World Scientific Research Journal (WSRJ) ISSN: 2472-3703
www.wsr-j.org

Location Optimization of Agricultural Cold Chain Distribution Center Based on K-means Clustering Algorithm and AHP

Meiling He[1, a, *], Jun Pu[1, b]

[1]School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang, China

[a]hemeiling@ujs.edu.cn, [b]lapjun2019@163.com

Abstract: The article mainly studied the location of the cold chain logistics distribution center of agricultural products. A location selection model based on k-means clustering algorithm and analytic hierarchy process (AHP) was proposed. Firstly, cluster analysis of demand points was carried out to obtain location options. Secondly, AHP was used to construct an indicator system that influences the location decision. And the comprehensive weight of each factor in the index layer was calculated so as to score the selection of the alternative scheme. Finally, the feasibility of the model was verified by an example analysis.
Keywords: Cold Chain logistics, distribution center location, K-means clustering algorithm, Analytical hierarchy process (AHP).

1. INTRODUCTION

With the improvement of living standards and the continuous development of social economy, the demand for agricultural products in the off-season becomes larger and larger. With people's awareness of "cold chain logistics" constantly improving, higher requirements have been placed on the quality and freshness of agricultural products. Cold chain logistics of agricultural products refers to the logistics process that agricultural products are always in the low temperature environment after being purchased or fished from the origin, through storage, transportation, distribution, retail and finally reaching the consumption terminal [1]. Cold chain logistics are required, due to a large variety of agricultural products with seasonal, large environmental impact, short shelf life, corruption and other characteristics. Through low temperature environment, it can reduce the metabolism of agricultural products and the growth of microorganism, as to inhibit the rate of corruption [2]. These processes ensure the quality and reduce the loss.

The cold chain logistics system of agricultural products is an organic combination of transport, storage, handling, transportation, packaging, circulation processing, distribution, information processing and other basic functions. If all aspects of the system are properly configured, the logistics will reach the optimal level. As the bridge of manufacturers and retailers [3], logistics distribution center is indispensable in the cold chain transportation of agricultural products. It will have a long-term impact since established, and its location is directly related to the logistics cost and customer service level [4]. Therefore, the research on the location and optimization of logistics distribution center of cold chain logistics is a key work of logistics system, which has important practical significance.

At present, Guo Yanan et al. proposed measures to improve the efficiency of cold chain logistics of agricultural products from three aspects: further strengthening infrastructure technology, constructing a complete cold chain logistics network system, and increasing capitalization investment by means of capital power [5]. Sobhi Mejjaoulia et al. established a decision-making model for the decision-making process of perishable products in cold chain logistics, and implemented it in the cloud through temporary virtual machines (VM) related to goods, based on the assessment of actual transportation conditions and the location of transportation products to make an operational decision about the shortest path and lowest cost [6]. On the basis of food safety, Wang Jing et al. used the principle of fresh food decay to establish a mathematical model to minimize logistics costs [7]. Yuan Qun et al. constructed an optimization model for the location of cold chain logistics distribution center considering factors such as cost and cargo rate of damage of goods, and improved the crossover operator's hybrid genetic algorithm with greedy algorithm [8]. Ren Chunyu et al. proposed to use the simulated annealing genetic algorithm and AHP combined solution to get the most preferred address [9]. Abo-Elnaga et al. proposed a combination of activity set strategy and punishment method, and trust domain technology for warehouse location problem (WLP) [10].

In summary, in the existing research results, foreign optimization criteria for warehouse location generally adopt the following forms: minimum storage cost, cost of goods conversion through warehouse facilities, cost of goods transported to warehouse facilities, etc. Domestically, the efficiency is improved mainly from the optimization of traditional warehouse location or the optimization of distribution routes.

In this paper, we construct cold chain logistics distribution center location model, combining k-means clustering algorithm and AHP to solve the location problem, so as to improve the overall efficiency of cold chain logistics.

