

RESEARCH ARTICLE

Network Site Optimization and Clustering Study Based on Simulated Annealing Algorithm

LIN-SHEN YANG¹, BIN WEN¹, AND JIE-JUN YAN¹

School of Artificial Intelligence, Tiangong University, Tianjin 300387, China

Corresponding author: Jie-Jun Yan (719196517@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 12372020.

ABSTRACT Nowadays, communication networks are becoming increasingly complex. This paper aims to demonstrate an effective method to achieve the intelligent planning for network base stations (BSs). The various parameters such as BS coordinates (x, y), the collaboration of multiple types of BS, and the density of BS construction are taken as design parameters for BS placement. We construct the objective function using the lowest total cost and the total minimum workload of BS to 90%. To solve the problem of siting planning with large data volume and mixed placement of multiple BS, we propose a new practical three-step model for BS siting planning: (I) roughly selecting the alternative coordinates for the BS using the DBSCAN algorithm; (II) correcting and further refining the alternative BS coordinates using the K-means algorithm; (III) determining the optimal BS construction solution to meet the requirements using simulated annealing algorithm (SAA). The real data of a 2500×2500 area have been used for the simulation test. The simulation result shows that BS placement covers 90.03% of the workload, confirming that the proposed method can handle site planning for large orders of magnitude of data and use a mix of BS to achieve the best economics for the demand. This paper provides basic support for future research on network site optimization.

INDEX TERMS DBSCAN algorithm, K-means algorithm, simulated annealing algorithm, base station (BS) planning.

I. INTRODUCTION

Mobile communication technology is proliferating, and the scale of operation is getting bigger, bringing more complex communication networks. With the development of 5G, communication bandwidth is getting bigger. However, the area that BSs can cover is getting smaller, making the number of BSs needed to cover the same area more. In addition, the types of BSs and antennas have also become more varied. Various BSs make the planning of communication networks, especially the problem of station site selection, more complicated. Rational BS placement plays a key role in the massive data exchange and communication within cities, which can reduce government overhead, coordinate the development of communication quality across regions and improve environmental quality. Apart from cities, optimal BS placement is also important: BS location planning at sea [1] can improve the accuracy of landing point, calculation speed of positioning and positioning accuracy of high-speed targets at sea [2]; Most of the BSs were in ruin during rescue and relief in disas-

ter areas, and planning and research on the placement method of UAVs(Unmanned Aerial Vehicles) BSs was needed [3] This will speed up the rescue process and maximize the safety of personnel.

Planning an urban communication network is intricate and complex, and its cost and complexity are closely related to the location and number of BSs. The problem of station site selection is: according to the coverage of antennas of the existing network, a certain number of points are given to the weak coverage area of the existing network so that the coverage problem of the weak coverage area can be solved. Many factors influence it: for example, too many reference placement points of BSs, the selection of BS types, whether the total service volume achieves sufficient coverage, and whether the cost is the lowest. Therefore, a set of suitable methods for urban communication BS planning under practical constraints with a large amount of data is needed to solve the problem of high cost and complex operation of traditional planning [4].

The site selection and construction of base stations (BSs) entails three planning objectives: coverage, user demand, and signal quality [5]. The scientific planning requires using the

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang¹.

BS location and coverage information, combined with the capacity and transceiver characteristics of the proposed BS [6]. In addition, the construction cost is a significant factor in meeting the user service demand while ensuring the BS radius and coverage design. Therefore, to solve such problems, there are two main approaches: using the coverage area [7] or the size of the service volume [6] for deciding whether to meet the demand. The solution to the coverage area approach can be divided into two categories: Set Covering Location Problems (SCLP) [8], which aim to minimize the cost of meeting the demanding across all areas, and Maximal Covering Location Problems (MCLP) [9], which aim to achieve as large an area as possible at a given cost. Given its practical applications in fields such as electronic vehicle BSs [10] and Unmanned Aerial Vehicle (UAV) BSs [11], MCLP has been paid considerable attention by many scholars.

Various methods and tools have been used practically to determine the optimal construction location of BS for communication network planning. In the past, researchers often utilized multi-objective linear programming methods to solve these problems; however, over time various approaches such as simulated annealing algorithm [12], particle swarm optimization algorithm [13], genetic algorithm [5], discrete fireworks algorithm [6], and greedy algorithm [14] have gradually become the mainstream solutions to such problems. Among them, multi-objective linear programming methods provide higher-quality solutions by relaxing some constraints [15]. The performance of population-based evolutionary algorithms, such as genetic algorithms, extensively depends on the efficiency of coding [6]. Hence, developing new algorithms or taking measures to improve the existing methods is crucial.

In the previous period, many scholars have proposed various models for constructing BS, among which, the two-step model proposed by Downs and Camm [16] opened up new ideas for solving such problems. The first step of this model is to determine the candidate locations set (CLS), and the second step is to select the alternative sites that satisfy the conditions in the corresponding CLS using a heuristic algorithm to achieve the maximum coverage. This method can find the optimal solution in most cases. Instead of the two-step model, He et al. proposed a three-step model [17], which selects CLSs as the initialization phase in the first step, like the two-step model. In the second step, they remove some alternative BS coordinates with low coverage service from the selected CLSs as the correction phase. Further, they select alternative sites based on the correction phase in the third step to achieve maximum coverage. Besides, Amine et al [18] considered several factors for BS construction in a realistic situation, constrained the traffic demand, and used the optimal coverage with the lowest economic overhead as the planning objective. Genetic algorithms were applied to solve this multi-objective planning problem.

Recently, many scholars have considered several practical factors in the existing BS siting problems. A distance-

decaying function for the facility coverage was considered by Hagi et al. [19]. Arana-Jiménez et al. [20] modeled this problem from a fuzzy perspective producing suitable fuzzy solutions. The study of Baldomero-Naranjo et al. [21] considered the demand unknown and distributed along the edges. In the study of Tedeschi and Andretta [22], the problem of selecting the optimal site for a BS with elliptical coverage is considered. Rodríguez et al. [23] analyzed the effect of vehicles' average utilization on the siting in a study on facility location and equipment emplacement led by. Nwelih et al. [24] introduced a weighted fitness function that combines coverage, capacity, and transmit power parameters in this field. However, in most of the articles, only the influence of the selection of coordinates for a single type of BS was considered, neglecting the complexity of the practical situation where multiple types of BSs work in combination [25], [26]. Besides, only a few papers provide solutions for practical situations of a large order of magnitude [17], [27].

In summary, the current BS placement methodology has several challenges and limitations, as follows:

- (1) Most of them contain only a single type of base station construction and do not consider a mixture of multiple BSs.
- (2) The stability problem of coexistence between BSs, i.e., the threshold between BSs, is not ignored.
- (3) Most of the siting schemes have optimization space. Moreover, only partially utilize the optimization algorithm to reach the optimal solution for a given situation.

Based on the above analysis, this paper intends to consider a network planning model that applies to multiple BS working together and it is suitable for large-scale data computation adopting the proposed improved three-step model to meet the multiple requirements of BS planning (including coverage, user requirements, and signal quality).

