

队伍编号	MC2204316
题号	D

## 基站站址规划和区域聚类问题

### 摘要

随着 5G 技术的全面普及，通信所需的带宽越来越大，原有基站能够覆盖的范围越来越小，从而需要建立新基站减少弱覆盖区域。本文围绕基站与弱覆盖点，建立基于改进免疫算法的 0-1 规划模型、基于赋权 K-means 算法的决策模型、基于 DBSCAN 算法的聚类模型，得出新基站建立的最优方案、基站的扇区主方向角度和弱覆盖点的聚类情况。

针对问题一，建立基于 **0-1 规划** 的数学规划模型，采用**改进免疫算法**对规划模型进行求解。根据条件与要求构造数学规划模型的目标函数、决策变量与约束条件，对于候选基站是否选择作为新基站的问题采用 0-1 规划。运用改进的免疫算法对数学规划模型进行求解，其中的初始种群基于**贪心算法**进行选取。最终分别得出宏基站与微基站的站址坐标，总计宏基站 **407** 个，微基站 **3968** 个，覆盖的业务量占比为 **90.0016%**。

针对问题二，基于**赋权的 K-means 算法**对各个基站覆盖范围的扇区主方向进行决策。根据问题一所求得的最优站址，分别对各个基站覆盖范围内的弱覆盖点的业务量进行**赋权**，将带有权重的弱覆盖点应用 K-means 算法进行聚类，文中聚类参数取  $k = 3$ 。在满足主方向间夹角不超过 45 度的情况下，聚类的中心点坐标能同时反应该区域弱覆盖点的坐标分布以及业务量分布，从而可以决策 3 个扇区主方向。最终得出各个基站的 3 个扇区角度。

针对问题三，本文采用复杂度较低的 **DBSCAN 算法**对弱覆盖点进行聚类。由于聚类要求为距离不大于 20 的两点聚为一类，且该聚类具有传递性。相比较于基于中心点位置的 K-means 算法，DBSCAN 算法是基于空间密度进行聚类，有着点间距的限制参数以及极优的传递性。本文 DBSCAN 算法参数的数值取  $Eps = 20, MinPts = 1$ 。最终求得所有弱覆盖点聚类为 **898** 类，该方法的总时间复杂度为  **$O(N\log N)$** 。

最终对本文模型进行分析总结，同时对存在的缺点提出改进的方案，并将模型的应用进行了推广。

**关键词：**改进的免疫算法；0-1 规划；贪心算法；赋权 K-means 算法；DBSCAN 算法

# 目录

<b>一、问题重述</b>	1
1.1 问题背景	1
1.2 问题重述	1
1.2.1 问题一	1
1.2.2 问题二	2
1.2.3 问题三	2
<b>二、问题分析</b>	2
2.1 问题一的分析	2
2.2 问题二的分析	3
2.3 问题三的分析	4
<b>三、模型假设与符号说明</b>	5
3.1 模型假设	5
3.2 符号说明	5
<b>四、问题一的模型建立与求解</b>	6
4.1 数据预处理	6
4.1.1 候选基站的数据预处理	6
4.1.2 弱覆盖点的数据预处理	7
4.2 基于 0-1 规划建立数学规划模型	7
4.2.1 数学规划模型的建立	7
4.2.2 数学规划模型的解释	8
4.3 基于免疫算法求解数学规划模型	9
4.3.1 免疫算法体系构建	10
4.3.2 免疫算法求解过程	10
4.4 新基站规划结果	12
<b>五、问题二的模型建立与求解</b>	13
5.1 基站新覆盖范围分析	13
5.2 基于权重赋值进行数据预处理	14
5.3 基于 K-means 算法决策扇区主方向	14
5.2.1 K-means 算法具体步骤	14
5.2.2 基于含权重的聚类中心点决策扇区主方向	15
5.4 扇区角度结果	16
<b>六、问题三的模型建立与求解</b>	17
6.1 基于 DBSCAN 算法进行区域聚类	17
6.1.1 DBSCAN 算法所需参数	17
6.1.2 DBSCAN 算法具体步骤	18
6.2 弱覆盖点聚类的求解	18
<b>七、模型的评价、改进与推广</b>	19
7.1 模型的优点	19
7.2 模型的不足	20
7.3 模型的改进与推广	20
7.3.1 模型的改进	20
7.3.2 模型的推广	20

参考文献 .....21

附录 .....22

    附录 1.....22

    附录 2.....23

    附录 3.....25

# 一、问题重述

## 1.1 问题背景

为深化落实习近平总书记有关加速 5G 发展的重要指示并加速推进 5G 赋能各行各业的宏伟计划，全民 5G 的计划被提上日程。然而，随着移动通信技术的高速发展，例如 5G 技术的全面普及，通信所需的带宽越来越大，原先的基站能够覆盖的范围越来越小，使得要达到与几年前相同水平的覆盖区域所需要基站数量大幅上升。如何在有限成本条件下实现基站覆盖范围以及信号强度最大成为了炙手可热的议题。

复杂问题的规则可被简化为如下所述：

- (1) 现实中的地区可被映射为众多相邻的方格，取方格中心点作为方格的代表，即给定任意区域，都可以将其划分为有限个点。
- (2) 每一个映射的坐标点都具有一定的属性值，包括但不限于：坐标，业务量以及是否为弱覆盖点等。
- (3) 为了尽可能降低成本以及设备之间相互的干扰问题，任意两个基站之间的距离应当满足特定门限要求。
- (4) 假设基站覆盖范围为  $d$ ，映射点基站的坐标为  $P_0(x_0, y_0)$ ，则任意满足条件  $\|P - P_0\|_2 \leq d$  的点  $P(x, y)$  都处于该基站的覆盖范围之内，反之则认为其不在基站覆盖范围内。

## 1.2 问题重述

### 1.2.1 问题一

问题一给定一个大小为 2500\*2500 的方块区域（共 2500\*2500 个坐标点，其中横坐标为 0 到 2499，纵坐标同上），同时建模者具有两种基站选择，分别为：

- (1) 宏基站（成本 10/站，覆盖范围 30/站）
- (2) 微基站（成本 1/站，覆盖范围 10/站）

同时本题限定基站门限为 10，即任意两个基站之间的距离  $d$  应当满足  $d > 10$  的条件。

本题要求基站的建立应当首先满足弱覆盖点总业务量的 90% 被规划基站覆盖，同时应当尽量降低成本。

### 1.2.2 问题二

本题基于问题一的基础，进一步考虑现实问题，即在实际中每一个基站的覆盖范围并不是圆形。出于技术难题，每个基站上有 3 个扇区，每个扇区都指向各自的一个方向，在主方向左右 60 度范围内可覆盖，然而覆盖范围按线性逐渐缩小，在 60 度时覆盖区域为主方向覆盖区域的一半。若超过 60 度，则无法被扇区覆盖。

基于上面的叙述，再考虑每个站的任意 2 扇区的主方向之间夹角不得小于 45 度，在这种情况下探讨是否能够新建基站使得其覆盖弱覆盖点的业务总量的 90% 以上，并给出最优基站位置和扇区角度。若不能则需要给出最多可覆盖的弱覆盖点总业务量占比。

### 1.2.3 问题三

继续深化考虑现实问题，为了更好的解决弱覆盖带来的问题，需要对被标记为弱覆盖的点进行区域聚类，将距离相对较近的弱覆盖点聚集成一类，从而得到弱覆盖区域，这使得管理者能够将弱覆盖区域分开管理并更加好的解决弱覆盖问题。

聚类要求为：两个弱覆盖点的距离小于等于 20，则这两个弱覆盖点应当被聚集为一类，并要求考虑到聚类的传递性质（即点 A 与点 B 为一类，点 B 与点 C 为一类，则 A、B、C 点均为同一类别）。在满足上述要求的情况下，聚类算法的时间复杂度应当尽量低。

