EAI Endorsed Transactions
on e-Learning
Research Article EAI.EU

# Research on Location Selection of Mobile Communication Weak Coverage Network Based on Hierarchical K-means Clustering Algorithm

Yuan Huang[1], Zezhong Huang[1], Shijun Liu[2], Liying Lan[3,*]

[1]Faculty of Electronic Information Engineering, Gannan University of science and technology, Ganzhou 341000, China
[2]Faculty of Intelligent Manufacturing and Automotive Engineering, Gannan University of science and technology, Ganzhou 341000, China
[3]Faculty of Humanity and Law, Gannan University of science and technology, Ganzhou 341000, China

## Abstract

At present, the problem of weak coverage of mobile communication still exists. Firstly, this thesis determines the scale of the base station and the location of the network with the hierarchical clustering algorithm. Secondly, this thesis also determines the base station coverage with the sector area of the Archimedes screw, the location of the weak coverage point with the discriminative model of weak coverage point, and the angle of the sector with the coverage model of gradient descent method. Finally, this thesis provides better solutions to the weak coverage problem with the propagation model of cluster analysis, and optimizes the site selection with the correction model of the propagation pass loss. Within the certain cost, we aim to increase the volume of coverage business, decrease the time complexity of the model, and improve the efficiency of the site selection.

Received on 07 May 2023; accepted on 30 May 2023; published on 23 June 2023

**Keywords:** Network site planning, Weak coverage, $k - means$ clustering, Dynamic Cluster Analysis Algorithm



doi:10.4108/eetel.v8i3.3319

## 1. Introduction

The planning of web addresses in weak coverage areas[1] is closely related to mobile communication. The rapid development of mobile communication technology has entered the development stage of the fifth generation mobile communication (5G). 5G mobile communication technology (fifth-generation, 5G) is the next generation of wireless mobile communication technology developed to meet the rapid popularization of intelligent terminals and the rapid development of mobile Internet after 4G. And a wireless mobile communication system[2] for the needs of the human information society after 2020.

The planning of mobile communication network[3–5] is to follow certain principles, which can be roughly divided into the following aspects: Firstly, in the planning process, the quality of service should be guaranteed to meet people's needs for network or communication; at the same time, a certain range of services should be ensured to better meet the needs of most densely populated areas; Secondly, it is also necessary to control the cost of construction to reduce the cost and waste of building base stations; Then, when planning, designing and arranging the network, use sustainable development, and observe the planning process from a comprehensive perspective; Finally, for users of different natures and in different application environments, they should be treated differently in network planning, and a planning scheme that is suitable for various users and in line with their characteristics should be proposed to meet the needs of users of different natures as much as possible.

The bandwidth of 5G communication[6] increases, but the coverage area of base stations is reduced, resulting in more base stations required to cover the same area. In the actual network planning and site

*Corresponding author. Email: lanliying318@126.com

selection[7], the following site selection principles[8] should be followed:

1. For the construction cost and coverage of base stations[9], when it is not possible to solve all weak coverage areas[10], it is best to prioritize solving areas with high business volume, so as to achieve the optimal construction cost and maximize the effective coverage of business volume.

2. The distance threshold between newly built sites and between newly built sites and existing sites[11] needs to be considered as 10.

3. Full consideration should be given to safety issues, convenience in construction and maintenance, selection of safe, hygienic, and non strong interference station sites, and implementation of lightning and flood prevention measures[12].

4. There are 3 sectors on each base station and each sector has a main firing direction. It can be covered within a range of about 60 degrees in the main shooting direction, and the coverage gradually decreases linearly. Moreover, when it exceeds 60 degrees, it cannot be covered by the sector.

5. For special areas such as public green spaces, the transmission cables of base stations and the introduction cables of urban power need to be buried and constructed.

In order to intuitively understand the relationship between the model and the actual base station, this paper regards the area as a uniform grid, each grid is regarded as a point, and the construction of the base station can be carried out on each point. One is a macro base station[13] (coverage 30, cost 10), and the other is a micro base station[14] (coverage 10, cost 1).

Firstly, this article improves and studies the classic k-means clustering algorithm [15], establishes a location selection model based on hierarchical k-means clustering algorithm[16], and uses the clustering center to preliminarily determine the website address. Then, based on the specific distance threshold constraints between base stations, base station size constraints, the total amount of base station coverage business as large as possible, and the cost as low as possible, an optimization model based on mobile communication network planning is established, And determine the scale and network location of the base station.

In order to determine the three sector areas of the coverage area of each base station in practice, a sector area model based on the Archimedes spiral is established. In order to determine whether the weak coverage points fall within the coverage range of the base station, a discriminative model based on the weak coverage points is established. In order to solve the problem of each sector having a main direction, which cannot be covered if it exceeds 60 degrees, and the coverage range of the sector gradually decreases linearly, a coverage model based on gradient descent is established[17]. Finally, through the above optimization model, we can determine the optimal location of weak overlay network in real life.

In practical work, in order to better solve the problem of weak coverage, the propagation environment features of the weak coverage area are first extracted, and the centroid (site) is determined through dynamic clustering analysis algorithms and clustering criterion functions, thereby establishing a propagation model based on clustering analysis. In order to demonstrate the transitivity of clustering properties, a propagation model correction model based on path loss is established[18]. This model not only optimizes network location, but also has a time complexity of $O(l * \log l + n * \log n + m * \log m + k * \log k)$ for dynamic clustering analysis algorithms, which is lower than the previous $O(l * n * k * m)$ and thus improves the efficiency of problem solving.

