

Optimization on Fire Station Location Selection for Fire Emergency Vehicles Using K-means Algorithm

Yunqian Wang^{1,a}

¹University of Science and Technology of China, Hefei, China

^aSA615457@mail.ustc.edu.cn

Keywords: Fire Station Planning, K-means, Shortest Path.

Abstract. Fire accident happens frequently with high density of population and large scale of production in cities and it will directly or indirectly cause incalculable loss. This paper evaluates the existing city fire fighting plan with historical fire alert call-out data, using the Baidu Map API to get the fire rescue path length, and then the K-means algorithm and its improved version are adopted to obtain a reasonable fire station layout according to the old ignition position data. This paper uses Baidu map API to get the shortest time, and uses K-means algorithm to screen potential fire station location, and reduces the amount of calculation. The data analysis and simulation experiment are carried out in this paper. The fire data sets in Suzhou were screened and useful data were obtained. The visualization of data is displayed on the map, and some conclusions of fire planning are obtained by intuitionistic method. The location of the fire station determined by various methods is compared with the experimental results, and the location of the fire is verified by the location of the alarm in the following month, which verifies the conclusion.

1. Introduction

In the early twentieth Century, the study of modern geographic location originated from the Weber problem. In 1964, Hakimi proposed the P- median problem and the P- central problem. At present, the location problem has become a complicated and comprehensive business decision process. It is widely used in the location of logistics centers, and even the retail industry is also known as *industrial location*, such as large e-commerce platforms Taobao and Jingdong, which have applied geographical location clustering to build large warehouse and improve logistics efficiency, and the improvement is obvious. The location model and algorithm are the focus of the research, which can effectively carry out the central location. The center location method is widely applied in the center location, numerical analysis, cluster analysis, linear programming and heuristic algorithm. Among them, clustering algorithm has the characteristics of scalability, processing different types of attributes and processing high-dimensional data, so it is very competitive in location algorithm. K- means clustering algorithm is the most famous clustering algorithm. Its characteristics are simplicity and high efficiency, making it a widely applied algorithm in all clustering algorithms.

2. Preliminary

In the study of facility location and vehicle routing problem (VRP), it is usually possible to use the Euclidean distance between the fire and the fire station as the data base for the shortest path problem. Although the Euclidean distance between the nodes is positive correlation to the actual fire engine travel path. In reality, there are also the intention driving, road bending and other operations, or the congestion in rush hours and the speed limit of the road, so the ideal and actual conditions differentiate greatly. And the result time is usually small and it is difficult to be used in actual application. In this paper, we call the Baidu map API to get the actual distance and actual fire engine running time between the fire station and fire. The Baidu map Web service provides a HTTP/HTTPS interface for developers, and developers can start search requests in the form of HTTP/HTTPS, and then retrieve data from the returned JSON or XML format. Users can develop applications such as JavaScript, C#, C++, Python, Java and other languages based on this.

In order to simplify the notation of horizontal axis in later figures, we use label instead of real name of fire station, shown in Table 1.

Table 1. Fire Station Label List

Label	Address
1	Suzhou Public Security Fire Tai Lake Resort full-time team
2	Suzhou Public Security Fire Detachment North Bridge full-time Team
3	Suzhou Public Security Fire Detachment vehicle workshop full-time team
4	Suzhou Public Security Fire Detachment East Zhu Squadron
5	Suzhou Public Security Fire Detachment Dushu Lake full-time team
6	Suzhou Public Security Fire Detachment Horizontal Pond Squadron
7	Suzhou Public Security Fire Detachment Lake East full-time team
8	Suzhou Public Security Fire Detachment Tiger Palace Road Squadron
9	Suzhou Public Security Fire Detachment Hu Guan Industry Park Full-time Team
10	Suzhou Public Security Fire Brigade, Hu Guan Professional team
11	Suzhou Public Security Fire Brigade yellow Dai full-time team
12	Suzhou Public Security Fire Detachment Jinchang Secret Service Squadron
13	Suzhou Public Security Fire Detachment Jinji Lake Squadron
14	Suzhou Public Security Fire Detachment Development Zone Squadron
15	Suzhou Public Security Fire Brigade Li Kou Squadron
16	Suzhou Public Security Fire Detachment Lou Feng professional team
17	Suzhou Public Security Fire Detachment, Mu du Squadron
18	Suzhou Public Security Fire Detachment Green Sword Lake Full-time team
19	Suzhou Public Security Fire Detachment Sand Lake Allied Team
20	Suzhou Public Security Fire brigade wins Pu full-time team
21	Suzhou Public Security Fire Detachment, Shishan Squadron
22	Suzhou Public Security Fire Detachment Suwan Squadron
23	Suzhou Public Security Fire Detachment Su Zhan Road Squadron
24	Suzhou Public Security Fire Detachment Special Service Two Squadron
25	Suzhou Public Security Fire Brigade Special Service Three Squadron
26	Suzhou Public Security Fire Brigade Special Service Squadron
27	Suzhou Public Security Fire Detachment Lookout Pavilion Full-time team
28	Suzhou Public Security Fire Brigade only pavilion full-time team
29	Suzhou Public Security Fire Detachment Logistics Park Professional team
30	Suzhou Public Security Fire Detachment Causal Lane Squadron
31	Suzhou Public Security Fire Detachment, Chang Qiao Squadron

