
ARIMA (0,1,1) 模型	RMSE: 375.39	RMSE: 504.9	RMSE: 401.7
多元线性回归模型	总样本R方: 0.429		总样本RMSE: 581.6

在表4中, R方为拟合优度, 由于非线性的特点, 多元线性回归模型的拟合优度只有0.429, 其RMSE高达581.6, ARIMA (0, 1, 1) 模型的总样本RMSE同样高达401.7, 训练样本和预测样本的RMSE均较高, 而BP神经网络模型的总样本RMSE只有163.12, 预测样本和训练样本的误差均较小, 明显优于多元线性回归模型和ARIMA (0, 1, 1) 模型。

四、基于BP神经网络模型的骑行量仿真及分析

(一) 以“日”为单位的骑行量仿真及分析

根据以上建立的改进的BP神经网络模型, 分别在给定日降雨量和日均温度下进行共享单车骑行量预测。将20 mm确定为恒定日降雨量, 以步长为2从2℃~42℃进行日均温度的取值, 共21个温度值。同理, 将22℃设定为日均温度的恒定温度, 以步长为2从2 mm~34 mm进行日降雨量的取值, 共17个降雨量值。将上述两组天气数据作为输入数据, 利用所建立的改进的BP神经网络预测模型进行共享单车日骑行量的预测, 结果如图4所示。

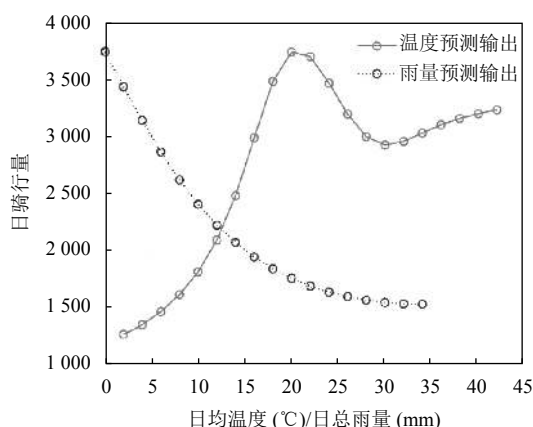


图 4 不同日均温度/降雨量下日骑行量预测

由图4可看出, 日均温度在2℃~20℃范围内, 随着温度的上升, 日骑行量明显上升, 尤其约在12℃之后, 日骑行量开始迅猛增长, 第一个日骑行量峰值出现在20℃左右, 在超过20℃之后, 日

骑行量开始逐渐下降, 在30℃时日骑行量出现极小值拐点, 之后随着温度的上升, 日骑行量再次增加。以上数据表明, 在12℃~20℃范围内, 日骑行量受温度的影响最为敏感, 骑行的最适温度范围为18℃~24℃。关于超过30℃之后, 骑行量再次增加的原因, 结合实际情况进行推测是由于高温刺激了更多用户需求, 其选择骑行代替步行, 从而减少高温灼晒时间。

同样, 通过图4可以清晰看到降雨量对骑行量的影响呈明显的指数下降趋势, 降雨量较小时日骑行量快速下降, 随着降雨量逐渐增大的同时, 骑行量下降幅度减缓。这点结合实际情况进行推测是由于在初始降雨时, 用户需求易受降雨的影响, 其会选择其他出行方式代替骑行, 从而骑行量逐渐递减; 但当降雨量超过20 mm时, 骑行量变化则趋于平稳。

(二) 以“时”为单位的骑行量仿真及分析

以下基于时刻和空间两种维度进行骑行量仿真及分析, 简称为基于“时空”的预测, 其预测方法与以“日”为单位的方法类似, 不同之处是前者输入变量增加了时刻和地点变量, 且均以小时为单位进行温度、降雨量和骑行量的统计, 故基于“时空”维度下的预测包含有更多信息, 其细粒度也更高。采用上述改进的BP神经网络预测模型, 对目标区域内不同地点的不同时刻骑行量进行预测。

由图1可知, 地点参数为8个兴趣地点。以地点1代表的电子科大食堂为例, 共收集到1116个样本, 取其中1050个作为训练样本, 剩余66个则作为预测样本, 相关预测结果如图5所示。在图5中, 横坐标为预测样本序号, 纵坐标数值为地点1 (电子科大食堂) 在特定时刻的每小时单车骑行量预测值, 从图5可发现66个预测样本的预测结果与实际输出值较为吻合, 每个峰值的骑行量也有较好地反映出来。

