International Journal of Hybrid Information Technology
Vol.8, No.3 (2015), pp.387-394
http://dx.doi.org/10.14257/ijhit.2015.8.3.35

Study on the Logistics Center Location Problem Based on CURE

Yu Yi [1] and Liu Xin-liang[1]

1 Beijing Technology and Business University, Computer and Information Engineering, Beijing 100048
E-mail: yuyi727@sina.cn

Abstract

According to the cluster analysis of the logistics demand points, it can help logistics conveniently select reasonable logistics distribution center, then to reduce their transportation costs, and improve transport efficiency. This thesis firstly introduces the CURE algorithm and UFLP model, then, proposes an improved algorithm based on CURE, and describes the specific process. Finally, it also proves the algorithm's effectiveness and superiority for choosing Logistics Centre based on the numerical experiments.

Keywords: *CURE, logistics location, UFLP*

1. Introduction

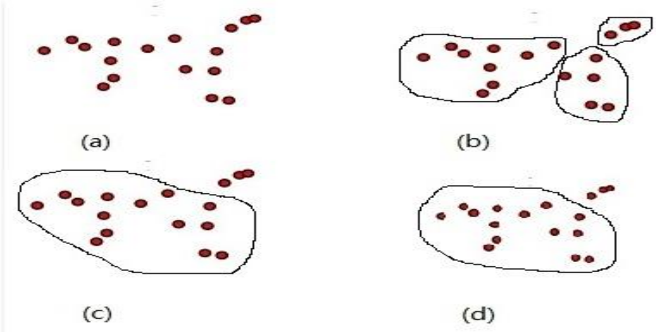
In recent years, with the rise of online shopping and economic globalization, logistics activities were frequently active in people's lives. They promoted the change of the mode of enterprise production and improved the core competitiveness. The location of logistics center is closely related to the development of the logistics industry. Traditional logistics center location is generally balanced on geographic location, transportation costs, storage capacity, the balance between customer demand and supply, response time and service efficiency factors for qualitative and quantitative analysis. It is based on the analysis of gravity model analysis method and factor score method. In the process of site selection, the methods are also different because of the different key factors. Transportation cost is the primary factor to be considered in the establishment of the entire logistics network, because it is related to the logistics company's overall revenue. In this paper, it introduces an improved CURE clustering algorithm and it mainly analyzes the logistics center location selection problem under the premise of considering transportation cost.

2. CURE

Logistics center location is a very complicated problem. In different situations, it can be analyzed and researched on different algorithms. And clustering algorithm is one of the typical representatives among the different algorithms. CURE (Clustering Using Representatives) algorithm is a kind of clustering algorithm based on hierarchy. Hierarchical method is to merge or split of sample collection until to meet the specific condition. The clustering result is to make the sample as a clustering tree. Clustering is a process of dividing data object into classes or clusters, the different objects in the same cluster have high similarity, then the different objects in different clusters are different [1]. In the clustering method, the K-means method which is proposed by McQueen [2] is the classic method to solve the problem of clustering, although its algorithm is simple and fast calculation speed, it is sensitive to noise and outlier data [3], and cannot find the cluster which has difference in big size. And Guha's CURE algorithm [4] is not using a single center or object to represent a cluster, but using a fixed number of representative points in data space to represent a class that can identify different classes of the complex

ISSN: 1738-9968 IJHIT

shapes and sizes, and can be well filtered isolated points. But the algorithm have some problems on the choice of merging points, if the merging decision is not good, it may decline the quality of clustering. Figure 1 shows the clustering process of the CURE algorithm. In the algorithm there are some representative points for the cluster in each step. First of all, Figure 1 (a) lists the sample data, each sample point acts a separate cluster. In Figure 1 (b), the sample data has been gathered into three clusters, each cluster has two representative points. The selected points should be chosen as far as possible away from another. In Figure 1 (c), two close clusters have been combined, and two new representative points are selected. Finally, these points in Figure 1 (d) those points have been shrank to the average direction of cluster, and then the final clustering result comes out. If a cluster centroid was chosen as a representative point, the smaller cluster will be merged with the underlying cluster, rather than the upper cluster.



