

基于使用行为分析的共享单车管理优化研究*

傅哲¹, 辛泓润², 余力¹, 徐冠宇³

(1.中国人民大学 信息学院, 北京 100872)

(2.北京邮电大学 国际学院, 北京 100876)

(3.北京理工大学 徐特立学院, 北京 100081)

摘要 近年来, 共享单车作为一种典型的共享经济应用的代表, 越来越受到大众的欢迎, 如何基于用户的使用行为分析来对共享单车管理进行优化是未来共享单车发展的重要问题。本文以 2016~2017 年纽约市的 Citi Bike 共享单车的使用作为研究数据集, 从时间和空间的维度详细分析全市各区域内共享单车的使用行为及特点; 采用长短期记忆 (long short-term memory, LSTM) 神经网络预测分析各站点在高峰时段的共享单车存量和净流入量情况; 在此基础上, 针对共享单车时空的分布失衡问题, 采用运筹学中的运输问题模型, 研究高峰时段的调度策略, 探索共享单车在各站点间调度的最小代价方案, 以缓解共享单车时空分布失衡的问题。本文研究对优化共享单车管理具有一定借鉴意义。

关键词 共享单车, 管理优化, LSTM 神经网络, 运输模型, 预测

中图分类号 TP311

1 引言

随着智能手机的兴起, 共享单车已经成为城市中一种重要的交通工具。截至 2017 年, 中国公共自行车系统的分布在全世界排名第一, 许多城市都有一个或者多个互相竞争的共享单车项目。随着共享单车规模的进一步扩大, 提升和优化共享单车的管理成为当今共享单车应用的关键^[1]。

共享单车管理的研究起步于 2010 年前后, 现在随着共享单车在全世界范围的快速推广, 共享单车运营管理与优化也开始越来越受到关注^[2-4]。在共享单车使用情况的分析 and 站点车辆平衡策略研究方面, Raviv 等针对需求不确定性和站点容量限制而导致的不平衡性^[5], 提出了两种站点共享单车再平衡策略, 分别是利用夜间共享单车使用率低的时间段, 进行静态的自行车重分配, 以及在白天进行动态的自行车重分配, 基于 SBRP (school bus routing problem, 校车路径问题) 模型来寻找调度共享单车数量计划的路线规划方案; O'Mahony 和 Shmoys 提出在每天用车高峰时段进行实时调度的再平衡策略^[6]; 宋明珠把共享单车的调度问题类比为旅行商问题 (travelling salesman problem, TSP), 通过蚁群算法来搜索共享单车车辆调度的最短路径^[7]。但相对来说, 上述研究多是从在不平衡状态发生后实施调度策略的角度出发, 目前还缺乏在不平衡状态发生前进行预测和提前调度的研究。同时, 除了传统的线性回归模型^[8], 越来越多的研究者现在也开始应用机器学习模型^[9]对共享单车的使用情况进行预测, 如贝叶斯网络^[10]、K 邻近算法^[11]、ARIMA (autoregressive integrated moving average, 差分自回归移动平均) 模型^[12]等, 但是总体上使用的模型方法比较简单, 也没有充分考虑用户使用共享单车行为的时间特点和空间特点。

* 基金项目: 国家自然科学基金项目 (71271209, 71331007)。

通信作者: 余力, 中国人民大学信息学院, 副教授, E-mail: buaayuli@ruc.edu.cn。

本文在对共享单车用户使用行为的时间维度和空间维度进行分析的基础上,对未来一段时间内共享单车使用状况进行预测,并预先采取调度策略来优化共享单车的管理。

本文通过对纽约市共享单车使用数据的分析,探索影响其使用数量的主要因素,并基于探索到的影响因素,对未来共享单车的使用数量进行预测并进行站点间的车辆调度。本文首先对数据从时空维度进行使用分析,寻找主要的影响因素并据此设计基于 LSTM 神经网络的共享单车预测模型。之后针对预测结果所反映出的共享单车站点间数量不平衡的情况,采用有转运的运输问题模型,对共享单车站点间的车辆调度提出优化策略,并进行实验验证。本文整体框架如图 1 所示。

