

## *Retraction*

# **Retracted: Research on Location of Logistics Distribution Center Based on K-Means Clustering Algorithm**

### **Security and Communication Networks**

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] P. Wang, X. Chen, and X. Zhang, "Research on Location of Logistics Distribution Center Based on K-Means Clustering Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 2546429, 9 pages, 2022.

## Research Article

# Research on Location of Logistics Distribution Center Based on K-Means Clustering Algorithm

Ping Wang,<sup>1</sup> Xianjun Chen ,<sup>1</sup> and Xuebin Zhang<sup>2</sup>

<sup>1</sup>School of Economics and Management, Chongqing Three Gorges Vocational College, Wanzhou 404155, Chongqing, China

<sup>2</sup>College of Business Administration, Ningbo Polytechnic, Ningbo 315800, Zhejiang, China

Correspondence should be addressed to Xianjun Chen; 2005180182@cqsxzy.edu.cn

Received 25 April 2022; Accepted 8 June 2022; Published 9 July 2022

Academic Editor: Hangjun Che

Copyright © 2022 Ping Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The basic task of the logistics distribution center is to achieve the storage and distribution of materials, and to plan, implement, and manage the effective flow of materials from the supplying place to the consumption place. Scientific location of logistics distribution center can effectively reduce logistics cost, improve the speed of circulation, increase the profits of enterprises, and enhance the core competitiveness of enterprises. Combining the advantages of K-means clustering algorithm, this paper applies it to the location problem of logistics distribution center and proposes a logistics distribution center location method combining K-means clustering theory and D-S reasoning, which provides a better solution for the location problem of logistics distribution center. Through case analysis, K-means clustering algorithm can obtain reasonable location of logistics distribution center, which can be applied to the location of multilevel logistics distribution network, and has certain practical application value.

## 1. Introduction

With the vigorous development of e-commerce, the logistics industry for ordinary consumers has become a new growth point. From 2010 to 2020, the business volume of national express enterprises increased from 2.34 billion pieces to 83.36 billion pieces, with a compound annual growth rate of about 42.9%. Moreover, in 2021, the business volume of express enterprises is 95.5 billion pieces. From January to November, 2021, the income of national express enterprises totalled 941.47 billion yuan, a year-on-year increase of 19.6%. As of now, the escalated level of the strategies business has improved, yet there are still a few issues, for example, in reverse activity mode, jumble between coordinated factors supply and operations interest, and so forth. Hence, further developing effectiveness of planned operations transportation and decreasing coordinated factors cost have turned into the essential objectives of endeavors [1, 2]. Dispersion focus is the critical hub during the time spent express, which has a direct impact on the choice of transportation routes, the transit time of express mail, and the logistics cost; in addition, it is closely related to the service quality and users' experience of logistics

enterprises. With the rapid increase of logistics, the traditional distribution method has not considered the nonrepetition of traffic. Without considering the location, the problems of the interests between sorting center and customers, and the irrationality of traditional distribution methods, the problems of low utilization rate of urban logistics distribution resources, poor decision-making ability, low scheduling efficiency, and high error rate are caused [3]. It is an important part of logistics distribution service system, which is the key to saving cost of distribution. Therefore, the location of logistics distribution centers based on K-means clustering algorithm is studied in this paper, and a location scheme of logistics distribution centers is put forward, which can provide new ideas for the research and application of optimization in location and realizes the sustainable development of logistics industry.

## 2. Evaluation System of Location of Logistics Distribution Center

As the transit station of logistics distribution, the logistics distribution center is the key to the whole logistics system

planning in the logistics industry. Choosing the factors of location is a comprehensive problem, which aims to make the process of location more scientific, standardized, and practical. On the basis of previous research, from the aspects of traffic conditions, laws and policies, resource conditions, operating environment, natural environment, cost, information quality, etc., the evaluation system of location of logistics distribution center is constructed.

*2.1. Traffic Conditions.* The distribution center needs to have strong speed of response, be able to provide services to customers in time, and have certain reliability and convenience in service, which mainly includes road facilities, accessibility, and public facilities [4]. When choosing a site, it is necessary to consider the transportation of the selected location. Location determines the value, and it should not only help to improve the economic benefits of enterprises but also pay attention to providing convenient, efficient, and low-cost services for customers.

*2.2. Legal Policy.* The distribution center occupies a large area. Considering the land price, natural environment, energy, and other conditions, it should also comply with various logistics policies and regulations [5]. Especially commercial scale, lease terms, zoning, etc, for example, whether it is allowed to build logistics distribution centers in some areas, whether the relevant policies logistics, custom duties and tax policies are beneficial to the establishment of logistics distribution centers. Because of the incomplete unification of logistics regulations and policies, there exists commonly incongruities and conflicts. Therefore, to choose areas that are beneficial to the construction and development of logistics distribution centers, the construction must be coordinated with the urban plan.

