浙江大学

数 学 建 模 期 末 论 文

题目 基于遗传算法的骑电单车空间聚类与调配模型

所在小组 第 25 组

指导老师 陈 剑

所在学院 控制科学与工程学院

所在专业 自动化（控制）

完成时间 2018 年 5 月

目 录

基于遗传算法的骑电单车空间聚类与调配模型

曾之宸　林润泽　冯凯琦

（浙江大学控制科学与工程学院，浙江 杭州　310027）

摘　要：

共享电单车的与传统单车的运营模式有很大区别，因此，共享电车对还车点设置与车辆调配的要求更高，需要通过建立数学模型得到合理的规划策略。

本文旨在通过分析骑电单车的实际位置数据，聚类出最佳的还车点空间位置分布，以及局部区域内还车点之间骑电单车的调度策略。考虑到骑电单车数据量大、变量耦合性强，用传统数学规划方法不能有效解决，通过 K-Means 聚类分析算法和基于轮盘赌法的遗传算法，来建立其空间聚类与调配模型。

根据所得骑电单车的实际运行数据，分析了各种聚类算法的合理性和可行性，最终采用了基于 K-Means 的空间聚类模型对选址问题进行求解。聚类的目标函数（SSE）采用了基于欧式距离的度量，对第 k 个质心 c_k 求解，寻找不同 k 值情况下的最小的 SSE。

通过指定典型 k 值，得到了特殊情况下的还车点经纬度坐标设置，验证了模型的准确性和适用性。为了求解最优的聚类 k 值，更新优化了 K-Means 空间聚类模型，通过代价函数自适应地确定最佳聚类数目 k 值。其中，代价函数与所有单车坐标同第 k 个聚类点的距离之和有关，同时与聚类 k 值有关。

通过遍历求解每个 k 值对应的代价函数值，选择最小代价对应的最优 k 值，即为杭州城区骑电单车还车点设置的最优解。进一步，采用可视化的方法，将聚类中心点与聚类区域在经纬度图中显示，得到了最终的结果。

对于骑电单车调度策略模型，经过问题的简化与抽象，可归结为 Hamiltonian Cycle 的最优路径规划问题。为了求解骑电单车调配的最优路径，采用了遗传算法作为基本的方法，同时将遗传算法与代价函数进行了结合，得到 TSP 问题的初步求解方案。

最后，综合运营成本和预期收益等因素，通过对代价函数进行一系列的改进与优化，利用 MATLAB 等工具绘制出了骑电单车调配的最优路线图。

关键词：K-Means 算法；聚类分析；遗传算法；调配模型

1 问题背景与重述

（1）研究背景

随着国内外绿色环保、低碳出行理念的诞生，单车共享这一新兴项目正在蓬勃发展。单车共享体现的是人们对生态环保及自身健康的又一次觉醒。现阶段，公共自行车数量最多的城市是中国杭州市，总计 6 万辆自行车可供市民使用。

但是，相较传统人力车辆长距离骑行对人们体力带来的挑战，骑电单车在 3 千米以上用车需求的市场，拥有绝对优势——省力，骑行者不必再为体力和骑行环境而费心。骑电单车以“共享电单车”为核心产品，基于智能硬件（LBS 地理位置）和 App（O2O 体验）开发，是目前国内前沿的分布式移动分时出行解决方案。享骑电单车采用分布式点对点借还车的经营模式，通过共享电单车连接城市内的小区和交通枢纽、写字楼、商圈以及学校、医院等人群聚集地，形成一张巨大的交通出行网。用户还车时，不可随意停放，而是需要到达还站点进行还车。独特的“点对点”模式，可以使享骑更好的在后台监控车辆的信息，并利用成熟的物联网技术，及时对车辆进行有效的调度，为用户带来更好的服务。此外，集中管理也有利于为政府部门管理市容市貌提供良好的协助，很大程度上防止乱停乱放的现象。

如果说共享单车解决了最后 1 公里的代步需求，那么电单车就解决用户 2-5 公里的出行需求的方式，二者相辅相成，将成为网约车、公交地铁之外的一种补充。但是骑电单车和共享单车的运维体系完全不同，共享单车的运维体系不可复用在骑电单车领域，尤其是换电运营，骑电团队的核心能力在于换电运营，这是比较复杂且具备竞争门槛的一件事。目前，骑电单车采用以分布式换电柜为基础的换电运营。换电柜可以实时充电循环利用，因此车和电池配比可以做到 1：1.1，而采用集中仓储进行充电的模式，则至少需要 1：1.6 的配比，也就是每运营一万辆车就可以减少 5000 个电池；随着规模的提升，单车、电池、柜子达到最佳配比后，运营效率可以提升至每辆车每天的换电成本低于 1 元。

