

城市公共自行车服务系统运行状况和效率分析

基于温州市鹿城区公共自行车系统运营实践的研究

摘 要

本文基于温州市鹿城区公共自行车系统的 20 天内借车和还车原始数据,构建模型对公共自行车的使用频次分布、供求状况和自行车租赁的站点位置特征进行了分析,进而探讨了公共自行车系统的有效性,挖掘其背后的系统运行规律和机理,并分析了这一系统的其他特征,如潮汐现象等,最终建立在长期可持续发展的基础上提出解决方案。

针对累计借还车频次、各日借车卡数量和累计借车次数分布的计算,我们主要采用 MATLAB 遍历过所有数据并进行统计运算,得到累积分布结果并二维直方图,采用拟合的方法得到其近似分布。我们还构建了自行车用户群体的效用函数模型,探讨高于某一保留效用阈值的城市居民采用公共自行车出行偏好程度。发现用户选择公共自行车出行的效用函数关于出行时间的变化近似为一个很宽脉冲函数,其拐点约为 28 分钟。当出行时间少于 28 分钟时,用户更倾向于选择公共自行车出行;在 28 分钟之后,用户倾向于选择其他的交通工具,仅在随机因素的作用下选择公共自行车。

在公共自行车服务系统站点设置和锁桩数量的配置的分析中,我们在鹿城区公共自行车管理中心网站上找到每个自行车站点的位置,在电子地图服务提供商网站上查询出该站点位置的坐标。对于距离的界定,我们采用欧氏距离、直角距离和地图实际距离三种方法度量站点间距。第三种方法较为精确,我们采用颜色替换,形态学处理,道路生长细化的方法克服了第三种方法的度量困难,最终获得城市道路信息,用于两地之间的实际道路路径求取。

我们将一天中的 24 小时划分成以 0.2 小时为间隔的若干时段,分别统计落入每个时段的借还车频次,即可用来反映借还车频次时刻分布,我们发现许多站点都存在较明显的早晚高峰现象。反推出的效用函数具有稳定性和可靠性,表明我们的研究是可信的。在峰值的搜寻中,我们借助 MATLAB 使用均值滤波的方法减弱随机因素造成的影响,对站点的借还车高峰时刻进行统计。通过计算等效时间峰值统一不同站点的时间峰值维数,然后采用 K-means 算法进行聚类分析,从而根据借还车高峰时刻对站点进行了归类。

基于上述初步结果,我们提取了数据中蕴含的信息,进而对鹿城区的公共自行车系统配置效率和运营状况进行分析和评价。我们发现,公共自行车系统存在明显的“潮汐现象”,在高峰时段可能有供不应求的问题,亟需建立良好的调度系统已平衡供求关系。我们拓展了前面的研究,从站点选址的社会福利优化问题以及合理的配套调度机制设计入手,进一步分析了公共自行车服务系统的运行规律,并提出了相应的改进建议。

关键词: 公共自行车 效用函数 分布估计 统计分析 聚类分析

一、问题提出

1.1 题目背景与研究意义

在倡导健康、节能、环保的生活理念的今天，公共自行车得到了越来越多城市居民的青睐，在国内掀起了一场“绿色革命”^[1]。作为同城短区间非机动车载人设备，公共自行车是具有可持续发展性的重要代步手段，日益成为人们生活中不可或缺和替代的主要交通工具之一。作为公共交通末端交通工具的公共自行车不仅弥补了公共交通线网密度不足的缺陷，而且向公共交通提供客源，如图 1 所示。我国现已有众多城市主动将自行车纳入公共交通领域，意图让公共自行车交通与公共交通实现无缝对接，破解交通末端“最后一公里”难题，实现低碳出行，美化城市。

随着公共自行车的普及使用，如何构建合理、高效的公共自行车服务系统是一个非常关键的问题。自行车租赁的站点位置及各站点自行车辆数和自行车数量的配置，对系统的运行效率与用户的满意度有重要的影响。浙江省温州市鹿城区公共自行车系统的数据能够为我们的机制设计提供参考。根据温州市鹿城区公共自行车管理中心网站（<http://www.wzbicycle.com>）提供的数据，鹿城公共自行车的投放量约为 5000 辆，设立 180 个服务网点。启动当天，首先在五马街区和新城街区推出 80 个服务网点，其余 100 个服务网点计划在 10 月设立。服务网点主要位于机关、企事业单位、商业楼宇、公共场所、住宅小区等人员密集地区，间距一般在 300 米至 500 米^[2]。

本文试图基于温州市的地形环境和公共自行车站点建设的现有情况，构建数学模型探讨温州市市民公共自行车的使用水平和分布规律，并对城市自行车服务系统进行评价。目前，我国公共自行车的系统建设尚处于初步阶段，我们的研究不仅针对特定城市的公共自行车系统进行了系统性分析，而且能够为全国其他城市的公共自行车系统优化建设提供参考。

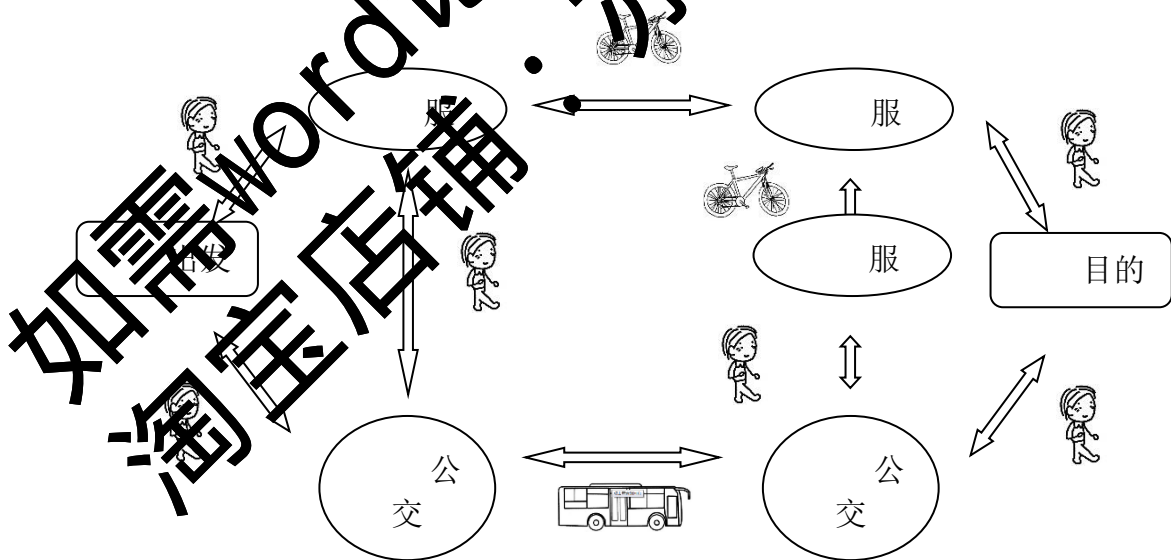


图 1：城市公共自行车系统

1.2 问题重述

借助鹿城区公共自行车客流数据和站点分布图，本文试图解决以下几个问题：首先，分析公共自行车的使用频度和每次使用的时长：我们统计各站点 20 天中每

天及累计的借车频次和还车频次,并对所有站点按累计的借车频次和还车频次分别给出它们的排序,并统计分析每次用车时长的分布情况。

其次,基于借车卡数据对使用主体区分,统计 20 天中各天使用公共自行车的不同借车卡(即借车人)数量,并统计数据中出现过的每张借车卡累计借车次数的分布情况。

然后,我们试图寻找所有已给站点合计使用公共自行车次数最多的一天,并按照以下几个角度探讨站点客流量特点和规律:

1) 我们合理界定两站点之间的距离,在此基础上找出自行车用车的借还车站点之间(非零)最短距离与最长距离,对借还车是同一站点且使用时间在 1 分钟以上的借还车情况进行统计。

2) 我们选择借车频次最高和还车频次最高的站点,分别统计分析其借、还车时刻的分布及用车时长的分布。

3) 寻找各站点的借车高峰时段和还车高峰时段,在给定的温州市鹿城区地图上标注或列表给出高峰时段各站点的借车频次和还车频次,并对具有共同借车高峰时段和还车高峰时段的站点分别进行归类。

进而建立在前面分析的基础上,我们探讨上述统计结果携带了哪些有用的信息,并由此对目前公共自行车服务系统站点设置和锁桩数量的配置做出评价。

最后,我们试图找出公共自行车服务系统的其他运行规律,提出改进建议。

二、问题分析

城市公共自行车系统是一个复杂的系统,“牵一发而动全身”,其机制的设计和模式的优化是城市公共基础设施建设服务的核心。在这一部分中,我们对要解决的问题进行梳理整合,为后文的研究构建逻辑架构。

2.1 基本思路

针对本文的几个关键问题点,我们首先从数据出发,通过对公共自行车使用频率和时长的数据观察统计,分析其中规律,进而采用效用函数合理刻画用户需求和选择模式,合理预测出行时间,建立模型刻画用户偏好。然后我们采用题目中给定的地图,借助电子地图寻找借还车频次高峰,并基于数据和初步结论探索最后的开放式问题——公共自行车系统的运行规律和政策建议。