## 二、问题分析

本文需要解决的问题是确定新基站建立的坐标与种类，根据基站覆盖范围的变化进行决策调优，对弱覆盖的点进行区域聚类。依据附件 1 和附件 2 所给现有基站与弱覆盖点的信息，进行新基站建立的规划决策。

### 2.1 问题一的分析

针对问题一，在给定区域内规划选取坐标进行新基站的建立，能够使得弱覆盖点的总业务量 90% 被新基站覆盖，同时需要满足各个基站之间的距离需要大于 10，并且成本最小化。新建基站有宏基站与微基站两种类型可供选择。

首先对数据进行预处理。根据附件 1，对业务量极低的弱覆盖点进行预处理，删除后对结果的影响差异很小；根据附件 2，筛除现有基站和以现有基站为圆心 10 为半径的圆心范围的点，剩余的所有点作为候选基站点。

随后运用基于 0-1 规划的免疫算法求解基站建立的最优方案。基于 0-1 规划模型，对新

基站问题建立数学规划模型。以总建站成本为目标函数，问题一的所有要求作为约束条件。基于免疫算法求解数学规划模型，该算法中的初始种群采取贪心算法进行获取。最终得出新基站的选址、基站类型以及覆盖业务量的百分比。

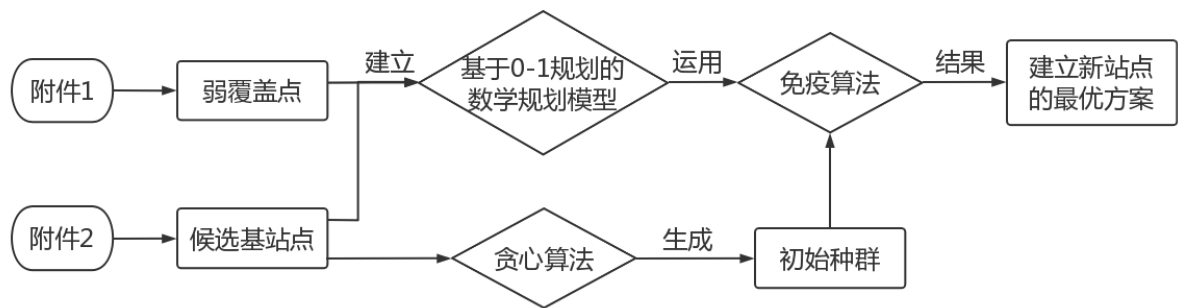


图 1 问题一的思维导图

## 2.2 问题二的分析

基于问题一的求解出的新基站最优规划决策，分析覆盖范围的变化时，弱覆盖点的被覆盖情况以及覆盖业务量的占比。因覆盖范围从主方向向两侧按线性减小，所以扇区的主方向应指向弱覆盖点密集且业务量较大的方向。因此采用聚类的方式，对扇区的主方向进行决策。

首先对每一新基站点覆盖范围内的弱覆盖点进行赋权，运用 K-means 算法对其进行聚类。本文将覆盖范围内的弱覆盖点聚为 3 类，同时分别求出该 3 类弱覆盖点的 3 个中心点，基站点与该 3 个中心点的连线作为初始的 3 个扇区主方向。其次根据问题二的要求，根据业务量以及主方向间的角度不小于 45 度要求，对各个主方向进行统筹调整。最终得出所有基站的各个扇区主方向的角度情况，以及最多可覆盖的弱覆盖点总业务量占比。

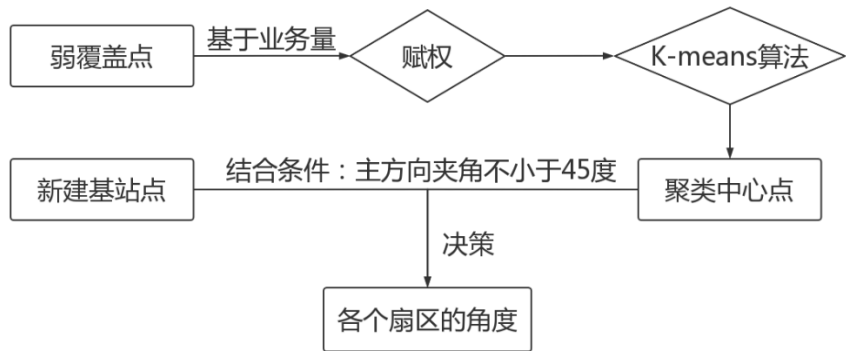


图 2 问题二的思维导图

## 2.3 问题三的分析

针对问题三，需要最终得出所有弱覆盖点的聚类情况，同时需要保证较为低的总时间复杂度。弱覆盖点之间的聚类约束条件为，若  $A$ 、 $B$  两点之间的距离  $\|A - B\|_2 \leq 20$  时，那么这两点聚为一类，同时该聚类具有传递性，并保证较低的总时间复杂度。因此本文采用 DBSCAN 算法对弱覆盖点进行聚类分析。

选取 DBSCAN 算法的三个原因：

- (1) 能够通过 DBSCAN 算法中的参数  $Eps$ ，即以某点为圆心所划分圆形区域的半径大小，从而可限定两点之间距离，满足问题三的聚类约束条件。
- (2) DBSCAN 算法的原理，亦是各个核心点与之间的传递对目标进行聚类，是一种基于空间密度有传递性质的聚类模型。
- (3) 问题三是定点的二维数据聚类分析，因此该算法在此处的总时间复杂度较低。

对比 K-means 算法，虽然可以将目标对象进行客观聚类，但是以下三个不选取的原因：首先，K-means 算法需要事先人为确定聚类种数，起始点的选择会极大影响聚类结果，因此具有很大程度的主观性。其次，该算法难以发现任意形状的簇，而地图上的弱覆盖点显然具有形状性。而且，K-means 算法的原理是基于各个点每次迭代的中心点进行聚类，因此该算法缺乏考虑各个弱覆盖点聚类的传递性。

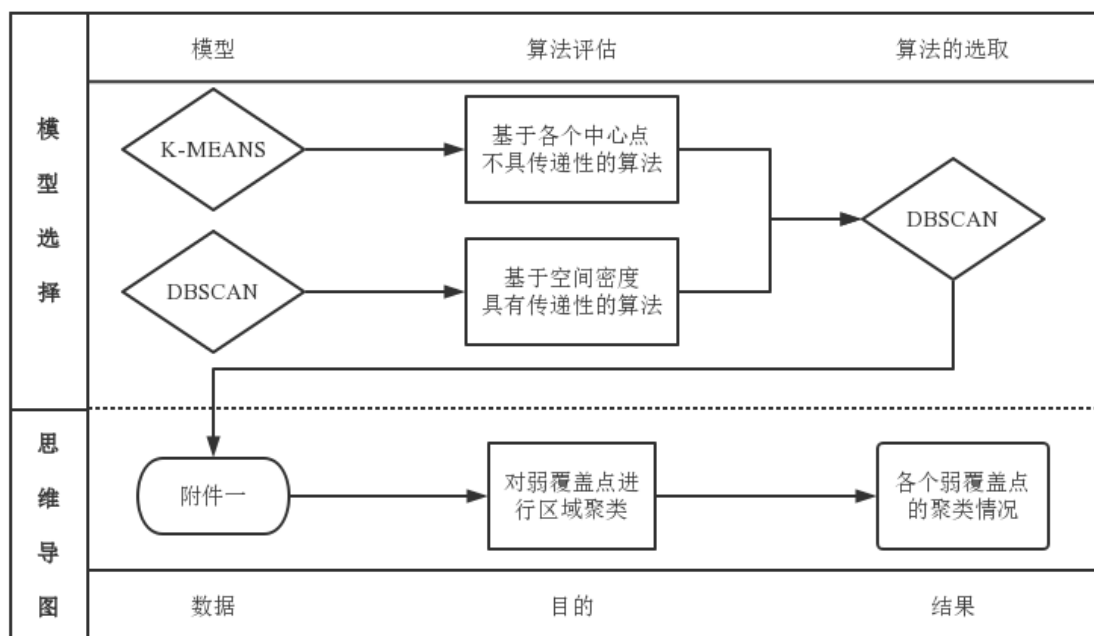


图 3 问题三的思维导图

### 三、模型假设与符号说明

#### 3.1 模型假设

(1) 本题 2500\*2500 的区域中，每一个点都可以建立通讯基站，即不考虑自然地理以及人为因素的影响。

(2) 本题中建设的所有基站在其自身信号覆盖范围内在任意时间均可做到均匀投射，即不考虑外界环境对于基站的影响，基站覆盖范围内所有位置都能够公平的接受到基站的信号。

