

# 同濟大學

TONGJI UNIVERSITY

## 毕业设计（论文）

课题名称 共享单车出行时空特征分析与优化调度研究

副 标 题 \_\_\_\_\_

学 院 交通运输工程学院

专 业 交通工程

学生姓名 李炘玲

学 号 1753298

指导教师 沈煜

日 期 2021 年 6 月 1 日

### 共享单车出行时空特征分析与优化调度研究

#### 摘 要

共享单车调度是提高共享单车系统的运行效率的重要手段之一，通过将单车从极度过饱和的区域运送到欠饱和区域，可以实现不同区域的单车供需平衡。本文首先基于共享单车 GPS 位移数据，解析共享单车出行的时空分布特征，采用可解释机器学习方法建模共享单车出行需求。在需求预测模型基础上，本文创新性地提出了基于网络流模型的共享单车调度方法，并以新加坡义顺区的实际共享单车使用情况为案例进行分析，验证了此模型的合理性和有效性；并继续利用此模型对调度成本、调度用车数量、共享单车数量进行了对于网络最终收益的敏感性分析。分析结果显示（1）可解释性的机器学习模型在保证预测精度的情况下对各因素的影响效果进行解释，与传统基于一般化线性回归模型结果保持一致（2）建立的网络流模型对调度计划制订可以提供理论依据，具有一定的有效性（3）系统收益与所分析影响因素间有明显联系。本文中所讨论的方法为共享单车网络中的单车投放数量以及定价提供了参考，并给出了有关共享单车网络管理的一些建议。

**关键词：**共享单车，调度问题，网络流量模型

# The Research on the Bikesharing Relocation Problem and Influential Factors

## ABSTRACT

Bikesharing relocation is one of the most important strategies to improve the operation efficiency of the bikesharing systems. Through redistributing bikes from extremely bike sufficient areas to bike shortage areas, better bike-rider balance across different areas can be achieved. In this study, we first use GPS data to do spatial-temporal analysis of bikesharing usage patterns and use the interpretable machine learning model to make bike demand prediction. On the base of the prediction model, a innovative network flow model is proposed to solve the optimal relocation problem of shared bikes, and is implemented with the actual dockless shared bike usage data from Yishun, Singapore, to demonstrate its effectiveness. Then a series of sensitivity analyses are performed to test the impact of the relocation cost, the number of bikes and truck trikes, and the usage price on bike relocation. The results show that (1)interpretable machine learning model can explain the influence of the variables on the result when also guaranteeing the accuracy, the explanation is consistent with the traditional generalized liner regression model. (2)the network flow model we developed can provide insight for making relocation operation scheme (3) there is an apparent connection between the profitability of the system and the analyzed factors. This work offers a modeling framework to start and operate a bikesharing service by determining the number of bikes and trikes as well as price schemes. Some bikesharing regulation policies are also suggested.

**Keywords:** (Dockless) Bike-sharing, Fleet Relocation, Network Flow Model

### 目 录

1 绪 论 .....	1
1.1 研究背景 .....	1
1.2 研究概况 .....	2
1.3 本文的工作框架 .....	4
2 共享单车出行特征与影响因素分析 .....	6
2.1 单车使用总体情况 .....	6
2.2 共享单车空间分布 .....	7
2.3 需求随时间变化规律 .....	9
2.4 新加坡天气情况 .....	9
3 需求预测 .....	13
3.1 线性回归 .....	13
3.2 广义线性回归（GLM） .....	13
3.3 分类算法 .....	14
3.4 单车需求预测 .....	14
3.4.1 数据准备 .....	14
3.4.2 类泊松回归 .....	15
3.4.3 KNN .....	16
3.4.4 方法比较 .....	16
4 调度模型建立 .....	20
4.1 问题阐述 .....	20
4.2 问题抽象与建模 .....	21
5 结果分析 .....	24
5.1 研究区域选取 .....	24
5.2 实例和结果 .....	25
5.3 基本实例：调度行为分析 .....	26
5.3.1 敏感性分析 .....	28
5.3.2 网络规模敏感性分析 .....	29
5.3.3 调度成本敏感性分析 .....	32
5.3.4 定价敏感性分析 .....	33
5.3.5 模型延申：增加站点容量的限制的共享单车系统 .....	33
6 总结与讨论 .....	35
参考文献 .....	37
致 谢 .....	40

### 1 绪论

#### 1.1 研究背景

为了减少尾气排放，并解决城市出行第一和最后一公里的问题，过去的十几年里，共享单车项目的建设在全世界内快速增长。总结来说，共享单车网络运营模式的发展过程可以分为以下三个阶段：

- （1）无锁免费共享单车
- （2）有桩基于押金的共享单车
- （3）基于信息系统的无桩共享单车<sup>[1-3]</sup>。

其中，第一、第二代共享单车系统通常被看作最基础的共享单车系统，他们缺乏有效的信息记录和管理设备，无法有效的进行单车管理和防止单车盗窃，所以最终大多走向失败。在第三代共享单车系统中，停车桩和自动缴费使得共享单车更加易于管理，而信息设备也使单车的追踪成为可能。伴随着科技的发展，这些设备同样也在不断的升级换代，从而提供了越来越优质的服务。近年来，随着移动信息服务的发展，新一代的共享单车系统——无桩共享单车系统逐渐占领了共享单车市场。这种系统中的车辆配备了移动支付模块，GPS 芯片，成为了目前共享单车系统的主流运营模式。

无桩共享单车系统能够追踪系统中所有车辆的位置和状态，从而使得车辆的管理和维护更具实时性和有效性。然而，在现实中，无桩共享单车项目还是常常难以长期运营，最终走向破产。在中国，自 2017 年以来，带有 GPS 的共享单车项目发展尤其迅猛，各运营商对市场的竞争十分激烈。2017 年，摩拜和 ofo 进入中国市场，并开始了激烈的市场竞争，他们都大量投放共享单车。到 2018 年，这两个运营商共同占有了 90% 的市场份额。但是到 2019 年初期，大多数共享单车运营公司，包括 ofo 在内，却已经倒闭或面临倒闭的危机。这些项目的失败说明，在不同运营商都投放了大量共享单车到市场中时，为了更好的管理，运营商需要对共享单车的运营模式和规律有更深入的理解。共享单车的过度投放会带来诸多问题，例如造成公共空间资源的浪费（如下图 1.1、1.2 所示），且并不能保证高利用率，但是投放不足又会减少运营商在市场中的份额和盈利能力，最终也会导致破产。如何利用尽量少的共享单车实现尽可能多的盈利是一个重要议题。



图 1.1 共享单车过度投放（图片来源于网络）



图 1.2 共享单车过度投放（图片来源于网络）

平衡单车数量和盈利空间的方式之一就是使用单车调度。相比于购买更多的单车，调度可以更加灵活的调整共享单车系统提供服务的能力，在单车量适当的情况下，达到最大化运营商的盈利空间的目的。然而，调度计划的制订也需要进行仔细的考虑，因为共享单车系统通常分布面积广，过多的无效调度会给系统带来多余的运营成本，同样可能导致项目的破产。在这种背景下，对调度计划制订的讨论就显得尤为重要。

在共享单车系统中，调度已经被广泛用于解决在单车数量有限的情况下，不同区间单车分布不平衡造成的单车利用率低的问题。在之前的研究中，很多学者已经提出了许多用于指定单车调度策略的方案，其中既包括针对早期的有桩共享单车系统的方法，也包括针对新式的无桩的单车系统的方法。但事实上，共享单车的有桩和无桩性质在学者们的研究过程中并没有显示本质区别，因为即使在无桩共享单车的建模过程中，研究者们通常也需要将空间连续变量离散化来表示某位置停靠的单车数量，否则问题的规模将会无限大，所以在有桩和无桩单车建模过程中，最显著一个区别是桩的数量是否是系统中单车分配的限制条件。在无桩共享单车系统中，不同单车停放点可以看作没有数量约束的停车桩。在下一节中，我们将先对单车调度模型研究领域的研究重点和成果进行总结。在之后的文章中，我们也将通过从新加坡最大的共享单车运营商之一的 Obike 公司获得的真实数据，讨论单车过度投放的问题，作为调度模型构建的基础。

### 1.2 研究概况

过去几十年来，已有大量研究着眼于对于传统的有桩共享单车系统的调度问题分析。由于停车桩的占用率和单车的可用性代表着系统的服务水平以及单车分布的合理性，大部对于有桩共享单车系统的研究都会将停车桩的占用率作为系统服务的一个评价指标。除此之外，过去的大部分研究都以路径规划问题的视角考虑调度问题，特别是单一货物、单一车辆的路径规划问题（One-commodity vehicle routing problem）。其基本思想为预先计算出每个站点需要的单车数量（货物数量），再计算如何使一辆调度用车在最短的时间（或距离）内满足这些站点的需求。

总结起来，若以调度的动态性划分，以往对于共享单车调度问题的研究可以大致分为以下两大类：

- （1）只在夜间发生的静态调度（忽略在调度过程中消费者的使用对于单车分布的影响）
- （2）动态调度（将消费者的使用也纳入考虑范围的运营时间内调度）

有关以上的两类调度问题的研究通常关注以下几点问题：

- （1）需求预测
- （2）车辆路径规划



由于在静态调度问题并不考虑消费者的使用对于调度计划的影响，所以在静态调度相关的研究中，大部分研究者都不会讨论需求预测这一步。而且正如前面提到的，共享单车系统通常涉及到的区域范围广，所以问题规模通常很大，这就导致一部分研究者相比于系统建模过程更加关注寻找更高效的模型求解算法，所以这一类的研究通常旨在为调度模型提供高效且使用的求解算法。例如，对于调度问题，就有学者提出过用 Benders 分解扩展传统的分支定界<sup>[4]</sup>、混合大规模临域搜索<sup>[5]</sup>、松弛算法<sup>[6]</sup> 这些算法进行求解。除了在传统的确定性混合整数规划模型的基础上进行改进，启发式算法在求解大规模问题中也运用广泛<sup>[7, 8]</sup>。这些启发式算法既有旨在在求解阶段对求解进行加速的，也有在问题分析层面对原问题进行分解的。除了直接从算法层面对求解进行改进，一些研究者还从不同的角度理解调度问题进行建模，利用旅行商问题的建模思路对问题进行抽象<sup>[9]</sup>，此研究与传统的以注消费者满意度和调度成本为优化目标不同的是，目标函数是找到可用单车的方便程度，所有未被满足的需求和调度花费的时间的加权平均值。

虽然静态调度在建模和求解上都更简单，也更易实施，但是对于大部分的城市而言，一天中总是存在高峰和平峰时段，对于单车的需求也就同样存在高峰与平峰。在高峰时段，不同站点间的共享单车需求具有高度的不平衡性，高峰时段过后也很可能会导致单车分布不符合下一高峰的需求。在这种情况下，动态调度是必要甚至是必须的。由于在动态调度中，需求是不可或缺的模式输入，所以在有关动态调度研究中通常也会对需求分析进行讨论。需求分析的方法很多样，有利用马尔可夫过程对系统状态转移进行建模，用以刻画系统中需求的产生和变化规律的<sup>[9]</sup>，也用利用仿真模拟消费者的期望起终点选择的<sup>[10]</sup>。无论是利用马尔可夫还是仿真模拟，这些方法都充分考虑了需求产生的原理及其不确定性。为了得到更加弹性、能适应更多现实情况的需求预测方法，有一些研究者对需求和环境、社会因素之间的联系进行了细致的分析。这些因素通常包括温度、天气情况、距离、交通状况等<sup>[11, 12]</sup>。

除了以上希望通过对共享单车需求的产生过程进行细致刻画和分析对共享单车需求进行预测，还有一些研究关注如何利用历史数据对未来需求进行直接预测。具体方法主要集中在线性回归、机器学习和决策树三部分，其中使用线性回归或广义线性回归进行需求预测使用的最为普遍。

直接需求预测中最早的研究为基于单点的共享单车需求预测。Kaltenbrunner 等在 2010 年提出了一种完全基于同期历史数据的需求预测方法<sup>[13]</sup>，通过求过去同一时间段每个站点进出单车数量的变化率，可以求出连续时间内任意时刻站内共享单车的数量，从而决定单车是否过饱和或欠饱和。其它的一些研究集中于通过定义时间、环境（主要是天气）不同的变量表示利用线性回归确定参数从而进行需求预测。在线性回归前往往需要对时间、天气状况这样的变量进行定义和处理，由于难以保证变量间的独立性以及时间、天气与需求间的线性关系，这一类直接回归的模型准确度往往较低。所以在此基础上，Shen.Y 等在对需求影响因素分析的基础上对变量进行了‘spatial log’的非线性处理，再利用线性回归求解，结果也证明了这种方法相比直接线性回归具有更高的精确性<sup>[14]</sup>。由于线性回归出色的可解释性和可以接受的精确度，很多研究都采用这种方法作为需求预测的基础。当然，以上这些模型都假设了单车需求能通过一个具体的函数进行预测，但是在现实生活中，需求的产生往往十分复杂，且具有高度随机性，所以基于机器学习<sup>[15]</sup>或基于决策树<sup>[16]</sup>的需求预测研究也因其更高的精确度得到越来越多的研究，但相应的计算成本也更高。

随着站点间需求的相关性和单车需求的动态程度越来越显著，人们逐渐意识到，基于单点的需求预测准确度较低的一个重要原因就是它忽略了需求的不确定性与动态性。所以在此基础上，很多研究者提出了基于“群”的需求预测方法<sup>[17, 18]</sup>，通过站点间需求的相似性先将站点划分为站点群，从而缩小问题规模，再预测群之间的单车流动关系。相比于直接利用线性回归进行需求预测，这种研究方法需要先自定义场景和指标划分站点类型，所以需要的数据种类也更多，除时间

和天气状况外，可能还需要额外的社会事件乃至交通状态信息。

以上讨论的研究都是针对有桩共享单车系统的研究。近年来，随着无桩共享单车系统的快速发展，对于无桩共享单车调度问题的研究也得到了越来越多的关注。然而，相比于已经发展了几十年的有桩共享单车系统而言，无桩共享单车系统的可用数据以及相关研究基础更少，所以仍缺乏对于无桩共享单车系统调度问题的全面讨论和升入研究。但值得注意的是，已经有一些学者在关注对于无桩共享单车系统特点的讨论<sup>[14, 19, 20]</sup>。基于用户的调度在无桩共享单车调度方法的讨论中也得到了进一步发展。类似的研究有助于促进对调度问题进行更深入、更多样化的探讨。

对于调度问题的研究思路已经越来越突破传统路径规划问题的约束。在动态调度问题的研究中，除了与静态调度相似的采用路径规划作为建模基础的分析思路外，还有一些学者提出了利用奖励机制作为调度手段的方法<sup>[21, 22]</sup>。这些研究仔细分析了消费者选择起终点的过程，并基于此给出设置奖励机制鼓励消费者从期望出发站点周围的站点取车或者在期望目的站点周围的站点还车的方法。具体的，有学者通过训练新的强化学习网络解决调度问题，这个网络设计的目的就在于构建一套最优定价算法，激励消费者在高需求区域需要用车时选择在附近区域取车以提高系统的服务水平<sup>[23]</sup>。在此基础上，另一些学者考虑了激励消费者不要在停车车辆过多区域还车的机制<sup>[24, 25]</sup>。除了改变调度主体之外，还有学者提出了在更大尺度上考虑调度问题二点思路，比如首先构建一套将研究区域进行动态聚类的方法，将研究区域划分为一个个分区域，在这些区域间讨论调度问题将能大大降低问题规模<sup>[26]</sup>。

以上讨论的调度建模思路和方法最终给出的都是具体到时间的确定性调度方案，适用于在网络运行过程中制定具体的调度计划。近两年来，还有部分研究者提出一种新的建模方法：基于网络排队的建模方法。Rick 和 Marco 已将这种方法用在网约车调度系统中<sup>[27]</sup>。在这个系统中，用户为服务台，网约车为被服务者，模型求解可以得到网约车网络中需要的调度员数量和调度频率。这种模型构建方式得到的结果对运营前共享单车系统的车辆部署、调度用车辆部署有借鉴意义，Maleki Vishkaei 等就在共享单车网络中利用网络排队模型进行了最佳站点容量计算以匹配各站点的需求，从而达到减少调度需求的目的<sup>[28]</sup>。

