Clustering Optimized Genetic Algorithm-Based 5G Communication base Station Site Selection

Biru Ren [1, a], Wei Liu [2, b]

[1] Engineering Management. School of Management. Northwestern Polytechnical University. Xi'an, China

[2] Communication Engineering. School of Electrical and Electronic Engineering. Anhui Institute of Information Technology. Wu'hu, China

[a] rbr2021@mail.nwpu.edu.cn, [b] liuwei.aiit@hotmail.com

Abstract. To minimize the negative impact of the base station construction, and at the same time make the base station cover as many users as possible, the Genetic Algorithm optimized Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method is proposed. Initially, the number of points is fixed and encoded into a gene sequence with 4 different features at each point. After that, the process of the DBSCAN is stated and then it constructs the cluster adaptation function. Then, with the combination of the cluster adaptation function and the coverage function, it can select the individual, cross the gene sequence, and change the slight value of the sequence. Ultimately, by iteration, it can output the optimal point location and the result shows that the optimal coverage is 95%.

Keywords: 5G Base Station; Location Selection; Genetic Algorithm; Clustering Algorithm.

1. Introduction

With the development of industrial production as well as social life, the demand for high performance as well as high transmission speed communication networks has been increasing. The fifth generation mobile communication technology (5G), a new generation of mobile communication technology introduced in the 1920s, has some improvements over the previous generation of communication technology. In terms of transmission speed and delay efficiency, the fifth generation communication technology has improved more significantly compared to the previous generation [1]. However, due to the relatively limited time of its introduction, the application of the fifth-generation mobile communication technology is currently at a low level. However, from a large number of signal data samples and literature, it can be found that the coverage of 5G signals will be incomplete, i.e., it may change to a non-equal radius sector shape [2]. How to control the non-equal radius sector to cover as many users and devices as possible is a problem that needs to be considered in the construction of 5G communication base stations.

However, this problem cannot be solved by increasing the number of 5G communication base stations because the density of the station, the type of the station, and the construction cost of the station have an impact on the coverage. To tackle this issue, Johansson has proposed a method to estimate the cost of base stations by analyzing the degree of influence on the cost of base station density: when the density of communication sites is relatively high, the operation cost and transmission cost are relatively the main cost expenditure, however, this law is too abstract to apply in real engineering [3]. Yu has proposed a method to select the answer from the given sentence, which uses the encoding technique and can be transferred into the site selection problem. However, the selection model needs multiple sentence input, which is not suitable for the site sele ction because it should have only the location input [4]. Another method is to apply the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, however, it is hard to consider the coverage of the site station because the cluster zone should be a circle [5]. One solution is to apply the computer vision-based method, which can visualize the coverage by representing the black or white zone so that it can then apply the DBSCAN algorithm to perform the cluster, though it considers the coverage, it needs to map the point coordinate into a pixel-based graph, which may consume lots of computing power due to the float coordinate of the point [6].

Based on the analysis above, the current research on site selection or related problems has the issue of transforming the data and the constraint considerations. To tackle this problem, the genetic algorithm is applied to assist the DBSCAN to consider the constraint. The total modeling framework is demonstrated in Fig.1.

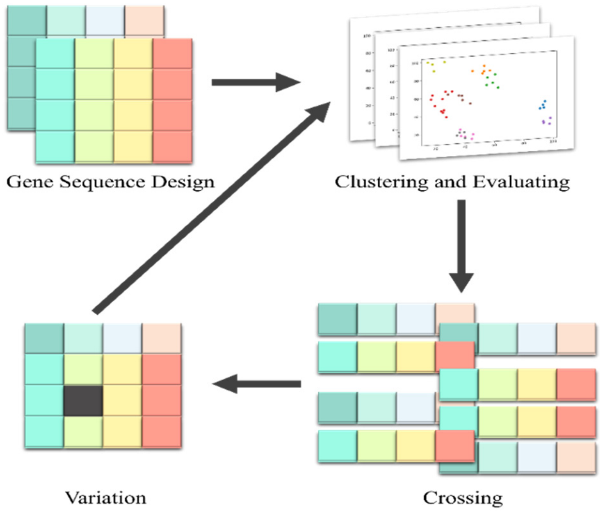

Fig 1. The Modeling Framework

2. DBSCAN Algorithm and Evaluation Method

The station selection problem can be summarized that using a certain number of points, and finding the optimized solution to obtain the largest region coverage. In this process, the cluster-based method remains a necessity because it can evaluate the distribution of the points. Compare with other clustering algorithms, such as K-Means, the DBSCAN is able to percept the distribution of the spatial density, which can evaluate the quality of the solution because the final distribution of the point should be uniform.