3. Location selection by using clustering algorithm

3.1 K-means algorithm

Machine learning can be divided into unsupervised learning and supervised learning. Unlike supervised learning, classification or grouping information represented a category of data doesn't exist in clustering. Clustering is a technique for discovering the inner structure, so it belongs to unsupervised learning. Its workflow is to assign the objects with similar property in dataset to the same cluster, elements in the same cluster are similar, but different from the elements in other clusters. This section describes K-Means algorithm, which is the most widely used in clustering. It is mature, understandable and highly efficient.

The principle of the algorithm is to find K clusters from the given dataset. The specific work flow is as follows:

- (1) Find k elements in data sets as the k centers;
- (2) For each of the remaining elements, find the nearest center, assigns it to the cluster the nearest center belonged.
- (3) For each cluster, calculate the new mean value of all elements in the cluster after last iteration, set the new mean as the cluster center. Then continue iterating until the distribution is stable.

The pseudocode is represented as following:

Input: K: Number of clusters

D: Data set contained n objects

Output: Collection of K clusters

Algorithm *K-means (k, D)*

choose k data points as the initial centroids(cluster centers)

repeat

for each data point $x \in D$ do

compute the distance from x to each centroid;

assign x to the closest centroid // a centroid represents a cluster

end for

re-computer the centroid using the current cluster memberships

until *the stopping criterion is met*

Among them, the so-called *the most similar center*, that is, to calculate a certain distance. This distance can be a measure of any distance, such as length, distance, and time.

Each iteration of the algorithm calculates and stores the error. The distance between the current fire and the center of the cluster is treated as error, then the clustering effect is evaluated by the error and the call-out time of each fire station. This derives the SSE (Sum of squared Error), which is measurement of clustering effect. The closer the point is to the center of each cluster, the smaller SSE value is, and the better the result of clustering is. In order to take care those points away from the center, the error is squared. If we increase the number of clusters, it can lower the SSE. The goal of clustering is to improve the quality of clusters under the stable number of clusters.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist(c_i, x)^2 \quad (1)$$

For fires described in this paper, they are grouped to the nearest fire station. Not only clarify the relationship of responsibility, but also can improve fire efficiency and reduce the average time of call-out and fire extinguishing. The K-means algorithm will be affected by the distance calculation method and the selection of initial centroid.

3.2 The improvement of K-means clustering

To apply the K-means algorithm in fire accident situation, we consider some methods to extend the advantage and make up the defect.

3.2.1 Change randomly selected position to original fire station position as the initial centroid

For the evaluation of Fire station planning, the original fire station has been running for many years, the actual experience has proved that the system is still running normally, so the original fire station position is actually reasonable to be initial value of K-means algorithm. The original values, not being evaluated comprehensively, are still more reliable than the randomly initial positions. The K-means algorithm is very sensitive to the initial position selection and the amount of calculation is huge. Therefore, the reasonable selection can not only reduce the compute, shorten the running time, but also optimize the result.

3.2.2 Change the distance measurement of K-means clustering

The dataset's characteristic is the geographical location coordinate data set. It is reasonable to use the Euclidean distance as the measurement. Actually it is more accurate to choose the fire engine's traveling time from the center to the ignition position in the cluster. In order to calculate the traveling time, the program unceasingly calls the Baidu map API to obtain travel time regardless of road condition. It is easy to replace the distance calculation method by traveling time in code. That is, after the initial centroid is selected, the point of the shortest traveling time to an initial centroid is classified as a cluster, then we use the Euclidean distance to determine the next cluster center, repeating the work until the end of the iteration, which results in a more accurate and reasonable fire station arrangement.