(3) 特定区域的通信基站用户使用量稳定，即用户使用量不会随着时间变化发生较大调整。

(4) 所有通讯基站的性能，不随着时间变化而发生变动，即本题中出现的所有通讯基站性能稳定且不会发生故障。

#### 3.2 符号说明

符号	含义	备注
$n$	候选基站的个数	/
$m$	弱覆盖点的个数	/
$T = \{1, 2, \dots, n\}$	候选基站点集	/
$F = \{1, 2, \dots, m\}$	弱覆盖点集	/
$w_i$	弱覆盖点 $i$ 的业务量	$\forall i \in F$
$W = \{w_1, w_2, \dots, w_m\}$	弱覆盖点的业务量	/
$(x_i, y_i)$	候选基站 $i$ 的坐标	$\forall i \in T$
$u_{is}$	候选基站 $i$ 覆盖弱覆盖点 $s$ 的情况	$u_{is} \in \{0, 1\}, \forall i \in T, \forall s \in F$
$v_i$	候选基站 $i$ 作为新基站的情况	$v_i \in \{0, 1\}, \forall i \in T$
$r_{is}$	候选基站 $i$ 覆盖弱覆盖点 $s$ 之间的覆盖关系的存在性	$r_{is} \in \{0, 1\}, \forall i \in T, \forall s \in F$
$H = \{S \mid r_{is} = 1\}$	新建基站所能覆盖到的弱覆盖点	$\forall i \in T, \forall s \in F$
$W' = \{w_h \mid h \in H\}$	新建基站覆盖弱覆盖点的业务量	/
$a_i$	新建选基站 $i$ 为宏基站的情况	$a_i \in \{0, 1\}$
$b_i$	新建选基站 $i$ 为微基站的情况	$b_i \in \{0, 1\}$
$L$	基站参数矩阵	/
$D$	抗体编码矩阵	/
$\varphi_i$	弱覆盖点 $i$ 的权重	$\forall i \in F$



## 四、问题一的模型建立与求解

### 4.1 数据预处理

首先对现有基站点与弱覆盖点进行可视化，直观认识整体的数据信息，如下图所示。

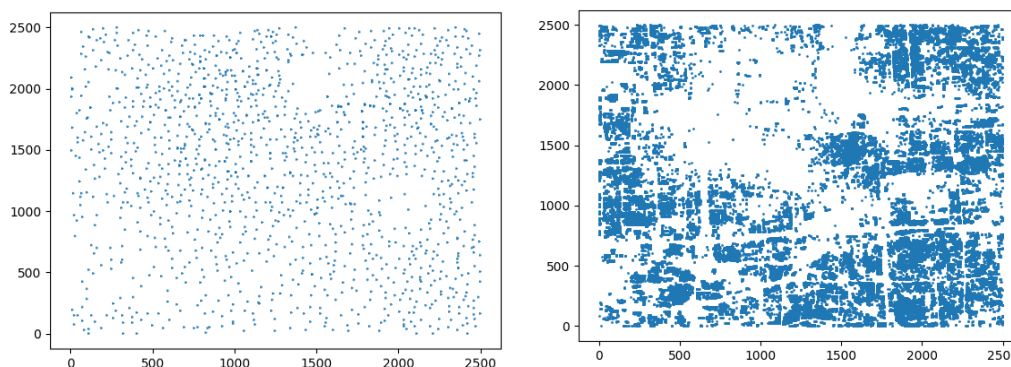


图 4 现有基站散点图（左）和弱覆盖点散点图（右）

#### 4.1.1 候选基站的数据预处理

根据问题一，现有基站与新建基站之间的距离门限是 10。以该约束条件进行部分数据的清洗，减少后期的计算量与数学模型的复杂程度。

在给定区域  $2500 \times 2500$  的点集中，运用附件 2，将现有基站为圆心 10 为半径的圆内所有点（包括圆心）进行剔除，公式如下

$$\sqrt{(x - x_i)^2 + (y - y_i)^2} \leq 10, \quad (1)$$

其中  $(x_i, y_i)$  为现有基站  $i$  的坐标， $i = 1, 2, \dots$

经过数据预处理后，清洗掉了 380,799 个不符合约束条件的坐标，得到可作为新基站的候选坐标 5,869,201 个，同时对每一候选坐标点进行顺序标号，如下表所示。

表 1 预处理后的候选基站数据

X坐标	Y坐标	X坐标	Y坐标	X坐标	Y坐标
0	0	5	991	1699	590
0	1	5	992	1699	591
0	2	5	993	1699	592
.....	.....	.....	.....	.....	.....
4	1392	420	2032	2499	2490
4	1393	420	2033	2499	2491
4	1394	6420	2034	2499	2492
.....	.....	.....	.....	.....	.....

#### 4.1.2 弱覆盖点的数据预处理

问题一要求，最终被覆盖的弱覆盖点总业务量占总量的 90%。弱覆盖点的数据总量较大，本文筛除部分业务量较低的弱覆盖点数据，即对总业务量的影响较小的弱覆盖点，这些弱覆盖点对最终结果有着很小程度上的影响。

根据附件 1，弱覆盖点的总数为 182,807 个，对业务量为 0.9 以下的弱覆盖点进行筛除，最终剩余 128,112 个弱覆盖点数据。数据删除前与删除后，对最终业务量分析计算的误差为 0.001，置信度为 99.9%。在保证最终结果准确性与误差小的同时，减少影响因素较小的样本，从而能够使得模型求解更加迅速。最终弱覆盖点数据预处理结果如下表所示。

表 2 预处理后的弱覆盖点数据

X坐标	Y坐标	业务量	X坐标	Y坐标	业务量
844	1962	47795	864	2013	21939
1356	2271	43295	1335	2206	21331
869	2292	32200	1357	1086	20138
881	1256	24688	675	1957	19781
1096	1658	23715	932	2210	18680
683	2198	22274	1019	1593	18394
865	2012	21939	935	1665	18241

#### 4.2 基于 0-1 规划建立数学规划模型

由于新基点的建立还需要满足：新建基站之间的距离门限为 10；规划后的新基站，能够覆盖弱覆盖点总业务量的 90%；建立基站的总成本尽量低。建立数学符号，如 3.2 符号说明所示。

##### 4.2.1 数学规划模型的建立

基于 0-1 规划，建立带约束的基站规划模型如下：

目标函数：

$$\min f = \sum_{i=1}^n (10v_i a_i + v_i b_i) \quad (2)$$

约束条件：

$$a_i + b_i = 1 \quad (3)$$

$$\sqrt{(x - x_i)^2 + (y - y_i)^2} > 10 \quad (4)$$

$$\sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 30 \quad (5)$$

$$\sqrt{(x - x_h)^2 + (y - y_h)^2} \leq 10 \quad (6)$$

$$\sum w_h \geq 6350607 \quad (7)$$

其中  $i \in T$ ,  $s \in F$ ,  $h \in H$ ,  $H = \{s \mid r_{is} = 1\}$ ,  $r_{is} = u_{is} \cdot v_i$ , 式(5)满足  $r_{is} \cdot a_i = 1$ , 式(6)满足  $r_{is} \cdot b_i = 1$ 。

