

## Research on Location Selection Model of Base Station based on Improved Genetic Algorithm

Aifei Pei<sup>1, a</sup>, Shuai Qin<sup>2, b</sup>, Qi Chen<sup>3, c</sup>, Wenyuan Lu<sup>4, d</sup>, Pengcheng Liu<sup>4, e</sup>

<sup>1</sup> School of Electrical Engineering, Tiangong University, Tianjin, China

<sup>2</sup> School of Mathematical Sciences, Tiangong University, Tianjin, China

<sup>3</sup> School of Chemical Engineering and Technology, Tiangong University, Tianjin, China

<sup>4</sup> School of Textile Science and Engineering, Tiangong University, Tianjin, China

<sup>a</sup> 1910140108@tiangong.edu.cn, <sup>b</sup> 1910140108@tiangong.edu.cn, <sup>c</sup> haloyolo07@163.com

<sup>d</sup> lwy18587043635@163.com, <sup>e</sup> 18294243809@163.com

**Abstract.** With the increase in the number of users accessing the 5G network in the future, how to choose the location of 5G base stations to ensure effective network coverage of the service area, so as to provide reliable communication and transmission services, is a key issue to be considered. By establishing an improved genetic algorithm model, the site planning problem is solved. According to the circular coverage area of the base station, the coverage requirement and the minimum cost optimization goal are completed. First, denoising according to business volume, and finally get 35,915 weak coverage points. The sum of the cost of setting up macro base stations and micro base stations is the optimization goal, and the coverage rate of the weak coverage point is more than 90%, is set as the constraint condition. At the same time, the lethal factor is set in the selection process, and the individuals who do not meet the threshold conditions of the original base station are deleted. The improved operator is adopted in the crossover link and mutation link, which speeds up the running speed of the algorithm and ensures the population diversity and the accuracy of the results. Finally, 689 macro base stations and 269 micro base stations are established, with a coverage rate of 93.51% and an optimal cost solution of 7159. By visualizing the results on the coordinate map, it can be clearly seen that most of the weak coverage points have been covered, which shows that the improved genetic algorithm shows fast convergence speed and good optimal value, and verifies the effectiveness of the model.

**Keywords:** Improved Genetic Algorithm; Lethal Factor; Improved Crossover Variation Operator; Location of Base Station; Optimization Model.

### 1. Introduction

In recent years, the information and communication technology of China has been developing very rapidly, among which information and communication infrastructure is the key content to strengthen national strength, safeguard national security and enrich people's lives. With the rapid development of communication technology, 5G technology is gaining momentum from 2G era to 4G era [1]. However, the number of locations suitable for the base station layout is becoming fewer [2]. With the development and popularization of 5G technology, the construction of 5G base stations will be important work for future communication infrastructure. Improving the breadth and depth of mobile communication network coverage is the focus of building a new generation of communication network. Therefore, the location selection of the base station will become an important work of the base station construction.

The selection of base station location is a very important part of base station network planning. A reasonable base station location scheme can effectively improve the coverage of the communication network, and at the same time, it can effectively save the cost, promote the progress of the project construction, reduce the difficulty of the engineers, and facilitate the later maintenance [3]. However, the process of selecting the base station location is often difficult. In the early stage, the base station scheme is often determined according to experience [4], which is inaccurate and inefficient. Survey

on inaccurate candidate schemes often leads to the base station location which satisfies the survey requirements not meeting the predetermined requirements, so it may be necessary to re-designate the pre-selected scheme and re-survey, which leads to the low efficiency of base station location selection.

The problem of base station location selection began to be solved by mathematical modeling. Because of the complexity of the base station location problem, there are many conditions to consider, and the model is complex and it is difficult to get the optimal solution as a whole. Therefore, it is very important to design a practical base station location model and improve the algorithm.

## 2. Establishment of the Optimization Model of Base Station Location Selection

In terms of the base station sample dataset, this paper first cleaned and denoised. According to the actual situation of operators, the coverage rate is usually less than 100%. According to the business volume, the data are screened, and the weak coverage points are sorted according to the business volume. When the business volume is accumulated to 10% of the total business volume, the accumulated part is discarded. In this way, it not only ensures the coverage rate but also reduces the operating cost.