2.1 The Flow of the DBSCAN Algorithm

The DBSCAN algorithm is a typical representative algorithm of density algorithms, and the DBSCAN algorithm needs to be set in advance to set the global parameters domain distance and threshold before proceeding. The threshold is set by the variation of the minimum number of sample points within the range radius.

The first step in executing the DBSCAN algorithm is to input the minimum radius E_{ps} with the density threshold $P_{\min}$ and then read the coordinates of the object dataset.

Subsequently, the point in the object dataset is judged to be a core point and the distance between the point and other points is calculated, and the squared Euclidean distance is generally used to eliminate the absolute value, and the definition of the squared distance between the general points p_1 and p_2 can be defined as shown in Equ. 1.

$$d_{p1p2} = (x_1 - x_2)^2 + (y_1 - y_2)^2 \tag{1}$$

If the distance between two points A and B is less than or equal to the set E_{ps} value, then these mutually exclusive points are marked. If the distance is greater than the set E_{ps} value, then the point is ignored and the distance is calculated for the next point until all the points are traversed.

The clusters in the core points are compared and if there are identical elements, the two clusters with core points can be merged into a new cluster until no new cluster is created.

After the clustering is completed, it can be found that some points do not belong to the core points but are included in the region of another core point, and such points are defined as boundary points. In addition, the points that do not belong to the boundary points and core points are called noise points. Therefore, a point is most likely to be a core point when its local density is the largest, and the definition can be expressed as Equ. 2.

$$\xi_p = \max_{x \in D} \mathrm{Dist}(p,x) \tag{2}$$

Where Dist denotes the distance between given point and a certain cluster, which can be measured by the minimum Euclidean distance between the given point and points within a class.

2.2 DBSCAN-Based Evaluation Method

DBSCAN clustering algorithm can effectively reduce the influence of noise points on clustering results, and can find clusters of arbitrary shape and size in the data set. Based on this attribute, it can percept the potential non-uniform distribution problem in the dataset. Specifically, it can fix the radius E_{ps}, and change the number of the minimum involved points $P_{\min}$ to obtain the number of the cluster $\mathbf{N} = \left(N_{p1}, N_{p2}, \ldots, N_{pn}\right)$. After that, considering that the re-uniform dataset still has certain inhomogeneity in some local fine positions, it should apply weight average as the evaluation function F_1 between the number of the cluster and the maximum cluster area $S_{\max}$:

$$F_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{S_{\max,n}}{N_{pn}} \tag{3}$$

This method can be used because if the area of the largest cluster is smaller under a certain parameter, the distribution of the clusters is more uniform, but in turn this is a manifestation of uneven data, because according to the definition of density clustering, uniform data It can be directly grouped into a class under the same parameter.

3. The Genetic Algorithm of the Points

To generate the distribution of the point, it can apply the genetic algorithm because it is easy to program and iterate. Initially, as shown in Fig.1, it can divide the algorithm into 4 steps: the design of the gene sequence, the selection, the cross method, and the variation Method.

3.1 The Gene Sequence Design

Given the assumption of the research that the number of the total point M is fix and the sector constraint for each point ρ, θ is also fixed. Considering that there is no other factor to impact on the base station, it can just use the coordinate of the point and the parameters of the sector to represent each point:

$$\mathbf{p}_i = \left(x_i, y_i, \rho_i, \alpha_i\right) \tag{4}$$

Notice that the angle α_i denote the angle of rotation from the positive direction of the x-axis to the starting edge of the sector. Ultimately, it can obtain the gene sequence of the system with the length of M:

$$\mathbf{P} = \left(p_1, p_2, \ldots, p_M\right) \tag{5}$$

3.2 The Adaptation Function Design

The adaptation function for a system can be divided into two parts, one part is the DBSCAN-based function F_1, and another one is the coverage F_2. It needs 2 functions because when the number of clusters in each size of the DBSCAN would be one, which indicates that the function F_1 would be the same. In this case, it needs both these 2 functions.

