

基于K-Means++算法的网络基站选址分析

邓 敏, 肖 琳, 吴建平*

湖南科技学院理学院, 湖南 永州

收稿日期: 2025年4月14日; 录用日期: 2025年5月7日; 发布日期: 2025年5月14日

摘 要

本文主要研究5G通信基站的建设及其站址规划, 针对现网弱覆盖区域的覆盖问题, 综合考虑基站覆盖半径、基站类型以及信号覆盖范围等因素。采用K-means++聚类算法, 建立站址选择模型, 得出了新基站建立的最优方案。

关键词

K-Means++算法, 0-1规划模型, 基站选址

Analysis of Network Base Station Location Selection Based on K-Means++ Algorithm

Min Deng, Lin Xiao, Jianping Wu*

College of Science, Hunan University of Science and Engineering, Yongzhou Hunan

Received: Apr. 14th, 2025; accepted: May 7th, 2025; published: May 14th, 2025

Abstract

This paper mainly focuses on the construction and site planning of 5G communication base stations. In response to the coverage issues in weak coverage areas of the current network, various factors such as the coverage radius of base stations, types of base stations, and signal coverage range are comprehensively considered. The K-means++ clustering algorithm is adopted to establish a site selection model, and the optimal plan for the establishment of new base stations is obtained.

Keywords

K-Means++ Algorithm, 0-1 Programming Model, Selection of Base Station Locations

*通讯作者。

文章引用: 邓敏, 肖琳, 吴建平. 基于 K-Means++算法的网络基站选址分析[J]. 应用数学进展, 2025, 14(5): 91-99.
DOI: 10.12677/aam.2025.145237



1. 需要解决的问题

本文研究的问题来源于 2022 年第十二届 MathorCup 高校数学建模挑战赛 D 题。题目要求根据收集到的弱覆盖点数据,使得弱覆盖点总业务量的 90%被规划基站覆盖,同时尽量降低成本;另外,根据问题所给数据,考虑每个站的任意 2 个扇区的主方向之间的夹角不能小于 45° 的情况下探讨是否能够新建基站使得其覆盖弱覆盖点总业务量达到 90%以上。本论文结合数学建模优化方法建立站址选择模型,基于 K-means++ 算法得出满足预期的最优基站选址方案。部分重要参考数据如下表(见表 1、表 2) (具体数据见:第十二届 MathorCup 高校数学建模挑战赛 D 题)。

Table 1. Weak overlay raster data**表 1.** 弱覆盖栅格数据

x	y	traffic
66	1486	140.58139
67	1486	140.518829
177	1486	48.919178
187	1486	4.322495
284	1486	71.528404

宏微基站覆盖范围、建设成本如下表 2 所示。

Table 2. Types, coverage range and cost of buildable base stations**表 2.** 可建设基站的种类、覆盖范围与成本

基站种类	覆盖范围/ km^2	成本/亿元
宏基站	30	10
微基站	10	1

2. 基站选址的数学模型

2.1. 基站选址的数学模型分析

现有基站与新建基站之间的门限是 10, 根据二范数公式[1]:

$$\|P - P_0\| = \sqrt{(x - x_i)^2 + (y - y_i)^2} \leq 10$$

其中, (x_i, y_i) 为现有基站 i 的坐标, 并对业务量为 1 以下的弱覆盖点进行筛选。

目标函数要求基站选址结果的建设成本之和最小: $\min f = \sum_{i=1}^n (10v_i\omega_i + v_i\nu_i)$ 。

约束条件包含以下几个方面:

约束一 新建基站的类型选取情况:

$$\omega_i + \nu_i = 1$$

约束二 弱覆盖点 90%的业务被覆盖:

$$\sum_{h=1}^n w_h \geq 6350607$$

约束三 两个新建基站之间的门限大于 10:

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

约束四 微基站的覆盖范围为 10 km²:

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

约束五 宏基站的覆盖范围为 30 km²:

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

其中，符号所表示的含义如下表 3 所示。

Table 3. Symbol meaning table
表 3. 符号含义表

序号	符号	符号说明
1	ω_i	新建基站 i 为宏基站的情况
2	ν_i	新建基站 i 为微基站的情况
3	ν_i	候选基站为 i 为新基站的情况
4	n	候选基站的个数
5	m	弱覆盖点的个数
6	w_i	覆盖点 i 的业务量
7	w_h	被覆盖到的弱覆盖点 h 的业务量

2.2. 基站选址模型

s.t

$$\begin{aligned} \min f &= \sum_{i=1}^n (10\nu_i\omega_i + \nu_i\nu_i) \\ \omega_i + \nu_i &= 1 \\ \sum_{h=1}^n w_h &\geq 6350607 \\ \sqrt{(x-x_i)^2+(y-y_i)^2} &> 10 \\ \sqrt{(x-x_h)^2+(y-y_h)^2} &\leq 10 \\ \sqrt{(x-x_h)^2+(y-y_h)^2} &\leq 30 \\ \nu_i, \omega_i, \nu_i &\in \{0,1\} \\ i &= 1, 2, \cdots, n \\ s &= 1, 2, \cdots, m \end{aligned}$$

3. 基于 K-Means++ 算法的基站选址模型

3.1. k 值的选择

聚类算法是通过量化样本间距离和聚类的紧密性来评价聚类效果的好坏。具体表示为：同一个类内数据间的距离越小越好、不同类间的距离越大越好。本文将同一类簇内的差异值

$$E_{in} = \sum_{x \in C_i} \frac{1}{n} \sum_{i=1}^k \|x - \bar{Z}_i\|^2$$

和不同类簇间的差异值的比值定义为评价聚类效果优劣的标准[2]。

n 是表示数据集中所有的样本数量, k 表示类的个数, C_i 表示第 i 个类簇, x 表示在第 i 簇类聚类点的数值, $\bar{Z}_i$ 表示第 i 个类的中心点。类簇内距离和 E_{in} 值越小, 意味着所有类簇中的数据样本越紧凑, 聚类的紧密型越好。

另外, 还需要计算各个类簇中心点相互间的距离作为类簇间距离, 类簇间距离定义为 E_{out} , 类簇间距离 E_{out} 的计算公式为:

$$E_{out} = \frac{1}{k} \sum_{i=1}^k \|C_i - \bar{C}\|^2$$

其中, k 表示类的数目, 即类簇数, C_i 表示第 i 个类簇的聚类中心点, $\bar{C}$ 表示各个类簇的聚类中心平均值, 类簇间距离 E_{out} 的值越大, 意味着类和类之间的相似度小。

用 EE 表示聚类效果优劣程度的评价标准, 则 EE 表示为:

$$EE = \frac{E_{in}}{E_{out}}$$

由上式可以得出, EE 值为先逐渐变小然后会慢慢变大, EE 值越小, 表示聚类的效果越好, 当 EE 达到局部最小值时, 聚类效果最佳, 由此可以确定聚类数目值 k 。

3.2. 距离度量

将对象点分到距离聚类中心最近的那个簇中需要最邻近的度量策略, 在欧式空间中采用的是欧氏距离。

3.3. 新质心的计算

先选择数据集中任意 k 个初始的聚类中心点, 计算数据集剩余样本到离地最近聚类中心的距离, 定义该距离为 $D(X)$, 为每个类重新选择聚类中心点, 此时选取概率与 $D(X)$ 的大小成正比。

3.4. 求解过程

分析图 1 所示的流程图, 按业务量降序排列聚类中心点 p_i 后, 从最大业务量点开始依次选取聚类中心点 p_i , 以 p_i 为圆心, 分别建立宏基站和微基站, 计算出宏基站和微基站的总业务量分别为 n_1 、 n_2 。依据尽可能多的覆盖面积的原则, 设计判断条件。判断条件如下: 如果 $n_1/n_2 < 10$ 并且 $n_1 - n_2 < 100$, 则选择微基站标记为 1, 否则选择宏基站标记为 0。

3.5. 模型一求解

经规划, 可以得出需要建设的宏基站个数为 3303 个, 在此地区的分布较为均匀。需要建设的微基站个数为 217 个, 主要分布在区域的右下角。在此时新建基站的业务量占弱覆盖点总业务量的 90.532%, 同时总成本达到最小值 33,247。