5G mobile communication weak coverage problem During the COVID-19 epidemic, when building shelter hospitals, quickly solve the problem of 5G weak coverage, make Transmission delay as low as possible on the equipment, and build a perfect network system[19–21].

The structure of this article is as follows. Section 2 introduces the establishment of weak overlay network location model from the coverage of the base station in the circular domain and three sectors. Section 3 is about model improvement, analyzing the model and method of weak coverage points for regional clustering, thereby improving the algorithm's time complexity. Section 4 presents the corresponding experimental results and analysis. Section 5 is the summarize.

## 2. Establishment of location model for weak overlay network

### 2.1. The coverage range of the base station is circular domain

For 2500 × 2500 grids, i.e. 2500 × 2500 points, with a horizontal coordinate range of 0 to 2499 and a vertical coordinate range of 0 to 2499. Attachment 1 contains information on the weak coverage points in the selected area, including the coordinates and traffic volume of each point. There are currently two types of base stations, namely: macro base stations (coverage range 30, cost 10) and micro base stations (coverage range 10, cost 1). Attachment 2 also provides the coordinate points of the existing network base stations, and the threshold for the distance between the newly built site and the existing site is 10 (https://github.com/Sunlight312/Data-support-materials.git). Now proceed with site planning to ensure that 90% of the total business volume of weak coverage points is covered by the planned base stations. This article provides the coordinates of the selected site and the type of base station selected for each site.

Under the idea of grid, the weak coverage area is placed in the coordinate grid, and the visual scatter diagram is drawn by MATLAB to obtain the distribution of the weak coverage area as shown in Figure 1.

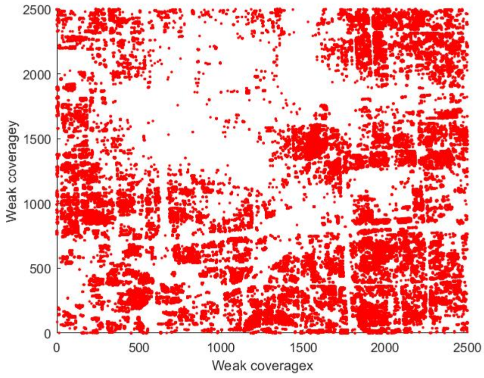


**Figure 1.** Weak coverage distribution map

Looking at the picture, we can see that most of the weak coverage areas are concentrated in clusters.

**Location model based on k-means clustering algorithm.** First of all, this paper adopts the method of random selection to set as the initial number of clusters.

Let $U = \{u_1, u_2, \cdots u_v \cdots u_n\}$ be the address of $n$ weak coverage points of $R^d$ space. Randomly select $k$ addresses from $n$ addresses as the initial clustering centers, so there are also $k$ clustering centers for class $k$.

Assuming that $a_j$ is the cluster center point of the $j$-th class, the Euclidean distance between $u_i$ and $a_j$ is obtained as:

$$d(i,j) = \sqrt{\left(u_i^1 - a_j^1\right)^2 + \cdots + \left(u_i^d - a_j^d\right)^2} \quad (1)$$

Among them, $u_i^d$ is the address of the $i$-th weak coverage point in $U$ space.

Then use the reciprocal of the distance from the weak coverage area to the cluster center to represent the similarity between $x_i$ and $a_j$:

$$s(i,j) = \frac{1}{d(i,j)} \quad (2)$$

Second, in order to get the cluster centers in the continuously updated class. At this time, it is assumed that the sample value in the $j$-th class is $\{u_{j1}, u_{j2}, \cdots, u_{jn_j}\}$, there are a total of $n_j$ samples. The samples are weighted and summed, then divided by the number of total samples to get the average of the samples, that is: the cluster center of this class is $\left(a_j^1, a_j^2, \cdots a_j^k, \cdots, a_j^d\right)$, and the obtained formula is as follows:

$$a_j^k = \frac{u_{j1}^k + u_{j2}^k + \cdots + u_{jn_j}^k}{n_j} \quad (3)$$

In the above formula, $u_{jn_j}$ is represented as the $n_j$-th sample of the $j$-th class, and $a_j^k$ is represented as the $k$-th attribute of the class center $a_j^k$.

According to formula (1) and (3), the cluster center of a certain class is obtained by continuous iteration, and then the mean square error is used to continuously measure the similarity (distance) between $u_{jnj}$ and $a_j$ until it tends to be stable, so as to obtain the standard clustering the class measure $G$ is:

$$G = \sqrt{\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_j} \left(u_{ij} - a_j\right)^2}{n-1}} \quad (4)$$

The interior of each class finally obtained is very aggregated, and the Euclidean distance from the cluster center is very small. The situation between different classes is opposite to that between classes. The classes are very scattered, and the Euclidean distance of different classes is very large.

This algorithmic model also has flaws and limitations: The first is that the initial distribution of samples is inconsistent before clustering, and the second is that this method may cause the obtained clustering result to fall into a local minimum, resulting in an unsatisfactory algorithm execution effect.

