# 共享单车的成本分析与网络分布

摘要

关键词

Abstract

Keyword

## 目录

## 第一章 引言

### 1.1 研究背景

[共享单车](https://baike.baidu.com/item/%E5%85%B1%E4%BA%AB%E5%8D%95%E8%BD%A6)是指[企业](https://baike.baidu.com/item/%E4%BC%81%E4%B8%9A)在[校园](https://baike.baidu.com/item/%E6%A0%A1%E5%9B%AD/1600458)、[地铁](https://baike.baidu.com/item/%E5%9C%B0%E9%93%81/21266)站点、[公交](https://baike.baidu.com/item/%E5%85%AC%E4%BA%A4)站点、[居民区](https://baike.baidu.com/item/%E5%B1%85%E6%B0%91%E5%8C%BA)、[商业区](https://baike.baidu.com/item/%E5%95%86%E4%B8%9A%E5%8C%BA)、[公共服务](https://baike.baidu.com/item/%E5%85%AC%E5%85%B1%E6%9C%8D%E5%8A%A1)区等提供自行车单车共享服务，是一种分时租赁模式。共享单车具有操作方便、低碳环保等优点，解决了“最后一公里”出行的问题，，并在一定程度上缓解了交通问题。然而，在许多城市，共享单车的过量投放一定程度上加重了城市的交通负担，影响了城市环境的美观，同时造成了城市用地面积和厂家维修成本的浪费。2016年以来，涌现的共享单车品牌有15-20家，已在全国30多个城市投放，全国累计投放超过1000万辆，目前互联网租赁自行车的用户数量超过了1亿人次。根据调研，北上广深四个一线城市的人口数共计7132万，现有的共享单车投放量达108万辆，若未来继续投放共享单车，可以达到181万辆，但这远远消耗不了3000万的单车产能。即便是共享单车下沉到二、三、四线城市，一共需要约1745万辆车，这距离ofo与摩拜今年3000万辆的订单总和仍有不小的差距。而ofo和摩拜两家公司最新一轮融资就合计高达13亿美元。长此以往，共享单车市场将产生大面积亏损。

