Application of Hierarchical Clustering and K-means Clustering in Location Selection of Rural Logistics Center

Wensheng He and Zhirui Hu

ABSTRACT

For rural reasonable logistics center site selection, combined with the characteristics and conditions of the rural road network in rural areas of logistics, based on Python hierarchical clustering algorithm and implementation of K-means clustering algorithm, established a logistics center selected rural road network reality Address model. Clustering points need to be positioned. Firstly, the number of suitable initial cluster centers is determined by hierarchical clustering; then, the number of selected initial cluster centers is used, and the K-means algorithm is used to cluster clusters based on Euclidean distance similarity. Iterative clustering is completed after three actual iterations. As a result, the cluster center is three logistics centers, and the problem of visualizing the location of the rural logistics center is realized. Finally, an example is given to verify that the method of solving the location problem of rural logistics center is feasible.

1. INTRODUCTION

With the implementation of China's, Extending Radio and TV Broadcasting Coverage to Every Village Project, the promotion of The construction of new socialist countryside, the continuous popularization and application of the Internet and smart phones in rural areas, e-commerce such as online shopping has also developed greatly in rural areas, bringing rural logistics demand. growth of. Rural logistics is a general term for transportation, packaging, handling, loading and

Wensheng He, School of Economics and Management, Southwest Petroleum University, Chengdu 610500, China
Zhirui Hu, School of Economics and Management, Southwest Petroleum University, Chengdu 610500, China
unloading, distribution, warehousing and other related activities for rural residents' production, life and other economic activities. It is a concept relative to urban logistics [1]. The location of rural logistics centers is one of the important issues in the development of the logistics industry. [2] studied the public logistics center location problem, and draw the best addresses and optimal size; [3] in order to minimize the total cost of logistics objective function, based on the reality of the rural road network distribution center options Address optimization model; [4] established a mixed integer programming model, which takes into account multiple influencing factors of two-way logistics; [5] most clustering algorithms inevitably need to specify the number of clusters, initial center, etc. In the actual planning, it is often difficult to determine; [6] based on the characteristics of rural China, based on the analysis of rural logistics characteristics, based on AP clustering algorithm to study the "logistics center location problem" in rural logistics; [7] proposed a multi-level equilibrium location algorithm based on BIRCH clustering.

The paper studies the location problem of rural logistics center, combines the characteristics of rural road network and the characteristics of rural logistics, based on Python implementation, uses hierarchical clustering algorithm and K-means clustering algorithm to solve the problem of rural logistics center location, and uses examples to verify The method is feasible.

2. APPLICATION OF CLUSTERING ALGORITHM IN LOCATION SELECTION OF LOGISTICS CENTER

2.1 Hierarchical Clustering Algorithm and K-means Clustering Algorithm

Hierarchical clustering algorithms can be divided into two categories according to different strategies. One is a condensed hierarchical clustering that gradually produces clustering results from the lower layer to the upper layer; the other is a split hierarchical clustering that gradually produces clustering results from the upper layer to the lower layer. There are not many studies on split hierarchical clustering, which is not common in engineering practice. Condensed hierarchical clustering initial state Each data object is a sub-cluster, and then these sub-clusters are gradually merged according to some principles, and finally all data objects are combined into one cluster [10]. The K-means algorithm belongs to the partitioning clustering algorithm. The K-means algorithm is based on Euclidean distance and has high operational efficiency, which can adapt to the needs of large-scale data clustering. The traditional K-means algorithm first assigns a data object to each partition as the cluster center, and then lists other data objects one by one into the partition with the greatest similarity to its cluster center; then updates the cluster according to the mean of all data objects of the partition class. Center; repeated division and update to achieve the purpose of optimizing clustering results.
2.2 Description of the location problem of rural logistics center

There are many types of rural logistics services, and the distribution routes are complex. There is no direct data available for each point of demand. These pose great inconveniences for the location analysis of rural logistics centers.

Assuming that any logistics node of the logistics system is represented as Node i, all logistics nodes of the logistics system form a set M, i.e. Node i ∈ M. Firstly, the Euclidean distance between the various logistics nodes Node i in the rural area is obtained and saved in the form of a two-dimensional matrix; since K-means is sensitive to the selection of the initial cluster center, and needs to be initial subjective selection of the cluster center. The paper first uses hierarchical clustering to cluster 13 data points. Through experiment and observation, and according to the constraints (cost, policy, etc.) of the actual problem, select the appropriate number of cluster centers k, and then use K-means clustering. The method combines the selected initial cluster centers, clusters according to the Euclidean distance between the points, and iterates three times of clustering, and finally obtains the specific location of the cluster center, that is, the location of the final selected logistics center.

2.3 Algorithm for problem solving

The hierarchical clustering algorithm and K-means clustering algorithm are introduced to solve the location of rural logistics center. For the K-means clustering algorithm, because the selection of the initial clustering points is difficult to determine, the hierarchical clustering algorithm is used first to observe the clustering number. The clustering is more in line with the conditions of address selection, and as the next step. Selection of initial clustering points for K-means clustering. Thereby obtaining a new solution method.

The AGNES algorithm is a bottom-up hierarchical clustering algorithm. Initially initialize each sample in the dataset to a cluster, then find the two closest clusters, merge them, and repeat the process until the preset number of clusters.