#### 4.2.2 数学规划模型的解释

决策变量：新建基站的坐标  $(x, y)$ 。

目标函数：新建基站的总成本最低。

约束条件：式(3)为新建基站的类型选取，即选择宏基站或者微基站；

式(4)为新建基站之间的距离门限为 10；

式(5)为被宏基站覆盖的弱覆盖点与宏基站之间的距离不大于 30；

式(6)为被微基站覆盖的弱覆盖点与微基站之间的距离不大于 10；

式(7)为新建基站所能覆盖到弱覆盖点的总业务量大于所有业务量的 90%。

**目标函数**式(2)为建立成本最低化，其中  $v_i, a_i, b_i = 0$  或  $1$ ，宏基站成本为 10，微基站成本为 1。 $v_i = 0$  表示在候选基站  $i$  不建立新基站，反之  $v_i = 1$  表示建立新基站； $a_i = 0$  表示在候选基站  $i$  不建立宏基站，反之  $a_i = 1$  表示建立宏基站； $b_i = 0$  表示在候选基站  $i$  不建立微基站，反之  $b_i = 1$  表示建立微基站。 $v_i a_i$  表示候选基站  $i$  作为新基站并选取宏基站的情况，当且仅当  $v_i a_i = 1$  时候选基站  $i$  建立宏基站；同理  $v_i b_i = 1$  时候选基站  $i$  建立微基站。 $v_i, a_i, b_i$  三者以及相互联系的含义如下所示：

$$v_i = \begin{cases} 0 & (\text{候选基站 } i \text{ 被选中}) \\ 1 & (\text{候选基站 } i \text{ 没被选中}) \end{cases} \quad (8)$$

$$a_i = \begin{cases} 0 & (\text{候选基站 } i \text{ 建立宏基站}) \\ 1 & (\text{候选基站 } i \text{ 不建立宏基站}) \end{cases} \quad (9)$$

$$b_i = \begin{cases} 0 & (\text{候选基站 } i \text{ 建立微基站}) \\ 1 & (\text{候选基站 } i \text{ 不建立微基站}) \end{cases} \quad (10)$$

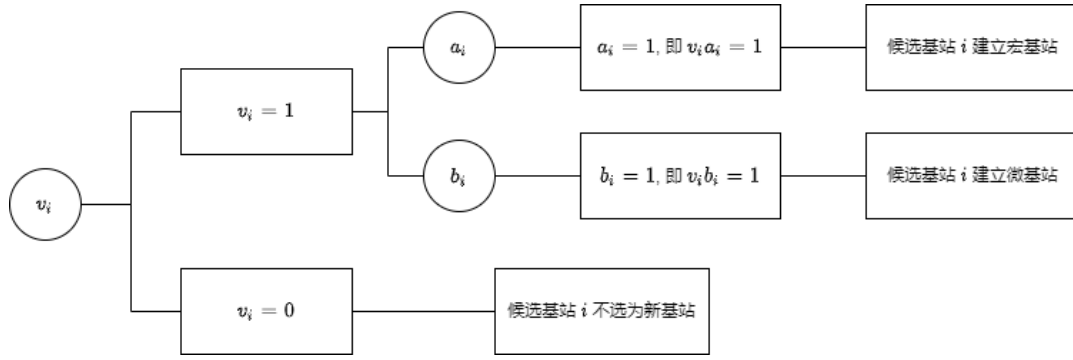


图5 候选基站  $i$  建立新基站的情况

**约束条件中**，式(3)为  $a_i + b_i$  为 1，表示一个选为新基站的候选基站  $i$ ，在宏基站与微基站两种基站类型中二选一。式(4)为两个不同新建基站之间的距离应大于 10。式(5)需满足  $r_{is} \cdot a_i = 1$ ，意为所有能被宏基站覆盖的弱覆盖点；其中  $u_{is} \cdot v_i = 1, r_{is} = u_{is} \cdot v_i$ ，当  $r_{is} = 1$  意为候选基站  $i$  覆盖弱覆盖点  $s$  之间的存在覆盖关系；当  $u_{is} = 1$  表示候选基站  $i$  覆盖到弱覆盖点  $s$ ，反之  $u_{is} = 0$  表示不覆盖；该覆盖关系的存在性取决于  $u_{is}$  和  $v_i$  的取值，同时为 1 时表示存在覆盖与被覆盖的关系。式(6)需满足  $r_{is} \cdot b_i = 1$ ，意为所有能被微基站覆盖的弱覆盖点。式(7)为能够被新建基站覆盖的弱覆盖点总业务量，要大于等于总业务量 7,056,230 的 90%，即 6,350,607；其中  $H$  为新建基站所能覆盖到的弱覆盖点的集合， $w_h$  为第  $h$  个弱覆盖点业务量( $h \in H$ )。

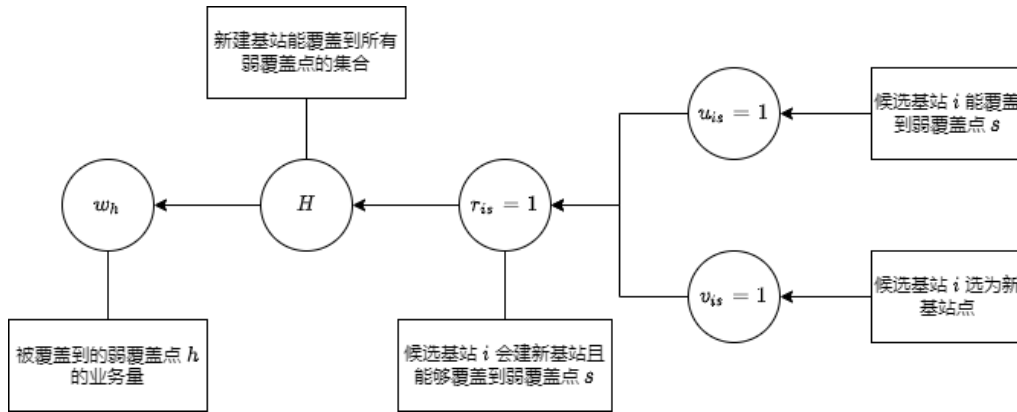


图6  $w_h$ （能够被新建基站覆盖的弱覆盖点总业务量）的思维导图

#### 4.3 基于免疫算法求解数学规划模型

免疫算法是基于遗传算法、以人工的免疫系统为理论基础构建的一类算法，其保留了遗传算法的优良特性，可以产生适应度更高的子代，并且免疫算法能够保留适应度较高的父代。

免疫算法既考虑各个免疫元素之间的竞争，又考虑外部环境的作用于影响。问题一新建基站的全局最优规划，也需从内外两方面，考虑个体基站之间的相互影响以及弱覆盖点的分

布情况。

#### 4.3.1 免疫算法体系构建

免疫算法是实现了相似于生物的免疫系统运作过程, 首先将免疫算法与免疫系统进行类比, 如下表所示。

表 3 类比免疫算法与免疫系统

免疫系统	免疫算法
抗原	新建基站的问题
抗体	建基站的目标函数和约束
抗原识别	问题的识别
从记忆细胞产生抗体	联系过去的最佳方案
记忆细胞分化	维持最优方案
T 细胞抑制	消除多余的候选基站
抗体增加 (细胞克隆)	运用遗传算子产生新的最优方案