2.2 具体分析

我们首先进行数据的预处理,利用 Matlab 导入数据,对数据进行整合,剔除异常值和缺漏值,得到用于分析的样本数据。我们分以下几个步骤解决:

1) 针对问题一,即计算公共自行车的使用频度和每次使用的时长,我们遍历过所有数据之后,即可以得到各站点每天的借车还车频次的统计数据。对每个站点的全部 20 天的借车与还车数据分别求和,既可以得到每个站点的累计借车与还车频次。用还车时间减去借车时间可以得到用户的用车时间,精确到秒。

2) 利用用户效用函数来刻画用户的选择行为,建立自行车出行时间分布的效用函数模型。求解效用函数模型的具体形式,对用户群体的偏好特征进行分析。

3) 针对问题二,我们按照借车卡区分不同用户个体,对用户群体的需求模式进行探索,方法类同问题一。通过累计各个借车人在 20 日中累计借车次数,我们做出频率

分布图，拟合得到借车频次的近似分布特征。

4) 针对问题三，共有三种方法界定站点距离——欧式距离，直角距离和实际地图距离，我们可以分别用三种方法求解。第三种方法最为精确，但实现难度较大，我们尝试采用 MATLAB 画图的方法，利用题目给出的地图计算实际距离。在此基础上，寻找借还车频次高峰，并对站点位置、锁桩数量的优劣进行评价。

5) 第四、第五问要求我们发散思维，寻找数据背后的运行规律，我们建立在数据特征的基础上，进一步挖掘公共自行车系统的运行特点，并回顾和借鉴了一些现有研究的方法，最终提出政策建议。

三、基本假设

H3.1: 假设该城区所有的用户都是同质的，即有相同的效用函数。由于生活地理位置和风俗人情接近，影响出行状况的外界不可抗因素相近，这一假设较为合理，并可以简化我们的分析。当用户同质时，我们可以采用代表性用户的效用函数完成推导过程。

H3.2: 不同用户出行需求不同，我们假设用户需要出行时长(T)近似服从指数分布。

H3.3: 假设公共自行车站点设置在研究时间区间内没有发生变更，借车站点序号没有较大的变化，我们可以在一个相对静态的公共自行车硬件系统环境下进行分析。

四、符号说明

符号	意义
$A_{m \times n} = [a_{ij}]$	第 i 个站点第 j 天的借车次数
$B_{m \times n} = [b_{ij}]$	第 i 个站点第 j 天的还车次数
m	站点数目的最大值
n	天数的最大值
a_0	第 i 个站点的累计借车频次
b_0	第 i 个站点的累计还车频次
T	用户需要出行的时长
$U(t)$	用户自行车出行的效用
$Q(x)$	标准正态分布的累计分布函数

五、数据预处理

分析发现公共自行车系统的数据记录中存在错误的数据记录。因此，在对系统使用情况进行统计分析之前，首先要对数据进行预处理，以去除这些错误数据。分析发现，数据记录错误的形式包括以下几种：

1) 借车站点序号错误，例如 2012 年 11 月 8 日的序号为 No.16548 的记录，借车站点为“调试站 1”，站点序号为 1000；

2) 还车站点序号错误, 例如 2012 年 11 月 8 日的序号为 No.30983 的记录, 还车站点名称为空白, 站点序号为 29999;

3) 借还车时间错误, 例如 2012 年 11 月 8 日的序号为 No.30983 的记录, 借车时间为“2012/11/5 17:50:07”, 还车时间为“2012/11/5”, 还车时间只有日期没有时刻;

4) 用车时间间隔错误, 例如 2012 年 11 月 5 日的序号为 No.19436 的记录, 借车时间为“2012/11/5 14:24:43”, 还车时间“2012/11/5 14:31:23”, 实际用车时间为 6 分 40 秒, 记录用车时间为 0 分钟。

综上所述, 制定筛选条件对错误数据进行滤除。正确的数据需要满足的条件为:

- 1) 借车站点序号小于 1000;
- 2) 还车站点序号小于 1000;
- 3) 借还车时间格式正确, 且还车时间晚于借车时间;
- 4) 还车时间减去借车时间所得到的时间间隔 (精确到秒) 与记录中的时间间隔的误差小于 2 分钟。

我们使用 MATLAB 导入数据, 并编写程序根据上述条件对全部 20 天的数据记录进行筛选, 最终得到一共 637336 条记录, 其中借车时间范围为 2012/11/1 5:59:38 至 2012/11/20 21:33:57, 还车时间范围为 2012/11/1 6:02:15 至 2012/11/21 7:37:20。数据一共有 180 个借车站点, 站点序号分别为 1~107, 109~180; 180 个还车站点, 序号范围和借车站点相同。

六、公共自行车总体使用情况统计分析

6.1 借车还车频次统计

针对导入的数据, 使用 MATLAB 编写程序对各站点每天的借车还车频次进行统计, 并分别使用矩阵 $A_{m \times n} = [a_{ij}]$, $B_{m \times n} = [b_{ij}]$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$ 来表示借车频次和还车频次的统计结果。其中 a_{ij} 表示第 i 个站点第 j 天的借车次数, b_{ij} 表示第 i 个站点第 j 天的还车次数。 m 和 n 分别表示站点数目的最大值和天数的最大值, 根据数据记录的范围, 分别取 $m = 181$, $n = 20$ 。

程序初始化后建立 $A_{m \times n}$ 和 $B_{m \times n}$ 两个矩阵并把它们置为零矩阵。程序遍历所有数据记录, 读每一条数据记录之后, 都对该条数据记录中借车日期、借车站点、还车日期、还车站点所对应的矩阵中的元素加 1。遍历过所有数据, 即得到各站点每天的借车还车频次的统计数据。对每个站点的全部 20 天的借车与还车数据分别求和, 可得每个站点的累计与还车频次, 第 i 个站点的累计借车与还车频次分别用 a_{i0} 和 b_{i0} 来表示, 显然有:

$$a_{i0} = \sum_{j=1}^n a_{ij}$$

$$b_{i0} = \sum_{j=1}^n b_{ij}$$
(6.1)

各个站点的每天与累计借车与还车频次如表 1 和表 2 所示，限于篇幅，表 1 和表 2 中仅列出了前 10 个站点前 10 天的借还车频次以及累计频次，完整的数据请参见附录。

表 1：各个站点的每天与累计借车频次（部分）

站点	a_{i1}	a_{i2}	a_{i3}	a_{i4}	a_{i5}	a_{i6}	a_{i7}	a_{i8}	a_{i9}	a_{i10}	累计
1	91	148	35	152	124	107	88	15	9	6	1760
2	114	112	25	82	135	135	120	41	21	15	1744
3	179	188	47	60	194	189	188	3	65	13	2672
4	235	286	86	255	243	322	259	12	65	66	5232
5	139	165	67	167	161	164	108	53	33	16	2412
6	105	100	50	102	117	118	151	47	19	13	1760
7	106	109	14	55	143	137	157	24	35	5	1675
8	95	80	15	47	106	116	114	35	41	10	1533
9	353	365	177	303	400	460	393	57	124	86	6483
10	354	359	198	311	265	334	276	165	82	53	5821

表 2：各个站点的每天与累计还车频次（部分）

站点	a_{i1}	a_{i2}	a_{i3}	a_{i4}	a_{i5}	a_{i6}	a_{i7}	a_{i8}	a_{i9}	a_{i10}	累计
1	93	145	46	167	116	103	98	23	11	4	1821
2	117	110	27	87	125	136	127	53	22	16	1811
3	179	183	65	199	187	187	219	115	54	9	2915
4	235	297	108	283	242	310	297	115	87	61	5582
5	151	153	80	166	182	176	190	61	49	16	2613
6	112	97	41	116	119	119	132	56	33	15	1829
7	106	106	23	59	140	134	157	67	19	3	1798
8	84	89	15	41	101	110	110	54	23	4	1478
9	324	351	171	315	408	447	400	178	121	72	6416
10	341	332	216	314	361	326	347	150	84	52	5741