The three-step model entirely considers the challenges and limitations proposed above and expects to use clustering ideas to solve the stability problem of BS coexistence and optimization ideas to solve the problem of mixed construction of multiple BSs and the optimal solution of the scheme.

The ideas of the model mainly consist of the following three points:

- (1) Preliminary clustering of a large number of BS coordinates data points based on the DBSCAN algorithm, where (the radius parameter) ε is set to a given threshold of 10 between BSs and (the neighborhood density threshold) parameter is 1. Computer simulation can obtain the set of coordinates of the proposed station site;
- (2) Optimizing the set of proposed site coordinates obtained by the DBSCAN algorithm in (1) and using the K-means algorithm to cluster the obtained set of site coordinates can avoid some of the data clusters in (1) being too large to ensure the stability of the BS operation (the radius of the cluster exceeds a given threshold distance), where the parameter K of the algorithm is one-tenth of the maximum value of the horizontal and vertical distance of the cluster in which it is located.

(3) Considering the actual situation of mixed construction of multiple BSs, the set of coordinates of the proposed BSs obtained in (2) is optimized again, and the station coordinates are used to select which BS to be constructed as the decision variable, the total cost of construction is minimized as the objective function, and the service volume in the coverage area is greater than 90% of the total service volume as the constraint for planning. Finally, based on the simulated annealing algorithm (algorithm parameters: initial temperature of 1000, maximum number of iterations of 1000, number of iterations at each temperature of 10000, and temperature decay coefficient of 0.95), the final station plan is obtained.

The planning model can be applied to various BS network planning models. It represents a breakthrough from most traditional BS planning models that include only a single type of BS and can combine different types of BSs to achieve optimal planning objectives. Also, the model has a flexible structure, and the BS site planning for any requirement can be accomplished by converting the specified constraints and established objectives of the actual problem into the corresponding parts of the model. Besides, with the aid of the clustering algorithms, the method proposed in this paper (combining initialization in step 1 with correction of optional station coordinates in step 2) can greatly reduce the amount of data to be processed during the optimization of the planning scheme (in step 3). The effectiveness of the model is verified in a 2500×2500 test data set.

This paper uses a real dataset with a size of 2500×2500 as an example to explain how the model works. The remaining parts of this paper are organized as follows. Section II discusses the BS data pre-processing. Section III demonstrates how the three-step model works to obtain the best plan for BS arrangement. Section IV summarizes the final scheme and model characteristics and presents ideas for future research.

II. BASE STATION DATA MUNGING

This paper uses a fixed area of size 2500×2500 raster as an example to construct new base stations (BSs). Information on the indicators of each raster area includes whether it is a weak coverage point, the service volume in the raster area, and the latitude and longitude coordinates (simplified to geographical horizontal and vertical coordinates).

Firstly, for the given data, careful observation was made, and it was found that the amount of weakly covered raster data needed to be more robust and conducive to the subsequent solution of the model. Fig. 1 shows the location of the 182,807 data points in the planning area.

Considering the temporal complexity of the calculation, it is decided to pre-process the data. Since the plan requires coverage of more than 90% of the services, this paper sorts all the weak coverage data in ascending order by service volume. It is found that the number of data points accounts for 70.01% of the total data when the service volume accumulates to 5%. Although these data are extensive, they have little impact on demand, and will hardly appear in the later planning coverage calculation. Therefore, we decide to eliminate this part of

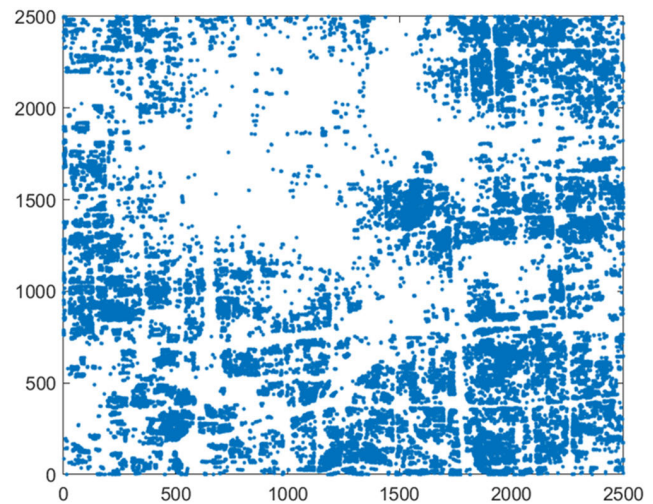


FIGURE 1. Visualization of the location of weak coverage points.

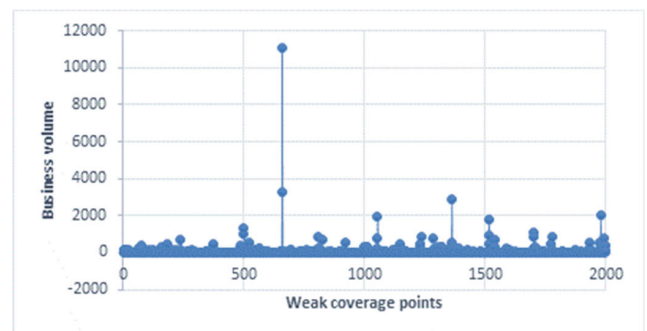


FIGURE 2. Descriptive statistics chart for business volume sampling.

the data and complete the data simplification. A random sample of 2,000 data points from the BS data source is analyzed descriptively, and a significant degree of dispersion between the data is found. Fig. 2 displays that eliminating some points with minimal service volume not only facilitates the calculation but also facilitates the subsequent clustering of the features presented, thereby ensuring the coverage of the service requirements. Thus, this processing is reasonable and scientific.

The data were then checked for outliers and missing values, and all data were found to be present and reasonable. At this point, the analysis and pre-processing of the data are completed. The locations of the 54,817 data points filtered after pre-processing the data are illustrated in Fig. 3.

III. MODELLING AND SOLVING

When investigating the demand data of Base Stations (BSs) in the selected area, three things need to be taken into consideration: latitude and longitude coordinates, signal coverage strength, and total service volume in this grid. To improve the signal coverage strength of the area, the coverage demand should reach at least 90% of the total service volume when the network BSs planning is completed. In addition, to ensure

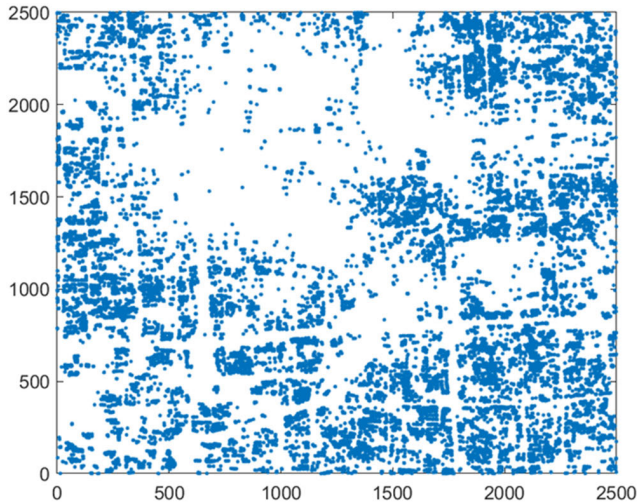


FIGURE 3. Visualization of weak coverage point locations after filtering.

signal stability, the distance between the BSs must be greater than the given threshold value. In the actual network planning, the signal coverage strength and stability, the cost of building BSs, and some other practical factors must be taken into consideration. This paper divides BSs into macro and micro BSs. In this article, the coverage area of macro BSs is 30 grid units in radius, and the cost is 10 value units. In comparison, the coverage area of micro BSs is 10 grid units in radius, and the price is 1 value unit. The following part of this section illustrates how our model works to deploy BSs.