（2）问题重述

①分析电单车的“定点还车”规则对用户出行在方面层面的限制，请在杭州城内设置还车的“定点”。

②调查杭州城内的还车的“定点”，并设计共享电动车在不同“定点”之间的调配规则，以防止某个“定点”用户很多单没车的情况。

③在已知每一电车电量的前提下，设计杭州城内运维团队更换电池的策略，寻找在收益与用户良好体验之间的平衡点。

④相比于无“定点还车”的规则，该规则肯定可以减少大量的运维人员，提

升企业收益，但这同时也给用户带来了不便。分析“定点还车”为企业带来的收益，寻找其他还车策略，并在企业收益和用户方便性等方面与“定点还车”规则进行对比。

我们小组选择的是问题①和问题②。

2 问题分析与基本假设

2.1 空间选址与调度的基本假设

2.1.1 还车点选址问题的基本假设

①还车点选址的核心原则：

1. **最优还车点：**车辆与离其最近还车点的距离之和最小；
2. 用车需求通过某一空间范围内的车流密集程度确定；
3. 车流密集程度根据骑电单车的经纬度坐标数据获取。

②还车点选址的问题简化：

1. 不考虑地铁、公交等对骑电单车需求量的影响；
2. 不考虑区域范围内其他共享单车的竞争；
3. 不考虑选址点与建筑物的冲突问题，只考虑车流量的因素。

2.1.2 站点间调配策略的基本假设

①站点间调配的核心原则：

1. **最佳调配策略：**使站点间调配的代价最小、收益最大；
2. 运输代价与运输距离成正相关，与预期收益成负相关；
3. 每次投放必须遍历所有的还车点，且每点仅通过一次。

②站点间调配的问题简化：

1. 不考虑交通状况（堵车、限行等）对运输的影响；
2. 不考虑实际道路的路线，即两点之间的调配沿直线进行；
3. 不考虑单次投放车辆数量的限制，仅研究调配的最优路径，即 Hamiltonian Cycle 问题的求解。

2.2 空间选址与调度的算法选取

2.2.1 还车点选址问题的算法实现

（1）考虑到“最佳”还车点的定义，我们将还车点的选址问题，归结为所有坐标点之间的聚类问题；

（2）选取了基于欧式距离的 K-Means 聚类算法。

2.2.2 站点间调配策略的算法分析

（1）考虑到“最佳”调配策略的定义，我们将站点间调配策略的问题，归结为 Hamiltonian Cycle 的求解问题；

（2）选取了基于“最优个体遗传”策略及“轮盘赌”策略的遗传算法对还车点进行最优路径规划。

3 问题 1：基于 K-Means 空间聚类模型的选址求解

3.1 求解算法：K-Means 聚类算法

我们采用 K-Means 聚类算法对杭州城区所有的骑电单车的位置进行聚类分析，下面先介绍 K-Means 聚类算法的基本原理。

首先，随机选取 k 个点作为质心，将个数据点与最近质心归为同一簇，计算各簇的质心作为新的质心，重复迭代上述操作，直至质心位置不再变化。

数据点间的距离采用欧式距离度量，聚类的目标函数为：

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} dist(c_i, x)^2$$

其中，k 表示 k 个聚类中心，c_i 表示第 $i(1,2,..,k)$ 个中心，$dist$ 为欧式距离。

我们需要选择最小 SSE 所对应的聚类情况。

下面，对第 k 个质心 c_k 求解，寻找最小的 SSE，不妨对 SSE 求偏导：

$$\begin{aligned}\frac{\partial}{\partial c_k} SSE &= \frac{\partial}{\partial c_k} \sum_{i=1}^{K} \sum_{x \in C_i} (c_i - x)^2 \\ &= \sum_{i=1}^{K} \sum_{x \in C_i} \frac{\partial}{\partial c_k} (c_i - x)^2 \\ &= \sum_{x \in c_k} 2(c_k - x_k)\end{aligned}$$

令

$$\frac{\partial}{\partial c_k} SSE = 0 \text{，即} \sum_{x \in C_k} 2(c_k - x_k) = 0 \text{，}$$

可以得到，

$$m_k c_k = \sum_{x \in C_k} x_k \Rightarrow c_k = \frac{1}{m_k} \sum_{x \in C_k} x_k \text{ 。}$$

