

5G base station positioning model based on optimization algorithm

Yibo Yang¹, Yuxuan Han¹, Xinyu Zhang¹, Haonan Zhang², Jingjing Yi^{3,*}

¹College of electrical engineering, Northeast Electric Power University, Jilin, China

²School of software, Henan University of Science and Technology, Henan, China

³School of Energy and Power Engineering, XI'AN University of Architecture and Technology, Xian, China

*Corresponding author: yi1903020213@163.com

Abstract. Based on the rapid development of 5G networks, the wider the bandwidth, the more limited the coverage. The problem of site selection is becoming more and more prominent. According to the coverage of the existing 5G network, this problem gives the weak coverage area of the existing network. It requires selecting a certain number of locations so that after the new base stations are built at these points, the weak coverage of the existing network can be solved. Network coverage issues in the coverage area. This paper establishes a related multi-objective programming model, and a heuristic algorithm is used to solve it. We first cleaned the data, judged the Euclidean distance between the existing base station and the weak coverage point according to the required minimum threshold of 10, and eliminated the points that did not meet the requirements. Due to the vast amount of data, the business volume was huge. Small dots are blurred. Set the coverage area of the base station as a sector. By adjusting the relationship between the calculation coverage radius and the angle, the angle between the three main directions is added as a constraint, and the planning is carried out again based on problem one location selection. The angle change is used as the decision variable, and the greedy algorithm is used to solve the problem. Moreover finally, the coordinates of each base station and the optimal sector angle that satisfy more than 90% of the traffic volume are: (232.6573, -111.668, -29.7508). Using the k-means clustering algorithm, the weak coverage points are clustered according to the requirements of less than 20, and the frequency and percentage of clusters are analyzed. The time complexity is calculated by testing the DVI index and the number of iterations. On this basis, Optimize our model. Finally, the evaluation and promotion of our model are carried out, and the detailed distribution map and clustering effect map of various base stations are given in the appendix.

Keywords: Multi-objective optimization model, greedy algorithm, K-Means clustering.

1. Introduction

With the rapid development of 5G modernization, the scale of mobile communication has become larger and larger, and the difficulty of network coverage has become more and more difficult^[1-3]. Based on the more significant the communication bandwidth, the smaller the coverage area, so more base stations are needed to cover the original size of the area. In addition, the type of base station is also related to the coverage area, which means that the selection of the type and location of the base station becomes more and more complicated. Based on this, it is necessary for areas with weak network coverage to consider the cost of covering new base stations in the area to solve the problem of weak network coverage.

Under the premise that the base station does not consider the influence of height, terrain, and topography, and the signal is not affected by other factors, it is considered that the angle between the main directions of any two sectors of each station cannot be less than 45 degrees. Other conditions in, under the conditions of the optimal site and sector angle, the newly built station covers 90% of the total traffic volume of weak coverage points, and the results of optimal site and sector angle are given^[4-6].

The distance between two weak coverage points is not greater than 20. These two weak coverage points should be clustered into a class, and considering the transitivity of the clustering property, that is, if point A and point B are of a class, point B and point C are of the same kind, then points A, B, and C are all of the same kind. Clustering all weak coverage points requires the total time complexity of the clustering method to be as low as possible^[7-10].

2. Method

We quantitatively analyze and compare the supply characteristics of each supplier, and determine the most important suppliers on this basis. Established "0-1 planning model" and "overall planning model", from the perspective of enterprises and suppliers, the relevant data were analyzed from the overall and individual aspects respectively, and the relevant conclusions were obtained.

2.1. Data Preprocessing

First, according to the two-norm formula, the coordinates of the existing weak coverage points are cleaned, and according to the formula Euclidean distance:

$$\|P - P_0\|_2 = \sqrt{(x - x_0)^2 + (y - y_0)^2} \leq 10 \tag{1}$$

Clean up the coordinate points whose threshold distance from the existing base station and the weak coverage point is less than 10.

Considering each weak coverage point's business volume and computational efficiency, we screened out the weak coverage points with the minimum traffic volume (less than 1), removed and cleaned them, and deleted 56925 pieces of data from 182807 pieces of data.