**Hierarchical K-means clustering location model.** On the basis of the $k-means$ clustering model, a hierarchical $k-means$ clustering method is adopted, and the number of clusters is determined by itself according to the hierarchical structure in the clustering process. By dynamically judging whether the current clustering class of the sample set $U$ is suitable, it is decided whether to perform the next clustering iteration at the level, and the final clustering effect obtained in this way can ensure that the clustering measure function maintains a small value to a certain extent.

**The algorithm for $k-means$ means[15] is as follows:**

- First, set the initial number of clusters $k_i$ by selecting a sample set $U = \{u_1, u_2, \cdots, u_i, \cdots, u_n\}$ containing $n$ data objects, where $i = 1, 2, \cdots, n$. The objective function of initializing clustering is $G^{(0)} = 0.01$, and the number of clustering iterations $t$ is initialized to 1.

- Then randomly select $k_i$ cluster centers, according to measuring the distance between each sample $u_i(i = 1, 2, \ldots n)$ and each class center $a_j(j = 1, 2, \ldots k)$, $u_i$ is classified as the class with the most similar class center, and calculate the clustering measure function value $G^{(1)}$ after the current clustering.

- Next, iterative clustering is performed. Let $u_j \in U_j$, according to the Euclidean distance formula (1), the class $u_j$ with the largest class radius and its class center $a_j$ are obtained, thereby establishing the class radius formula:

$$r_i = \max \left\| u_j - a_i \right\|, j = 1, 2, \cdots, n_i \tag{5}$$

  In the above formula, $r_i$ represents the largest class radius, $u_j$ represents the class with the largest class radius $a_j$ represents the class center.

- Finally, select the sample point $u_{i1}$ that is farthest from class $a_j$ in this class, and then select the sample point $u_{i2}$ that is farthest from $u_{i1}$ in this class. Re-do $k - means$ clustering with these two points and other cluster centers as the initial cluster centers. In order to make the algorithm repeat iteratively, the iterative formula can be obtained:

$$\lambda = \frac{G^{(t)} - G^{(t-1)}}{G^{(t-1)}} \tag{6}$$

  If $\lambda > \Delta$ , return to the previous step to continue iterative execution, otherwise the clustering algorithm ends, and the clustering result $U \to \{U_1, U_2, \cdots, U_k\}$ is output.

  Each clustering result corresponds to a cluster center. Compared with the traditional $k - means$ clustering method, the improved hierarchical $k - means$ clustering method can dynamically obtain a more appropriate final number of clusters, so as to achieve better clustering results.

**Optimization model based on mobile communication network planning.** The goal of network planning optimization is to ensure large coverage while making the cost small. The weak coverage area of a certain place is planned, and the area of the area in Figure 1 is nearly 7,056,230, thereby establishing an optimization model based on mobile communication network planning.

**Decision variables**

This quiz asks whether to establish a macro base station or a micro base station, so the decision variable is whether to establish a macro base station, $g_i = 1$ to establish a macro base station, $p_i = 1$ to establish a micro base station.

**Objective function**

For the construction of macro base station, the cost $v_1 = 10$, the cost of micro base station $v_2 = 1$, try to achieve the minimum construction cost $Z$, so as to obtain the objective function with the minimum construction cost:

$$\min Z = v_1 \sum g_i + v_2 \sum p_i \tag{7}$$

Let the traffic volume of the weak coverage point covered is the largest $W$, the traffic volume covered by the macro base station and the micro base station are $y_i$ and $y_j$ respectively, and the objective function of the largest covered traffic volume is:

$$\max W = \sum_{i=1}^{n} q_i y_i + \sum_{j=1}^{n} p_j y_j \tag{8}$$

**Constraints**

1. For the distance between the weak coverage point $P(x, y)$ and any base station $P_0(x_0, y_0)$, the Euclidean distance $d_u$ formula is used to obtain the distance between them:

$$d_u = \sqrt{(x - x_0)^2 + (y - y_0)^2} \tag{9}$$

2. Coverage constraints of macro base stations:

$$d_{ui} \le 30, i = 1, 2 \cdots n \tag{10}$$

   where $i$ is the weak coverage point.

3. Coverage constraints of micro base stations:

$$d_{uj} \le 10, j = 1, 2 \cdots n \tag{11}$$

   where $j$ is the weak coverage point.

4. The new site and the distance between the new site and the existing site cannot be less than or equal to 10. There are a total of $n$ sites, let their coordinates be $(x_n, y_n)$, so that the constraints are as follows:

$$\sqrt{(x_n - x_0)^2 + (x_n - y_0)^2} \ge 10 \tag{12}$$

5. In order to make 90% of the total traffic on the weak coverage point covered by the planned base station,that is: the minimum total business volume needs to be 6350607. According to the weighted summation of the traffic volume of the weak coverage point covered by the macro base station and the traffic volume covered by the micro base station, the objective function with the largest covered traffic volume is converted into a constraint condition:

$$\sum_{i=1}^{n} q_i y_i + \sum_{j=1}^{n} p_j y_j \ge 6350607 \tag{13}$$

In summary, the optimal model for the type of base station selected for each site is:

$$\min Z = v_1 \sum g_i + v_2 \sum p_j$$
$$\begin{cases} d_{ui} \le 30, i = 1, 2 \cdots n \\ d_{uj} \le 10, j = 1, 2 \cdots n \\ \sqrt{(x_n - x_0)^2 + (y_n - y_0)^2} \ge 10 \\ \sum_{i=1}^{n} q_i y_i + \sum_{j=1}^{n} p_j y_j \ge 6350607 \\ g_i = 0, 1 \\ p_i = 0, 1 \end{cases} \tag{14}$$