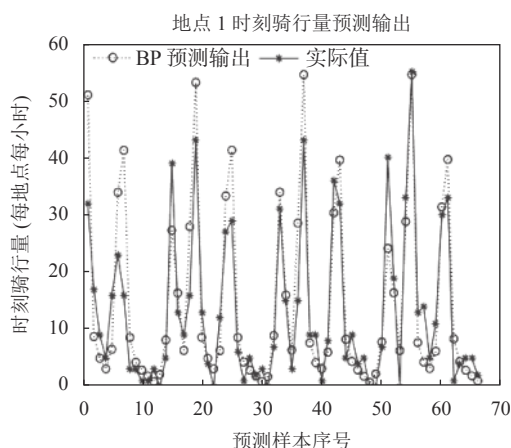


图 5 地点1时空参数下时刻骑行量预测对比

(三) 其他条件下的骑行特征分析

1. 温度影响下的不同地点骑行特征

为了研究不同地点受温度影响的骑行量变化特征, 本文在具有代表性的8个兴趣地点中选取了其中2个地点进行对比分析, 分别是地点1(电子科大食堂)和地点4(电子科大第二教学楼)。时刻参数选取编号为8的早上08:00~08:59, 此时段为上课高峰期, 温度的选取从12℃~40℃以步长为2进行统计, 时刻雨量取恒定值0 mm, 预测结果如图6。

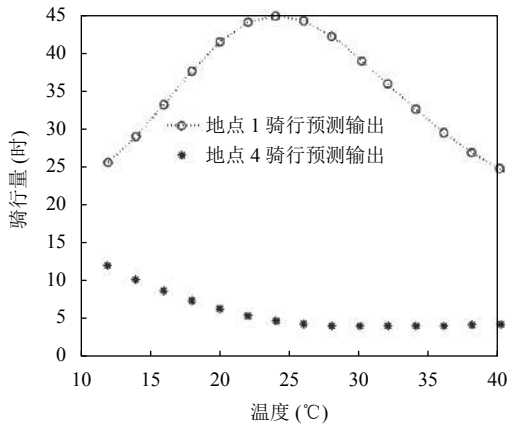


图 6 不同地点受温度影响的时刻骑行量预测

由图6可见, 地点1(电子科大食堂)每小时的骑行总量明显多于地点4(电子科大第二教学楼), 且地点1(电子科大食堂)大体呈现倒U分布, 在24℃左右时, 达到骑行量高峰。地点4(电子科大第二教学楼)的骑行量受温度影响较小, 在温度逐渐上升的过程中, 总体呈现缓慢下降趋势。结合实际情况分析, 食堂的人口流动量比教学楼多, 其骑行量固然大于教学楼的单车骑行量。

2. 雨量影响下的不同地点骑行特征

同样的处理方法, 时间上选取12:00~12:59时段(编码12), 将温度设为恒定值20℃, 降雨量以步长为2, 在0~40 mm范围内进行取值, 其预测结果如图7所示。同时为更详细了解低降雨量(椭圆圈出部分)下不同地点的骑行量特征, 图8展示了图7中降雨量从0~4 mm, 步长为0.2的骑行量更细致的预测结果。从图7和图8可以看到, 降雨量在1 mm及以下时, 地点1(电子科大食堂)和地点4(电子科大第二教学楼)的骑行量均随降雨量的增加而减少, 地点1(电子科大食堂)的下降速度尤为急剧, 骑行量从50下降到10左右。地点4(电子科大第二教学楼)的骑行量从13下降到5。而雨量在1 mm以上时, 两地骑行量受降雨量的影响均较小。

3. 不同温度下不同地点的骑行量特征

同样, 以地点1(电子科大食堂)和地点4(电子科大第二教学楼)为例, 降雨量取恒定值0 mm, 时间上以步长为1在06:00~23:59进行取值, 为更详细分析不同地点不同时刻的骑行量特征, 加入温度维度信息, 分别取14℃和20℃, 预测结果如图9所示。

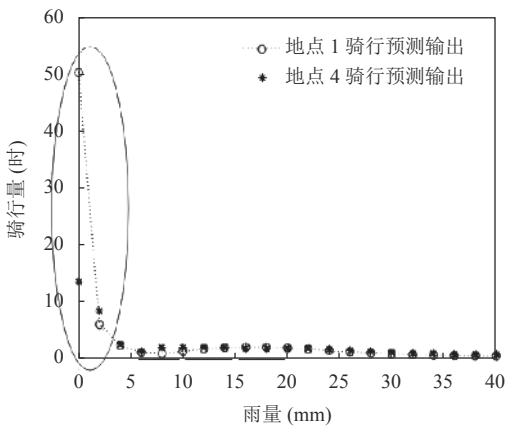


图 7 不同地点受雨量影响的时刻骑行量预测

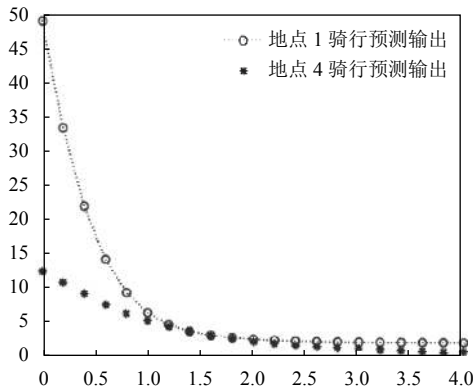


图 8 不同地点受雨量影响的时刻骑行量预测
(降雨量在4 mm以下)