Traveling time is a more accurate measurement than Euclidean distance, but the traveling time cannot be used to calculate the center of the cluster. Once the position changed the shortest traveling time of all points in the cluster change and the center would be indeterminate. So we still choose Euclidean distance in the calculation of the new cluster center.

According to the actual clustering process, the center of the cluster moves almost in a fixed range after each iteration, and the adjacent data points only change between several fixed clusters. For example, the north of the fire point is almost impossible to divide into the south of the cluster, and the east of the fire point is unlikely to be divided into west clusters. So, each time the all points are iterated to calculate Euclidean distance or traveling time, 90% compute is unnecessary. Therefore, there is improvement strategy: For each iteration, calculate the points in the range of twice times of former radius to the resulting cluster center. For this study, each iteration, the number of measurement calculation will not exceed 1/3 of the former amount of calculation. Although the number of iterations will be increased, but the overall compute is reduced.

4. Data analysis

4.1 Preliminary screening of data

In order to facilitate the analysis, the data were excluded suburban areas and some of the obvious wrong data before the study, taking into account the impact of urban traffic. The bar chart of 31 fire stations answering alert, is shown in figure.

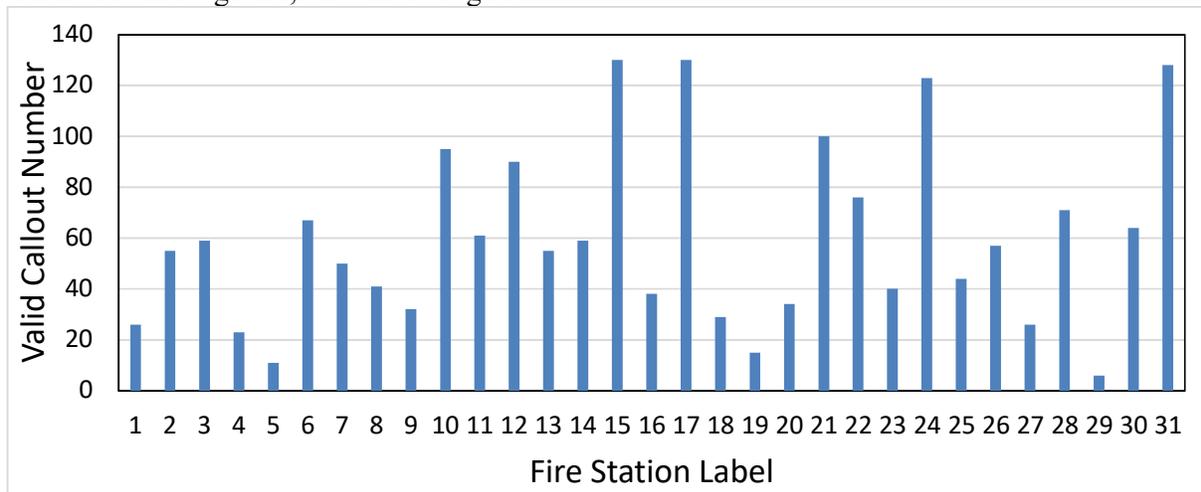


Fig. 1. Valid data

4.2 A comparison of call-out time

Fast call-out is an important basis for the completion of the fire task, and the call-out process needs to be completed in 45 Seconds to, according to national standards. In addition to the response time of the alarm information, the rapid route planning is also an important factor affecting the efficiency of the fire police. Therefore, the maximum call-out time, the average call-out time and the shortest call-out time of each fire brigade were counted to be compared. The longest call-out time is highly uncertain, and some is up to a few hours, which can be deleted in case it is wrong data. The longest call-out time is nearly 20 minutes, and some of the station's longest call-out time exceeded 30 minutes, shown in Fig. 2, Fig.3.