### 1.3 本文的工作框架

本文将首先基于新加坡的共享单车数据对单车的使用情况进行刻画，以期对新加坡共享单车的使用以及有关环境因素情况进行大致刻画，之后再对共享单车需求预测的方法进行讨论，并依据现有的数据情况选用最适宜已有数据的需求预测方法。在此基础上，我们构建一个从有桩共享单车网络中衍生出的用于解决单车调度的网络流模型，结合从新加坡得到的共享单车实际使用数据以及前述需求预测模型，对调度模型进行了测试检验，证明其合理与有效性。之后，为了加深对不同单车数量、调度卡车数量等因素对收益的影响的理解，利用模型与数据对这些因素进行敏感性分析。最后，对共享单车运营商和政府有关部门提出了一些管理商的建议，以期在保证共享单车系统持续盈利能力的基础上，为社会公众提供更加便捷的共享单车服务。整体的工作路线如图1.3所示。

本文旨在提出一套完整的调度设计流程，从需求预测，到模型的构建，为实际调度计划的制订提供理论支撑。同时为了加深对于共享单车网络运营的理解，利用模型进行网络中各项因素的敏感性分析，以期对共享单车网络的更高效运营提供参考的切入点，使运营商在保证盈利的情况下有能力提供更好的出行用户服务。



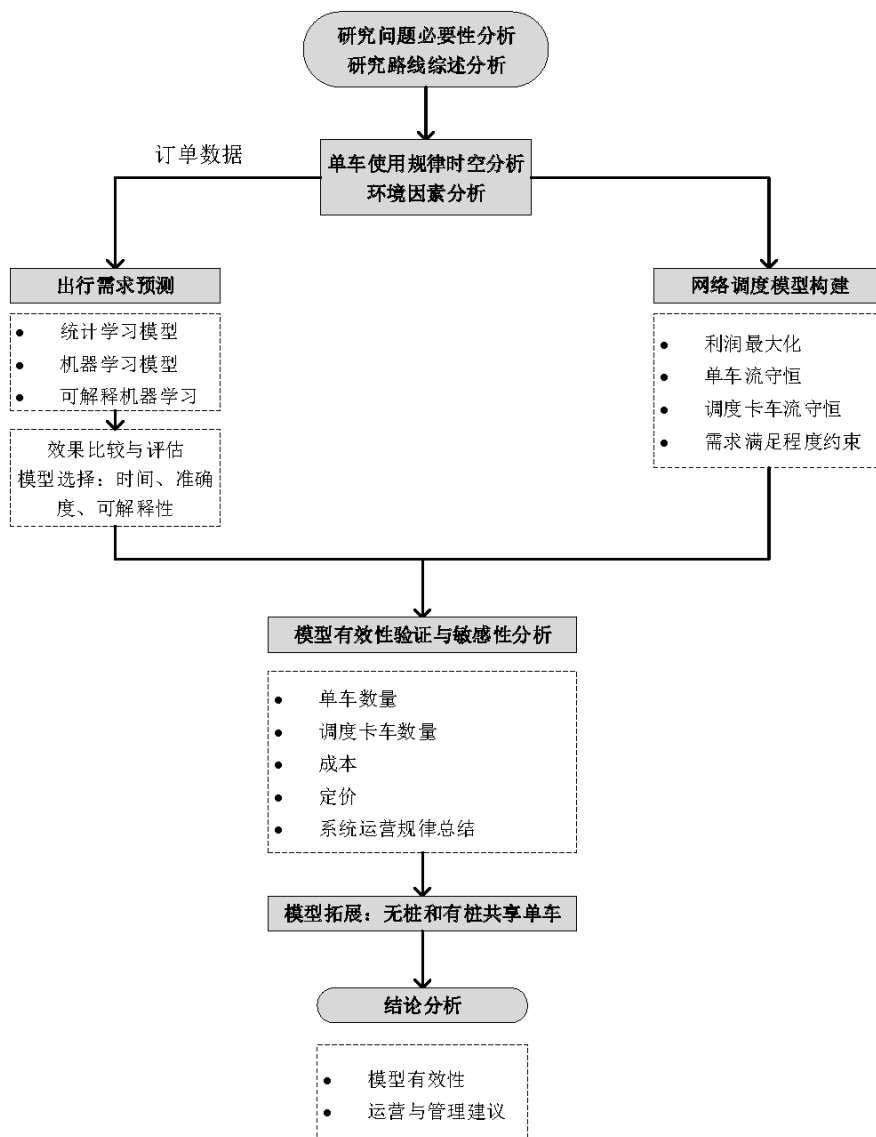


图 1.3 工作路线

## 2 共享单车出行特征与影响因素分析

本章将对共享单车整体使用情况、新加坡共享单车空间分布、共享单车需求随时间变化情况以及有关环境因素进行分析。

无桩共享单车在 2017 年首次进入新加坡的共享单车市场，Obike 就是最早开始的一批项目之一。在 Obike 的共享单车系统中，所有单车都配备了 GPS 定位仪，定位仪记录的数据使得对共享单车时空轨迹的刻画，以及对单车分布的分析成为了可能。本文的所有有关单车位置、使用情况的分析都是基于从 Obike 公司获取的 2017 年 9 月 30 日至 11 月 1 日的订单数据，每条订单数据中都包含起终点的位置以及时间信息。

除了来自 obike 的订单数据外，本研究还从新加坡开源数据网站 <https://data.gov.sg/> 获取了全新加坡监测站记录的有关环境的数据进行与单车使用有关的环境情况分析。

### 2.1 单车使用总体情况

从十月的出行订单中，统计所有用户骑行的距离和时间分布如下图2.1所示。由图中可以看出，单车骑行时间基本在 20 分钟以下，距离基本在两公里以内，属于正常的出行行为。以下数据分布情况可以在之后的数据处理过程中使用，即对于部分极端值，可以采用阈值法过滤处理。

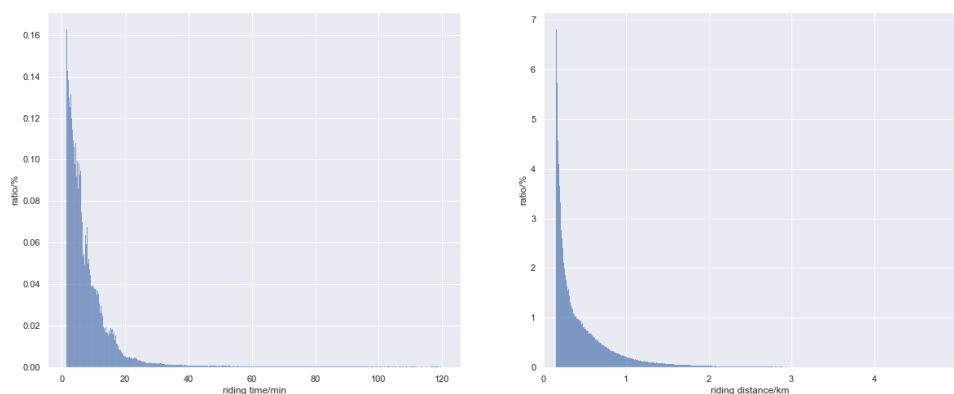


图 2.1 使用情况

除以上每次使用的统计情况之外，选取十月中的一周统计总车辆数和日单车使用次数、日平均单车使用时间如下表2.1所示，表中的列分别代表日期、此日期内被使用的单车数量、每辆被使用过的单车平均使用时长、每辆被使用过的单车使用的平均距离。

从下表中可以看出，每天内被使用的单车数量差异比较大，说明每天都有部分单车处于完全未被使用的状态。每辆单车每天的平均使用时间和距离都很短，而且平均时长、平均距离都是在当天被使用过的单车的尺度内计算的，事实上由于有完全没被使用过的单车，这个数值在所有单车范围内的值会更小，这些数据都直观的说明单车的使用率比较低。

除了从表中可以直接推断出的使用率低之外，从订单数据中还可以发现，每天被使用的最频繁的单车使用总时长超过两小时，这说明不同单车使用时间之间还存在较大的差异，而单车的投放位置是决定其使用的最大影响因素，这说明了单车合理分布对充分利用系统运营能力的重要性。

表 2.1 单车使用情况

Date	Used bikes	Average use time/min	Average use distance/km
2017-10-09	14260	11	0.9
2017-10-10	13867	11	0.9
2017-10-11	14836	12	0.8
2017-10-12	12750	11	0.9
2017-10-13	14790	12	1.0
2017-10-14	17444	14	1.0
2017-10-15	17192	13	0.9
2017-10-16	15301	12	0.9
2017-10-17	15426	13	0.9
2017-10-18	16464	14	0.8
2017-10-19	15233	13	0.8

### 2.2 共享单车空间分布

本文通过对 2017 年 10 月 17 日一天的订单数据统计进行共享单车空间分布的展示。由于单车分布具有瞬时性，所以只选取一天中早上九点这一时刻进行分布分析。根据订单数据得到单车分布的统计方法为：选取在早上九点正在使用的所有单车在此趟行程中产生的订单，由于此刻单车正在运动过程中无法决定其位置，所以取订单开始位置为这辆单车在早上九点的位置。对于剩下的没有正在被使用的单车，取其在早上九点前的最后一条订单，该订单的终止位置即为此单车在早上九点的位置。由于空间的连续性，所以统计前先将空间进行离散化，以方格划分整个新加坡区域，单车的计数以划分的土地块为范围。

通过以上步骤得到的共享单车的日分布如图2.2所示。图中的颜色越深，表示单车分布的数量越多。除了单车分布之外，下图中还显示了新加坡地铁站的位置和地铁线路的走向。从图中可以看出，共享单车在地铁站周围的分布密度明显高于其它地区，且单车数量有随距地铁站距离增大而减小的趋势。这种分布选择对于运营商来说是合理的，因为地铁站周围通常有大量的交通需求，而高需求意味着更高的盈利额。图中单车分布的范围也大致表示了新加坡有人群居住的地域范围。

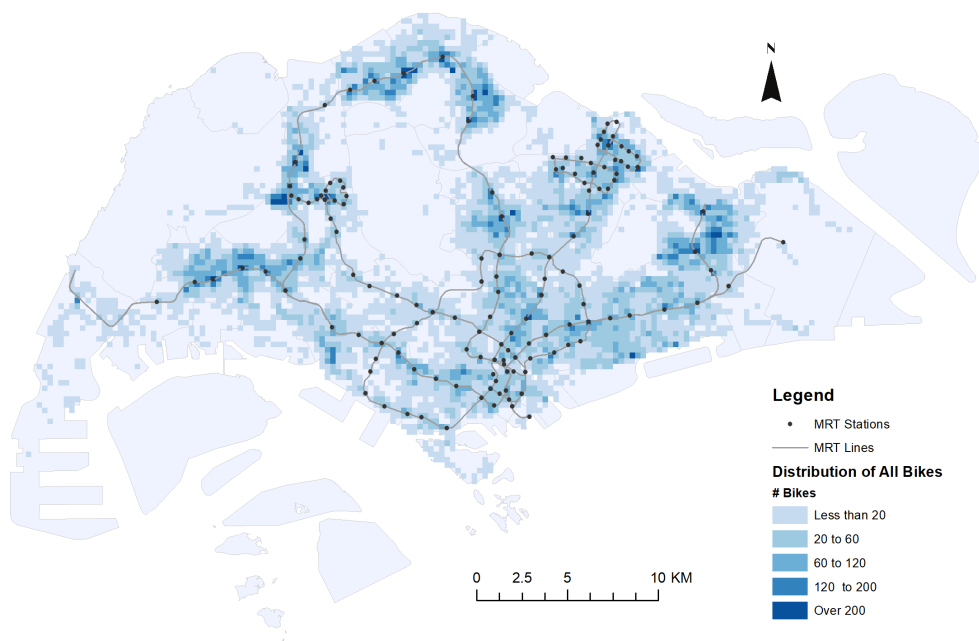


图 2.2 共享单车日分布

图2.3展示了 10 月内超过一周没有被使用的车辆比例分布，这个比例可以视为显示该区域中单车过度投放程度的指标。从图2.2、图2.3两张图中可以很直观的看出，高过度投放程度的区域和高投放量的区域是互补的。除此之外，过度投放程度比较高的区域分布在远离地铁站的区域。这些现象证明单车并没有合理分布，不合理的分布导致了不充分的使用，所以在这个系统中，将单车从那些地利用率的区域调度到需求更高的区域是十分必要的。

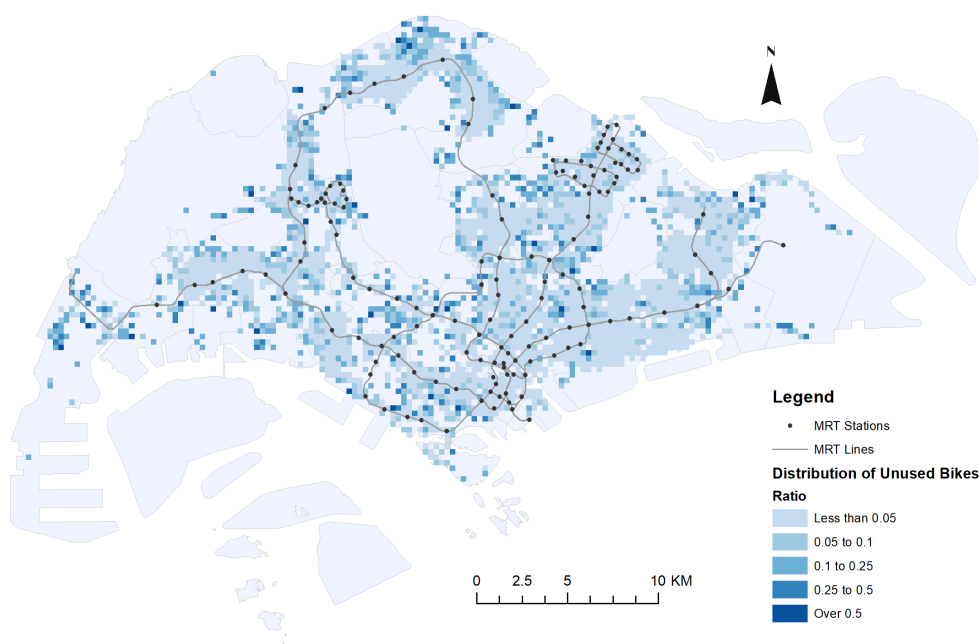


图 2.3 未使用车辆空间分布

## 2.3 需求随时间变化规律

与人们的出行规律一致，共享单车的使用需求会随时间变化，且同样具有一定的规律性。选取几天的需求随时间变化规律展示如图2.4所示。图中显示，工作日和周末的使用规律并不完全相同。在工作日，造成七点至十点、晚上五点到八点存在明显的使用高峰，而周末这种高峰规律会相对较弱，但在晚间仍然存在使用高峰。由于工作日和周末的规律不同，所以进行需求预测时，我们将两者分开建模，以提高模型预测的准确度。