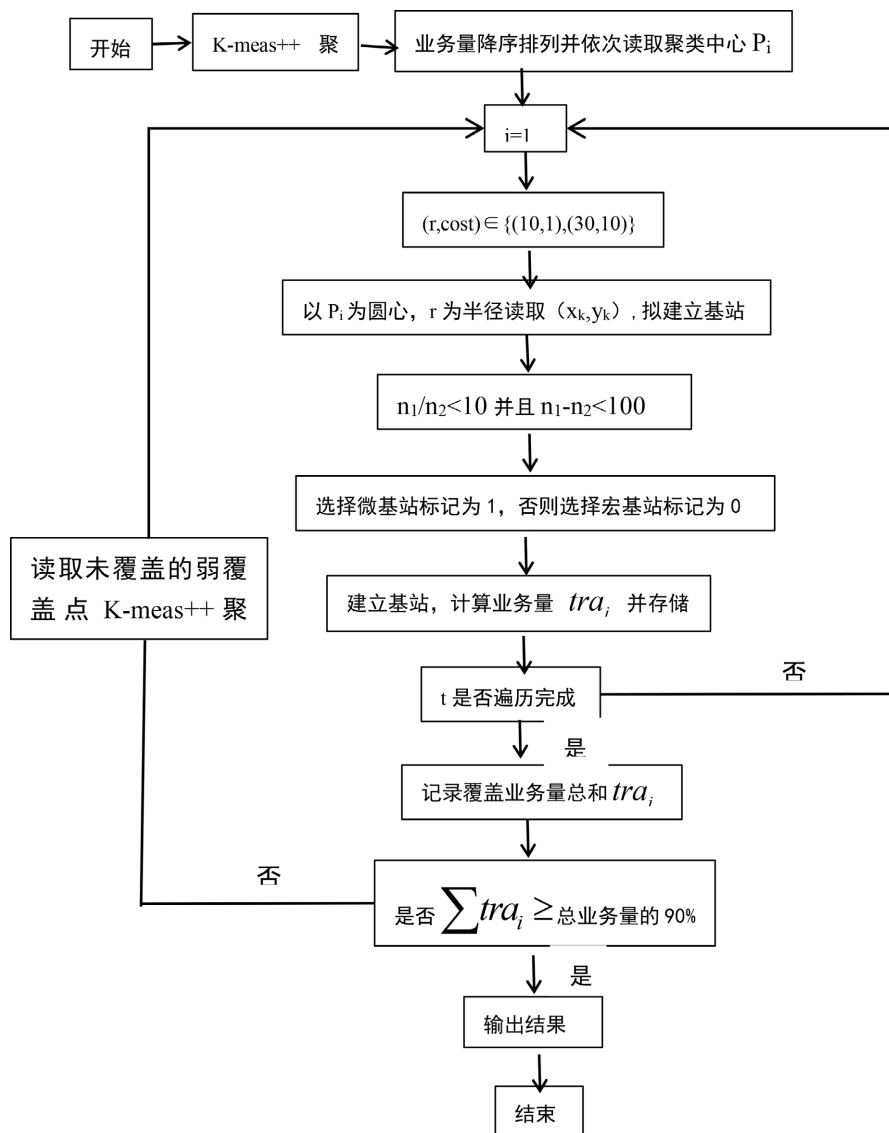


Figure 1. Algorithm flow
图 1. 算法流程

使用 K-means++ 聚类算法进行 18 万组弱覆盖点的处理之后, 其全体的分类如下表 4 所示。

Table 4. Base station coordinates (partial)
表 4. 基站坐标(部分)

宏基站选址	(1340, 2285)	(866, 1226)	(616, 2416)	(700, 422)	(521, 1965)
微基站选址	(480, 1306)	(204, 1762)	(260, 996)	(1684, 270)	(522, 204)

4. 基于赋权的 K-Means++ 算法主方向决策模型

模型二需要在模型一的基础上对所选基站的坐标进行深入分析, 即覆盖区域由原来的圆形变成三个扇形区域, 对角度变化范围进行判断, 观察覆盖的业务量占比是否仍然大于 90%, 以角度为自变量建立规划模型, 根据题目的要求可以绘制出各基站的最广覆盖角度与最小覆盖角度, 如下图 2 所示。