*2.3. Resource Conditions.* The logistics distribution center should be equipped with complete resource conditions, including water and electricity, heating, sewage, etc. The periphery of the distribution center should be equipped with complete resource conditions [6]. Sufficient heating, water supply, electricity supply, and discharge capacity to provide basic living guarantee for staff's daily life. It also needs enough fuel and energy, and the area should have certain capacity of sewage and waste treatment. At the same time, supply of basic security for the daily operation of the logistics distribution center counts a lot. These factors will not become the limiting conditions for the development of the logistics distribution center in the future but can guarantee the long-term healthy development.

*2.4. Business Environment.* Considering the location and density of logistics distribution center, the distribution of surrounding customers and the structure and layout of nearby logistics industry are the prerequisite for the location of logistics distribution center [7]. The distribution center should carefully investigate the distribution of customers

when selecting the site. At the same time, the coordination and contradiction between the logistics distribution center and other nearby logistics enterprises should be managed by avoiding vicious competition, and making full use of the existing logistics channels of other companies.

*2.5. Natural Environment.* The distribution center is a place where a large number of products gather, and the location should have strong capacity of soil bearing which needs to consider terrain, landform, water quality, climate, and other indicators [8]. For example, it is necessary to select a place with flat terrain and relatively high terrain, and it must have a suitable shape and size, which is suitable for building; meanwhile, it is best to choose a place where the terrain is completely flat. Slightly undulating areas are the second choice. Rectangular shape is the best choice, and irregular shape is not suitable.

*2.6. Cost.* Logistics distribution center location needs to consider costs, mainly transportation costs, operating costs, and infrastructure costs. The construction of logistics nodes generally takes up a lot of land, so the land price will be directly related to the size of logistics node construction [9]. Generally, the construction of logistics nodes takes up numerous lands, so the price will directly affect the construction scale of logistics nodes. At the same time, it needs to pay various taxes and insurance during its operation. Such as property tax, business tax, vehicle and vessel use tax, and personal income tax withheld by employees. Beyond that, there are differences in the amount of partial tax payment in different regions, in which the vehicle and vessel use tax is paid according to the vehicle, and the local regulations are not the same. Once the location of the distribution center is determined and the construction has been completed, it is not easy to relocate, so unreasonable location will make the enterprise pay a long-term cost for mistakes.

*2.7. Information Quality.* Information quality is mainly measured by information infrastructure and information accuracy. In today's information explosion era, the emergence of e-commerce has had a great impact on traditional industries. With the advent of the Internet age, massive data is pouring in, and all kinds of information, such as orders, processing, transportation, etc. [10], must be processed accurately and effectively to organize the division of labor and form a whole integration. In addition, information infrastructure, as a hardware facility, is directly related to the ability to collect and process various data. Inaccurate information will affect judgment, make unreasonable decisions, and cause serious losses. Therefore, it is significant to ensure the accuracy of information, respond to the problems encountered in logistics and distribution randomly, and improve the distribution efficiency.

Based on the analysis above, the evaluation index system of logistics distribution center location is obtained, as shown in Figure 1.