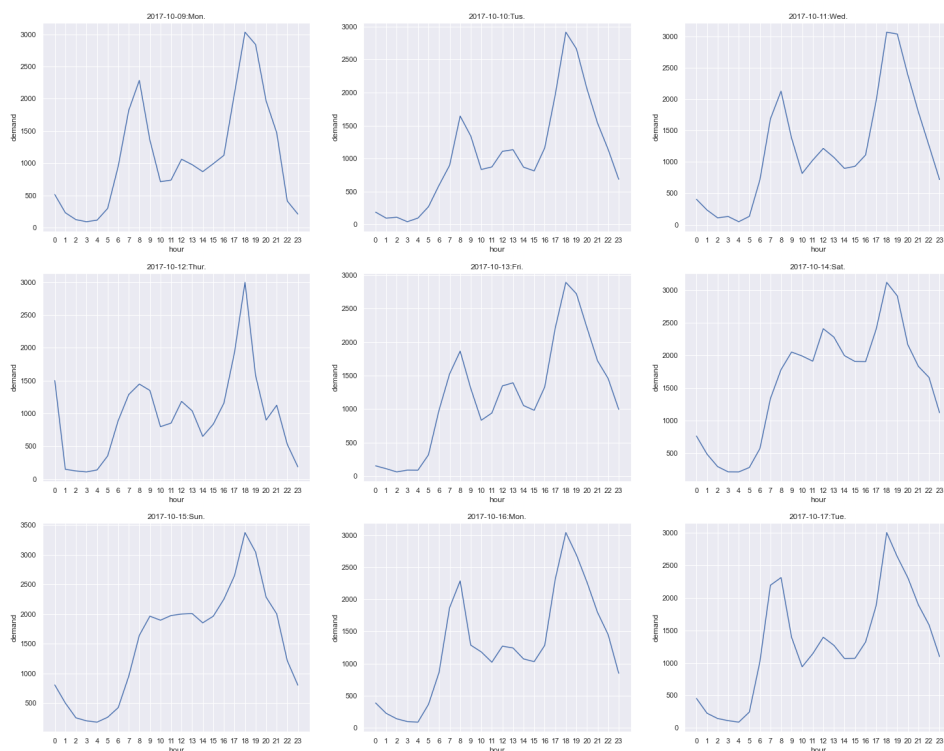


图 2.4 需求随时间变化

## 2.4 新加坡天气情况

由于温度和降雨是影响单车出行的两个重要因素，所以接下来，我们对新加坡的温度变化和降雨情况进行分析。仍然选取与 2.3 中相同的时间，通过从新加坡开源数据网站 <https://data.gov.sg/> 获取的监测站温度数据，得到新加坡一天中温度的变化如图2.5所示。由图中可以看出，新加坡的天气基本在中午十二点至下午四点达到一天中的最高峰，但一天中温度波动并不大，基本在五摄氏度以内。事实上，新加坡全年的温度大多都分布在 20-35 摄氏度（如图2.6所示），月最低和最高气温在全年内都没有显著的季节区分，所以温度变化对单车使用的影响在新加坡可能不如其它地区显著。

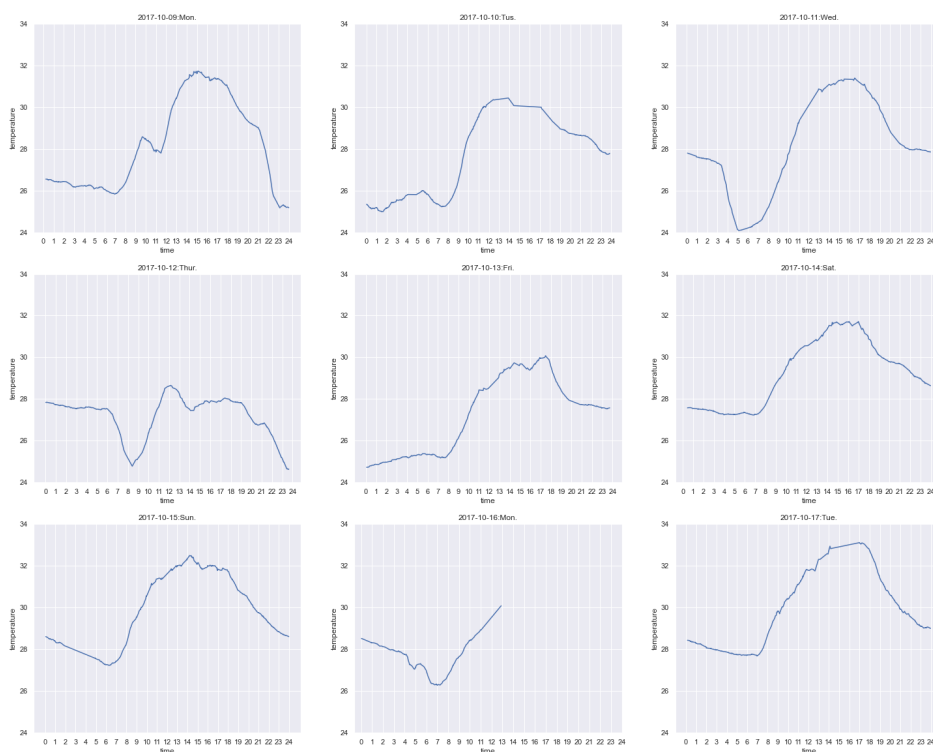


图 2.5 温度随时间变化

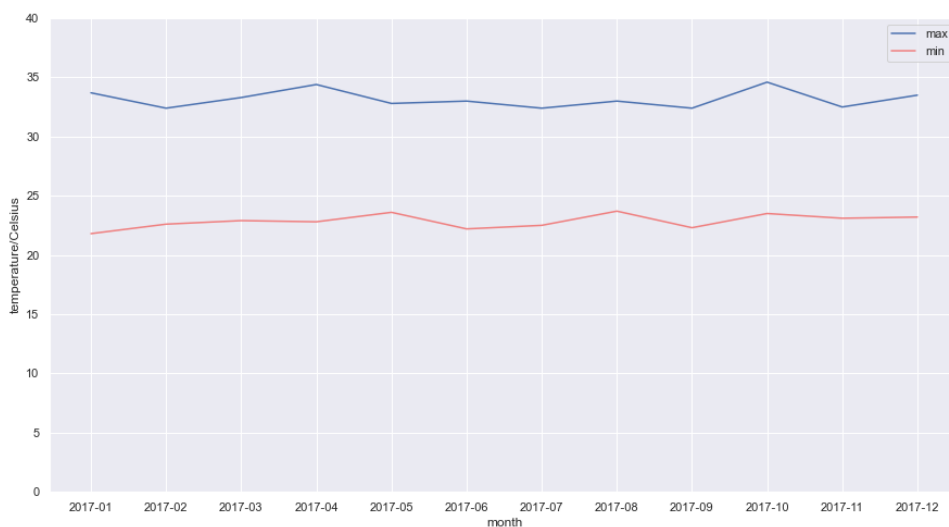


图 2.6 一年中温度变化

在以上选取的日期内，新加坡在十月九日、十月十二日、十月十三日有少量降雨，降雨等级都为小雨。从全年来看，新加坡属于降雨量比较大的区域，新加坡降雨情况如下图2.7所示。两图



分别为月降雨量、每月降雨天数。从图中可以看出，新加坡每月都存在一定的降雨，但在每天内，新加坡的降雨属于阵雨，通常只在每个时段内下，并不太会全天处于雨天天气。

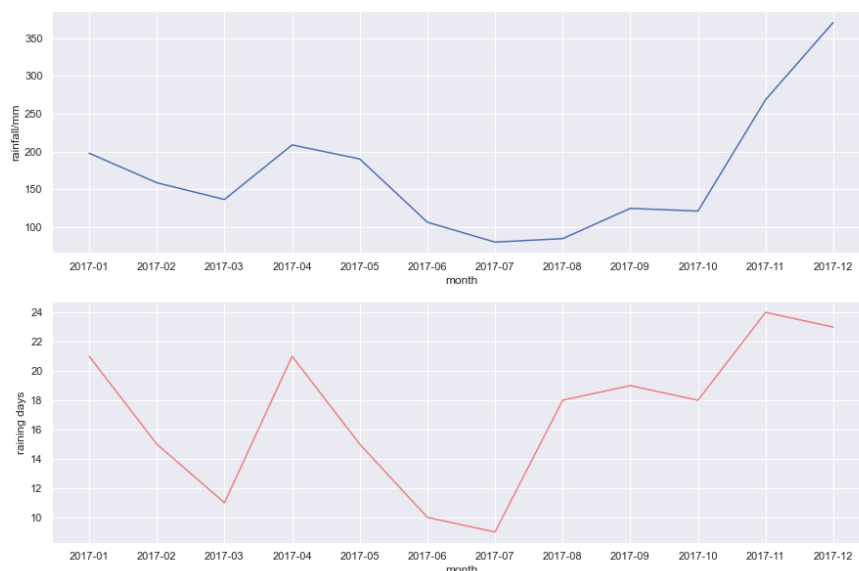


图 2.7 新加坡降雨情况

虽然降雨量比较大，但新加坡的降雨还存在比较明显的地域差异。任意选取某两个小时内全新加坡监测站的降雨数据，进行克里金插值，得到新加坡各个区域内的降雨情况如图2.8、2.9所示。

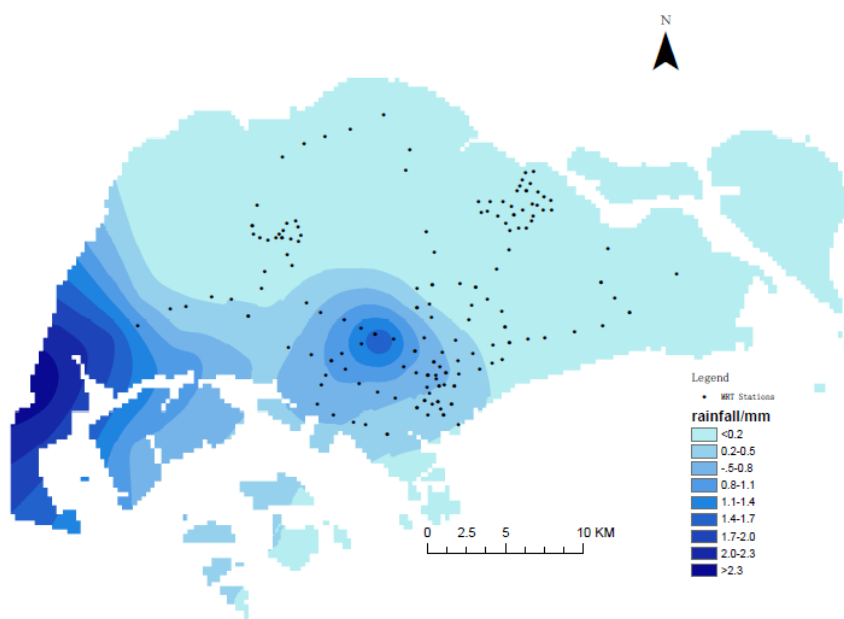


图 2.8 新加坡不同区域降雨情况（1）

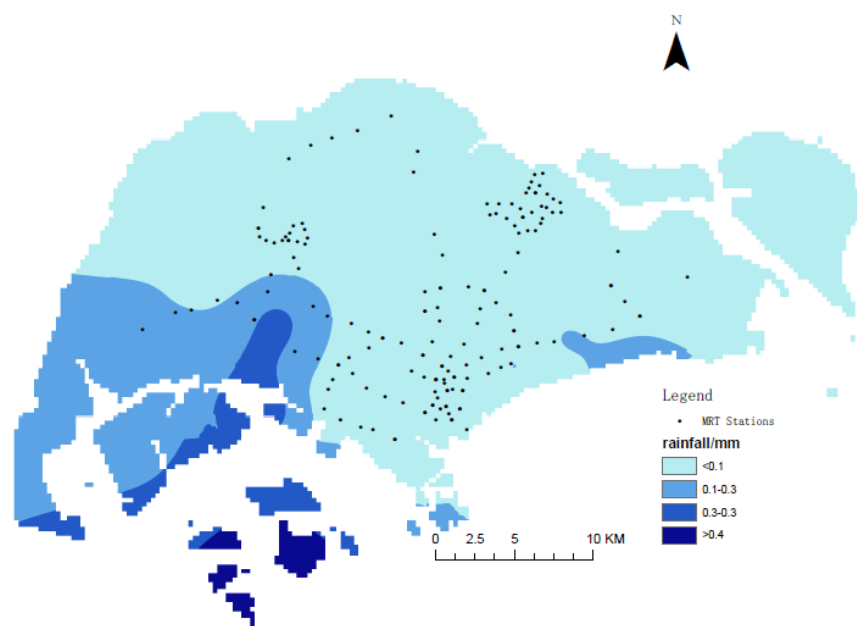


图 2.9 新加坡不同区域降雨情况（2）

由以上两图可以看出，虽然降雨量比较大，但降雨量大的区域主要分布在西南部分，主城区的主要部分日常降雨量并不大。

总结以上分析，新加坡全年的天气情况都比较相似，没有太大差异，所以一个月的数据基本上能够代表全年的情况。另外，由于新加坡的降雨比较集中，所以对于某不同区域内的单车使用，降雨的影响可能会不同，但由于降雨量大的区域主要集中在基本上无人员聚集的区域，所以剩下的大部分地区情况大致是相似的。

### 3 需求预测

在单车调度模型中，站点间的需求通常是需要预先知道的输入条件。在现实生活中，制订调度计划通常是在调度需求真正发生前，即需要先对未来可能有的需求进行预测才能利用模型进行有效求解。本节将首先对单车需求预测的几种常用方法进行总结，并根据已有数据在这些方法上的测试效果，选取最适合本文使用数据集的预测模型，为之后的数值实验提供支撑。

#### 3.1 线性回归

线性回归是一种简单的预测模型，由于其结果的高度可理解性，被广泛运用在共享单车的需求预测中。线性回归的表达式如下：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.1)$$

虽然线性回归存在每个变量对结果的影响可以通过线性回归的结果系数直接表示的优点，但是线性回归存在以下突出问题：

- （1）线性回归假设因变量服从特定自变量下的高斯分布。很多实际问题显然不符合这个假设，例如在共享单车需求预测中，共享单车的需求量恒为非负数，而服从高斯分布意味着预测结果可能为负数值。
- （2）线性回归假设自变量相互独立。在共享单车需求预测中，“天气”和“温度”是两个常被考虑的自变量，而天气和温度显然并非完全独立的变量。
- （3）线性回归假设了每个自变量和因变量间的线性关系。现实世界中的大部分问题都不是线性的。

由于以上问题都涉及到对问题定义的理解这个本质问题，所以通常线性回归在复杂问题中并不能产生令人信服的解释。

#### 3.2 广义线性回归（GLM）

正如上一节中提到的，线性回归总是假设变量与结果间的线性相关关系，而现实中的问题往往具有高度的非线性特性。为了使线性回归能够适应这一点，研究者们提出了广义线性回归（GLM）的概念。广义线性回归本质上就是为自变量与因变量间的映射关系加上一层非线性函数变换，假设这个函数为  $f$ ，广义线性回归的表达式可以表示为：

$$f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (3.2)$$

接下来要讨论的单车需求特别是 OD 对之间的需求具有一定的特征，即数据集中在较小值并且高度分散，有这种特征的数据可以被称为计数数据。计数数据存在的领域很广泛，比如事故发生次数，某种疾病的发病次数等。由于与一般的随机分散的数据有明显的特征差异，计数数据有一套相应的方法论进行分析。在道路风险评估领域，通常用泊松回归或类泊松回归解决这一类问题<sup>[29]</sup>。泊松回归也是广义线性回归的一种，它的连接函数为以自然对数为底数的对数函数，并假设因变量服从泊松分布。由于泊松回归还有均值等于方差的假设，而现实中的数据很多情况下都不满足这个条件，所以在泊松回归的基础上，还衍生出了负二项回归、零膨胀泊松回归等模型，这些模型更适用于过度分散或零数据过多的真实数据集。

### 3.3 分类算法

具有计数特征的数据除了可以被视为计数数据外，还可以当作类别数据处理<sup>[30]</sup>。对于分类数据处理常用的分类算法包括 logistic 回归、决策数分类、KNN 等，其中 logistic 回归只适用于二分类问题，而具体采用哪一种分类算法需要根据预测逻辑和数据情况选择。这些分类算法已经具有完备的理论基础，并且在分类问题研究领域已经有了广泛的运用。

还有一类分类算法是基于深度学习的分类算法，这类算法具有处理大规模数据以及高非线性拟合度的优点，是近年来广泛运用于数据处理的一种有效算法。但由于本文不仅关注预测结果，也关注结果的可解释性，而深度学习网络的参数多，可解释性较差，所以本文将不对深度学习进行讨论。

### 3.4 单车需求预测

由于具有计数数据特征，对于单车需求的预测可以尝试类泊松回归的方法；对单车需求预测的基本逻辑是寻找与需要预测的状况最相近的历史状态，即相似的时间段和环境、社会条件下可能产生的单车需求，所以也可以用 KNN 分类算法进行预测。以下将分别利用类泊松回归和 KNN 进行数据建模并进行比较。建模过程分为数据准备、数据集划分、模型训练、模型测试这几个阶段。以下先讨论数据准备过程，再分别讨论这两个方法得到的最终结果。

#### 3.4.1 数据准备

本文使用的原始数据为 2017 年十月的单车使用订单数据，包括的字段有 {车辆 ID，时间戳，起始精度，起始纬度，终止精度，终止纬度，骑行时间，骑行距离}，在需求预测阶段，对原始数据进行如下处理：