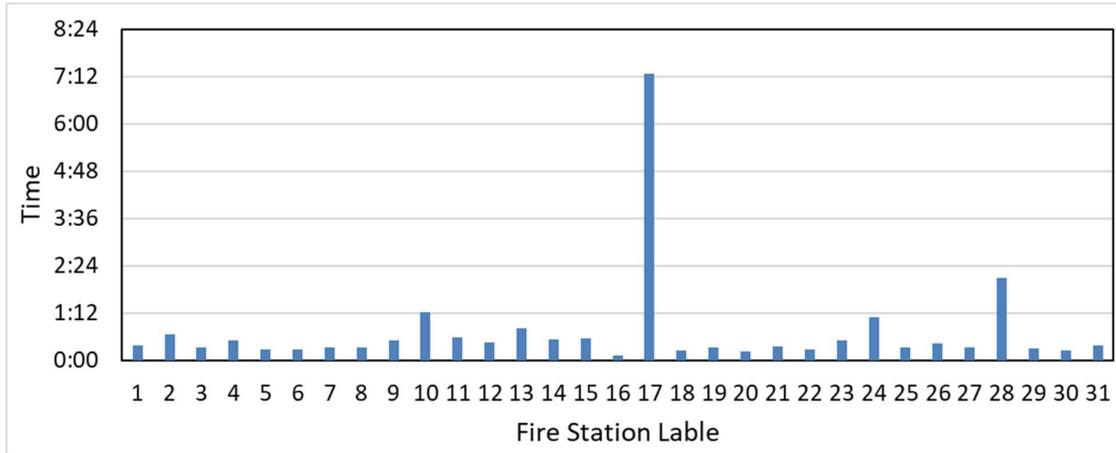


Fig. 2. The maximum call-out time for each fire brigade

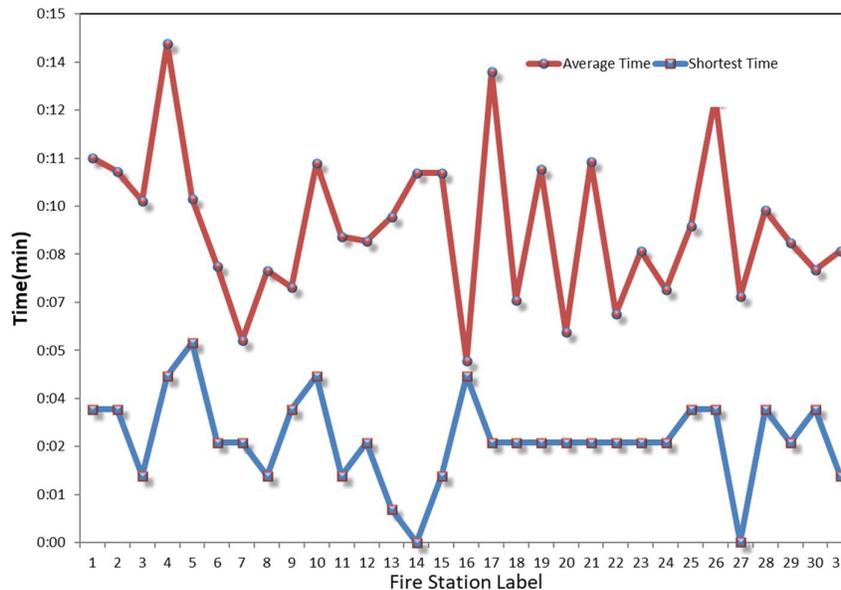


Fig. 3. Comparison of the average call-out time and the shortest call-out time of each fire brigade

According to the provisions of the "Standards of Fire Station Construction in City", the 14th entry tells that the area of fire station should be designed according to the following principles:

Fire station in city are designed on two classes, the first class of fire station should be not greater than 7 square kilometers, and the second class of fire station should be not greater than 4 square kilometers. The small fire station should be not greater than 2 square kilometers, and the common station located near the suburbs should be not greater than 15 square kilometers. The area of the fire station can also be determined according to the city's fire risk.

According to statistics published by the Ministry of Housing and Urban and rural construction in 2015, area of Suzhou city is 458.29 Sq. km. Suzhou has 92 fire brigades. If the fire brigade is in charge of the city's firefighting mission, each fire brigade should take care of 4.98 square kilometers on average. However, the 31 fire brigades in main city of Suzhou are responded for area of far beyond 400 square kilometers, thus the average area for each fire brigade in the urban area is much more than 7 square kilometers.

However, by analyzing the actual data, the average fire time exceeds 5 minutes, and fluctuates around 10 minutes. Therefore, Suzhou main urban area fire station layout and dispatch still have very big promotion space.

4.3 Determining Location Relationships

By determining one-to-one corresponding relations of the physical location of the fire and the fire brigade, and intuitively analyzing some cases, we find out whether there is unreasonable place, and show the data on the map by different time division. Unreasonable data points are mainly as follows: Sometimes, the fire is far away from its fire station. It is also possible that fire station in the east side of the city responds the fire located in the west side of the city. Maybe the fire station is closer to the expressway, or the fire station closer to the west side is less convenient for local roads or traffic jams than a fire station in the east side, or the West Side fire brigade is focusing on other fires, and cannot spare more force. The reason is the unreasonable dispatch. By conducting SUMO Simulation or adopting Dijkstra algorithm can determine the reasonable ignition position.