Notice that the function F_2 can be calculated by the Inclusion and exclusion principle:

$$F_2 = \sum_{i=1}^{M} T_i - \sum_{i=1}^{M} \sum_{j=1, j \neq i}^{M} T_i \cap T_j - \ldots - \bigcap_{i=1}^{M} T_i \tag{6}$$

Where T_i denotes the area of each sector, which can be calculated by:

$$T_i = \alpha \frac{\pi \rho_i^2}{360} \tag{7}$$

Ultimately, it can calculate the complete adaptation function:

$$F = F_1 + F_2 \tag{8}$$

Individuals with greater fitness can be selected, because the larger the adaptation function is, the better the system would be.

3.3 The Cross Design

According to the definition of the gene sequence, it can know that the final result is independent of the position of the point in the gene sequence. Based on this, it can just cross the parts of the gene sequence like the common genetic algorithm.

3.4 The Variation Design

Based on the analysis above, it can know that the type of the feature of the point is all the float types. In addition, the effective radius of the sector cannot be changed for a certain point.

Therefore, the variation of the gene sequence should be continuous, which means, it can apply to add/subtract a certain random float to the coordinate or the angle for a certain point. Other variation scheme includes symmetry and mirroring global rotation, etc. can also be applied.

4. Experiment

In order to facilitate visualization, the area that the base station needs to cover is selected as a square, its side length is 100 km, the given number of base stations $M = 500$, the effect radius of each base station is $\rho_i = 3$ km, and the effected sector angle is $\theta_i = 120°$. The population number of genetic algorithm is set to 100, the selection probability is 30%, and the cross probability is 30%. The clustering radius E_{ps} in the DBSCAN algorithm is set to 10 km, and the number of included points $P_{\min}$ is 1, 2, 4, and 8, respectively.

4.1 The Result of The Coverage

Table 1. The Coverage Rate

Iteration/Method	Coverage Rate
0/Ours	41.18%
200/Ours	44.25%
400/Ours	45.73%
600/Ours	46.58%
800/Ours	47.12%
Computer Vision-Based Method	47.07%
Direct DBSCAN	42.69%
Theory Value	47.12%

By implement and iterate the genetic algorithm, it can obtain the coverage rate in Tab.1.

From the table, it can know that our method can achieve the theory value, which means, there is no station to overlap other stations' coverage. However, our method needs iteration while another method has no iteration scheme.

In addition, direct DBSCAN performs the worse because it only relies on the random state, rather than the iteration.

4.2 The Result of Station Location Selection

For better visualization, it can draw the location selection result over the iteration in Fig.2.

