Article

Base Station Planning Based on Region Division and Mean Shift Clustering

Jian Chen [1] , Yongkun Shi [2], Jiaquan Sun [2], Jiangkuan Li [3] and Jing Xu [1],*

1 School of Mechanical Engineering, Yangzhou University, Huayang West Road 196, Yangzhou 225127, China; jian.chen@yzu.edu.cn
2 College of Electrical, Energy and Power Engineering, Yangzhou University, Huayang West Road 196, Yangzhou 225127, China; yongkunshi@outlook.com (Y.S.); jiaquan.yzu@outlook.com (J.S.)
3 School of Information Engineering (School of Artificial Intelligence), Yangzhou University, Huayang West Road 196, Yangzhou 225127, China; jiangkuan.li@outlook.com
* Correspondence: jingxu@yzu.edu.cn

Abstract: The problem of insufficient signal coverage of 5G base stations can be solved by building new base stations in areas with weak signal coverage. However, due to construction costs and other factors, it is not possible to cover all areas. In general, areas with high traffic and weak coverage should be given priority. Although many scientists have carried out research, it is not possible to make the large-scale calculation accurately due to the lack of data support. It is necessary to search for the central point through continuous hypothesis testing, so there is a large systematic error. In addition, it is difficult to give a unique solution. In this paper, the weak signal coverage points were divided into three categories according to the number of users and traffic demand. With the lowest cost as the target, and constraints such as the distance requirement of base station construction, the proportion of the total signal coverage business, and so on, a single objective nonlinear programming model was established to solve the base station layout problem. Through traversal search, the optimal threshold of the traffic and the number of base stations was obtained, and then, a kernel function was added to the mean shift clustering algorithm. The center point of the new macro station was determined in the dense area, the location of the micro base station was determined from the scattered and abnormal areas, and finally the unique optimal planning scheme was obtained. Based on the assumptions made in this paper, the minimum total cost is 3752 when the number of macro and micro base stations were determined to be 31 and 3442 respectively, and the signal coverage rate can reach 91.43%. Compared with the existing methods, such as K-means clustering, K-medoids clustering, and simulated annealing algorithms, etc., the method proposed in this paper can achieve good economic benefits; when the traffic threshold and the number of base stations threshold are determined, the unique solution can be obtained.

Keywords: base station planning; mean shift clustering; region division; single objective nonlinear programming; unique solution

MSC: 90C90

1. Introduction

In recent years, with the development of the internet, more and more devices have been connected to the mobile network, and new services and applications have been launched [1]. The number of mobile broadband users in the world is increasing rapidly, and the surge of data traffic poses serious challenges to the network. Firstly, from the current development of the mobile communication network, the capacity is insufficient to support thousands of times the traffic growth, and the network energy consumption and cost are unbearable [2]. Secondly, in order to improve the network capacity, it is necessary to make intelligent use of network resources, such as intelligent optimization of business

and user personality. Finally, in the future, the network will inevitably become complex and complicated with a large amount of data. To increase the capacity of the network, it is necessary to manage each network efficiently and improve the user experience.

To solve these challenges and meet the growing demand for mobile communication, it is necessary to develop a new generation of 5G communication network. With the new round of technological revolution, 5G network communication technology has developed rapidly. As a result, 5G bandwidth is increasing, but the range that base stations can cover is decreasing. At the same time, the types of base stations and antennas are also gradually increasing, which makes the planning problem of 5G network communication become gradually more complex, especially the base station location problem. The location problem is: according to the existing network signal coverage, determine the current network of weak coverage area, choose a certain number of points, by building new base stations on these points, to solve the signal coverage problem of the current weak coverage areas. Therefore, an effective method for solving the base station location problem is urgently needed. In the case of clear objectives and constraints, it is relatively easy to build a mathematical model, and the selection and design of the solution algorithms have a great impact on the solution results. The location of the base station must be such that the signal covers as many users as possible. A feasible scheme is to use cluster analysis to find the center of the users and then build the base station.

Commonly used clustering methods include K-means clustering, mean shift, KNN, DBSCAN, hierarchical clustering, and so on [3]. K-means clustering [4] is a clustering algorithm based on sample set partitioning. The principle of implementation is simple and the clustering effect is better, but the number of clustering centers is uncertain. KNN algorithm [5] is a classification algorithm in supervised learning. It is simple and effective, and can automatically classify large samples, but it also needs to determine the number of center points. Hierarchical clustering does not need to specify the number of clusters in advance, and can discover the hierarchical relationship of clusters, but the computational complexity is high, and the singular value has a great influence on the results. Mean-shift and DBSCAN [6] are both density-based clustering methods which also do not need to specify the number of clusters in advance. However, DBSCAN performs poorly on large sample data, whereas mean-shift is a sliding window-based algorithm that also performs well on large samples. Considering that the base station site selection needs to calculate a large number of samples, the selection of the user volume center is also more inclined to density division, so the base station site selection based on mean shift has high credibility.

This paper summarizes the experiences of previous studies and divides the signal coverage area into three categories. Combined with the single-objective nonlinear programming model established in this paper, a new and improved algorithm based on mean shift clustering is designed. Using the given data, a complete and detailed base station layout scheme is given. The main contributions of this paper are as follows:

(1) According to the strength of signal coverage, the whole study area was divided into three categories, and the research focus was placed on the study of redundant area, which improves the accuracy of calculation.

(2) Based on the differentiation of three types of signal coverage types, a single-objective nonlinear programming model was established to solve the optimal base station layout scheme with the lowest cost as the objective and the distance requirements between base stations, etc., as the constraints.

(3) A newly proposed location algorithm based on mean shift clustering was applied to base station deployment for the first time, the solution of large-area base station layout was realized, and a unique solution was obtained.

The rest of this paper is organized as follows: Section 2 is the summary of the relevant literature in the field of mean shift clustering and the base station optimization models. Section 3 introduces the preprocessing of the given data. Section 4 presents the region partitioning, the establishment of the single objective nonlinear programming model, and the design and solution of the algorithm in detail. Section 5 shows the results obtained by

the proposed method and discusses the results. Section 6 concludes the paper and provides an outlook on future feasibility studies.

2. Related Works

This part presents the relevant literature from two perspectives. On the one hand, the mean shift clustering algorithm and its current application are introduced. On the other hand, the relevant mathematical models and the state-of-the-art in base station optimization layout are presented.

2.1. Literature on Mean Shift Clustering Methods

For the application of mean shift clustering, in 2010, Fu et al. [7] proposed an anti-occlusion mean shift tracking algorithm combining Kalman filtering and least squares support vector machine (LSSVM) target trajectory prediction, aiming at the problem that the classical mean shift algorithm easily causes tracking failure when the target is occluded. In 2012, Ting et al. [8] proposed an improved mean shift algorithm for isolator image segmentation. Based on the habitat suitability index (HSl) model and the mean shift algorithm, the age segmentation of isolators was performed, the theory of the mean shift algorithm was introduced, the edge detection algorithm was used to explain the morphological processing, and the isolator image contour was obtained.

In 2018, the "blurred" mean shift algorithm was studied by Chen et al. [9], in which the convergence and the consistency of the blurred mean shift were investigated. In 2019, based on the mean shift algorithm, a dynamic gesture capture and tracking method for human-computer interaction system was proposed by Gong et al. [10]. In this method, the rectangular box representing the hand position was used as the control signal input and the initial detection target, which effectively processes the dynamic gesture video. In 2019, Demirovic [11] proposed the implementation and analysis of mean shift algorithm, which is a general non-parametric mode finding/clustering procedure widely used in image processing and analysis and computer vision technology, such as image denoising, image segmentation, motion tracking, etc. A mean shift outlier detector was proposed by Yang et al. [12], which uses a mean-shift technique to modify the data and remove the bias caused by the outliers. It also detects outliers based on the shifted distance. The experiments show that the proposed method works well regardless of the number of outliers in the data. The mean-shift-based target algorithm was summarized by Yao [13], wherein the development and improvement of the mean shift target tracking algorithm in anti-jamming (occlusion, illumination), fast moving target, scale and direction estimation were reviewed in detail. A new clustering algorithm, α-mean shift++, was proposed by Park [14], which improves mean shift++ in terms of runtime and image segmentation quality. The performance of α-mean shift++ was validated on image segmentation benchmark datasets, which achieves better image segmentation quality. Based on the neighborhood granular operation, Chen et al. [15] proposed the granular vectors relative distance and granular vectors absolute distance, and the neighborhood granular mean shift clustering algorithm. The effectiveness of neighborhood granular mean shift clustering was proved by both internal and external metrics. It is found that the granular mean shift clustering algorithm has a better clustering effect than the traditional clustering algorithms, such as K-means, Gaussian mixture, and so on. By applying the mean shift algorithm to the medical field, Cui et al. [16] proposed an enhanced image feature extraction and data mining method to extract simplified rules of the medical images, which is more beneficial to people's understanding than the raw data, and helps doctors to quickly understand patients' conditions. In 2022, Cariou et al. [17] proposed a new data mean shift method, which iteratively updates the equation and mixes the equations of standard mean shift (MS) and fuzzy mean shift (BMS).