Some of the ignition positions are in the waters, and some are on the highway. The ignition position in the water area is generally unreasonable, because there is no alert from water area after check. The fires on highway should be taken care specifically because there are some alarm positions on the highway, but the description is not accurate, the point may be on the map on the contrary, but not on the highway exactly. These ignition positions have been corrected by hand.

In fact, the exact coordinates of the points are roughly acquired by crawler tool, the high-German map API, Baidu map API, and Xgeocoding tool. And all kinds of map database and various software tools are using different coordinate system, and they cause relatively high error. Another reason for the larger error is that alarm location description in the original data set is not very detailed, resulting in a large deviation in coordinate data.

For these deviations, by comparing its position in each system, then determine the location of the deviation, and after the manual rectification, and deletion of some uncertain location of the fire, the map of the ignition positions and fire stations are shown in the following color graph by calling map interface, shown in Fig.4.

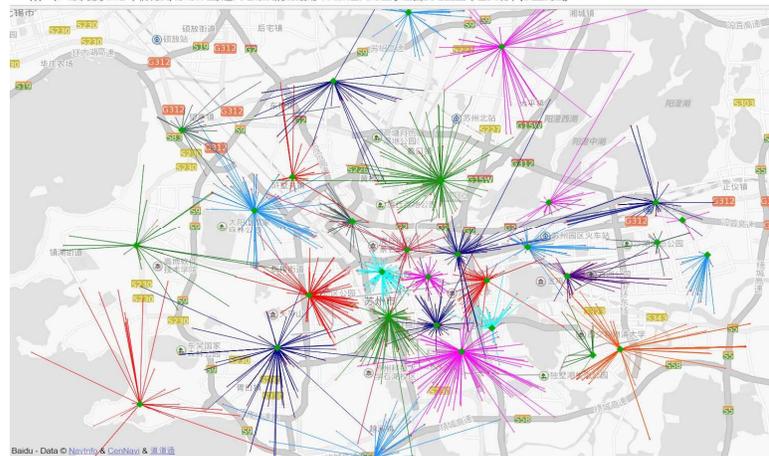


Fig. 4. The corresponding relationship between fire point and fire brigade

5 Experiment

This experiment is conducted by coding with HTML and Python with IDE named Pycharm in Windows10 Operating system.

5.1 Distance as measured by Euclidean distance

The closer the fire station locates to the possible fire, the quicker fire engine can arrive. The first experiment uses Euclidean distance as a measure of K-means algorithm, and we set the K value 31, which represents 31 fire stations. The initial value is chosen randomly by computer, and then the average arrival time of each fire station to the fire is calculated to evaluate whether the original results

are optimized. After running the program, we get that the number of iterations is up to 58 times, the mean square error is 62.5042068068, the distance calculation count is 2013690.

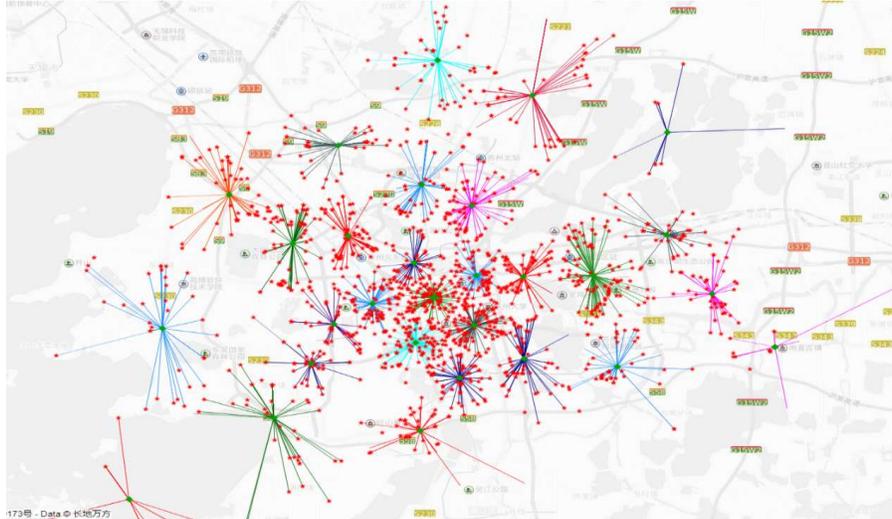


Fig. 5. Comparison of average fire time between fire stations

Fig.5 shows the visualized result. Green Diamond is for the fire station location, and red Pentagram is for the point of fire, and different colors of the line shows for the fire station coverage of the area. And Compared with the original fire Station fire-fighting diagram, the result of the cluster calculation has a clear boundary of the scope of coverage.