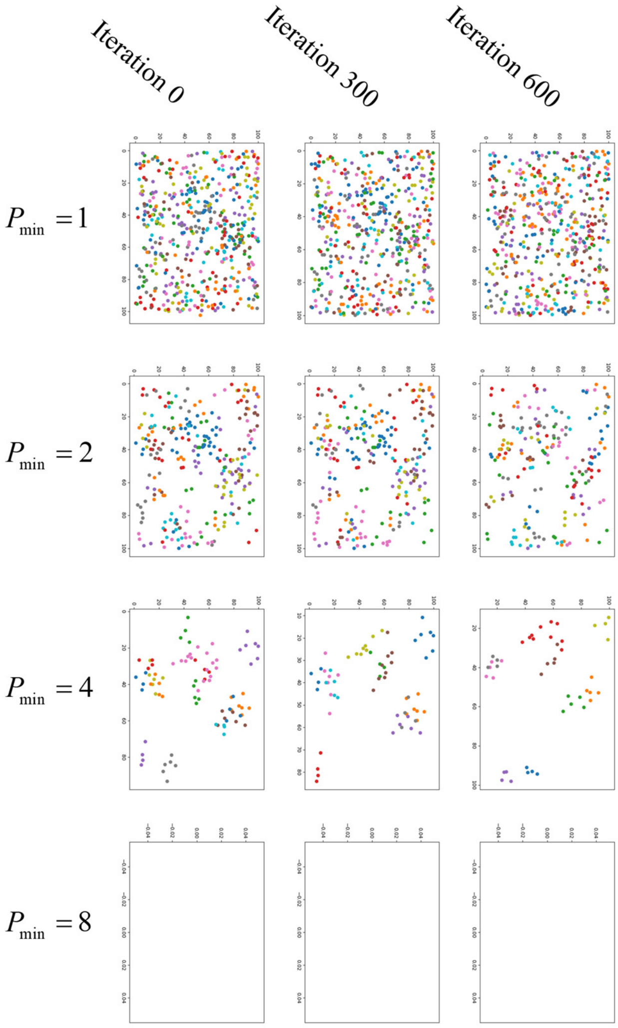

Fig 2. The Result of Location Selection

From the figure, it can be easily witnessed that there is no point in the right picture, which indicates that the DBSCAN recognize all the point in the noise with $P_{min}=8$. In this case, it can also discover that the number of points with $P_{min}=4$ increases with the iteration, which is positive because according to the adaptation function, the larger the area of the cluster is, the more adaptation would be seen. From this view, more points have a larger area, thus increases the adaptation.

In addition, it can be witnessed that the point is slightly in order in the left picture, which indicates that the square of the largest cluster is occupied more with $P_{min}=1$, which is still positive to the adaptation function.

Specifically, it can list the value of the DBSCAN adaptation function in Tab.2.

Table 2. The Adaptation Function of DBSCAN

Iteration	F_1 Value
0	0.101492
200	0.150892
400	0.153670
600	0.173813
800	0.176578

It can be seen from the table that as the number of iterations increases, the fitness of DBSCAN also increases, and after the last iteration of 800 steps, its fitness function is twice the initial fitness function. However, the fitness function still grows slowly, and there is a certain random growth characteristic, which is determined by the inherent characteristics of the genetic algorithm.

Combined with Tab. 1, it can also be found that although the fitness function increases slowly, it has reached the theoretical optimal solution in terms of coverage, and the subsequent adjustment of the fitness function is to adjust on the optimal solution, but It is not meaningful.

5. Conclusion

To tackle the location selection of the 5G base station, a DBSCAN-based optimized genetic algorithm is proposed with 4 steps, which can summarize the work and result as follows:

(1) This method constructs the adaptation function based on DBSCAN and the adaptation function based on the coverage area. The results show that as the iteration progresses, both function values show an upward trend. After 800 iterations, the optimized base station coverage has reached the theoretical optimum value of 47.12%, that is, there is no overlapping area between base stations, and based on Adaptation of DBSCAN. Compared with other algorithms, its effect is improved by about 5%.

(2) This method proposes the construction method, cross method and variation method of gene sequence, and uses genetic algorithm to optimize and iterate. With the deepening of the iteration, the points in the dataset show the characteristics of homogenization, which also proves the effectiveness of the algorithm.

References

[1] Sun J, Yu Q, Niyazbek M, et al. 5G network information technology and military information communication data services[J]. Microprocessors and Microsystems, 2020: 103459.

[2] Bartholomew C. China and 5G[J]. Issues in Science and Technology, 2020, 36(2): 50-57.

[3] Johansson K, Furuskar A, Karlsson P, et al. Relation between base station characteristics and cost structure in cellular systems[C]//2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No. 04TH8754). IEEE, 2004, 4: 2627-2631.

[4] Yu L, Hermann K M, Blunsom P, et al. Deep learning for answer sentence selection[J]. arXiv preprint arXiv:1412.1632, 2014.

[5] Zola P, Ragno C, Cortez P. A Google Trends spatial clustering approach for a worldwide Twitter user geolocation[J]. Information Processing & Management, 2020, 57(6): 102312.

[6] Falanga D, Kleber K, Scaramuzza D. Dynamic obstacle avoidance for quadrotors with event cameras[J]. Science Robotics, 2020, 5(40): eaaz9712.