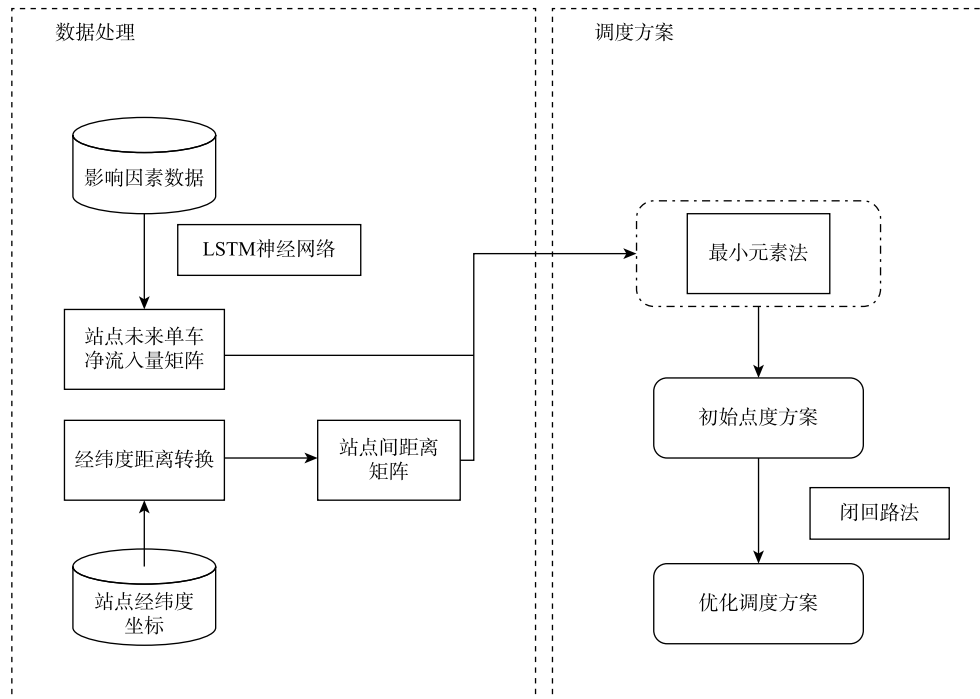


图 1 本文整体框架

本文组织如下:第 2 节,简要介绍共享单车数据集;第 3 节,分别从时间维度和空间维度对共享单车的使用行为进行较为详细的分析,并总结影响共享单车使用的关键因素;第 4 节,构建 LSTM 神经网络模型,对共享单车使用进行预测,并与其他算法进行比较;第 5 节,采用运输问题模型建立共享单车调度模型,并进行实证优化。

2 共享单车数据集

本文采用的是 2016 年 9 月~2017 年 9 月纽约市 Citi Bike 的共享单车使用的历史数据^①,该数据集包含了该市所有使用 Citi Bike 共享单车的详细记录,包括起始时间、起始站编号、起始站的经度和纬度等,同时,我们还获取了同时段的纽约天气数据。具体数据属性如表 1 和表 2 所示。

^① <http://www.citibikenyc.com/system-data>。

表 1 纽约市 Citi Bike 共享单车数据属性表

属性列	描述
start time	租用共享单车的起始时间
stop time	租用共享单车的停止时间
start_station_ID	起始站 ID 编号
start_station_longitude	租用共享单车的起始站点经度
start_station_latitude	租用共享单车的起始站点纬度
end_station_ID	终点站 ID 编号
end_station_longitude	租用共享单车的借车站点经度
end_station_latitude	租用共享单车的还车站点纬度

表 2 天气数据属性表

属性列	描述
times	日期
max_temperature_F	当日最高温
min_temperature_F	当日最低温
precipitation	降水量

3 基于时空视角的共享单车使用分析

为准确建立共享单车的使用流量数据,分析影响共享单车使用的关键因素是基础。本节我们从时间、空间两个维度对共享单车的使用进行较为具体的分析。

3.1 时间维度的共享单车使用分析

从时间维度进行分析,可以发现共享单车在各个时间段的使用情况。不同时间粒度的时间维度分析可以反映不同的影响因素。具体来讲,通过对共享单车单日的使用情况进行分析,可以观察共享单车每天的使用是否存在极度不平衡的高峰时段;通过对共享单车单月的使用情况进行分析,可以观察共享单车的使用是否受天气因素变化的影响;通过分析共享单车一年内各月的使用情况,可以观察共享单车的使用是否受到温度因素变化的影响。基于上述三个方面的考虑,本文接下来进行详细的时间维度分析。

3.1.1 单日的使用分析

在已知各个共享单车站点各时段的借车与还车数量的条件下,一个站点在短时间内是否会出现用车紧张的情况,主要取决于该站点在单位时间内的借车数量。图 2 是沿时段轴的方向进行观察的纽约市 1 月每天每小时的共享单车使用总量的变化情况图。可以发现,2017 年 1 月每天共享单车的使用总量有两个非常明显的波峰,也就是存在两个非常显著的用车高峰时段,在图中分别是 8:00~9:00 和 18:00~19:00 两个时间段,可以推断这两个时间段是用车最紧张的时候。

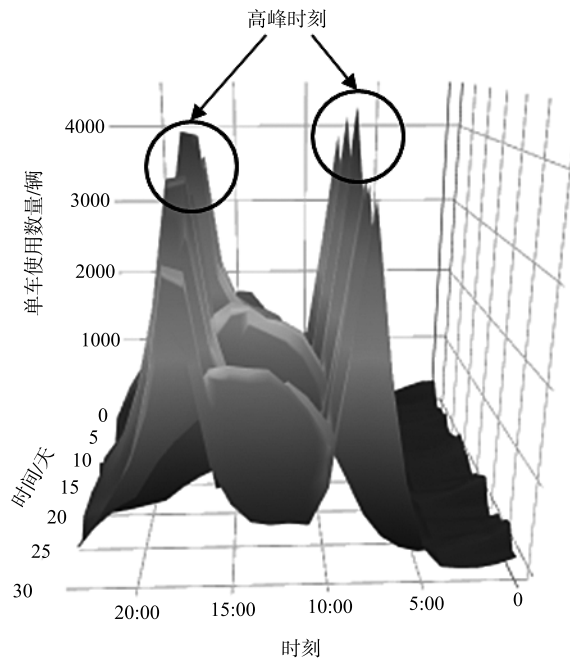


图2 2017年1月在不同时段共享单车总体使用情况

3.1.2 单月的使用分析

进一步研究各天共享单车使用数量的差别，以2017年1月为例，我们将图2调整为沿日期轴的方向进行分析的纽约市2017年1月每天的共享单车使用总量的变化情况图，如图3所示。