#### 4.3.2 免疫算法求解过程

##### (1) 抗体编码

在免疫算法之中, 抗原作为要解决的问题, 抗体作为问题的解。因此每一抗体对应了一种新基站规划的优化方案, 而每一候选基站所对应的基因座的值为该基站的选中情况。由于在基站的选择与优化的过程中, 每一候选基站只存在选中与没被选中这 2 中情况, 因此根据 4.2 的 0-1 规划, 设候选基站集合为  $T$ , 每一候选基站  $i$  的选中情况如下所示, 其中  $i \in T$ :

$$v_i = \begin{cases} 0 & (\text{候选基站 } i \text{ 被选中}); \\ 1 & (\text{候选基站 } i \text{ 没被选中}). \end{cases}$$

从而形成候选基站状态矩阵

$$V = (v_1, v_2, \dots, v_n) \quad (11)$$

基站的参数含有候选基站的坐标与覆盖范围, 本文采用实数二维矩阵的编码形式。其中每一行作为一个基站的参数情况, 矩阵行数作为基站的个数, 形成基站的参数矩阵, 如下所示:

$$L = (l_1, l_2, \dots, l_n)^T \quad (12)$$

其中,  $l_i$  为候选基站  $i$  的基站数据行向量, 包括有  $x$  坐标、 $y$  坐标与该基站的覆盖范围。

例如, 某参数矩阵为

$$L = \begin{pmatrix} 355 & 123 & 30 \\ 355 & 124 & 10 \\ \dots & \dots & \dots \end{pmatrix}$$

那么  $l_1 = (355, 123, 30)$  表示基站坐标为  $(355, 123)$ ，基站覆盖范围为 30。

将候选基站状态式(11)与基站的参数矩阵式(12)相乘，得到**抗体编码矩阵**  $D$  为

$$D = (d_1, d_2, \dots, d_n), \quad (13)$$

其中  $d_i = v_i \cdot l_i, i = 1, 2, \dots, n$ 。

## (2) 种群初始化

免疫算法起始需要一个初始种群，该算法能够以迭代的方式，改进现有的规范方案，那么可以根据问题的本身的先验知识，以及弱覆盖点的大致分布情况以及密集程度，得到部分潜在的较优解加入初始种群。

本文在首次运行算法时，在候选基站集中采用**贪心算法**产生初始种群，使得初始新基站尽可能多的覆盖业务量较大的弱覆盖点。随后再次运行时，保留历史的求解结果中有的优势解集，对下一轮的初始种群进行填充，剩余部分仍然以随机的形式生成。不仅仅提高收敛速率，同时保证了优势解的不流失，而且持续了初始化种族的多样性。

## (3) 抗体的亲和度函数

为了降低新基站的成本，建立基站建立的成本目标函数，即抗体亲和度的评价函数：

$$f(D) = \sum_{i=1}^n (10v_i a_i + v_i b_i) \quad (14)$$

其中  $v_i$  为候选基站  $i$  作为新基站的情况， $a_i$  为新建选基站  $i$  为宏基站的情况， $b_i$  为新建选基站  $i$  为微基站的情况，宏基站建立的成本为 10，微基站建立的成本为 1。

## (4) 抗体的浓度

抗体的浓度是指在抗体的种群中看题邻近数量与抗体总规模的比值。

两个抗体的欧氏距离如式(15)所示

$$Distance(d_p, d_q) = \sqrt{\sum_{i=1}^n \sum_{j=1}^m (d_p[i][j] - d_q[i][j])^2}, \quad (15)$$

其中  $n$  为基站的数目（即抗体矩阵行数）， $m$  为基站的参数（即抗体矩阵的列数）。

如果两个抗体的欧氏距离不大于阈值  $\theta$  时，该两个抗体为邻近抗体，如式(16)所示

$$NB(d_p, d_q) = \begin{cases} 1, & Distance(d_p, d_q) < \theta \\ 0, & \text{其他} \end{cases} \quad (16)$$

因此抗体的浓度 *density* 为

$$density(d_p) = \frac{1}{size} \sum_{q=1}^{size} NB(d_p, d_q) \quad (17)$$

其中  $size$  表示抗体的种群规模。

#### (5) 免疫算法步骤

首先对抗体编码矩阵  $D$  进行识别，运用贪心算法生成初始的基站种群群体，根据抗体亲和度的评价函数计算出抗体亲和度。随之通过式(17)计算抗体浓度数值，对抗体进行免疫系统操作：克隆、变异、克隆抑制的操作，最终得出新的种群群体。重复迭代操作，最终输出满足条件的基站建立最优策略。免疫算法步骤如下图所示。

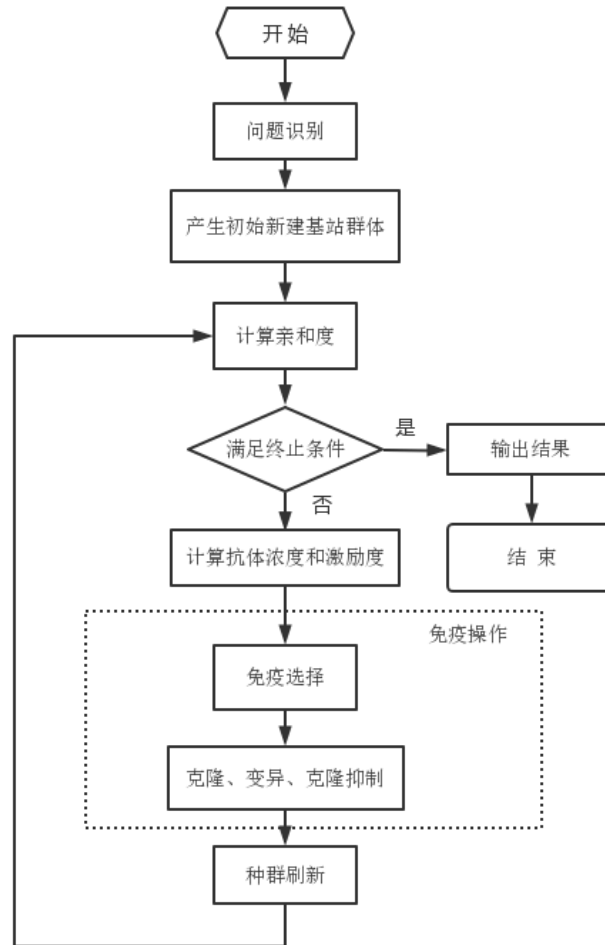


图 7 免疫计算流程图

#### 4.4 新基站规划结果

根据上述所有算法和模型求解出的最终结果为：建立 407 个宏基站，3968 个微基站，覆盖的业务量占比为 90.0016%。每个新基站的站址坐标参见**支撑材料 1**“问题一：新基站建立情况”，部分基站坐标如下表所示。

表 4 部分站址坐标

微基站选址				宏基站选址			
X坐标	Y坐标	X坐标	Y坐标	X坐标	Y坐标	X坐标	Y坐标
844	1962	864	2013	1335	2206	515	1987
1356	2271	1357	1086	675	1957	1379	2312
869	2292	932	2210	430	1643	1041	1586
881	1256	1019	1593	863	2186	1102	1126
1096	1658	935	1665	1367	1501	1323	1389
683	2198	1489	1228	2238	1788	1257	2372
.....	.....	.....	.....	.....	.....	.....	.....