The data visualization after cleaning is shown in the following figure 1 and 2:

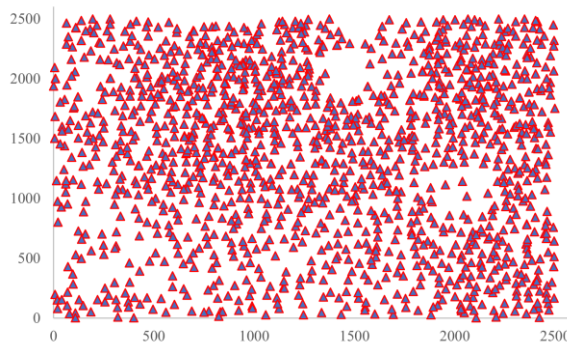


Fig. 1 Weak coverage point visualization

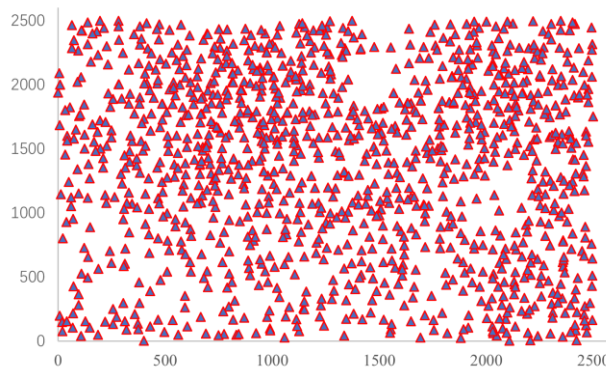


Fig. 2 Data cleaning and visualization comparison

In order to conveniently and intuitively display the relationship between traffic volume and weak coverage points, a 2500-square heat map is drawn.

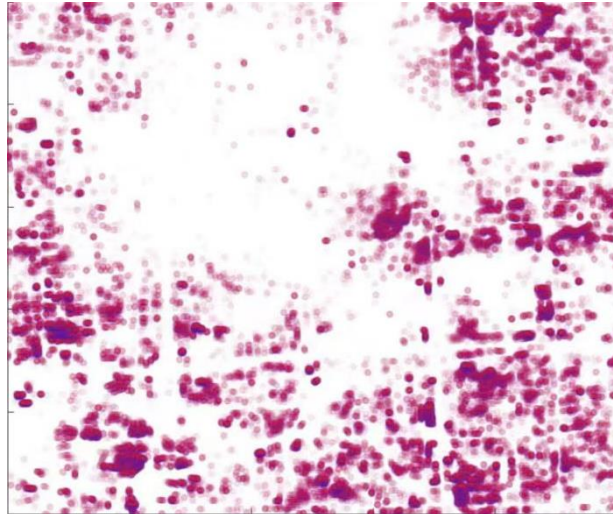


Fig. 3 Business volume heat map

2.2. Consider Sector Range

Further consideration, in practice, each station is not entirely circular coverage, but there are three sectors on each station, each sector pointing in one direction. Each sector has the most extensive coverage in the main direction (30 for the macro base station and 10 for the micro base station), which can be covered within 60 degrees left and right of the main direction, and the coverage is gradually reduced linearly. At 60 degrees, the coverage is Half of the coverage in the main direction. If it exceeds 60 degrees, it cannot be covered by the sector, so we analyze the coverage of different angles, and we establish a polar coordinate system as shown in the following figure 4:

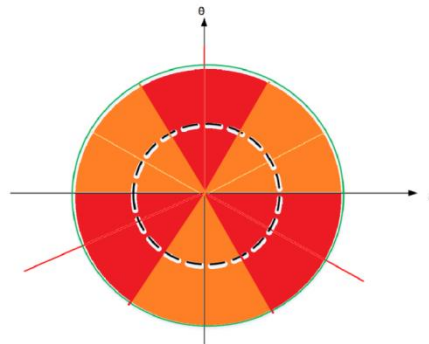


Fig. 4 Business volume heat map

Next, we establish a $\rho - \theta$ polar coordinate system, taking the horizontal axis as 0° the horizontal axis. From this question, we can see that the range of the main direction angle of the sector area is: $(-360^\circ, 360^\circ)$, it is required that the angle between any two main directions should be less than 45° , we assume that the radius in the unit circle is R , the angle between the main direction and the horizontal axis is θ , then the radius formula is:

$$R = k\theta + t \tag{2}$$

Since at 60 degrees the coverage is half of the coverage in the main direction, we can get:

$$\begin{aligned} \frac{R}{2} &= \frac{\pi}{6}k + t \\ R &= \frac{\pi}{2}k + t \end{aligned} \tag{3}$$