Next, it is necessary to plan the location selection of the base station. Meanwhile, it is necessary to ensure the coverage of the weak coverage area while achieving the minimum cost of building the base station. In addition, it also needs to meet the conditions of a given threshold, that is, it is impossible to build the base station within a certain range of the existing base station. Because of the overlap of such ranges, it is easy to cause a many-to-one phenomenon, that is, multiple base stations simultaneously complete the business volume of a service area, resulting in a waste of resources. The present study calculates the initial threshold constraint range through the coordinates of the existing base station and the threshold range. The objective function is the total cost of setting up the base station, and the constraint condition is that 90% of the total business volume of the weak coverage point is covered by the planned base station. Because the basic genetic algorithm solves this kind of combinatorial optimization problem with slow evolution speed and easy to fall into the local optimal solution, this paper improves the basic genetic algorithm by improving the crossover variation operator and setting the lethality rate, which can further increase the global search ability of the genetic algorithm and improve the accuracy of the algorithm. At the same time, because of the large amount of data, the convergence speed of the improved algorithm is also accelerated, which is suitable for processing a large amount of data.

### 2.1 The Establishment of the Model

#### (1) Constraints of coverage rate

The previous studies showed that the location planning of the communication base station is a multi-objective combination optimization problem. The main factors considered in the construction of the base station include cost target, quality target, and coverage target, among which the coverage target is the most important factor [5].

Coverage rate is the first consideration in base station planning, and it is the ratio of the total value of all covered user business volume in the area to all user business volume in the area. When calculating the number of users covered by each base station, in order to avoid repeated counting, only the number of user business volumes with the shortest distance is calculated within the coverage overlap range of a certain base station [6].

By generating a random matrix, the initial position of the base station, the number of macro base stations and micro base stations is randomly set. By accelerating the optimization process of the genetic algorithm, the distribution parameters of each base station are obtained when the fitness function obtains the optimal solution. The planning area of base stations is two-dimensional, and  $N$  base stations are planned in this area, including  $i$  macro base stations and  $j$  micro base stations. According to the type of base station, the radius is divided into  $r_1$  and  $r_2$ , and the coverage model of each base station can be expressed as a circle with the base station coordinate as the center and  $r_1$  or

$r_2$  as the radius. The macro base station set is  $C_1=(c_{11}, c_{12}...c_{1i})$ , where  $c_{1i}$  is the coverage model of the  $i^{th}$  macro base station,  $(x_i, y_i)$  is the coordinate of the  $i^{th}$  macro base station. The micro base station set is  $C_2=(c_{21}, c_{22}...c_{2j})$ , where  $c_{2j}$  is the coverage model of the  $j^{th}$  micro base station, and  $(x_j, y_j)$  is the coordinate of the  $j^{th}$  macro base station. The event that point  $(x, y)$  is covered by the macro base station is defined as  $r_i$ , and the event that point  $(x, y)$  is covered by the micro base station is defined as  $r_j$ , then the probability that  $(x, y)$  weak coverage point is covered by the base station is  $p\{r_i \cup r_j\}$ , that is,  $p\{r_i \cup r_j\} = p(x, y, c_{1i} \cup c_{2j})$  and this probability is a binary function:

$$p(x_k, y_k, c_{1i} \cup c_{2j}) = \begin{cases} 1 & \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \leq r_1 \\ 1 & \sqrt{(x_k - x_j)^2 + (y_k - y_j)^2} \leq r_2 \\ 0 & \end{cases} \quad (1)$$

Specifically,  $S_k$  is the business volume of the  $k^{th}$  weak coverage point,  $S$  is the sum of business volumes, and  $M$  is the sum of weak coverage points. In this paper, the coverage rate  $p_{cov} \geq 0.9$  is taken as the constraint condition.

**(2) Objective function of base station cost**

Among the operating costs of mobile communication operators, the operating costs of base stations account for about two-thirds of the total costs. Therefore, this is a very important optimization goal. The cost includes not only the rent and construction cost of the base station but also the later maintenance cost. The less the number of base stations, the lower the construction and maintenance cost of base stations, which is the ultimate goal of optimization. In this paper, the number of base stations is obtained by constantly adjusting the length of individual chromosomes. The number of base stations determines the length of chromosomes. After finding the shortest chromosome length that can meet the requirements of the above objective function, the minimum number of base stations is counted.

The cost of the objective function is the lowest. According to the above modeling process,  $i$  is the number of macro base stations and  $j$  is the number of micro base stations, and the objective function is:

$$f = 10 * i + j \quad (2)$$

Through the above analysis, the cost is the first consideration in the planning of base stations. The cost mainly depends on the type and number of base stations, which is the sum of the costs of macro base stations and micro base stations, while the total number of base stations depends on the individual chromosome length. According to the requirements, the coverage rate needs to reach 90%, so this paper takes the coverage rate as a constraint. Therefore, in the process of genetic algorithm optimization, the optimization objective function of the cost objective is mainly considered, and this function can be used as the fitness function in the process of genetic algorithm operation to evaluate the fitness of individuals, so as to select individuals and carry out genetic operations.