对累计借车频次和还车频次进行排序，得到的结果如表 3 所示，表 3 中仅给出借车与还车频次最高的前 10 名站点，完整的排序结果见附录。

表 3：站点的借还车频次排序

借车站点累计频次排序	还车站点累计频次排序
------------	------------

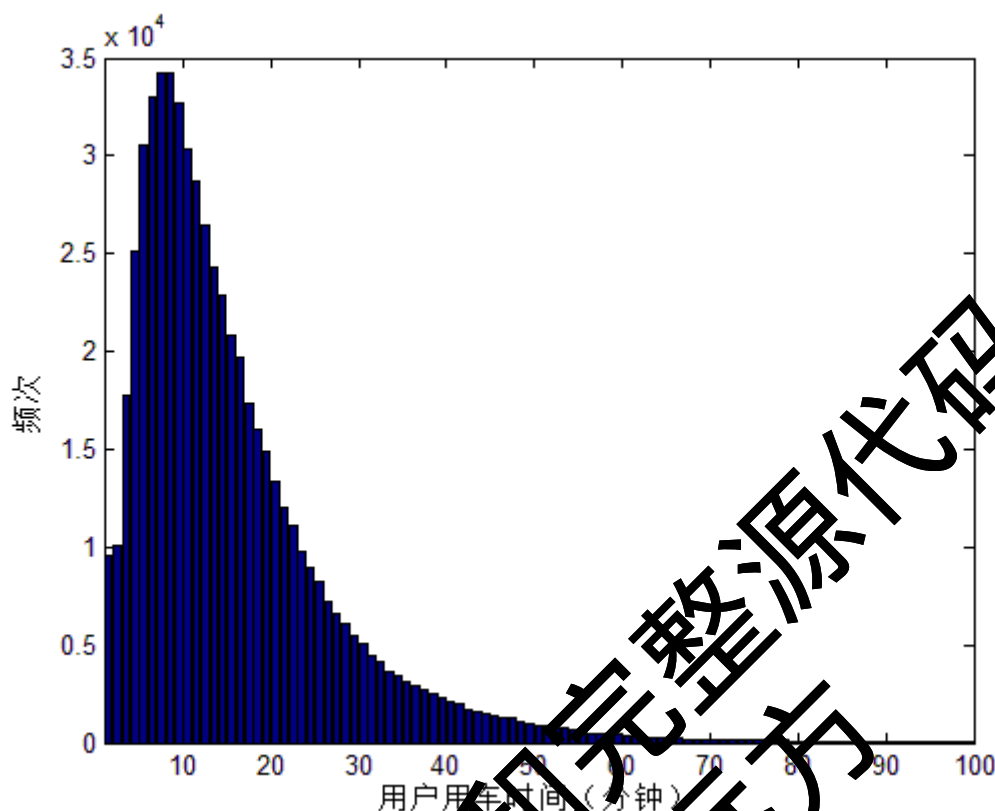


图 2：用户用车时间的频次分布

6.3 效用函数模型

公共自行车的推广主要针对的人群有三类：短距离出行者、长距离换乘出行者和非常规目的出行者等。短距离出行者具体是指自行车的出行适宜的范围（5km）内，用公共自行车的单一交通方式完成的出行的，包括通勤、购物以及娱乐为目的的交通出行者。长距离换乘出行者指使用公共自行车作为公共交通的一个换乘工具的日常通勤者或者节假日出门娱乐、探访亲友或者购物的人群。非常规目的出行者是指以健身或者旅游为目的的出行者。

用户选择公共自行车出行与否显然和出行的时间相关，出行时间间接反映了出行距离的远近。自行车是一种人力驱动的出行交通工具，在短时间条件下，出行者体力充沛，其速度优势得到体现，但是在时间超过一定的程度，其速度优势就完全暴露。因此出行时间是自行车方式选择的重要因素。表 8 反映了统计资料中披露的部分城市自行车出行时耗分布。从表 8 来看自行车交通的出行时间大部分都控制在 30 分钟以内。绝大部分城市自行车出行时间控制在 20 分钟以内的比例达到了 60%以上，时间控制在 30 分钟以内的比例都达到了 85%左右（郑州除外），出行时间超 30 分钟的出行比例较小，我们可以认为，自行车的最佳出行时耗在 30 分钟以内。^[3]

表 8：部分城市自行车出行时耗分布（%）¹

出行时耗	<10min	<20min	<30min	<40min	<50min	统计年份
郑州	4.2	42.5	62.3	89.2	93.8	2000

¹ 数据参见单晓峰（2007）：城市自行车交通合理方式分担率及其路段资源配置研究，东南大学博士论文，2007

苏州	18.4	62.7	81	87	94	2000
无锡	28.1	65.9	87.1	93.3	96.9	1996
常熟	40.2	80.8	95.6	96.8	97.7	2001
昆山	7.4	62.4	84.8	96.5	97.4	2001

我们构建了一个模型来描述用户用车时长的分布规律。我们假设用户需要出行的时长是一个随机变量，用 T 来表示，并用 $p(T=t)$ 表示 T 的概率密度函数。用户根据自己的出行需求选择要采用的交通工具。显然，用户选择出行交通工具需要考虑一些因素，例如出发点、出发终点、出行距离、出行用时、出行时段以及交通状况、天气等一系列因素。用户会综合考虑这些因素以确定出行所采用的交通工具。

我们引入效用函数来描述用户选择自行车出行的偏好。效用函数是对用户使用公共自行车获得的综合收益的量化，显然，当收益较高的时候，用户更倾向于使用公共自行车。设用户采用公共自行车出行的效用函数为 $U(t)$ ，则可以设用户使用公共自行车的条件 S_1 为效用函数大于某一阈值 δ ，这一阈值可以视为用户选择出行方式的保留效用。如式(6.2)所示：

$$S_1: U(t) > \delta \quad (6.2)$$

由于除了 t 之外其他因素也可能影响用户使用公共自行车的效用，所以 $U(t)$ 应该是以 t 为参数的随机变量。为了简化起见，设效用函数是和距离相关的确定部分 $u(t)$ 与一随机变量的和，即

$$U(t) = u(t) + n \quad (6.3)$$

式(6.3)中， $u(t)$ 为 t 的函数， n 是一个随机变量，用以刻画影响用户选择的随机因素。由于随机因素可以看成许多正、负的随机分布的综合效应，根据中心极限定理，我们可以假定 n 服从正态分布，即 $n \sim N(\mu, \sigma^2)$ 。

将式(6.3)代入式(6.2)中可得

$$\begin{aligned}
S_1: U(t) &> \delta \\
\Leftrightarrow S_1: u(t) + n &> \delta \\
\Leftrightarrow S_1: \frac{u(t) - \delta + \mu}{\sigma} + \frac{n - \mu}{\sigma} &> 0 \\
\Leftrightarrow S_1: u_0(t) + n_0 &> 0
\end{aligned} \quad (6.4)$$

其中， $u_0(t)$ 和 n_0 分别为归一化之后的确定效用函数分量和随机效用函数分量，并

且有 $n_0 \sim N(0,1)$ 。用随机变量 Y 来表示用户使用公共自行车出行的用时。用户选用公共自行车出行需要满足两个条件，第一，用户需要出行的时间达到一定水平；第二，在此基础上，并综合各种其他因素，用户最终选用公共自行车。因此， Y 的密度函数可以表示为：

$$p(Y=t) = p(S_1)p(T=t) \quad (6.5)$$

由于 $n_0 \sim N(0,1)$ ，故 n_0 的概率密度函数为：

$$p(n_0 = x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (6.6)$$

把式(6.6)代入式(6.5)可得

$$p(Y=t) = p(T=t) \cdot \int_{-u_0(t)}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (6.7)$$

式(6.7)为使用公共自行车出行时间分布的效用函数模型。

6.4 出行时间的效用函数模型求解

下面我们求解用户出行时间的效用函数模型，通过求解模型，可以解出效用函数的表达形式和用户出行时间的分布。设 $Q(x)$ 为标准正态分布的累计分布函数，即

$$Q(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$$

根据对称性我们可知 $\int_{-u_0(t)}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = Q(u_0(t))$ ，则有：

$$p(Y=t) = p(T=t) \cdot Q(u_0(t)) \quad (6.8)$$

解出 $Q(u_0(t))$ 可以表示为：

$$Q(u_0(t)) = \frac{p(Y=t)}{p(T=t)} \quad (6.9)$$

由于 $Q(u_0(t))$ 为单调非减连续函数，我们必可以求解其反函数，从而解出用户的效用函数表达式：

$$u_0(t) = Q^{-1}\left(\frac{p(Y=t)}{p(T=t)}\right) \quad (6.10)$$

我们可以利用样本数据的统计结果近似代替 Y 的分布，即用样本的出现的频率分

布来代替随机变量的密度函数,如图 1 所示。因为数学期望在样本趋近于无穷大的时候收敛于数学期望,因此这种近似是有效的。

对于用户需要出行的时间 T ,其分布是独立于该问题的,并且只要任给一个 T 的分布就可以解出一个效用函数的表达式, T 分布的具体形式并不影响解决问题的方法的一般性。因此我们可以假设一种 T 的分布。常识性的知识表明用户需要出行某一时间的概率随时间的增加而下降,即用户有更大的可能在更近的范围内活动。因此不妨设 T 服从指数分布,即 T 的密度函数为:

$$p(T=t)=\lambda e^{-\lambda t} \quad (6.11)$$

可以设 $\lambda=1/100$ 。由于仅统计 100 分钟以内的出行时间,因此需要对 p 的概率密度函数进行归一化,归一化之后概率密度函数如图 3 所示。