2.2. Literature on Base Station Optimization Models

An integer linear programming model has been proposed by Mathar et al. [18] to solve the base station layout planning problem. The simulated annealing algorithm was used as an approximate optimization technique, and the performance of the different approaches was compared through extensive numerical tests. Zimmermann et al. [19] established a multi-objective optimization model, taking into account signal interference, traffic demand, regional coverage, and other issues. Using the proposed evolutionary algorithm, the method can realize the processing of more than 700 candidate sites. However, the above two methods have higher data quality requirements, and it is difficult to give a good scheduling scheme for different data streams. Resende et al. [20] outlined some of the most important optimization problems in the planning of the second and third generation cellular networks. The main mathematical models corresponding to these problems as well as some solution methods used to solve these models, were briefly presented. However, due to the lack of simulation data, it is not possible to present the specific design and compare it with the existing design. For the design of base stations in fourth generation (4G) cellular networks, Mai et al. [21] proposed a multi-objective mathematical model in 2013. The objective of the model is to minimize construction costs while maximizing coverage and capability, and factors such as orthogonal frequency division multiplexing, co-channel interference, reference receive power, base station density, and cell edge rate were considered in detail in the planning. In order to reduce costs and increase efficiency, it is difficult for the model to achieve ideal results when dealing with the weak coverage point problem of the base station. Singh et al. [22] proposed an effective algorithm for solving the base station layout problem, which can determine the optimal location of the base station without an exhaustive search. In addition, this algorithm reduces the number of base station installations as much as possible to make the location of base station location feasible, and provides full area coverage to reduce overlap. Due to the high time complexity caused by the exhaustive method, the efficiency of this method is low when dealing with large amounts of data. Valavanis et al. [23] proposed a multi-objective genetic algorithm (NSGA-II) that satisfies the three criteria of coverage, capacity, and total network cost, and studied the optimal location of base stations that satisfy certain coverage and capacity constraints in the planning of cellular networks. The paper by Kang et al. [24], considered the 5G networks with different heterogeneous cells and different traffic types, respectively. An efficient energy saving scheme for base stations (BSs) was proposed, an optimization problem for the proposed energy saving scheme was formulated, and the solution was obtained using particle swarm optimization (PSO). Numerical results show that the proposed energy saving scheme has better energy efficiency and lower total delay compared to the conventional energy saving schemes in both basic and modified separated network architectures. Although the PSO method can iteratively find a better scheme, it can easily overfit and cannot obtain a unique scheme.

In 2018, a comprehensive optimization framework based on multi-objective evolutionary algorithms (MOEAs) was proposed by Goudos et al. [25]. The equilibrium considers the effective maximization of average user rate, average area rate, and energy efficiency, to solve such multi-objective problems in 5G networks and obtain the best solution. Although comprehensive optimization is considered, it is difficult to obtain the best economic benefits using this method. In 2020, a 5G base station layout method considering cost and signal coverage had been proposed by Wang et al. [26]. To solve the optimization problem of 5G base station positioning, the implementation process of 5G base station collaborative operation and an optimized location deployment scheme was carried out with the aim of reducing installation costs and improving signal coverage. The effectiveness of the proposed method was verified by a series of numerical examples, and the optimal deployment scheme of the number of macroscopic and microscopic base stations was determined by cost-benefit analysis. In 2021, Lopes et al. [27] proposed a multi-objective optimization method that prioritizes locations with the largest number of users, satisfies the maximum coverage, and reduces a large amount of computation time. In 2021, Shakya et al. [28]

studied various analytics and machine learning techniques to accurately identify 5G needs and used these techniques to plan the best 5G locations. Although the above literature have good theoretical performance, they do not simulate a large amount of data for specific base station design planning.

Base station placement and configuration with an optimization approach has been studied by Amine et al. [29]. A mathematical model based on the set coverage problem was proposed to solve the base station positioning, with the main objectives of maximizing the coverage and minimizing the financial cost. The non-dominate sorting genetic algorithm (NSGA II) was applied to find an appropriate solution. In addition, it was applied for service operators in motorway scenarios where extreme density of base stations is required to provide uninterrupted coverage. A framework to characterize the inherent trade-off has been developed by Beschastnyi et al. [30]. A technique to maximize the latter by addressing the "last hop problem" was proposed and compared to the set of alternative solutions.

In 2022, Chen et al. [31] prioritized the construction of new stations in weak coverage areas with heavy traffic, taking into account the construction cost and other factors, and used a K-means clustering algorithm to cluster weak data. For two different signal coverage forms, taking the minimum total construction cost of the new base station as the objective function, and the distance requirement of adjacent base stations and the coverage requirement of the total traffic, etc., as constraints, a single objective nonlinear programming model was established, to obtain the macro and micro base station layout, respectively. This method can give a good result, but it cannot give a unique solution. Similarly, the K-medoids clustering method was proposed by Guo et al. [32] to solve the same problem. Chen et al. [33] adopted the 0–1 program knapsack algorithm, heuristic algorithm, simulated annealing algorithm, and other methods, combined with MATLAB software to determine the location and the type of base station selection, and gave the optimal site selection result. For the same problem discussed in this paper, with the objectives of the minimum construction cost and the minimum number of overlapping coverage points, a bi-objective nonlinear function model was established by Ding et al. [34], under the constraints of base station coverage, inter-base stations threshold, base station throughput, etc., which was solved by means of an unknown optimization search method.

Comparisons between the methods used in the existing literature and in this paper have been listed in Table 1. In summary, although many scientists have conducted research on base station location, the relevant research is mainly limited to a small range and a few base stations, and the lack of data support cannot accurately complete the large-scale calculation, the search of the center requires constant hypothesis testing, there is a large systematic error, and it is difficult to give the unique design result of the base station. To solve these problems of base station location, the mean shift algorithm with the addition of a kernel function has been proposed in this paper. The types of base stations are selected by multiple drifts, the number of clustering centers is determined by machine learning, and global adjustment is performed based on the established single objective nonlinear programming model, which can largely solve the problems of premature convergence to local extremes and excessive computation caused by excessive iteration time. Tens of thousands of weak coverage points are simulated and the specific site planning scheme is given. The test results show that the improved algorithm proposed in this paper can give more realistic results with higher economic benefits, and has some application value.

Table 1. Comparison between the methods used in the existing literature and in this paper.

Problems in Existing Literature	Advantages of This Article
Cannot accurately complete the large-scale calculation.	For any scale of base station data, the ideal planning scheme can be given.
The search for the center requires constant hypothesis testing, and there is large systematic error.	The unique and optimal location of the center point can be automatically determined using the improved mean shift method.
Difficult to give the unique design result.	It can provide the unique optimal solution of the base station design scheme.

3. Data Preprocessing

The data used in this paper are from the MathorCup Mathematical Contest in Modeling [35]. To facilitate the calculation, the given area is divided into small grids and only the center point of each grid is considered in the calculation, i.e., the center point is used to replace the small grid area. For some points, it is difficult to be completely covered by the signal range of the existing base station, so these points are defined as weak coverage points in this paper.

For the area studied in this paper, the range of horizontal and vertical coordinates are defined as integers from 0 to 2499, resulting in 2500×2500 points. A large number of weak coverage points are distributed in the coverage area. The existing base stations cannot solve the signal coverage problem of these points, so new base stations must be built to re-cover them again. The types of base stations are divided into macro and micro base stations. For ease of calculation, the coverage areas and costs of base stations are scaled proportionally according to the actual situation. It is assumed that the construction cost of the macro base station is 10 and its coverage area is 30. The construction cost of the micro base station is 1, and its coverage area is 10.

It should be noted that in actual network planning, in addition to re-covering weak signal coverage points, it is also necessary to consider the benefit of constructing base stations. In this case, the weak coverage points with high traffic should be covered first. In general, in the weak coverage area of the current network, the layout requirements of the base station construction scheme can be met when the total service volume to be re-covered reaches 90%.

In practice, it is often necessary to give up some weak points of coverage that cause negative returns to the whole, in order to maximize interest. But the classification of "noise points" becomes the key to denoising. In this paper, "noise points" are firstly divided into two categories. One is isolated weak coverage points, i.e., with this point as the center, there are no other weak coverage points in the area with a coverage radius of 30. The other is that the amount of business contained in this point is so small that it can be eliminated to increase revenue, taking into account the economic benefits. However, after refining the data, it is found that there are isolated points with high traffic volumes. If this is regarded as "noise", it will have a major impact on the construction of the base station.

As you can see, it is clearly inappropriate to treat the isolated points in Table 2 as "noise". In reality, these points correspond to a large number of traffic users, but they are located in areas with poor traffic signal coverage. It is therefore more practical to reset base stations for signal coverage in these areas.

Table 2. Some of the isolated points with high data volume.

Serial Number	1	2	3	. . .
Coordinate	(844, 1962)	(1356, 2271)	(869, 2292)	. . .
Volume of business	47,795	43,295	32,201	. . .