3.2 模型求解：K-Means 聚类算法在选址问题中的应用

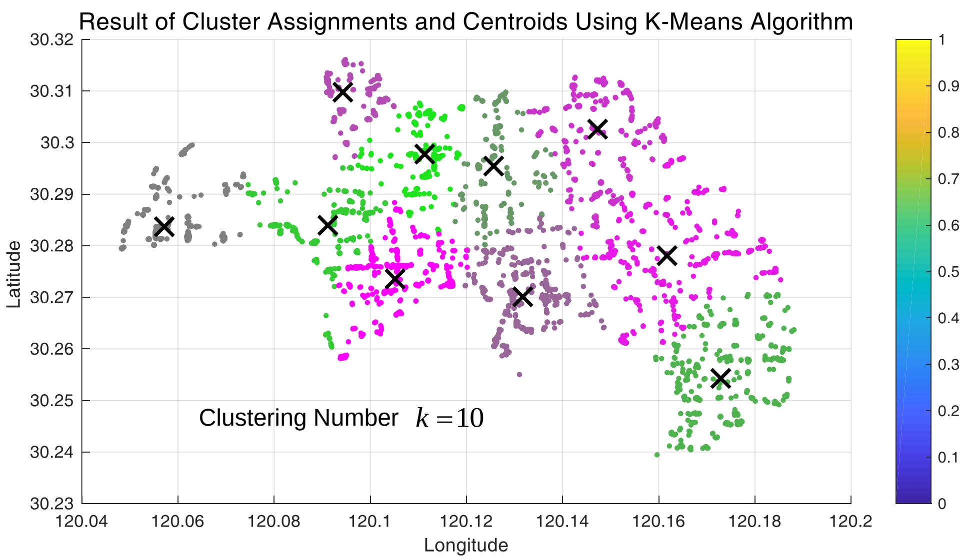
3.2.1 基础模型：指定 k 值情况下的还车点分布

由于上文假设还车点设置仅和某一范围内的车流密度有关，忽略了其他次要影响因素，我们在 MATLAB 中实现了 K-Means 聚类算法的求解。

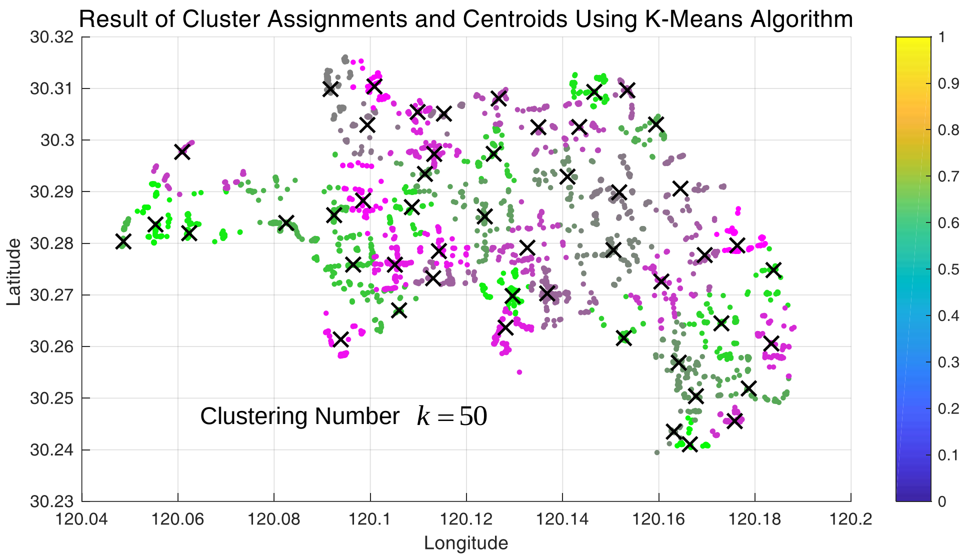
源码： 输入指定 k 值确定还车点空间分布

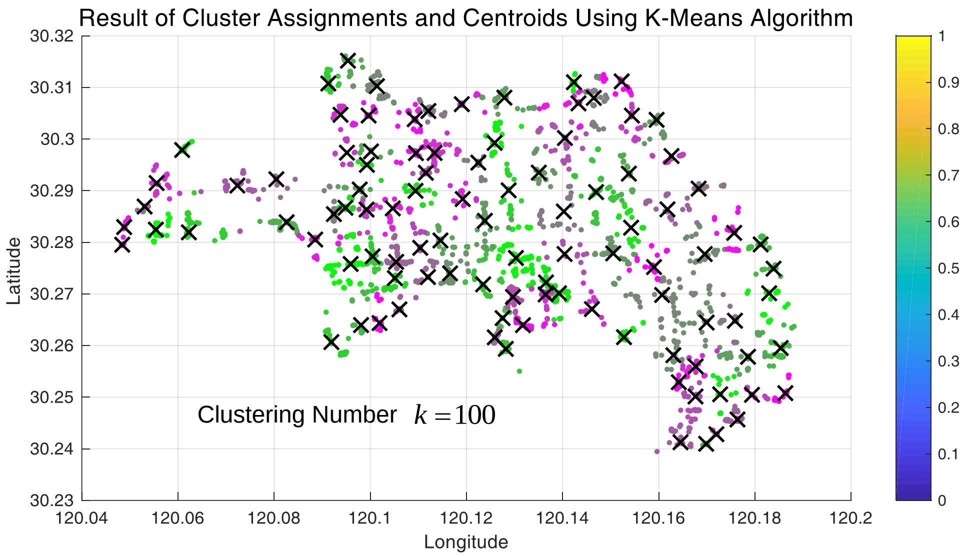
```
clear
clc
Heat = xlsread('qeebike.xlsx');
opts = statset('Display','final');
k = input('聚类数目k =');
[idx,C] = kmeans(heat,k,'Distance','cityblock',...
    'Replicates',5,'Options',opts);
figure;
hold on
for i = 1:k
    plot(heat(idx==i,1),heat(idx==i,2),'.',...
        'Color',[1/k*i,1-1/k*i,1/k*i],'MarkerSize',9)
end
plot(C(:,1),C(:,2),'kx','MarkerSize',15,'LineWidth',2)
title 'Result of Cluster Assignments and Centroids
Using K-Means Algorithm'
```