### 1.2 研究意义

节约社会资源；帮助共享单车产业更好运营。（太少）

### 1.3 论文组织结构

## 第二章 共享单车与网络的成本分析

### 2.1 关于成本的影响因素

#### 2.1.1 假设条件

1）城区内人口密度视为均匀的。

2）居民从出发地选择自行车出行的概率一致。

3）用户会选择离自己最近的一个投放点。

4）用户的目的地是投放点，没有新投放点生成。

5）每个类型的投放点影响范围是相同的。

6）车辆损坏率只与数量有关。

7）只考虑“大投放点-中型投放点-自然生成的小点”两个层级。

#### 2.1.2 单车成本的影响因素

1）购买自行车的成本。

2）投放点占地的成本。

3）维护成本。包括维修自行车、管理投放点自行车的成本。

4）运输成本。包括将自行车运至投放点和投放点间相互运输的成本。

## 第三章 不考虑用户密度的简单网络成本分析

### 3.1 网络覆盖的近似图形选择

假设用户能接受的从出发地到达中大型自行车站点的最长时间为t，用户步行速度为v，则中大型站点的服务半径固定为为L=vt。

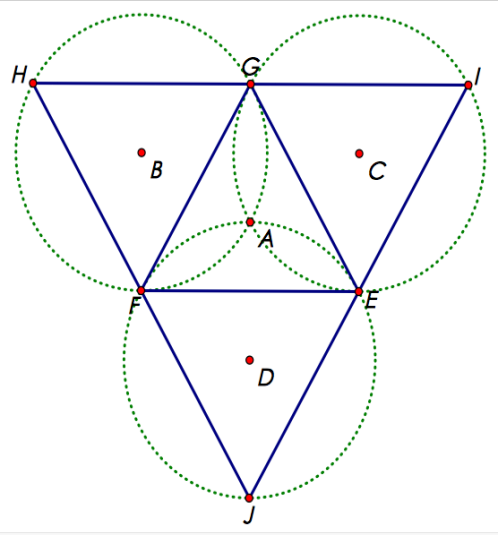
但由于圆无法密铺平面，我们需要将站点服务范围的形状优化为可以密铺的图形。

已知可以密铺平面的正多边形有三角形、正方形和六边形。

令站点利用率A=单位图形内不在其他任意站点服务范围内的面积/单位图形

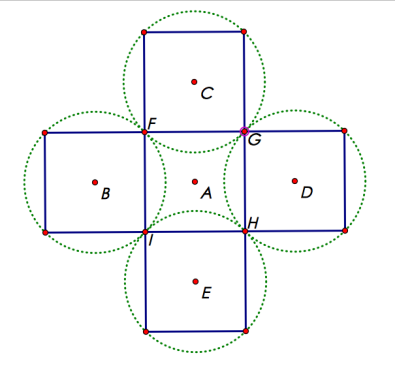
我们将采取对站点利用率最高的形状。

#### 1 三角形



若将站点服务范围的形状近似成正三角形，如图，则单位三角形EFG内不在其他任意站点服务范围内的面积为0，故三角形对站点的利用率A1=0。

#### 2 正方形

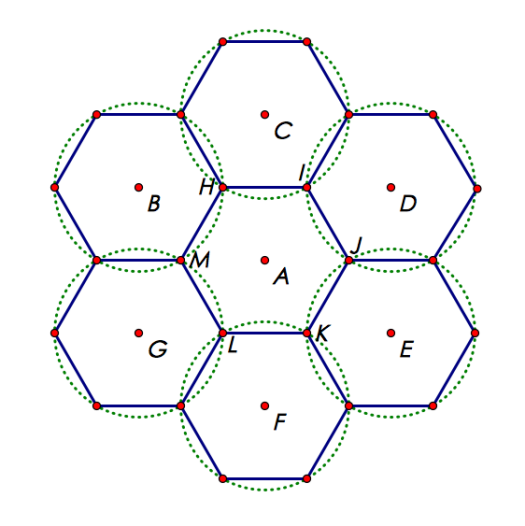


若将站点服务范围的形状近似成正方形，如图，则单位正方形FGHI内不在其他任意站点服务范围内的面积为



故利用率为

#### 3 六边形



若将站点服务范围的形状近似成正方形，如图，则单位六边形HIJKLM内不在其他任意站点服务范围内的面积为



故利用率为



综上，A3>A2>A1，故选择将服务站点形状优化为正六边形。

### 3.2 基于六边形的网络建设与成本

六边形

假设六边形边长为L公里

一个大站或中站的服务范围成为单位六边形。

#### 1 用户距离成本

用户距离成本即用户从出发地步行至自行车站点的成本。

由于假设人口密度均匀、所有居民选择自行车出行的概率一致，用户距离成本与用户到自行车站点的步行距离成正比。步行距离近似为直线距离。

设一个中型或大型站点服务范围内的用户距离成本为C1，C1等于固定系数乘以服务范围内所有点到站点的距离之和，即固定系数乘以正六边形内所有点到中心的距离积分。

将正六边形六等分为六个边长为L的正三角形，如图 。则六边形内所有点到中心的距离积分等于六个正三角形内所有点到一个顶点的距离之和。

以正三角形OAB为例。设C在边OA上，D在边OB上，CD∥AB。则CD上所有点到顶点O的距离之和等于正三角形OCD的面积。设正三角OCD边长为x（x∈[0, L]），则