(a)Sample data (b) Clustering
(c)Merging clusters (d) Shrinkage

Figure 1. CURE Clustering Process

3. UFLP Logistics Location Model

In the process of research on logistics center Location, problems should be modeled, and Uncapacitated Facility Location Problem, referred UFLP is one of the most classic models [5]. UFLP is widely applied in study of logistics location problem, in UFLP models, known demand point sets D = {1, 2... , m},$\forall j \in D$, $(\bar{x}_i, \bar{y}_i)$ and q_j represent the needs of the point j coordinates location and demand, the pre-center for logistics facilities sets F = {1, 2... , p},$\forall i \in F$,$(\bar{x}_i, \bar{y}_i)$and f_i are respectively the location of the center i and construction cost. The distance between demand point and logistics center, is calculated by the Euclidean distance.

$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ The costs from the j-th demand point to the i-th shipping logistics center is c_{ij}. UFLP model aims to meet the needs of each point, and at the same time, minimize the total cost. And this model is exactly proceed from two key parts, construction costs and transportation costs.

The first one is the cost to build a logistics center. $C(F) = \sum_{i=1}^{p} f_i y_i$, $y_i \in \{0, 1\}$. If the logistics center is built in the location of i, $y_i = 1$; Otherwise $y_i = 0$. Total transportation cost, $T(D) = \sum_{j=1}^{m} \sum_{i=1}^{p} c_{ij} x_{ij}$, according to the above assumptions, UFLP model is showed as below:

$$\min \sum_{i \in F} \sum_{j \in D} c_{ij} x_{ij} + \sum_{i \in F} f_i y_i$$

$$\text{s.t.} \quad \sum_{i \in F} x_{ij} = 1, \ \forall j \in D \qquad (1)$$

$$x_{ij} - y_i \leq 0, \ \forall i \in F, \ \forall j \in D \qquad (2)$$

$$x_{ij} \in \{0, 1\}, \ \forall i \in F, \ \forall j \in D \qquad (3)$$

$$y_i \in \{0, 1\}, \ \forall i \in F \qquad (4)$$

In the model, (1) shows that each demand point has only one central point to provide service, (2) lists that any demand point can only have one open logistics center to provide services, (3) and (4)explains that restrictions x_{ij} and y_i turn to 0-1 variables.

In UFLP model, the key point is that the cost calculation for the demand points and logistics centers' transportation costc_{ij}, the traditional calculation generally carries out in accordance with the radial point-to-point distance. And it's just considering from local so it's failed to make the overall optimized. In the following part the thesis will be calculated by the improved CURE algorithm and verified its feasibility and superiority.

4. Improved CURE Algorithm

CURE algorithm sees each data object as a separate clusters. In the process of the merger each step will choose the nearest cluster, but the algorithm's complexity is high, so it isn't conducive to handle large data, in addition if it's not chosen reasonable representative points what has a great influence on the result of clustering.

Aiming at shortcoming of CURE algorithm this paper divides the improved algorithm into two phases. The first stage is to pre cluster to those sample data. In the pre-clustering process, it will use K-means algorithm with low complexity algorithms and faster. The sample data is divided into n classes, n is bigger than the ultimate expectation of cluster number k. Specific steps are as follows:

Step1 It will be randomly selected n samples from S data points as initial cluster centers;

Step2 according to the mean of each object clustering, it calculates the Euclidean distance between each object and the center object. And re-divides the corresponding object according to smallest cluster;

Step3 re-calculate the mean of each cluster, and determine the new cluster center;

Step4 cycle and Step2 Step3, until the standard measurement function began to converge, and there is no data elements being re-allocated.

The second stage is used CURE algorithm to handle the preprocessed class.

Step1 select the representative point set of the new cluster after preliminary clustering, and choose the point which is furthest from cluster center as the first representative point. Then, select a point which is farthest from the first representative point as the second representative point.

Step2 find the nearest two clusters then combine them;

Step3 re-select a new cluster representative set of points, shrinking representative point set, and eliminate effects that most of the outlier make on cluster;

Step4 cycle Step2 and Step3 until the clustering number is k, and remove out the obvious less number of objects.