2. METHODS

2.1 K-means clustering algorithm

Cluster analysis is a statistical analysis method to study classification problems, including division method, hierarchical method, density-based method, network-based method and model-based method. Its analysis process is a process to divide all data in the data set into multiple categories according to similarity. Among them, the division method takes the distance as the similarity measurement between different data in the data set, and divides the data set into multiple clusters to make the generated clusters as compact and independent as possible, including k-means, k-medoids and other methods.

The principle of the k-means algorithm can be described as: input a data {Z} containing N samples and the number of clusters to be selected M. First, randomly select M clustering objects as the initial position of the clustering center, and then according to the remaining objects to each cluster center. The distance is assigned to the nearest cluster, and the cluster center of the new cluster is obtained. The above process is iterated until the sum of the distances of each sample data to the cluster center reaches a minimum value, and the clustering result is obtained.

The establishment process of k-means clustering analysis model is as follows: given a data set {Z} containing N samples and the number of clusters to be selected M (Mj, j=1,2,3…M), and the k-means clustering analysis model is established by minimizing the distance from each sample in each cluster to the objective point [11].

The target function is as follows.

$$\min \sum_{j=1}^{m} \sum_{N \in M_j} y_{ji} \left\| Z - m_j \right\| \tag{1}$$

In the formula, m_j epresents the mean distance between each sample in the cluster and the clustering point. y_{ji} represents the unique cluster and clustering center for each sample data. $y_{ji} = 1$ means that the pattern sample j is assigned to the cluster center i; otherwise, 0 is set. The calculation formula is below.

$$m_j = \frac{1}{\sum_{i=1}^{N} y_{ji}} \sum_{j=1}^{N} y_{ji} Z_j \, (j = 1, 2, 3 \cdots N) \tag{2}$$

$$\sum_{i=1}^{N} y_{ji} = 1 (j = 1, 2, 3 \cdots N)$$

$$y_{ji} = \begin{cases} 1, Sample\ i\ is\ distributed\ to\ the\ po\text{int}\ j \\ 0, \ \ else \end{cases}$$

2.2 AHP

The Analytic Hierarchy Process (AHP) was proposed by American operational research expert of the University of Pittsburgh, Professor T.L. Saaty, in 1970s [12]. This method decomposes the factors that affect the decision-making of complex problems into several components, and calculates the hierarchical single order and total order by using the method of fuzzy quantification of qualitative indicators, and presents them in the form of data, so as to determine the importance of each factor to the overall scheme decision. The implementation steps of the Analytic Hierarchy Process are as follows:

Identify the problem. By defining the scope of the problem, the specific requirements put forward, the elements involved and the relationships among the elements. Clarify what problems need to be solved and whether the information has met the needs.

Establish a multi-level hierarchical structure. According to the understanding and analysis of the problem, the elements involved in the evaluation system are arranged according to the nature. Generally, it is divided into target layer, criterion layer and index layer from top to bottom.

Establish judgment matrixes. Judgment matrix is the basic information of AHP, and also the basis of relative importance degree calculation and single hierarchical arrangement.
Relative importance degree (weight coefficient) is calculated and consistency test is performed. Based on the judgment matrixes, the relative importance of the total hierarchy is calculated and the consistency is tested. Calculate maximum feature root $\lambda_{\max}$ λ_{max}, importance degree w_i, relative importance degree vector w_i^0, consistency index $C.I.$, mean random consistency index $R.I.$ etc. Among them, the commonly used random consistency index values are shown in table 1, the calculation formula of required parameters is shown in equations (3) - (6).

Table 1. Common mean random consistency indicator values

n-dimensional matrix	3	4	5	6
$R.I.$	0.52	0.86	1.12	1.26

Calculation formula of importance degree:

$$W_i = \left(\prod_{j=1}^{n} a_{ij} \right) \frac{1}{n} \tag{3}$$

Calculation formula of relative importance degree vector:

$$W_i^0 = \frac{W_i}{\sum_i^n W_i} \tag{4}$$

Calculation formula of maximum feature root:

$$\lambda_{\max} \approx \frac{1}{n} \sum_{i=1}^{n} \frac{(AW)_i}{W_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{j=1}^{n} a_{ij}}{W_i} W_j \tag{5}$$