In summary, the correct definition of "noise" is the point where the traffic volume is extremely low. After the second round of data analysis, it was found that there were a large number of weak coverage points with a traffic volume of less than 10. These points have a small proportion of traffic volume, but the number of points is very large, which in reality corresponds to areas with few data traffic users and poor traffic signals. Combined with the construction requirements of base stations, a reasonable omission of signal coverage at some of these points can not only simplify the data scale, but also improve the economic benefits. Figure 1 compares the number of weak coverage points with the traffic volume between 0 and 10 and the loss of traffic coverage after removing this data:
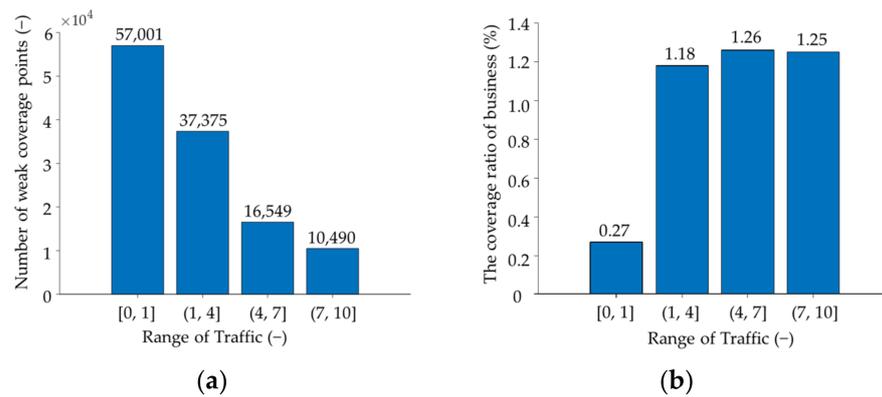
Figure 1. Data comparison of weak coverage points in different traffic sections. (**a**) corresponds to number of the weak coverage points, (**b**) corresponds to the coverage ratio of the business.

The weak coverage points in the traffic range of 0 to 1, 4, 7 and 10 are stacked and filtered one after the other. After comparative analysis, the weak coverage points with traffic volume less than 4 are defined as a "noise point".

$$s < 4 \tag{1}$$

where s is the traffic volume. Visualization of the weak coverage points before and after denoising is performed, and Figure 2 is obtained.
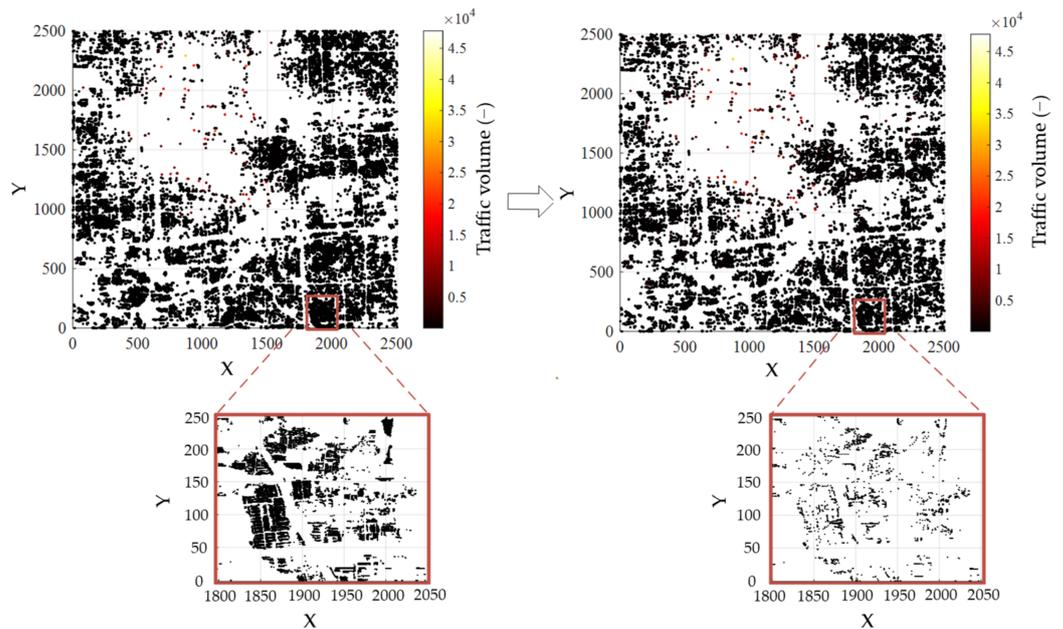

Figure 2. Comparison of signal points before and after noise filtering.

As can be seen in Figure 2, a large number of weak coverage points with very low traffic volume are filtered before and after denoising. In the end, the available weak coverage points are filtered down from 182,807 to 88,431.

From the relationship table (Table 3), between the business volume interval of weak grid points and the number of weak coverage points before and after data processing, it can be seen that the number of weak coverage points is mainly concentrated in the range [0, 1000], and the number is very small in the range of other intervals, but there are also weak grid points with high traffic.

Table 3. Comparison of business volume interval and number of weak coverage points before and after data processing.

Business Volume Interval	[0, 1000)	[1000, 2000)	[2000, 3000)	[3000, 4000)	[4000, 5000)	[5000, 10,000)	[10,000, 15,000)
Number before	182,140	376	109	57	27	62	17
Number after	87,764	376	109	57	27	62	17
Business volume interval	[15,000, 20,000)	[20,000, 25,000)	[25,000, 30,000)	[30,000, 35,000)	[35,000, 40,000)	[40,000, 45,000)	[45,000, 50,000)
Number before	9	7	0	1	0	1	1
Number after	9	7	0	1	0	1	1

4. Methodology

The most important factor in planning the construction of a base station is the choice of site coordinates and site types. We designed the system model diagram shown in Figure 3 to illustrate the applied method.
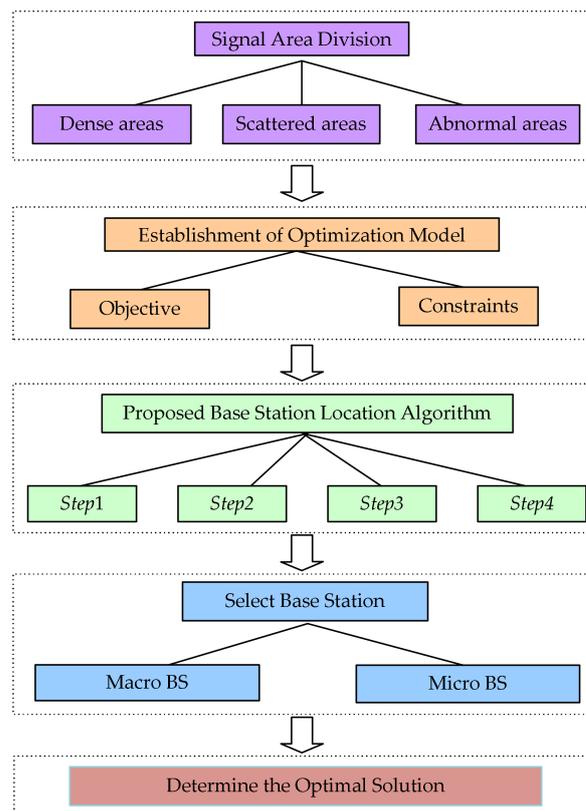

Figure 3. Schematic diagram of the system model used in this paper.

As shown in Figure 3, the whole system model diagram can be divided into the following steps. Firstly, the study area is divided into three categories according to signal intensity: dense area, scattered area, and abnormal area. Then, assuming that the signal coverage area is circular, a single objective nonlinear programming model is established with the number of macro and micro base stations as the decision variables, and the distance requirement of station construction, the condition of weak coverage points to be re-covered, the proportion of total service, etc., as the constraints. For the solution algorithm, a location algorithm based on mean shift clustering with the addition of a kernel function is proposed, and then the determination of the base station planning scheme is

completed. The application of the improved algorithm can be adjusted according to the actual needs of different coverage traffic, and finally the unique solution of the base station planning is obtained, which has very good practical value. In the following Sections 4.1–4.3, the methodology used for this problem will be described in detail.

4.1. Signal Area Division

In addition to the "noise" mentioned in the Section 3, the choice of macro and micro base stations is also critical to the signal coverage of weak coverage points. In this paper, the study area is again divided into dense, scattered, and abnormal areas, denoted n_1, n_2, and n_3, respectively. The macro base station is used to cover all dense areas. Micro base stations are used for signal coverage in all scattered areas. The layout of the base stations in the abnormal area is used to adjust the coverage of the traffic, ultimately achieving the deployment requirements of the construction scheme are. It is assumed that the number of micro base stations needed to be constructed in abnormal areas is k.