The improved algorithm flow chart is showed as follows:

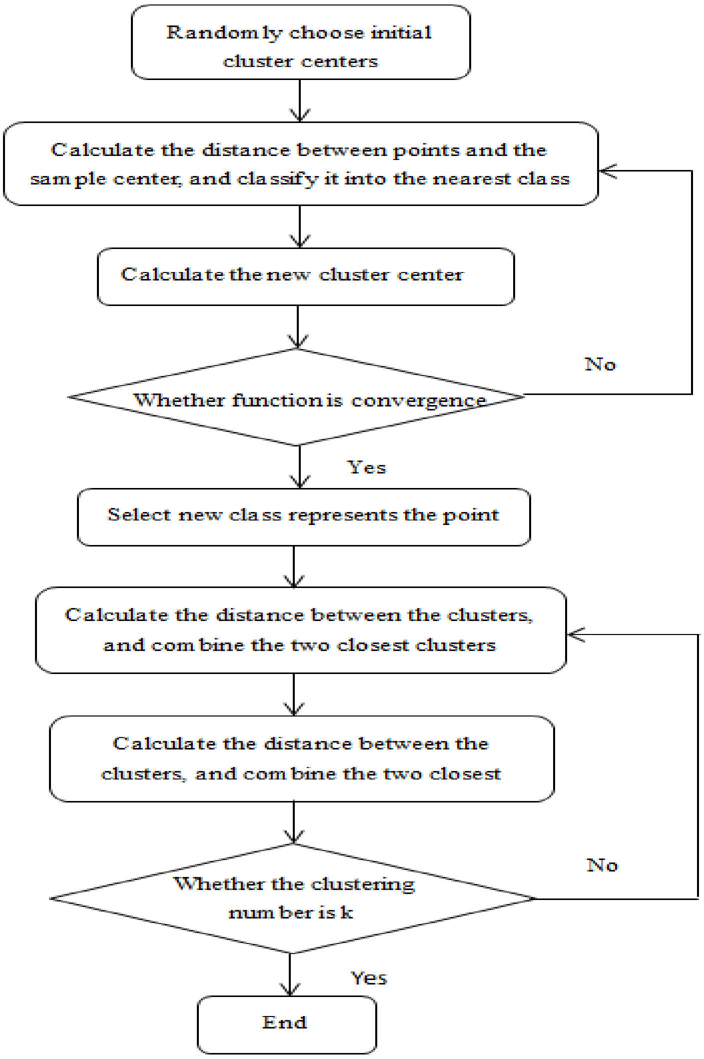


Figure 2. Flow Chart of Improved CURE Algorithm

5. The Experiment and its Results

In the following part the logistics center location will be shown by numerical simulation tests to verify the superiority of the algorithm. It is using the Python language to build a three-dimensional coordinate system of 100 * 100 * 100, and randomly generated 200 points as demand points set, just as Figure 3. The (x, y) shows that point's location, it is derived from 3D map projection in the XY plane, as in Figure 4. It is used for distance calculation. z shows that the result getting in the influence of other factors, such as construction of logistics center. Assumptions in these 200 points need to build 10 logistics center.

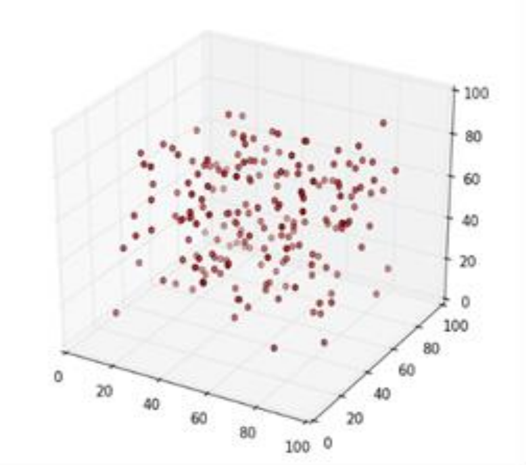


Figure 3. Randomly Generate Three-dimensional Figure Point set

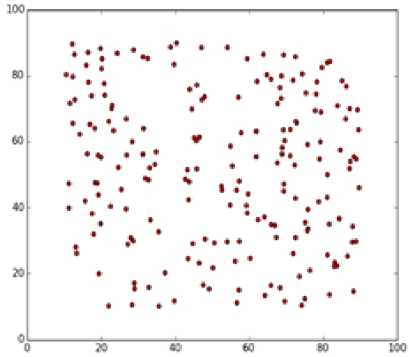


Figure 4. XY Plane Projection Points in the Chart

First, put the 200 points into UFLP model. In the LINGO11.0 software it calculates according to the radial line, then obtains the optimal solution. Like the Table 1 below and it is represented by square points in Figure 5, the optimal value of the objective function is 6892.5.