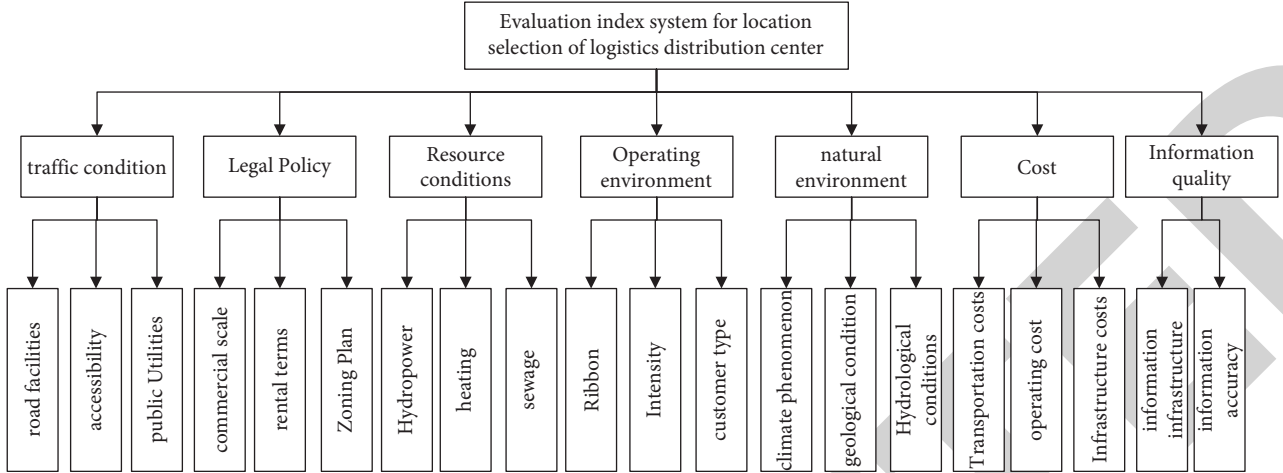


FIGURE 1: Evaluation index system of logistics distribution center location.

### 3. K-Means of Location of Logistics Distribution Center

**3.1. K-Means Algorithm.** Macqueen put forward K-means algorithm in 1967, which is also called Fast Clustering Method. The main idea is to gather each sample into its most similar subclass. The similarity is generally measured by the distance between the sample and the centroid of the class, which is generally the Euclidean distance [11]. The principle of K-means algorithm is generally described as follows:  $k(k \leq n)$  as the parameter,  $n$  objects are divided into  $k$  classes, so that there is a high degree of similarity within classes, while the similarity between classes is low.

Supposing  $Z = \{x_1, x_2, \dots, x_n\}$  is a collection of  $n$  sample points, samples  $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$  are composed of  $m$  attributes or features.  $A = \{a_1, a_2, \dots, a_m\}$ . K-means is the formula (1) to minimize the nonconvex function  $F$  with constraint conditions and to obtain the division of  $Z$  consisting of  $K$  classes, that is, dividing  $Z$  into  $K$  classes, and then minimizing the sum of squares of the distance from each sample to the center of the membership cluster. This optimization can be described as

$$F(W, U) = \sum_{i=1}^K \sum_{j=1}^n w_{ij} d(x_j, u_i). \quad (1)$$

Constraints need to be met:

$$\left\{ \begin{array}{l} d(x_i, u_t) = \sum_{j=1}^m a_j(x_i, u_t), \\ w_{li} \in \{0, 1\}, 1 \leq l \leq K, 1 \leq i \leq n, \\ \sum_{l=1}^K w_{li} = 1, 1 \leq i \leq n, \\ 0 < \sum_{i=1}^n w_{li} < n, 1 \leq l \leq K, \end{array} \right. \quad (2)$$

where  $w_{li}$  is a binary variable as 0 or 1.  $W$  is the matrix of membership between sample  $x_i$  and each class;  $W = [w_{li}]$  is

$K$ -order  $\{0,1\}$  matrix;  $n$  represents the attribute dimension of the sample;  $a_j$  represents the  $j$ -th attribute of the sample;  $u_l$  is the center of class  $l$ ,  $u_l = [u_{l1}, u_{l2}, \dots, u_{lm}]$  is composed of  $M$  components;  $U$  is the class center matrix,  $U = [u_1, u_2, \dots, u_k]$ . In formula (2),  $d(x_i, u_l) = \sum_{j=1}^m a_j(x_i, u_l)$  is used to calculate the dissimilarity measure between samples  $x_i$  and class center  $u_l$ ,  $a_j(x_i, u_l)$  represents the difference value on samples  $x_i$  and class center  $u_l$  in attribute  $a_j$ . If  $a_j$  is of numeric attribute, then  $a_j(x_i, u_l) = \|x_{ij} - u_{lj}\|^2$ . At this point,  $d$  becomes measure of Euclidean distance.

**3.2. Effectiveness of Clustering.** It is used to measure whether the result produced by clustering algorithm reaches the optimal standard, and the classification number under the result of optimal clustering is taken as the optimal clustering number. With reference to the effective index of clustering, the ratio BWACR between the minimum value of the average cosine value of the inter-class included angle and the average cosine value of the intraclass included angle is used to evaluate the results of K-means algorithm [12, 13]. Finally, the best clustering result is determined, that is, the number of sites.

Assuming that  $n$  sample points are clustered into  $K$  classes, and  $2 = K \leq 5$ , it defined as follows:

**Definition 1:** the average cosine value of the minimum included angle between classes at the  $i$ th sample point of class  $l$  is  $bc(l, i)$ .

**Definition 2:** The average cosine value of the included angle within the class at the  $i$ th sample point of class  $L$  is  $wc(l, i)$ .

**Definition 3:** The cluster validity index BWACR of the  $i$ th sample point of the first class is the ratio of the minimum value of the average cosine value of the interclass angle to the average cosine value of the intraclass angle BWACR( $l, i$ ).