以部分区域为例，将该区域的弱覆盖点、新基站点及其覆盖范围进行可视化，如下图所示。

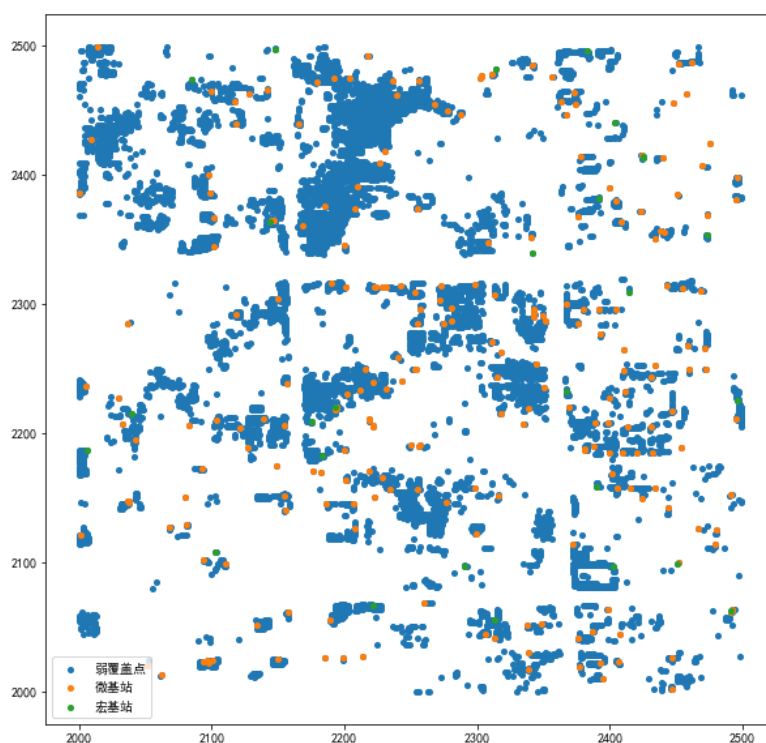


图 8 新建基站与弱覆盖点分布图

## 五、问题二的模型建立与求解

### 5.1 基站新覆盖范围分析

由于覆盖范围在主方向上最大，即宏基站在主方向为 30、微基站在主方向为 10，随着角度偏离主方向，覆盖范围以线性的变化速度减小，直至 60 度时覆盖范围为原覆盖范围的



$\frac{1}{2}$ 。以宏基站为例，其覆盖范围大小的变化情况如下图所示。

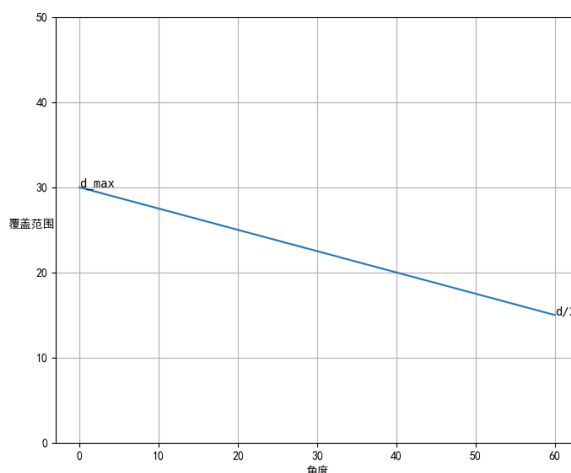


图 9 宏基站的覆盖范围变化情况

## 5.2 基于权重赋值进行数据预处理

由于 K-means 算法仅将弱覆盖点的各个坐标进行聚类分析，忽略了弱覆盖点业务量这一影响因素。因此首先对弱覆盖点进行赋权，依据业务量对每一弱覆盖点进行等数量的克隆，从而弱覆盖点的各个坐标得到了与业务量成正比的权重值。弱覆盖点的权重与业务量的关系如下式所示：

$$\varphi_i = \frac{w_i}{10}, \quad (18)$$

其中  $w_i$  为弱覆盖点  $i$  的业务量， $\forall i \in F$ ， $F$  为弱覆盖点集。

## 5.3 基于 K-means 算法决策扇区主方向

K-means 算法是一种基于距离来对样本进行分类的方法，考虑到实际上基站的覆盖范围不规则，而是由三个覆盖范围从主方向往两侧递减的扇形决定，再通过对第一问选出的基站覆盖范围内的弱覆盖点分析，我们可以采用 K-means 算法将这些点进行分类处理。

### 5.2.1 K-means 算法具体步骤

首先计算各样本点离各个中心点的距离，并将该样本点分配给最近中心点代表的类别，当一次迭代完成后，根据聚类结果更新每个类别的中心点，然后重复之前操作再次迭代，直到前后两次分类结果没有差别。

步骤一： $k = 3$ ，随机选取 3 个中心点，计算各样本点到中心点距离。

步骤二：将每个样本点分配给距离最近的中心点。

步骤三：根据聚类结果更新中心点。

步骤四：重复步骤二，将每个样本点重新分配给距离最近的中心点。

步骤五：重复步骤三，更新中心点。

步骤六：重复步骤四，判断聚类结果是否与上次结果一致。若一致则算法终止，否则返回步骤五，继续进行迭代。

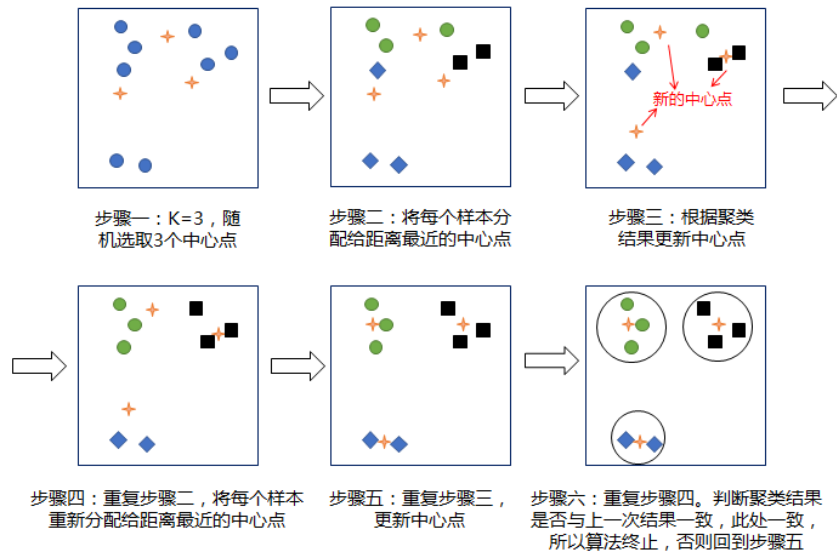


图 10 K-means 算法具体步骤示意图

### 5.2.2 基于含权重的聚类中心点决策扇区主方向

#### (1) 对弱覆盖点进行聚类

基于 K-means 算法，采用 Python 编程语言，首先，导入问题一求解的各基站点数据，运用 K-means 算法对数据进行聚类分析，由于实际上基站由三个覆盖范围从主方向往两侧递减的扇区决定，设定参数  $k = 3$ 。