In the actual setting of regional classification, the classification conditions of different regions are very important. The larger the regional radius is, the larger the calculation error. Considering that the maximum coverage radius of the macro base station is 30 and the maximum coverage radius of the micro base station is 10, the mean shift algorithm aims to find the point with the maximum number density in an area. To maximize the number of weak coverage points covered by the macro and micro base stations, the radius should not be less than 30 when defining the dense area and 10 when defining the scattered area. To reduce calculation errors, the search coverage radius is set to 30 for dense areas and 10 for scattered areas. The traffic threshold s_0 and the number threshold n_0 of weak overwrite points must be set when determining the conditions for area division. In this paper, the range of different weak coverage points and traffic is first reduced, and then the traversal search method is used to find out the value of the threshold when the cost is lowest.

The following are the conditions for defining different regional classifications: Dense area refers to the area with a coverage radius of 30, where the number of all weak coverage points is greater than n_0 and the total traffic is not less than t_0, the macro station can be selected for signal coverage. Scattered area refers to the area where the coverage radius is 10, the number of all weak coverage points is less than n_0 and the total traffic is not less than t_0, micro base stations can be set up for signal coverage. In addition to the dense and scattered areas, it is also necessary to find the abnormal area. The abnormal area is defined as the area covered by the remaining weak coverage points outside the dense and scattered areas, and the sum of the number of all weak coverage points in this area n is greater than 0. To accurately classify the area under study, it is necessary to first separate the dense area, then find the scattered area, and finally divide the abnormal area.

Three 0–1 variables, M_a, M_b, and M_c are defined in this paper, where M_a ($a = 1, 2, \ldots, n_1$) represents the division of the dense areas. If the value of M_a is 1, it means that the area belongs to the dense areas; if the value of M_a is 0, it is not divided:

$$M_a = \begin{cases} 0, & 0 < n < n_0 \; or \; s < s_0 \\ 1, & n \geq n_0 \; and \; s \geq s_0 \end{cases} \tag{2}$$

M_b ($a = 1, 2, \ldots, n_2$) represents the division of the scattered areas. If the value of M_b is 1, it means that the area belongs to the scattered area; if the value of M_b is 0, it is not scattered:

$$M_b = \begin{cases} 0, & n \geq n_0 \; or \; s < s_0 \\ 1, & 0 < n < n_0 \; and \; s \geq s_0 \end{cases} \tag{3}$$

M_c ($c = 1, 2, \ldots, k$) represents the distribution of abnormal areas. If the value of M_c is 1, it means that the area belongs to the abnormal area; if the value of M_c is 0, it is not an abnormal area:

$$M_c = \begin{cases} 0, & n = 0 \\ 1, & n > 0 \end{cases} \tag{4}$$

4.2. Establishment of Single Objective Non-Linear Programming Model

It is assumed that the coverage area of the base station signal is circular. As for the specific planning scheme of the site, the single objective nonlinear programming model is used to determine the base station deployment, which can obtain the optimal planning scheme of the theoretical construction of the base station.

4.2.1. Objective Function and Decision Variable

For the construction of a new base station, one of the most critical indicators is the construction cost, so this paper takes the minimum value of the construction cost W, as the objective function. The total cost consists of two parts: the cost of the macro base station and the cost of the micro base station. Assuming that the costs of constructing one macro base station and one micro base station are 10 and 1, respectively, the number of macro and micro base stations required to be constructed are s_i and s_j respectively, the total cost can be expressed as:

$$W = 10s_i + s_j \tag{5}$$

The number of macro and micro base stations to be built is taken as the variable to be decided.

4.2.2. Constraint Determination

Five constraints are considered in this study, the following describes each of these constraints.

(1) When building a new site, it is necessary to ensure that the weak coverage point can be surrounded by the signal of the new base station. A_i is used to present the location of the new macro station with row number I ($I = 1, 2, \ldots, n$). B_j is the location of the new micro station with row number j ($i = 1, 2, \ldots, m$). C_o is the location of the weak coverage points with series number o ($o = 1, 2, \ldots, 88{,}431$). Since the coverage radius of the micro base station is 10 and that of the macro base station is 30, the formula can be obtained:

$$\begin{cases} \|C_o - A_i\|_2 \le 30 \\ \|C_o - B_j\|_2 \le 10 \end{cases} \tag{6}$$

(2) In the actual construction, it is necessary to consider the land occupancy, base station utilization, and other factors. The distance between the new base station and the original base station should not be less than the given limit of 10. Assume that A_i is the location of the new base station with serial number a ($a = 1, 2, \ldots, m + n$). E_b is the location of the existing station with serial number b ($b = 1, 2, \ldots, 1474$). D_a is the coordinate position of the newly constructed base station numbered a ($a = 1, 2, \ldots, m + n$). c and d are two random integers from 1 to $m + n$, and D_c and D_d are the coordinate positions of two randomly selected stations, respectively. The following equation can be obtained:

$$\begin{cases} \|D_c - D_d\|_2 \le 10 \\ \sum\limits_{a=1}^{m+n} D_a = \sum\limits_{i=1}^{n} A_i + \sum\limits_{j=1}^{m} B_j \\ \|D_a - E_b\|_2 \le 10 \end{cases} \tag{7}$$

(3) Assuming that s_a is the business volume covered by the newly built station with series number a ($a = 1, 2, \ldots, m + n$), the business volume s_o thus included in the weak coverage point with series number o ($o = 1, 2, \ldots, 88{,}431$) should satisfy the following formula:

$$\sum\limits_{a=1}^{m+n} s_a \ge \sum\limits_{o=1}^{88{,}431} s_o \times 0.9 \tag{8}$$

(4) To calculate base station construction cost of s, 0–1 variables M_a, M_b, and M_c can be used as counters to represent the number of macro and micro base stations constructed:

$$\begin{cases} s_i = \sum_{a=1}^{n_1} M_a \\ s_j = \sum_{b=1}^{n_2} M_b + \sum_{c=1}^{k} M_c \end{cases} \tag{9}$$

(5) The setting of the variable 0–1 is used to divide different areas, and different types of base stations will be built in different areas. Specific constraints are shown in Equations (1)–(3) as the final constraints.

4.2.3. Final Single Objective Nonlinear Programming Model

Finally, a single-objective nonlinear programming model is established with the number of macro and micro base stations as decision variables, the condition of weak coverage points to be re-covered (Equation (6)), the limit of base station construction distance (Equation (7)), the proportion of total business (Equation (8)), the constraint related to the numbers of macro and micro base stations (Equation (9)), and the judgement conditions of whether to build base stations (Equations (2)–(4)) as constraints, and the lowest construction cost (Equation (5)) as the objective.

$$\text{Min} \quad W = 10s_i + s_j \tag{10}$$

Subject to:

$$s.t. \begin{cases} \sum_{a=1}^{m+n} D_a = \sum_{i=1}^{n} A_i + \sum_{j=1}^{m} B_j \\ s_i = \sum_{a=1}^{n_1} M_a \\ s_j = \sum_{b=1}^{n_2} M_b + \sum_{c=1}^{k} M_c \\ \|C_o - A_i\|_2 \leq 30 \\ \|C_o - B_j\|_2 \leq 10 \\ \|D_c - D_d\|_2 \leq 10 \\ \|D_a - E_b\|_2 \leq 10 \\ \sum_{a=1}^{m+n} s_a \geq \sum_{o=1}^{88,431} s_o \times 0.9 \\ M_a = \begin{cases} 0, \, 0 < n < n_0 \text{ or } s < s_0 \\ 1, \, n \geq n_0 \text{ and } s \geq s_0 \end{cases} \\ M_b = \begin{cases} 0, \, n \geq n_0 \text{ or } s < s_0 \\ 1, \, 0 < n < n_0 \text{ and } s \geq s_0 \end{cases} \\ M_c = \begin{cases} 0, \, s \geq s_0 \\ 1, \, s < s_0 \end{cases} \end{cases} \tag{11}$$

4.3. Design of the Base Station Location Selection Algorithm

In machine learning, a kernel refers to a method that allows us to apply linear classifiers to non-linear problems by mapping non-linear data to a higher dimensional space without having to visit or understand that higher dimensional space. In order to find the center point of the base station construction more reasonably, the kernel function is introduced in mean shift.

4.3.1. Mean Shift Clustering with Adding Kernel Function

Mean shift can refer to a general term for changing the location, position, or direction of something, or a specific technique for finding the maxima of a density function by shifting data points towards the high-density region. It is a density-based non-parametric and unsupervised method commonly used for cluster analysis and image segmentation.