- （1）利用四分位法对骑行时间，骑行距离、速度处理异常的订单数据进行删除。
- （2）以一小时的时间间隔划分时间片，将订单中的起止时间拆分为起止日期，时间片编号这两个字段。
- （3）将起终点经纬度转化为对应土地块编号，并以起止土地块编号、出发时间为分类字段统计对应时间片内，从对应起点到终点的单车数量，次时数据字段变为 {起始 cell 编号，终止 cell 编号，日期，时间片编号，单车使用数量，起终点距离，起始土地块熵，终止土地块熵}。
- （4）在每条记录上添加起止土地块间距离、气质土地块静态信息（代表土地块利用情况的信息）字段
- （5）从新加坡开源数据网站 <https://data.gov.sg/> 上获取分析时段内新加坡各个监测站测得的温度和降雨信息，并通过克里金插值，得到每个起始土地块在对应时刻的温度和降雨情况，并入每条记录中。

经过以上处理，最终用作建模和预测的数据集包含的字段为 {起始 cell 编号，终止 cell 编号，日期，时间片编号，单车使用数量，起终点距离，起始土地块熵，终止土地块熵，温度，降雨}。由于无论是采用负二项分布还是 KNN，自变量的尺度都会对系数或距离的计算产生影响，所以在训练前还需要对自变量字段进行标准化处理。

在 2.2 节中我们已经指明，单车使用具有周期性特征，例如每个工作日间的使用特征相似，所以在划分训练集和测试集阶段，并不随机划分，而是在四周的数据中挑选前三周作为训练集，后一周作为测试集。另外，周末和工作日的单车使用具有差异性，所以在预测时，将工作日和周末分别用两个模型进行预测。

### 3.4.2 类泊松回归

处理完成的数据中单车使用数量字段的值得统计图如图3.1所示。从图中可以看出，两站点间得单车使用数量基本集中在 6 以下得较小值，且从分布上来看符合 3.2 中提到的计数数据的分布特征。这个字段的均值为 0.18，方差为 0.56，属于过度分散的数据集，所以接下来利用负二项分布模型对需求进行预测。

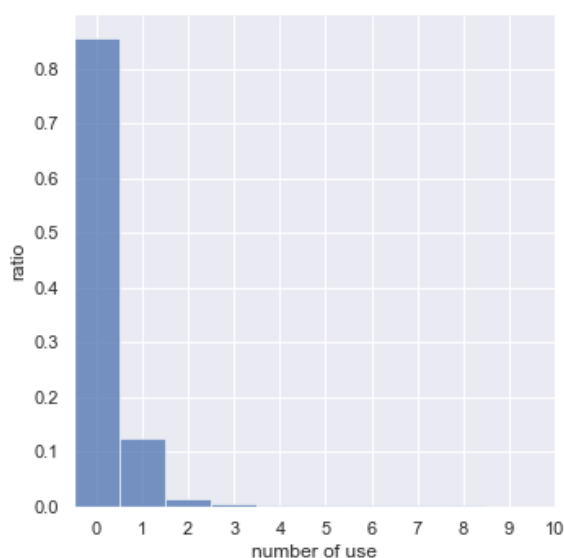


图 3.1 单车使用次数统计

负二项分布对均值和方差间关系的假设为：

$$variance = mean + \alpha \cdot mean^n \quad (3.3)$$

本文取  $n$  为 2，即采用 NB2 型负二项分布进行训练和预测。

在建模过程中，当将时间、起终点间距离、起终点土地利用情况、温度、降雨都作为自变量输入负二项分布时，由于降雨量的方差太小，模型会将自变量矩阵判断为不可逆，所以在负二项回归中删去降雨列，得到的工作日拟合结果如表3.1所示。工作日用于训练的数据共有 2676319 条，结果  $r^2$  值为 0.001。

表 3.1 工作日

variable	Coef	std err	z	p> z	0.025	0.975
const	-1.7135	0.002	-1032.607	0.000	-1.717	-1.710
hour	-0.0229	0.002	-13.089	0.000	-0.026	-0.019
distance	0.0474	0.002	29.404	0.000	0.044	0.051
entropystart	0.0013	0.002	0.806	0.420	-0.002	0.005
entropyend	-0.0073	0.002	-4.402	0.000	-0.011	-0.004
temperature	0.0712	0.002	41.338	0.000	0.068	0.075
alpha	1.8021	0.009	193.343	0.000	0.068	0.075

从表格中可以看出，温度和距离对需求的影响较大。在工作日，距离、温度、起点处土地利用熵系数为正，说明起终点间距离越大，温度越高，起点繁华程度越高，单车的使用数量越多。由于义顺区的面积比较小，直径在五公里左右，所以在这个区域内，共享单车能够骑行的最大距离也就在三只四公里，在共享单车比较适宜的骑行距离之内，所以产生需求的两土地块间距离对于义顺区的尺度来说比较大，从而使得结果体现出距离越大，需求越高的情况。

利用以上模型对 664107 条测试集数据进行测试，得到预测结果整体的准确度为 51%。

对周末数据进行和以上相同的分析后得到的结果分别如表3.2，??所示。周末模型拟合结果的  $r^2$  值为 0.02，相比工作日高。从表3.2可以看出，周末的终点熵值对需求的影响比工作日更加明显，且起终点的熵值越大，需求越大，这说明在周末消费者常活动在较繁华的区域周围。

表 3.2 周末

variable	Coef	std err	z	p> z	0.025	0.975
const	-1.8374	0.003	-625.403	0.000	-1.843	-1.832
hour	0.2621	0.003	80.209	0.000	0.256	0.268
distance	0.1035	0.003	39.120	0.000	0.098	0.109
entropystart	0.0216	0.003	7.605	0.000	0.016	0.027
entropyend	0.0176	0.003	6.225	0.000	0.012	-0.023
temperature	0.2579	0.003	95.537	0.000	0.253	0.263
alpha	1.5884	0.015	105.146	0.000	1.559	1.618

在测试集上，周末预测的整体准确度为 29%，明显低于工作日，这说明周末需求的产生较工作日随机。

### 3.4.3 KNN

由于共享单车使用规律存在周期性特征，所以对共享单车进行需求预测的逻辑为期望通过寻找相似场景下的历史需求，得到当前场景需求的可能值，这与 KNN 算法的逻辑是一致的，KNN 的预测方法即为找到与待预测点值相近的训练集中的点，利用这些点的情况对待分类点进行预测。

利用 KNN 得到的在工作日和周末的测试集上的预测准确度分别为 87%，82%。

### 3.4.4 方法比较

本节将从运行时间、效果和结果的可解释性三个方面对用负二项回归和 KNN 分类对需求进行预测的方法进行比较。

从运行时间上看，在针对较大训练数据集时，KNN 相比于负二项回归具有较快的训练速度，但是其预测速度比负二项回归慢。由于在现实生活中，历史数据集的量往往很大，较快的训练速度可以达到在有限时间内更充分的利用这些数据的目的；在调度问题中需要预测的时间段往往比较固定，可能是一天、一周或一个月，数据量相对较小，所以预测速度稍慢并不会产生太大影响。

从准确度上看，利用 KNN 进行预测的准确度高于负二项回归。

从结果的可解属性来看，直观上看，上利用负二项分布可以直接得到每个变量的系数，从而



可以直接对每个变量对结果的影响进行解释和理解，而利用 KNN 构建的模型难以做到这一点。

机器学习算法的可解释性之所以重要，是因为人们除了希望获得具有高准确度的模型之外，往往还期望能够通过这个模型对问题的本质进行探究和理解，这种理解有助于模型参数的调优，以及在模型无法很好的解释现象时，对可能的原因进行分析。

虽然机器学习的可解释性可以带来诸多好处，对于大部分高效的机器学习算法而言，想要对其参数进行解释并不容易。例如，在神经网络中有成百上千的参数，由于在神经网络计算过程中的高度非线性性质，我们很难从训练好的神经网络中通过每个参数的值计算对应的自变量的系数，所以很难像诸如线性回归这一类的算法一样直接通过系数值的大小对问题进行解释。

为了解决这个问题，一些学者提出了可解释机器学习的概念。可解释机器学习的方法很多样，Molnar 在他的书中对常用的方法进行了总结和解释<sup>[31]</sup>。从基本原理上看，可解释机器学习就是试图拆分每个参数对结果的影响，从而衡量这个参数在此问题中所起的作用。接下来本文将采用 Shapely 值法，对 KNN 模型进行解释。

Shapely 值计算的基本思想为，通过改变一个变量的值，保持其它变量在平均值，计算这个变量改变时结果的改变程度，这个改变程度即代表了这个变量对结果的影响。例如在我们的单车需求问题预测中，假设预测的所有需求的平均值为 2，我们需要解释的是，当时间、距离、起终点土地利用情况、温度和降雨共同预测出一个为 3 的需求时，每个自变量对 3 和 2 间的差异，即 1，贡献了多少。以时间为例，当计算时间的 Shapely 值时，首先我们给定一个时间值，从数据集中随机选取除时间以外的其它自变量的值，产生一条新的数据，再利用模型对这条数据进行预测得到需求值；接下来，再从数据集中随机选取时间值代替一开始选定的时间值，再次用模型预测需求。这两次需求的差值即为时间这个变量对结果的贡献量。为了提高计算准确度，重复以上第二步中的采样过程，并对所有过程计算平均值作为最终的 Shapely 值。

从 Shapely 值的计算原理中可知，不仅可以计算单个实例的 Shapely 值，还可以计算模型整体的 Shapely 值，这两种 Shapely 值对应的就是可解释机器学习中的局部解释和全局解释。局部解释侧重于分析在单个实例中，各个自变量是如何共同作用产生预测结果的，全局解释则可以帮助人们理解所研究的问题的整体逻辑结构<sup>[32]</sup>。

对于工作日 KNN 模型计算的两个实例的 Shapely 值如图 3.2、3.3 所示。从两图中可以看出，对于不同的实例，不同的自变量对结果的影响不同。不过，对于单车需求预测而言，我们并不关系单个预测是如何产生的，更关心的是每个自变量对需求产生的整体影响。

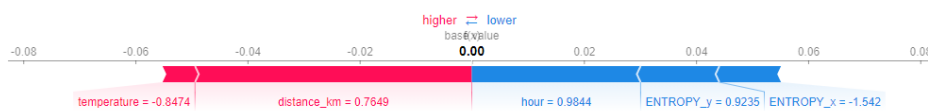


图 3.2 实例 1



图 3.3 实例 2

整体 Shapely 值如下图 3.4 所示。从图中可以直接看出，距离、时间对预测结果的影响比较大，温度和起终点土地利用情况对预测结果的影响较小，这与之前用负二项回归训练结果中参数的解释是一致的。

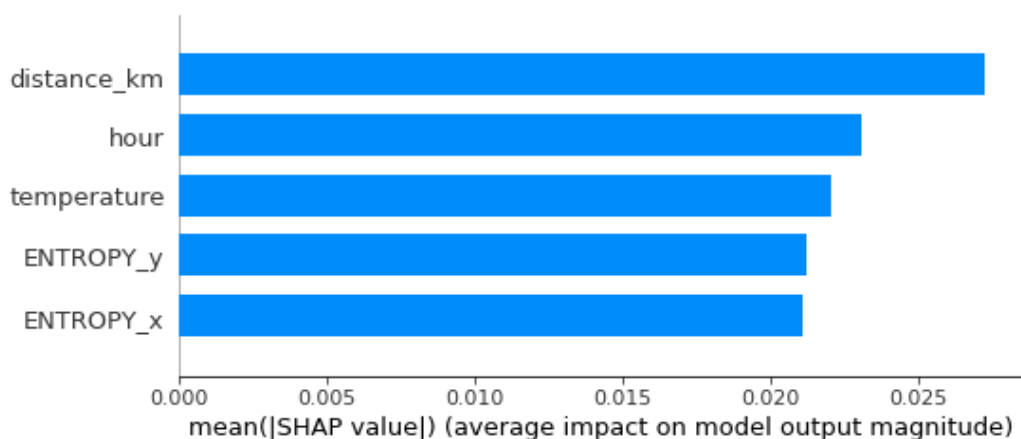


图 3.4 工作日 Shapely 值

下图3.5是对类似图3.2、图3.3的每个实例的汇总。从图中的分布可以看出，站点间距对预测结果的影响整体为正值，即站点间距使得模型对需求的预测增加；时间对预测结果的影响更偏向负值，这与在利用负二项分布拟合时对结果的分析是相似的。

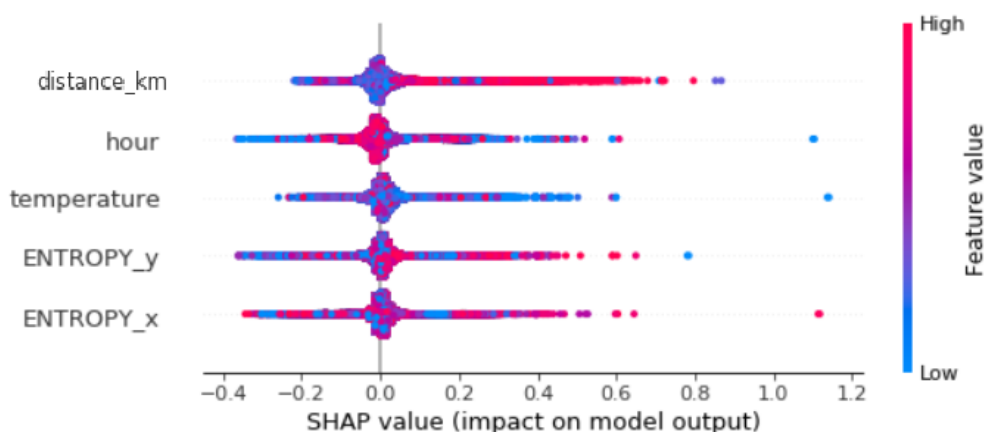


图 3.5 工作日 Shapely 值分布

相同的，对周末的 KNN 模型计算 Shapely 值，结果如下图3.6、3.7所示。图中显示距离和时间仍然是影响需求预测最大的两个因素，且影响多集中在正值区域。

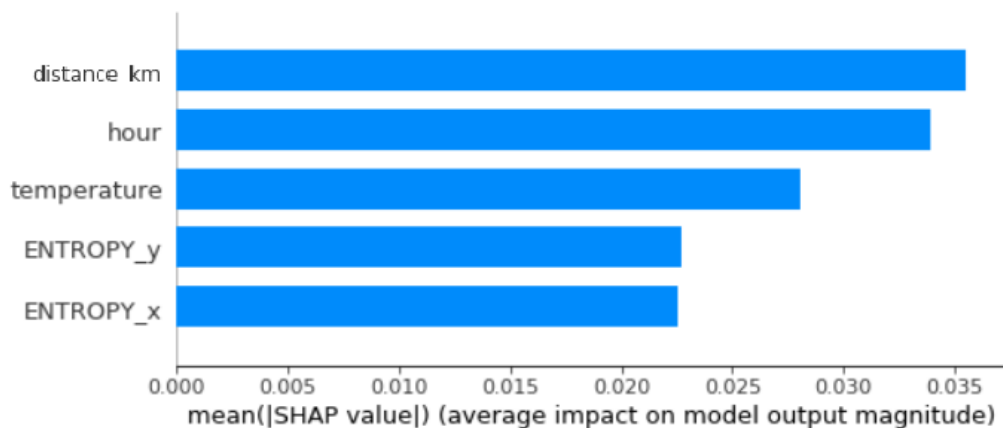


图 3.6 周末 Shapely 值

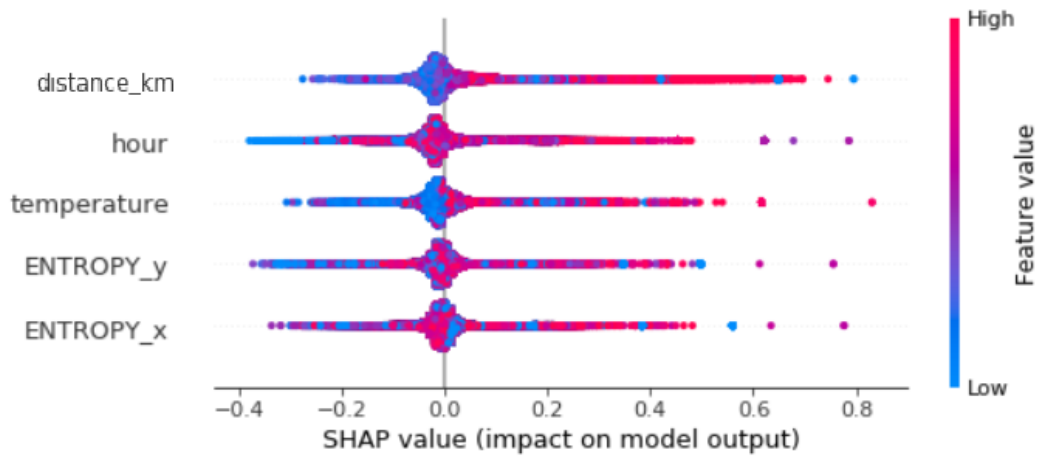


图 3.7 周末 Shapely 值分布

总结以上对比，可以发现虽然统计学习具有更高的直接可解释性和更快的预测速度，但相比于机器学习，其训练时长更长，训练效果也更差一些，这是由其训练原理导致的，只能对问题中的非线性关系进行有限的拟合。而且随着可解释机器学习的发展，机器学习模型的可解释性也会越来越高，从而弥补其对帮助人们理解问题上的欠缺。通过本章的分析，总体而言，对于需求预测这样比较复杂的问题，利用机器学习模型会比统计学习模型更加适合。具体的，在本文讨论的方法中，利用 KNN 比负二项回归更加适合。