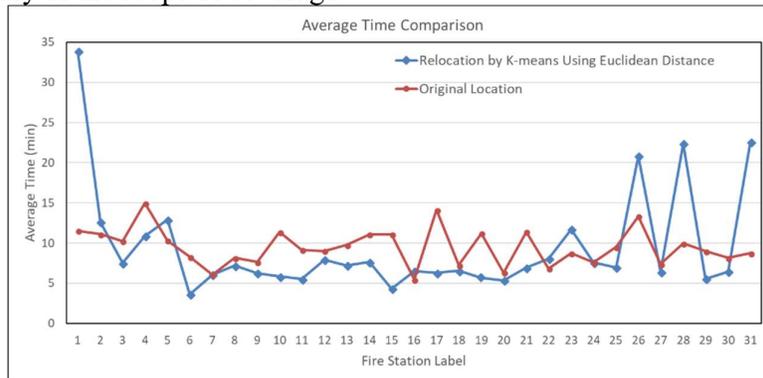


Fig. 6. Comparison of average fire time between fire stations

Except several fire stations in the suburbs, the whole call-out time has decreased markedly, especially the station in the densely populated downtown area, call-out time was reduced 34.6%, and is close to 5 minutes, shown Fig.6.

5.2 K-means algorithm with fixed original centroid

For the evaluation of Fire station planning, the original fire station has been running for many years, the actual experience has proved that the system is still running normally, so the original fire station position is reasonable to be initial value of K-means algorithm. There is little difference between the new position and the original one, the algorithm only needs 17 iterations to complete all tasks, the number of calculation is reduced to 98 millions. The result is shown in Fig.7.

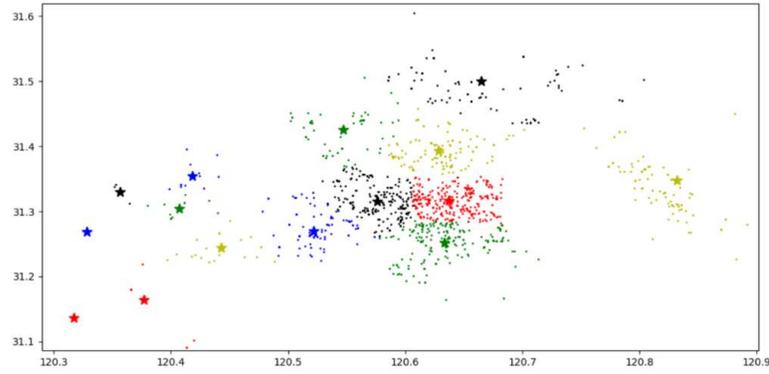


Fig. 7. Using K-Scatter graph (initial value fixed) obtained by means of mean value algorithm

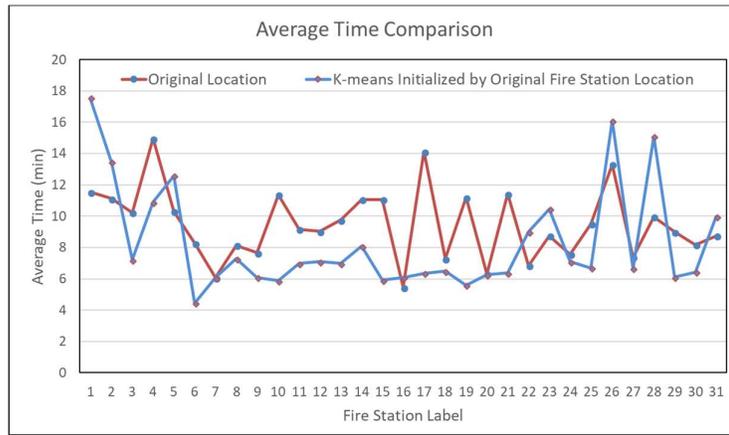


Fig. 8. Average time comparison with K-means initialized by original Fire Station Location

Thus, as Fig.8 shows the average time of call-out is significantly less than that of the original fire station location. The selection of initial position of K-means algorithm is very important. Reasonable initial position selection can reduce the execution time of the program and improve the efficiency in running.