本文对某一个宏基站进行分析，在进行聚类时，主要考虑一下几方面：

第一，每个聚类中的点应该相对集中，这样就能在扇区覆盖范围内尽可能多地覆盖弱覆盖点，从而提升每个基站覆盖的业务量。

第二，聚类的数量。这个数是需要预先设定的，考虑到实际情况，根据一个基站有三个扇区得到  $k$  值为 3。

第三，业务量大的点优先考虑，这样就在保证数量的情况下充分选择了质量高的点。

该基站覆盖范围内弱覆盖点聚类如下图所示，其中星号代表该宏基站，三种不同颜色的点分别代表三类弱覆盖点，圆为该宏基站的覆盖范围。

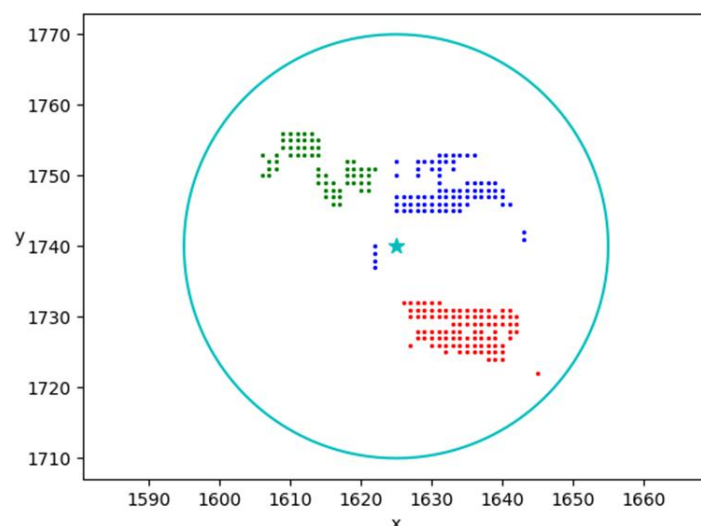


图 11 某宏基站范围内弱覆盖点聚类示意图

## (2)扇区主方向角度分配原则

基于覆盖最多弱覆盖点和最大业务量的原则，本文选取 K-means 算法聚类后的三个中心点与基站的连线为初始主方向，从主方向往两侧线性递减建立扇区。对主方向之间夹角小于 45 度的决策方案再次细化聚类，从而调整主方向角度。由于经过对大业务量弱覆盖点的复制处理后已经确保三个聚类满足最大业务量的原则，考虑到中心点所在区域弱覆盖点最密集，主方向上覆盖范围最大可以保证覆盖点数量最大化。

## 5.4 扇区角度结果

采用问题一的建站方案结果，对于各基站点的扇区划分，以东南西北四个地理方向，表示 3 个扇区的主方向角度，最终覆盖弱覆盖点总业务量的 78.8546%，详细扇区角度的结果参见支撑材料 2 “问题二：基站与扇区角度情况”。部分基站的扇区角度如下表所示。

表 5 部分宏基站的扇区角度

宏基站坐标	扇区 1 主方向	扇区 2 主方向	扇区 3 主方向
(515, 1987)	南偏西 83.29 度	北偏西 55.80 度	北偏东 23.75 度
(1379, 2312)	北偏东 37.41 度	北偏西 27.05 度	南偏西 61.07 度
(1102, 1126)	北偏西 88.59 度	南偏东 63.43 度	南偏西 75.96 度
(1323, 1389)	北偏东 70.87 度	南偏西 3.18 度	南偏西 53.13 度
(1257, 2372)	北偏西 15.52 度	南偏东 45 度	南偏西 35.0 度
(1522,1536)	南偏东 47.27 度	北偏东 78.28 度	南偏西 68.11 度
.....	.....	.....	.....

表 6 部分微基站的扇区角度

微基站坐标	扇区 1 主方向	扇区 2 主方向	扇区 3 主方向
(1195, 2290)	正西	南偏东 75.96 度	北偏东 71.57 度
(1129, 1894)	北偏西 79.70 度	北偏西 45.0 度	北偏西 82.41 度
(1611, 2233)	南偏东 7.59 度	北偏东 44.50 度	南偏东 9.46 度
(1370, 2333)	南偏西 29.36 度	北偏西 14.04 度	南偏西 33.69 度
(1194, 2290)	北偏东 75.96 度	正东	南偏东 80.54 度
(1363, 2303)	南偏东 18.43 度	北偏东 70.56 度	南偏东 68.20 度
.....	.....	.....	.....

下图为某一个宏基站的扇区主方向示意图。

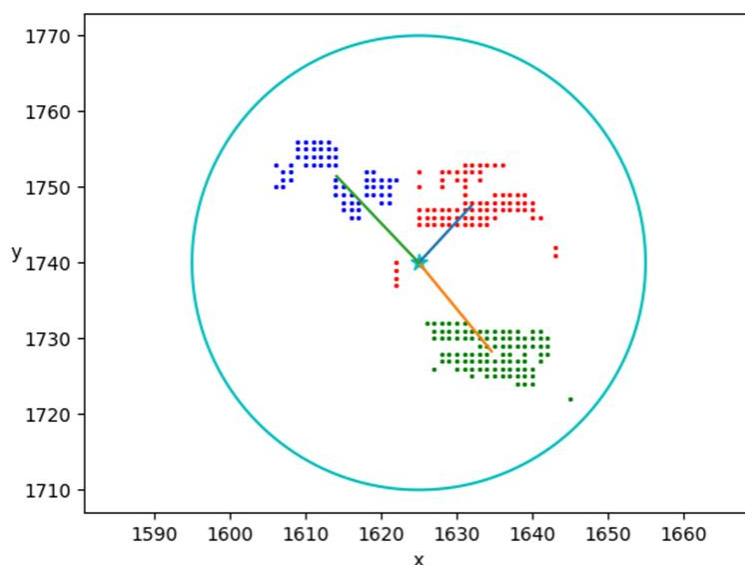


图 12 某宏基站的扇区主方向示意图

## 六、问题三的模型建立与求解

### 6.1 基于 DBSCAN 算法进行区域聚类

DBSCAN 算法是基于空间密度的一种聚类算法。它与 K-means 算法不同，不需要事前确定聚类最终数量，起始点的选择比影响聚类结果，同时可以识别出噪声点（即离群点）。

DBSCAN 算法以数据本身聚类的结果，显示聚类数量，适用于拥有大量数据点的弱覆盖点的聚类问题。而 K-means 算法，需事先确定数据聚类数量，起始点的选取会较大程度地影响聚类结果，并且难以发现形状型的簇。因此本文采用 DBSCAN 算法进行聚类分析。

#### 6.1.1 DBSCAN 算法所需参数

DBSCAN 聚类算法需确定两个参数：

- (1)  $Eps$  是以一个样点为圆心的邻近周围区域的圆半径。根据问题三，要求聚类后的弱覆盖点集合内，任意两个弱覆盖点间的距离小于等于 20。因此  $Eps$  取值为 20。
- (2)  $MinPts$  是一个样点的圆形邻近区域内的最小样点数（包含该样点）。因是弱覆盖点间两两相互比较，所以可能存在某样本点距离其余样本点均大于 20，此时该点作为单独一个簇。因此  $MinPts$  取值为 1。

### 6.1.2 DBSCAN 算法具体步骤

步骤 1：随机选取 1 个弱覆盖点，以该弱覆盖点为圆心，按照  $Eps = 20$  绘制圆。当圆内样本点数大于等于  $MinPts$  时，该点为核心点，并将这些圆内样本点聚为一类。

步骤 2：继续以核心点为圆心的圆中，其他样本点作为圆心，持续进行步骤 1 的操作，最终连通所有的核心点形成一个簇，归为一个聚类。

步骤 3：重复上述两个步骤，最终将所有弱覆盖点进行聚类。

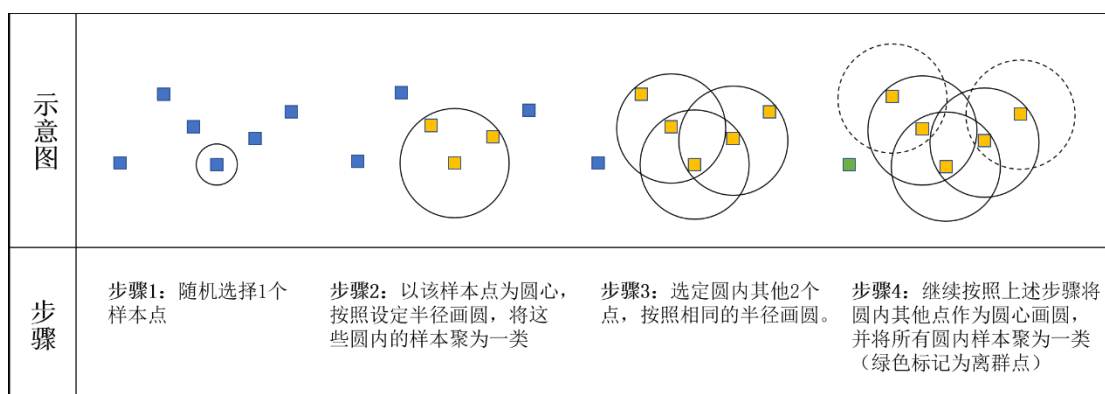


图 13 DBSCAN 算法具体步骤示意图

## 6.2 弱覆盖点聚类的求解

基于 DBSCAN 聚类算法，采用 Python 编程语言，运用附件 1 的 182 807 个弱覆盖点的数据，对弱覆盖点进行聚类分析。本文所用聚类方法的总时间复杂度为  $O(N\log N)$ ，时间复杂度较低。