A. STEP 1: ROUGHLY SELECTING THE ALTERNATIVE COORDINATES FOR THE BS BASED ON DBSCAN ALGORITHM

The site selection and construction of base stations (BSs) have the following characteristics:

1. The number of base stations to be constructed is still being determined. More base stations may be required in areas with high traffic density. In contrast, areas with low traffic density may require fewer base stations.

2. There is a constraint between the base station coverage radius and the coordinates of weak coverage points. If the service volume demand of two weak coverage points is satisfied by the same BS, the distance between the two weak coverage points should be less than twice the distance of the BS coverage radius (i.e., the diameter of the BS coverage area). The DBSCAN clustering algorithm has the following characteristics:

1. Compared with other clustering algorithms, the DBSCAN algorithm does not need to determine the number of clusters in advance.

2. The parameter that needs to be determined for the DBSCAN algorithm is the maximum coverage radius. The idea of this algorithm is that if the distance between two data points is less than or equal to a certain threshold, then these two points belong to the same cluster.

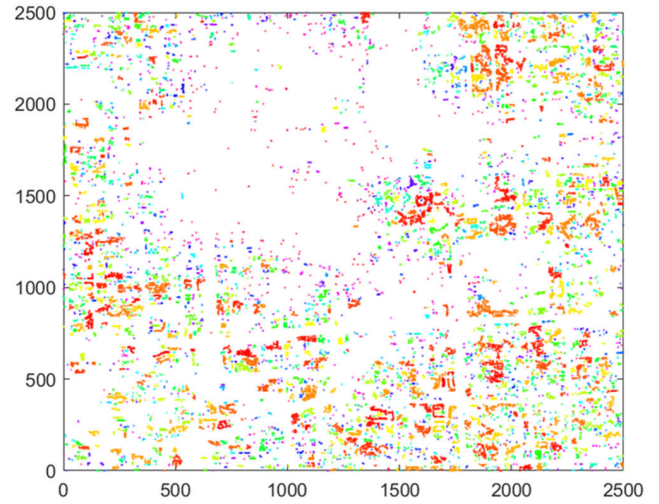


FIGURE 4. DBSCAN clustering results.

By continuously performing simulation iterations, the DBSCAN algorithm can determine the number of clusters actively and combine the coordinates of weak coverage point locations with the base station coverage radius. Therefore, we choose the DBSCAN algorithm in the initial base station construction coordinates selection.

To simplify the planning complexity, this paper first considers only micro BSs when selecting BS sites. A single macro BS is subsequently used to replace multiple similar micro BSs to save the cost. Since the coverage of the micro BSs is 10, we set ε (the radius parameter) to 10 and $MinPts$ (the neighborhood density threshold) parameter to 1, i.e., no noise points by default [28], [29]. By inputting pre-processed weak coverage raster data, preliminary clustering results can be obtained, and the results are shown in Fig. 4.

Next, the center coordinate of each cluster in the clustering result is denoted as (X_{center}, Y_{center}) for a cluster containing n data points:

$$X_{center} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

$$Y_{center} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (2)$$

where X_i represents the horizontal coordinate of the i -th point in a certain cluster, and Y_i represents the vertical coordinate of the i -th point in a certain cluster [28].

It is observed from Fig. 5 that the clustering effect is relatively significant, and the goal of clustering the weak coverage points is initially achieved. After continuous iterative calculations in Matlab, the results of some clustering centers and the number of points they contain are shown in Table 1.

Table 1 demonstrates that, after clustering, the largest cluster contains 556 coordinates, which is larger than the maximum value of the coverage (314) for a single micro BS. It can be seen in Fig. 6 that the DBSCAN clustering method

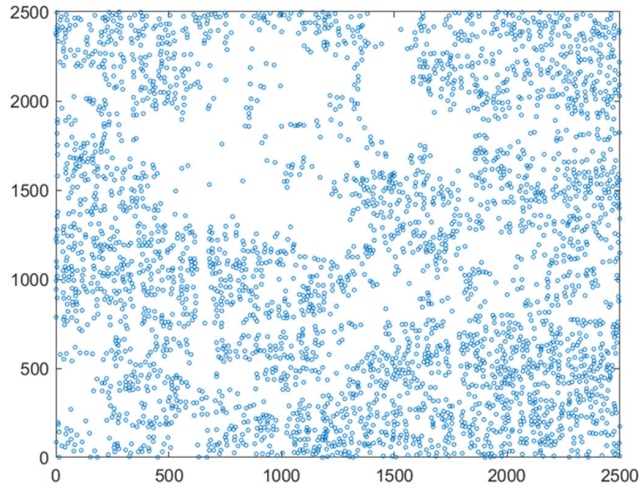


FIGURE 5. Scatter plot of DBSCAN clustering centers.

TABLE 1. Results of the DBSCAN clustering center calculation.

No.	x coordinate	y coordinate	Number of points in the cluster
1	1963	1341	556
2	1901	874	442
3	1956	2146	433
4	1098	421	402
5	1844	2432	352
6	1872	2214	337
...
3353	869	2292	1
3354	1356	2271	1
3355	844	1962	1

only allows the existence of two points whose distance is less than the given radius parameter. However, not all the points within the cluster are necessarily within a circle whose radius is the parameter (ϵ) [30]. The cluster centers generated only by the iterative DBSCAN clustering method may not be used as alternative coordinates for the initial micro BSs, as they may ignore some points and may not meet the 90% service demand requirement. Therefore, further clustering is required for some clusters with a more substantial coordinate coverage.

B. STEP 2: CORRECTING AND FURTHER REFINING THE ALTERNATIVE BS COORDINATES BASED ON K-MEANS ALGORITHM

The K-means clustering algorithm is used for the secondary clustering, which is to further separate the clusters that exceed the coverage of the micro BS. This algorithm divides the sample into K clusters and minimizes the sum of the distances between all objects in the sample space and their cluster centers [31], [32], [33]. The clusters that do not meet the predefined threshold in the initial clustering are subjected to

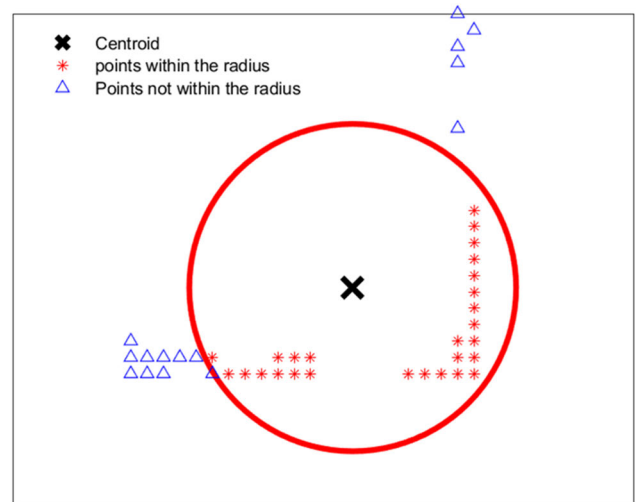


FIGURE 6. Schematic diagram of DBSCAN clustering result.