5.3 Take the actual driving time as a measurement in K-means algorithm

We used the Euclidean distance as measurement in K-means algorithm, but the real situation is driving time of fire engine. Therefore, the driving time between fire and fire station is the measurement of K-means algorithm. Baidu Map API offers the function to get the driving time in Suzhou city.

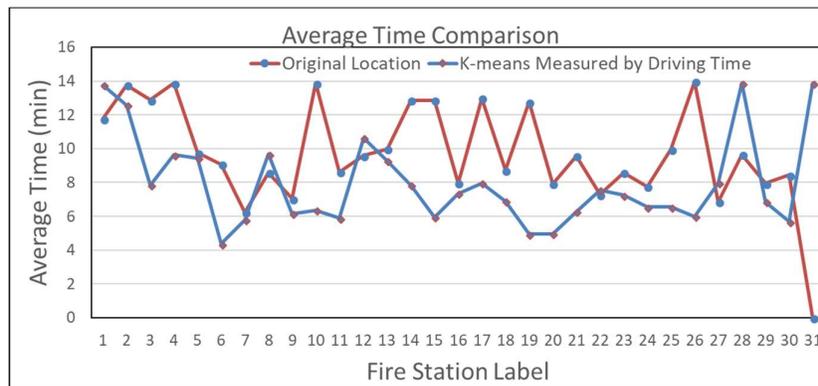


Fig. 9. Average time comparison with K-means measured by driving time

From the Fig.9, K-means clustering measured by driving time showed improvement compared with original data.

5.4 Bisecting K-means clustering

From the above experiments, K-means clustering has achieved good results, but the K-means algorithm has a problem is to converge to the local minimum value, so for the facility location problem, the bisecting k-means algorithm can solve this problem better. The algorithm first considers all the points as a class. The class is then divided into two. Then choose one of them to continue dividing. The basis of the selection is to minimize the SSE value. The Division process is iterative until the specified K value is reached.

The advantage of the bisecting K-means algorithm is that it is fast and can be considered as an improved K-means algorithm, and the initial value selection has less influence on the algorithm. Fig.10 and Fig.11 show the result.

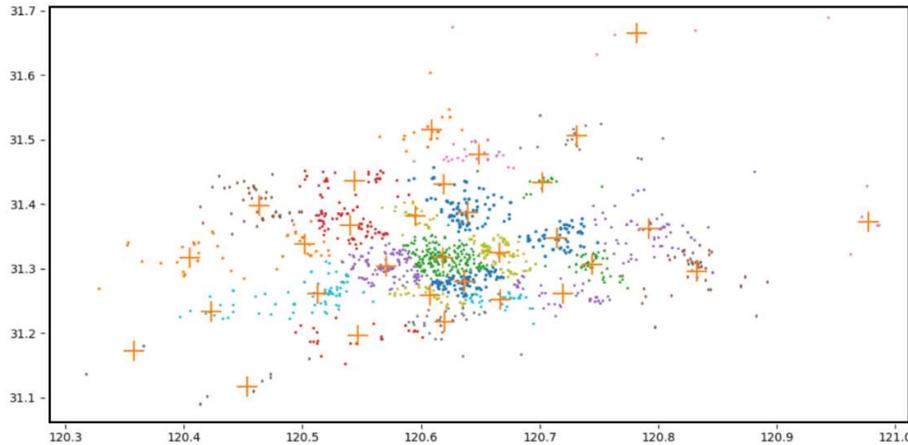


Fig. 10. Scattering Diagram using bisecting K-means algorithm (initial value fixed)

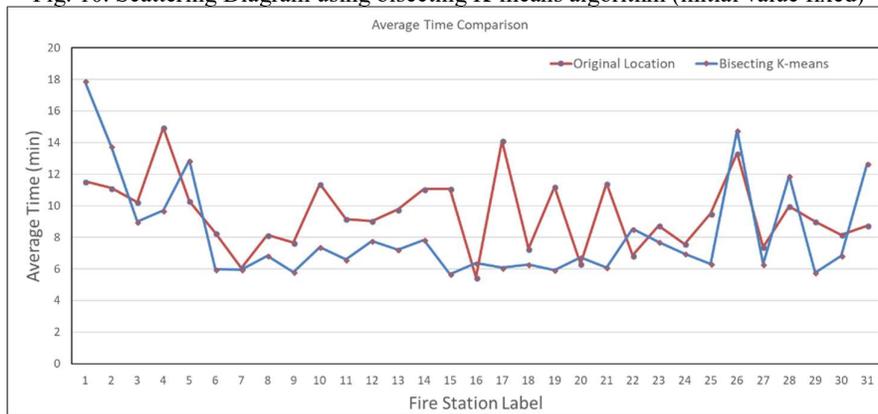


Fig. 11. Average time comparison with bisecting K-means algorithm (initial value fixed)

5.5 Comparison of several planning methods

After conduct all these kind of K-means algorithm, put all call-out time line together and we can have the comparison and find out that changed version of K-means showed excellent improvement in relocating fire station(Fig.12 shown).