运用 DBSCAN 算法对数据进行聚类分析，将所有弱覆盖点聚为 **898** 类，部分聚类情况如下表所示，完整聚类情况参见**支撑材料 3**“问题三：弱覆盖点聚类情况”。

表 7 弱覆盖点的聚类情况列举

弱覆盖点坐标		聚类标签
X 坐标	Y 坐标	组号
66	1486	0
67	1486	0
177	1486	1
187	1486	1
.....	.....	.....
1495	124	39
1829	124	43
.....	.....	.....

通过将不同类别的弱覆盖点以不同颜色进行可视化,可更加直观观测到各类样本点的大致聚类情况。由于 Python 中颜色种类有限,不同聚类之间可能会使用相同颜色。值得注意的是,在大面积同一颜色的区域中,会存在有个别点状的另一颜色;以该点为圆心 20 为半径的圆中,不存在任何除自身以外的其余弱覆盖点,因此单独成一簇;因此该现象为符合聚类要求的正常聚类情况。弱覆盖点聚类的可视化如下图所示。

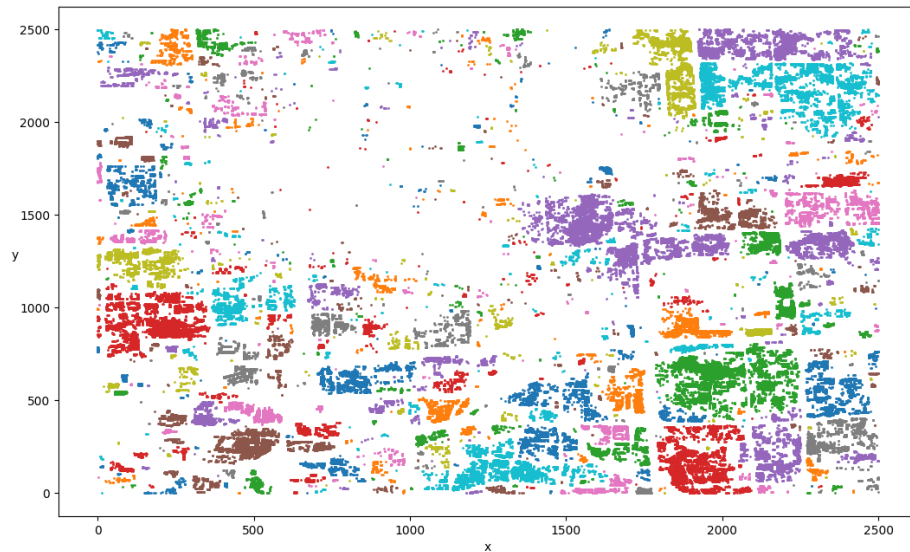


图 14 弱覆盖点的聚类可视化

## 七、模型的评价、改进与推广

### 7.1 模型的优点

(1) 本文在第一题中采用免疫算法作为主要解题思路,相比于传统的遗传算法,免疫算法

具有能够保留优秀父代个体的能力,使算法运算过程中产生的信息损失尽可能的减少。同时,本文采用贪心算法来选取免疫算法中的初代基体,这种方法比简单的随机选取更加有可能筛选出能对整体产生较大影响的个体,使得算法整体的运算速度大幅上升。

(2) 本文在第二问中使用 **K-means** 聚类算法之前对弱覆盖点进行了赋权操作,这一操作的优秀之处在于他不仅考虑了弱覆盖点的空间位置,同时将弱覆盖点的业务量大小这一属性也纳入了聚类算法的考虑之中。

(3) 本文在第三问中采用 **DBSCAN** 算法作为主要解体思路,这种算法的时间复杂度相比于 **K-means** 等常见聚类算法更低,并且该算法的聚类准确率也相对较高。同时, **DBSCAN** 算法基于空间密度进行分类的特质使得他有一定的传递性,且 *Eps* 参数的设置使本算法能够成功解决本题的空间限定要求。

## 7.2 模型的不足

(1) 本文在实验初期进行的数据清洗可能会造成信息缺失的问题,进而对模型的训练产生一定的影响。

(2) 本文建立的模型对于基站之间相互影响程度的关注相对较少,可能在基站建立的过程中产生一定的资源浪费。

## 7.3 模型的改进与推广

### 7.3.1 模型的改进

(1) 模型自身在训练过程中对于计算资源的消耗较大,在未来的研究中我们将试图降低其资源消耗

(2) 考虑到自然环境以及认为影响等外界因素影响,现实中可能会存在某些特定地区无法顺利假设信号基站,在未来模型的改进中我们将把这些变量纳入考虑中。

### 7.3.2 模型的推广

(1) 考虑到小型信号发射装置 (**WIFI**) 的建立与本题的背景有着高度的相似性,本题建立模型的过程可以在进行一定的条件转化之后被运用到企业或大型建筑内的无线信号部署流程中。

(2) 本题中覆盖范围的限定与军事领域中雷达站的部署有着一定的相似性,本文所提出的模型可以在一定程度上推广至防空识别区中雷达站的部署流程。

(3) 农业领域中自动灌溉问题与本题中信号基站覆盖问题有着异曲同工之妙,在进行适当的转化后本文的模型可被应用至农业领域。

## 参考文献

- [1]张英杰,毛赐平,俎云霄,孙先佑.基于免疫算法的 TD-SCDMA 网络基站选址优化[J].通信学报,2014,35(05):44-48.
- [2]席裕庚,柴天佑,恽为民.遗传算法综述[J].控制理论与应用,1996(06):697-708.
- [3]陈实. 基于改进的人工免疫算法的 NB-IoT 基站选址研究 [D]. 南京邮电大学,2020.DOI:10.27251/d.cnki.gnjdc.2020.001256.
- [4]胡铮,张国翊,田辉.基于人工免疫的聚合重构资源自治分配[J].北京邮电大学学报,2010,33(02):82-87.
- [5]郑日荣,毛宗源,罗欣贤.基于欧氏距离和精英交叉的免疫算法研究[J].控制与决策,2005(02):161-164+169.DOI:10.13195/j.cd.2005.02.42.zhengrr.009.
- [6]宋建厚,陈良琼,刘道华.改进的免疫算法参数自适应调整的优化设计[J].计算机测量与控制,2013,21(05):1297-1300.DOI:10.16526/j.cnki.11-4762/tp.2013.05.078.
- [7]丁晓群,周春泉,杨保,姜相明.基于多种群免疫算法的分布式电源选址与定容[J].电测与仪表,2014,51(04):29-34.
- [8]常友渠,肖贵元,曾敏.贪心算法的探讨与研究[J].重庆电力高等专科学校学报,2008(03):40-42+47.
- [9]周丽娟,王慧,王文伯,张宁.面向海量数据的并行 KMeans 算法[J].华中科技大学学报(自然科学版),2012,40(S1):150-152.DOI:10.13245/j.hust.2012.s1.022.
- [10]周水庚,周傲英,曹晶.基于数据分区的 DBSCAN 算法[J].计算机研究与发展,2000(10):1153-1159.
- [11]段明秀. 层次聚类算法的研究及应用[D].中南大学,2009.