It can be solved by calculation, which can be seen in figure 5:

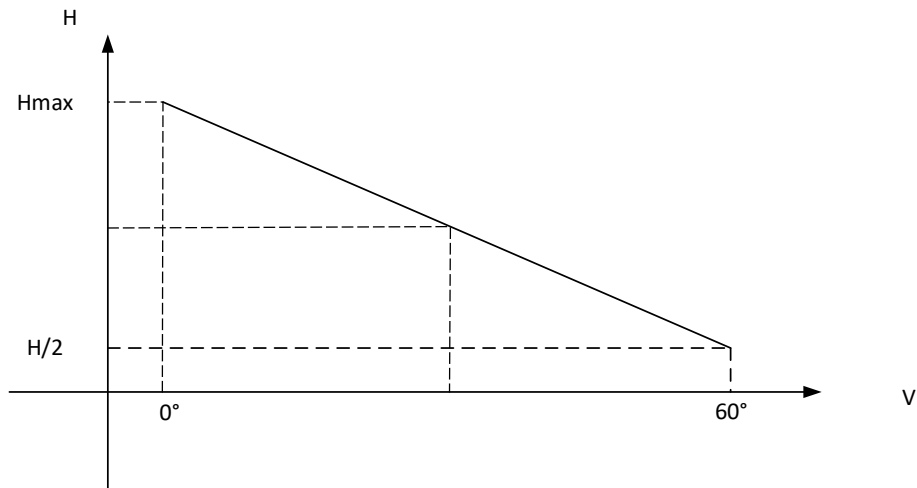


Fig. 5 Changes in angle and coverage

Therefore, we take the macro base station numbered one as an example to calculate and solve the angle change between the macro base station and the weak coverage point within the threshold and coverage constraints. The visual analysis is shown in the following figure 6:

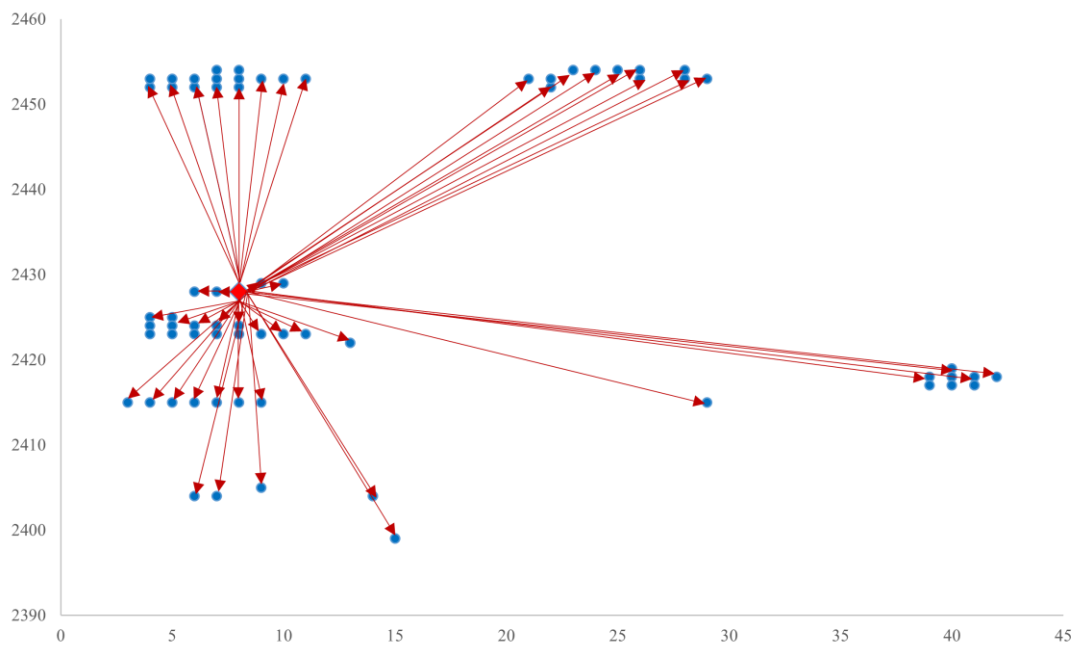


Fig. 6 The angle between macro base station 1 and the corresponding weak coverage point