Three formulas for calculating distance:

\[ d_{\text{avg}} (C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{C_i} \sum_{C_j} \text{dist}(x, z) \]

\[ d_{\text{min}} (C_i, C_j) = \min_{x \in C_i, z \in C_j} \text{dist}(x, z) \]

\[ d_{\text{max}} (C_i, C_j) = \max_{x \in C_i, z \in C_j} \text{dist}(x, z) \]

The hierarchical clustering algorithm is based on the above three different formulas, and the corresponding ones are called all links, single links and full links. The steps of the AGNES algorithm are as follows:

Each sample is a set of initialization data cluster, and into a collection C of. Calculating the distance between any two sets, the coexistence of the M in. Set the current number of clusters q=m.
When \( q \) is greater than \( k \), perform the following steps:

1. Find the two most recent collections from \( C_i \) and \( C_j \), the \( C_i \) and \( C_j \) merger. And assigned to \( C_i \). (2) in the set \( C \) will \( C_j \) delete, update \( C_j + 1 \) to \( C_q \) subscripts. (3) Delete the \( j \)th and \( j \)th columns of \( M \). Update the \( i \)-th row and the \( i \)-th column of \( M \). (4) \( q = q - 1 \). Returns the set of clusters \( C \). Cluster termination condition: The number of clusters \( k \) is limited. When the number of clusters obtained has reached \( k \), the cluster terminates. After the test paper \( k \) equals 3, 4 or 5, according to the constraint considerations, finally taking \( k = 3 \).

K-means clustering is as follows:
Input: sample data set \( D \) and the cluster number of clusters \( k \), obtained by the above \( k \) takes 3 ;
Output: A collection of clusters.
Methods: (1) randomly select \( k \) sample points from the sample as the initial mean vector \( \{ \mu_1, \mu_2, \ldots, \mu_k \} \); (2) Cycle the following steps until the stop condition is reached; (3) the order is not equal to the empty set, of which \( C_i \); (4) Calculate the distance between them to the \( k \) mean vectors for all sample points, take the mark of the mean vector corresponding to the shortest distance as the cluster mark of the store, and then add the point to the corresponding cluster; (5) Calculate their new mean vector for each cluster. If there is a change from the previous vector, update it and use it as the new mean vector. If there is no change, it will not change. This article sets the iteration 5 times. When the iteration reaches the 3rd stop, the clustering is completed.

\[
\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x.
\]

3. APPLICATION AND ANALYSIS OF EXAMPLES

A rural logistics system consisting of 13 townships in a rural area was selected for simulation experiments. As shown in Figure 3.1, the system contains 13 logistics nodes.
The actual distance traveled here is used instead of the straight line distance between the nodes.

Scale 1cm=2.75km (the scale of the coordinate axis in the figure is 1 cm)

The image coordinates (unit: cm) and actual coordinates (unit: km) of each township node are as follows:

Legend coordinates:
Node9 : ( 2.2 , 2.9 ), Node13 : ( 3.1 , 3.9 ), Node8 : ( 3.6 , 1.9 ),
Node7 : ( 3.7 , 6.2 ), Node1 : ( 5.3 , 1.4 ), Node6 : ( 5.1 , 7.2 ), Node2 : ( 6.4 , 0.8 ),
Node3 : ( 6.7 , 2.1 ), Node12 : ( 7.3 , 5.8 ), Node10 : ( 7.6 , 1.0 ), Node11 : ( 8.2 , 2.4 ),
Node4 : ( 8.1 , 4.6 ), Node5 : ( 8.5 , 6.8 ).

Firstly, using hierarchical clustering, it is still necessary to specify the number of clustering k, which is 3, 4 or 5 and observe the number of clusters. The clustering is more in line with the conditions of address selection, which is used as the initial step of K-means clustering. The choice of cluster points. Through the test, From the perspective of economy and convenience, the number of clusters suitable for the location of rural logistics centers is 3. When the number of clusters is set to 3, the clustering effect is shown in Figure 3.2:
Next, when the initial number of clusters is 3, K-means clustering is used: the iteration is set 5 times, and the experiment is iterated 3 times to reach the target. The final iteration results are as follows:

The logistics center location of the 3 iterations is shown in Figure 3.3:

Figure 3.3. Logistics Center Location in 3 Iterations.
Ticks portion is the center point of the cluster. 3 logistics center coordinates (unit: cm) is:
(6.84, 1.54), (6.54, 6.12), (2.96666667, 2.9)
Average_distance = 2.38236

Logistics center location in the actual map, the yellow star position is 3 logistics centers in. As shown in Figure 3.4.

![Figure 3.4. Logistics Center Site Selection.](image)

From figure 3.4 can be seen in selected rural areas of logistics center reasonable layout, and are located in State Road, near the provincial or county roads. The division of radiation in the logistics center is also very reasonable.

4. CONCLUSIONS

In conclusion, the paper studies the problem of reasonable location of rural logistics center. Based on the theory of logistics and cluster analysis, the research on the location of logistics center is reviewed. Combined with the characteristics of rural road network and the characteristics of rural logistics, the location model of logistics center in rural reality road network was established. Based on Python implementation, the combination of hierarchical clustering algorithm and K-means clustering algorithm overcomes the shortcomings of the traditional K-means algorithm in the initial clustering point selection. The simulation experiment was
carried out with a rural logistics system consisting of 13 townships in a rural area, and the method proposed in the paper was valid and effective.

REFERENCES