**2.2 Description of Improved Genetic Algorithm**

Through the above analysis, we can know that the objective function is  $f=10*i+j$ , then  $f(poh)$  represents the fitness value of the  $h^{th}$  individual in the population. The chromosome number of an individual is determined by the number of base stations, and the number of micro base stations and macro base stations in the chromosome is randomly generated, and the sum is  $N$ . The larger  $f(poh)$  is, the better the  $h^{th}$  individual is. When the number of iterations reaches the maximum, the individual with the largest fitness function value in the optimal population is the optimal solution. The genetic algorithm program for base station location is implemented in MATLAB. The parameters to be entered in the algorithm include the number of additional base stations, the coordinates of demand points, the coordinates and types of existing base stations, the initial population size, and the number

of iterations. After the program is executed, the output content is the base station coordinates of newly added base stations. The steps to improve the genetic algorithm are as follows:

Step 1: code. In the process of algorithm implementation, the first step is coding.

Assuming that the two-dimensional coordinates of the  $k^{th}$  base station in the newly added base station are  $x_k$  and  $y_k$ , the chromosome code of the  $h^{th}$  individual in the population is:

$$P_{poh} = \{g_{h1} \dots g_{hi} \dots g_{hj}\} \quad (3)$$

Each individual chromosome code is encoded by the two-dimensional coordinates of each base station. The length of the chromosome is determined by the number of newly added base stations. The length of the individual chromosome is twice that of the assumed base stations, ranging from 0 to 1.

Step 2: initialize the population. Each individual parameter of the population is a random number generated in the initial range so that a sufficient search range is kept in the early stage of the evolution iteration. The initial population in the improved genetic algorithm is established by utilizing coding mapping. Set the target area as:

$$\{(x, y) : 0 < x \leq L, 0 < y \leq W\} \quad (4)$$

$L$  and  $W$  are the length and width of the target area, respectively. The population consists of  $h$  individuals, and the chromosome of each individual in the algorithm consists of  $N$  genomes,  $N=i+j$ .

Initialize the population as:

$$\begin{aligned} P_{p01} &= \{g_{11} \dots g_{1i} \dots g_{1j}\} \\ P_{p02} &= \{g_{21} \dots g_{2i} \dots g_{2j}\} \\ P_{p03} &= \{g_{31} \dots g_{3i} \dots g_{3j}\} \\ &\dots \\ P_{poh} &= \{g_{h1} \dots g_{hi} \dots g_{hj}\} \end{aligned} \quad (5)$$

Step 3: replicate and select. The individual bit string replicates itself according to the objective function, and the first step of replication is to calculate the fitness. In this paper, the fitness of an individual is its objective function, as shown in the above formula. In simple genetic algorithms, the random selection method is usually adopted. In this paper, the roulette method is adopted, and its structure is as follows: the independent candidates are encircled into a circle, and the probability of each individual being selected is directly proportional to the fitness of the individual. The higher the fitness, the higher the selection probability.

When the individual contains a base station that is less than the distance threshold from the existing base station, we will select the base station with a probability of 0, that is, set the lethality rate.

- 1) Calculate the fitness function value  $f(p_{oh})$  of each individual;
- 2) Calculate the selection probability of each individual:

$$p\{p_{oh}\} = \begin{cases} \frac{f\{p_{oh}\}}{\sum_1^n f\{p_{oh}\}} \\ 0 \end{cases} \quad (6)$$

Less than the distance threshold, the individual will die.

- 3) Calculate the cumulative probability;
- 4) Randomly generate 0-1 random number to select.

Step 4: Cross. Randomly pairing the generated individual bit strings in pairs, randomly selecting the crossing points, and performing cross propagation on the paired bit strings to generate a pair of new chromosome sequences, that is, new base station sequences. The specific crossing mode is shown in Figure 1. The crossing point is randomly generated, and a crossing position is a random number

between 0 and chromosome length minus 1. Then using the crossover operation between chromosomes, the population is replaced and updated.