结果： 聚类数目 $k = 10,50,100$ 情况下的还车点空间分布



上图中聚类数目为 $k=10$，虽然不符合杭州城内骑电单车还车点的分布规律，但是聚类数目对于后续调配模型的求解和分析没有本质的影响。为了简化后面的求解运算，不失一般性，我们在后面都采用基于 $k=10$ 的还车点空间分布。





3.2.2 优化模型：运用代价函数求解最佳的 k 值

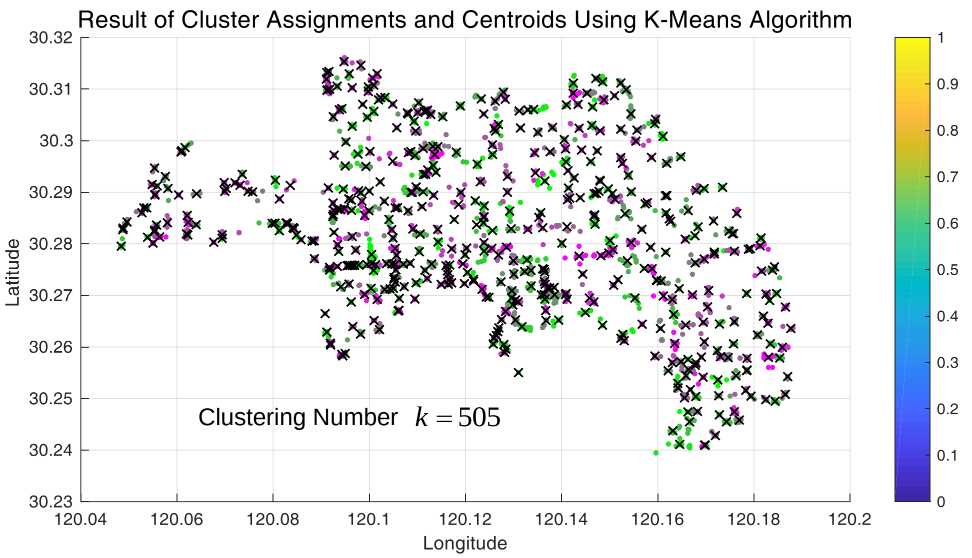
K-Means 算法存在一些缺陷，其中，最大的问题就在于 k 值的选取。在应用该算法的时候，k 值需要预先给定，属于预先知识，很多情况下 k 值的估计是非常困难的。因此，我们试图通过代价函数自适应地确定最佳聚类数目 k 值。

假设： 对特定的 k 值而言，设第 $i(i=1,2,\ldots,k-1,k)$ 类中的车辆数目为 $m_i(i=1,2,\ldots,k-1,k)$。将第 i 类中的第 $j(1,2,\ldots,m_i)$ 辆车离聚类中心点的距离，记为 $d_{ij}(i=1,2,\ldots,k-1,k)$；将所有骑电单车与离其最近聚类点的距离之和，记为 $D_k=\sum_{i=1}^{k}\sum_{j=1}^{m_i}d_{ij}(k=1,2,\ldots,1000)$；那么，对于特定 k 值而言，其代价函数可以表示为 $J_k=k\cdot D_k=k\cdot\sum_{i=1}^{k}\sum_{j=1}^{m_i}d_{ij}(k=1,2,\ldots,1000)$。