By introducing the kernel function into the mean shift algorithm, the contribution of the offset to the mean offset vector changes with the distance between the sample and the offset point, thus achieving the purpose of machine learning. The shifting process is shown in Figure 4:
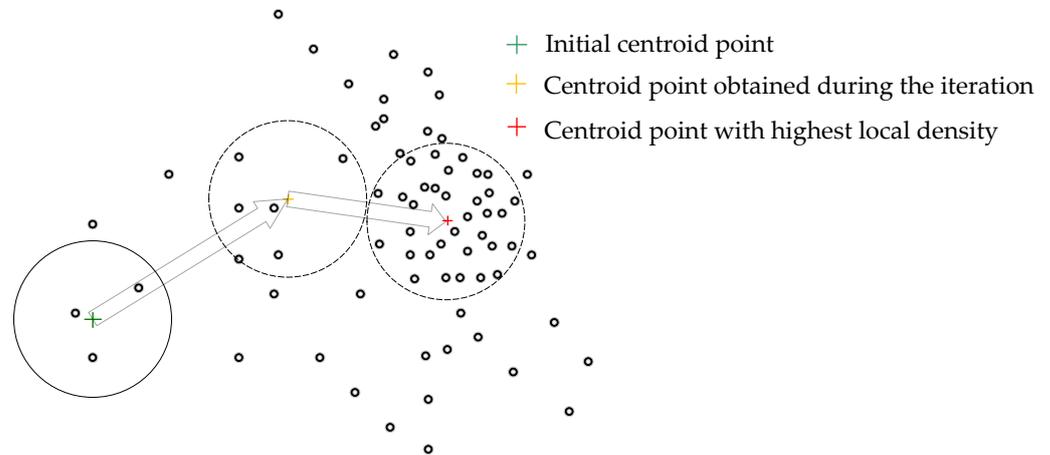

Figure 4. Schematic diagram of mean shift clustering.

In this paper, the exponential function has been used to construct the weighting function. Points far away from the center are given a small weight, so as to avoid excessive shifting of the density center caused by large gross errors. The weighting function is as follows:

$$q(P_i) = e^{-\frac{d_i}{h}} \tag{12}$$

where d_i is the distance between the vector P_i and the center of the density. The initial random density center is moved along the vector $M_h(P)$ and after several rounds of iteration, the maximum point of the density function is reached.

Suppose that a data set $X = \{P_1, P_2, \ldots, P_o\}$ is a set of experimental samples with number o in the d-dimensional space. K is an arbitrary integer from 1 to o. At each point P_k, the vector form of the mean shift is given by:

$$M_{P_k} = \frac{1}{K} \sum_{P_k \in X} (P_o - P_k) \tag{13}$$

To iteratively find the center of the maximum density region in machine learning, a kernel function must be added to the mean shift. Assuming that $P_k = [\lambda_1, \lambda_2, \ldots, \lambda_d]$ is the eigenvector of x_k in the experimental sample, h is the kernel width, and $q(P_i)$ is the sample weight, then the kernel density estimation function can be expressed as follows:

$$\rho_k(P) = \frac{\sum_{i=1}^{o} K\left(\left\|\frac{P_i - P}{h}\right\|^2\right) q(P_i)}{h^d \sum_{i=1}^{o} q(P_i)} \tag{14}$$

where $K(P)$ is the kernel function, i.e., the vector in the original space as the input vector, and returns the function of the dot product of the vectors in the feature space, taking the derivative of Equation (14), and setting $G(P) = -K'(P)$, the compensation vector $M_h(P)$ can be written as:

$$M_h(P) = \frac{\sum_{i=1}^{o} G\left(\left\|\frac{P_i - P}{h}\right\|^2\right) q(P_i) P_i}{\sum_{i=1}^{o} K\left(\left\|\frac{P_i - P}{h}\right\|^2\right) q(P_i)} - P \tag{15}$$

Then, the local mean moves iteratively towards the sample density concentration area, and the density center point is represented as follows:

$$x_{t+1} = \frac{\sum_{i=1}^{o} G\left(\left\|\frac{P_i - P}{h}\right\|^2\right) q(P_i) P_i}{\sum_{i=1}^{o} K\left(\left\|\frac{P_i - P}{h}\right\|^2\right) q(P_i)} \tag{16}$$

4.3.2. Proposed Location Algorithm Based on Mean Shift Clustering

Previously, Chen et al. [31] determined the center point of the base station by K-means clustering, but the traditional K-means clustering could not obtain the unique solution. For the single objective nonlinear programming problem of base station location, this paper proposes an algorithm based on mean shift clustering, which can independently select the location point of the base station construction according to the given coverage radius of the base station by mean shift, and obtain the unique optimal solution under the condition of profit maximization.

The solution of the base station planning scheme is completed in four steps. This method can be well adapted to different actual service volumes. The flowchart of the algorithm is shown in Figure 5:
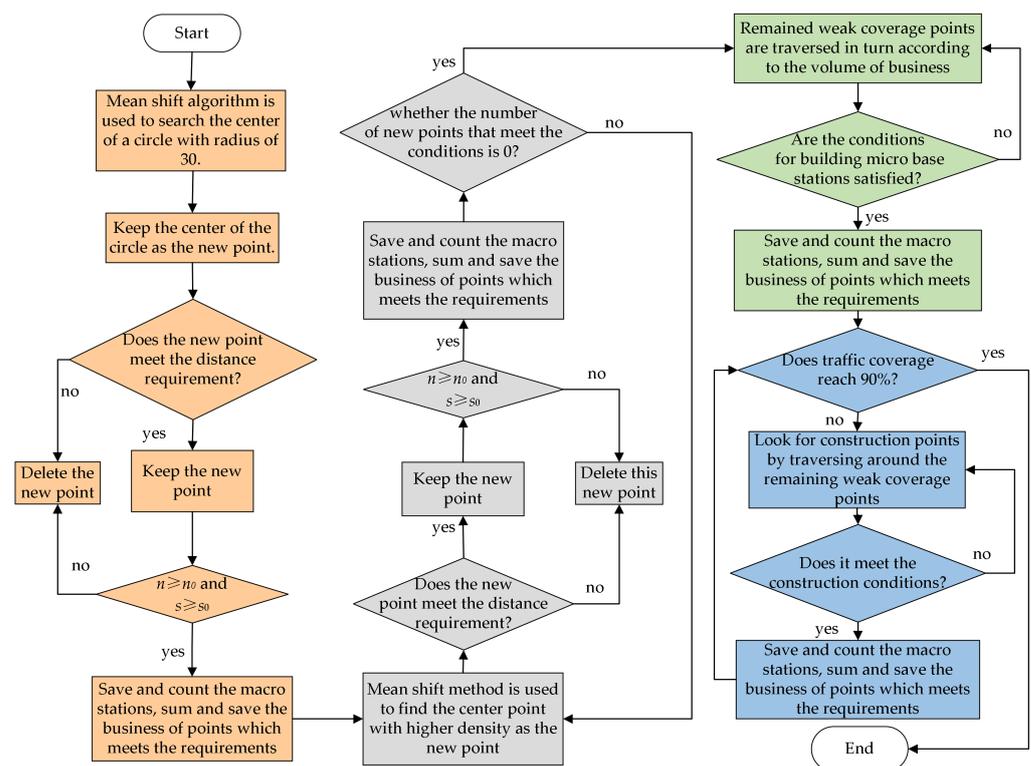

Figure 5. Flowchart of the base station site selection algorithm. Different colors correspond to the four different steps.

As shown in Figure 5, the steps of base station planning are as follows:

Step 1: Using the means of mean shift algorithm, a center point is randomly selected from all the weak grid coordinates after the initial filtering. The center point is taken as the center of the circle, and the radius is 30 to make a circular region. The vectors between all points and the center point in this circular region are calculated, then the mean value of all vectors is taken as the offset mean value, and the center point is moved to the offset mean position point. From there, iterate and repeat until the area with the highest concentration of points is reached. Take the coordinates of the center points of all the offset locations as the location coordinates of the new planned base station to be constructed.

Judge whether the points meet the conditions, calculate the distance between the newly built stations, and the distance between the existing and new stations, eliminating the new stations with a distance of less than 10. Then, the weak grid points in the circular coverage area of the new base station are evaluated. If the sum of the number of weak grid points in the region is greater than n_0, and the sum of the traffic is greater than s_0, the coordinates of the new base station are reserved and counted. Then drift points are continuously taken until the number of points satisfying the condition is 0. The points satisfying the conditions are reserved, and the sum of the traffic volume in the covered area is calculated and stored. Finally, macro stations are constructed at the reserved points. The pseudo-code of Step 1 is shown in Algorithm 1.

Algorithm 1: Algorithm for determining the number of macro base stations

---

1: **Input:** initial number of macro base stations si = 0;
2:      initial number of micro base stations sj = 0;
3:      coordinates and traffic arrays of weak coverage points D1;
4:      array of coordinates for existing base station locations D2;
5: **Output:** total covered business S;
6:      number of macro base stations si;
7:      number of micro base stations sj;
8:      coordinate array of old and new base station locations D2;
9:      coordinate array of the center point obtained by means of mean shift algorithm A;
10: **while** i < length(A) **do**//The length of array A
11:    **for** j 1 to length(D2) by 1 **do**
12:        judge whether the center point meets the distance constraint condition;
13:        **if not**: exit the for loop; (Equation (7))
14:    **end for**
15:        **if** the central point satisfies the condition; go to the next step;
16:        **if not**: jump out of the while loop;
17:    **while** j < length(D1) **do**//The length of array D1
18:        traversal to find the weak coverage points that meet the condition, record the number of
19:        weak coverage points (n) and the total amount of business (s) covered in the region with
20:        the center point as the center of the circle with a radius of 30. At the same time, remove
21:        the weak coverage points covered in D1; (Equation (6))
22:        j ← j + 1;
23:    **end while**
24:        **if** s ≥ s_0 and n ≥ n_0 then (Equation (2))
25:        s_i++; store this center point in D2; S ← S + s;
26:        **else**:
27:        restore the covered weak coverage points in D1;
28:        **end if**
29:        i ← i + 1;
30: **end while**
31: M ← s_i;
32: assign an initial value to the flag that exits the loop to Step 2.