Calculation formula of consistency index:

$$C.I. = \frac{\lambda_{\max} - n}{n-1} \tag{6}$$

In the formula, W_i W_i represents the importance of the i-th component, and the arithmetic mean of the i-th sub-vector. W_i^0 represents the relative importance degree vector, and the normalized value of each factor relative to a factor in the upper layer. $(AW)_i$ is the i-th component of the vector AW, and represents the ratio of the arithmetic mean of the i-th component of vector

AW to the importance degree of the i-th component. a_{ij} is an element of the i-th row and j-th column in the matrix. n represents the index number in this layer. When $C.R.=\frac{C.I.}{R.I.}<0.1$

$C.R.=\frac{C.I.}{R.I.}<0.1$, it is considered that the judgment matrixes pass the consistency test.

Then, calculate the comprehensive importance degree (overall weight of the system) and conduct a consistency test to obtain the optimal solution.

3. EXAMPLE ANALYSIS

A city has a company B engaged in the cold chain logistics of agricultural byproducts. It plans to establish three distribution centers among the 70 demand points in the city. These demand points are stable, large, and demanding for agricultural byproducts. Due to the management optimization of warehousing logistics and the saving of transportation costs, it is necessary for company B to make the location decision of cold chain logistics and optimize it to maximize the company's interests.

3.1 Cluster analysis of site selection

As the basis of solving the modeling problem of the optimal location for the cold chain logistics of agricultural byproducts of company B, 70 demand points are placed in the rectangular coordinate system according to the actual position. Each demand point corresponds to A point in the coordinate system, which is represented by the coordinate (x, y), and the number is 1, 2, 3, and...70. K-means clustering analysis is carried out on the data by MATLAB programming. Since the number of distribution centers to be selected is 3, only M=3 is considered. According to the cluster classification results, 70 demand points can be divided into three new clusters, and the coordinates of j_1, j_2 and j_3 in the cluster and the distance between each point in the cluster and the cluster point can be obtained. The results are shown in Table 2 and Figure 1.