求解： 遍历寻找最优 k 值的代码（由于迭代运行时间长，不附上 M 文件）

```
clear
clc
heat = xlsread('qeebike.xls');
opts = statset('Display','final');
for k = 1:1000
    [idx,C,sumd,D] = kmeans(heat,k,'Distance',...
        'cityblock','Replicates',5,'Options',opts);
    Temp = 1000;
    Sum = 0;
    for i = 1:4086
        for j = 1:k
            if(D(i,j)<temp)
                temp=D(i,j)
            end
        end
        sum = sum+temp;
    end
    sum = sum*k;
```

结果： 聚类数目 $k=505$ （最佳）情况下的还车点空间分布



3.2 模型验证：考虑实际地形、道路情况后的调整

为简化模型计算，且不失一般性，我们不妨取聚类数目 $k=10$ ，可以得到每个聚类中心的经纬度数据，即还车点坐标。接下来，结合实际情况以及还车点设置的基本原则，我们将还车点标注在地图中，如下图所示：

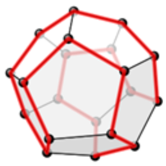
通过上面的地图，可以发现，我们得到的还车点位置非常合理，主要分布在浙大玉泉校区、黄龙体育馆、西溪湿地北门、西溪湿地文二路出口等人流量相对密集的的区域，比较符合实际情况，说明聚类模型得到了较好的验证。

4 问题 2：基于遗传算法的骑电单车调度策略模型

4.1 问题抽象：基于 Hamiltonian Cycle 的最优路径规划

根据上面的假设与分析，骑电单车调配的最优路径，可以看作是对上述得到的 k 个还车点顺序的一种排列（不允许重复通过一个还车点）。

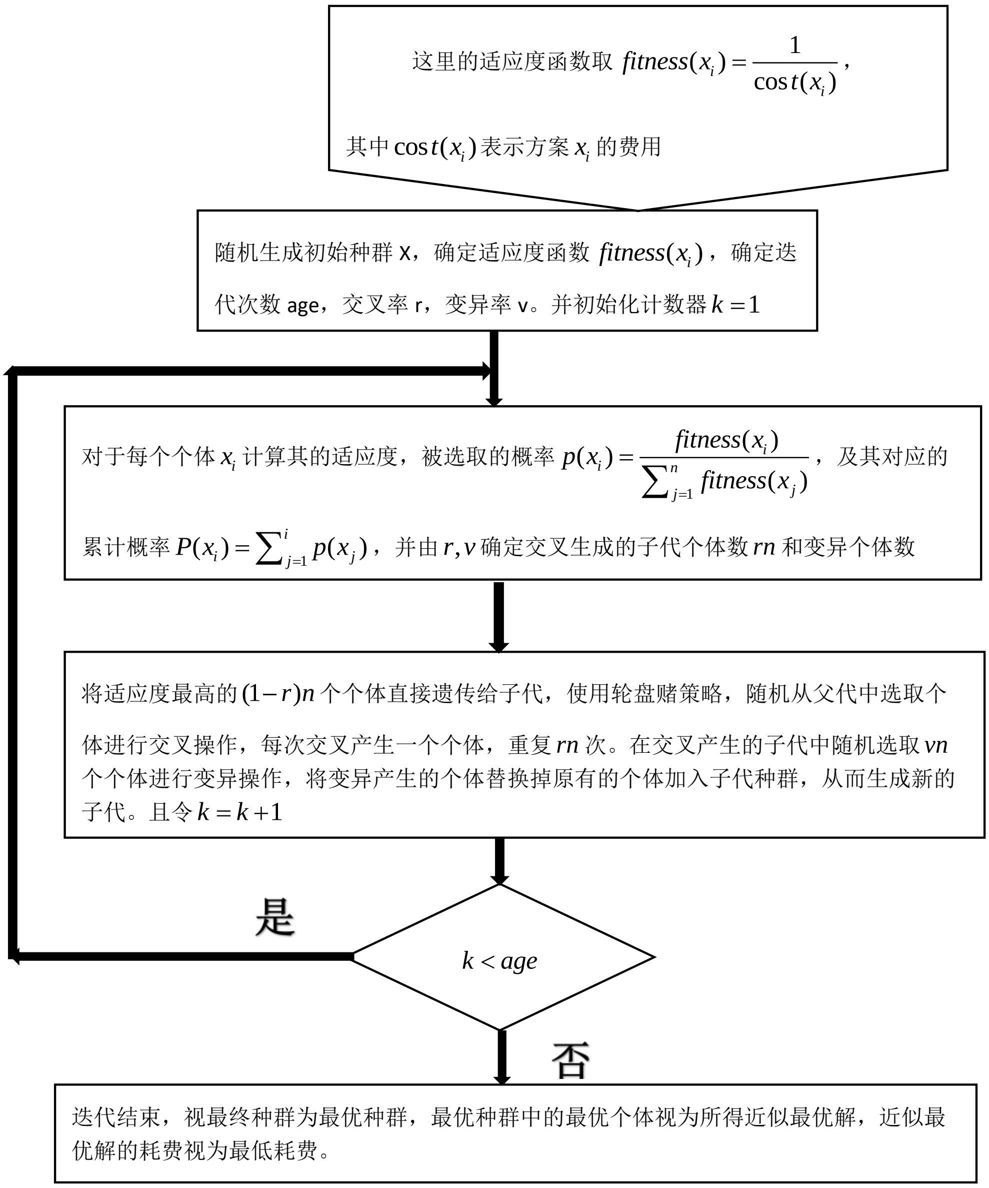
在图论的数学领域中，Hamiltonian Cycle（哈密顿路径，或可追踪路径）是无向或有向图中的一条路径，每个顶点只遍历一次。哈密尔顿循环（或哈密尔顿循环）是一个循环的哈密尔顿路径。