Step 2: The mean shift algorithm is used to randomly select the center points among the remaining weak grid points. With the selected center point as the center of the circle and the radius of 10 as the area of the circle, the vector between all the points and the center point in the area of the circle is calculated, and the average value is taken as the mean shift value. The center of the circle is constantly moved to the next average offset point until the highest density is reached. Take the center points of all the offset locations as the position coordinates of the new base station to be built. Calculate the distance between the newly built stations and the distance between the existing and new stations, eliminating the new stations with a distance of less than 10. Eliminate the points whose total number of weak grid points in the region is not greater than n_0 and whose traffic volume is not less than s_0, reserving the coordinates and counting them. The sum is then added to the amount of traffic reserved in Step 1. Then, the drift continues and the point coordinates that satisfy the condition are kept until the number is 0. Finally, micro base stations are built at the reserved points. If the total service volume is not less than 90% of the service volume of the

weak coverage point, the algorithm is terminated. If the conditions for jumping out are not met, go to Step 3. The pseudo-code of Step 2 is shown in Algorithm 2.

Algorithm 2: Algorithm based on mean shift clustering to determine the number of micro base stations

---

1: **Input:** coordinates and traffic arrays of weak coverage points D1;
2:      array of coordinates for existing base station locations D2;
3:      the sign M that jumps out of the loop in step 2;
4:      total covered business S;
5:      number of macro base stations si and number of micro base stations sj obtained in Step 1;
6: **Output:** total covered business S;
7:      number of macro base stations si;
8:      updated number of micro base stations sj;
9:      array of coordinates for existing base station locations D2;
10: **while** M! = 0 **do**
11:      M ← 0;
12:      coordinate array of center point obtained by means of the mean shift algorithm A;
13:      **while** i < length(A) **do**//The length of array A
14:          **for** j 1 to length(D2) by 1 **do**
15:          judge whether the center point satisfies the distance constraint condition;
16:          **if not**, exit the for loop; (Equation (7))
17:          **end for**
18:          **if** the central point satisfies the condition, it goes to the next step;
19:          **if not**, exit the while loop;
20:              **while** j < length(D1) **do**//The length of array D1
21:              traversal to find the weak coverage points that meet the condition, record the
22:              number of weak coverage points (n) and the total amount of business (s) covered in
23:              the region with the center point as the center of the circle and the radius of 10. At
24:              the same time, remove the weak coverage points covered in D1; (Equation (6))
25:              j ← j + 1;
26:              **end while**
27:          **if** s ≥ s0 and n ≤ n0 then (Equation (3))
28:          sj++; store this center point in D2; S ← S + s; M ← M + 1;
29:          **else**:
30:          restore the covered weak coverage points in D1;
31:          **end if**
32:          i ← i + 1;
33:      **end while**
34:      judge whether the traffic volume S satisfies the conditions;
35:      **if yes**, stop the algorithm;
36:      **if no**, continue the algorithm;
37: **end while**

Step 3: In the above step, after using the drift algorithm, some weak grid points are missed. These weak grid points are sorted according to the traffic volume. The points that meet the conditions for building the micro base station are searched successively by traversing, and the coordinates of the weak coverage points that meet the conditions are reserved and counted, and the sum is made with the previously stored traffic. Finally, the micro base stations are built on the reserved points. If the total service volume is not less than 90% of the service volume of the weak coverage point, the algorithm is terminated. If the jump out condition is not satisfied, go to Step 4. The pseudo-code of Step 3 is shown in Algorithm 3.

Step 4: If the above steps still cannot satisfy the total traffic volume of more than 90% of the traffic volume of the weak coverage point, then continue to find points near the remaining weak grid points to establish the micro base station. The remaining weak grid points are taken as the center of the circle, and within the radius of 10, the points that satisfy the conditions for establishing the micro base station are traversed. They are extracted one-by-one according to the size of the service volume, their coordinates are kept and counted, and then added to the previously stored service volume. The algorithm is terminated when the total stored traffic is more than 90% of the total traffic of the weak overlay point. The architecture of the algorithm is shown in Algorithm 4:

Algorithm 3: Algorithm for determining the number of micro base stations based on traffic sequencing

---

1: **Input:** coordinates and traffic arrays of weak coverage points D1;
2:      array of coordinates for existing base station locations D2;
3:      total covered business S;
4:      number of macro base stations si and number of micro base stations sj obtained in Step2;
5: **Output:** total covered business S;
6:      number of macro base stations si;
7:      number of micro base stations sj;
8:      array of coordinates for existing and newly constructed base stations D2;
9:      array D1 is sorted by traffic volume;
10: **while** i < length(D1) **do**//the length of array D1
11:      judge whether the business amount S satisfies the requirement;
12:      **if** satisfied, stop the algorithm;
13:      **if not** satisfied, continue the algorithm; (Equation (8))
14:          **for** j 1 to length(D2) **do**
15:          judge whether the remaining weak coverage points satisfy the distance constraint;
16:          **if not**, exit the for loop; (Equation (7))
17:          **end for**
18:          **if** the central point satisfies the condition, it goes to the next step;
19:          **if not,** exit of the while loop;
20:              **while** j < length(D1) and i ! = j **do**//The length of array D1
21:              traversal to find the weak coverage points that meet the condition, record the
22:              number of weak coverage points (n) and the total amount of business (s) covered in
23:              the region with the center point as the center of the circle with a radius of 10. At the
24:              same time, remove the weak coverage points covered in D1; (Equation (6))
25:                  j ← j + 1;
26:              **end while**
27:          **if** n > 0 then (Equation (4))
28:          sj++; store this center point in D2; S ← S + s;
29:          **else**:
30:          restore the covered weak coverage points in D1;
31:          **end if**
32:      i ← i + 1;
33:      judge whether the traffic volume S satisfies the conditions;
34:      **if yes**, stop the algorithm;
35:      **if not**, continue the algorithm; (Equation (8))
36: **end while**

Algorithm 4: Algorithm for determining the number of micro base stations based on grid points

---

1: **Input:** coordinates and traffic arrays of weak coverage points D1;
2:      array of coordinates for existing base station locations D2;
3:      total covered business S;
4:      number of macro base stations si and number of micro base stations sj obtained in Step 3;
5: **Output:** total covered business S;
6:      number of macro base stations si;
7:      number of micro base stations sj;
8:      array of coordinates for existing and new built base stations D2;
9:      array D1 is sorted by traffic volume;
10: **while** i < length(D1) **do**//The length of array D1
11:      judge whether the business amount S satisfies the requirement;
12:      **if** satisfied, stop the algorithm;
13:      **if not** satisfied, continue the algorithm; (Equation (8))
14:          **for** j 1 to length(D2) by 1 **do**
15:          judge whether the remaining weak coverage points satisfy the distance constraint;
16:          **If not**, exit the for loop; (Equation (7))
17:          **end for**
18:          **if** the central point satisfies the condition, it goes to the next step;
19:          **if not**, exit the while loop;
20:              **while** j < length(D1) **do**//The length of array D1
21:              traversal to find the weak coverage points that meet the condition, record the
22:              number of weak coverage points (n) and the total amount of business (s) covered in
23:              the region with the center point as the center of the circle with radius of 10. At the
24:              same time, remove the weak coverage points covered in D1; (Equation (6))
25:                  j ← j + 1;
26:              **end while**
27:          **if** n > 0 then (Equation (4))
28:          si++; store this center point in D2; S ← S + s;
29:          **else**:
30:          restore the covered weak coverage points in D1;
31:          **end if**
32:      i ← i + 1;
33:      judge whether the traffic volume S satisfies the conditions;
34:      **if yes**, stop the algorithm;
35:      **if no**, continue the algorithm; (Equation (8))
36: **end while**

5. Results and Discussion

The solution of the result can be divided into two parts: the selection of the threshold value and the planning of the base station site. After selecting the optimal plan of the base station deployment plan, the results are further discussed.

5.1. Selection of Regional Classification Threshold

When planning the base station layout, it is necessary to find the most suitable partition threshold by traversal search. According to the data in this paper, it is found that if the threshold of the number of weak coverage points is greater than 185, the number of macro stations will be greatly reduced. If the threshold is less than 177, the number of macro stations increases significantly. The main reason is that the threshold setting is not reasonable and does not match the distribution of the actual data. According to the actual data, the threshold should be set randomly for the pre-experiment. According to the pre-experiment results, the value of the threshold should be increased or decreased to determine the optimal traversing step size, which can greatly reduce the traversing range and ensure the rationality of the threshold. A part of the pre-test results that meet the base station layout conditions are shown in Table 4.