## 4 调度模型建立

### 4.1 问题阐述

在之后的文章中，我们将在动态的共享单车调度系统中利用时空网络模型模拟单车的使用和停放、卡车的调度行为的过程并分析他们对于系统运行结果的影响。由于调度卡车的数量有限，而共享单车的数量往往远大于调度车辆的数量，如果将单车的数量视为离散变量会大大增加模型的求解难度，所以在模型构建过程中，将每辆调度卡车视为独立的个体，即代表卡车数量的变量为整数变量，而将单车视为连续流，即代表单车数量的变量为连续变量。最终得到的模型是一个混合整数优化问题。为了更好的理解建模过程，本节先对建模过程中需要用到的一些符号的含义进行解释说明，具体如表4.1所示。

表 4.1 符号表示

变量类型	符号	释义
集合	$N$	站点（土地块）集合
集合	$A$	两站点间最近连线集合
集合	$T$	时间切片
参数	$c_{ij}^v$	从站点 $i$ 运送一辆单车到站点 $j$ 的成本，包括在 $i$ 装载单车和在 $j$ 卸载单车，统称为运送成本
参数	$c_{ij}^u$	从 $i$ 到 $j$ 的过程中一辆单车的磨损成本
参数	$e_{ij}^d$	消费者将单车从 $i$ 骑到 $j$ 需要的费用
参数	$a$	调度卡车容量
参数	$n$	系统中的单车总数
参数	$r_{ij}^t$	在时间片 $t$ 期望将单车从 $i$ 骑到 $j$ 的消费者数量
参数	$P$	其它固定成本，例如购买单车成本
连续变量	$b_{ik}^{t,t_k}$	在 $[t, t_k]$ 被从 $i$ 运送到 $j$ 的单车数量
连续变量	$b_i^t$	时间片 $t$ 时停在 $i$ 的单车数量
连续变量	$d_{ij}^{t,t_j}$	在 $[t, t_j]$ 被消费者从 $i$ 骑到 $j$ 的单车数量
离散变量	$w_{ik}^{t,t_k}$	在 $[t, t_k]$ 从 $i$ 到 $j$ 的卡车数量
离散变量	$u_i^t$	时间片 $t$ 时停在 $i$ 的卡车数量

上表中， $n$  代表系统中调度卡车数量，每辆卡车的运载容量为  $a$ ，即一辆卡车一次最多只能运送  $a$  辆单车。在空间维度上，由于在无桩共享单车系统中并没有固定的车辆停放点，为了表示车辆停放位置，我们需要先将整个共享单车服务区域划分为一些更小的分区，这些分区的划分应以共享单车消费者常用的停车区域为依据，并且能被视为没有停放数量的限制。以下过程中将这些分区称为土地块或站点。 $N$  代表研究区域中划分的土地块的数量。在时间维度上，需要将连续的时间变量分割为从 1 到  $T$  的时间片。 $b_i^t$  代表在时间片  $t$  停放在土地块  $i$  的单车数量， $u_i^t$  代表在

时间片  $t$  在土地块  $i$  等候的调度卡车数量。在定义的这些基础变量的基础上，接下来定义与单车和卡车在不同时间片内移动有关的决策变量如下：

- (1) 在时间片  $[t, t_k]$  内被调度卡车从土地块  $i$  运送到  $k$  的单车数量  $b_{ik}^{t, t_k}$
- (2) 在时间片  $[t, t_k]$  内被用户从土地块  $i$  骑到  $k$  的单车数量  $d_{ij}^{t, t_j}$
- (3) 在时间片  $[t, t_k]$  内被从土地块  $i$  运送单车到  $k$  的调度卡车数量  $w_{ik}^{t, t_k}$
- (4) 在时间片  $[t, t_k]$  内被从土地块  $i$  空载到  $k$  的调度卡车数量  $u_{ik}^{t, t_k}$

$t$  和  $t_k$ ,  $t$  和  $t_j$  间的差值由对应 OD 对间的距离和单车或卡车的速度决定。在模型中，我们还假设调度卡车在站点进行单车搬运操作的时间为一个常数，即这个时间不会随需要搬运的单车的数量变化。虽然这个假设是为了简化模型，但也同时是这个模型的一个局限，不过想要解决这个问题，可以在模型中添加一个表示操作时间的变量。

为了更好的阐述模型的构建逻辑，以下先通过一个小型网络中的例子来描述本文对调度过程的刻画逻辑。图4.1展示了一个由三个站点和三个时间切片组成的小型时空网络，图中只展示了站点1在时间片1的单车与卡车流流入流出情况，其它站点同理。流入和流出站点1的单车流由两部分组成：卡车调度过程中的运送和消费者的使用，图中黄色表示单车流，蓝色表示卡车流， $b_{21}^{0,1}$ 、 $b_{31}^{0,1}$  指被调度卡车从站点2、3调度到站点1的单车数， $d_{21}^{0,1}$  指被消费者从站点2骑到站点1的单车数，从站点1流出的单车数量同理。由于站点1在时刻1、时刻2的单车数为  $b_1^1$ 、 $b_1^2$ ，相邻两个时间切片间同一站点的单车数量关系就可以由这两部分的流量完全表示，即：

$$b_1^2 = b_1^1 + b_{21}^{0,1} + b_{31}^{0,1} + d_{21}^{0,1} - b_{12}^{1,2} - b_{13}^{1,2} - d_{12}^{1,2} - d_{13}^{1,2} \quad (4.1)$$

对网络中的所有时间片和站点都可以列出以上表达式，此表达式即为网络流模型的核心表达式。还需要特别提出的是，在图4.1表示的例子中，从站点1到3的卡车和单车流耗时不同，因为卡车和单车的速度不同。而且在考虑两站点间的运行时间时，我们只关注站点间的距离，而不关注具体的路线，这意味着无论图中的站点2是否位于站点1和3之间都不会影响模型的最终结果。最后，模型中所有的时间都以时间切片为单位，不足一个时间片的按一个时间切片计。

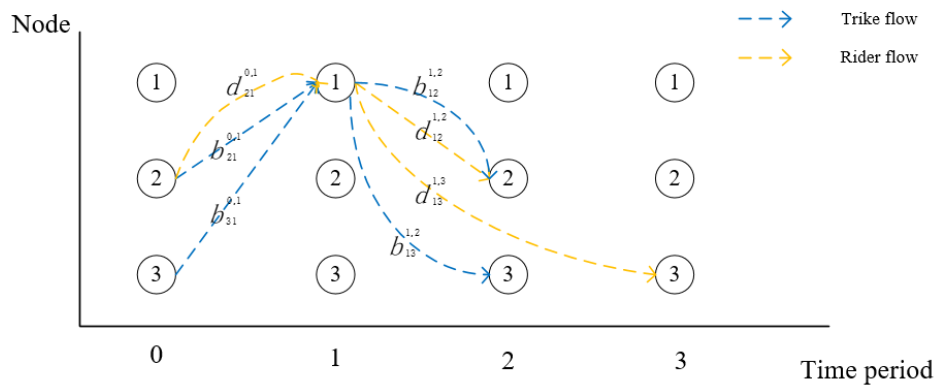


图 4.1 小型时空网络示例  
(站点 1 在时间片 1 的流入流出)

## 4.2 问题抽象与建模

本节将讨论将实际调度问题抽象成可解的数学模型的过程，我们将提出一个混合整数线性规划模型用于描述和解决单车调度问题。建立这个模型的目的是为了动态的为卡车调度计划的制订提供理论指导，以使得共享单车网络在现有规模下能获得可能的最优收益。模型的目标收益仅指消费者使用共享单车带来的收益与调度成本（包括搬运、运输单车过程中的磨损成本）和卡车调

度的成本（燃油）间的差值。其它成本，如车辆贬值、卡车司机工资等成本受诸多其它社会因素的影响，更适合在讨论网络长期稳定运营的问题中考虑，所以本模型中并未将这些成本纳入模型考量范围，只重点关注如何利用有限数量的卡车和共享单车得到最大的现实收益，最后得到的结果是在每个时间片内，应在网络中执行什么样的调度操作，具体到某时间卡车应从何处将多少单车调度到何处。

为了尽可能的降低模型构建过程中一些过于复杂且难以量化的现实情况造成的建模困难，使模型尽可能只关注共享单车调度过程中那些影响调度行为的最本质的问题，在模型构建前，我们需要先做出的假设和简化。对这些假设、简化以及模型使用的注意事项总结如下：

- （1）模型构建前，需要先将整个研究时间段需要被切分为具有相等间隔的时间切片。
- （2）每个 OD 对间在每个时间切片内的需求都是需要预先预测的，即需求为已知量。
- （3）每个消费者都可以直接从他们的期望骑行出发地到达目的地，不需要绕行到其它位置，即他们的骑行时间是可以计算的。
- （4）尽管单车和卡车的速度受实时交通状况的影响，但是在模型中，将两者的速度视为可知的定值。
- （5）单车骑行依据使用时间收费，这个费用在确定起终点和速度后就可以直接估计出来。
- （6）共享单车流为连续流，调度卡车流为离散值。
- （7）如果消费者在他们的期望始发点没有找到可用单车，则他们应直接离开。换言之，消费者不会在原地等待直至有可用车辆。
- （8）单车被搬运上卡车和被从卡车搬运下来的时间忽略不计。如果有考虑这部分时间的需求，在模型中加上代表这一部分时间的变量也不会明显改变模型的复杂性。
- （9）每个站点的停车容量不受限制。

在以上简化和假设的基础上，构建出的含有时空网络结构的混合整数线性规划模型如下所示：

$$\text{maximize} \quad \sum_t \sum_i \sum_j e_{ij}^d d_{ij}^{t,t_j} - \sum_t \sum_i \sum_k c_{ik}^u w_{ik}^{t,t_k} - \sum_t \sum_i \sum_k c_{ik}^v b_{ik}^{t,t_k} - nP \quad (6.1)$$

$$\text{subject to} \quad b_i^{t+1} = b_i^t - \sum_k b_{ik}^{t,t_k} - \sum_j d_{ij}^{t,t_j} + \sum_k b_{ki}^{t,t_k} + \sum_j d_{ji}^{t,t}, \quad \forall i \in N, t \in T, \quad (6.2)$$

$$u_i^{t+1} = u_i^t - \sum_k w_{ik}^{t,t_k} + \sum_k w_{ki}^{t,t_k}, \quad \forall i \in N, t \in T, \quad (6.3)$$

$$b_{ik}^{t,t_k} \leq a w_{ik}^{t,t_k}, \quad \forall i, k \in A, t \in T, \quad (6.4)$$

$$d_{ij}^{t,t_j} \leq r_{ij}^{t,t_j}, \quad \forall i, j \in A, t \in T, \quad (6.5)$$

$$b_i^t \geq 0, \quad \forall i \in N, t \in T, \quad (6.6)$$

$$u_i^t \in \{0, 1, 2, \dots\}, \quad \forall i \in N, t \in T, \quad (6.7)$$

$$b_{ij}^{t,t_j} \geq 0, \quad \forall i, j \in A, t \in T, \quad (6.8)$$

$$d_{ij}^{t,t_j} \geq 0, \quad \forall i, j \in A, t \in T, \quad (6.9)$$

$$w_{ik}^{t,t_k} \in \{0, 1, 2, \dots\}, \quad \forall i, k \in A, t \in T \quad (6.10)$$



目标函数 (6.1) 最大化系统收益，收益通过骑行总收益减单车调度成本和卡车调度成本求得；约束 (6.2) 和 (6.3) 分别为在每个站点的单车和卡车数量守恒约束，其构建逻辑已在图4.1给出的例子中说明，单车和卡车的逻辑相同，只是缺少了由消费者使用带来的流量变化这一部分；约束 (6.4) 保证了由卡车调度的单车总数总是不超过卡车的最大容量；约束 (6.5) 表示需求满足限度约束，即某站点在某时间切片的单车使用量不能超过该时间切片内该站点存有的单车数量；约束 (6.6) 和 (6.7) 保证了所有单车数量和卡车数量都是非负的；约束 (6.8), (6.9) 和 (6.10) 保证了单车流、卡车流的非负性以及卡车流受整个系统中卡车总量的约束。

将前述处理后得到的需求数据作为以上模型的输入，确定参数值后通过 Gurobi 9.0.2<sup>[33]</sup> 在一台配有 Intel215 Core i7-9750H 处理器的电脑上进行求解，可以得到结果。关于实例的构建和结果的展示将在下一章中详细讨论。

装

订

线

## 5 结果分析

### 5.1 研究区域选取

新加坡在地理上被划分为五十个选区，如下图5.1所示。其中，位于北部有一个叫义顺（YISHUN）的选区。义顺是一个人口密集的区域，对于共享单车的使用具有较高的需求，这也使这个区域内共享单车调度成为必要。由于共享单车的高使用量，这个选区内有大量的历史使用数据可供使用。

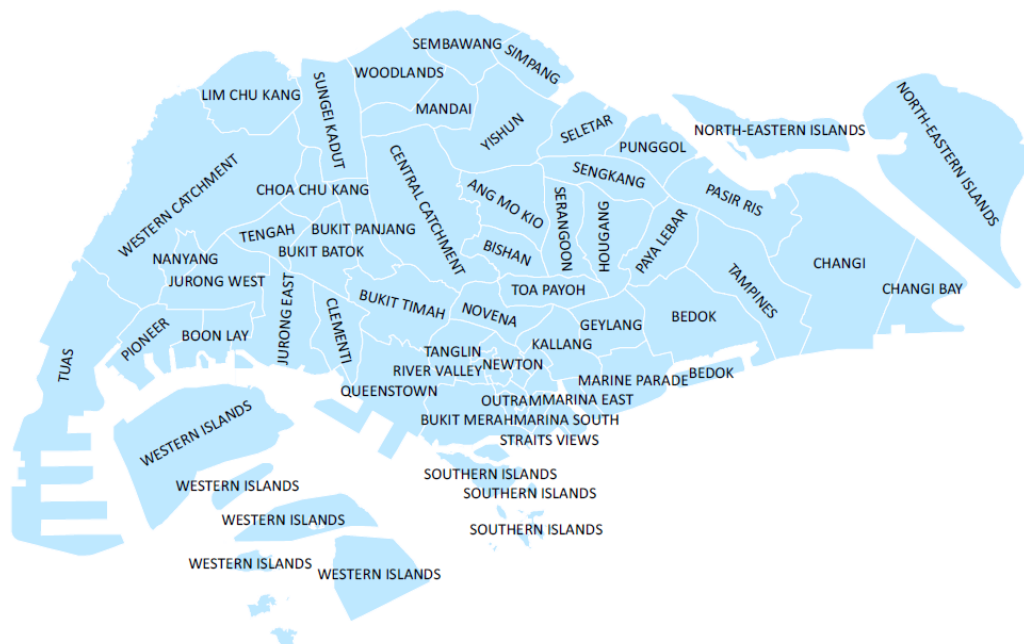


图 5.1 新加坡选区划分

如果在整个新加坡的尺度上进行调度问题建模，那么问题规模将会变得巨大。但事实上，在每个城市内部，经常使用共享单车的区域可以划分为不同的分区，需求常常在这些分区内部产生。对于新加坡，在 Shen<sup>[34]</sup> 的研究中，研究者们已经证明了新加坡大部分的单车需求都发生在各个选区内部，这就使得新加坡在空间上天然地形成了不同的共享单车使用分区。由于各个分区内共享单车使用的独立性，每一个分区可以看作一个更小的相互独立的研究区域，从而针对义顺区内的单车使用数据就可以为我们提供足够的信息用于分析共享单车调度细节和不同因素对收益的影响规律。由于前面已经提到，义顺区的历史数据多，单车使用量大，因此，在接下来的数值实例中，我们选取义顺作为研究区域。义顺的大致位置如图5.2所示。