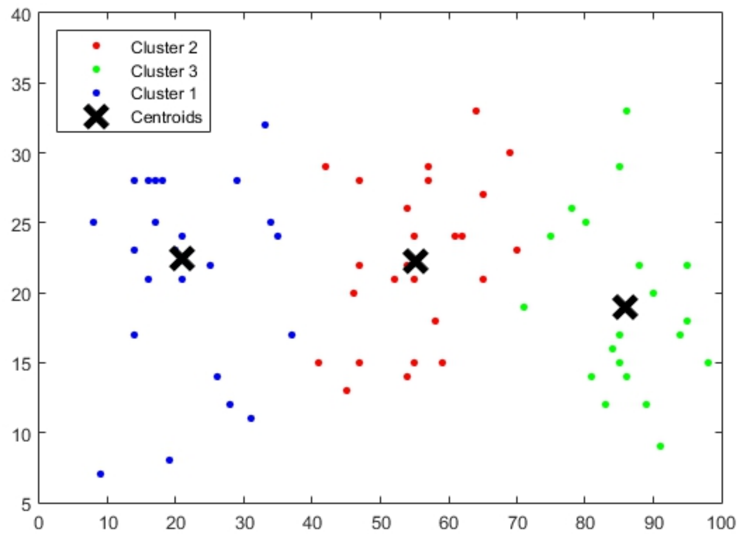

Fig 1. Distribution map of location cluster analysis results

Table 2. The clustering results of 70 demand points of company B

region number	number of demand points in the region	coordinates of demand points in the region	coordinates of clustering points
1	25	E_1(5,35),E_2(8,25),E_3(9,7),E_4(14,28),E_5(14,23),E_6(14,17),E_7(14,36),E_8(16,21),E_9(16,28),E_{10}(17,28),E_{11}(17,25),E_{12}(18,28),E_{13}(19,8),E_{14}(20,23),E_{15}(21,24),E_{16}(21,21),E_{17}(26,14),E_{18}(26,14),E_{19}(28,12),E_{20}(29,28),E_{21}(31,11),E_{22}(33,32),E_{23}(34,25),E_{24}(35,24),E_{25}(37,17)	(20.84,22.48)
2	25	F_1(41,15),F_2(42,29),F_3(45,13),F_4(46,20),F_5(47,15),F_6(47,28),F_7(47,22),F_8(52,21),F_9(54,26),F_{10}(54,14),F_{11}(54,22),F_{12}(55,24),F_{13}(55,15),F_{14}(55,21),F_{15}(57,29),F_{16}(57,28),F_{17}(58,18),F_{18}(59,15),F_{19}(61,24),F_{20}(62,24),F_{21}(64,33),F_{22}(65,21),F_{23}(65,27),F_{24}(69,30),F_{25}(70,23)	(55.24,22.28)
3	20	G_1(71,19),G_2(75,24),G_3(78,26),G_4(80,25),G_5(81,14),G_6(83,12),G_7(84,16),G_8(85,15),G_9(85,29),G_{10}(85,17),G_{11}(86,33),G_{12}(86,14),G_{13}(88,22),G_{14}(89,12),G_{15}(90,20),G_{16}(91,9),G_{17}(94,17),G_{18}(95,18),G_{19}(95,22),G_{20}(98,15)	(85.95,18.95)

3.2 Analysis by AHP

The E14, E15 and E16 demand points in region 1, F11 and F14 demand points in region 2 were taken as the alternative distribution centers respectively. Based on the results of k-means clustering analysis, hierarchical analysis was conducted to obtain the optimal solution. Firstly, the factors affecting the location of the cold chain logistics distribution center are analyzed.

The cold chain logistics of agricultural byproducts not only has the problems in the development of general logistics, but also has its unique characteristics. For example, due to the perishability and seasonality of agricultural byproducts, the construction of cold chain logistics is stricter in terms of time, technology, safety and fixed facilities than general supply chain logistics. Therefore, this paper summarizes the factors affecting the location of cold chain logistics distribution center in the following three aspects.

The first factor is traffic geography. The main traffic geography factors influencing the location of agricultural byproducts cold chain logistics distribution center are traffic conditions, transportation radiation range, distribution center area, and distance from the supply location.

The second factor is the cost. Minimizing costs and maximizing benefits are the goals pursued by every logistics enterprise. Due to the particularity of its distribution products, cold chain logistics has high requirements for storage equipment and conveyances.

Therefore, the cost factors that affect the location of agricultural byproducts cold chain logistics distribution center mainly include land cost, labor cost, transportation cost and distribution center construction cost.

The third factor is economic policy. Cold chain logistics is closely related to people's livelihood. Do local governments have preferential policies for developing cold chain logistics industry? Do enterprises have impacts on the local environment, the average income of residents and the permanent population? These factors will directly or indirectly affect the sustainable development or economic benefits of logistics enterprises.

The fourth factor is the operating environment. The main operating environment factors influencing the location selection of cold chain logistics are the stability of water supply and power supply, service level, geological conditions and topographic conditions.

3.2.1 Build a hierarchy chart

Based on the above analysis, the hierarchical analysis structure of the location problem of the cold chain logistics distribution center is obtained, as shown in Figure 2.