## 2.2. The base station covers 3 sectors

In real life, each station is not completely covered by a circle, but rather has three sectors on each station, each pointing in a direction. Each sector has the largest coverage range in the main direction (30 for macro base stations and 10 for micro base stations), and can be covered within a range of about 60 degrees in the main direction. The coverage range gradually decreases linearly, and at 60 degrees, the coverage range is half of the coverage range in the main direction. If it exceeds 60 degrees, it cannot be covered by the sector. Considering that the angle between the main directions of any two sectors of each station cannot be less than 45 degrees, and other conditions are the same as for circular domains. Under the conditions of optimal station address and sector angle, the coverage of weak coverage points by new stations can reach 90% of the total traffic volume.

Since the coverage is 60 degrees left and right of the main direction, the base station is solved in the polar coordinate system (as shown in Figure 2), and the coverage decreases linearly with the angle, so the Archimedes spiral is introduced to calculate the sector area.

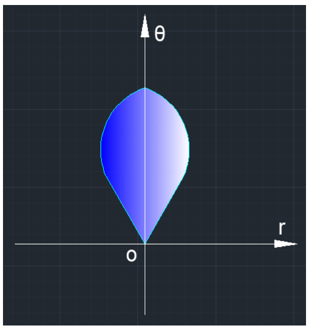


**Figure 2.** Sector diagram in polar coordinates

**Sector area model based on Archimedes spiral.** The coverage is gradually reduced linearly, and the relationship between the midline segment (coverage) and the angle is obtained, and the polar coordinate equation is used to obtain the Archimedes spiral formula:

$$r = a - b\theta \tag{15}$$

Among them, $a, b > 0, 0 \le \theta \le 30°$, $a$ is the distance from the initial coordinate point to the polar coordinate origin, $b$ is the coefficient of linear change on the polar coordinate, and $\theta$ is the angle between the polar diameter and the polar axis.

With the change of the polar angle, at a certain angle $L$, the arc length corresponding to the angle is:

$$L = \frac{\pi r \theta}{180°} \tag{16}$$

Then the definite integral idea is used to obtain the formed half-sector area:

$$S_0 = \int_{\theta_1}^{\theta_2} \frac{Lr}{2} d\theta \tag{17}$$

Finally, the area of a sector is obtained:

$$S_m = 2S_0 \tag{18}$$

**Discriminant Model Based on Covering Weak Covering Points.** Divide the sector arc into infinite points, the weak coverage point and the two vertices on the sector will form a triangle. Assuming that the side lengths of this triangle are $a_i, b_i$, and $c_i$ respectively, using Heron's formula, the area of the triangle can be obtained as:

$$S_i = \sqrt{p_i (p_i - a_i)(p_i - b_i)(p_i - c_i)} \tag{19}$$

Where $p$ is the half perimeter, i.e. half the perimeter:

$$p_i = \frac{a_i + b_i + c_i}{2} \tag{20}$$

Then sum the obtained triangle areas to determine whether it is equal to the area of the sector of the base station. If it is equal to the area of the sector of the base station, the weak coverage point is within the sector covered by the base station, otherwise it is not. The discriminant model is as follows:

$$\sum S_i = S_m \tag{21}$$

In the above formula, $S_i$ is the area of the triangle formed by the weak coverage point to any two points; $S_m$ is the sector area of the coverage area of the base station.

**Gradient descent based coverage model.** For the same type of base station, the covered sector area is the same, and the state of the same sector angle of the base station is graphically visualized to draw the base station state based on the cellular model (as shown in Figure 3).

Obviously, the fan angle state of the base station cannot be in the above ideal state, but on the same site, looking for different angles to maximize the total

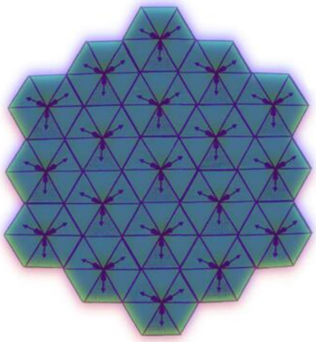

**Figure 3.** Same sector angle diagram of ideal base station

traffic coverage. Therefore, under the premise that the angle between the main directions of the two sectors cannot be less than 45 degrees, the direction of the sector angle is a changed value, and the sector angle of each base station is different. Therefore, gradients are introduced to find different fan angles to maximize the total coverage of services.

First, let the coverage index of sampling point $U_i$ be $f_k$, $m$ is the number of weak coverage points that reflect coverage, and $w$ is the traffic volume covered, which is defined as:

$$f_k \triangleq \min(m, w) \tag{22}$$

Let $P = \{P_1, P_2, \cdots, P_s\}$ be the set of all sampling points, where $s$ is the number of sampling points. Assuming that each grid area is the same in the service area, the coverage can be approximately estimated as:

$$C = \frac{|\hat{p}|}{|p|} \tag{23}$$

Among them, $\hat{p}$ is the set of sampling points whose signal quality meets the requirements.