前面已经提到，为了使用模型计算调度结果，需要先将研究区域进行进一步划分。在接下来的实例中，我们将义顺区划分为 300m 乘 300m 的网格，每一格代表一个分区，即模型中的站点；在对时间片的切分上，以五分钟为间隔切割时间片。做好这些准备后，就可以利用数据开始计算数值实例了。

装

线

订

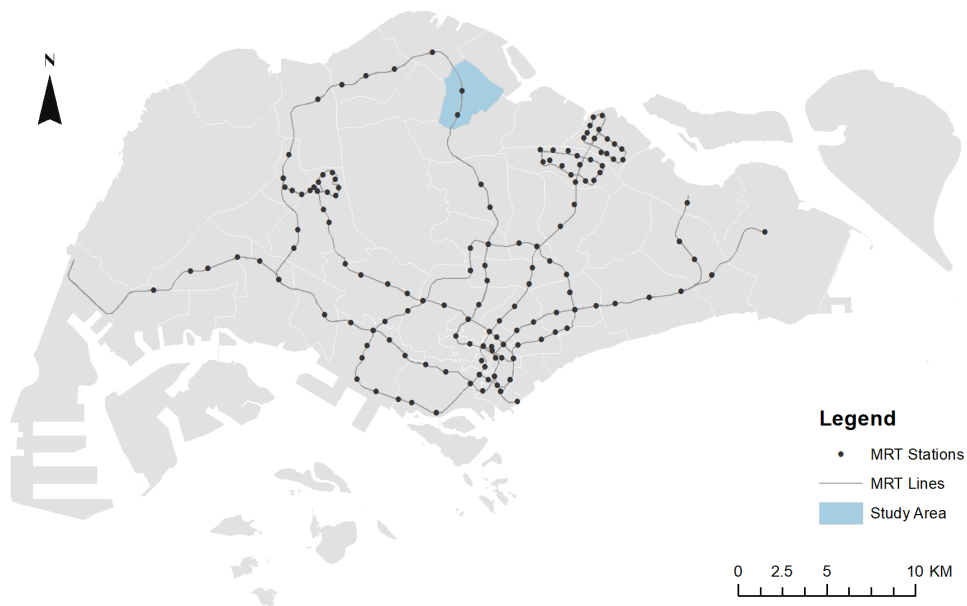


图 5.2 研究区域选取

5.2 实例和结果

本节将展示如何利用已经构建的网络流模型解决现实世界的问题，并对求解得到结果进行进一步分析。首先，我们先在一个选取了十个站点、三十六个时间切片的小时空网络中利用模型进行求解，以方便对结果进行可视化，从而分析结果的合理性；接着，我们将实例扩大到包含更多站点，并对卡车和单车数量，调度成本，价格进行敏感性分析，这些分析将有助于我们建立对这些因素可能对收益造成的影响的更清晰认识。

在第一部的小型网络实例中，我们挑选了工作日早高峰七点至十点的需求数据用作模型输入。假设在这个时间段内的任何时间切片中都可以进行调度。因为时间以五分钟进行切片，每个站点的单车需求也以五分钟为时间段进行计数。虽然由于可能有未被满足的需求的存在，即消费者到达期望出发点时没有找到可用单车，所以从单车使用数据中得到的单车需求量并不严格等于真实需求，但是由于从开始的单车分布结果中可以看出，义顺区内的大部分区域单车存在过饱和，所以需求大于供给的情况可以忽略不计，所以在进行输入预测时仍将从数据中统计得到的使用量视为需求量。在选取十个站点时，选取三个小时内使用量最高的十个站点作为研究范围，以减少随机需求的影响。

在敏感性分析中利用到的所有参数的值选取情况如表5.1所示。

表 5.1 敏感性分析参数设置

参数	值
调度卡车数量	0,1,2,3,4,5
单车数量	225,488,602,899
燃油成本 (\$\$)	0.4,0.6,0.8,1
定价 (\$\$)	1,1.5,2,2.5,3

在进行敏感性分析的过程中，每次计算只改变一个参数的值，同时保持其它所有参数的值为默认值。卡车数量、单车数量、成本和定价的默认值分别为 2, 899, 0.5, 1。其它未提及的参数值设置均与第一步中包含十个站点的实例中设置相同。

消费者骑行带来的收益是骑行时间的函数，这个函数中的单位价格是从新加坡目前真实的骑行定价中得到的。站点间的距离是根据站点的路网距离计算出来的。单车调度成本与每辆单车的调度都有关，具体包括单车调度过程中的磨损以及调度此辆单车需要耗费的劳动力。卡车调度成本主要指卡车油耗成本。我们还从义顺的数据中得到卡车的初始分布作为模型中初始站点停放车辆的值。

### 5.3 基本实例：调度行为分析

本节对含有十个站点的小型网络计算结果进行分析，之后的文章中我们都将称这个实例为基本实例。本节中，我们还将利用时空调度图将利用模型计算得到的调度计划进行可视化。

基本实例求解中使用的参数值如下：

- (1) 收益为 S\$1 每 15 分钟，不满 15 分钟按 15 分钟计。
- (2) 卡车调度成本为 S\$0.4/km。
- (3) 单车调度成本为 S\$0.5 每辆每次。
- (4) 每辆卡车的容量为 20。
- (5) 系统中调度卡车数量为 3。
- (6) 每辆卡车的购买成本均摊到每天为 S\$70（依据卡车的平均使用年限计算）。

在基本实例中，为了防止问题过于复杂难以可视化，从义顺挑选十个站点、设置调度卡车数量为三辆进行计算。这十个站点分布在义顺的两个地铁站、工业园区和社区中心附近，这些区域都是有较高共享单车需求的区域（见图5.3），图中各自颜色的深浅代表每个站点使用的频繁程度。其中，义顺工业园和社区中心是义顺区的两个重要人口聚集地。

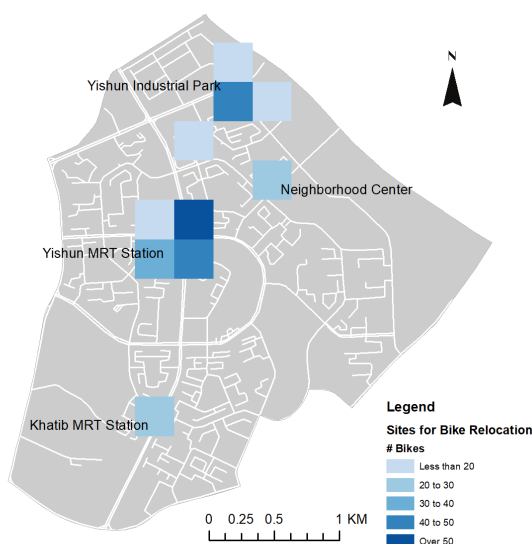


图 5.3 研究区域选取

下图5.4中只展示了开始十二个时间切片内的计算结果。图中 Yishun-1 到 Yishun-4 表示在义顺站周围的四个站点，Industrial Park-1 至 Industrial Park-4 表示在义顺工业中心附近的四个站点。每个圆圈的含义解释如图5.5所示。圆圈内的数字表示该时刻停在该站点的单车数量，圆圈上方的

数组表示该时刻流入和流出该站点的单车数量。每个圆圈都含有时间和站点位置两个固定属性，在接下来的分析中，将每个圆圈称作一个“状态”。

如图5.4所示，消费者使用单车存在明显的聚集现象，例如，在时间切片 7 和站点 Yishun-2，被用户骑走的单车数就很大，而在其它状态下骑走的数量就相对随机。除此之外，由于社区中心离义顺工业园和义顺地铁站都在步行可达距离内，住在社区中心的人不太需要利用共享单车解决出行的第一公里和最后一公里问题，这也是在这些站点间的单车流入和流出量都很小的原因。单车调度总结起来主要发生在义顺工业园、社区中心、Khatib 和 Yishun 两个地铁站之间。值得注意的是，调度在义顺工业园和义顺地铁站间发生的最为频繁，这是因为义顺地铁站是距离义顺工业园最近的地铁站，需要在早晚高峰通过搭乘地铁到达义顺工业园上下班的人需要频繁的来往于这两个区域之间。在早高峰，人们从义顺地铁站下地铁后需要从地铁站骑车到工业园；在晚高峰，从工业园下班的人需要骑车到义顺地铁站搭乘地铁返回。由于这两个站点间的大量共享单车需求，这两个站点间的调度也相对频繁。相比于义顺地铁站，Khatib 地铁站距离社区中心和工业园都更远一些，所以从图中可以看到 Khatib 地铁站周围的单车使用量较少，执行的调度操作也就比较少。

对调度结果进行进一步分析可以看出，调度主要发生在早上七点至七点三十分，从义顺地铁站到义顺工业园，以及早上七点至八点，在义顺地铁站周围区域。从七点至七点半的调度高峰的出现是由于前述的早高峰时段从地铁站至工业园的上班需求，可以理解为从七点半到八点，从义顺地铁站出发的单车需求比较大，所以需要提前做好准备好足够的单车使用。从七点半至八点，由于很多消费者骑车到达 Yishun-1，而从 Yishun-2 和 Yishun-4 出发的需求比较大，所以需要在义顺地铁站周围的四个区域内进行单车调度以满足这些需求。

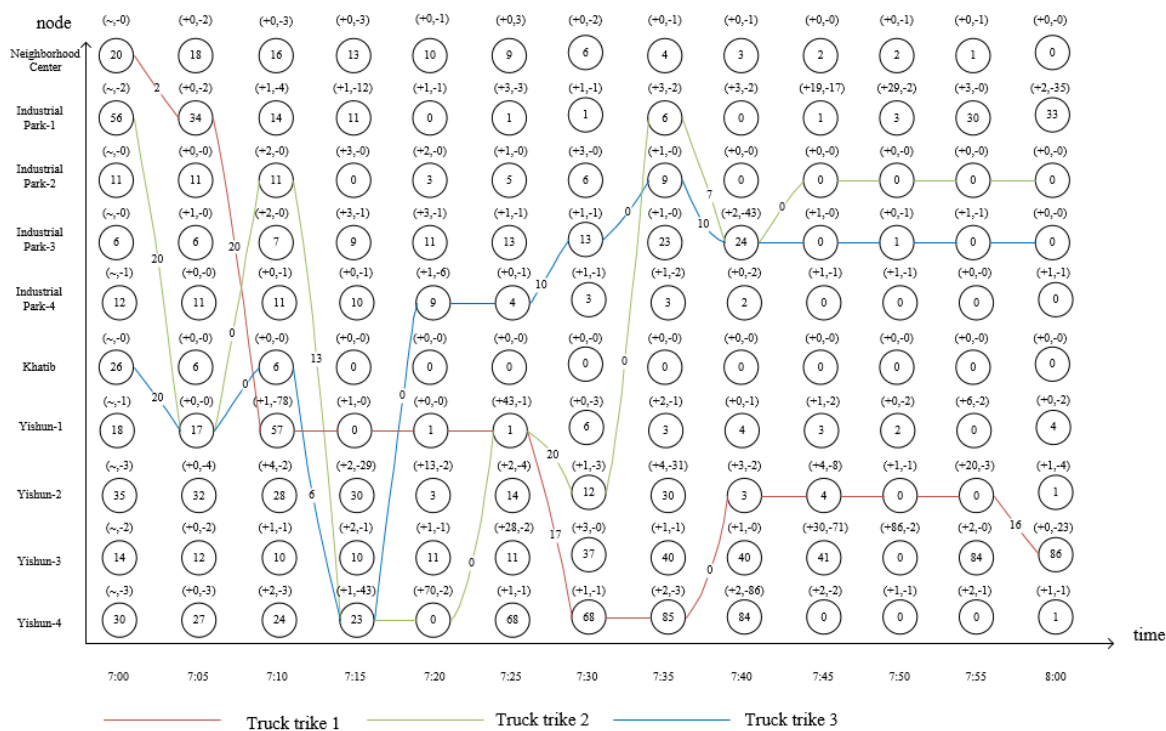


图 5.4 求解结果可视化

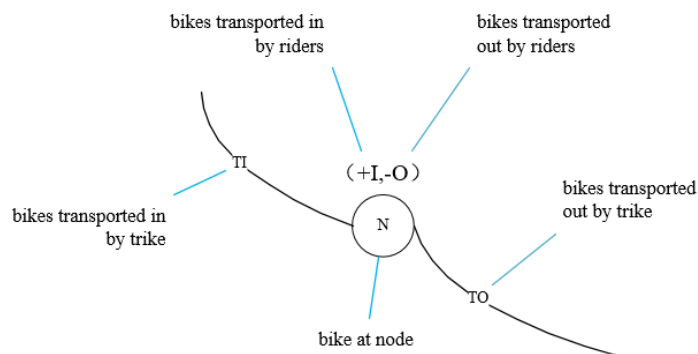


图 5.5 元素含义解释

总体来说，图5.4中展示的调度操作是符合现实情况的，并不存在异常的调度行为，如从需求量的区域调度车辆到需求量小的区域或者在两个站点间频繁来回调度车辆。通过以上基本实例的求解和分析可以得知，通过建立的模型得到的调度计划是合理并且可执行的。

### 5.3.1 敏感性分析

本节主要讨论以下因素对收益的影响的敏感性分析：

- (1) 调度卡车数量和系统中单车数量
- (2) 单车数量
- (3) 运送成本
- (4) 定价

在成本中，我们并没有讨论单车磨损成本这一部分变化对收益的影响，因为单车在运送过程中的磨损程度相对固定，不受明确的因素影响，而运送成本收交通状况影响很大，是一个常变化的参数，所以对于成本，我们只讨论运送成本变化对结果的影响。

进行敏感性分析的顺序组织如下：

- (1) 分析单车网络规模，即调度卡车和单车数量对收益的影响，以为如何在不同的情况下合理的决定单车和卡车数量提供一些理论支持。卡车数量取值范围为 0-5，单车数量以从数据中得到的 899 辆为基准值，由于 899 存在过饱和，所以取原数量的 2/3、1/2、1/4 计算。
- (2) 对成本和定价这两个直接影响收益的核心因素核心因素对最终收益的影响进行敏感性分析。正确理解单车数量、调度卡车数量、成本和定价对收益的影响是决定运营的共享单车网络是否能稳定维持的关键。

敏感性分析中涉及到的参数值取值已总结在表5.2中给出。表中每一行代表一次计算。经过测试，发现在计算时间显示为 30 分钟时，基本所有实例的结果误差都在 1% 以内，为了体现模型计算的时效性，对所有实例都设置计算时间限制 30 分钟。

进行所有实例的计算时，除了对收益进行敏感性分析，本文同样对消费者满意度进行了敏感性分析。收益代表的是网络的短期盈利能力，想要长期的稳定占有市场，消费者满意度是不可以忽略的。消费者满意度的定义为：

$$\text{满意度} = \frac{\text{满足的需求数量}}{\text{总需求}}$$

其中，满足的需求数量是指在期望出发点找到可用单车的消费者数量。



表 5.2 研究场景

	卡车数量	单车数量	燃油成本 (S\$/km)	定价 (S\$)
对卡车数量 敏感性分析	0	899	0.5	1
	1	899	0.5	1
	2	899	0.5	1
	3	899	0.5	1
	4	899	0.5	1
	5	899	0.5	1
对单车数量 敏感性分析	2	225	0.5	1
	2	488	0.5	1
	2	602	0.5	1
	2	899	0.5	1
对成本 敏感性分析	2	899	0.4	1
	2	899	0.6	1
	2	899	0.8	1
	2	899	1	1
对定价 敏感性分析	2	899	0.5	1
	2	899	0.5	1.5
	2	899	0.5	2
	2	899	0.5	2.5
	2	899	0.5	3