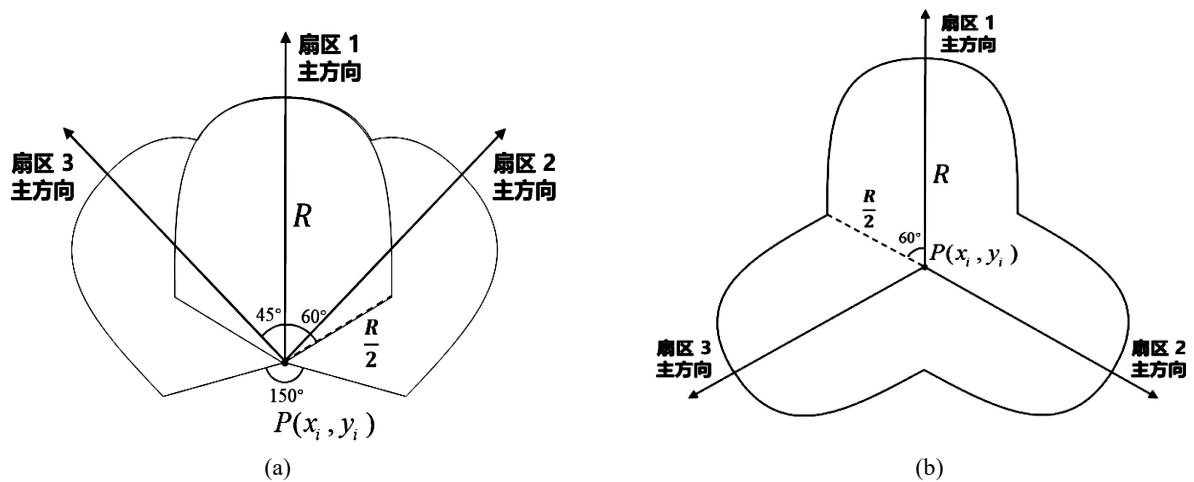


Figure 2. Interval end points for the range of base station coverage angle: (a) Minimum covering angle; (b) Maximum coverage angle

图 2. 基站覆盖角度范围的区间端点情况: (a) 最小覆盖角度; (b) 最广覆盖角度

4.1. 基站扇区主方向角度的计算

通过使用加权 K-means 算法求解分别得到三个聚类区域, 设聚类中心的坐标为 (x_k, y_k) , 已知基站坐标为 (x_i, y_i) , 根据欧式距离公式可得基站到聚类中心的欧式距离为 s_1 , 则设弱覆盖点 (x_h, y_h) 与基站的欧式距离为 s_2 , 聚类中心点与弱覆盖点的欧式距离为 s_3 , 可得:

$$s_1 = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}$$

$$s_2 = \sqrt{(x_i - x_h)^2 + (y_i - y_h)^2}$$

$$s_3 = \sqrt{(x_k - x_h)^2 + (y_k - y_h)^2}$$

设 s_1 和 s_3 之间的夹角为 θ , 由任意三角形余弦公式和反三角公式[3]可得:

$$\cos \theta = \frac{s_1^2 + s_2^2 - s_3^2}{2s_1s_2}, \quad \theta = \arccos \frac{s_1^2 + s_2^2 - s_3^2}{2s_1s_2}$$

4.2. 基站覆盖范围的计算

接着, 建立 $\rho-\theta$ 极坐标系, 以水平轴为 0° , 由本题可知扇形区域的主方向角度变化为: $(0^\circ, 360^\circ)$, 任意两个主方向之间的夹角应大于 45° , 且覆盖范围按线性逐渐缩小, 假设与主方向 60° 角的半径为 R , 则半径公式[4]为:

$$R = -k\theta + \rho, \quad \theta \in [0, 2\pi]$$

其中, k 为半径 R 的线性衰减系数, θ 为半径 R 的自变量, ρ 为基站扇区主方向最大覆盖距离, 宏基站时 $\rho = 30$, 微基站时 $\rho = 10$ 。