According to formula (23),the probability of covering business volume can be deduced as:

$$C = \frac{|\hat{p}|}{s} \tag{24}$$

Then according to the guidance of the coverage index $f_k$, the coverage rate $c$ is obtained:

$$c = \frac{\hat{s}}{s} \tag{25}$$

According to the above formula (25), it can be deduced that:

$$C = \frac{\sum_{k=1}^{s} f_k}{s} \tag{26}$$

$f_k$ takes several values, and the coverage $c(\theta)$ is a function of discrete values, and it is not differentiable on $i$. When the number of weakly covered points is large, its gradient is:

$$\nabla C(\theta) = \left(\frac{\partial c(\theta)}{\partial \theta_1}, \frac{\partial c(\theta)}{\partial \theta_2}, \cdots, \frac{\partial c(\theta)}{\partial \theta_i}\right) \tag{27}$$

The objective function $c(i)$ can then be evaluated directly with small changes in $i$, expressed as[18]:

$$\frac{\partial c(\theta)}{\partial \theta_i} \approx \frac{\sum_{k=1}^{s} f_k(\theta_1, \cdots, \theta_i + \Delta\theta_i, \cdots, \theta_n)}{s \cdot \Delta\theta_i} - \frac{\sum_{k=1}^{s} f_k(\theta_1, \cdots, \theta_i, \cdots, \theta_n)}{s \cdot \theta_i} \tag{28}$$

Calculating $s$ coverage metrics for each numerical gradient is time-consuming and inefficient. A new type of derivable coverage index is proposed to approximate the original index. The main idea is to change from discrete coverage metrics to continuous coverage metrics. The former is a hard-coverage indicator, and the latter is a soft-coverage indicator. In order to continue using the gradient-based method, the soft coverage index is derivable or sub-derivable, and the hard coverage rate is the cumulative sum of the hard coverage indexes of all weak coverage points, which can be converted into the soft coverage rate $C(\theta)$, which is expressed as:

$$C(\theta) = \frac{1}{s}\sum_{k=1}^{s} \tilde{f}_k(\theta) \tag{29}$$

**Optimization model based on coverage area.** When the base station coverage is 3 sectors, the above optimization model is improved.

**Objective function**

As the sector angle changes, the angle affects the total cost of building a base station. The total cost needs to reach the minimum value $Z$ as the sector angle changes, so as to establish the objective function as follows:

$$\min Z(\theta) = \sum n\theta_i c_1 + \sum n\theta_j c_2 \tag{30}$$

With the change of the sector angle, the influence of the angle on the coverage index, the total amount of business $W$ in the coverage area reaches the maximum value with the change of the sector angle, so as to establish the objective function, as follows:

$$\max W(\theta) = \frac{1}{s}\sum_{k=1}^{s} f_k(\theta) \tag{31}$$

**Constraints**

1. The sector is left-right symmetrical and can cover a range of 60° left and right in the main direction, as follows:

$$0 \leq \theta \leq 30^\circ \tag{32}$$

2. For the problem that the included angle between the main directions of the sectors cannot be less than 45°. First, make an angle bisector for any two main directions, and then make a perpendicular line 0 from the original coverage of the main direction line (30 or 15) to intersect the angle bisector. Using the Euclidean distance, the length from the original coverage of the main direction line to the vertical foot can be obtained as $d_i$ ($d_1$,$d_2$,$d_3$ exist in the main directions 1, 2, and 3). Thus, the angle of the inverse trigonometric function is obtained, and the angle needs to be greater than or equal to 22.5°. The constraints are as follows:

$$\arctan \frac{d_i}{r_0} \geq 22.5° \tag{33}$$

$r_0$ is the original coverage length in the main direction of the sector.

3 In order to make 90% of the total traffic on the weak coverage point covered by the planned base station, that is, it needs to meet the minimum total business volume of 6350607. Thus, the objective function with the largest covered business volume is transformed into a constraint condition:

$$\frac{1}{s}\sum_{k=1}^{s} f_k(\theta) \geq 6350607 \tag{34}$$

In summary, the overall formula is obtained, as follows:

$$\min Z(\theta) = \sum n\theta_i c_1 + \sum n\theta_j c_2$$
$$\begin{cases} 0 \leq \theta \leq 30° \\ \arctan \frac{d_i}{r_0} \geq 22.5° \\ \frac{1}{s}\sum_{k=1}^{s} f_k(\theta) \geq 6350607 \end{cases} \tag{35}$$

## 3. Model improvements

In practical work, in order to better solve the problem of weak coverage, we now cluster the weak coverage points in a region. Close weak coverage points are grouped together to obtain weak coverage areas, which can be managed separately and better solve the problem of weak coverage. If the distance between two weak coverage points is not greater than 20, then these two weak coverage points should be clustered into one class. According to the transitivity of clustering properties, if point A and point B are of the same class, and point B and point C are also of the same class, then points A, B, and C are all of the same class. Now clustering all weak coverage points requires the total time complexity of the clustering method to be as low as possible.

### 3.1. Selection strategy of propagation model

Regional clustering analysis is mainly to cluster different characteristics with similar characteristics, but the propagation environment is the key factor that determines the propagation characteristics. The weak coverage points with different characteristics are classified into one category by cluster analysis, so that the weak coverage points in the same weak coverage area have high similarity.

The propagation model[20] is improved, and a propagation model selection strategy based on cluster analysis is proposed to select the most appropriate corrected propagation model for each type of propagation environment, so as to minimize the error caused by improper selection of propagation model. The process is shown in Figure 5.