Table 1. General Approach for Optimal Solution Set Table

X	Y	Z
78.0	77.6	76.0
54.8	8.2	28.8
47.3	92.5	99.8
32.2	15.6	60.8
59.2	50.1	47.3
84.8	37.1	91.3
18.1	71.1	55.7
89.3	86.6	70.4
44.9	69.6	10.0
25.2	30.5	14.1

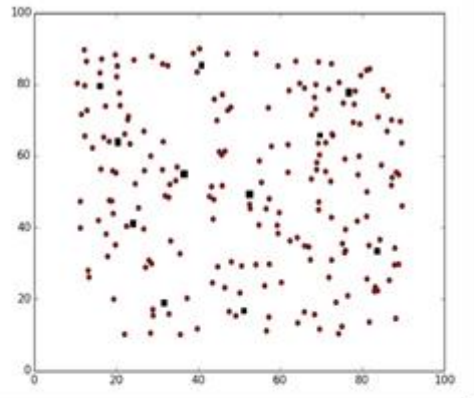


Figure 5. General Approach Optimal Set of Chart

Then, in the Python environment, use the improved algorithms CURE to cluster calculate for point set. Then the center point can be obtained, and then use TSP Problem (Traveling Salesman Problem) respectively solve for 10 subclasses of point set, get the shortest transport path, then put it into UFLT model, in the LINGO11.0 software calculate the global optimal solution as the following Table 2, meanwhile it is represented by square points in Figure 5, the optimal objective function value is 6030.1.

Table 2 improve CURE Clustering for Optimal Solution Set Table

X	Y	Z
73.8	27.5	26.6
44.9	79.6	9.5
31.9	20.8	89.8
58.7	35.5	18.5
71.0	80.4	22.9
23.5	83.2	98.1
63.8	45.9	8.9
90.6	52.2	62.1
20.3	42.5	60.8
25.5	60.3	77.6

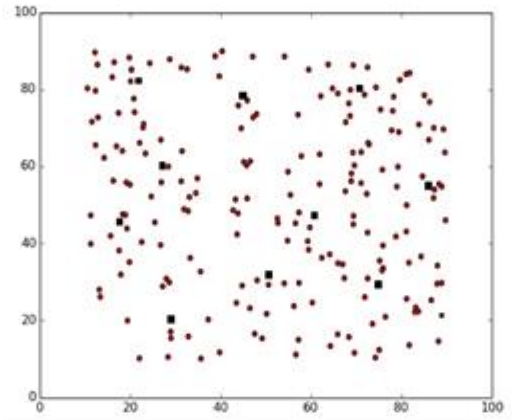


Figure 6. Improved CURE Clustering for Optimal Solution Set Figure

By comparing the above two, it clearly proved that the improved algorithm to CURE is superior to the traditional algorithm in the logistics center location selection.

6. Conclusion

Logistics center location problem is a key factor in optimizing the layout of the logistics network, but also an important factor to reduce the cost of logistics enterprise and improve the efficiency of logistics and transport. In this article, it has introduced the CURE clustering algorithm. analyze the logistics center location selection problem, build UFLP model, make numerical experiment, it proved that the improved CURE clustering algorithm has the effectiveness and superiority for logistics location problem, and it can thus extend to the multidimensional structure. It improves the reliability of this method in practical application by considering more factors in the location.

References

[1] J. Han and M. Kamber, "The concept of data mining and technology", Fan Ming, Meng Xiaofeng, translation. Beijing Mechanical industry press, **(2006)**.
[2] J. MacQueen, "Some methods for classification and analysis of multivariate observations", Proc. of the 5th Symposium on Mathematical Statistics and Probability, Berkeley, **(1967)**, pp. 281-297.
[3] H. Jia-Wei and M. Kamber, "Data mining: concepts and techniques", [S.l.]: Higher Education Press, Morgan Kaufmann Publishers, **(2002)**.
[4] S. Guha, R. Rastogi and K. Shim, "CURE an efficient clustering algorithm for large database", Information Systems, vol. 26, no. 1, **(2001)**, pp. 35-58.
[5] J. Jaroslay and B. Lubos, "An acceleration of Erlenkotter-K rkel's algorithms for the uncapacitated facility location problem", Annals of Operation Research, vol. 164, **(2008)**, pp. 97-109.

Authors

YuYi, master of school of Computer and Information Engineering Research interests: logistics information system

Liu Xin-liang, associate professor, master instructor Research interests: Management Information System; Mobile Internet