a secondary clustering analysis. The K value for each cluster is selected individually. The two furthest data distance from the x coordinate in a single cluster is called Δx_{\max} , and the two furthest data distance from the y coordinate is called Δy_{\max} . Since the micro base radius parameter ϵ is 10, the number of clusters K for a single sample space is expressed as follows.

$$K = \max\left\{\frac{\Delta x_{\max}}{10}, \frac{\Delta y_{\max}}{10}\right\} \quad (3)$$

By updating the DBSCAN clustering results with the clustering results of K-means, 5,815 clustering centers can be chosen for BS sites. If all 5,815 clustering centers are built with micro BSs, the entire service volume can be covered. Alternatively, if some of the 5,815 clustering centers are selected to establish macro BSs, some to build micro BSs, and the rest to remain BS-free, the construction cost can be significantly reduced while still reaching the planning target of a 90% service volume. The 5,815 clustered centers have been identified as potential sites for new BSs and are shown in Fig. 7 for visualization purposes.

Some summary results of the final clustering centers are given in Table 2.

C. STEP 3: OBJECTIVE OPTIMIZATION MODEL BASED ON SAA

Nowadays, the architecture of a 5G network is more complex than that of a 4G network, and the emergence of 5G microbes can effectively improve the cost-effectiveness of network BS construction and optimize the service capacity of the system. Since the above two clustering steps have ensured signal coverage and stability, the third optimization step can be performed by combining the results of the above two clustering steps with the minimum total cost as the objective function and the minimum workload as the constraint to form a planning model. The intelligent optimization based on a

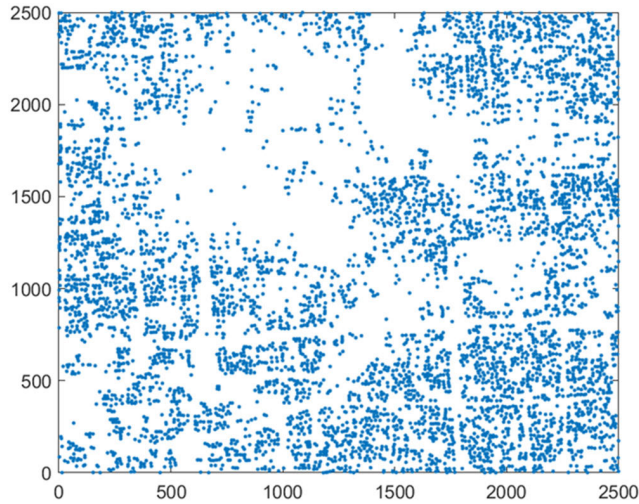


FIGURE 7. Site planning model visualization results.

TABLE 2. Summary results of the final clustering centers.

Serial number	x coordinate	y coordinate
1	444	640
2	123	1053
3	938	678
4	303	663
5	1672	768
6	711	254
...
5813	833	1011
5814	273	555
5815	660	753

simulated annealing algorithm is used to further improve the network base station siting scheme by adjusting the number of stations and macro and micro BS types and changing the station layout.

In addition, it should be emphasized that the signal coverage of the network BS construction scheme and the signal interference generated by the coexistence of macro and micro BSs are also factors and objectives that we need to consider. However, in the first two steps of clustering, the DBSCAN algorithm in the first step ensures the signal coverage, and the K-means algorithm in the second step ensures the stability of the coexisting signals of macro and micro BSs, so both of them play a simplified contribution to the overall optimization objective.

To obtain the optimal construction plan for BSs, this paper considers the selection of new BSs as the decision variable, the minimum total cost of new BSs as the objective function, and the constraint that the service volume within the coverage area is greater than 90% of the total service volume. In summary, the following model can be obtained.

Objective function:

$$\min P = num_{MicroBaseStation} + 10num_{MacroBaseStation} \quad (4)$$

TABLE 3. Initialization parameter settings for the SAA.

Parameters	Value
Initial temperature	1000
Maximum number of iterations	1000
Number of iterations at each temperature	10000
Temperature decay coefficient	0.95

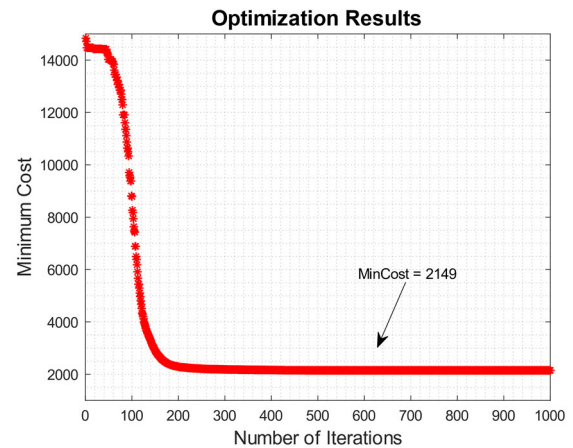


FIGURE 8. The plot of the number of iterations versus minimum cost.

Binding conditions:

$$\sum_{i=1}^n Traffic_i \geq Total(Traffic) \times 90\% \quad (5)$$

The work described in this paper chooses possible station coordinates randomly, and the range of choices is as follows.

$$\begin{cases} 0, & \text{no base stations} \\ 1, & \text{Micro base stations} \\ 2, & \text{Macro base stations} \end{cases} \quad (6)$$

At the same time, to avoid the new-generated macro BS coverage including the surrounding micro BS coverage, the alternative sites within both the macro BS and the micro BS coverage are set as the first choice by Matlab, i.e., no BS is established. According to this model, this problem can be solved using the SAA, which is a heuristic optimization algorithm that theoretically finds the optimal global solution [34], [35]. Therefore, it is employed to provide the optimal siting solution. The initialization parameters for SSA in this paper are set as shown in Table 3.

In this paper, macro BSs are established at all optional sites as the initial condition. Finally, the construction cost of these BSs converges to 2,149 after about 482 iterations. During this process, 102 macro BSs and 1129 micro BSs are established, covering 90.03% of the total service volume and achieving the target of providing service coverage to 90% service volume. The change in the cost of each iteration and the siting options for some BSs is shown in Fig. 8.

The final visualization of the BS is depicted in Fig. 9.

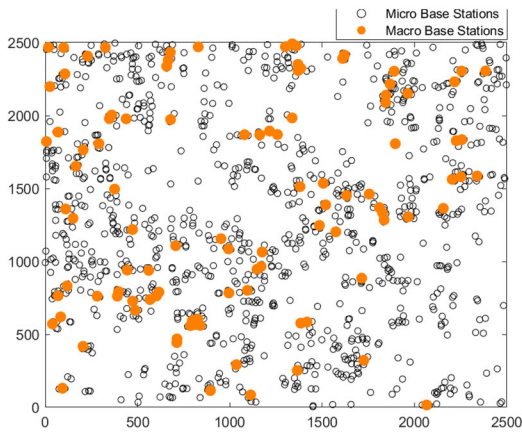


FIGURE 9. Final site plan.