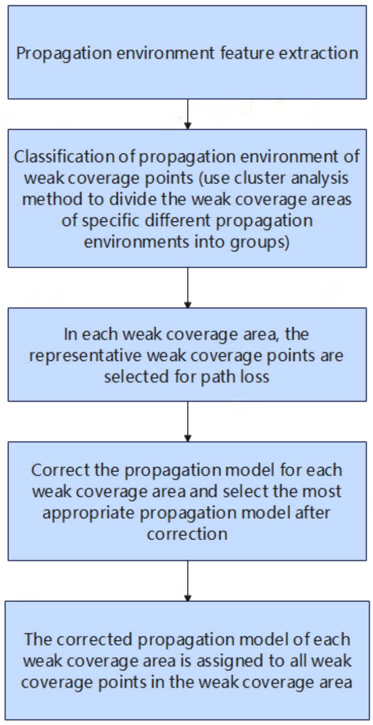


**Figure 4.** Flowchart of the propagation model

### 3.2. Propagation model based on cluster analysis

The features extracted[22] from the propagation environment are: business volume, construction cost, weak signal coverage, signal quality, communication density at weak coverage points, population density at weak coverage points, economic development at weak coverage points, and network attention at weak coverage points, received signal loss at weak coverage points. This is divided into 9 categories, regarded as the sample points of 9 attributes, expressed as $E = (e_1, e_2, \cdots, e_9), i = 1, \cdots, N$ where $N$ is the number of weak coverage points number. All weak coverage points form a sample set $B = \{b_1, b_2, \cdots, b_n\}$, and by performing cluster analysis on $b$, the weak coverage points with similar propagation environments can be divided into the same class to obtain clusters $H_1, H_2, \cdots, H_M$, where $M$ is the number of weak coverage areas.

As shown in Figure 5, in the classification of weak coverage point environment types, the dynamic clustering analysis algorithm $k-means$ algorithm is mainly used. The algorithm[18] steps are as follows:

**Step1:** Take the number of clusters as $M$;

**Step2:** $M$ sample points are arbitrarily selected in the sample set B as the initial centroids $Q_j(0)(j = 1, \cdots, M)$ of the $M$ weak coverage areas $H_j(j = 1, \cdots, M)$;

**Step3:** Loop through all sample points $b_1, b_2, \cdots, b_n$ each sample point finds the nearest centroid $Q_k(1 \leq k \leq M)$ to the sample point $b_n$, and set $B_i \in H_k$ until all sample points are used;

**Step4:** Iteratively find $H_j(j = 1, \cdots, M)$ centroids $Q_j(j = 1, \cdots, M)$;

**Step5:** Determine whether $Q_j(j = 1, \cdots, M)$ is the same as the centroid obtained in the previous iteration. If it is the same, stop the iteration, otherwise repeat **Step3** until the optimal solution is obtained by iteration.

On the basis of the inherent centroid stable convergence criterion of the $k-means$ algorithm, the clustering criterion function[22] can be used at the same time:

$$T_e = \sum_{i=1}^{M} \sum_{E_j \in H_i} \left| E_j - Q_i \right|^2 \quad (36)$$

The degree of convergence of the algorithm is examined by Equation (36), where $E \in H_i$, $Q_i$ is the centroid of $E_i$.

### 3.3. Propagation model correction based on path loss

For each weak coverage area, one or several weak coverage points closest to the centroid (Euclidean distance) are selected to optimize the propagation model.

Therefore, the optimization model of the propagation model[18] is established as follows:

When $d_A < D_f$,then

$$\eta = \beta_1 + \beta_2 \log d + \beta_3 \log f + \beta_4 * \eta_B + Q \quad (37)$$

When $d_A > D_f$,then

$$\eta = \beta_1 + \beta_2 * \log \frac{d}{D_f} + \beta_3 \log f + \beta_4 * \eta_B + Q \quad (38)$$

In the above formula, are feature correction factors,$Q = \sum_{i=1}^{9} Q_i e_i$.

For the clusters $H_1, H_2, \cdots, H_M$ obtained through the classification of the propagation environment, select the weak coverage points with the nearest Euclidean distance from the centroid $Q_1, Q_2, \cdots, Q_M$ to optimize the current network drive test and broadcast model. Thus, the optimized $M$ propagation models $[P_{M1}, P_{M2}, \cdots, P_{MM}]$ are obtained.

Finally, after $M$ weak coverage areas, the optimized propagation model $P_{Mi}$ of the ith $(i = 1, \cdots, M)$ weak coverage area is allocated to all weak coverage points in the weak coverage area, and the propagation model selection process ends.

## 4. Solution of the model

### 4.1. Model Results and Results Analysis

Using MATLAB to perform clustering, the center points of 3000 classes are obtained, and the contour values of the clusters are all greater than 0.6, so the clustering effect is good, and 563 micro base stations and 1436 macro base stations are obtained. At this time, the total number of services covered is 6356123, which is exactly about 90%, and some data are as follows (Detailed data can be found at the following website (https://github.com/Sunlight312/Data-support-materials.git).