故三角形OAB内所有平行于AB的线段上的点到顶点O的距离之和等于该线段端点与O组成的三角形的面积，即线段一端点到O的距离的平方的四分之根号三倍。

因此，三角形OAB内所有点到O的距离之和等于三角形OAB内所有平行于AB的线段上的点到O的距离之和，即

正六边形ABCDEF内所有点到中心点O的距离之和为

设α为用户距离成本固定系数，则单位六边形内的用户距离成本C1可表示为



#### 2 运输成本

开始运行后共享单车系统的熵会增加，故需要在固定周期内通过运输调节单车的不均匀分布。

##### 2.1 大型站点与中型站点间的运输成本

由于大型站点与中型站点采用大型交通工具运输，且站点固定，故不考虑运输单车数量，只考虑运输次数和距离。

因每个中型站点周围距离最近的大型站点有三个，每个大型周围距离最近的中型站点有六个，相当于每个大型站点要定期接收6\*1/3=2个中型站点的多余单车数量。由于大型站点与中型站点之间的运输成本只与运输次数和距离相关，为了尽可能减少运输次数，不采用将中型站点的多余单车三等分，分别运送至相邻三个大型站点的方法，而直接由一个大站负责两个中型站点的多余单车。所以每个周期需考虑一个大站与两个相邻中站的运输成本。

设一个大型站点与其相邻两个中型站点间的运输成本为C2.1，一个大站与相邻中型站点间距离为√3L公里，运输价格为β元/公里，每周期内大站与其服务范围内的每个中站之间运输次数为n。



##### 2.2 单位六边形内的运输成本

假设自然生成的小站点是均匀分布的，C2.2为单位六边形中心站点与其服务范围内自然生成的小站点间的运输成本，每辆单车的运输价格为χ元/公里，每周期一个中型站点的服务范围内共需运输m辆单车。

参考1可得：



##### 2.3 单位运输成本

由2.1可得，一个单位的站点系统包括一个大站和其周围的两个中型站点，共三个单位六边形。

故单位运输成本包含一个大站两个中站间的运输成本与三个单位六边形内部的运输成本。

即



## 第四章 考虑用户密度的简单网络成本分析

假设每个点的人口密度与该点到市中心的距离成正比，其比例系数为δ。

以某单位六边形中心为原点建立坐标系.设市中心坐标为(a, b)，六边形内某点坐标为(x, y)。则该点的人口密度为



该点用户到对应单车站点的距离成本为

其中α为距离成本的固定系数。

于是可以求得单位六边形内的距离成本。

## 第五章 离散用户密度与浮动网络站点

### 5.1 公式推导

在第四章中，本文讨论了基于连续用户密度与固定网络的成本。但在现实环境中，由于统计困难，求出对数据贴合的函数更为复杂，很难得到一个连续的密度函数。同时，基于对现实生活的观察，由于受到各种因素的影响，共享单车的站点并非是一个理想的六边形分布情况：对于用户密度较大的地区，会分部更密集的站点提供服务，以降低用户成本；而在偏远地区，则很少有站点。在本章中，本文将使用离散型随机变量对用户进行模拟，以获得更为真实的结果。同时考虑浮动而非固定的站点，通过调整站点的分布，降低高用户密度地区的成本。

根据第三章和第四章的部分结论，对于一个大型站点下覆盖的多个小型站点，本文假设其成本是与自行车数量成正比，与距离大型站点的距离成正比，同时包含一个维护的固定成本。即：

(5-1)

此处N表示站点的车辆数量，D表示站点到大型站点的距离，F为固定成本。

同时考虑到站点影响范围不会太大，因此假设每个用户的距离成本正比于用户最近的站点的距离，即：

(5-2)

其中Mmin是该用户到最近的站点的距离。

对于某一站点i，带入5-1和5-2式，该站点的总成本为：

(5-3)

其中P为用户数量。

对5-3式求和，则所有站点的总成本为：

(5-4)