TABLE 4. Comparison of different models.

Method	Number of micro BSs	Number of macro BSs	Total Cost	Coverage area
SAA[36]	7168	1566	22828	98.23%
DBSCAN[37]	1058	868	9738	/
K-means + PSO[38]	942	365	4592	91.28%
Three-step model	1129	102	2149	90.03%

IV. MODEL COMPARISON

The optimal BS construction plan and its required cost are finally obtained through the optimization calculation of the above steps. On this basis, we selected some calculation results of the same context as this problem, used the same data set for comparison, and confirmed the superiority and practicality of the three-step model. The specific results are shown in Table 4.

Compared with the optimization scheme using SAA alone, it is found that without the constraint of pre-clustering, the calculation results show more serious deviations, and the coverage areas of the BSs overlap significantly. Although the coverage areas are very substantial, the resulting cost increase is not reasonable for applying the practical problem.

The BS site planning using the DBSCAN algorithm combined with the exhaustive method is the opposite of the model solution process proposed in this paper, in which the type of BSs that can be built at each point is first determined. Then the location of the new BSs is determined by clustering using the DBSCAN algorithm. From the comparison of the results of the two planning schemes, it is easy to see that the use of the DBSCAN algorithm for site planning leads to a high total cost. Although the problem of the coverage threshold of the

BS is well solved, the traversal algorithm using only methods such as the exhaustive enumeration method will make the results not optimal, and the results have strong randomness. In contrast, the use of some combinatorial optimization algorithms and intelligent optimization algorithms can solve this problem well.

In this method, the weak coverage points to be planned are first clustered using the K-means clustering method to determine the approximate location of the BS construction points. Then PSO is used to optimize the solution of specific station sites considering the cost and benefit factors. The total cost of this method is lower than the above two methods, and it combines the ideas of clustering and optimization to solve the problem. However, it needs to fully consider the threshold of BSs and the waste of resources caused by repeated planning, while the three-step model of two-layer clustering can solve such problems very well.

Algorithm 1 DBSCAN

Input: The set of coordinates of the weakly covered points $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$; Neighborhood parameters: $(\epsilon, MinPts)$.

Process:

- 01: Initialize the core object set: $\Omega = \emptyset$
- 02: **for** $j = 1, 2, \dots, m$ **do**
- 03: Determine the ϵ -neighborhood $N_\epsilon(x_j)$ of the weak coverage point (x_j, y_j) ;
- 04: **if** $|N_\epsilon(x_j)| \geq MinPts$ **then**
- 05: Set (x_j, y_j) into $\Omega = \Omega \cup \{x_j\}$
- 06: **end if**
- 07: **end for**
- 08: Initialize the number of clusters: $k = 0$
- 09: Initialize the set of uncovered coordinates: $\Gamma = D$
- 10: **while** $\Omega \neq \emptyset$ **do**
- 11: Record the current set of unvisited weak coverage points: $\Gamma_{old} = \Gamma$;
- 12: Randomly select a core object $o \in \Omega$ and initialize the queue $Q = \langle o \rangle$;
- 13: $\Gamma = \Gamma \setminus \{o\}$;
- 14: **while** $Q \neq \emptyset$ **do**
- 15: Take the first weak coverage point q in queue Q ;
- 16: **if** $|N_\epsilon(q)| \geq MinPts$ **then**
- 17: Set $\Delta = N_\epsilon(q) \cap \Gamma$;
- 18: Take the weak coverage points in Δ to queue Q ;
- 19: $\Gamma = \Gamma \setminus \Delta$;
- 20: **end if**
- 21: **end while**
- 22: $k = k + 1$, and $C_k = \Gamma_{old} \setminus \Gamma$;
- 23: $\Omega = \Omega \setminus C_k$
- 24: **end while**

Output: All clusters $C = \{C_1, C_2, \dots, C_k\}$

V. CONCLUSION

In this paper, we present a new three-step model with practical implications for planning optimal BS construction solutions.

Algorithm 2 K-means

Input: Set coordinates after initial clustering
 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$;
and set the number of proposed sites as k .

Process:

01: Randomly select k clustered point coordinates from D as the initial site coordinate vector $\{\mu_1, \mu_2, \dots, \mu_k\}$

02: *repeat*

03: Set $C_i = \emptyset$ ($1 \leq i \leq k$)

04: *for* $j = 1, 2, \dots, m$ *do*

05: Calculate the distance between (x_j, y_j) and i -th mean vector μ_i ($1 \leq i \leq k$): $d_{ji} = \|\mathbf{x}_j - \mu_i\|_2$;

06: Determine the cluster labeling of (x_j, y_j) based on the closest mean vector:
 $\lambda_j = \operatorname{argmin}_{i \in 1, 2, \dots, k} d_{ji}$

07: Assign (x_j, y_j) to the corresponding clusters:
 $C_{\lambda_j} = C_{\lambda_j} \cup \{(x_j, y_j)\}$;

08: *end for*

09: *for* $i = 1, 2, \dots, k$ *do*

10: Calculate the new i -th mean vector as the new site coordinate vector: $\tilde{\mu}_i = 1/|C_i| \sum_{x \in C_i} \mathbf{x}$;

11: *if* $\tilde{\mu}_i \neq \mu_i$ *then*

12: Update the current site coordinate vector μ_i to $\tilde{\mu}_i$

13: *else*

14: Keep the current site coordinate vector unchanged

15: *end if*

16: *end for*

17: *until* None of the current site coordinate vectors have been updated

Output: All clusters: $C = \{C_1, C_2, \dots, C_k\}$

The model consists of initialization, correction, refining as well as optimization. The design parameters for BS placement are BS coordinates (x, y) , the collaboration of multiple types of BS, and the density of BS construction. We construct the objective function using the lowest total cost and setting the total minimum workload of BS to 90%. The real data of a 2500×2500 area are used for the simulation test and the results are reasonable, confirming the effectiveness of this method for large orders of magnitude of data and multi-types BS. This study also leads to the following conclusions:

1. Combining the DBSCAN and K-means algorithms can yield more accurate results. The DBSCAN clustering algorithm identifies more concentrated points in the alternative area and distinguishes more discrete points; however, it may lead to the merging of adjacent clusters. In contrast, the K-means clustering algorithm can refine the decomposition of connected clusters, but it is more affected by noise points.

2. The SAA can provide an optimal solution for BS siting and construction relatively quickly to minimize costs and satisfy the constraint of covering at least 90% of the service volume. 1129 micro BSs and 102 macro BSs are established

Algorithm 3 Simulated Annealing

Input: Initial solution: The initial solution to start the search;
Initial temperature: The initial temperature for simulated annealing (1000);
Max iterations: The maximum number of iterations (1000);
Iterations per temperature: The number of iterations to perform at each temperature (10000);
Temperature decay coefficient: The rate at which the temperature is reduced (0.95);
Stopping condition: When the function value descent curve becomes flat.

Process:

01: Initialization

02: Generate a random solution X and calculate its fitness value $f(X)$.