**Table 1.** Partial results of base station site and type

| x | y | scale | x | y | scale |
|---|---|---|---|---|---|
| 1 | 769 | macro | 1600 | 1206 | micro |
| 300 | 1611 | macro | 1708 | 896 | micro |
| 600 | 1047 | macro | 1802 | 198 | micro |
| 1000 | 1858 | macro | 1901 | 663 | micro |
| 1203 | 137 | macro | 2100 | 353 | micro |
| 1501 | 514 | macro | 2400 | 1908 | micro |

In the Table 1, scale represents the base station scale, macro represents the macro base station and micro represents the micro base station.

### 4.2. Solution for base station coverage of 3 sectors

The two-dimensional space is divided into hexagonal areas, and then the base station is established according to the Lee cellular model[24]. At a certain point in time, the angle transition of the sector on a certain base station presents a graph (As shown in Figure 4).

Use MATLAB for clustering and get 4000 cluster centers. After optimization, 1,524 macro base stations, 625 micro base stations, and a total of 2,149 base stations are obtained. At this time, the total number of services covered is 6,350,745, which is almost the same as 90% of the weak coverage area (6,350,607). The data of the base stations are as follows(Detailed data can be found at the following website (https://github.com/Sunlight312/Data-support-materials.git):

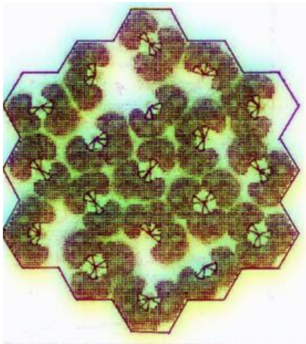

**Figure 5.** Distribution and coverage of base stations

**Table 2.** Base station distribution in type diagram

| x | y | scale | angle1 | angle2 | angle3 |
|---|---|---|---|---|---|
| 4 | 1743 | macro | 170.7 | 111.6 | 161.2 |
| 4 | 852 | macro | 113.9 | 125.8 | 93.9 |
| 30 | 898 | macro | 69.5 | 176.1 | 151.7 |
| 36 | 1049 | micro | 94.5 | 65.3 | 156.1 |
| 71 | 1796 | micro | 64.8 | 64.8 | 64.8 |

### 4.3. Results of improved regional clustering

This paper analyzes 182,802 weak coverage points, because these weak coverage points are distributed in various directions. After the feature extraction and classification of the propagation environment, the 182,802 weak coverage points are divided into 100 weak coverage areas. After 25 iterations, the centroid becomes stable.

At the same time, the number of weak coverage points contained in 100 weak coverage areas (100 categories) was obtained through MATLAB (part of the data is shown in Table 3).

**Table 3.** List of weak coverage points included in each weak coverage area

| Weak cover | number | Weak cover | number |
|---|---|---|---|
| $H_1$ | 1243 | $H_{42}$ | 3786 |
| $H_9$ | 1243 | $H_{54}$ | 2754 |
| $H_{18}$ | 2398 | $H_{73}$ | 5846 |
| $H_{25}$ | 3009 | $H_{90}$ | 2164 |
| $H_{37}$ | 1084 | $H_{100}$ | 1785 |

Then use MATLAB to get the point distribution diagram of Class 1 and Class 100, as shown in Figure 6 and Figure 7 (Detailed data can be found at the following website (https://github.com/Sunlight312/Data-support-materials.git)).

As can be seen from the above figure, when the clustering is 100 categories, the area is distributed

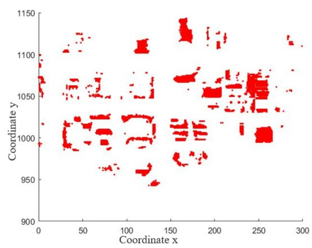


**Figure 6.** Distribution of the first type of points

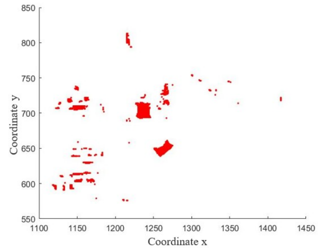


**Figure 7.** Distribution of points in the 100th category

in blocks, and there are almost no scattered points. Therefore, as long as continuous iterative correction is performed through the model algorithm in this paper, when the final regional clustering[25] is stable, it will show a blocky distribution, and the model clustering propagation model algorithm is effective for weak coverage problems.The algorithm has a better effect on clustering. If the algorithm is not used, the time complexity is $O(l * n * k * m)$,which is determined by the number of iterations $l$, the amount of data $n$, and the number of fields of $m$ elements, and $k$ is the type of clustering number decides. The time complexity of the dynamic clustering analysis algorithm is $O(l * \log l + n * \log n + m * \log m + k * \log k)$.The time complexity of the algorithm is much reduced, so the result is valid with the model.

## 5. Summarize

In this paper, the K-means algorithm is used to select the location of the network base station. Through the sector area model of the Archimedes spiral, the problem that the coverage decreases linearly with the angle can be better solved, and the angle selection problem of the three sectors can also be better solved. Then establish a discriminative model covering weak coverage points to determine whether the weak coverage points fall within the coverage range of the base station, so as to determine the amount of traffic in the coverage range. Secondly, the coverage rate of business volume is determined through a gradient descent coverage model, combined with an optimization model: the size and quantity of base stations can be determined with the objective function of minimizing costs and maximizing business volume coverage. In order to better solve the problem of weak coverage, regional clustering is performed on weak coverage points. Close weak coverage points are clustered into one group. By establishing a propagation model for clustering analysis and a propagation correction model based on path loss, the results of regional clustering are better determined, and the time complexity of the algorithm is essentially reduced, improving the efficiency of network location selection.

