Research on Base Station Site Planning Based on Cluster Analysis and Azimuth Analysis Model

Haobin Song [1], Jing Yang [2], Yue Liu [1], Zhihao Jiang [3], Shi Wang [3], Limi Chen [3], Xiaochun Zhao [1,*]

[1] School of Urban Construction, Hainan Vocational University of Science and Technology, Haikou, Hainan, 571137

[2] School of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, Hainan, 571126

[3] Public Courses Teaching Department, Hainan Vocational University of Science and Technology, Haikou, Hainan 571126

* Corresponding Author Email: 344674871@qq.com

Abstract. With the rapid development of the information age, the construction of 5G mobile network has improved the speed of network transmission and the utilization rate of spectrum. However, at the same time, the coverage of base stations is correspondingly less and less. In order to improve the quality of mobile communication, this paper uses the normal distribution 3-standard deviation method, Euclidean distance and 0-1 planning site selection model to analyze and divide the macro and micro base stations and make the optimal solution of site planning. Furthermore, the azimuth analysis model and K-means clustering algorithm were used to optimize the site planning and coverage area. Finally, 331 macro stations and 6459 micro base stations needed to be built, with an optimized Angle of 47 degrees and a coverage area of 93.3%. This paper provides some reference ideas for solving the problem of selecting and planning the base station site in the communication network.

Keywords: site selection, 0-1 programming model, K-means clustering algorithm, sector coverage planning.

1. Introduction

With the development of The Times, mobile communication technology has become an inevitable trend of development, along with the communication bandwidth is getting larger and larger, but the coverage of the base station is getting smaller and smaller, so that the coverage of the same area, the need of the base station also becomes more and more extensive. With the rapid development and progress of optical fiber technology, chip technology and so on, the types of base stations and antennas are also increasing to better support the corresponding information products, such as mobile social networking, mobile games and so on, and promote their popularity and application [1]. At the same time, along with the communication network planning and site selection problems become more and more complex, causing academic and industrial attention.

The site selection is generally based on the analysis of the coverage of the live network antenna, and a certain number of points are selected to solve the problem of the weak coverage area in case of weak coverage area. In this paper, the planning model of mobile site selection is built, the site selection is planned, the coordinates are given and the corresponding types of base stations are selected, so that the coverage points account for 90% of the total business volume, and the weak coverage points are studied by further introducing the sector area angle.

2. Model construction

Site selection planning is an important link of communication network. For the construction of communication base stations, the base stations are required to be selected first, and then the bottom points suitable for construction can be determined through terrain and environmental analysis. For

the selection of base stations, the 0-1 backpack planning problem can be used to slightly adjust the location of base stations in the same subarea, and determine the general range of each base station. Finally, the final location of the base station is determined by fine tuning.

Firstly, the problem is classified as zero-one integer backpack programming problem, and the constraint conditions such as threshold and distance, and the base stations should be connected to each other, considering the uneven terrain, so the signal may be blocked, so the establishment of two base stations must require that the distance between them is not blocked by the terrain and does not exceed the maximum distance of directional antenna.

To sum up, this paper, through grid division, discrete ratio, then build polygon model, detailed analysis of the optimal site, in the model of the specific application of the macro station and micro base station are discussed, and after the data noise, the corresponding results can be obtained.

2.1. Coordinate noise reduction and K-Means clustering model

2.1.1 Coordinate denoising

The image information is regarded as a spatial function F(z), and the image noise function N(z) is the factor that degrades the information expressed by the function F(z), that is, under the influence of noise, the image degrades to G(z). According to the relationship between image and signal, image noise can be divided into additive noise and multiplicative noise [2], whose expression is shown in Formula 1 and 2. Equation (1) represents the effect function of additive noise on image information, and equation (2) represents the effect function of multiplicative noise on image information.

$$G(z) = F(z) + N(z) \tag{1}$$

$$G(z) = F(z) * N(z) \tag{2}$$

Noise brings many difficulties to image segmentation, feature extraction, image recognition and texture analysis, and directly affects the image post-processing. Therefore, real-time acquisition of images need to be filtered to eliminate the noise in the image, in order to maintain the clarity of the image. This paper uses Euclidean distance for visual analysis (Eq. (3)).