03: Set $X_{best} = X$, $k = 0$, $t = \text{initial temperature}$.

04: *while not* stopping_condition:

05: The search loop under the temperature threshold (t)

06: *for* $i = 1$ to inner loop iterations:

07: Generate a new solution X_{new} based on the current solution X

08: Generate a new solution X_{new} from X .

09: Calculate its fitness value $f(X_{new})$.

10: Check if the new solution is better

11: *if* $f(X_{new}) < f(X)$:

12: *if* $f(X_{new}) < f(X_{best})$:

13: Set $X_{best} = X$

14: *continue*

15: *end if*

16: Calculate the acceptance probability $\exp(-\Delta \text{energy}/t)$

17: Accept the new solution with a probability

18: *if* $\text{random}(0, 1) < P$:

19: Set $X = X_{new}$.

20: *end if*

21: *end for*

22: Reduce the temperature t .

23: Increment k by 0.95.

24: *end while*

Output: Best solution

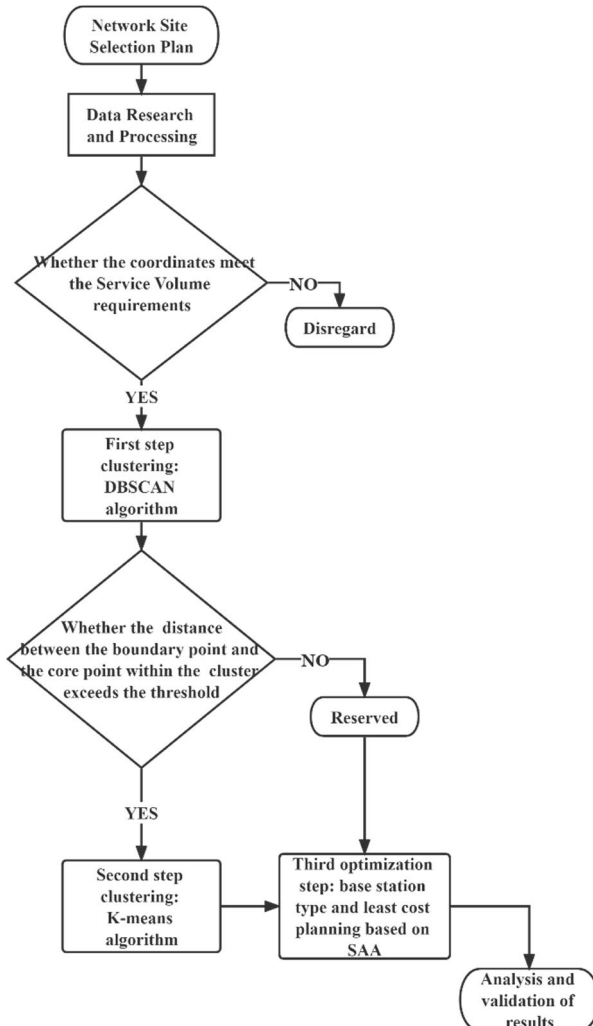
in this paper, at a total cost of 2,149, covering 90.03% of the total service volume in the area.

In the current scenario of constructing BSs, we can import the coordinates of the current area that have not yet met the demands for business, meanwhile import the types of BSs to be established (BSs with differing coverage radius) This will enable us to perform site planning for the new BSs using the method outlined in this paper. The present model requires fewer variables, leading to higher computational efficiency and the ability to handle large-scale data,

ultimately reducing operational expenses. Moreover, it allows for a rough determination of the types and coordinates of the newly constructed BSs.

Introducing clustering algorithms in station site optimization reduces spatial and temporal complexity. The SAA, using heuristic ideas, allows the model to be solved accurately. However, our experiments are based on Matlab simulations, and no field experiments are conducted, so we have neglected the height of actual geographical locations and the issue of frequency interference between signals in real-world scenarios. Also, in the algorithmic approach, the selection of initial values may affect the final alternative BS coordinates due to the multiple uses of clustering algorithms [39]. In addition, the stability of the clustering algorithm as an unsupervised pattern recognition method [40] is also worth considering [41]. Therefore, to meet this experiment's needs, we over-constrained the model. In the future, we will gradually relax certain constraints to evaluate the model, and we will also try to improve the model by randomly adding loss functions such as altitude and signal interference.

**APPENDIX A
ALGORITHM FLOW CHART**



The flowchart depicted above illustrates the entire algorithm, which is primarily employed for the problem of base station siting planning. The algorithm consists of three main components and the specifics of each step are outlined below:

- **The DBSCAN algorithm used in the first step of clustering**

Application object: coordinate points of BSs that reach a specific service volume;

Parameter settings: ϵ (the radius parameter) is 10, which is the minimum coverage radius of the BS, and *MinPts* (the neighborhood density threshold) parameter is 1 to ensure that all points are considered.

- **The K-means algorithm used in the second step of clustering**

Application object: the set of BSs whose core points and boundary points in the cluster in the clustering result in the first step exceed the BS threshold distance;

Parameter setting: K (number of clusters) is the maximum value of the horizontal and vertical distance within the selected cluster, and the formula is as follows. The 10 in the formula depends on the stability distance of the BS threshold.

- **SAA used in the third optimization step**

Application object: a planning model in which the station coordinates are used to select which BS to build as the decision variable, the least total cost of construction as the objective function, and the service volume in the coverage area is more significant than 90% of the total service volume as the constraint;

Parameter settings: The specific parameter settings of the simulated annealing algorithm are shown below. The values of these parameters are obtained by comparing multiple simulations.

Parameters	Value
Initial temperature	1000
Maximum number of iterations	1000
Number of iterations at each temperature	10000
Temperature decay coefficient	0.95

APPENDIX B

See Algorithms 1–3.

APPENDIX C

The data in this article represents the existing network coverage in a specific city area. The data has been rasterized for computational convenience, dividing the area into a 2500 × 2500 coordinate grid.

The dataset contains information about 182,808 weak coverage points within this 2500 × 2500 coordinate grid. The information includes each weak coverage point's X and Y coordinates and the corresponding traffic volume. In this area, the minimum traffic volume among the weak coverage points is 0.000192, the maximum traffic volume is 47,795.01, and the average traffic volume of the weak coverage points is 38.59.

APPENDIX D

The abbreviations of terms used in the article are as follows:

Full name	abbreviation
Base Station	BS
Simulated Annealing Algorithm	SAA
Set Covering Location Problems	SCLP
Maximal Covering Location Problems	MCLP
Unmanned Aerial Vehicle	UAV
Candidate Locations Set	CLS
Particle Swarm Optimization	PSO

APPENDIX E

The explanations of the parameters in this article are as follows:

Parameter	Explanation
ϵ in DBSCAN algorithm	neighborhood radius
$MinPts$ in DBSCAN algorithm	minimum number of points
K in K-means algorithm	Number of clustering

APPENDIX F

In the experimental process, we use a computer with Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz (12 CPUs), ~2.6GHz, NVIDIA GeForce GTX 1650 graphics card, and 16G of RAM. The processing time in the third step of the simulated annealing algorithm using the data mentioned in this paper is about 250 seconds. At the later stage of the actual program, when the data are gradually stabilized, it is possible to decrease the maximum number of iterations within a single epoch. This adjustment aims to expedite the speed of each epoch and facilitate the calculation of a substantial amount of data.