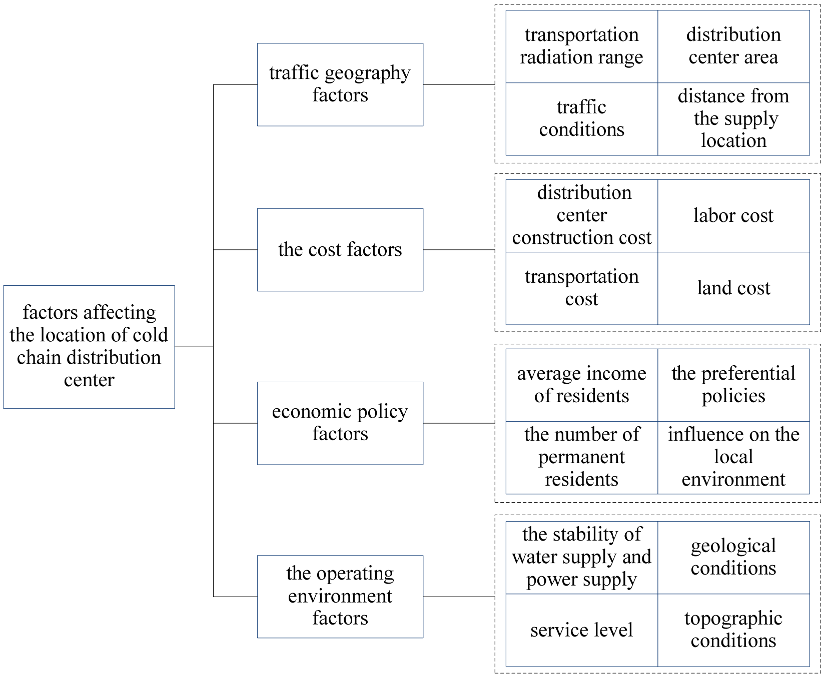

Fig 2. Hierarchical structure analysis diagram

3.2.2 Establish a judgment matrix

According to the location principle of cold chain logistics distribution center, B company has extracted the key factors and established the evaluation index of qualitative analysis according to its development strategy and years of operation experience. In addition, a large number of questionnaires were distributed to site selection experts and enterprise managers as the objects of the survey, the pairwise comparison results of various factors were calculated, the consistency of the obtained data was tested, and some abnormal data were discarded. Finally, the importance scale of the comparison between the two factors was determined by the arithmetic mean method and the 1-9 scale proposed by the American operations research scientist Saaty [11], and the judgment matrix was constructed. The criteria are set up to determine the matrix and to determine the matrix and to determine whether to pass the consistency test. The calculation results show that $C.R. = 0.0556 < 0.1$, which passes the consistency test.

3.2.3 Criterion layer to target layer pairwise judgment matrix

The judgment matrices within the factors of traffic geography, cost, economic policy and business environment were established respectively, and the second-level weight was calculated, and the consistency test was judged. The calculation results show that C.R.= 0.0140,0.0861,0.0734,0.0725<0.1, which passes the consistency test.

3.2.4 Calculate the overall importance

The weight coefficient of each single layer is comprehensively evaluated to determine the overall importance, and the order of importance of each influencing factor of each scheme layer is obtained.

4. CONCLUSION

Based on the above classification results, the shorter demand points of j1 in the middle distance clustering center of region 1 are E14 (20,23), E15(21,24) and E16 (21,21). The short demand points of region 2 medium-distance clustering center j2 (55.24,22.28) are F11 (54,22) and F14 (55,21). The shortest demand point in region 3 is the cluster center j3 (85.95,18.95), which has only one demand point, i.e. G10 (85,17). Based on the principle of shortest distance, G10 is selected as the optimal location for the cold chain distribution center in region 3. The site selection of region 1 and region 2 is further analyzed and optimized.

The influencing factors of the index layer were sorted and the total weight was calculated, as shown in Table 3 [13]

Table 3. Total weight ranking of influencing factors

target layer	criteria level	first-level weights	indexes level	Secondary weights	Total weight
location of cold chain logistics distribution center	traffic geography factors	0.3333	traffic conditions	0.2800	0.0933
			transportation radiation range	0.2800	0.0933
			distribution center area	0.2400	0.0799
			distance from the supply location	0.2000	0.0667
	the cost factor	0.2917	distribution center construction cost	0.3571	0.1042
			transportation cost	0.2857	0.0833
			land cost	0.2143	0.0625
			labor cost	0.1429	0.0417
	economic policy factors	0.2083	average income of residents	0.3214	0.0669
			the number of permanent residents	0.2587	0.0539
			preferential policies	0.2143	0.0446
			influence on the local environment	0.1818	0.0379
	the operating environment	0.1667	the stability of water supply and power supply	0.3529	0.0588
			service level	0.2941	0.0490
			geological conditions	0.1765	0.0294
			topographic conditions	0.1765	0.0294