在基本实例中，为了限制网络规模，我们只选取了需求总量排名前十的站点作为研究区域。作为基本实例的延申，在进行敏感性分析时，我们在义顺区选择了 23 个站点进行计算，这 23 个站点是所有站点中使用最为频繁的前 23 个站点。这些站点的位置如图5.6所示。如图中所示，这些位置仍然集中在义顺区的地铁站、居民区和工业园附近。在接下来的敏感性分析中，这 23 个站点将作为我们关注的主要区域，站点的单车使用数据也是模型输入的基础。

5.3.2 网络规模敏感性分析

共享单车网络的管理需求可以分为以下两两个场景：

- （1）在固定网络规模（单车数量）的情况下，共享单车网络运营商期望通过尽可能少的调度实现利润最大化。这种场景适用的情况为某运营商已经占据了一定共享单车的市场，期望获得更多利润。
- （2）共享单车运营商希望在一片新的区域建立共享单车网络，期望尽可能少的投放单车实现

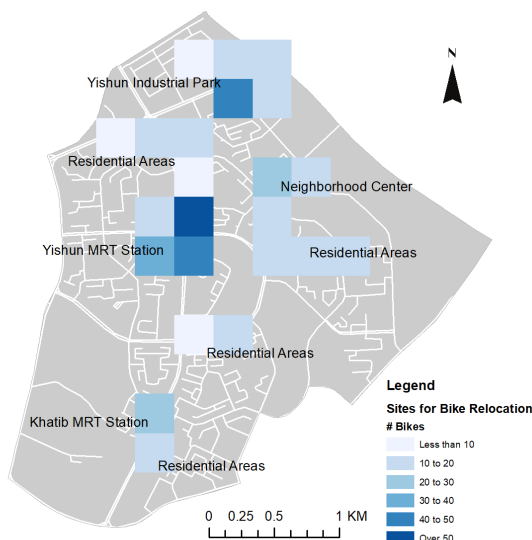
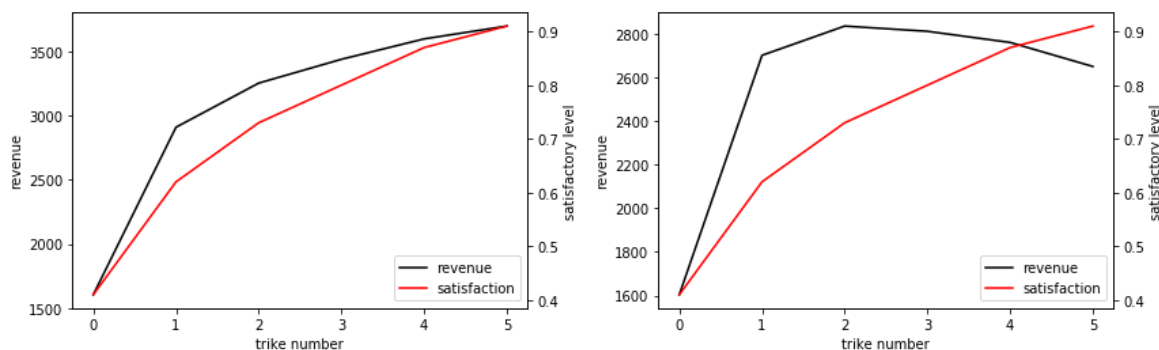


图 5.6 新研究区域选取

利润最大化。这种场景适用的情况为运营商希望构建一个新的共享单车网络，需要一定的理论指导。

接下来我们将在这两个场景中进行敏感性分析。

在第一个场景下，我们进行在固定单车数量时，卡车数量对于利润的敏感性分析。利用义顺的实际数据计算得到的结果如图5.7所示。图中每个坐标点代表一个实例。如图5.7a所示，当不考虑购买卡车的成本时，收益随着卡车数量增加持续上升，但上升速度逐渐减慢最终趋于稳定；在图5.7b中，收益和乘客满意度先随卡车数量的增加而上升，当卡车数量为 2 时达到峰值，之后收益随卡车数量的上升而下降。图5.7a中的峰值点之所以存在是因为卡车数量增多带来的调度能力增加可以部分缓解某些站点单车数量不足的情况，使得网络能够满足更高的单车需求量，但同时，购买更多的卡车意味着更高的运营成本，所以一味地通过购买卡车来提高网络运载能力以提高收益是不合理的。卡车的数量需要综合网络的现实情况决定才可能达到图中的峰值。



(a) 不考虑卡车购买成本

(b) 考虑卡车购买成本

图 5.7 卡车数量敏感性分析

为了进一步分析在卡车数量对不同规模的系统，即单车数量不同的系统收益的影响差异，我们分别在单车数量为 899（实际规模）、499 时计算卡车数变化时系统收益的变化。其中，减小单车数量通过将每个站点的初始共享单车量同比例缩小达到。计算结果如图5.8所示。从图中可以看出，有 448 辆共享单车，两辆调度卡车的共享单车系统的盈利高于有 899 辆单车，无调度用车

的系统，这正说明在合适选择卡车数量的情况下，拥有更少共享单车的系统的盈利能力可以超过有更多共享单车的系统。所以，不需要为了提高收益一味的增加共享单车投放量，调度是除了增加单车投放量之外的一种有效提升系统满足需求能力的手段。

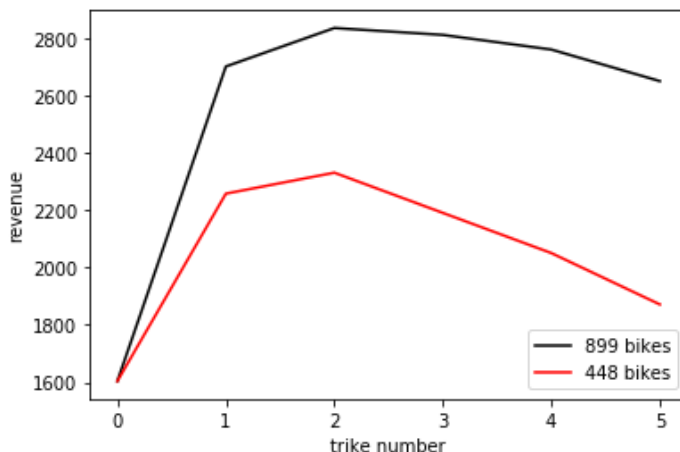


图 5.8 不同单车数量分析

在第二个场景中，我们进行在调度卡车数量一定时，单车数量对于收益影响的敏感性分析。将卡车数量设为 2，敏感性分析得到的结果如图5.9所示。图中显示，消费者满意度随单车数量的上升持续上升。这一点很好理解，因为从消费者的角度看，系统中有更多的单车就意味着他们有更高的机率能在自己的期望出发点找到可用的单车，这正是我们对消费者满意度的定义。另外，尽管从图中看来系统的收益也随单车数量的增多而上升，但这并不意味着更多的单车能够带来更多的利润，这是因为我们在场景二下进行对单车数量的敏感性分析，这意味着我们的假设是被研究的系统是一个新的或者需要改变规模的系统，所以此时单车购买的成本不能忽略，而应计入总成本中，但前述模型的目标函数并没有加上这一部分成本。由于单车数是在优化前就给定的，所以这部分成本为固定值，这里我们直接在求解结果中减去这部分固定成本求出实际收益。假设一辆共享单车的使用寿命为一年，那么购买一辆单车的花销可以均摊到每天，设为 S\$1.5。减去这部分成本后的结果如表5.3所示。从表中可以看出，单车数量和系统收益间没有明显的单一正相关或者负相关关系。事实上，对于一个在建的共享单车系统，最终的收益和很多因素都有关，比如单车的使用寿命、价格以及消费者对待单车的方式。由于这些因素在不同地区的差别可以很大，难以给出比较具有普适性的参数值，所以在决定单车数量时，应先全面考虑这些因素，以及结合单车系统所处的不同状态，确定合适参数值以及期望达到的目标后再进行计算。

单车系统所处状态不同，运营这个系统期望达到的目标也就不同，具体表现为：对于一个新的共享单车系统，消费者满意度通常是决定这个新的运营商是否能取得消费者信赖的关键指标；但当共享单车系统的运营已经进入了一个比较稳定的阶段，已经在消费者心中建立了信任体系后，系统的盈利水平会是维持运营的一个更重要因素。

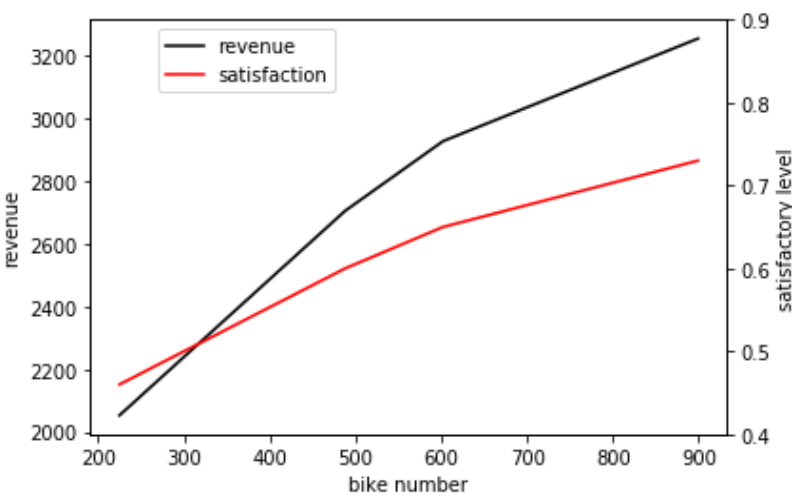


图 5.9 单车数量敏感性分析

表 5.3 纯利润对单车数量的敏感性

Number of bikes	Number of trikes	Profit
899	2	1905
602	2	2024
488	2	2033
225	2	1715

5.3.3 调度成本敏感性分析

在本节中，我们讨论调度成本对于模型结果的影响。这里的成本指的是燃油成本，因为这个成本会随着交通状况变化，而单车损耗的成本相对固定。本节中的所有计算实例都假设系统有899辆单车和2辆调度卡车。计算结果如图5.10所示。由结果计算可得。当燃油成本翻倍时，系统收益只下降了2%，而消费者满意度基本维持不变。这是由于一辆车的燃油成本是需要均摊到运送的每辆单车上的，例如，如果一辆卡车上运载了20辆单车，那么每辆单车带来的燃油成本就是总成本除以20，所以调度成本的增加对收益的影响没有前面讨论的系统规模对收益的影响那么显著。

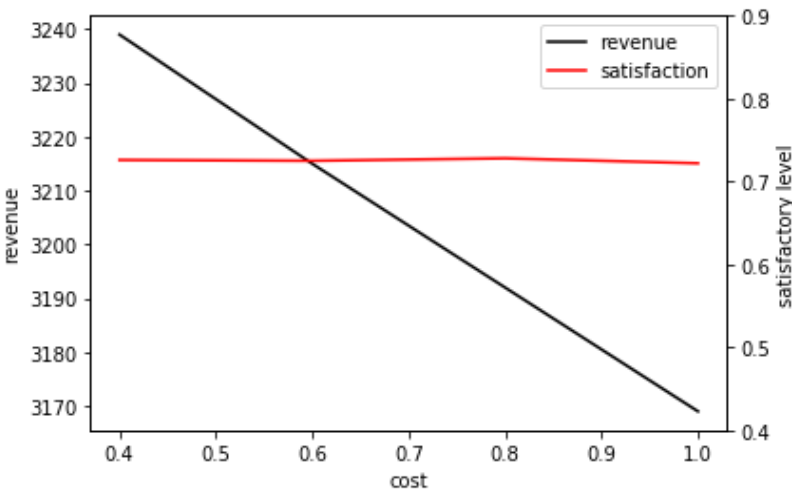


图 5.10 成本敏感性分析

### 5.3.4 定价敏感性分析

由于定价是决定共享单车系统是否能长期稳定运营的一个核心因素，所以在本节中，我们讨论定价对收益的影响规律。和前面几节一样，在改变单车使用价格时，我们假设这个系统中拥有 899 辆单车和 2 辆调度卡车。

在前面进行的敏感性分析中，由于定价总是固定的，所以我们仍认为消费者满意度是满足的需求与总需求的比值，且乘客对价格不敏感。但在现实世界中，消费者通常情况下都是对价格敏感的，价格的上升会使需求量下降。所以在本节中改变定价时，我们利用 Kaviti 等的文章中阐述的需求与定价的关系<sup>[35]</sup>，在价格改变时同样改变输入的需求量。图 5.11 展示了得到的结果。由图中可知，定价与系统收益的关系并不是单一的正相关或负相关关系，这是由于在价格变化时，需求量也会相应变化。但是，消费者满意度会随着定价持续上升，这是因为价格上升时需求量下降，在系统单车数量不变的情况下，这意味着每个有需求的用户在期望出发点找到可用车辆概率上升。和 4.4.1 相同，在不同阶段，对收益和乘客满意度这两者的考虑需要分别权衡，得到合适的定价策略。

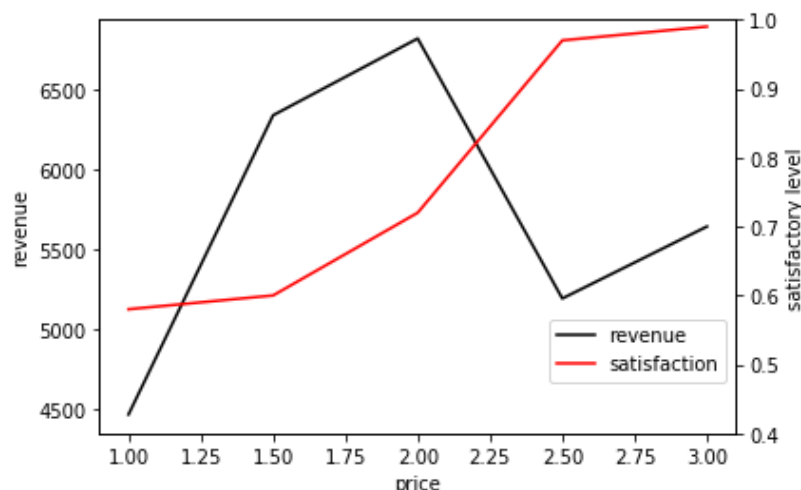


图 5.11 定价敏感性分析

### 5.3.5 模型延伸：增加站点容量的限制的共享单车系统

正如本文开头所说，有桩共享单车的历史比无桩共享单车长很多，并且在现有研究中，有桩共享单车常常被当作共享单车系统的典型运营模式进行研究。为了使我们的模型适用于有桩共享单车系统，只需要在原来的模型基础上增加一个每个站点的容量约束即可。以往对有桩共享单车系统的研究中，大多数都将调度问题看作一个车辆路径规划问题，强调如何规划最优的调度路线以消除期望单车分布和实际单车分布间的差别<sup>[6, 36, 37]</sup>，而本研究从不同的角度分析调度这个问题。在我们的模型中，我们并不关注调度的具体路线，而关注整个系统的调度计划。除此之外，我们的模型始终关注的是系统收益，而之前的研究主要关注如何利用调度提升用户服务，也就是说，提高消费者在期望始发地找到可用车辆，在期望目的地有空余停车桩的概率。

将以下约束加入模型中，得到适合有桩共享单车网络的模型：

$$b_i^t \leq cap, \forall i \in N, t \in T \quad (5.1)$$

$cap$  代表每个站点给定的桩的数量。

为了比较有桩和无桩两种共享单车系统在利用此模型求解时的结果差异，我们使用与基础实

例中相同的参数值和初始单车分布、需求值求解有桩共享单车和无桩共享单车网络的模型实例并进行比较。由于从新加坡的真实数据中得到的单车分布具有高空间不平衡的特征，比如在一些点存放的共享单车数量超过另一些点的十倍，在有桩共享单车实例中，我们需要设置站点容量为很高的值以保证从数据中获得的单车初始分布满足容量限制约束。然而，设置站点容量值过高会使得站点容量的约束成为多余的约束，因为在之后的求解中容量可能总大于需要停放的单车数量。为了解决这个问题，我们把每个站点的初始单车分布等比例缩放至均小于 30，再将站点容量从 30 提高到 70 观察收益变化。在所有实例中，选择系统中的十个站点，并设置调度卡车数量为三辆，其它参数和基础实例中设置的相同。求解结果如图5.12所示。