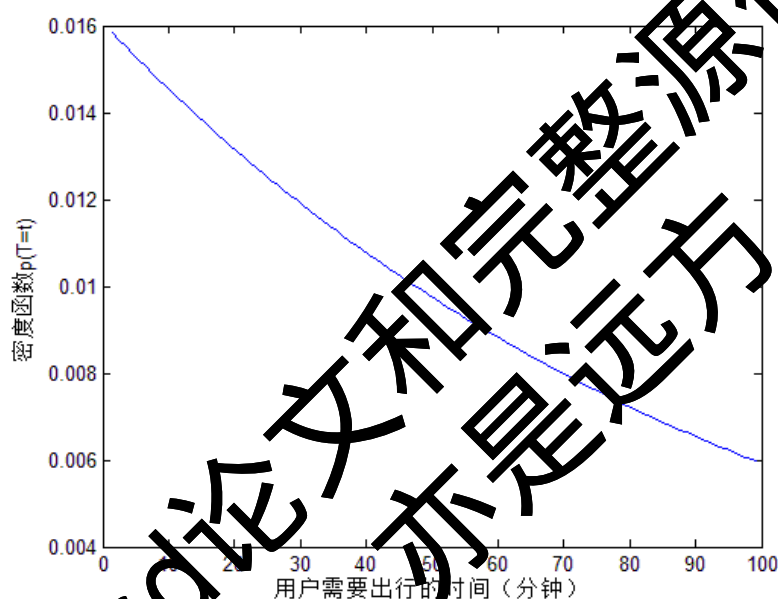


图 3: 用户需要出行的时间 T 的密度函数

Q 没有解析解,但是正态分布的累积分布函数有使用数值方法计算出的函数值表,可以使用反查该表的方法来得到 $Q^{-1}(x)$,其函数图象如图 4 所示。

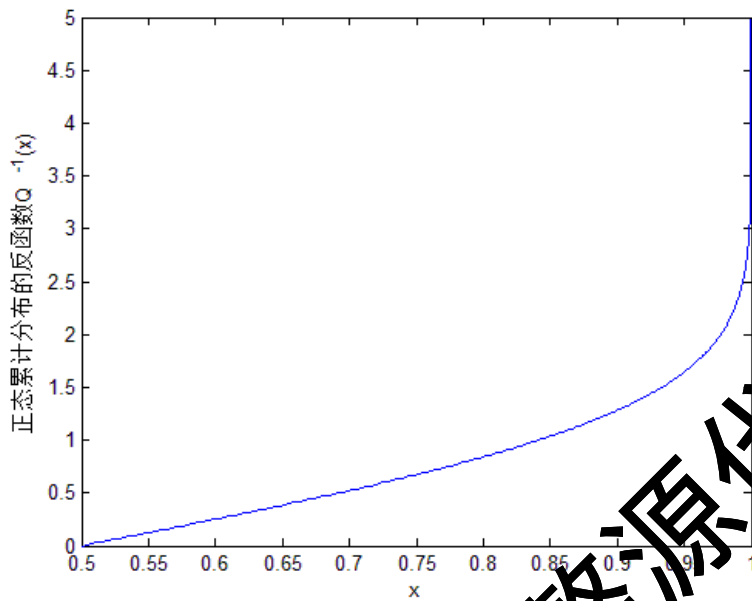


图 4：正态累计分布的反函数 $Q^{-1}(x)$ 的函数图象。

综上所述，将 $p(Y=t)$ 、 $p(T=t)$ 和 $Q^{-1}(x)$ 的值代入式 6.10，即可解得效用函数 $u_0(t)$ ， $u_0(t)$ 的函数图象如图 5 所示。

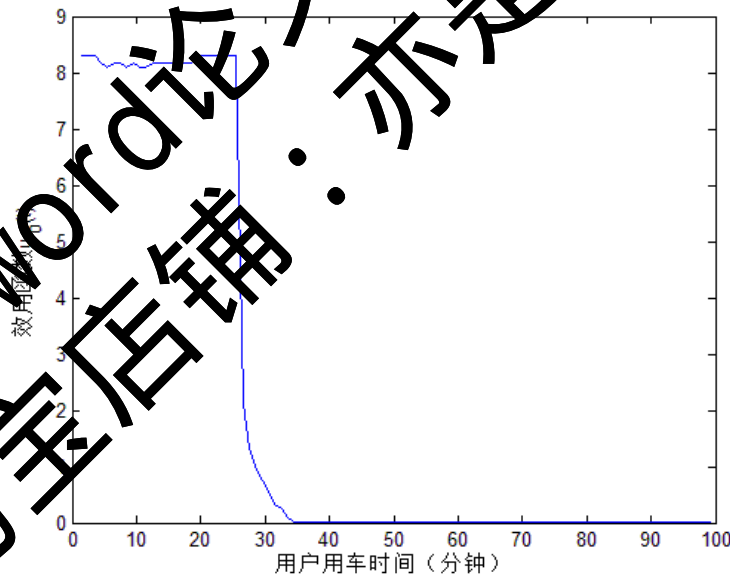


图 5：效用函数 $u_0(t)$ 的函数图象

从图 5 中可以看出，用户选择公共自行车出行的效用函数近似为一个矩形脉冲函数，其拐点约为 28 分钟。当出行时间小于 28 分钟时，用户更倾向于选择公共自行车出行；在 28 分钟之后，用户倾向于选择其他的交通工具，仅在随机因素的作用下选择

公共自行车。

综上，即可用效用函数 $u_0(t)$ 以及式 6.7 来描述用户选择公共自行车出行时间的分布。

6.5 借车卡累计借车次数统计分析

在这一部分的分析中，我们按照借车卡区分不同用户个体，对用户群体的需求模式进行探索。根据温州市鹿城区借车卡办理方法规定，办卡采用实名制，一张借车卡，只能借一辆公共自行车。^[2]因此我们假定用户和借车卡之间是一一对应的关系，即各位用户各采用一张借车卡。通过计算 20 天中各天使用公共自行车的不同借车卡（即借车人）数量，计算统计数据中出现过的每张借车卡累计借车次数的分布情况，我们可以看出借车人的偏好规律。

类似 6.1 部分中的分析，我们采用 MATLAB 统计每日使用公共自行车的借车人数量，如表 9 所示：

表 9：各天使用公共自行车的借车人数量

天数											1
j	1	2	3	4	5	6	7	8	9	0	
借车	1	1	9	1	1	1	1	1	7	4	
人数量	6840	7462	671	467	7985	8708	8887	0600	038	152	
天数	1	1	1	1	1	1	1	1	1	2	
j	1	2	3	4	5	6	7	8	9	0	
借车	1	1	1	1	1	1	1	1	1	2	
人数量	5097	8194	9531	9462	8075	1335	5412	5311	9188	0020	

通过累计各个借车人在 20 日中累计借车次数，我们做出如下频率分布图：

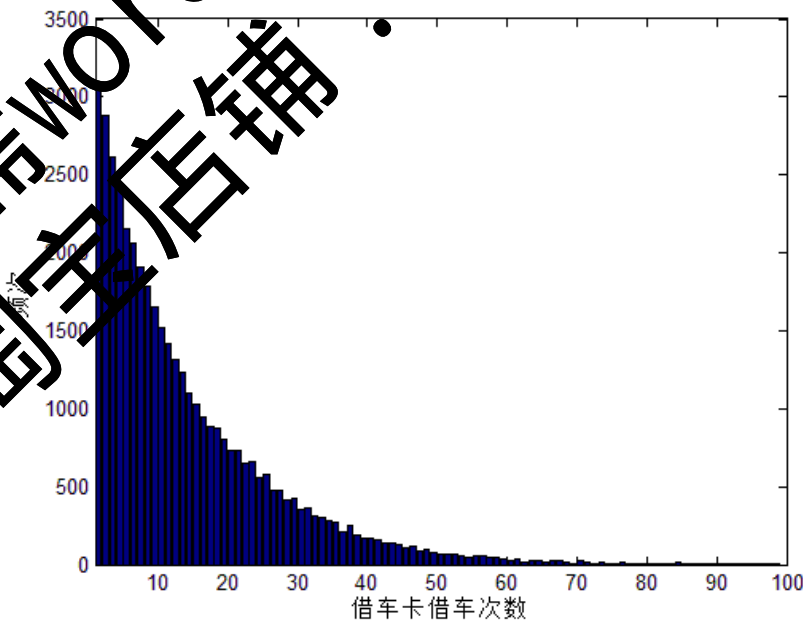


图 6：累计借车次数

横坐标表示各个借车卡累计借车次数，纵坐标表示各个次数出现的频数。从图 6 中我们可以看到，大多数借车人的借车次数在 30 次以下，随着借车次数的逐渐增加，相应借车次数的借车人数逐渐下降。我们依然采用简单统计方法分析各用车频次：

表 10：累计借车次数的区间频次统计

次数区间	[0,50)	[50,100)	[100,150)	[150,200)	[200,250)	[250,300)	[300,350)
累计频率	0.978202	0.999295	0.99989	0.999956	0.999956	0.999978	0.999978
次数区间	[350,400)	[400,450)	[450,500)	[500,550)	[550,600)	[600,650)	[650,700)
累计频率	0.999978	0.999978	0.999978	0.999978	0.999978	0.999978	0.999978