由于在 60° 的时候, 覆盖范围为主方向覆盖范围的一半, 因此可以得到宏基(微基站)站半径衰减:

$$R_{\text{宏}} = -\frac{45}{\pi}\theta + 30, \quad R_{\text{微}} = -\frac{15}{\pi}\theta + 10$$

每个站的任意两个扇区的主方向之间的夹角不能小于 45° , 对主方向之间夹角小于 45° 的决策方案再次细化聚类。设每个站的任意两个扇区的主方向之间的夹角为 α , 假设第 i 个扇区的主方向直线斜率为

k_i , 可得:

$$k_i = \frac{y_k - y_i}{x_k - x_i}$$

两条直线之间的角度, 由正切公式和反三角公式可得[4]:

$$\alpha = \arctan \frac{k_2 - k_1}{1 + k_1 \cdot k_2}$$

最终的模型为:

$$\begin{aligned} \min f &= \sum_{i=1}^n (10v_i\omega_i + v_i v_i) \\ \text{s.t.} \quad &\omega_i + v_i = 1 \\ &\sum_{h=1}^n w_h \geq 6350607 \\ &\sqrt{(x-x_i)^2 + (y-y_i)^2} > 10 \\ &\sqrt{(x-x_h)^2 + (y-y_h)^2} \leq 10 \\ &\sqrt{(x-x_h)^2 + (y-y_h)^2} \leq 30 \\ &\theta \leq 60^\circ \\ &\alpha \geq 45^\circ \\ &v_i, \omega_i, v_i \in \{0,1\} \\ &i = 1, 2, \dots, n \\ &s = 1, 2, \dots, m \end{aligned}$$

4.3. 模型二求解

由于 K-means 算法仅将弱覆盖点的各个坐标进行聚类分析, 结合数学建模的模型假设, 本文忽略弱覆盖点业务量这一影响因素, 为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[5]。弱覆盖点的权重与业务量的关系如下式所示:

$$\varphi_i = \frac{\omega_i}{10}$$

其中, φ_i 是弱覆盖点 i 的权重, ω_i 为弱覆盖点 i 的业务量。

基于 K-means++ 算法, 首先, 导入模型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, 主要考虑以下几方面:

1) 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;

2) 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 $k=3$;

3) 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 3 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。

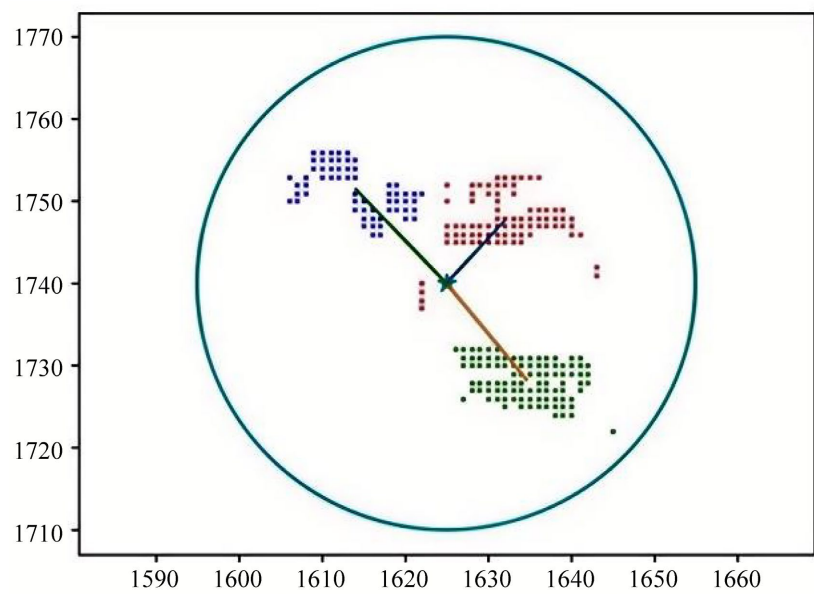


Figure 3. Sector main direction diagram of a macro base station
图 3. 某宏基站的扇区主方向示意图