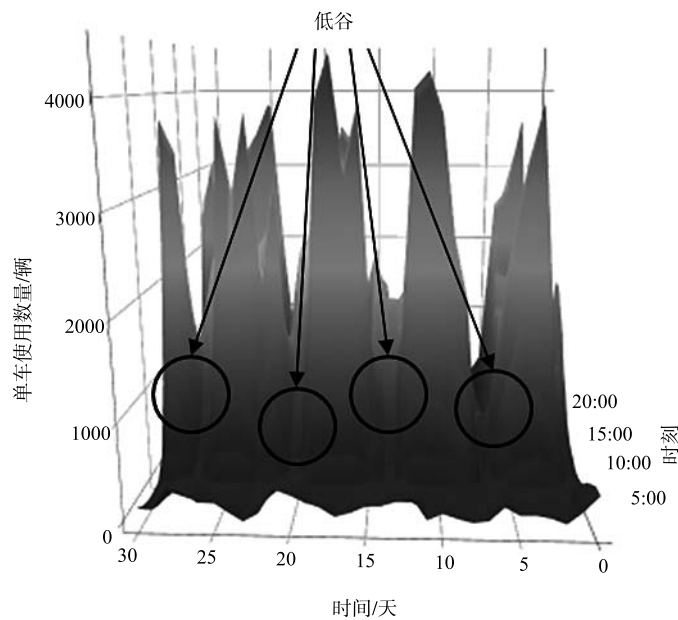


图3 2017年1月在不同日期共享单车总体使用情况

可以观察到，并不是每一天的共享单车使用总量都能保持在高位，其中有几天的共享单车使用数量相较于其他各天明显偏低。为了进一步探索原因，我们采用纽约2017年1月的天气数据进行分析。我

们对共享单车使用数量和天气相关指标进行相关性检验。结果显示,降水量(包括降雨量、降雪量和地面结冰量)与共享单车使用数量的相关系数达到 -0.61 ,呈较强的负相关关系。这在一定程度上说明降水量对共享单车的使用有着比较显著的影响。因此,我们把降水量作为共享单车预测和调度优化的依据之一。

3.1.3 各月的使用情况分析

进一步考察不同月份共享单车的使用情况,我们分析了2016年9月~2017年9月一年内各月的共享单车使用情况,如图4所示。

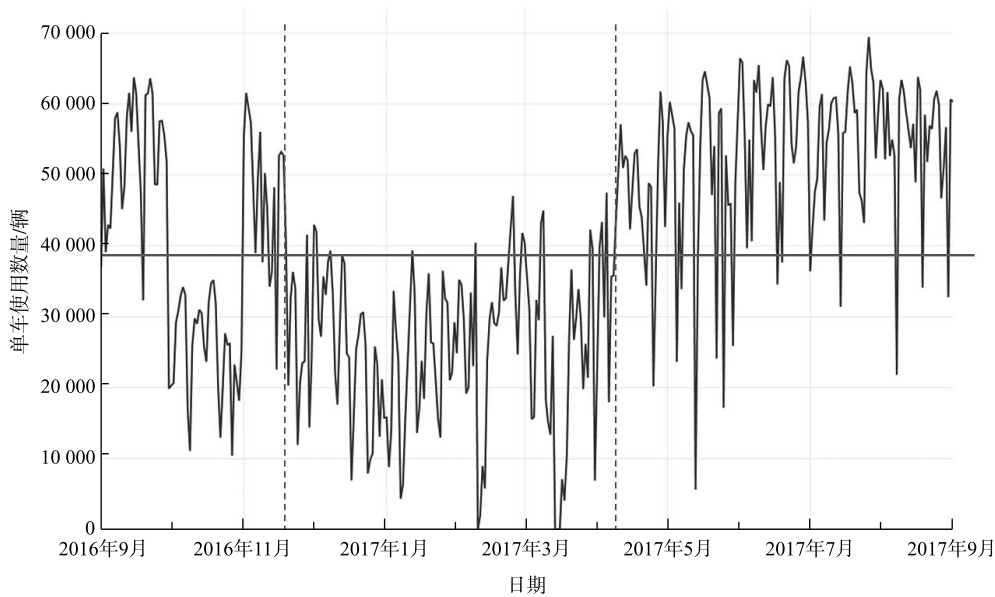


图4 2016年9月~2017年9月各月共享单车使用情况

如图4所示,在2016年9月~2017年9月各月的共享单车使用数量统计中,使用数量在2016年11月~2017年4月约五个月的时间段中,大都少于4万辆,而其余时间使用数量大都大于4万辆。实际上不难发现,中间五个月正好进入冬季降温的时段,使用数量明显小于其他时间。为此,我们对每月的气温(最高温和最低温)与当日共享单车使用数量进行相关分析。结果显示,最高温、最低温与共享单车使用数量的相关系数分别为 0.90 和 0.88 ,呈较强的正相关关系,这说明气温是影响共享单车使用数量的一个重要因素。因此,我们把每日的气温也作为影响因素纳入后文对共享单车使用数量的预测和投放量的优化方案中来。

3.2 空间维度的共享单车使用分析

从空间维度对共享单车的使用情况进行分析,可以了解不同地域共享单车使用情况的差异,为共享单车的区域调度提供参考。图5(a)和图5(b)给出了2017年1月20日8:00~9:00、18:00~19:00两个用车高峰时段的借车量从大到小排名前十的站点分布图。