我们发现存在有离群值，样本点 99.2% 的用户借车次数都在 100 次以下，只有极少数人超过 100 次，根据常识判断这些值可能是统计错误，我们作为异常值去除。观察频数直方图，我们发现其分布近似于指数分布。基于这一假设，我们利用 MATLAB 拟合，得到拟合图如图 7 所示：

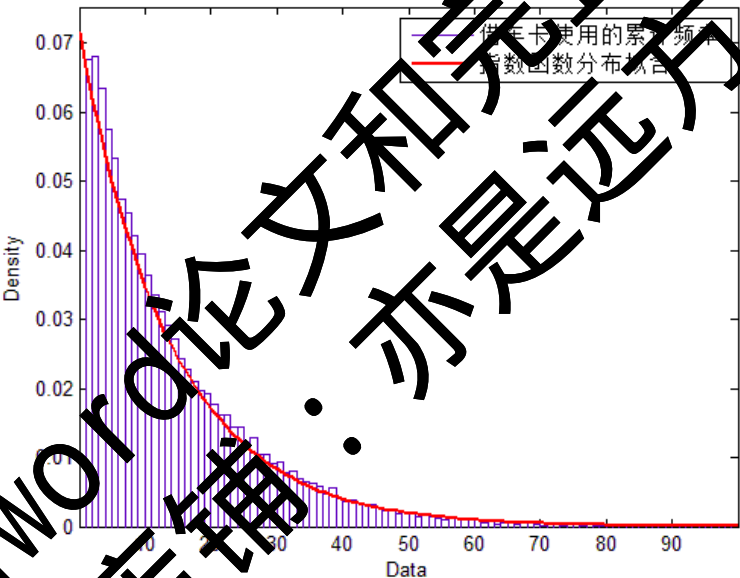


图 7：指数分布拟合图

从表 11 中可以看到，拟合结果良好，标准差 0.065，表明指数分布能够较好地拟合该借车频数的累计概率分布。

表 11：拟合结果

分布 (Distribution)： 指数分布 (Exponential)		
对数似然率 (Log likelihood)： -164938		
区间 (Domain)： 0 <= y < Inf		
均值 (Mean)： 13.9364		
方差 (Variance)： 194.225		
参数估计		
系数	估计值	标准差

Parameter	Estimate	Std. Err.
mu	13.9364	0.0654207

参数估计协方差

Estimated covariance of parameter estimates:

mu 0.00427987

七、地理信息数据的获取

7.1 公共自行车站点的坐标信息

对城市公共自行车的空间使用情况进行统计，需要已知城市地理信息以及公共自行车站点的地理位置信息。题目附件 2 中给出的鹿城区公共自行车站点分布图，仅描述了自行车站点的分布，并没有每个自行车站点的具体坐标。我们在温州市鹿城区公共自行车管理中心网站上找到每个自行车站点的位置，再在电子地图服务提供商（例如谷歌地图）网站上查询出该站点位置的坐标^[4]。为了表示方便，我们定义北纬 $28^{\circ}2'17.68''$ ，东经 $120^{\circ}38'19.13''$ ，以正南为 y 轴正方向，以正东为 x 轴正方向，以 1m 为单位建立直角坐标系。由于鹿城区面积较小，故可以不考虑地球曲率。将各站点的地理坐标映射到直角坐标系下，得到站点的坐标如表 12 所示。完整的表格参加附录中。

表 12：公共自行车站点的坐标（部分）

站点序号	站点名称	y 刻度	x 刻度
1	科技馆	8112.064	4378.759
2	温州大剧院	8155.842	4209.716
3	吴桥路加油站	3838.912	3882.536
4	银泰百货	4515.084	2475.662
5	星河广场	2453.85	916.104
6	绣山卫生院	8539.398	3735.305
7	区政府西	7972.286	3713.493
8	区政府东	8217.671	3697.134
9	小南门立交桥	3718.946	2333.884
10	市九中	4340.588	1134.224
11	鹿城区审批中心	3528.091	3909.801
12	桥儿头公交站	5414.829	3511.732
13	公共自行车中心	3757.117	3740.758
14	南浦医院	5453	4209.716
15	温州建国医院对面	4596.879	5005.854
16	金色家园	4580.52	4187.904
17	区政府西	3549.903	1706.789
18	区政府东	3626.245	1706.789
19	开太百货	4324.229	2044.875
20	南浦桥	5529.342	3762.57

将这些公共自行车站点的坐标标记在地图上，如图 8 所示。



图 8：公共自行车站点的在地图上的位置

7.2 城市道路地理信息

计算两个公共自行车站点时将会使用到城市道路信息。城市道路信息可以使用由现有的 GIS 信息，也可以通过图片格式的地图获得，使用图片格式的地图获得道路信息的方法主要包括颜色替换，形态学处理，道路生长细化等等，最终获得城市道路信息，用于两地之间的实际道路路径求取。颜色替换之后的城市地图如图 9 所示。



图 9：经过颜色替换得到的包含道路信息的城市地图

八、公共自行车时间、空间使用情况统计分析

8.1 站点距离的界定

对公共自行车站点的空间规划研究是建立在合理界定站点距离的基础上的。在本研究中，我们主要采用三种方法来对公用自行车站点间的距离进行刻画。

1) 欧氏距离

度量地理空间中的距离最简单的方法是欧氏距离 (Euclid Distance)，也称欧几里得度量、欧几里得距离，它是在 m 维空间中两个点之间的真实距离。本研究基于鹿城区的城市交通地图，属于二维空间中的欧氏距离，即为两点之间的直线段距离。

2) 直角边距离

我们注意到，欧氏距离虽然计算方法简单，但是不符合公共交通的一般特点。在城市规划和道路建设中，两个站点往往不能通过直线连接相互连通，因此，采用近似直角边距离和的方式近似求距离对现实的拟合度更高。但是，这种方法假设道路全部是“横平竖直”的，这也与实际有一定的差异，会存在计算偏差。

3) 城市道路实际距离

这一方法建立在实际的城区道路网络基础上，衡量的是站点之间城市道路的真实距离，因此是最为精确的。其优点是对空间距离的衡量拟合度最高，但缺点是操作难度较大。

8.2 借/还车的最短距离和最长距离

分别统计每天所有站点的借车和还车频次，并将两天的加和最大的一天作为公共自行车使用最大的一天。经统计发现，2012 年 11 月 1 日公共自行车的使用次数最大，借还总频次为 84441 次。本章以下问题均针对此日的数据进行讨论。

根据 8.1 节定义的距离计算方法，分别采用欧氏距离法和直角边距离法计算各站点之间的距离。根据数据记录，有的站点之间没有连接，将这些没有连接站点之间的距离视为 0。分别统计不同站点之间距离的最小距离，并对其进行排序，即可得到用车借还站点之间的最短距离和最长距离（非零距离，并且只统计时间在 1 分钟以上的借还车情况），如表 13 所示，完整的表格参见附录。

表 13：借车站点的最近与最远还车站点及距离（部分）

借车 站点	欧式距离				直角边距离			
	最近还 车站点	最近还车 距离 (米)	最远 还车 站点	最远还车 距离 (米)	最近 还车 站点	最近还车 距离 (米)	最远 还车 站点	最远还车 距离 (米)
1	2	220.6	12	5223.4	2	310.8	18	7159.8
2	6	472.3	142	6482.8	8	550.8	142	8081.3
3	2	312.0	148	4499.8	11	338.1	111	5578.4
4	71	259.4	72	5026.0	71	376.3	67	6472.7
5	175	382.1	114	8008.0	175	398.1	114	9602.7
6	8	324.0	140	6364.0	8	359.9	140	7067.1
7	7	245.9	131	5520.0	8	261.7	38	6549.1
8	7	245.9	133	5656.3	7	261.7	133	6778.1
9	43	123.0	73	5979.1	129	141.8	72	7399.7
10	21	239.0	115	5272.1	21	321.7	1	7018.0
11	3	312.0	116	6214.2	3	338.1	116	7001.7
12	78	208.4	160	4352.7	78	229.0	160	5720.2
13	11	284.7	72	5490.8	11	398.1	72	5954.7
14	79	359.9	157	5830.9	79	365.4	157	8010.5

15	75	531.2	144	4517.9	75	730.7	144	6309.1
16	166	319.3	58	4653.4	166	359.9	138	5692.9
17	18	76.3	76	5564.6	18	76.3	76	7028.9
18	17	76.3	114	6707.1	17	76.3	114	7639.7
19	71	215.2	113	5940.1	71	245.4	113	6925.3
20	23	270.9	160	4560.9	12	365.4	160	6085.5

8.3 借/还车频次最高站点

通过统计可知，2012 年 11 月 20 日的借车频次最高的站点为 42 号站点，还车频次最高的站点为 56 号站点。对这两个站点统计借还车时刻分布以及用车时长分布。对于用车时刻，将一天中的 24 小时划分成以 0.2 小时为间隔的若干时段，分别统计落入每个时段的借还车频次，即可用来反映借还车的时刻分布，统计的结果如图 10 所示。