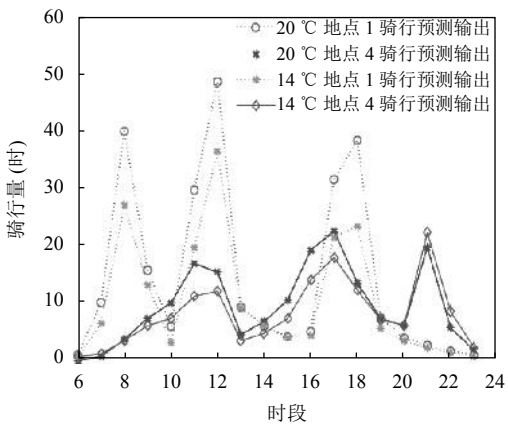


图 9 不同温度不同地点的时刻骑行量预测
(14℃、20℃)

从图9可看到,地点1(电子科大食堂)的骑行量远高于地点4(电子科大第二教学楼),其在08:00~08:59、12:00~12:59和18:00~18:59三个时间段均达到了骑行量高峰,反映了食堂就餐时间对骑行量的影响。地点4(电子科大第二教学楼)在11:00~11:59、17:00~17:59和21:00~21:59三个时间段均达到了骑行量高峰,反映出午餐和晚餐时间,教学楼(地点4)的骑行时间要早于食堂(地点1)1个小时,并且在21点存在一个开锁高峰,与实际下课或下自习情况相符合。此外,地点4(电子科大第二教学楼)的骑行量受温度影响程度明显低于地点1(电子科大食堂),特别是在骑行高峰期。

五、结论

本文基于BP神经网络建立了针对天气因素(温度和降雨量)对骑行量影响的预测模型,并通过五折交叉验证法对比证明了BP神经网络模型的可行性与可靠性,利用成都市成华区电子科技大学(沙河校区)及其周边的真实骑行数据与天气数据验证了所建模型准确度约为81%。之后,本文分别以“日”和“时”为单位分析了该区域内兴趣地点基于上述天气因素的骑行量预测结果。在以“日”为单位的骑行量仿真分析中发现,不同温度下骑行量的变化特征为,在12℃~20℃,日骑行量受温度的影响最为敏感,骑行的最适温度范围为18℃~24℃,超过30℃之后,骑行量会再次增加;不同降雨量下骑行量的变化特征为,初始降雨时,用户需求易受降雨的影响,选择步行代替骑行,从而骑行量逐渐递减;但降雨量一旦超过20 mm时,骑行量变化则趋于平稳。在以“时”为单位的骑行量仿真分析中发现,不同地点在不同时刻的骑行量受温度和降雨量的影响程度差异较大,需不同地点区别管理。从共享单车运营的角度出发,每个聚集地点和典型时段需要区别管理,并设置相应的管理优先级,如高频骑行或敏感度较高的地点和时段进行优先投放和调度,以确保这些地点和相应时段供应充足。

由于本文研究收集仅为2018年8~10月该区域内的数据,若能获得一年甚至两年的相关骑行和气象数据,经过大规模数据训练后的预测模型将具有更好的预测效果。此外,以“时”为单位进行预测的情况下,部分地点的某些时段骑行量为个位数,预测精度会有所降低。未来研究将基于跨度更大的数据,进一步完善基于BP神经网络的共享单车骑行量预测模型的准确性。