This article focuses on the situation where the coverage range of 5G base stations is significantly reduced. By establishing a cooperation between micro base stations and macro base stations, the coverage range of business volume is maximized and the cost is reduced as much as possible.

## References


[1] Bi Q, Zysman G L, Menkes H. Wireless mobile communications at the start of the 21st century[J]. IEEE communications magazine, 2001, 39(1): 110-116.

[2] Zhou S, Zhao M, Xu X, et al. Distributed wireless communication system: a new architecture for future public wireless access[J]. IEEE Communications Magazine, 2003, 41(3): 108-113.

[3] Chen S, Law R, Zhang M, et al. Mobile communications for tourism and hospitality: a review of historical evolution, present status, and future trends[J]. Electronics, 2021, 10(15): 1804.

[4] Qin Weijin Planning and optimization of mobile communication network [J] Journal of Anhui Mechanical and Electrical University, 2002, 17 (1): 62-65.

[5] Tutschku K. Demand-based radio network planning of cellular mobile communication systems[C]//Proceedings. IEEE INFOCOM'98, the Conference on Computer Communications. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Gateway to the 21st Century (Cat. No. 98. IEEE, 1998, 3: 1054-1061.

[6] C.-H.Wang,C.-J.Lee and X.Wu,"A Coverage-Based Location Approach and Performance Evaluation for the Deployment of 5G Base Stations," in IEEE Access, vol.8,pp.123320-123333,2020,doi:10.1109/ACCESS.2020.3006733.

[7] Menli. Research on site planning and business innovation of China Tower Corporation [D]. Xi'an University of Electronic Science and Technology, 2018

[8] El Ansary A M, Shalaby M F. Evolutionary optimization technique for site layout planning[J]. Sustainable Cities and Society, 2014, 11: 48-55.

[9] Habibi J, Mahboubi H, Aghdam A G. A gradient-based coverage optimization strategy for mobile sensor networks[J]. IEEE Transactions on Control of Network Systems, 2016, 4(3): 477- 488.

[10] Li L, Zhang B, Shen X, et al. A study on the weak barrier coverage problem in wireless sensor networks[J]. Computer Networks, 2011, 55(3): 711-721.

[11] Yang Weiqiang Research on protection distance between mobile base station and radio monitoring station [D]. Zhejiang University of Technology, 2012

[12] Li X, Wang X, Zheng N, et al. Enhanced location privacy protection of base station in wireless sensor networks[C]//2009 fifth international conference on mobile ad-hoc and sensor networks. IEEE, 2009: 457-464.

[13] Wang Siyu, Zhang Chunwang, Ma Baohua, Li Yuan, Wang Qian, Li Yongnan 5G macro base station planning method based on traffic density Research [C]//. 5G Online Innovation Seminar (2019) Proceedings, 2019:128-131.

[14] Liu H, Li W, Cai H, et al. Research on Location Selection Model of 5G Micro Base Station Based on Smart Street Lighting System[J]. Mathematics, 2022, 10(15): 2627.

[15] Wang Qian, Wang Cheng, Feng Zhenyuan, Ye Jinfeng. Overview of K-means clustering algorithm [J] Electronic designer Cheng, 2012,20 (7): 21-24

[16] Hu Wei. Improved hierarchical K-means clustering algorithm [J]. Computer Engineering and Application, 2013,49 (02): 157-159.

[17] Liu Yaxi Research on mobile communication network coverage calculation and optimization methods [D] Beijing University of Science and Technology Science, 2021. DOI: 10.26945/d.cnki.gbjku.2021.000123.

[18] Wei Zaixue, Yang Dacheng. A propagation model selection strategy based on cluster analysis [J]. Radio Engineering, 2005 (12): 20-22.

[19] Shui-Hua Wang;Deepak Ranjan Nayak;David S.Guttery;Xin Zhang;Yu-Dong Zhang.COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis[J].Information Fusion,2021,Vol.68: 131-148

[20] Shui-Hua Wang, Vishnu Varthanan Govindaraj, Juan Manuel Górriz, Xin Zhang, Yu-Dong Zhang.Covid-19 Classification by FGCNet with Deep Feature Fusion from Graph Convolutional Network and Convolutional Neural Network[J].An international journal on information fusion,2021,Vol.67: 208-229

[21] Shui-Hua Wang;Yu-Dong Zhang.DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification[J].ACM Transactions on Multimedia Computing, Communications, and Applications,2020,Vol.16: 1-19

[22] Li Hailin, Wan Xiaoji, Lin Chunpei. Research on thematic analysis based on keyword importance and neighborhood communication clustering [J]. Information Science Report, 2018,37 (05): 533-542.

[23] Zhang Xuefeng, Zhang Guizhen, Liu Peng. Improved K-means algorithm based on clustering criterion function [J]. Computer engineering and application Use, 2011,47 (11): 123-127.

[24] Roh W, Seol J Y, Park J, et al. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results[J]. IEEE communications magazine, 2014, 52(2): 106-113.

[25] Chen, J.; Tian, J.; Jiang, S.; Zhou, Y.; Li, H.; Xu, J. The Allocation of Base Stations with Region Clustering and Single-Objective Nonlinear Optimization. Mathematics 2022, 10, 2257. https://doi.org/10.3390/math10132257