Hamiltonian Cycle

要想对 k 个还车点进行遍历，则需要的遍历数目为 $(k-1)!$ 个，而实现该算法所需要的时间复杂度为 $O(n!)$，显然不是最优的方法。与此同时，遗传算法是一种求解类似问题的高效并行全局搜索方法，能够解决复杂的全局优化问题，解决 Hamiltonian Cycle 问题也是遗传算法的一个重要应用。

遗传算法借助于生物进化的思想和原理与计算机科学相结合，在解决实际问题中得到了很好的广泛应用。遗传算法一般由选择、交叉、变异构成。它通过不断地迭代，逐步改进当前解，直到最后搜索到最优解或满意解。

遗传算法的流程图如下所示：



4.2 模型建立：代价函数的引入与优化

4.2.1 调配过程中代价函数的引入

通过查阅相关资料，我们引入了调配过程中的代价函数，且该代价函数为与调配距离有关的一个 $k\times k$ 的矩阵。

基本模型：调配代价与运输距离成正比，将 10 个还车点记为 $i(i=1,2,\ldots,10)$，由此建立一个 10×10 的距离代价矩阵 $D=D_{i,j}=[d_{ij}]_{10\times10},(i,j)\in\mathbf{R}^{10\times10}$，且 $d_{ii}=0$，$d_{ij}=d_{ji}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2},(i,j)\in\mathbf{R}^{10\times10}$，其中 x_i,y_i 分别为第 i 个还车点的经度和纬度。

4.2.2 调配过程中代价函数的优化

接下来，我们结合骑电单车调度员的薪酬，进一步改进调配过程中的代价函数，使其更具合理性、普遍性和适用性。

优化模型：在上述距离代价矩阵的基础上，我们需要继续考虑骑电公司在调配环节中的潜在收益等因素，来优化调配模型。

假设：每个还车点所带来的收益与其附近用车需求的大小成正比，而用车需求又和周围的车流密度成正比。

（一）还车点周围用车需求的获取

首先，我们假设用以还车点为中心的某一区域内的骑电单车数量表征该点的用车需求量。接下来，需要确定精确的范围，经过相关的数据处理、分析与比较，我们最终计算各还车点附近单位平方千米内的单车数量，如下表所示：

还车点	原始经度	原始纬度	区域内车的数量	相对密度
1	120.1233315	30.2969335	**92**	0.096842
2	120.1531524	30.2726867	**38**	0.04
3	120.0992959	30.3059138	**71**	0.074737
4	120.0563883	30.28366745	**128**	0.134737
5	120.1727397	30.2535499	**52**	0.054737
6	120.1051891	30.2757145	**177**	0.186316
7	120.1315412	30.2701314	**178**	0.187368
8	120.1743103	30.2786834	**61**	0.064211
9	120.1485077	30.3019819	**35**	0.036842
10	120.0886206	30.2839937	**118**	0.124211
总计			**950**	1

由于杭州城内骑电单车总数量约为 4086 辆（在这里我们使用了第 8 组的数据，在此对他们表示衷心的感谢），上述得到的数据相当于总体中的 10 的样本，以此代表各还车点附近用车需求量的相对大小。

（二）优化代价函数的提出

分析：①骑电单车目前收费价格为 ￥$2/h$，考虑实际运营成本，我们认为每辆车每次使用最终可以收益 ￥1.8；②考虑经纬度的影响，地球表面每千米距离所对应的的经纬度变化约为 1/117；③经调查，骑电单车调度员的月薪约为 ￥4000，通过一系列复杂的计算、分析与求解，可以得到，每公里人工费约为 ￥20；等等。

因此，改进的代价函数为：

$$\mathbf{Cost}_{i\times j}=\left[\mathbf{amount}\times 1.8-\mathbf{distance}\times 117\times 20\right]_{i\times j}^{-1},(i,j)\in\mathbf{R}^{10\times 10}$$