Table 4. Results of the pre-test of the threshold values.

n_0	s_0	Total Cost
. . .	. . .	. . .
177	280	3772
178	280	3777
179	280	3774
180	280	3769
. . .	. . .	. . .

Finally, according to the data used in this paper, set the traversal interval of the quantity threshold of weak coverage points to be 180~185 and the substep length to be 1, the traversal interval of the total service threshold to be 280~290 and the substep length to be 1, the final traversal data are numbered from 1 to 12, and the list of results is shown in Figure 6.
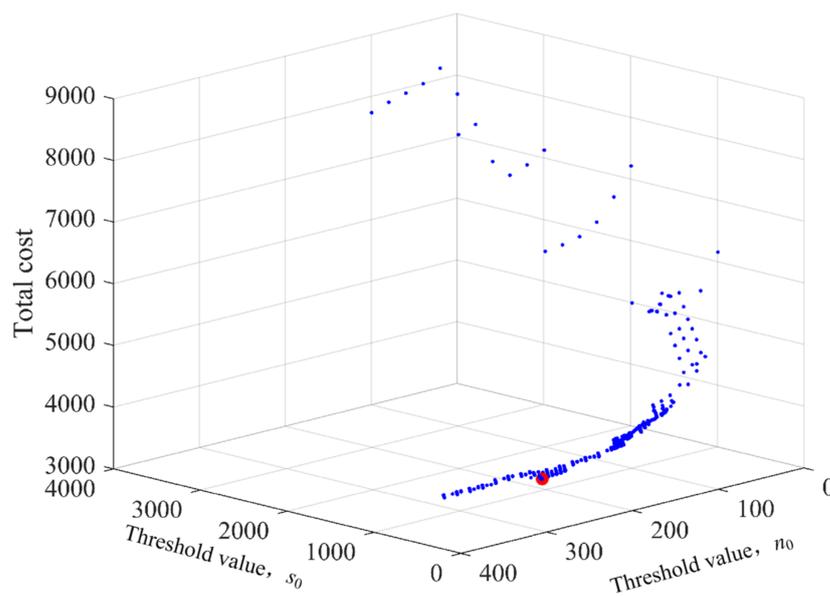

Figure 6. Diagram showing how to select the threshold. The red point corresponds to the best threshold with lowest construction cost.

As shown in Table 5, assuming that the base station layout requirements are satisfied, the base station construction cost is the lowest when the threshold value corresponds to row number 19. Therefore, in this paper, the number threshold of weak coverage points is 181, and the total traffic threshold is 287.

Table 5. The corresponding threshold for each number.

Series Number	n_0	s_0	Series Number	n_0	s_0
1	180	280	12	181	280
2	180	281	13	181	281
3	180	282	14	181	282
4	180	283	15	181	283
5	180	284	16	181	284
6	180	285	17	181	285
7	180	286	18	181	286
8	180	287	19	181	287
9	180	288	20	181	288
10	180	289	21	181	289
11	180	290	22	181	290

5.2. Results of the Optimal Base Station Planning

Aiming at the single-objective nonlinear programming model established in this paper, the improved algorithm based on mean shift clustering is used in Matlab to obtain the base station site selection planning results of the base station, as shown in Table 5. Draw the location map of all the new base stations, as shown in Figure 7, where the large red circle represents the macro base station, and the small dark blue circle represents the micro base station:
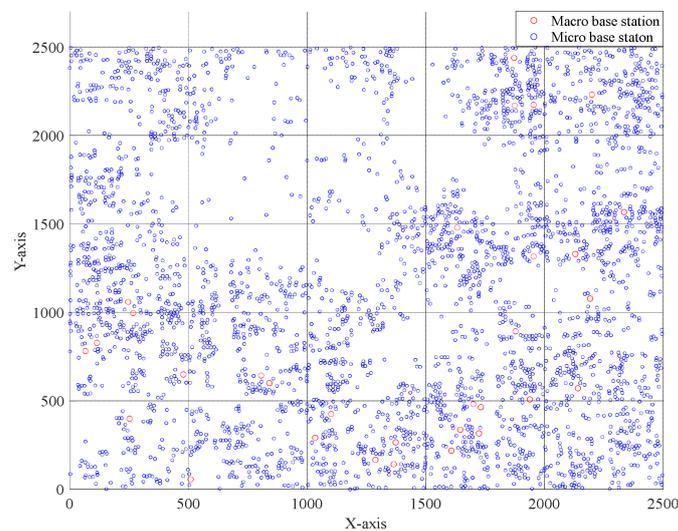

Figure 7. Diagram showing where to locate the new base station.

Finally, according to the method proposed in this paper, the design of the optimal site planning scheme is to establish 31 macro stations and 3442 micro base stations, so that the traffic coverage rate is 91.43% and the cost is 3752.

5.3. Discussions

First, the threshold value of the region partitioning is selected by traversal search on the given data. Through the comparative analysis of a large number of data, we find that if the threshold of the number of weak coverage points is greater than 185, the number of macro base stations will be greatly reduced. If the threshold is less than 177, the number of macro base stations increases significantly. This phenomenon is unreasonable. Therefore,

the sub-step size of the traversal search is reduced to 1 and a more appropriate search is performed in the range of 177–185. Finally, by analyzing the results of the program operation, it can be found that the cost is the lowest and the practical benefit is the best when the number of weak coverage points threshold is 181 and the total amount of business threshold is 287.

The method used in this paper can find the optimal solution of each partition, so as to achieve the purpose of global optimization, which can directly, quickly, and accurately obtain the only optimal answer.

Currently, there are various methods used to solve the problem in this paper, such as: the K-means clustering proposed by Chen et al. [31], the K-medoids clustering used by Guo et al. [32], the simulated annealing algorithm used by Chen et al. [33], as well as the unknown optimization search method carried out by Ding et al. [34]. We compared the results obtained by the above methods, as shown in Table 6.

Table 6. The comparison results of different methods.

Methods	Authors	Number of Macro Base Stations	Number of Micro Base Stations	Business Covered	Total Cost
K-means clustering	Chen et al. [31]	412	8856	93.53%	12,976
K-medoids clustering	Guo et al. [32]	99	1231	90.00%	2221
Simulated annealing algorithm	Chen et al. [33]	3302	216	91.09%	33,236
Unknown optimization searching method	Ding et al. [34]	162	2047	90.01%	3667
Improved mean shift clustering	Method proposed in this paper.	31	3442	91.43%	3752

For the K-means clustering algorithm adopted in [31], the number of macro and micro base stations to be built is 412 and 8856, respectively, and the minimum total cost is 12,976. However, the center point of the base station of K-means clustering needs to be confirmed in advance, so it is difficult to obtain the unique optimal solution. The result obtained is not the only solution, and the total cost is higher than the result in this paper.

For the simulated annealing algorithm proposed in [33], the result of base stations needed to be built are 3302 macro base stations and 216 micro base stations, which is obviously not reasonable.

For the K-medoids cluster analysis model proposed in [32], a K-medoids cluster-based method was proposed to find the main direction of the sector. DBSCAN density cluster analysis was used to divide the regions by distance. Finally, the result of 1330 new base stations was obtained, including 1231 micro base stations and 99 macro base stations. The total cost of new base stations is the lowest at 2221, and the coverage rate is 90.00%.

For the literature [34], the minimum construction cost and the minimum overlap of coverage points were taken as objectives, taking into account the sector distribution. An unknown optimal search method was used to solve the problem. The results show that 162 macro base stations and 2047 micro base stations need to be built if the constraints are met. The minimum total cost is 3667 and the coverage rate is 90.01%.

Comparing the results with this paper, the total construction costs of the above two methods [32,34] are slightly lower. Both of these two methods consider the sector distribution, but the final coverage rate of the base station is small. In addition, neither of these two papers provided the design of the algorithms, so it is difficult to complete the data reproduction according to the current content, and the accuracy of the results cannot be evaluated.

However, compared with the existing method, the advantages of this method are obvious, for the base station planning scheme designed in this paper, by finding the

optimal solution of each partition, the global optimization can be achieved. The algorithm proposed in this paper can significantly reduce the construction cost of the base station, and has stronger practicability and higher economic benefits in actual base station planning. The unique and optimal location of the center point can be automatically determined by using the improved mean-shift method.

In addition, the proposed method still has some shortcomings in the selection of the partition threshold. The method adopted to find appropriate threshold data by range shrinking and ergodic search takes a long time to realize, and the time complexity of the method is relatively high. If a model can be found that can realize fast search, the time problem caused by insufficient computing power can be further reduced, and scheme planning with higher efficiency and range can be completed in the actual construction of the base station deployment process.

6. Conclusions

First of all, this paper innovatively divides the simulation area where the given weak coverage point is located into dense area, scattered area, and abnormal area. The regionalization of a large number of complex data blocks improves the computational efficiency and has strong practicability in the actual situation, especially for a large number of base station data.