The x-coordinate of weak coverage point 1, the y-coordinate and category of weak coverage point 1 are shown in Table 1:

Table 1. Angle and category between macro base station 1 and the corresponding weak coverage point

The x-coordinate of weak coverage point 1	y coordinate of weak coverage point 1	θ	category
10	2429	0.463648	1
9	2429	0.785398	2
11	2423	-1.03038	2
13	2422	-0.87606	2
10	2423	-1.19029	3
7	2428	0	2
9	2423	-1.3734	1
8	2424	-	1
8	2423	-	3
6	2428	0	1
7	2424	1.325818	1
7	2423	1.373401	1
6	2424	1.107149	1
6	2423	1.19029	1
5	2425	0.785398	1
5	2424	0.927295	1
5	2423	1.030377	1
4	2425	0.643501	1
4	2424	0.785398	1
4	2423	0.896055	2
9	2415	-1.49402	1
8	2415	-	1
7	2415	1.494024	1
6	2415	1.418147	1
5	2415	1.343997	1
4	2415	1.272297	1
3	2415	1.203622	3
29	2415	-0.55431	2
9	2405	-1.52735	2
14	2404	-1.32582	1
8	2452	#DIV/0!	1
7	2404	1.529154	2
7	2452	-1.52915	1
11	2453	1.451367	1
6	2404	1.487655	2
6	2452	-1.48766	1
22	2452	1.042722	1
10	2453	1.490966	1
9	2453	1.530818	2
5	2452	-1.44644	1
8	2453	-	2
7	2453	-1.53082	1
21	2453	1.091277	2
4	2452	-1.40565	2
6	2453	-1.49097	1
22	2453	1.060308	2
5	2453	-1.45137	1
8	2454	-	2
4	2453	-1.41214	2
7	2454	-1.53235	3
39	2418	-0.31204	3
39	2417	-0.34098	1
23	2454	1.047518	3
40	2419	-0.27417	1
26	2453	0.946773	1
24	2454	1.019141	3
40	2418	-0.30288	1
25	2454	0.991722	3
40	2417	-0.3311	1
26	2454	0.965252	1
28	2453	0.896055	3
41	2418	-0.29423	2
15	2399	-1.33395	1
29	2453	0.872137	3
41	2417	-0.32175	1
28	2454	0.915101	3
42	2418	-0.28605	2

In the end, the angles of the three center lines we obtained are: 184.358°, -248.164°, -49.377°

Similarly, for all other base stations, we have established a planning model and solved it using a greedy algorithm.

Finally, we solve the optimal main direction angles of the sector as: 232.6573, -111.668, -29.7508. The corresponding service volume ratio is: 94.323.

3. Consider the situation of weakness

If the distance between the two weakly covered points is not greater than 20, then the two weakly covered points should be clustered into one class, and the clustering property is considered transitive, that is, if the point and the point are one class, the point and the point are one class, then the points, and are of one class, so we choose the K-means clustering algorithm to cluster the weakly covered points and calculate the time complexity.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

3.1. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Step1. Selection of K value

The choice of K is generally determined by the actual situation, or the value is directly given when the algorithm is implemented.

The number of centroids is given by the user, and is recorded as the final number of clusters. The number of centroids in each update is equal to the initial value, and the number of clusters in the final cluster is equal to the number of centroids specified by the user. One centroid corresponds to a cluster, each sample is clustered into only one cluster.

Step2. Distance metric

Dividing the object points into the cluster closest to the cluster center requires the nearest neighbor measurement strategy, which is judged by the Euclidean distance in the Euclidean space.

Step3. Calculation of new centroid

For the clusters generated after classification, the points with the smallest mean distance from other points in the cluster are calculated as the centroid (for clusters with coordinates, the mean of the coordinates of each cluster can be calculated as the centroid).