In this paper, the improved crossover operator is used to improve the probability of a genetic algorithm. The crossover probability is shown in this formula:

$$P_{cross} = \frac{1}{1 + e^{k(f_{avg} - f_{max})}} \quad (7)$$

K is a constant coefficient,  $f_{avg}$  is the average fitness value, and  $f_{max}$  is the maximum fitness.

Finally, calculate the fitness value of each population, and repeat steps 2, 3, 4, and 5 until the maximum number of iterations is reached. Check the fitness value of each individual in the final iterative population, and the highest fitness value is the optimal solution for adding base stations in the site selection process. That is, increasing the number and location of base stations can make the cost of macro base stations and micro base stations in the region optimal, and meet the requirements of coverage. Finally, the optimal chromosome vector obtained is the coordinates of the increased base stations, and half of the chromosome length is the number of newly increased base stations.

The flowchart of the whole algorithm is as follows:

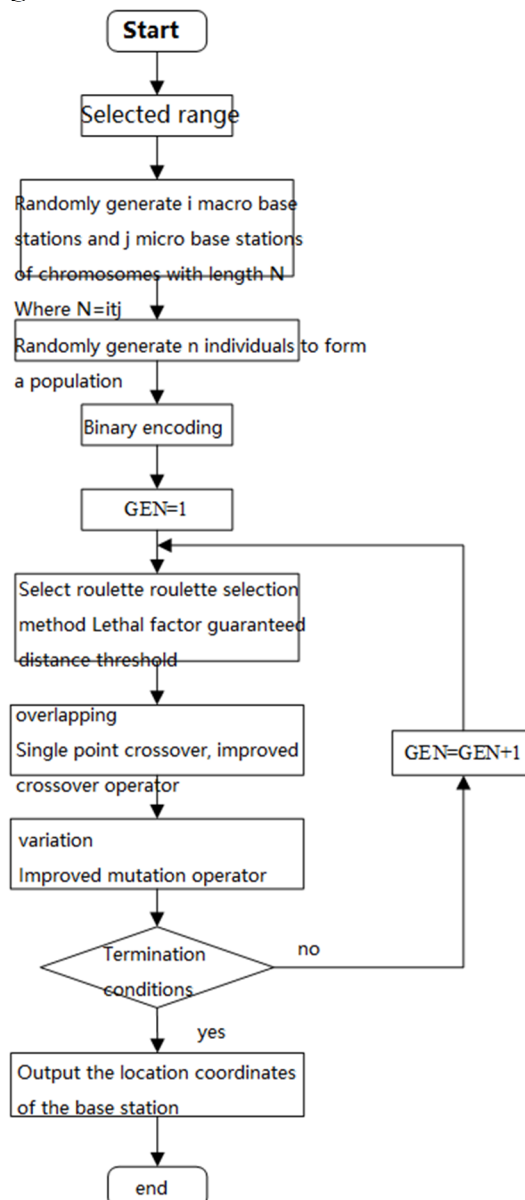


Figure 1. Flowchart of improved genetic algorithm

### 3. Solution and Result Analysis of the Optimization Model of the Base Station Location

The genetic algorithm can be used to get a better solution in a short time. The set U1 of the initial threshold constraint range is calculated by the coordinates and threshold range of existing base stations, and the set U1 of all ranges when dealing with this problem is subtracted from the set U of 2500\*2500 coordinate points. The establishment of base stations is regarded as a 0-1 planning problem for each coordinate point in a coordinate system. The objective function is the total cost of setting up base stations, and the constraint condition is that 90% of the total business volume of weak coverage points is covered by the planned base stations. The mathematical model with constraints is as follows:

(1) Decision variables:

$x_{ij}$ : the decision variable of the  $j^{th}$  point base station in class  $i$ , if the  $j^{th}$  point base station in class  $i$  is put into the chromosome segment,  $x_{ij}=1$ , otherwise  $x_{ij}=0$ .

(2) Objective function:

$$\min = f(x) = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} m_{ij} w_{ij} \quad (8)$$

Among them,  $m_{ij}$ : the cost of the  $j^{th}$  base station in class  $i$ , and  $w_{ij}$ : the category of the  $j^{th}$  base station in class  $i$ .