从图5的展示结果来看,在早高峰和晚高峰时段,共享单车使用数量大的站点主要集中在ID为521、519和318的站点附近。可以看出,站点也是影响共享单车使用数量的重要因素之一。

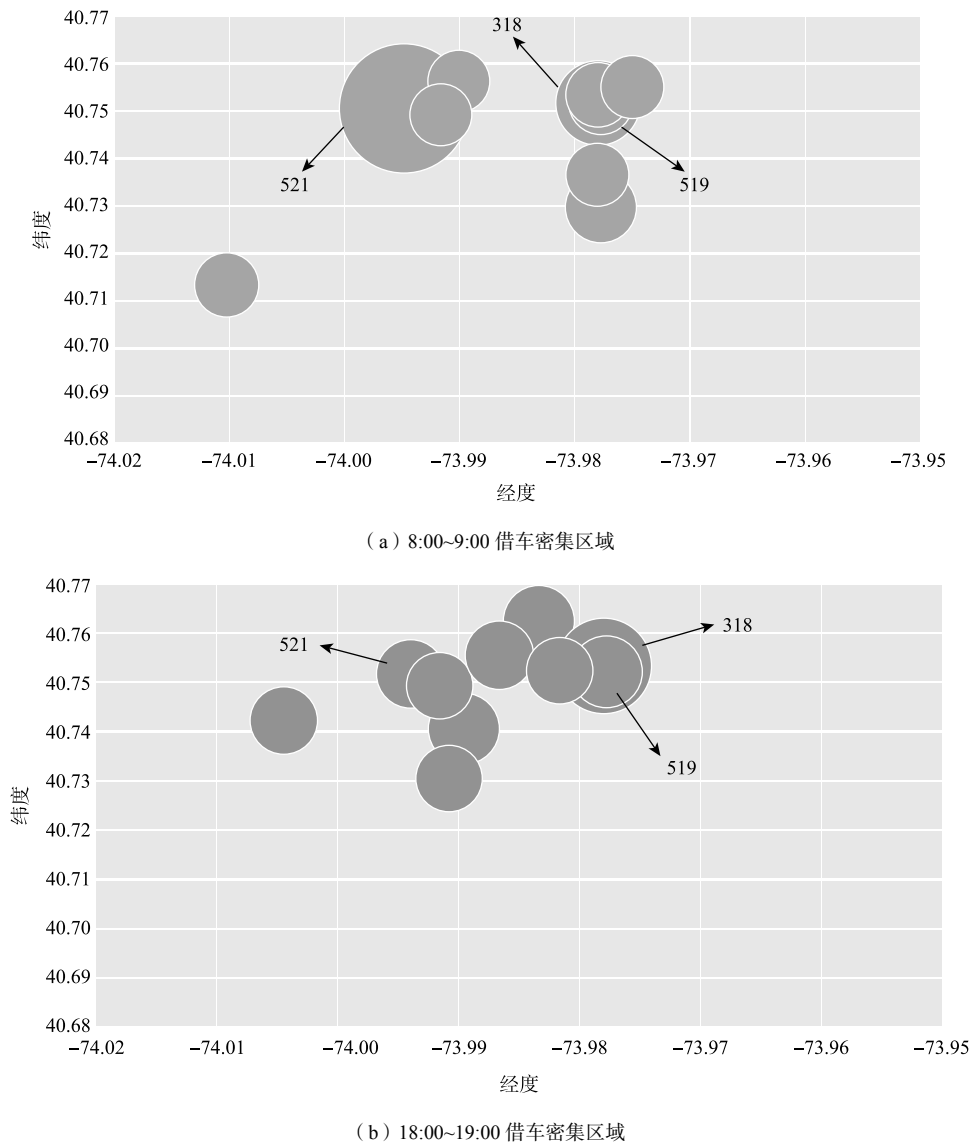


图5 高峰时段借车量前十站点分布图

3.3 影响共享单车使用的关键因素

通过从时间维度和空间维度对 Citi Bike 数据进行分析后,我们发现了 5 个影响共享单车使用的关键因素,分别是每天的时间段、降水量、最高温、最低温、站点。这些关键因素将作为特征变量,输入后文构建的共享单车深度预测模型中。

4 基于 LSTM 神经网络的共享单车流量预测

4.1 预测模型设计

根据第 3 节对共享单车使用数据的分析可以发现,共享单车的数据存在一定的周期性,每间隔一段时间会出现用车的高峰时段。在训练时,需要记忆类似高峰时段这类时间间隔长的长期信息。LSTM 神

神经网络是近年来广为使用的一种特殊的循环神经网络 (recurrent neural network, RNN), 广泛用于具有时间维度信息的预测。

相较于 RNN 的算法, LSTM 加入了一个用于判断信息是否有用的遗忘门 (forget gate), 它可以调节存储单元的自递归连接, 允许细胞根据需要记住或忘记其先前状态, 使 LSTM 神经网络可以对间隔大的长期数据中的一些特征进行记忆, 遗忘门函数为

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

式中, x_t 为 t 时刻的共享单车影响因素向量; W_f 、 U_f 为权重矩阵; h_{t-1} 为 $t-1$ 时刻的隐藏层输出向量; b_f 为偏移量。其中, 由前 $t-1$ 时刻生成的隐藏状态 h_{t-1} , 是一个用于记录 $t-1$ 时刻影响共享单车使用的重点因素的向量。在当前时刻 t 中, h_{t-1} 通过遗忘门重新调整长期记忆中各个影响因素的权重, 判断需要弱化记忆的因素和需要强化记忆的因素, 进而将时刻 $t-1$ 的信息记录到长期记忆中, 并形成新的长期记忆。LSTM 单元的内部结构如图 6 所示。