通过 Matlab 2020 对加权 K-means++ 算法求解得到的结果如下表 5、表 6 所示。

Table 5. Angle of the main direction of the macro base station (partial)
表 5. 宏基站主方向的角度(部分)

(X, Y)	角度 1/度	角度 2/度	角度 3/度
(0, 951)	1.835932	53.25669	97.75694
(1, 1281)	4.992284	57.47109	98.84858
(4, 853)	3.888901	63.92687	154.0484
(4, 1254)	4.412131	83.89204	115.0877
(4, 1350)	2.998948	66.42517	111.3229

Table 6. Angle of the main direction of micro base station (partial)
表 6. 微基站主方向的角度(部分)

(X, Y)	角度 1/度	角度 2/度	角度 3/度
(2227, 2442)	52.84472	131.3967	231.9516
(1872, 2441)	73.86334	139.7094	263.1159
(2166, 2440)	49.74416	106.0203	249.4436
(1848, 2439)	57.29269	131.37052	266.2685
(1926, 2439)	70.01989	122.304	246.6699

最终求解出的扇区主方向的角度分别为： 184.358° 、 -248.164° 、 -49.377° 。使用贪心算法求解在编号 1 的基站求解结束之后，去掉已经被 1 号基站的三个扇形覆盖的所有弱信号点。然后考察编号 2 的基站周围的所有弱信号点情况，并再次进行聚类，得到三个角度，同理，针对其他基站也做这样的步骤，最终得到的结果如下表 7 所示。

Table 7. Service volume and proportion in main direction
表 7. 主方向服务量以及占比情况

主方向 1/度	主方向 2/度	主方向 3/度	是否满足要求	服务量	比例
283.6777	-203.638	-8.2555	否	94,123	77.921%
232.6573	-111.668	-29.7508	否	21,547	79.053%
-41.5009	-141.595	-115.732	否	45,654	83.647%
272.1131	175.901	-82.0873	是	64,648	92.936%
81.3648	-190.54	-50.3116	是	-	-
212.2492	-237.215	33.80057	是	-	-

经规划，现有基站与新建基站共覆盖弱覆盖点总业务量为 642,783,008，占弱覆盖点总业务量的 90.09%。规划基站共有 1404 个，其中现有基站有 848 个，新建基站数量为 556 个。将模型一、模型二中的结果进行对比可以发现：在各基站覆盖范围由整圆减小为三个扇区后，仍然使得规划基站所覆盖的业务总量超过弱覆盖点业务量的 90%。

5. 结论

本文从数学建模的角度将 5G 宏基站和 5G 微基站的建设成本作为目标函数，将信号覆盖率、系统的容量以及信号干扰作为约束条件，在 5G 网络信号覆盖率以及网络容量较高的情况下，实现基站建站成本 5G 网络最小化。最后，以某待规划区域 5G 网络基站选址为例，通过 Matlab 2020 软件仿真实验，实现了 5G 网络基站选址寻优。结果验证了本文所设计的求解 5G 网络基站选址优化数学模型和所使用的 K-means++这一求解算法的有效性。

基金项目

永州市 2022 科技指导性科技计划项目；湖南科技学院科研项目(湘科院校发[2022] 108 号)；2023 年度湖南省大学生创新创业训练计划一般项目(湘教发[2023] 237 号)；湖南科技学院 2023 年大学生创新创业训练计划项目(湘科院校发[2022] 106 号-17)。

参考文献

[1] 郭文娟. 基于优化初始聚类中心的 K-means 聚类算法[J]. 科技风, 2022(4): 63-65.
[2] 沈栩竹, 李江云, 陈丽萍, 等. 5G 信号基站的选址和设备选择问题[J]. 昆明冶金高等专科学校学报, 2022, 38(6): 89-95.
[3] 史秀岭. K-means 聚类优化算法的研究[D]: [硕士学位论文]. 长沙: 长沙理工大学, 2011.
[4] 葛春悦. 基于 K-Means 农产品物流中心选址研究[J]. 中国储运, 2021(10): 152.
[5] 徐文进, 管克航, 寻晴晴, 等. 基于 KNN 算法的改进 K-means 算法[J]. 青岛科技大学学报(自然科学版), 2019, 40(5): 107-111.