(3) Constraints:

$$s.t. \begin{cases} g_i(x) = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij} \cdot c_{ij} \geq 0.9S \\ g_i(x) = \sum_{j=1}^{n_i} x_{ij} = 1 \\ x_j = 0 \text{ or } 1 \end{cases} \quad (9)$$

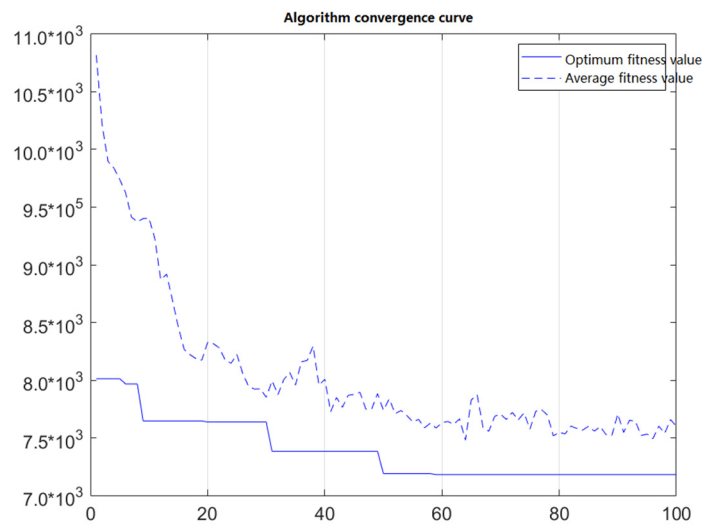
Among them,  $c_{ij}$ : the total amount of services covered by the  $j^{th}$  base station in class  $i$ .

After the operation of the dataset, further analysis shows that the set interval of this question is 2500\*2500, and if this interval is encoded into chromosomes, it will be very complicated and tedious. Therefore, in this paper, the coordinate system is divided into 500 100\*125 cells, and each interval is optimized. The objective function and constraint conditions are established on the improved genetic algorithm, and each interval is iterated 100 times. Finding the optimal solution for each interval in the iterative process, that is, setting the position and setting type of the base station. When summarizing the optimal solution set, we need to process the boundary data. In this paper, the processing method is to keep the base station with the largest business volume. Finally, summarize the optimal solutions of all intervals to get the optimal solution after 100 iterations as follows:

**Table 1.** Model optimization results

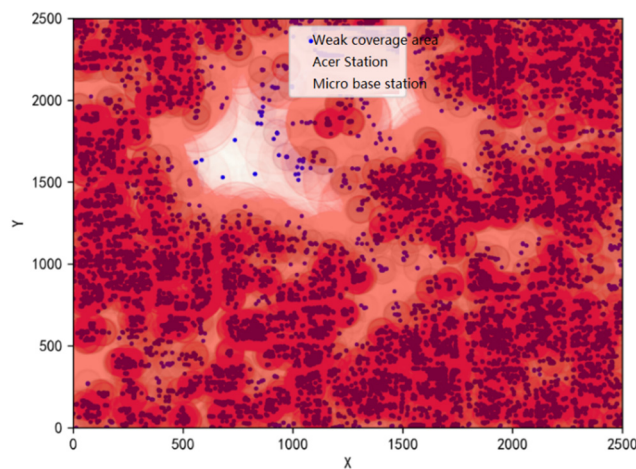
project	numerical value
Cost optimal solution	7159
Coverage business volume	5937875
Coverage rate	93.51%
Number of the macro base station	689
Number of the micro base stations	269
Fastest convergence times	57

After running the improved genetic algorithm program, the convergence curve and fitness of the algorithm are as follows:



**Figure 2.** Convergence curve

It can be seen that the convergence rate of fitness value is faster and smaller with the increase of iteration times, which accords with the objective optimization condition with the lowest cost. Visualizing the obtained base station coordinates, Figure 4 shows the coverage effect of 689 macro base stations and 269 micro base stations:



**Figure 3.** Coverage of all base stations

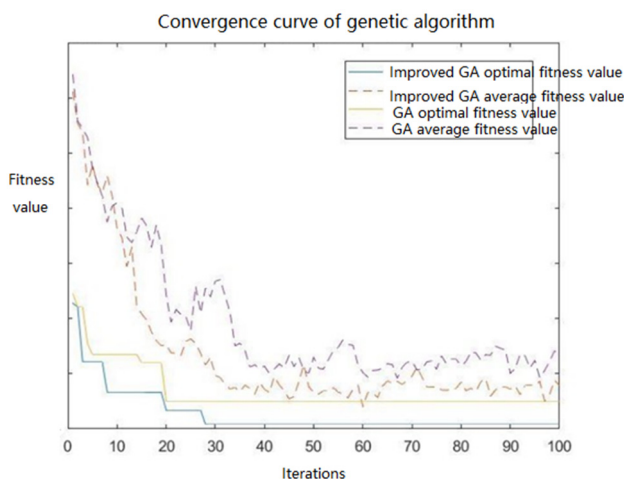
The dark blue dot in the figure is the weak coverage point after denoising, the dark red area is the coverage area of the newly-established micro base stations, and the light area is the coverage area of the newly-established macro base stations. It can be seen intuitively from the figure that the location selection effect of the base stations is relatively good, and most of the weak coverage points are within the coverage area of the newly-established base stations. The optimal solution of the cost calculated is 7159, the business volume coverage is 5937835, and the coverage rate is 93.51%, reaching the business volume coverage requirements of this question.