$$d_{IG} = \sqrt{\sum_{i,k=1}^{m,n} \left(\frac{x_{ik} - x_{Gk}}{S_k}\right)^2} \tag{3}$$

Where, d_{IG} represents the weighted Euclidean distance between system I and the target system, x_{ik} represents the kth evaluation index (the k-dimensional variable in mathematical concept) value of the evaluated system (the standardized value), and x_{Gk} represents the kth evaluation index value of the target system (the standardized value). S_k is the standard deviation of the kth index. The inverse of the variance is actually the weight of the index. Therefore, this formula can be regarded as a weighted Euclidean distance to obtain the result after noise reduction.

2.1.2 K-means cluster analysis

The organization in the data set is divided into multiple groups, and the corresponding tree graph is formed at the same time. This clustering algorithm is collectively referred to as hierarchical clustering algorithm. In the traditional K-means clustering analysis method, each data classification should be adjusted continuously, and then each iteration will be reduced once, until the algorithm does not adjust the classification of data objects and the corresponding classification of data objects does not change, the final correct classification is obtained, the iterative result of clustering algorithm meets the standard, and the algorithm is terminated [3].

For the evolved K-means algorithm, the clustering center will be modified for every data classification adjustment, which represents the completion of the iterative calculation process. The time complexity of this process is consistent with that of the original k-means algorithm. There is no significant difference in the complexity of computing the function value of the clustering criterion and the time spent to compute the new clustering center. Therefore, the operation of the whole clustering algorithm is carried out continuously, and the corresponding criterion function value is continuously reduced until it reaches the corresponding minimum value. A new clustering center can

be obtained by the above clustering algorithm. After the analysis of N data, it can be found that there is no substantial change in the clustering center before and after the iteration, so it is determined that the clustering algorithm has converged rather than diverged. Then, the criterion function is used to judge whether it is complete, and the value of the criterion function is modified accordingly [4]. Finally, in the iterative process of clustering algorithm, each data object should be separated into the closest cluster to the data object. The site selection process is shown in Fi. 1.

In terms of the input of k-means algorithm, it is assumed that the number of data objects contained in S of the current data set is n, so it needs to be divided into K class clusters, the number of data classification is set as K, and the number of sampling times is set as J. The output is K cluster center values and data elements in the corresponding classes. The steps are as follows:

1. Euclidean distance was used to calculate the distance between the corresponding samples, and the initial clustering center point (S, K, J) was set.

2. Then, k objects in high density region are selected from C of density set as the initial clustering center.

3. Finally, the deviation values of the corresponding data objects in the set are calculated, and then the distance of all data objects is calculated.

After the calculation is completed, the deviation from the mean distance is solved (Eq. (4)). If so, then $x_i \in \omega_i$.

$$D\left(x_i, Z_k(I)\right) = min\left\{D\left(x_i, Z_j(I)\right), i = 1,2, \dots, n\right\} \tag{4}$$

The set C of density values of all data objects in the data set is obtained. For the selected K initial clustering centers, the traditional K-means clustering algorithm is adopted to perform cluster analysis on the data objects, as shown in the Eq. (5):

$$Z_k(I+1) = \frac{1}{n}\sum_{i=1}^{n} x_i^j, j = 1,2, \dots, k \tag{5}$$

Then make improvement judgment as Eq. (6) shows.

$$Z_j(I+1) \neq Z_j(I), j = 1,2, \dots, k \tag{6}$$

If I=I+1, return step 2, Otherwise, the algorithm ends.

2.2. Sector coverage planning modeling

This paper needs to conduct in-depth analysis of the sector area and change the coverage area from the original circle to three sector areas. Besides, it also needs to judge the range of Angle changes, and carry out statistical analysis on the coordinates of the newly added base stations, as well as the Euclideian distance between the weak coverage points of the nearest coordinate points.