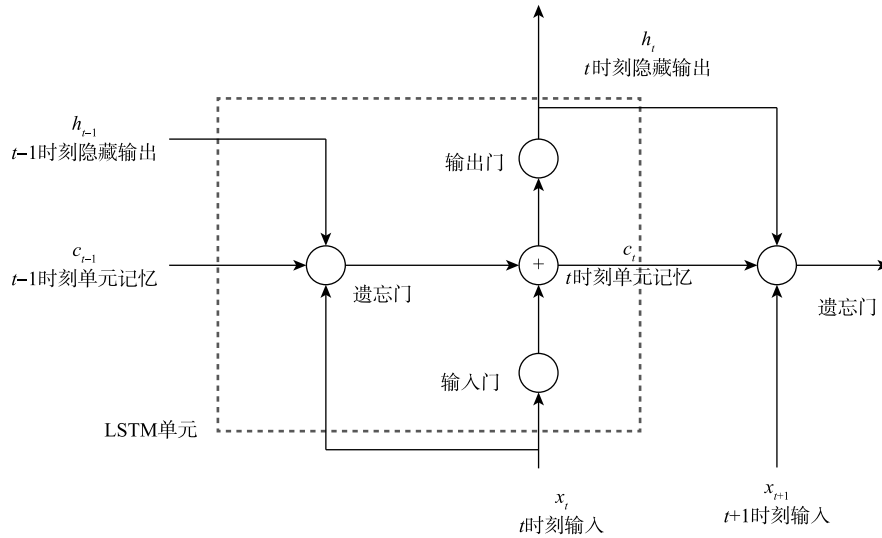


图 6 LSTM 单元内部结构

本文使用 LSTM 神经网络的方法, 对已有数据集进行训练并对未来时间点的情况做出预测, 基于 LSTM 神经网络的预测模型如图 7 所示。

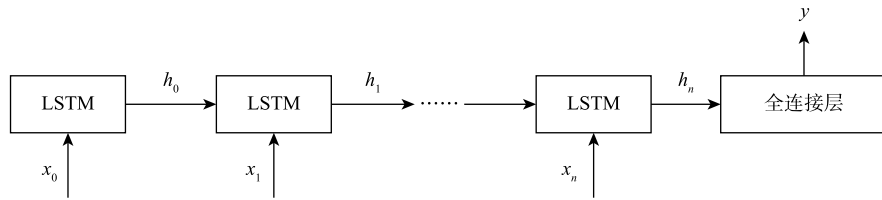


图 7 基于 LSTM 神经网络的预测模型

共享单车的预测模型算法流程如图 8 所示, 首先将共享单车数据与天气数据进行拼接; 其次对拼接后的数据做归一化处理, 并通过将时间序列向前移动一个时间步长的方式, 把时间序列问题转换为有监督的学习问题; 最后将处理后的数据输入基于 LSTM 神经网络的预测模型中进行预测。

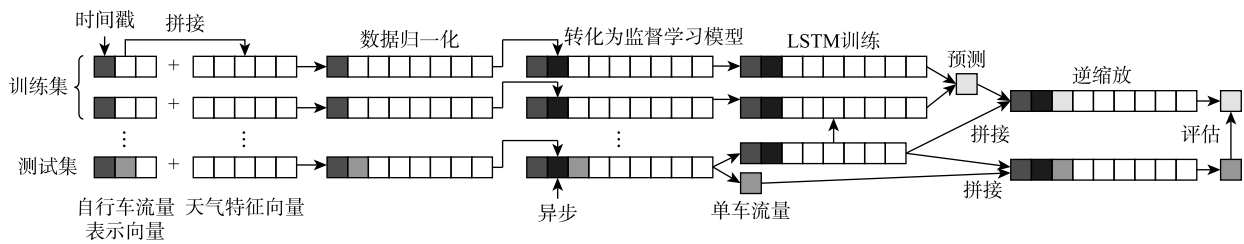


图 8 预测模型算法流程

4.2 共享单车使用流量预测实验

4.2.1 实验设计

由于共享单车站点的调度量首先取决于该站点的净流入，即某站点的一定时间和区间的还车数减去借车数，本文将会选取纽约市 2016 年 9 月~2017 年 8 月的数据作为训练集，对各站点的净流入量的数值进行预测，并将 2017 年 9 月的数据作为测试集进行检验。

为了评测 LSTM 的预测效果，本文设置三个对照组进行实验。

(1) ARIMA。ARIMA 模型是一种经典的时间序列预测模型，只学习共享单车数量的时序变化，对比增加了五种影响因素的 LSTM 预测模型，检验 LSTM 预测模型是否有更好的预测效果。

(2) RNN。RNN 结构与 LSTM 基本类似，可以很好地处理时序问题，但是缺少遗忘机制，无法有效地记忆长期的信息。将 RNN 与 LSTM 神经网络进行对照，检验 LSTM 预测模型的遗忘机制在共享单车预测中是否具有优势。

(3) RBF^①。RBF 神经网络是适用于非线性情况下的预测模型。将 RBF 神经网络与擅长处理时序问题的 LSTM 进行对比，检验适用于时序问题的 LSTM 预测模型是否比处理非时序问题的 RBF 神经网络更有效。