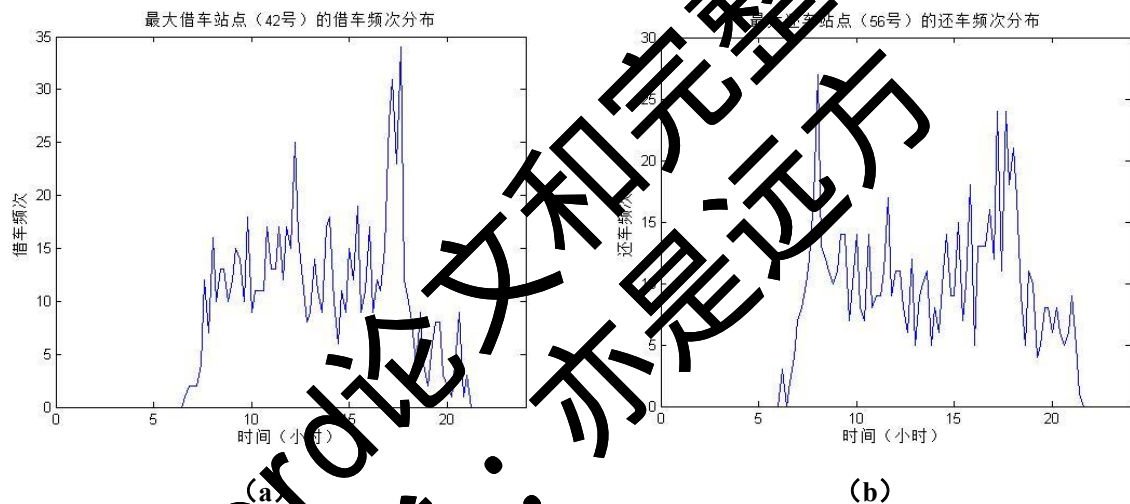


图 10: 最大借还车站点的借还车频次时刻分布

通过观察可以发现，借还车的时刻分布呈现一定的分布规律，例如 42 号站点的借车频次在 17:30 分左右出现峰值，在峰值左右的频次大于其他部分的频次，说明在该时刻前后借车行为比较集中；56 号站点的还车频次有两个峰值，分别出现在 8:00 和 17:00 左右，说明该站点在早晨和晚上都出现了还车比较集中的现象。

按照 6.2 节的方法来统计两个站点的用车时长，得到用车时长的分布如图 11 所示。根据 5.3 和 5.4 节的方法求解效用函数分布模型，得到的效用函数的图象如图 12 所示。

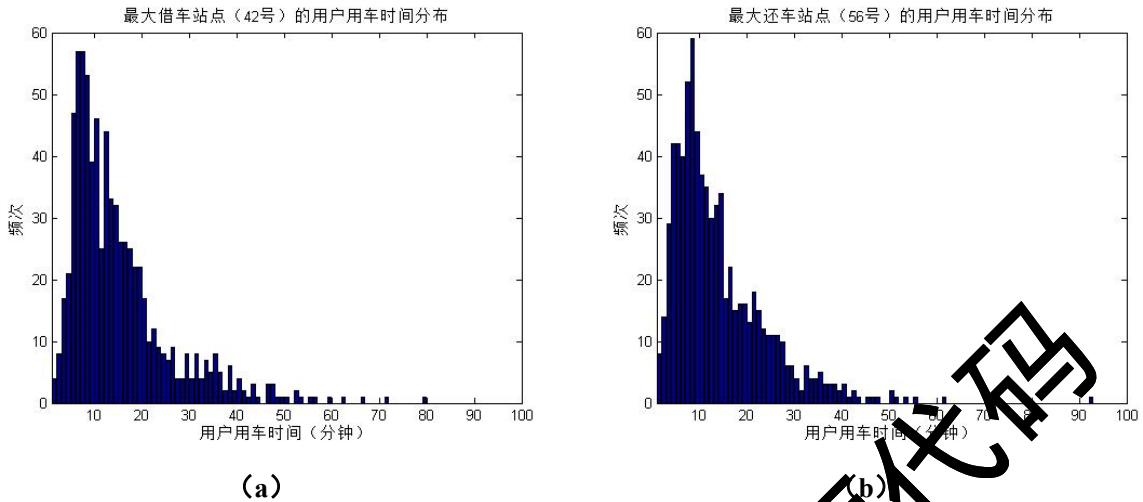


图 11：最大借还车站的用户用车时长分布

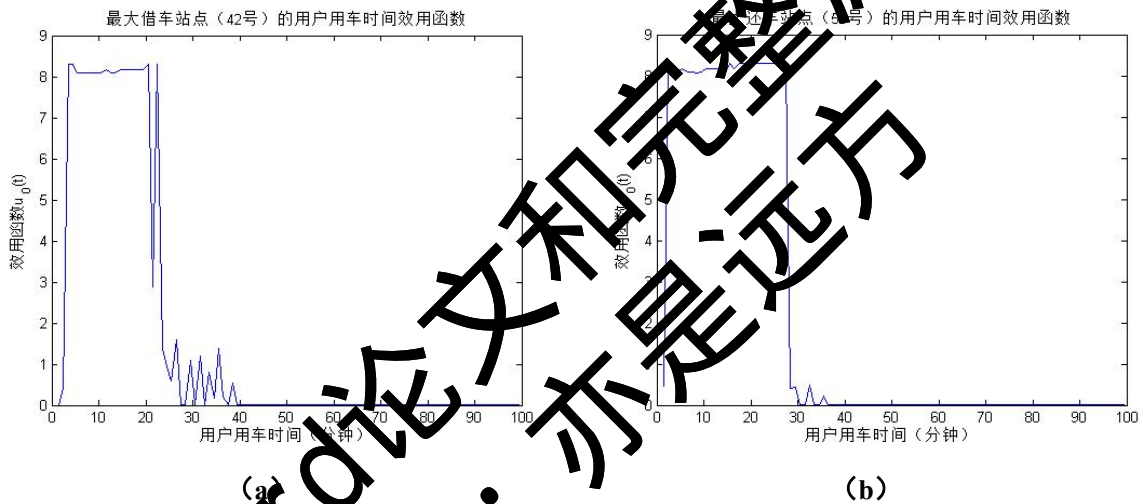


图 12：最大借还车站的用户用车效用函数

从图 11 和图 12 可以看出，42 号和 56 号站点的用车时长分布以及用户选择公共自行车的效用函数与 6.2 和 6.4 节中的总体情况大致相同，说明效用函数模型具有较好的稳定性。

4.4 峰值搜索算法

由 8.3 的分析可知，图 10 中的借车与还车时刻分布具有一定的规律，表现为存在一定的显著高于平均值的高峰时段。我们希望找到每个站点的这些峰值的位置，用来评价公共自行车系统的运行状况。

峰值一定是频率密度函数的极大值，但是极大值并不一定是峰值。随机因素造成的波动同样可能在频率密度函数中产生极大值。可以使用均值滤波的方法减弱随机因素造成的影响。均值滤波即使用某一点附近的一个小区间内的点的平均值来代替这个点的数值。图 13 为借还车时刻频率密度函数均值滤波的效果。

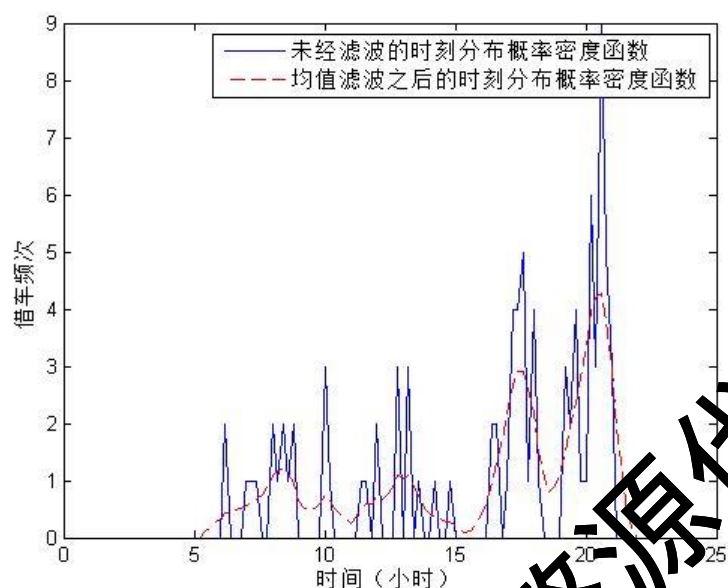


图 13:用车时刻分布频率密度函数的均值滤波效果

经过均值滤波之后，频率分布函数变得平滑了。对滤波之后的结果求极大值，定义当极大值大于整个取值范围的 60% 才为峰值时刻。当多个峰值之间的距离小于 2 个小时的时候，将它们视为同一个峰值，并用平均的时间代替峰值时间。使用 MATLAB 实现峰值搜索算法，其结果如图 14 所示。