Step4. Whether to stop K-means

The centroid no longer changes, or the loop Limit is given the maximum number of loops. When the centroid of each cluster does not change, k-means can be stopped; when the number of loops exceeds the loop Limit, k-means can be stopped; only one of the two conditions needs to be satisfied, k-means can be stopped. If Step4 does not end k-means, then execute step2-step3-step4; if Step4 ends k-means, then print (or plot) clusters and centroids. Select the number of clusters k to be 20 for analysis, and use Matlab to solve the results shown in the figure 7:

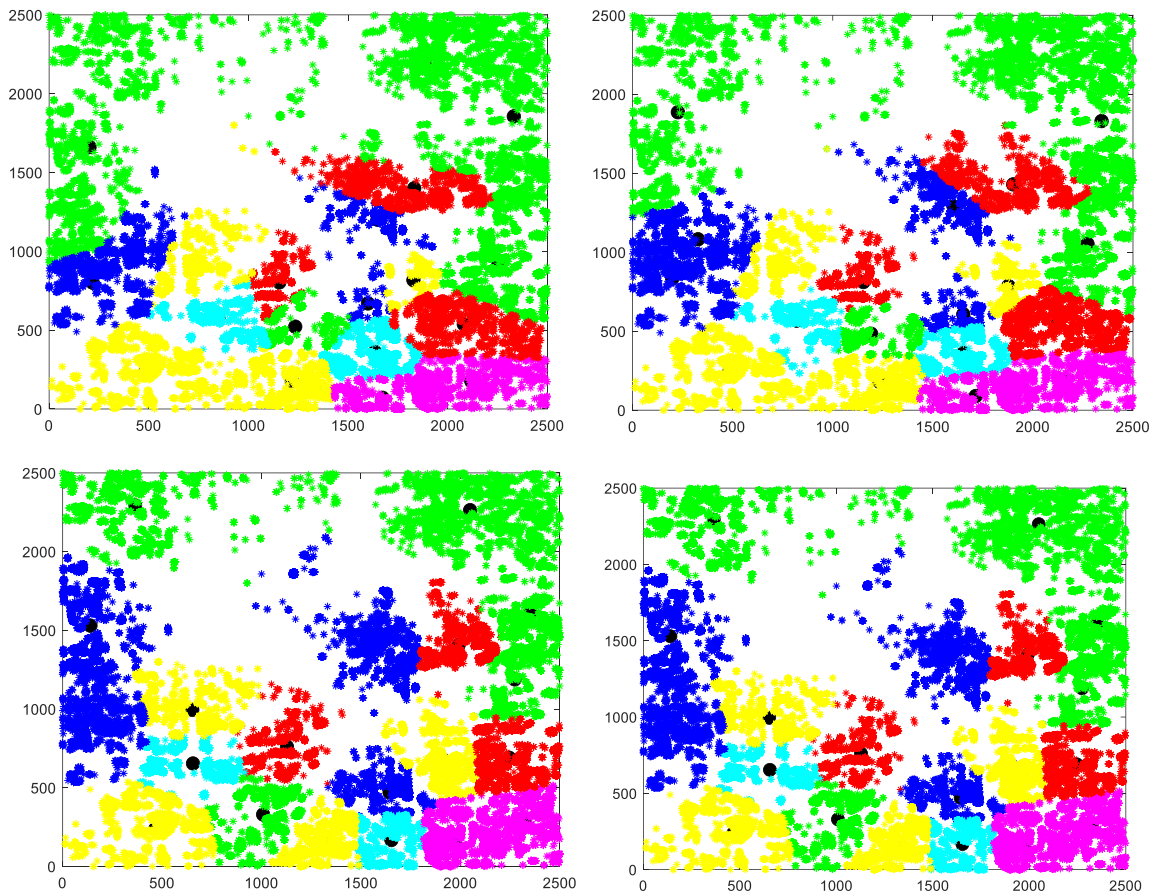


Fig. 7 k-means clustering results

3.2. Evaluation of K-means Clustering

The Davidson-Bouldin Index (DBI), also known as the classification fitness index, is proposed by David Davis and Donald-Bouldin to evaluate the pros and cons of clustering algorithms. First, assume that we have a time series clustered. The time series is set as the input matrix, and the classes are set as the parameters passed into the algorithm. Calculate using the following formula:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max \left(\frac{\bar{S}_i + \bar{S}_j}{\|w_i - w_j\|_2} \right) \quad (4)$$