Then, a single objective nonlinear programming model is established, and a new base station location algorithm based on the mean-shift clustering model is designed for the first time. Compared with the existing algorithms, this algorithm has certain advantages, which can effectively avoid the problem of excessive computation and local optimal solution dilemma.

After that, this paper does not consider the possibility that the coverage signal of the base station may be disturbed by external factors in the area where the base station is located, so that the coverage norm becomes smaller, such as the interference caused by terrain and other high-intensity electromagnetic wave signals.

Finally, by adding the kernel function to the mean shift clustering algorithm, the center point of the base station is determined, and the optimal threshold of the base station type classification is searched by traversal, and the unique optimal solution is determined under different data conditions, which has high accuracy and good economic benefits. However, the time complexity of the threshold search algorithm is high, as it takes a long time to plan the layout of a large area, so there is room for improvement of the algorithm.

Through the test results of this model, it can be seen that the base station location method proposed in this paper has strong practicability, and the model and algorithm can be widely used in all kinds of large data processing and location problems, and is not limited to base station selection, with good promotion benefits.

References

1. Tan, Y.; Liu, J.; Wang, H.; Xian, M. The Development Trend Analysis of 5G Network. In Proceedings of the 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China, 5–7 July 2019; pp. 382–385.
2. Taheribakhsh, M.; Jafari, A.; Peiro, M.M.; Kazemifard, N. 5G Implementation: Major Issues and Challenges. In Proceedings of the 2020 25th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 1–2 January 2020; pp. 1–5.
3. Lukauskas, M.; Ruzgas, T. A Review of Clustering Algorithms and Application. In Proceedings of the International Conference on Applied Analysis and Mathematical Modelling (ICAAMM2021), Istanbul, Turkey, 11–13 June 2021.
4. Putra, E.D.; Rifqo, M.H.; Deslianti, D. Analysis of the Theme Clustering Algorithm Using K-Means Method. *J. Komput. Inf. Dan Teknol. (JKOMITEK)* **2022**, *2*, 431–442. [CrossRef]
5. Sun, J.; Du, W.; Shi, N. A Survey of kNN Algorithm. *Inf. Eng. Appl. Comput.* **2018**, *1*, 770. [CrossRef]
6. Deng, D. DBSCAN Clustering Algorithm Based on Density. In Proceedings of the 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), Hefei, China, 25–27 September 2020; pp. 949–953.
7. Fu, H.; Sun, F.; Liu, S. Anti-occlusion Tracking Algorithm Based on LSSVM Prediction and Kalman-MeanShift. In Proceedings of the 2010 8th World Congress on Intelligent Control and Automation, Jinan, China, 7–9 July 2010; pp. 6031–6036. (In Chinese).
8. Ting, F.; Zhao, Y.; Hu, X. An Improved MeanShift Insulator Image Segmentation Algorithm. *Adv. Mat. Res.* **2012**, *634–638*, 3945–3949. [CrossRef]
9. Chen, T. On the Convergence and Consistency of the Blurring Mean-Shift Process. *Ann. Inst. Stat. Math.* **2015**, *67*, 157–176. [CrossRef]
10. Gong, S.; Li, C.; Hou, J.; Yu, L. Dynamic Gesture Capture, Location and Tracking Based on MeanShift Algorithm. In Proceedings of the 2019 Chinese Intelligent Systems Conference (CISC 2019), Haikou, China, 26–27 October 2019; pp. 66–74.
11. Demirović, D. An Implementation of the Mean Shift Algorithm. *Image Process. Line* **2019**, *9*, 251–268. [CrossRef]
12. Yang, J.; Rahardja, S.; Fränti, P. Mean-shift Outlier Detection and Filtering. *Pattern Recognit.* **2021**, *115*, 107874. [CrossRef]
13. Yao, H. A Survey for Target Tracking on Meanshift Algorithms. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 15–17 January 2021; pp. 476–479.
14. Park, H. α-MeanShift++: Improving MeanShift++ for Image Segmentation. *IEEE Access* **2021**, *9*, 131430–131439. [CrossRef]
15. Chen, Q.; He, L.; Diao, Y. A Novel Neighborhood Granular Meanshift Clustering Algorithm. *Mathematics* **2022**, *11*, 207. [CrossRef]
16. Cui, J.; Wang, Y.; Wang, K. Key Technology of the Medical Image Wise Mining Method Based on the Meanshift Algorithm. *Emerg. Med. Int.* **2022**, *2022*, 6711043. [CrossRef]
17. Cariou, C.; Le Moan, S.; Chehdi, K. A Novel Mean-Shift Algorithm for Data Clustering. *IEEE Access* **2022**, *10*, 14575–14585. [CrossRef]
18. Mathar, R.; Niessen, T. Optimum Positioning of Base Stations for Cellular Radio Networks. *Wirel. Netw.* **2000**, *6*, 421–428. [CrossRef]
19. Zimmermann, J.; Höns, R.; Mühlenbein, H. ENCON: An Evolutionary Algorithm for the Antenna Placement Problem. *Comput. Ind. Eng.* **2003**, *44*, 209–226. [CrossRef]
20. Resende, M.G.C.; Pardalos, P.M. *Handbook of Optimization in Telecommunications*; Springer: New York, NY, USA, 2006.
21. Mai, W.; Liu, H.; Chen, L. Multi-objective Evolutionary Algorithm for 4G Base Station Planning. In Proceedings of the 2013 Ninth International Conference on Computational Intelligence and Security, Emeishan, China, 14–15 December 2013; pp. 85–89.
22. Singh, W.; Sengupta, J. An Efficient Algorithm for Optimizing Base Station Site Selection to Cover a Convex Square Region in Cell Planning. *Wirel. Pers. Commun.* **2013**, *72*, 823–841. [CrossRef]
23. Valavanis, I.K.; Athanasiadou, G.E.; Zarbouti, D. Multi-Objective Optimization for Base-Station Location in Mixed-Cell LTE Networks. In Proceedings of the 10th European Conference on Antennas and Propagation (EuCAP), Davos, Switzerland, 10–15 April 2016.
24. Kang, M.; Chung, Y. An Efficient Energy Saving Scheme for Base Stations in 5G Networks with Separated Data and Control Planes Using Particle Swarm Optimization. *Energies* **2017**, *10*, 1417. [CrossRef]
25. Goudos, S.K.; Diamantoulakis, P.D.; Karagiannidis, G.K. Multi-Objective Optimization in 5G Wireless Networks with Massive MIMO. *IEEE Commun. Lett.* **2018**, *22*, 2346–2349. [CrossRef]
26. Wang, Y.; Chuang, C. Efficient eNB Deployment Strategy for Heterogeneous Cells in 4G LTE Systems. *Comput. Netw.* **2015**, *79*, 297–312. [CrossRef]
27. Lopes, R.L.F.; Gomes, I.R.; Gomes, C.R. Improved Multi-Objective Optimization for Cellular Base Stations Positioning. *J. Microw. Optoelectron. Electromagn. Appl.* **2021**, *20*, 870–882. [CrossRef]
28. Shakya, S.; Roushdy, A.; Khargharia, H.S.; Musa, A.; Omar, A. AI Based 5G RAN Planning. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; pp. 1–6.
29. Amine, O.M.; Sylia, Z.; Selia, K.; Mohamed, A. Optimal Base Station Location in LTE Heterogeneous Network Using Non-dominated Sorting Genetic Algorithm II. *Int. J. Wirel. Mob. Comput.* **2018**, *14*, 328–334. [CrossRef]
30. Beschastnyi, V.; Machnev, E.; Ostrikova, D.; Gaidamaka, Y.; Samouylov, K. Coverage, Rate, and Last Hop Selection in Multi-Hop Communications in Highway Scenarios. *Mathematics* **2023**, *11*, 26. [CrossRef]

31. Chen, J.; Tian, J.; Jiang, S.; Li, H.; Xu, J. The Allocation of Base Stations with Region Clustering and Single-Objective Nonlinear Optimization. *Mathematics* **2022**, *10*, 2257. [CrossRef]
32. Guo, R.; Zhang, J. Research on 5G Communication Station Location Planning and Regional Clustering Based on K-medoids and DBSCAN Algorithm. In Proceedings of the 2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Dalian, China, 11–12 December 2022; pp. 1097–1101. [CrossRef]
33. Chen, Y.S.; Fu, F.Q. Research on Site Planning of Mobile Communication Network. *J. Appl. Math. Phys.* **2022**, *10*, 2708–2719. [CrossRef]
34. Ding, J.-L.; Wang, M.; An, M.-Y.; Yuan, D.-L.; Shen, Y.-C.; Cao, X.-J. Sector-like Optimization Model of 5G Base Transceiver Stations Redeployment and the Generalization. *J. Comb. Optim.* **2023**, *45*, 71. [CrossRef]
35. MathorCup-University Mathematical Modeling Competition, Question 2022. Chinese Society of Optimization, Overall Planning and Economic Mathematics. Available online: http://mathorcup.org/detail/2378 (accessed on 20 March 2023).