Then, based on the network division of 2.1, the influence of different sector areas is analyzed, and the cluster analysis model is introduced. Finally, the most azimuth Angle of the base station is used for planning, and the total coverage is recalculated, and the results are obtained to observe whether it is greater than 90%. If not, the iterative calculation is continued.

Based on the rasterized coverage points and the comprehensive consideration of the target value and the center aggregation degree of the coverage targets, the optimal azimuth Angle planning method of the base station based on cluster analysis can be used again, and verified by practical application cases, so as to improve the planning efficiency and planning quality, and thus save the investment of sector resources. Based on the basic principle of ultrasonic phased array detection technology, the delay algorithm and imaging algorithm of sector scanning are studied. The research focuses on the calculation of delay time and imaging algorithm, and puts forward the calculation process of deflection focusing of the acoustic beam and the calculation of focusing delay in the process of sector scanning. Finally, based on the received fan scanning data, the fan scanning image is obtained [5]. In addition, during the debugging of the system program, the reliability and validity can be verified by combining the specific test blocks, and the influencing factors for the quality of sector scanning detection are proposed.

Sector scanning is S scanning or Angle scanning. In the selected area, the appropriate beam Angle range is selected, and the sound beam is moved within the scanning range at a certain depth of focus, so as to obtain echo data of the scan line at different angles and describe the ultrasonic detection method for the above scanning imaging (Fig. 2). Finally, in sector scanning, the range of Angle deflection can be changed and controlled, so that the direction of ultrasonic beam can be optimized perpendicular to the estimated defects, and good specimen morphology and defect detection images can be obtained.

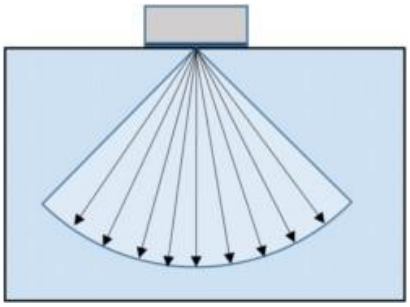


Fig 1. Schematic diagram of sector radiation

For the coverage point level planning, too large a target is not conducive to accurate analysis, so it is necessary to rasterize the target corresponding to the coverage point level planning and conduct the physical and chemical treatment of the grid. However, for the grid, the smaller the accuracy, the higher the corresponding accuracy. In this paper, aiming at the coverage point level planning goal of rasterization, site-based coverage capability is matched with the main service station at the same time, and site-level coverage capability is determined to comprehensively consider the scene type, the distance between grid and site and other factors. Therefore, the problem of base station planning is transformed into the problem of cluster analysis of the grid around the site. Finally, the planning of inclination Angle should be considered, so the planning and calculation can be carried out based on the optimal azimuth of cluster analysis by combining the antenna height and the farthest target coverage [6].

In order to complete the cluster analysis, this paper set the number of sectors as K, the iterative step of direction Angle as δ, and the antenna main lobe width as W. According to the planning experience, K=3, δ=10, and W=65 in general. Define x_i as the orientation Angle of the *i*th grid, and where i = {1... N}, N is the total number of grids served by the base station. Then initialize j = 0, for j, as an integer, and satisfy the 0 or less j acuities were $0 \le j \le C_{360/\delta}^{K}, C_{360/\delta}^{K}$, representing the number of combinations of K elements taken from $360/\delta$ different elements. Then assign the direction Angle, if $j = C_{360/\delta}^{K}$, it will terminate the operation, otherwise assign the base station azimuth Angle, $A_j = \{a_{j1}, a_{j2}, \dots, a_{jk} \dots, a_{jK}\}$. Where, a_{jk} is mutually different, and the value of a_{jk} is as Eq. (7) shows:

$$\begin{cases} a_{jk} \in \{0, \delta, \dots, 360 - K * \delta\} \\ a_{jk} = \{a_{k-1} + \delta, \dots, 360 - (K - k + 1) * \delta, 2 \le k \le K\} \end{cases} \tag{7}$$

The formula for calculating the Angle between sectors is as follows (Eq. (8)):

$$\begin{cases} m, n \in \{1,2, \dots, K\} \\ m \in n \end{cases} \tag{8}$$

Calculate the Angle θ_j^{mn} between two pairs in direction Angle A_j, $\theta_j^{mn} = |a_{jm} - a_{jn}|$ satisfies the above conditions.