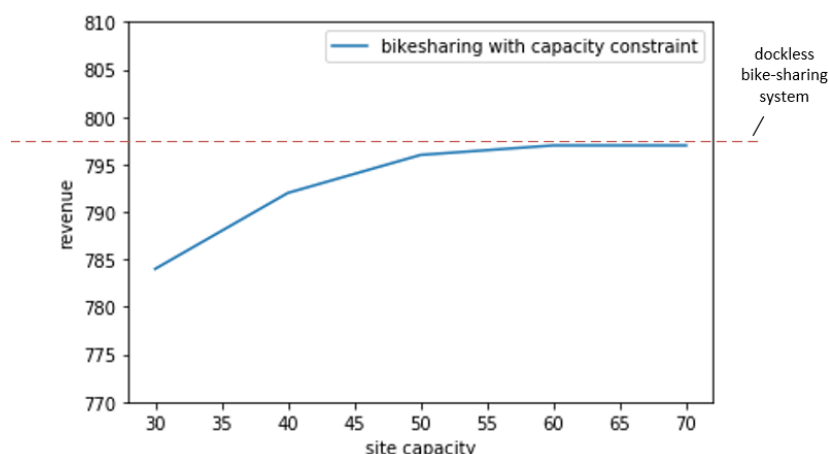


图 5.12 收益随站点容量变化

从上图中可以看出，随着站点容量的上升，有桩共享单车系统的收益也上升。当站点容量达到 60 时，有桩共享单车系统与相同规模、属性的无桩共享单车系统收益相等。这个结果证明通过在每个站点设置合适数量的停车桩，有桩共享单车系统的表现就能和无桩共享单车系统相同，并且兼有有桩和无桩共享单车系统的优势：（1）此系统能像无桩共享单车系统一样满足任何的归还车辆需求（2）此系统具有有桩共享单车系统的易于管理的优点。停车桩的建设也是成本的一部分，为了进一步在有桩共享单车系统中降低成本，可以在模型求解过程中，对每个站点的停车桩数量根据大致需求设置不同的值。这样一来，这个增加站点停车数量约束限制的模型就可以有效的为有桩共享单车网络的设计提供依据。

## 6 总结与讨论

在本研究中，我们首先讨论了统计模型和机器学习模型对单车需求预测的效果，经过对比之后，确定机器学习为一种更合适的方法，并选用KNN为本文的预测模型。之后，我们定义了一个混合整数优化问题和一个网络流模型用以解决共享单车网络的日常调度计划设计问题。我们利用了新加坡义顺区的实际共享单车使用数据作为输入基础，对构建的网络流模型进行了应用实例的测试。在证明了模型的有效性的基础上，我们又利用此模型对网络规模、成本和定价对网络收益的影响进行了敏感性分析，在分析的过程中，除了研究收益之外，我们对这些因素造成的消费者满意程度的变化也进行了讨论，因为消费者满意程度是衡量网络运营质量的关键指标。之后作为模型的延申，我们在原模型的基础上加入了站点容量约束，用于解决无桩共享单车的调度问题。最后，我们利用这个延申模型分析了有桩和无桩共享单车间的关系。

对于一个已知单车数量的共享单车系统，卡车数量对系统收益的影响可以直接利用本文中构建的模型进行分析，并帮助运营商管理调度。然而，当需要为一个新的共享单车系统决定合适的共享单车投放量时时，问题就会变得更加复杂。正如上一节所讨论的，对于一个新的或者需要改变规模的共享单车系统，可以通过两种途径尽可能提高系统收益。第一种途径是在系统中投放更多单车，第二种是在系统中投放更多卡车用于调度。由于无桩共享单车还处在发展阶段，为了提高市场占有率，很多运营商为了提高自己的市场竞争力一味地加大共享单车投放量，这是十分没有必要的。这种过度投放也使得一些共享单车项目最终走向破产，因为过多单车不但会提高购买成本，也会大大提高后期的运营成本。除了对受益的影响，共享单车的过度投放也会造成公共空间的浪费，这对城市管理者来说也是一个挑战。尽管共享单车的使用能够有效减少碳排放，其对空间资源的占用还是会造成环境问题。如果碳排放的减少需要以空间资源的浪费为代价，那么共享单车的使用是否是一个环保的出行方式就更有待商榷。

延申模型展示了一个很好理解的规律：通过增大站点的停车桩数量，一个有桩共享单车网络可以转变为无桩共享单车网络。但是，每个站点具体设置多少停车桩合理还需要更细致的研究计算。

通过本研究对共享单车需求预测以及网络运营规律和管理的分析和理解，最后，我们对整篇文章进行如下总结，并给出给出对网络运营和管理上的建议：

- （1）统计学习模型具有高度可解释性、快速预测的优点。针对其可解释性，目前已有可解释机器学习模型方法可与其达到相似效果，而针对其预测快速的优点，对于共享单车的需求预测，由于需要预测的数据量常常是有限的，比如未来一天、一周或一个月，数据规模不会很大，所以机器学习模型在预测速度上的劣势并不会造成太大问题。除此之外，机器学习模型还有预测准确性高的特点。所以对于单车需求预测，机器学习相比于统计学习是一种更为合适的方法。
- （2）共享单车网络的管理模式可以改进。在本研究中，我们并没有在所有的计算实例中考虑共享单车的购买成本，因为我们主要在固定网络规模的共享单车系统中考虑各个因素对收益的影响，所以可以看作共享单车购买这一步已经完成。在未考虑单车购买成本的情况下，共享单车项目看起来是很有盈利空间的。但是在现实世界中，共享单车项目走向破产的情况屡见不鲜，这是因为在实际运营中，车辆的损坏和丢失无法避免，所以在单车购买上通常需要持续投入。由于单车购买的投入太大，运营商难以长期维持系统运营，导致在现实中一些共享单车系统缺乏盈利能力。我们建议，缓解这个问题的途径之一可以考虑改变共享单车系统目前的运营结构。除了运营商自生之外，应让政府部门也参与到决定网络中投放的共享单车数量中，并且政府部门对单车运营商进行单车购买的



补贴，单车数量由系统的盈利能力和将占用的公共空间共同决定。在限制单车数量的调节下，运营商可以自主指定调度计划管理整个网络。网络最终的收益有运营商和政府共同享有。这种运营模式有诸多好处，例如，城市管理者可以通过设置单车数量限制减少单车将占用的公共空间资源，协调城市管理；运营商可以降低系统构建和运营成本，获得更多利润。这些综合起来将为城市居民提供高质量的共享单车服务，不仅方便出现，同样也环境友好。

- (3) 共享单车投放数量应受限制。从延申模型中得到的结果证明合理的设置每个站点停靠的单车数量并不会大大降低系统的盈利能力。共享单车过饱和不仅会提高系统运营成本，也会造成空间资源浪费，使得城市管理更加困难。所以，城市管理者可以设置共享单车的数量限制，设置数量限制并不会显著降低运营商的盈利或者降低服务质量。

随着共享单车系统在越来越多的城市落地，调度问题的场景也越来越多样、越来越复杂。未来的研究还需要继续关注这些不同的调度场景下的调度问题。以下提出一些研究点作为未来的研究讨论的方向的参考。在本研究中，为了简化模型构建，我们做了很多假设和简化。事实上，为了使模型贴合实际，更多更符合现实的因素应该被考虑，比如消费者在站点等待可用单车或者移动到附近站点寻找可用单车，这将会影响消费者满意率的计算。消费者满意率同样也需要更精细的定义。至于调度手段，除了运营商通过配置卡车进行统一调度，使用骑行激励，即利用骑车补贴鼓励消费者改变期望目的地实现调度的方式也是一种有效的调度策略。除以上这些之外，进一步的研究也可以着眼于寻找更高效的模型求解算法，因为随着问题规模的扩大，求解难度也会大大上升。无桩共享单车的模型规模由于其停车点的随机性要远大于有桩共享单车系统。最后，在本研究所讨论的网络流量模型中，每辆调度卡车的调度路线是被模糊化处理的，如果想要制订更精细的调度策略，未来的研究可以考虑把车辆路径规划问题整合到目前的模型中。

装

订

线



### 参考文献

- [1] DEMAIO P. Bike-sharing: History, impacts, models of provision, and future[J]. Journal of public transportation, 2009, 12(4): 3.
- [2] SHAHEEN S A, GUZMAN S, ZHANG H. Bikes sharing in Europe, the Americas, and Asia: past, present, and future[J]. Transportation Research Record, 2010, 2143(1): 159-167.
- [3] FISHMAN E. Bikeshare: A review of recent literature[J]. Transport Reviews, 2016, 36(1): 92-113.
- [4] ERDOĞAN G, BATTARRA M, CALVO R W. An exact algorithm for the static rebalancing problem arising in bicycle sharing systems[J]. European Journal of Operational Research, 2015, 245(3): 667-679.
- [5] PAL A, ZHANG Y. Free-floating bike sharing: Solving real-life large-scale static rebalancing problems[J]. Transportation Research Part C: Emerging Technologies, 2017, 80: 92-116.
- [6] CHEMLA D, MEUNIER F, CALVO R W. Bike sharing systems: Solving the static rebalancing problem[J]. Discrete Optimization, 2013, 10(2): 120-146.
- [7] HO S C, SZETO W. Solving a static repositioning problem in bike-sharing systems using iterated tabu search[J]. Transportation Research Part E: Logistics and Transportation Review, 2014, 69: 180-198.
- [8] FORMA I A, RAVIV T, TZUR M. A 3-step math heuristic for the static repositioning problem in bike-sharing systems[J]. Transportation research part B: methodological, 2015, 71: 230-247.
- [9] SCHUIJBROEK J, HAMPSHIRE R C, VAN HOEVE W J. Inventory rebalancing and vehicle routing in bike sharing systems[J]. European Journal of Operational Research, 2017, 257(3): 992-1004.
- [10] CAGGIANI L, OTTOMANELLI M. A dynamic simulation based model for optimal fleet repositioning in bike-sharing systems[J]. Procedia-Social and Behavioral Sciences, 2013, 87: 203-210.
- [11] FRADE I, RIBEIRO A. Bicycle sharing systems demand[J]. Procedia-Social and Behavioral Sciences, 2014, 111: 518-527.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] XU T, HAN G, QI X, et al. A hybrid machine learning model for demand prediction of edge-computing-based bike-sharing system using internet of things[J]. IEEE Internet of Things Journal, 2020, 7(8): 7345-7356.
- [16] VE S, CHO Y. A rule-based model for Seoul Bike sharing demand prediction using weather data[J]. European Journal of Remote Sensing, 2020, 53(sup1): 166-183.

- [17] CHEN L, ZHANG D, WANG L, et al. Dynamic cluster-based over-demand prediction in bike sharing systems[C]//Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016: 841-852.
- [18] HULOT P, ALOISE D, JENA S D. Towards station-level demand prediction for effective rebalancing in bike-sharing systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 378-386.
- [19] LI Y, ZHU Z, GUO X. Operating characteristics of dockless bike-sharing systems near metro stations: Case study in Nanjing city, China[J]. Sustainability, 2019, 11(8): 2256.
- [20] XING Y, WANG K, LU J J. Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai, China[J]. Journal of transport geography, 2020, 87: 102787.
- [21] FRICKER C, GAST N. Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity[J]. Euro journal on transportation and logistics, 2016, 5(3): 261-291.
- [22] GHOSH S, VARAKANTHAM P. Incentivizing the use of bike trailers for dynamic repositioning in bike sharing systems[J]., 2017.
- [23] PAN L, CAI Q, FANG Z, et al. A deep reinforcement learning framework for rebalancing dockless bike sharing systems[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 33. 2019: 1393-1400.
- [24] DUAN Y, WU J. Optimizing rebalance scheme for dock-less bike sharing systems with adaptive user incentive[C]//2019 20th IEEE International Conference on Mobile Data Management (MDM). 2019: 176-181.
- [25] JI Y, JIN X, MA X, et al. How Does Dockless Bike-Sharing System Behave by Incentivizing Users to Participate in Rebalancing?[J]. IEEE Access, 2020, 8: 58889-58897.
- [26] CAGGIANI L, CAMPOREALE R, OTTOMANELLI M. A dynamic clustering method for relocation process in free-floating vehicle sharing systems[J]. Transportation Research Procedia, 2017, 27: 278-285.
- [27] ZHANG R, PAVONE M. A queueing network approach to the analysis and control of mobility-on-demand systems[C]//2015 American Control Conference (ACC). 2015: 4702-4709.
- [28] VISHKAEI B M, MAHDAVI I, MAHDAVI-AMIRI N, et al. Balancing public bicycle sharing system using inventory critical levels in queueing network[J]. Computers & Industrial Engineering, 2020, 141: 106277.
- [29] MANNERING F L, BHAT C R. Analytic methods in accident research: Methodological frontier and future directions[J]. Analytic methods in accident research, 2014, 1: 1-22.
- [30] LUKUSA M T, PHOA F K H. A Horvitz-type estimation on incomplete traffic accident data analyzed via a zero-inflated Poisson model[J]. Accident Analysis & Prevention, 2020, 134: 105235.
- [31] MOLNAR C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable[M]. 2019.

- [32] PODUSKA J. SHAP and LIME Python Libraries:Part 2 - Using SHAP and LIME[EB/OL]. 2019 [2019-01-14]. <https://blog.dominodatalab.com/shap-lime-python-libraries-part-2-using-shap-lime/>.
- [33] Gurobi Optimization Inc. Gurobi optimizer reference manual[J]. URL <http://www.gurobi.com>, 2018.
- [34] SHEN Y, ZHANG X, ZHAO J. The mobility patterns of dockless bike sharing: A four-month study in Singapore[J]. Transportation Research Part D: Transport and Environment, forthcoming.
- [35] KAVITI S, VENIGALLA M M. Assessing service and price sensitivities, and pivot elasticities of public bikeshare system users through monadic design and ordered logit regression[J]. Transportation Research Interdisciplinary Perspectives, 2019, 1: 100015.
- [36] LIU J, SUN L, CHEN W, et al. Rebalancing bike sharing systems: A multi-source data smart optimization[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1005-1014.
- [37] PFROMMER J, WARRINGTON J, SCHILDBACH G, et al. Dynamic vehicle redistribution and online price incentives in shared mobility systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(4): 1567-1578.

装

订

线

### 致谢

转眼间四年的本科生涯即将结束，四年的日子说长不长，说短也不断，四年前初入校园的画面仿佛还在眼前。回首四年的同济生活，很感谢有各位老师和同学的陪伴，我的本科生活非常充实与难忘，在统同济的学习和生活也给了我足够的自信奔向下一段旅程。

感谢我的导师沈煜老师，感谢沈老师对我毕业论文的指导，更要感谢的是两年来在我面临很多选择感到迷茫的时候，您的耐心倾听和提供的宝贵建议与帮助，这使我在做出决定的时候从来不觉得彷徨和害怕。还要感谢您给我的机会，让我能更早的获得探索科研领域的机会，以及一路上给我的鼓励 and 信心。

感谢谢驰老师，在毕业设计的思路给我的种种启迪和指引，在我觉得遇到问题时总是耐心的解释和提出的真诚建议，为我打开了新知识的大门，让我有机会探索一个全新的领域。

感谢杜豫川老师，在我刚升入大二还有诸多不懂得的情况下仍然很用心的指导了我的创新项目，感谢您提供给我的机会和科研资源，带领我进入了科研的大门，感谢您在我留学路上提供的慷慨帮助。

还要感谢我的室友和朋友们，四年来因为你们的陪伴让我的大学生活增添了很多色彩。虽然来自不同的城市，有不同的成长轨迹，不同的生活习惯，但四年来我们从没有过争吵，记忆中全是快乐和感动，我们也是相互陪伴、可以倾诉心事的人。很感激我们一路一起走到了这里，希望在以后的人生中，我们可以一直是彼此生活中重要和亲近的人。

最后感谢我的父母，在我觉得失落的时候总是鼓励我，在我觉得烦闷的时候总是倾听，在我需要做出人生抉择的时候总是无条件的支持我的决定。感谢你们给我尽情做我自己的自由，成为你们的孩子是让我觉得最幸福和幸运的事。

四年是一段旅程的终点也是下一段旅程的起点，我已经准备好走向下一个终点，山高水长，希望来日再见。

装

订

线