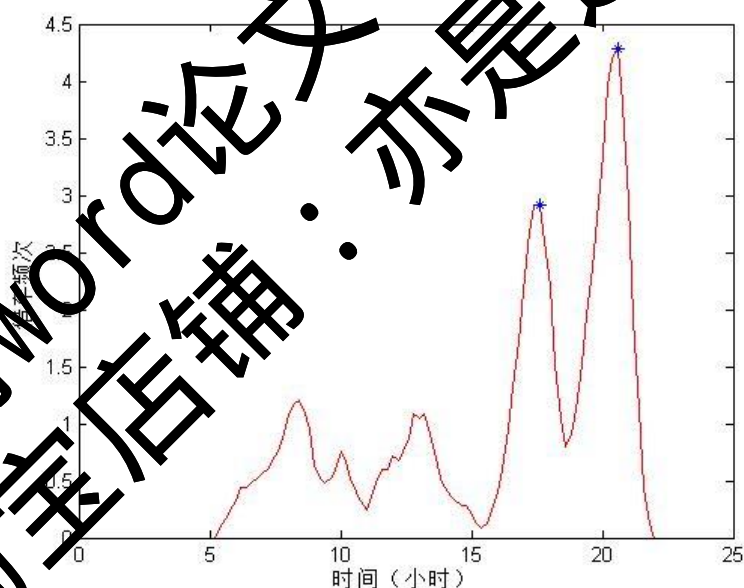


图 14: 峰值搜索算法的运行结果

8.5 用车高峰时段统计与归类

根据 8.4 节所述的峰值搜索算法对站点的借还车高峰时刻进行统计。得到各个站点的借车和还车高峰数据如表 14 所示，完整的表格见附表。

表 14: 各站点的借车和还车高峰时间表

站点序号	借车高峰	还车高峰
1	17:36, 20:36	19:23
2	8:12, 16:23	8:12, 15:35
3	17:11	8:12
4	7:48, 13:35, 14:23, 17:11, 19:11	9:12, 11:24, 12:47, 16:00, 17:36
5	17:11	10:12, 10:48, 18:00, 19:11, 19:48
6	8:12, 17:36, 18:36	17:23, 20:23
7	17:48	8:12
8	17:36	8:24
9	8:00, 13:47, 16:11, 17:48	8:24, 11:00, 15:47, 17:48
10	8:00, 15:24, 17:11	8:24, 17:23
11	17:11	8:12
12	7:48, 17:23	17:36
13	10:00	8:24, 10:00
14	7:48	7:48, 17:36
15	8:00, 10:36, 17:11	8:12, 17:23, 18:00
16	7:24, 12:35, 17:23	7:48
17	9:36, 17:23	8:12, 17:23
18	17:23	8:24, 17:23
19	8:00, 13:00, 15:35, 17:23, 19:23	8:24, 11:36, 12:47, 13:00, 15:47, 17:23, 19:11
20	8:00, 17:36	17:48

(续表)

8.6 高峰时段的聚类分析

可以使用聚类分析的方法对不同站点的高峰时段特征进行分类。聚类分析的思路是将参数空间中的点分成若干类，使得每类中的点的差异最小。经典的聚类方法如 K-means 算法，通过构造一个类中心，通过不断迭代的方法改变类中心的距离以及分类结构，使得同一类中的点距离类中心的距离之和最小。

由于不同的站点具有的峰值数目不同，即所在的空间的维数不同，因而无法直接使用 K-means 算法进行聚类，此时可以将不同峰值数目的点折算成具有相同数目的等效峰值的点。例如，设等效峰值数目为 20 个，对于不足 20 个峰值的站点，需要将一些峰值时间复制或若干倍。峰值时间复制的数目与峰值频次的大小成正比。

经过折算之后，即可使用 k-means 算法对 20 维等效峰值时刻空间进行聚类，使用欧式距离作为空间中点的距离，设置类数目为 5 个。得到结果后将同类的站点使用相同的颜色标号在地图上。借车站点高峰时刻聚类分析的结果如图 15 所示，还车站点高峰时刻聚类分析的结果如图 16 所示。

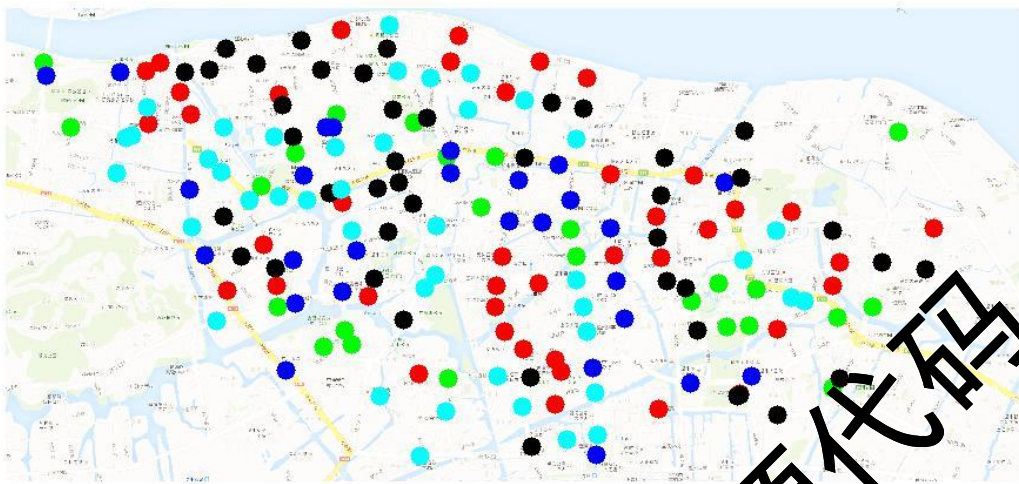


图 15: 借车站点高峰时刻聚类分析的结果



图 16: 站点高峰时刻聚类分析的结果

4. 其他信息和公共自行车系统设置评价

9.1 潮汐现象

我们观察上面的计算和统计结果，注意到最大借还车站点的借还车频次时刻分布（即图 10）显示，42 号站点的借车频次在 17:30 分左右出现峰值，高于最大自行车提供量即 30 辆，出现供给不足的“晚高峰”，而 56 号站点公共自行车也存在早高峰和晚高峰两个较为明显的峰值，供给非常紧张，这一现象也符合自行车交通的基本规律。而其他时段则不同程度出现低谷（例如，用户需求量低于 5 辆）等现象。不同的时间区间用户的用车需求波动较大，体现出“潮汐现象”的特征。^[1]

潮汐现象主要是指租车或还车不能持衡，出现“供不应求”或者“供过于求”的不均衡

现象,在一定时间段内租车量持续高于或者低于还车量,并且以一定时间为循环周期重复出现,严重影响了系统的运行效率。一系列新闻报道已经从侧面证实了这一困难,潮汐现象是自行车系统的设计和运行中普遍存在的规律。²从我们的统计分析来看,部分地段服务点的公共自行车在使用过程中有明显的峰值和低谷,基于用户的频率分布也有显著的差异,有较为明显的自行车需求和潮汐现象。表明鹿城市的公共自行车系统的效率依然有提高空间。

9.2 潮汐现象的表现

潮汐现象主要表现为:①租赁点还车数大于借车数,造成停车桩已满而无法还车的现象;②租赁点的借车数大于还车数,造成自行车全空,没有自行车可以租借的现象^{[5][6]}。这两种情况均影响了公共自行车的可持续使用。潮汐现象有一个特点就是当天内出现借还逆转,并且以日为周期重复出现。

城市边远地带的大型居住区、大型的工业园区、交通枢纽站、商务集中区等类型的服务点更容易出现潮汐现象,这些地区的用地功能复合度较低,城市居民出行需求趋同,出行时间集中,短时间内对车辆的需求量一面波动频繁是潮汐现象形成的主要原因。在本文的研究框架下,温州市鹿城区位于浙江省东南部、温州市中部,是温州市的政治、经济、文化中心,潮汐现象相对较弱,而城市公共自行车系统在其他城市的推行中,不可避免地会遇到不同程度的潮汐问题。

9.3 潮汐现象的解决

解决潮汐现象的最好方法是实现自行车的高效率、低成本调度,提高公共自行车流转量,减少因空架或满架造成的阻碍使用的情况,及时满足供给、需求的相对均衡,保证市民的正常使用的。目前我国公共自行车系统的应用尚处于初级阶段,调度模式落后,主要利用人工巡查的方式对服务点进行监控。目前的调度主要是静态调度,一般选就近的原则进行服务点间的调配,未形成系统科学、结构清晰的调度模式。这也是亟待解决的问题。在10.2部分中我们会有进一步的阐释。

十、公共自行车服务系统的其他运行规律和政策建议

在前面的研究中,我们从公共自行车的需求规律、用户的效用水平和站点的地理空间配置等几个角度对公共自行车服务系统的规律进行了分析和评价。除这些方面之外,公共自行车服务系统作为一个复杂、庞大的体系,还涉及许多其他方面,体现出一些独特的运行规律。我们选取有代表性的两个方面简要分析。

10.1 站点选址的社会福利最大化问题

城市公共自行车系统建设不仅是解决城市交通问题的重要手段,而且作为基础设施建设的重要组成部分,这一系统的平稳、可持续运行还有赖于良好的运营商业模式和策划。鹿城新闻网显示,鹿城要推行公共自行车系统,鹿城捷信小额贷款公司 20 名股东经过商议,决定承担前期建设的所需全部资金 1200 万元,并出资 500 万元设立“公