Fan-shaped area Angle judgment corresponds to the direction Angle of the Angle between the two within A theta θ_j^{mn}whether meet the theta $\theta_j^{mn} \ge 90°$, if satisfied, can continue to perform, or to continue the iteration, and for the next direction Angle combination step $j + 1$.

Grid number pause when the azimuth grid number is within their main lobe$C_j = \sum_{K}^{k=1} C_{ij}$. C_{ij} is calculated as shown in the Eq. (9)

$$C_{ij} = \begin{cases} 1, \left| 0, \left| x^i - a_{jk} \right| \le \frac{W}{2} \right| \\ 0, \left| x^i - a_{jk} \right| > \frac{W}{2} \end{cases} \tag{9}$$

After the calculation is completed, the azimuth Angle A and the grid number C_j are recorded, and the iteration continues, and the steps are performed for the next direction Angle group.

The azimuth Angle with the largest number of covered grids is output to generate each mountain area of the base station. When the number of grids within the range of the main lobe is the largest, the azimuth Angle is set as A_j and C_j. If the value of A is unique, the corresponding value of A can be recorded and used as the output of the optimal azimuth Angle. If the value of A_j is not unique, the step is continued.

When azimuth Angle A is calculated, the corresponding sector equalization is the equalization of grid distribution in the main lobe range of each sector of the base station (Eq. (10)).

$$O_j = p * W_j + q * Y_j \tag{10}$$

p and q are the weights of crowding and evenness, $p + q = 1$, W_j and Y_j are the crowding and evenness at azimuth Angle A.

Generate the optimal balance of each sector of the base station. For example, if the value is the smallest and the azimuth is set to A_j, it is the optimal azimuth output. If the requirements are not met, an optimization process is required (Eq. (11)):

$$X = N \frac{1+\beta_0}{1+\beta} \tag{11}$$

In order to eliminate the influence of different index dimensions, all cluster indicators are standardized, thus forming the sample set $X = \{x_1, x_2, \ldots, x_n\}$. Then, K sample points are randomly selected as the initial clustering center, and the distance between the class centers of each sample is calculated respectively, and each sample is assigned to the nearest center class to obtain the clustering results. The sample mean of each class is calculated as its new class center, and the above steps are repeated until convergence. Among them, Euclidean distance $d(x_i, x_j)$ is also used to represent the similarity between samples (Eq. (12) and Eq. (13)). The smaller this index is, the greater the similarity of the same kind is, and the better the clustering effect is.

$$d(x_i, x_j) = \sum_{k=1}^{m} (x_{ki} - x_{kj})^2 = \sum_{k=1}^{m} \|x_{ki} - x_{kj}\|^2 \tag{12}$$

$$WSS = \sum_{i=1}^{k} \sum_{C(i)=t} \|x_i - \bar{x_l}\|^2 \tag{13}$$

Where, K is the number of categories, m is the index corresponding to the clustering index, $\bar{x_l} = (\overline{x_{1l}}, \overline{x_{2l}}, \ldots, \overline{x_{ml}})^T$, the mean or center of the l class, $C(i)$ represents the sample set of the l class.

By using cluster analysis method, we can obtain the degrees of three main directions, cluster all angles into three categories, and make the system clustering method as a separate category. Under the specified distance, select the one with the smallest distance and combine it into a new category, then calculate the distance between the new category and other categories, and then combine the two categories with the closest distance. This will reduce one category at a time, until all the samples are in the same category.