注意：上述运算不是单纯的代数操作，而是矩阵的运算。

源码：下面为优化代价矩阵的求解与获取

```
clear
clc
chance =
    [0.096842105,0.04,0.074736842,0.134736842,0.054736
    842,0.186315789,0.187368421,0.064210526,0.03684210
    5,0.124210526];
%根据数据点得到框定范围内用车的可能性
amount = 4086*chance;
%根据每个聚合点的还车可能性，乘以车辆数据总量，得每个还车点周围
可能的车辆数
distance = zeros(10,10);
point = xlsread('qeebike_return.xlsx');
%读取还车点坐标，此坐标为问题一中聚类算法获得的还车点坐标
for i = 1:10
    for j = 1:10
        if(i~=j)
        distance(i,j)=sqrt((point(i,1)-...
            point(j,1))^2+(point(i,2)-point(j,2))^2);
        end
    end
end
%获取不同点间的距离矩阵
Cost = zeros(10,10);
for i = 1:10
    for j = 1:10
        if(i~=j)
            Cost(i,j)=1/(amount(1,j)*1.8...
                -distance(i,j)*117*20);
            Cost(j,i)=Cost(i,j);
        %纬度转化为公里，综合搬运距离与用车可能带来的收益计算代
价函数,Cost(i,j)为由i到j的代价
        end
    end
end
```

4.3 模型求解：遗传算法在调配问题中的应用

遗传算法是采用自定义的适应度函数，通过遗传中的选择、交叉和变异对个体进行筛选，使适应度值好的个体被保留，适应度较差的个体被淘汰。这样反复循环，直至满足条件。

选择操作：从旧群体中以一定概率选择个体到新群体中，个体被选中的概率跟适应度值有关，个体适应度值越好，被选中的概率越大。

交叉操作：从个体中选择两个个体，通过两个染色体的交换组合，来产生新的优秀个体。交叉过程为从群体中任选两个染色体，随机选择一点或多点染色体位置进行交换。

变异操作：从群体中任选一个个体，选择染色体中的一点进行变异以产生更优秀的个体。

遗传算法实现的基本步骤：

（1）**种群初始化：**对于该问题，初始化的参数有种群个数 n、迭代次数 age、变异概率 v、交叉概率 r。

（2）**适应度函数：**在该问题中，已知任意两个还车点之间的总代价 $Cost_{ij}$，因此可将一个随机全排列的总距离的倒数作为适应度函数，即总代价越短，适应度函数越好。

（3）**选择操作：**在遗传算法选择操作中，我们采用轮盘赌法。我们采用基于适应度比例的选择策略，即适应度越好的个体被选择的概率越大，同时在选择中保存适应度最高的个体。

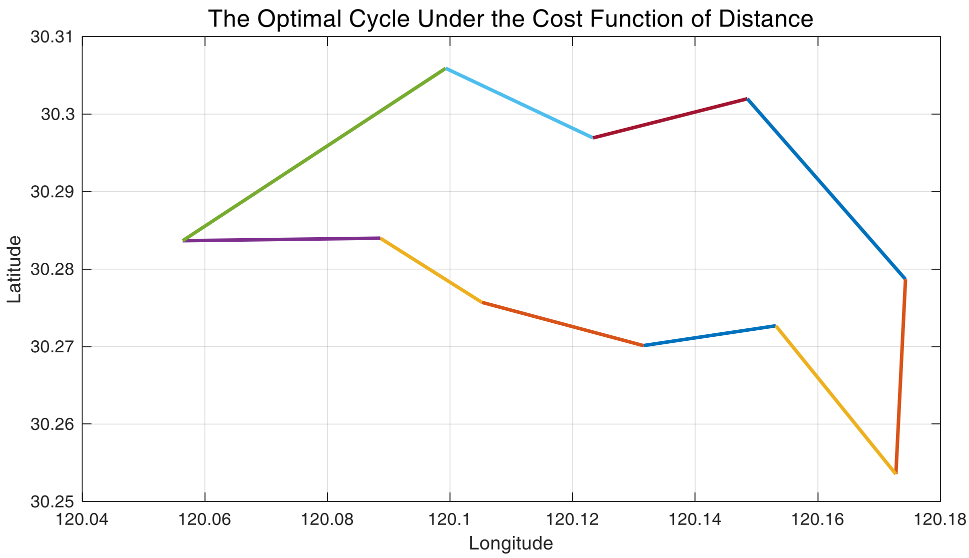
（4）**交叉操作：**我们的程序中对于个体，随机选择两个个体，在对应位置交换若干个基因片段，同时保证每个个体依然是 $1-n$ 的随机排列，防止进入局部收敛。