ACKNOWLEDGMENT

(Lin-Shen Yang, Bin Wen, and Jie-Jun Yan are co-first authors.)

REFERENCES

- [1] Z. Zhang, A. Sun, X. Zhang, D. Zhang, and Z. Li, "Optimal placement of base stations in falling point measurement at sea," in *Proc. 5th Int. Conf. Autom., Control Robot. Eng. (CACRE)*, Sep. 2020, pp. 468–472, doi: [10.1109/CACRE50138.2020.9230350](https://doi.org/10.1109/CACRE50138.2020.9230350).
- [2] Z. C. Zhan and L. Tong, "Precision analysis of passive location of 4-stations based on TDOA," *J. China Acad. Electron. Sci.*, vol. 5, pp. 582–585, Nov. 2010.
- [3] M. Ozasa, G. K. Tran, and K. Sakaguchi, "Research on the placement method of UAV base stations for dynamic users," in *Proc. IEEE VTS 17th Asia Pacific Wireless Commun. Symp. (APWCS)*, Aug. 2021, pp. 1–5, doi: [10.1109/APWCS50173.2021.9548763](https://doi.org/10.1109/APWCS50173.2021.9548763).
- [4] G. E. Athanasiadou, D. Zarbouti, and G. V. Tsoulos, "Automatic location of base-stations for optimum coverage and capacity planning of LTE systems," in *Proc. 8th Eur. Conf. Antennas Propag. (EuCAP)*, Apr. 2014, pp. 2077–2081, doi: [10.1109/EuCAP.2014.6902217](https://doi.org/10.1109/EuCAP.2014.6902217).
- [5] N. Lakshminarasimman, S. Baskar, A. Alphones, and M. W. Iruthayarajan, "Base station placement for dynamic traffic load using evolutionary algorithms," *Wireless Pers. Commun.*, vol. 72, no. 1, pp. 671–691, Sep. 2013, doi: [10.1007/s11277-013-1036-9](https://doi.org/10.1007/s11277-013-1036-9).
- [6] H. M. Ali, J. Liu, and W. Ejaz, "Planning capacity for 5G and beyond wireless networks by discrete fireworks algorithm with ensemble of local search methods," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, p. 185, Dec. 2020, doi: [10.1186/s13638-020-01798-y](https://doi.org/10.1186/s13638-020-01798-y).
- [7] O. Galinina, "5G multi-RAT LTE-WiFi ultra-dense small cells: Performance dynamics, architecture, and trends," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1224–1240, Jun. 2015, doi: [10.1109/JSAC.2015.2417016](https://doi.org/10.1109/JSAC.2015.2417016).
- [8] C. Toregas, R. Swain, C. ReVelle, and L. Bergman, "The location of emergency service facilities," *Oper. Res.*, vol. 19, no. 6, pp. 1363–1373, Oct. 1971, doi: [10.1287/opre.19.6.1363](https://doi.org/10.1287/opre.19.6.1363).
- [9] R. Church and C. ReVelle, "The maximal covering location problem," *Papers Regional Sci. Assoc.*, vol. 32, no. 1, pp. 101–118, Dec. 1974, doi: [10.1007/BF01942293](https://doi.org/10.1007/BF01942293).
- [10] S. Zu and L. Sun, "Research on location planning of urban charging stations and battery-swapping stations for electric vehicles," *Energy Rep.*, vol. 8, pp. 508–522, Jul. 2022, doi: [10.1016/j.egyr.2022.01.206](https://doi.org/10.1016/j.egyr.2022.01.206).
- [11] M. E. Morocho-Cayamcela, W. Lim, and M. Maier, "An optimal location strategy for multiple drone base stations in massive MIMO," *ICT Exp.*, vol. 8, no. 2, pp. 230–234, 2022, doi: [10.1016/j.ict.2021.08.010.v](https://doi.org/10.1016/j.ict.2021.08.010.v).
- [12] R. Mathar and T. Niessen, "Optimum positioning of base stations for cellular radio networks," *Wireless Netw.*, vol. 6, no. 6, pp. 421–428, 2000, doi: [10.1023/A:1019263308849](https://doi.org/10.1023/A:1019263308849).
- [13] H. M. Elkamchouchi, H. M. Elragal, and M. A. Makar, "Cellular radio network planning using particle swarm optimization," in *Proc. Nat. Radio Sci. Conf.*, Mar. 2007, pp. 1–8, doi: [10.1109/NRSC.2007.371389](https://doi.org/10.1109/NRSC.2007.371389).
- [14] O. Berman, J. Kalcsics, and D. Krass, "On covering location problems on networks with edge demand," *Comput. Operations Res.*, vol. 74, pp. 214–227, Oct. 2016, doi: [10.1016/j.cor.2015.04.005](https://doi.org/10.1016/j.cor.2015.04.005).
- [15] V. Blanco and M. Martínez-Antón, "Optimal coverage-based placement of static leak detection devices for pipeline water supply networks," 2023, *arXiv:2301.04707*.
- [16] B. T. Downs and J. D. Camm, "An exact algorithm for the maximal covering problem," *Nav. Res. Logistics (NRL)*, vol. 43, no. 3, pp. 435–461, 1996/04/01 1996, doi: [10.1002/\(SICI\)1520-6750\(199604\)43:3<435::AID-NAV8>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1520-6750(199604)43:3<435::AID-NAV8>3.0.CO;2-A).
- [17] Z. He, B. Fan, T. C. E. Cheng, S.-Y. Wang, and C.-H. Tan, "A mean-shift algorithm for large-scale planar maximal covering location problems," *Eur. J. Oper. Res.*, vol. 250, no. 1, pp. 65–76, Apr. 2016, doi: [10.1016/j.ejor.2015.09.006](https://doi.org/10.1016/j.ejor.2015.09.006).
- [18] O. M. Amine, Z. Cilia, and A. Khireddine, "Optimal base station planning using genetic algorithm for LTE network," in *Advances on Broad-Band Wireless Computing, Communication and Applications*, L. Barolli, F. Xhafa, and J. Conesa, Eds. Cham, Switzerland: Springer, 2018, pp. 731–740.
- [19] M. Haghi, O. Arslan, and G. Laporte, "A location-or-routing problem with partial and decaying coverage," *Comput. Oper. Res.*, vol. 149, Jan. 2023, Art. no. 106041, doi: [10.1016/j.cor.2022.106041](https://doi.org/10.1016/j.cor.2022.106041).
- [20] M. Arana-Jiménez, V. Blanco, and E. Fernández, "On the fuzzy maximal covering location problem," *Eur. J. Oper. Res.*, vol. 283, no. 2, pp. 692–705, Jun. 2020, doi: [10.1016/j.ejor.2019.11.036](https://doi.org/10.1016/j.ejor.2019.11.036).
- [21] M. Baldomero-Naranjo, J. Kalcsics, and A. M. Rodríguez-Chía, "Minmax regret maximal covering location problems with edge demands," *Comput. Oper. Res.*, vol. 130, Jun. 2021, Art. no. 105181, doi: [10.1016/j.cor.2020.105181](https://doi.org/10.1016/j.cor.2020.105181).
- [22] D. Tedeschi and M. Andretta, "New exact algorithms for planar maximum covering location by ellipses problems," *Eur. J. Oper. Res.*, vol. 291, no. 1, pp. 114–127, May 2021, doi: [10.1016/j.ejor.2020.09.029](https://doi.org/10.1016/j.ejor.2020.09.029).
- [23] S. A. Rodriguez, R. A. De la Fuente, and M. M. Aguayo, "A facility location and equipment emplacement technique model with expected coverage for the location of fire stations in the Concepción province, Chile," *Comput. Ind. Eng.*, vol. 147, Sep. 2020, Art. no. 106522, doi: [10.1016/j.cie.2020.106522](https://doi.org/10.1016/j.cie.2020.106522).
- [24] E. Nweli, J. Isabona, and A. L. Imoize, "Optimization of base station placement in 4G LTE broadband networks using adaptive variable length genetic algorithm," *Social Netw. Comput. Sci.*, vol. 4, no. 2, p. 121, Dec. 2022, doi: [10.1007/s42979-022-01533-y](https://doi.org/10.1007/s42979-022-01533-y).
- [25] H. Küçükaydın and N. Aras, "Gradual covering location problem with multi-type facilities considering customer preferences," *Comput. Ind. Eng.*, vol. 147, Sep. 2020, Art. no. 106577, doi: [10.1016/j.cie.2020.106577](https://doi.org/10.1016/j.cie.2020.106577).
- [26] A. M. Heyns and J. H. van Vuuren, "Multi-type, multi-zone facility location," *Geographical Anal.*, vol. 50, no. 1, pp. 3–31, Jan. 2018, doi: [10.1111/gean.12131](https://doi.org/10.1111/gean.12131).