3. Results

3.1. Construction of macro station and micro base station

The coordinate noise reduction effect is shown in the Fig.1. Finally, it is calculated that 331 macro stations and 6459 micro base stations need to be built (Fig. 2 and 3), and the business volume of newly built base stations accounts for 92.751% of the total business. The visualization analysis of the European distance is amplified 30 times, as shown in the figure. For selecting an Acer or microbase

station, this is shown in Table 1. It can be seen that the site selection scheme basically meets the requirements, and the total amount required is 16,389 yuan.

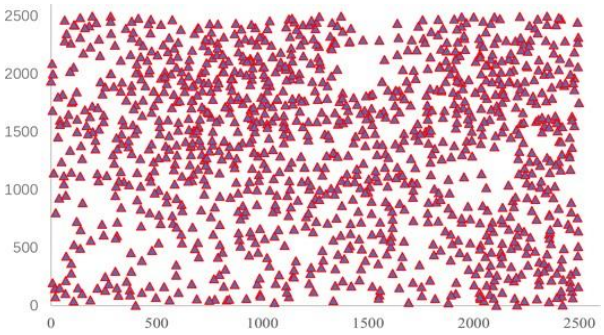


Fig 2. Effect picture after noise removal

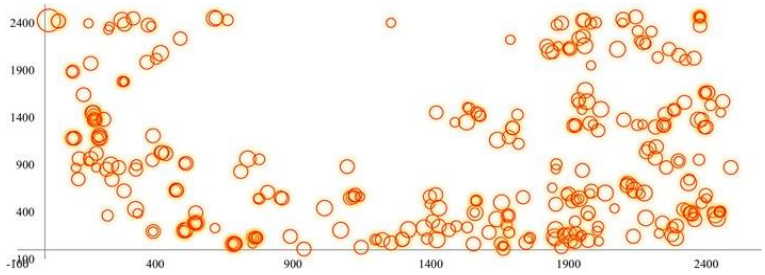


Fig 3. Schematic diagram of Acer station

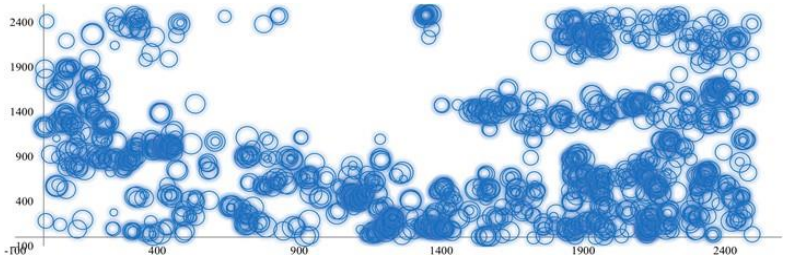


Fig 4. Schematic diagram of microbase station

Table 1. Comparison between macro and microbase stations

Basic information		Weight allocation				
Base Station Type	Scene	Coverage	Capacity	construction	performance	trouble
Acer Station	Wide coverage: new urban expansion	40%	10%	15%	10%	25%
	Wide coverage: towns and villages	35%	20%	10%	10%	25%
	Wide coverage: scenic area	45%	20%	0%	10%	25%
	Continuous coverage: urban areas	40%	10%	15%	10%	25%
	Continuous coverage: traffic trunk	45%	20%	5%	10%	20%
Micro base station	Campus	40%	40%	0%	10%	10%
	Shopping mall	40%	40%	0%	10%	10%
	Office building	40%	40%	0%	10%	10%
	Large gymnasium	40%	40%	0%	10%	10%
	Hospital	40%	40%	0%	10%	10%
	Residential area	40%	40%	0%	10%	10%

3.2. Site planning and coverage area optimization

Taking the Acer station as an example, the threshold and the Angle variation between the weak coverage point and the coverage limit are calculated (Fig. 4). It can be seen that the positioning is initiated from the terminal position to reach the positioning platform. In the positioning platform, authentication is carried out successively, positioning is initiated, sector data is updated, optimized sector data is used for positioning, and verification accuracy is improved. Then the positioning results are transmitted to the analysis system, and the optimized sector data is fed back to the former. Finally, the corresponding optimization can be completed by optimizing analysis and processing from the sector data.