（5）**变异操作：**我们的程序中对于变异操作，随机选取个体，同时随机选取个体的两个基因进行交换以实现变异操作。

4.3.1 最短路径情况下的 TSP 问题求解

TSP 问题可以归结为路径没有加权情况下的 Hamiltonian Cycle 问题。

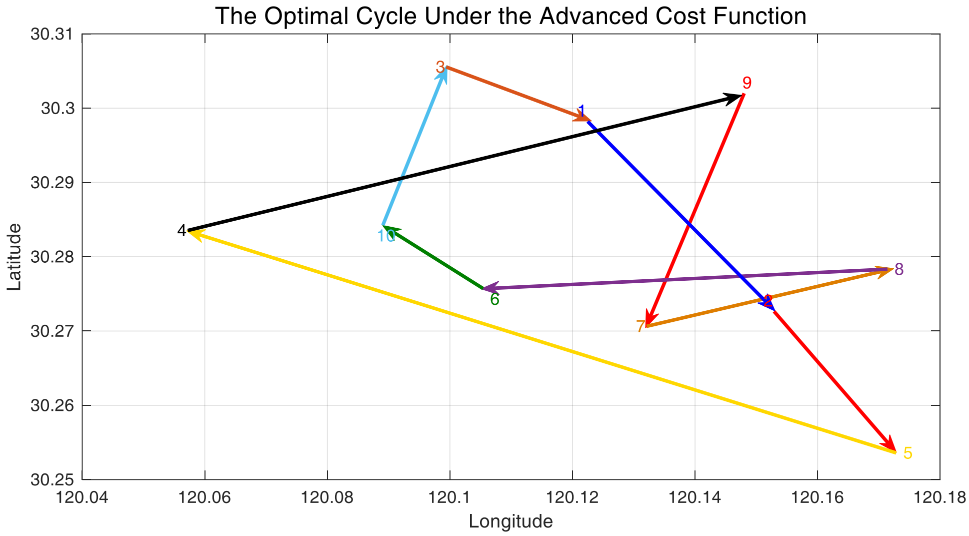
基本结果：下图为基于距离代价矩阵的结果路线



由上图可见，在仅考虑距离为代价函数参数的情况下，TSP 问题的求解结果就是线段无交叉的封闭多边形。显然，这种结果缺乏合理性，需要进行优化。

4.3.2 最小代价情况下的 Hamiltonian Cycle 求解

改进结果：下图为基于改进代价矩阵的结果路线



由于遗传算法本身具有随机性，不能保证每次的结果完全相同，因此，我们经过一系列的调试、运行与筛选后，选择了遗传算法的参数为：种群大小 $n = 200$，迭代次数 $age = 500$，变异率 $r = 0.8$，交叉率 $v = 0.2$。在这种参数条件下，得到的最小代价函数值为 15.3593，可以认为就是最优的轨迹。

根据上图的结果，可以得到，骑电单车调配的最优路线为：9→7→8→6→10→3→1→2→5→4→9。每次运行的路径起点不一定相同，但是最终的轨迹曲线完全一致，说明模型完全符合预期标准，收敛性较好，鲁棒性较高。

5 模型的优缺点

K-Means 聚类算法计算时间短、速度快、聚类效果佳，但是预先确定 k 值，对于异常数据比较敏感。

遗传算法覆盖面大、利于全局优化；可以同时并行处理群体中的多个个体，仅用适应度函数(即代价函数的倒数，在我们的模型中为矩阵的求逆运算)评估；过程简单，继承性好，易于集成；但是，需要对问题进行编码以及得到结果时的解码操作，参数中的交叉率和变异率等参数对解的品质影响较大，而且小依赖于先验知识，没有一个固定的程式。

6 结论

运用 MATLAB，做出了较为精确的还车点分布图，并采用可视化的方式展现骑电单车的选址、调配的情况。为合理解决骑电单车的选址与调度问题，通过聚类分析及遗传算法进行区块数理统计分析，建立合理指标求解了 K-Means 空间聚类模型以及骑电单车调度模型，得出了合理的分配方案，使骑电单车的选址与调度得到了解决与优化。

参考文献

[1] 赵曼. 共享单车网络分析及其优化调度研究[D]. 山东科技大学, 2017.

[2] 潘雪. 基于多目标优化的城市公共自行车租赁点选址研究[D]. 沈阳工业大学, 2016.

[3] 林剑柠, 吴慧中. 基于遗传算法的网格资源调度算法[J]. 计算机研究与发展, 2004, 41(12):2195-2199.

[4] 陶启萌. 基于聚类分析的共享单车时空分布模型[J]. 中国战略新兴产业, 2018(4).

[5] 文蝶斐, 戴亚兰, 郑莹,等. 共享单车的配置与调度优化[J]. 中国科技信息, 2018(6).