其中T为站点数量。

为计算方便，令总站为坐标原点(0, 0)。假设站点i的坐标为(xi, yi)，则

同样的，假设用户j的坐标为(xj, yj)，且用户j距离最近的站点为i站点，则：

### 5.2 数据生成

#### 5.2.1 用户概率密度

由于无法从网络上的公开资料中获得详细的人口密度或用户密度数据，这里本文根据作者实际生活经验结合地图假设用户概率密度。

首先从网上截取2km\*2km的城区地图，大小为1000像素\*1000像素，即每个像素点为2m\*2m。然后在图片上绘制20\*20的网格，并根据人口密度为网格划分权重。这里假设了用户密度和人口密度是线性相关的。将人口密度按0~5划分权重，根据经验为每个格子打分，设为wi。最后求出所有权重值的总和W，用每个格子的权重除以权重的总和，就可以看做每格近似的用户概率pi，除以每格的面积就是用户概率密度。实际计算时只使用每格的用户概率：

#### 5.2.2 用户位置生成

得到每个格子的用户概率数据后，需要用3个0~1的随机数来决定用户的位置。第一个随机数r1用于计算用户所处的格子i，这里采用累计概率的方法，即：

1. 第三个随机数r2、r3则用于计算用户的坐标，简单以格子左下角为原点，右上角为(1, 1)点，以(r2, r3)作为用户在格子内的坐标点，再加上格子本身的坐标即可：

#### 5.2.3 参数假设

### 5.3 模拟算法

在5.2中，本文讨论了如何在x-y平面上生成符合实际的用户点用以模拟。本节的主要内容是以5.2生成的用户点为输入，设计算法以计算成本最小的站点分布。站点位置分布主要包括两个部分：站点的数量和站点的位置。这里本文介绍K-means算法及其改进算法ISODATA，并结合上文公式对算法进行改造，计算出较优站点分布。

#### 5.3.1 K-means算法

K-means算法是一种简单的聚类算法，常用于机器学习，对数据集进行分类。K-means算法会将映射到空间中的点按距离分成K类，具体算法[[1]](#footnote-1)是：

1. 从数据集中随机选取*k*个数据点作为聚类质心。
2. 重复一下过程直至结果收敛：
   1. ，计算其所属的类：
   2. 对于每个类*j*，重新计算其质心:

可以看到K-means算法和上文中需求非常相似。通过将距离与成本进行映射，分类点映射站点，可修改为本文所需算法。不过K-means算法的缺点在于分类数K是人为规定的，同时最终分类K点的位置受到初始K点影响。在K-means改进算法K-means++中，通过提高各类初始质心点间的距离，减少初始选择点对结果的干扰。但这种算法的分类数仍是固定的，不符合本文的需求，因此需要进一步的改进。

#### 5.3.2 ISODATA算法

ISODATA是一种对K-means的优化算法。ISODATA算法相比于K-means算法，添加了分裂操作和合并操作，并通过这两个操作增加或减少类的数量。

相比于K-means算法，ISODATA算法在进行一轮计算后，对每一类会计算其标准差，如果某一维度的标准差大于一个设定值S，则对这个类进行分裂操作。分裂操作的具体算法为：

1. 计算类的标准差向量：

每个分量为：

其中为类的样本个数，为向量的维数。

1. 求出每类的最大标准差分量,
2. 如果，则把该类分为两个新的类，类的质心分别为：

如果两个类的质心小于某个设定值D，则对着两个类进行合并操作，其具体算法是：

1. 计算类的距离,
2. 如果，则把两个类合并为一个新的类，类的质心为：

其中 分别为两个类的样本各数。

在ISODATA中对类仍有一些其他的限制，但与本文需求并不相关，故不在此多做介绍。

有了分裂操作和合并操作，算法就可以对站点的数量进行修改。将成本映射为标准差，即可以将ISODATA算法应用于增加或减少站点的数量。结合基本的K-means算法，以5.2中生成的用户点作为输入，进行编程计算最佳的站点位置。程序代码部分请参见附录。下面将展示计算的结果。

### 5.4 模拟结果

## 第六章 结论

附录

1. 参考文献

K-means聚类算法的三种改进介绍与对比<http://www.cnblogs.com/yixuan-xu/p/6272208.html>

1. 代码

1. 具体算法的数学表达参考了《K-means聚类算法》一文，http://www.cnblogs.com/jerrylead/archive/2011/04/06/2006910.html [↑](#footnote-ref-1)