- [27] V. Blanco, R. Gázquez, and F. Saldanha-da-Gama, "Multi-type maximal covering location problems: Hybridizing discrete and continuous problems," *Eur. J. Oper. Res.*, vol. 307, no. 3, pp. 1040–1054, Jun. 2023, doi: [10.1016/j.ejor.2022.10.037](https://doi.org/10.1016/j.ejor.2022.10.037).
- [28] N. Hanafi and H. Saadatfar, "A fast DBSCAN algorithm for big data based on efficient density calculation," *Expert Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117501, doi: [10.1016/j.eswa.2022.117501](https://doi.org/10.1016/j.eswa.2022.117501).
- [29] L. Hu, H. Liu, J. Zhang, and A. Liu, "KR-DBSCAN: A density-based clustering algorithm based on reverse nearest neighbor and influence space," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115763, doi: [10.1016/j.eswa.2021.115763](https://doi.org/10.1016/j.eswa.2021.115763).
- [30] R. Zhang, J. Qiu, M. Guo, H. Cui, and X. Chen, "An adjusting strategy after DBSCAN," *IFAC-PapersOnLine*, vol. 55, no. 3, pp. 219–222, 2022, doi: [10.1016/j.ifacol.2022.05.038](https://doi.org/10.1016/j.ifacol.2022.05.038).
- [31] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci.*, vol. 622, pp. 178–210, Apr. 2023, doi: [10.1016/j.ins.2022.11.139](https://doi.org/10.1016/j.ins.2022.11.139).
- [32] A. A. Abdunassar and L. R. Nair, "Performance analysis of kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model," *Meas., Sensors*, vol. 25, Feb. 2023, Art. no. 100666, doi: [10.1016/j.measen.2023.100666](https://doi.org/10.1016/j.measen.2023.100666).
- [33] S. Huang, Z. Kang, Z. Xu, and Q. Liu, "Robust deep k-means: An effective and simple method for data clustering," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107996, doi: [10.1016/j.patcog.2021.107996](https://doi.org/10.1016/j.patcog.2021.107996).
- [34] R. W. Eglese, "Simulated annealing: A tool for operational research," *Eur. J. Oper. Res.*, vol. 46, no. 3, pp. 271–281, 1990, doi: [10.1016/0377-2217\(90\)90001-R](https://doi.org/10.1016/0377-2217(90)90001-R).
- [35] S. Atta, P. R. S. Mahapatra, and A. Mukhopadhyay, "Solving a new variant of the capacitated maximal covering location problem with fuzzy coverage area using metaheuristic approaches," *Comput. Ind. Eng.*, vol. 170, Aug. 2022, Art. no. 108315, doi: [10.1016/j.cie.2022.108315](https://doi.org/10.1016/j.cie.2022.108315).
- [36] W. Yang and C. Sun, "Research on communication network site planning and regional clustering algorithm," in *Proc. IEEE Conf. Telecommun., Opt. Comput. Sci. (TOCS)*, Dec. 2022, pp. 913–918, doi: [10.1109/TOCS56154.2022.10016030](https://doi.org/10.1109/TOCS56154.2022.10016030).
- [37] L. Jiang, B. Huang, L. Chen, and Z. Li, "Research on location planning of 5G base station based on DBSCAN clustering algorithm," in *Proc. IEEE 2nd Int. Conf. Electr. Eng., Big Data Algorithms (EEBDA)*, Feb. 2023, pp. 669–675, doi: [10.1109/EEBDA56825.2023.10090761](https://doi.org/10.1109/EEBDA56825.2023.10090761).
- [38] H. Lin, J. Lin, J. Lin, and K. You, "Mobile communication network site planning and regional clustering," in *Proc. IEEE 2nd Int. Conf. Electr. Eng., Big Data Algorithms (EEBDA)*, Feb. 2023, pp. 792–798, doi: [10.1109/EEBDA56825.2023.10090816](https://doi.org/10.1109/EEBDA56825.2023.10090816).
- [39] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognit.*, vol. 93, pp. 95–112, Sep. 2019, doi: [10.1016/j.patcog.2019.04.014](https://doi.org/10.1016/j.patcog.2019.04.014).
- [40] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: [10.1016/j.patrec.2009.09.011](https://doi.org/10.1016/j.patrec.2009.09.011).
- [41] C. X. Gao, D. Dwyer, Y. Zhu, C. L. Smith, L. Du, K. M. Filia, J. Bayer, J. M. Menssink, T. Wang, C. Bergmeir, S. Wood, and S. M. Cotton, "An overview of clustering methods with guidelines for application in mental health research," *Psychiatry Res.*, vol. 327, Sep. 2023, Art. no. 115265, doi: [10.1016/j.psychres.2023.115265](https://doi.org/10.1016/j.psychres.2023.115265).



LIN-SHEN YANG was born in Tianjin, China, in 2002. He is currently pursuing the B.E. degree in artificial intelligence with Tiangong University, Tianjin. His main research interests include pattern recognition and deep learning.



BIN WEN was born in Hubei, China, in 2002. He is currently pursuing the B.E. degree in artificial intelligence with Tiangong University, Tianjin, China. His main research interests include machine learning and data mining.



JIE-JUN YAN was born in Jiangsu, China, in 2001. He is currently pursuing the B.E. degree in artificial intelligence with Tiangong University, Tianjin, China. His main research interests include data clustering and regression calculations, modeling estimation, and data prediction.

• • •