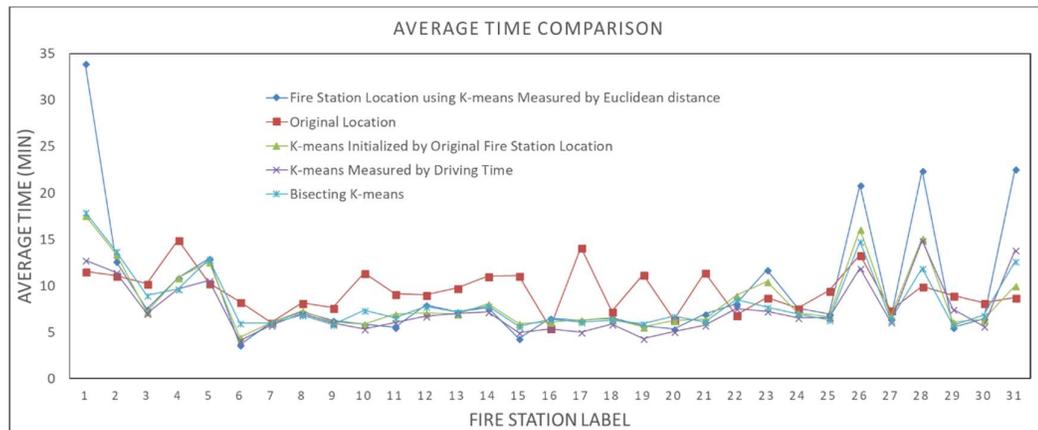


Fig. 12. Comparison of call-out time between different planning methods

6. Summary

In this paper, we propose to rearrange the fire station location in Suzhou city by several version of K-means algorithms by using real ignition data of Suzhou in recent years. We consider the fire engine traveling time as the measurement and compared with original call-out time data. The experiment showed the efficiency of planning method we proposed. As a future work, we plan to study the traffic situation and rearrange the fire stations according to the tendency of traffic situation.

Acknowledgement

This research was supported by the Suzhou Public Security in China and USTC.

References

- [1] M.W.Berry.Survey of Text Mining : Clustering, Classification, and Retrieval. New York : Springer, 2003
- [2] Anil K J. Data clustering : 50 years beyond K-Means[J]. *Pattern Recognition Letters*, 2010, 31 (8) : 651-666.
- [3] Aloise D, Deshpande A, Hansen P, et al. NP-hardness of Euclidean sum-of-squares clustering[J].*Machine Learning*, 2009, 75 (2) : 245-248
- [4] Roy D K, Sharma L K. Genetic K-means Clustering Algorithm for Mixed Numeric and Categorical Data Sets[J].*International Journal of Artificial Intelligence & Applications*,2010,1(2):23-28
- [5] Babu G P, Murty N M.A. Near-optimal Initial Seed Selection in K-means Algorithm Using a Genetic Algorithm[J].*Pattern Recognition Letters*,2003,14(10):763-769
- [6] David J. Hand, Veronica Vinciotti. Choosing K for L. Lu, A one-vendor Unbalanced classes[A]. *Pattern Recognition Letters* ,2003:1555-1562
- [7] Plllhoi Yoo, P Xiaohua Hu. A comprehensive comparison study of document clustering for a biomedical digital library. *Proceedings of the 6th ACM/IEEE-CS joint conference* . 2006
- [8] Silva J D A, Hruschka E R . Extending K-means Based Algorithms for Evolving Data Streams with Variable Number of Clusters. *International Conference on Machine Learning and Applications and Workshops (ICMLA)* . 2011

- [9] Z G WANG, Y S WONG, M. Rahman. Development of a Parallel Optimization Method Based on Genetic Simulated Annealing Algorithm. *Parallel Computing* . 2005
- [10] WANG XM, NIU RQ. Spatial Forecast of Landslides in Three Gorges Based on Spatial Data mining. *Sensors* . 2009
- [11] Li Y, Zheng Y, Ji S, et al. Location selection for ambulance stations: a data-driven approach[C]//Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2015: 85.