1. Calculate S_i

The S_i variable is first defined in the DBI calculation formula. S_i calculates the average distance from the data within the class to the cluster centroid, representing the dispersion degree of each time series in cluster class i . The calculation formula is: where X_j represents the j th in the cluster class i . data points, that is, a time series, A_i is the centroid of cluster I , T_i is the number of data in cluster I , and p is usually 2, so that the Euclidean distance between independent data points and centroids can be calculated (euclidean metric), of course, when examining manifold and high-dimensional data, Euclidean distance may not be the best distance calculation method, but it is also typical.

2. Calculate M_{ij}

After the numerator sum is calculated, the denominator M_{ij} needs to be calculated, which is defined as the distance between cluster class I and cluster class j .

3. Calculate R_{ij}

After calculating the numerator and denominator, DBI defines a value R_{ij} to measure the similarity.

4. Calculate DBI

With the basis of the above formula, we do a nested loop of n^2 based on the number of cluster classes n , and calculate the R_{ij} of the maximum value for each cluster class i , denoted as D_i , that is, the

cluster class i and other classes. The maximum similarity value of D_i , that is, the worst result is taken out. Then take the average of the maximum similarity of all classes to get the DBI index, and the calculation formula is:

$$DBI = \frac{1}{n} \sum_{i=1}^N D_i \quad (5)$$

Finally, we get that the clustering method has high accuracy and feasibility and dramatically reduces the total time complexity.

4. Conclusions

4.1. Advantages of the model

1. Based on the analysis of the polar coordinate system, each base station's optimal central direction angle is specifically analyzed.

2. The traditional greedy algorithm is used to ensure that the global optimum is expected to be achieved in the case of the local optimum.

3. Using the k-means clustering algorithm, a higher-precision clustering effect is obtained, and the total time complexity is reduced.

4.2. Disadvantages of the model

1. May converge to a local minimum and converge slowly on large-scale data

2. Too much reliance on heuristic algorithms makes it difficult to perceive optimal local processing.

3. The algorithm design process is more complicated. Moreover, it is challenging to use nesting intuitively.

References

- [1] Hassan, Najmul, Kok-Lim Alvin Yau, and Celimuge Wu. "Edge computing in 5G: A review." *IEEE Access* 7 (2019): 127276-127289.
- [2] Ahmad, Wan Siti Halimatul Munirah Wan, et al. "5G technology: Towards dynamic spectrum sharing using cognitive radio networks." *IEEE Access* 8 (2020): 14460-14488.
- [3] Shafi, Mansoor, et al. "5G: A tutorial overview of standards, trials, challenges, deployment, and practice." *IEEE journal on selected areas in communications* 35.6 (2017): 1201-1221.
- [4] Marler, R. Timothy, and Jasbir S. Arora. "Survey of multi-objective optimization methods for engineering." *Structural and multidisciplinary optimization* 26.6 (2004): 369-395.
- [5] Asokan, P., et al. "Development of multi-objective optimization models for electrochemical machining process." *The International Journal of Advanced Manufacturing Technology* 39.1 (2008): 55-63.
- [6] Ahluwalia, Poonam Khanijo, and Arvind K. Nema. "A life cycle based multi-objective optimization model for the management of computer waste." *Resources, Conservation and Recycling* 51.4 (2007): 792-826.
- [7] Feo, Thomas A., and Mauricio GC Resende. "Greedy randomized adaptive search procedures." *Journal of global optimization* 6.2 (1995): 109-133.
- [8] Caruana, Rich, and Dayne Freitag. "Greedy attribute selection." *Machine Learning Proceedings 1994*. Morgan Kaufmann, 1994. 28-36.
- [9] Krishna, K., and M. Narasimha Murty. "Genetic K-means algorithm." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 29.3 (1999): 433-439.
- [10] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the royal statistical society. series c (applied statistics)* 28.1 (1979): 100-108.