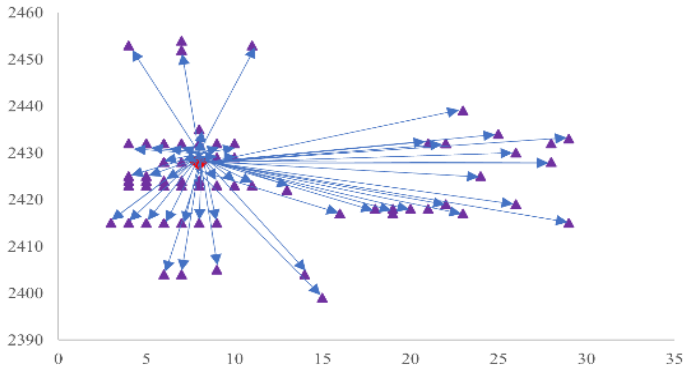


Fig 5. Visual analysis between weak coverage points

The optimized sector-level coverage planning results are shown in the Fig. 5. It can be seen that the optimized Angle is 47 degrees, and through the corresponding analysis of site coverage, the actual site can be covered by 216,900 grids, and the total number of grids involved is 231,700, accounting for 93.3% of the total grid, so it basically meets the requirements.

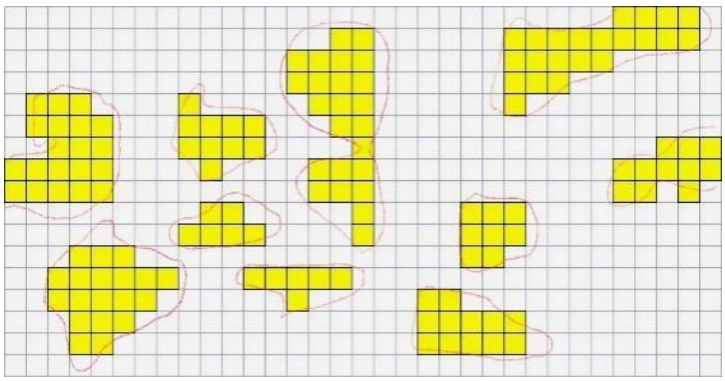


Fig 6. Schematic diagram of sector-level coverage planning after optimization

4. Conclusion

With the construction of 5G mobile networks, the speed of network transmission increases, and the coverage of base stations is correspondingly less and less. In order to improve the quality of mobile communication and make the optimal solution of site planning, this paper uses the normal distribution 3-standard deviation method, Euclidean distance and 0-1 planning site selection model to analyze the macro and micro base station area and divide the network, and further uses the azimuth analysis model and K-means clustering algorithm to optimize the site planning and coverage area.

References

[1] Wang Fanlin; Zhang Jide. Research on Information Strategy, IT Governance and Enterprise Performance [J]. Public Finance Research, 2018(12): 114-125.

[2] LI Yiliang; WANG Weiwei. Image Denoising Based on Wiener Filtering in Wavelet Domain and Anisotropic Diffusion [J]. Computer Engineering and Applications, 2006(29): 52-54.

[3] WANG Yong; TANG Jing; RAO Qinfei; YUAN Chaoyan. High efficient K-means algorithm for determining optimal number of clusters [J]. Journal of Computer Applications, 2014, 34(5): 1331-1335.

[4] LV Minglei; LIU Dongmei; ZENG Zhiyong. Novel Image Retrieval Method of Improved K-means Clustering Algorithm [J]. Computer Science, 2013, 33(S1): 195-198.

[5] TIAN Guoliang; HOU Jingang; WEI Daoxiang; LIU Enkai. Comparison of two imaging modes of sectorial scanning in ultrasonic phased array inspection [J]. Nondestructive Testing, 2019(12): 57-60.

[6] CAI Jiaqi; WAN Haibin; SUN Youming; QIN Tuanfa. Artificial bee colony algorithm based self-optimization of base station antenna azimuth and down-tilt angle [J]. Telecommunications Science, 2021(01): 69-75.