According to the total weight ranking of influencing factors of AHP, the analysis shows that: among the influencing factors of the above index layer, the final important influencing factors are the construction cost, traffic conditions, traffic radiation range and transportation cost of the distribution center, and other indicators have little influence on the location decision of cold chain logistics distribution center, so it is not considered for the time being. The above influencing factors are selected as the evaluation indexes for company B's site selection decision in region 1 and region 2, and the final result is: E16, F14 and G10 demand points in region 1, 2 and 3 are selected as the site selection of its cold chain logistics distribution center. In this paper, based on the example analysis, we get the regional cluster center as its alternative by using the k - means clustering algorithm to cluster analysis of demand point. Then we analysis of various influence factors of cold chain logistics distribution center location, using the analytic hierarchy process(AHP) to get the total weight of all the factors, ultimately determine the optimal location of the regional strategy. It proves the practicability of combining k-means clustering analysis with AHP, and provides a scientific reference and evaluation method for the location of cold chain logistics distribution center.

ACKNOWLEDGEMENTS

This paper was supported by “the Natural Science Foundation of Jiangsu Province” (Grants No BK20160512), “the Humanity and Social Science Youth foundation of Ministry of Education of China” (Grants No 16YJCZH027), “the Social Science Foundation of Jiangsu Province” (Grants No 15GLC004), “the Scientific Research Projects of College Students in Jiangsu University” (Grants No 17A352).

REFERENCES

[1] Rouviere E, Latouche K, “Impact of liability rules on modes of coordination for food safety in supply chain”, European Journal of Law & Economics, 2014, Vol. 37 (1), p111-130.

[2] Jahid I K, Ha S D, “A review of microbial biofilms of produce: future challenge to food safety”, Food Science and Biotechnology, 2012, Vol. 21 (2), p299-316.

[3] Song H L, Huang Y P, Zhang Y, Jiang M, “Study on Location Model of Fresh Food Distribution Center—Taking Distribution Center of Yantai Cherry as an Example”, Journal of Shandong Technology and Business University, 2018, Vol. 32 (3), p48-54.

[4] Zhang Jin, “Principles and methods of logistics”, Chengdu, Southwest Jiaotong University Press: 2009.10.

[5] Guo Y N, Hu Y K, “Research on the present situation and countermeasures of Chinese agricultural cold-chain logistics”, Logistics Engineering and Management, 2018, Vol. 40 (2), p4-7.

[6] Sobhi Mejjaouli, “Cold supply chain logistics: System optimization for real-time rerouting transportation solutions”, Computers in Industry, 2018, Vol. 95, p68–80.

[7] Wang J, Liu H T, Zhao R, “The optimization of cold chain operation based on fresh food

safety", Systems Engineer-Theory&Practice, 2018, Vol.30 (1), p22-33.
[8] Yuan Q, Zuo Y, "Selection of cold chain logistics distribution center location based on improved hybrid genetic algorithm", Journal of Shanghai Jiaotong University. 2016, Vol. 50 (11), p1795-1800.
[9] Ren C Y, Wang X B, Li T Y, "Study on location based on genetic algorithm combing with simulated annealing and AHP", Journal of Harbin University of Commerce (Natural Sciences Edition), 2006, Vol. 20 (1), p58-62.
[10] A. bo-Elnaga, B. El-Sobky a, L. Al-Naser b, "An active-set trust-region algorithm for solving warehouse location problem", Journal of Taibah University for Science, 2017, Vol. 11 (2), p353–358.
[11] Dou X B, Li L.Solving, "Location-based k-means clustering and genetic algorithm research on path optimization questions", Logistics Engineering and Management, 2017, Vol. 39 (5), p71-73.
[12] Saaty T L, "The anlytic hierarchy process", McGraw-Hill, New York, 1980.
[13] Zhang S, Zhu L M, "Research on location method of coal logistics node based on analytic hierarchy process", Journal of Nanjing University of Science and Technology, 2015, Vol. 39 (3): 301.