² 参见新闻报道:“湖东公共自行车频现“潮汐”现象”

<http://jsnews.jschina.com.cn/system/2012/04/09/013101671.shtml>, “市民指河西公共自行车早晚高峰借车难还车难” <http://jiangsu.sina.com.cn/news/m/2013-05-23/074956286.html> 等

共自行车发展专项基金”，用于下阶段的扩建、维护和管理等后续费用。³政府项目的市场化运作已经成为当今社会公共品提供的主流趋势。作为公共物品，公共自行车的站点设置应当建立在政府财政预算的基础上实现收益和社会福利最大化。我们构建简单的社会福利模型对这一问题进行分析。

社会福利最大化模型受制于政府的财政预算约束，最大化社会收益和投资公共自行车企业收益。相关的文献研究参见何博（2009）^[6]的站点选址模型：

$$\begin{aligned} \max W &= \sum_{i=1}^s \left(\frac{365P_i}{\gamma} N \right) + \sum_{j=1}^l (g_j Y_j N + q_j Y_j N \sigma) - \sum_{j=1}^l (t_1 X_j + t_2 Y_j + t_3 Y_j N) \\ \text{s.t.} \quad &\sum_{j=1}^l w_{ij} Y_j \geq 1 \\ &\sum_{j=1}^l w_{ij} Y_j \leq T \\ &Y_j \in \{0, 1\} \end{aligned} \quad (10.1)$$

s, l 分别为需求点的数目、备选租赁点的数目；

g_j, q_j 分别为备选租赁点 j 每年的商业收入、每年的政府补贴；

t_1, t_2, t_3 分别为基础设施单价、每个租赁点建设费用以及运营费用；

σ 为运营模式分类系数哑变量, 完全市场化模式取 0, 其他模式取 1；

γ 为公共自行车高峰小时系数；

T 为租赁点个数的最大规模；

X_j 为备选租赁点 j 的基础设施数量；

Y_j 如果在 j 处建设租赁, 取值为 1, 否则取值为 0；

P 为需求点 i 的出行者所需要支付的租赁费用；

w_{ij} 如果需求点与备选租赁点的距离小于最大值则取 1, 否则取 0；

对于基于预算约束的规划模型求解可以得出社会福利最优解。更广泛经济意义上来说，该机制的设置存在两个维度的问题：①权衡问题（trade off）：建立更为广泛的自行车站点的福利增进、经济利益与建造租赁地点成本和财政补贴、预算约束的权衡。这一机制设计可以用行为博弈论的方法求解。②委托-代理问题：以政府为主导的城市公共自行车系统建设市场化，面临作为委托方（principle）的政府与作为代理方（agent）的承建企业之间的利益分歧，政府的目标是社会效益最大化，而企业是以股东利益最大化为目标的组织，其主要目标是利润最大化，这可能与社会最优的目标之间存在一定冲突。

10.2 公共自行车系统运行的调度机制

公共自行车“想借时难还亦难”问题可以通过有效的自行车调度来解决。在推广城

³参见鹿城新闻报道：<http://www.wzbicycle.com/article.asp?Action=View&ArticleID=260&Catalog=2>

市公共自行车服务系统的同时，必须建立起配套的调度系统。公共自行车行系统主要通过后台控制中心和区域调度中心来完成车辆调配。通过各服务点自行车停放情况的跟踪统计,当车辆数低于某一个阈值时，启动调度，并对站点实行实时监控，保证调度最优化。调度系统主要包括两个方面的作用：①日内调度：每个服务点的公共自行车租赁需求存在季节差异和日内时间段差异；②需求低谷期的车辆入库：当出现极端恶劣天气时，将多余的车辆进行入库管理能够提高车辆的使用寿命。^[5]区域调度的层级实现方法如下图：

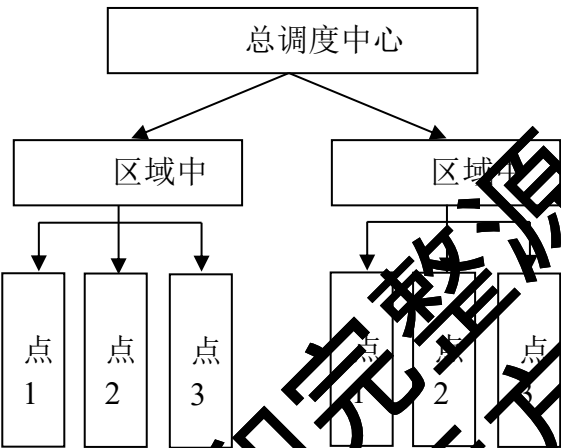


图 17：自行车区域调度系统

通过构建基于模糊时间窗的用户满意度函数、惩罚函数，叶丽霞（2013）^[5]系统地采用了基于 A-Priori 优化的方法和实时再优化方法计算最优调度方式和规模。另外，节约算法、插入算法、遗传算法、模拟退火算法等方法都可以用于解决最优调度问题。

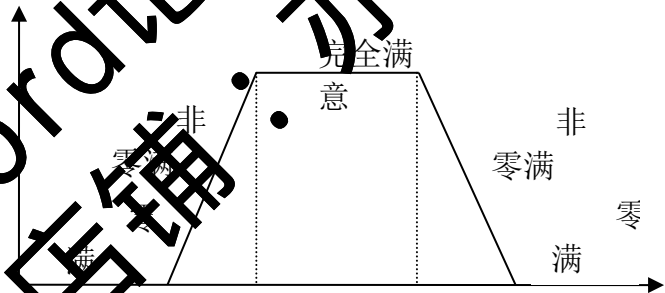


图 18：模糊时间窗的用户满意度区间图

十一、 结论

城市公共自行车为代表的新型公共交通计划实施后，城市交通拥堵现象带来了有效的缓解，为“最后一公里现象”找到了行之有效的解答^[7]。良好的公共自行车服务系统设计有助于这一举措的顺利实施和可持续发展。本文建立在温州市鹿城区的公交系统基础上对于城市公共自行车系统的客流量、供求关系和分布规律进行了研究，在总结数据的基础上对公共自行车系统一般规律进行探索，并提出了一些问题和可行的解决方法。本文的结论主要揭示了以下几个方面的问题：

- 1) 通过原始数据的初步整合处理、剔除异常值，我们求出了借还车频率、借车人

分布,并通过拟合方法发现,借车人的借车次数普遍在 30 次以下,并且群体的借车次数的近似服从指数分布。

2) 通过建立用户效用函数,我们可以求解城市居民采用公共自行车出行的偏好程度随出行时间的变化规律。我们发现,用户选择公共自行车出行的效用函数关于出行时间的变化近似为一个矩形脉冲函数,当出行时间少于 28 分钟时,用户更倾向于选择公共自行车出行;在 28 分钟之后,用户倾向于选择其他的交通工具,仅在随机因素的作用下选择公共自行车。

3) 我们采用了三种方法界定站点距离——欧式距离,直角距离和实际地图距离,我们借助于温州市实际电子地图通过 MATLAB 进行相应处理,获得市级道路信息,从而成功解决了距离的度量问题。

4) 我们的统计分析来看,部分地段服务点的公共自行车在使用过程有明显的峰值和低谷,基于用户的频率分布也有显著的差异,有较为明显的自行车需求和潮汐现象。另外,站点设置的社会福利经济学分析以及配套的调度机制设计也是公共自行车系统运营中需要解决的重要问题。

本文的研究对政府建设城市公共自行车租赁系统具有一定的参考价值。公共自行车系统是一个复杂的系统,在城市的运用过程中会涉及多个角色的相互作用,包括政府、企业、出行者,也会涉及到地理、政治、经济、交通等非常广泛的领域^[8]。而我们的这一研究尚粗浅,还有许多没有涵盖的部分,建立在社会福利、用户需求等多方权衡下的最优化的公共自行车系统的构建将是未来可以继续探索的方向。

十二、参考文献

- [1] 鲍娜,城市公共自行车租赁点选址决策及调度模型研究,长安大学硕士学位论文,2012。
- [2] 温州市鹿城区公共自行车管理中心网站, <http://www.wzbicycle.com>, 2013.09。
- [3] 单晓峰,城市自行车交通合理方式分担率及其路段资源配置研究,东南大学博士论文,2007。
- [4] 谷歌(google)地图, <http://ditu.google.cn/maps?hl=zh-CN&tab=w1>, 2013.09。
- [5] 叶丽霞,城市公共自行车调度系统研究,南京理工大学硕士学位论文,2013。
- [6] 孙楠,城市公共自行车系统的应用研究,西南交通大学硕士学位论文,2012。
- [7] 李正浩,城市公共自行车租赁站远期发展规模分析 [J][J], 交通节能与环保, 2010, 2, 015。
- [8] 王志高,刘皓,谢建华等,欧洲第三代公共自行车系统案例及启示[J], 城市交通, 2009, 7(4)。