为了评测结果，本文定义了预测的准确率 P ：

$$P = \frac{C}{S}$$

式中， C 为预测的使用数量与测试集的实际使用数量相等的个数； S 为该站点测试集的总数量。本文通过比较四个模型对各个站点的预测结果的平均准确率与均方根误差来评测预测的效果。

4.2.2 实验结果

为了进行实验，本文同时采用同样适用于时序问题预测的 RNN、适用于非线性数据预测的 RBF 神经网络，以及传统的 ARIMA 模型进行预测，并与 LSTM 神经网络的预测效果进行对比。最后将预测结果与测试集进行比较，表 3 展示了四种模型预测结果的平均准确率和均方根误差。

表 3 神经网络预测效果对比表

模型	平均准确率	均方根误差
LSTM	55.7%	2.42
RNN	47.2%	2.49
RBF	6.4%	6.60
ARIMA	55.2%	2.45

① RBF, radical basis function, 表示径向基函数。

通过四种方法的对比可以看出，LSTM 神经网络预测结果的平均准确率和均方根误差都是其中效果比较好的，因此后文对站点 2017 年 9 月 1 日的净流入量进行预测时，采用基于 LSTM 神经网络的预测模型进行预测。

5 基于运输问题模型的调度策略

5.1 调度模型的建立

基于站点净流量的预测，我们可以对高峰时段的共享单车进行有效调度。一般来讲，调度主要从调度成本的角度来考虑，也就是调出站点距离调入站点越近，成本越少、效率也越高。对于某个未来的时间点，在已知周围各个站点共享单车的预测存量的情况下，可以近似将车辆的调度问题看成是一个产品的运输问题。

对于运输问题的经典描述是：设某种产品有 m 个生产地点 P_1, P_2, \dots, P_m ，各个生产地点的生产量分别是 p_1, p_2, \dots, p_m ；又有 n 个销售地点 S_1, S_2, \dots, S_n ，各个销售地点的销售量分别是 s_1, s_2, \dots, s_n 。假设从生产地点 P_i 向销售地点 S_j 运输单个产品的运输成本是 c_{ij} ，寻找最优的方案使产品从生产地点运输到销售地点的运输成本最小。表 4 为调度运输问题的模型。

表 4 调度运输问题的模型

调度		调出站点		调入站点		调出量		
		1	\dots	m	$m+1$		\dots	$m+n$
调出站点	1	0	\dots	c_{1m}	$c_{1,m+1}$	\dots	$c_{1,m+n}$	$Q+p_1$
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	m	c_{m1}	\dots	0	$c_{m,m+1}$	\dots	$c_{m,m+n}$	$Q+p_m$
调入站点	$m+1$	c_{11}	\dots	c_{1m}	0	\dots	$c_{1,m+n}$	Q
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	$m+n$	c_{m1}	\dots	c_{mm}	$c_{m,m+1}$	\dots	0	Q
调入量		Q	\dots	Q	$Q+s_{m+1}$	\dots	$Q+s_{m+n}$	

为了使运输问题适用于单个站点共享单车的调度问题，本文对运输问题的描述进行了改进。将周围调出共享单车的调出站点看作生产地点，调入共享单车的调入站点看作销售地点，调出单车的调出站点到调入站点间的距离作为运输成本，并试图寻找最优的方案，使共享单车从调出站点到调入站点之间的运输总距离最小。假设车辆调出站点的富余单车总量大于等于调入站点的需求数量，即有

$$\sum_{i=1}^m p_i \geq \sum_{j=1}^n s_j$$

因此，可以将共享单车的车辆调度问题用如下的数学模型表示

$$\min z = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

$$\begin{cases} \sum_{j=1}^n x_{ij} \leq p_i (i=1,2,\dots,m) \\ \sum_{i=1}^m x_{ij} = s_j (j=1,2,\dots,n) \\ x_{ij} \geq 0 \end{cases}$$

同时,我们考虑到,在调度单车时,最经济的方案并不是直接将各个站点的富余单车运往调入站点,而是中间需要经过其他站点转运,然后再运输到调入站点。这样,就将原来的运输问题拓展为有转运的运输问题。假设站点的转运成本忽略不计,将车辆调出站点与车辆调入站点统一编号,并将车辆调入站点编号排在调出站点的前面。

令 x_{ij} 代表第 i 个调出站点运输到第 j 个调入站点的共享单车数量, Q 代表共享单车需求量,则共享单车调度运输问题如表 5 所示。

表 5 共享单车调度运输问题

调度		调出站点		调入站点		调出量		
		1	...	m	m+1		...	m+n
调出站点	1	x_{11}	...	x_{1m}	$x_{1,m+1}$...	$x_{1,m+n}$	$Q + p_1$
	\vdots	\vdots		\vdots	\vdots		\vdots	\vdots
	m	x_{m1}	...	x_{mm}	$x_{m,m+1}$...	$x_{m,m+n}$	$Q + p_m$
调入站点	m+1	x_{11}	...	x_{1m}	$x_{1,m+1}$...	$x_{1,m+n}$	Q
	\vdots	\vdots		\vdots	\vdots		\vdots	\vdots
	m+n	x_{m1}	...	x_{mm}	$x_{m,m+1}$...	$x_{m,m+n}$	Q
调入量		Q		$Q + s_{m+1}$		$Q + s_{m+n}$		

5.2 两阶段调度求解过程

根据有转运的运输问题模型求解的基本思路,我们提出两阶段调度求解算法,如表 6 所示。该算法总体包含两个阶段:首先采用最小元素法计算初始调度方案;其次采用闭回路法计算优化调度方案。