参考文献

- [1] 刘新宇,陈群.考虑动态多时段借还车需求的城市自行车租赁系统规划模型[J].系统管理学报,2020,29(1):83-91.
- [2] 郭鹏,林祥枝,黄艺,等.共享单车:互联网技术与公共服务中的协同治理[J].公共管理学报,2017,14(3):1-10+154.
- [3] 胡郁葱,黎俊廷,陈枝伟.基于双层规划的共享单车区域投放规划模型[J].中山大学学报(自然科学版),2020,59(2):48-59.
- [4] 上海市规范共享单车:中心城区暂停投放编制标准[EB/OL].(2018-05-10).<https://wenku.baidu.com/view/3929580fae1ffc4ffe4733687e21af45b307fe27.html>.
- [5] 李昆,肖计划.基于位置服务的共享单车政府监管与服务平台设计与实现[J].全球定位系统,2018,43(4):73-76.
- [6] 刘冬旭,董红召.共享单车系统动态调度时间域的获取方法[J].交通运输系统工程与信息,2018,18(2):177-182+200.
- [7] 刘冬旭,董红召.共享单车系统调度区域的分形树自平衡划分算法[J].浙江大学学报(工学版),2018,52(7):1275-1283.
- [8] 薛冰,李京忠,肖骁,等.基于兴趣点(POI)大数据的人地关系研究综述:理论、方法与应用[J].地理与地理信息科学,2019,35(6):51-60.
- [9] 王雪.共享单车押金性质与法律规制之探讨——从共享单车押金退还案件说起[J].电子科技大学学报(社科版),2019(5):45-51.
- [10] 张大涛,郝永勤.共享经济时代下公共服务体系的构建——以共享单车为例[J].电子科技大学学报(社科版),2018,20(6):12-18.
- [11] SHAHEEN S A, COHEN A P, MARTIN E W. Public bikesharing in North America: early operator understanding and emerging trends[J]. Transportation Research Record, 2018, 2387(1): 83-92.
- [12] ZHANG Y, BRUSSEL M J G, THOMAS T, et al. Mining bike-sharing travel behavior data: an investigation into trip chains and transition activities[J]. Computers, Environment and Urban Systems, 2018, 69: 39-50.
- [13] SHUI C S, SZETO W Y. A review of bicycle-sharing service planning problems[J]. Transportation Research Part C: Emerging Technologies, 2020, 117: 10264.
- [14] JIANG Z T, LEI C, OUYANG Y F. Optimal investment and management of shared bikes in a competitive market[J]. Transportation Research Part B: Methodological, 2020, 135: 143-155.
- [15] EREN E, VOLKAN EMRE U Z. A review on bike-sharing: the factors affecting bike-sharing demand[J]. Sustainable Cities and Society, 2020, 54: 101882.
- [16] 郝雅立,温志强.共建共治共享:大数据支持下共享

单车智能化治理路径[J]. 管理评论, 2019, 31(1): 249-254.

[17] 鄢章华, 刘蕾. 考虑服务水平与动态转移规律的共享单车投放策略研究[J]. 中国管理科学, 2019, 27(9): 195-204.

[18] 徐国勋, 李妍峰, 金大祥, 等. “红包车”机制下的共享单车调度问题[J]. 系统工程理论与实践, 2020, 40(2): 426-436.

[19] XU CH CH, JI J Y, LIU P. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets[J]. *Transportation Research Part C*, 2018, 95: 47-60.

[20] LIN L, HE Z B, SRINIVAS P. Predicting station-level hourly demand in a large-scale bikesharing network: a graph convolutional neural network approach[J]. *Transportation Research Part-C*, 2018, 97: 258-276.

[21] CHANG P CH, WU J L, XU Y H, et al. Bike sharing demand prediction using artificial immune system and artificial neural network[J]. *Soft Computing*, 2019, 23: 613-626.

[22] 宋鹏, 黄同愿, 刘渝桥. 基于SVM的共享单车需求预测[J]. 重庆理工大学学报(自然科学), 2019, 33(7): 187-194.

[23] 许淼, 刘宏飞, 初凯. 基于AM-LSTM模型的共享单车时空需求预测[J]. 湖南大学学报(自然科学版), 2020, 47(12): 77-85.

[24] WANG B, INHI K. Short-term prediction for bike-sharing service using machine learning[J]. *Transportation*

Research Procedia, 2018, 34: 171-178.

[25] 杨军, 赵继新, 易安军, 等. 基于BP神经网络算法的共享单车需求预测[J]. 交通工程, 2019(2): 155-158.

[26] AI Y, LI Z, GAN M, et al. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system[J]. *Neural Computing and Applications*, 2019, 31(39): 1-13.

[27] PAN Y, ZH R CH, ZHANG J X, et al. Predicting bike sharing demand using recurrent neural networks[J]. *Procedia Computer Science*, 2019, 147: 562-566.

[28] SOHRABI S, PALTI R, BALAN L, et al. Real-time prediction of public bike sharing system demand using generalized extreme value count model[J]. *Transportation Research Part A: Policy and Practice*, 2020, 133: 325-336.

[29] SATHISHKUMAR V E, PARK J W, CHO Y Y. Using data mining techniques for bike sharing demand prediction in metropolitan city[J]. *Computer Communications*, 2020, 153: 353-366.

[30] 焦志伦, 金红, 刘秉镰, 等. 大数据驱动下的共享单车短期需求预测——基于机器学习模型比较分析[J]. 商业经济与管理, 2018(8): 16-25.

[31] 曹旦旦, 范书瑞, 张艳, 等. 基于长短期记忆神经网络模型的共享单车短时需求量预测[J]. 科学技术与工程, 2020, 20(20): 8344-8349.

编辑 何婧