#### 4. Performance Comparison between Improved Genetic Algorithm and Traditional Genetic Algorithm

After the above analysis, the improved genetic algorithm in this paper adds a lethal factor in the selection process and excludes the base stations less than the threshold distance from the original base station, so as to ensure the rational utilization of resources. The crossover operator is improved in the

crossover link so that the gene part with high fitness can be better preserved and the survival of the fittest can be achieved. At the same time, the mutation operator is improved in the mutation link. In the early stage of the genetic algorithm, the mutation probability is reduced and the algorithm speed is accelerated. In the late stage of iteration, the mutation probability is increased in order to keep the population diversity. Our improved genetic algorithm is still fast and accurate in dealing with big data problems.

Compare the results of traditional genetic algorithm with those of improved genetic algorithm:



**Figure 4.** Comparison of convergence curves

By comparison, the improved genetic algorithm has converged when it has evolved to the 20<sup>th</sup> generation, which is earlier than the 30<sup>th</sup> generation of the basic genetic algorithm. Therefore, the improved genetic algorithm is more efficient, and its optimal fitness is lower than that of the traditional genetic algorithm, which is more in line with our goal of building a station at low costs. It can be seen that the improved genetic algorithm is more accurate and reasonable. Through the above analysis, it is obvious that the improved genetic algorithm is superior to the traditional genetic algorithm in all aspects. Using the genetic algorithm alone will have the problems of long-running time and high dependence on the initial population.

## 5. Summary

In this paper, an improved genetic algorithm model is established to solve the station location planning problem. According to the circular coverage area of the base stations, the coverage requirement and the minimum cost optimization goal are completed. First, denoise the data according to the business volume, and finally get 35,915 weak coverage points. The sum of the cost of setting up macro base stations and micro base stations is the optimization goal, and the coverage rate of the weak coverage point is more than 90% as the constraint condition. At the same time, the lethal factor is set in the selected link, and the individuals which do not meet the threshold conditions of the original base station are deleted. The improved operators are adopted in the crossover link and mutation link, which speeds up the running speed of the algorithm and ensures the population diversity and the accuracy of the results. Comparing the traditional genetic algorithm with the improved genetic algorithm, it can be seen intuitively that the improved genetic algorithm has a fast convergence speed and good optimal value, which can better complete the site planning. Simply using the genetic algorithm will lead to the problems of long-running time and high dependence on the initial population. Finally, 689 macro base stations and 269 micro base stations have been established, with a coverage rate of 93.51% and a cost-optimal solution of 7159. By visualizing the results on the coordinate map, it is clear that most of the weak coverage points have been covered, which shows

that the improved genetic algorithm has a fast convergence speed and a good optimal value. And it is conducive to solving the global optimal solution, can avoid local optimal solution, and shows good robustness and a good grasp of the integrity of base station site selection.

## References

- [1] Economic Information Daily. The information technology industry plan will be released: the 5G commercial network will be officially deployed in 2020 [J]. China Computer & Communication, 2016(19), 6.
- [2] Ling Juan. Research on site selection optimization of TD-LTE network base station based on hybrid immune algorithm [D]. Hangzhou Dianzi University, 2015.
- [3] Wang Wentao. Modeling and algorithm research of wireless communication network base station location optimization problem [D]. Northeastern University, 2012.
- [4] Zhou Yuguang. Improved particle swarm optimization algorithm and its application in optimal site selection of base station [D]. Guangdong University of Technology, 2014.
- [5] Xie Qingxi. Research and implementation of base station location optimization based on intelligent optimization algorithm [D]. Jinan: Shandong Normal University, 2018.
- [6] Chen Shi. Research on the site selection of NB-IoT base station based on improved artificial immune algorithm [D]. Nanjing University of Posts and Telecommunications, 2020. doi: 10.27251/ d. cnki. gnjdc. 20000.000000000005.