表 6 基于运输问题模型的共享调车调度求解算法

<p>#第一步:最小元素法计算初始调度方案</p> <p>Input: Station: 站点集合</p> <p>S_i: 站点 i 的单车供给量</p> <p>D_j: 站点 j 的单车需求量</p> <p>Output: 站点 i 到站点 j 的调度策略 res_{ij}</p> <ol style="list-style-type: none"> for $i, j \in \text{Station}$ do if $s_i > d_j$ then $res_{ij} \leftarrow d_j$ else $res_{ij} \leftarrow s_i$
<p>#第二步:闭回路法计算优化调度方案</p> <p>Input: $Ires$: 最小元素法获得的初始调度策略矩阵</p> <p>Cost: 运输成本矩阵</p> <p>Station: 站点集合</p> <p>Output: 优化后的调度策略矩阵 $Nres$</p> <ol style="list-style-type: none"> for $i, j, x, y \in \text{Station}$ do 对 $Ires_{ij}, Ires_{xy}, Ires_{iy}, Ires_{xj}$ 进行排序 if $Ires_{xy}=0$ or $Ires_{ij}=0$ then 对 $Ires_{xj}, Ires_{iy}$ 排序,并取最小值 min 进行调整 计算调整后的成本 d

6. if $d > 0$ then
7. 代入 \min 值对 I_{res} 进行调整得到 N_{res}
8. else if $res_{sj}=0$ or $res_{iv}=0$ then
9. 对 $I_{res_{xy}}, I_{res_{ij}}$ 排序, 并取最小值 \min 进行调整
10. 计算调整后的成本 d
11. if $d > 0$ then
12. 代入 \min 值对 I_{res} 进行调整得到 N_{res}

5.3 共享单车调度实证研究

首先选取纽约市 2016 年 9 月~2017 年 8 月的数据作为训练集对神经网络进行训练。根据上文对影响因素的分析, 选取这几个月中每天的时间段、降水量、最高温、最低温、站点五个因素作为特征, 按照站点顺序对数据用 LSTM 神经网络进行训练拟合, 并对 1 小时内共享单车站点的共享单车净流入量进行预测。

为了便于检验, 本文随机选取距离 ID 为 521 的站点 150 米范围内的 5 个活跃的站点作为研究样例, 以 ID 为 457、520、447、402、521 的站点为例, 利用纽约 2016 年 9 月~2017 年 8 月的共享单车数据, 对这些站点在 2017 年 9 月 1 日上午 8:00~9:00 的共享单车净流入量进行预测。同时, 我们通过查阅美国国家气象局 2017 年 9 月的数据可知, 2017 年 9 月 1 日当日的气温最低温为 55 华氏度 (约 12.8 摄氏度), 最高温为 71 华氏度 (约 21.7 摄氏度), 降水量为 0, 通过 LSTM 神经网络拟合后预测的该时段繁忙站点共享单车净流入量如表 7 所示。

表 7 2017 年 9 月 1 日繁忙站点共享单车净流入量预测值表

站点 ID	净流入量预测值
457	11
520	12
447	-1
402	12
521	-33

通过检验可以发现, ID 为 447 和 521 的站点在 8:00~9:00 的时候会出现共享单车大量流出而导致的紧缺状况, 属于调入站点; 而 ID 为 520、402 和 457 的站点则会出现共享单车流入富余的状况, 属于调出站点。首先将各个站点间的距离填入站点间运输距离代价表 (表 8) 中。

表 8 站点间运输距离代价表

站点距离/米		调出站点			调入站点		调出量
		520	402	457	447	521	
调出站点	520	0	150	56	60	131	46
	402	150	0	178	152	72	46
	457	56	178	0	30	135	45
调入站点	447	60	152	30	0	104	34
	521	131	72	135	104	0	34
调入量		34	34	34	35	67	

利用最小元素法给出初始的调度方案, 由于这是一个供销不平衡问题, 本文需要增加一个假想的调入站点 B 。利用闭回路法对上解进行检验和优化, 在经过 1 次迭代后, 得到最优解, 如表 9 所示。

表 9 最终调度方案表

单车调度/辆		调出站点			调入站点			调出量
		520	402	457	447	521	B	
调出站点	520	35	0	0	0	12	0	47
	402	0	35	0	0	12	0	47
	457	0	0	35	10	0	1	46
调入站点	447	0	0	0	26	9	0	35
	521	0	0	0	0	35	0	35
	B	0	0	0	0	0	35	35
调入量		35	35	35	36	68	36	

因此总调度代价为

$$\begin{aligned}
 z &= c_{15}x_{15} + c_{25}x_{25} + c_{34}x_{34} + c_{36}x_{36} + c_{45}x_{45} \\
 &= 12 \times 131 + 12 \times 72 + 10 \times 30 + 1 \times 0 + 9 \times 104 \\
 &= 3672(\text{米})
 \end{aligned}$$

此时, 最优的共享单车调度方案是: 由 ID 为 402 的站点直接运送 12 辆共享单车到 ID 为 521 的站点; 由 ID 为 520 的站点直接运送 12 辆共享单车到 ID 为 521 的站点; 由 ID 为 457 的站点先运送 10 辆共享单车到 ID 为 447 的站点, 再由 ID 为 447 的站点转运 9 辆共享单车到 ID 为 521 的站点。

进一步, 将上述调度算法用代码实现, 并拓展到纽约市所有的共享单车站点上, 最终得到站点间最优调度算法, 并得出了站点间车辆调度数量大于 15 的调度策略。

6 结论

本文为优化共享单车的管理, 着力研究了区域中存在的共享单车时空分布不均衡的问题, 在此问题的基础上利用人工神经网络, 建立了站点内共享单车存量和净流入量的预测模型, 并选取了研究样例进行测试研究, 最终在预测结果的基础上寻找出车辆调度优化的方案。本文的主要结论有: ①在进行站点共享单车数量预测前, 对影响站点共享单车使用数量变化的时空特点进行分析, 并最终找到了时间段、最高温、最低温、降水量、站点五个对共享单车使用数量有显著影响的因素。②在找到了影响站点共享单车使用数量的五个因素后, 对比 LSTM 神经网络、RNN、RBF 神经网络及纯时序的 ARIMA 四种模型, 对未来各时间段的各个站点共享单车净流入量进行预测, 并发现 LSTM 神经网络的预测效果最佳。③在通过人工神经网络预测出各站点的共享单车的净流入量后, 根据区域内各个站点的共享单车使用情况, 类比于有转运的运输问题, 利用最小元素法和闭回路法求解车辆调度的路径策略。但本文在研究纽约市共享单车调度方案时, 忽略了城市中具体的道路状况对调度代价的影响。同时在考虑夜间车辆再平衡策略时, 没有考虑一个站点的共享单车存放数量有限的情况。

参 考 文 献

- [1] Shaheen S, Guzman S. Worldwide bikesharing[J]. Access Magazine, 2011, 1 (39): 22-27.
- [2] Shaheen S, Guzman S, Zhang H. Bikesharing in Europe, the Americas, and Asia: past, present, and future[J]. Transportation Research Record: Journal of the Transportation Research Board, 2010, 1 (2143): 159-167.
- [3] Shaheen S, Zhang H, Martin E, et al. China's Hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing in Hangzhou, China[J]. Transportation Research Record, 2011 (2247): 33-41.
- [4] DeMaio P. Bike-sharing: history, impacts, models of provision, and future[J]. Journal of Public Transportation, 2009, 12 (4): 41-56.
- [5] Raviv T, Tzur M, Forma I A. Static repositioning in a bike-sharing system: models and solution approaches[J]. EURO Journal on Transportation and Logistics, 2013, 2 (3): 187-229.
- [6] O'Mahony E, Shmoys D B. Data analysis and optimization for (Citi)bike sharing[C]. Austin: Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: 687-694.
- [7] 宋明珠. 公共自行车调度模型理论分析与案例研究[D]. 北京: 清华大学, 2014.
- [8] Singhvi D, Singhvi S, Frazier P I, et al. Predicting bike usage for New York city's bike sharing system[C]. Austin: Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: 110-114.
- [9] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 97.
- [10] Froehlich J, Neumann J, Oliver N. Sensing and predicting the pulse of the city through shared bicycling[C]. Pasadena: 21st IJCAI, 2009: 1420-1426.
- [11] Liu J M, Sun L L, Chen W W, et al. Rebalancing bike sharing systems: a multi-source data smart optimization[C]. San Francisco: SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 1005-1014.
- [12] Kaltenbrunner A, Meza R, Grivolla J, et al. Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system[J]. Pervasive and Mobile Computing, 2010, 6 (4): 455-466.

Management and Optimization of Shared Renting Bicycle Based on User Behavior Analysis

FU Zhe¹, XIN Hongrun², YU Li¹, XU Guanyu³

(1. School of Information, Renmin University of China, Beijing 100872, China)

(2. International School, Beijing University of Posts and Telecommunications, Beijing 100876, China)

(3. Xu Teli School, Beijing Institute of Technology, Beijing 100081, China)

Abstract In recent years, as a typical representative of sharing economy application, renting bicycle is more and more popular. It is key issue about how to analyze the behavior of users based on the use of shared bicycle to improve their management. In this paper, we will focus on the problem of unbalanced distribution of bicycles in the region. We will use LSTM network to train and forecast the data of Citi Bike and use the minimum element method to find the most minimum cost scheme. The research contents includes the follows: Exploring the distribution characteristics of shared bicycles in time and space; Using LSTM neural network to predict the stock and net inflow of shared bikes in stations during peak hours; Based on the forecast of the use of bicycles in each station, finding the scheduling strategy of peak season. Through the study of the results of the New York Citi Bike, we can not only further optimize the operation of the bike sharing management and alleviate the problem of unbalanced distribution of sharing bicycle, but share the experience to operation of China's bike sharing system as well.

Key words Sharing renting bicycle, Management and optimization, LSTM neural network, Transportation modeling, Prediction

作者简介

傅哲 (1995—), 男, 中国人民大学信息学院硕士研究生, 研究方向: 管理信息系统、大数据分析。
E-mail: fuzhe@ruc.edu.cn。

辛泓润(1997—),男,北京邮电大学国际学院本科生,研究方向:电子商务、信息管理。E-mail: xinhongrun@bupt.edu.cn。

余力(1973—),男,中国人民大学信息学院副教授、博士生导师,研究方向:大数据分析、信息管理、机器学习等。E-mail: buaayuli@ruc.edu.cn。

徐冠宇(1999—),男,北京理工大学徐特立学院本科生,研究方向:大数据分析、机器学习。E-mail: axuguanyu@163.com。