**Definition 4:** Define the BWACR average value of all sample points under the cluster number  $K$  as  $avg_{BWACR}(K)$ .

The mathematical formula of BWACR index is as follows:

$$K_{\text{opt}} = \max_{2 \leq K \leq S} \{ \text{avg}_{\text{BWACR}}(K) \}, \quad (3)$$

which meets the following conditions:

$$\left\{ \begin{array}{l} bc(l, i) = \min_{1 \leq j \leq K} \left( \frac{1}{n_j} \sum_{p=1}^{n_j} \frac{\sum_{q=1}^m x_{pq}^{(j)} x_{iq}^{(l)}}{\sqrt{\sum_{q=1}^m x_{pq}^{(j)}} \sqrt{\sum_{q=1}^m x_{iq}^{(l)}}} \right), \\ wc(l, i) = \frac{1/n_l - 1 \sum_{t=1, t \neq i}^{n_l} \sum_{q=1}^m x_{tq}^{(l)} x_{iq}^{(l)}}{\sqrt{\sum_{q=1}^m (x_{tq}^{(l)})^2} \sqrt{\sum_{q=1}^m (x_{iq}^{(l)})^2}}, \\ \text{BWACR}(l, i) = \frac{wc(l, i)}{bc(l, i)}, \\ \text{avg}_{\text{BWACR}}(K) = \frac{1}{n} \sum_{l=1}^K \sum_{i=1}^{m_l} \text{BWACR}(l, i). \end{array} \right. \quad (4)$$

In which  $J$  and  $i$  represent class labels,  $M$  represents sample attribute dimension,  $x_{pq}^{(j)}$  represents the  $q$ -th attribute of the  $p$ -th sample in class  $j$ ,  $x_{iq}^{(l)}$  represents the  $q$ th attribute of the  $i$ th sample in class  $l$ ,  $n_j$  indicates the number of samples of class  $j$ ,  $x_{tq}^{(l)}$  represents the  $q$ -th attribute of the  $t$  sample in class  $l$ , and  $t \neq i$ ;  $n_l$  indicates the number of samples in class  $l$ . The more dispersed the classes are, the more compact the classes are, which indicates that the clustering effect is better. In formula (4), to verify the effect under different  $k$ , the BWACR average value of all sample points is calculated under different  $k$ , that is, formula (4). Finally, the BWACR average value  $k$  is compared under different, that is, formula (3), to determine the best clustering number  $K$ .

**3.3. D-S Reasoning Method.** Dempster–Shafer evidence theory, referred to as D-S reasoning, is an activity that analyzes the basic attributes of evidence and uses evidence to identify the facts of a case. The D-S evidence method based on fuzzy rules is an uncertain multiattribute evaluation method, which is used to solve the multiattribute evaluation or decision-making problem under incomplete information [8]. Every index evaluation level in D-S evidence method corresponds to a utility, where the quality of the scheme is measured by calculating the comprehensive evaluation utility value of each scheme.

**3.3.1. Basic Assumptions.** Suppose a system with two levels of indicators is established, with one  $Y$  for the top (parent) indicator and  $Z$  for the bottom (child) indicators, that is, the basic indicator, which means that only the parent indicator has no bottom indicator, and  $e_i$  ( $i = 1, \dots, L$ ). Then, the basic index set represents:  $E = \{e_1, e_2, \dots, e_i, \dots, e_L\}$ ; Weight set represents:  $\omega = \{\omega_1, \omega_2, \dots, \omega_i, \dots, \omega_L\}$ ,  $\omega_i$  which

indicates the weight of  $i$ th indicator  $e_i$ , and  $0 \leq \omega_i \leq 1$ ; supposing that an index has  $N$  different evaluation levels, and its set is represented  $H = \{H_1, H_2, \dots, H_n, \dots, H_N\}$ , where  $H_n$  means the  $n$ th evaluation level, and  $H_n$  is better than  $H_{n+1}$ . For example: an evaluation set of certain indicators is expressed by {Best, Good, Average, Poor, Worst}, among which Best is better than Good. The evaluation of indicators  $e_i$  ( $i = 1, \dots, L$ ) known is shown in formula:

$$\begin{aligned} S(e_i) &= \{H_n, \beta_{n,i}\}_{n=1, \dots, N} | i = 1, \dots, L, \\ B &= \{\beta_n, n = 1, \dots, N\}. \end{aligned} \quad (5)$$

In which,  $B$  indicates that the index is evaluated as the confidence set of the grade  $H$ .

Among them,  $\beta_{n,i}$  indicates index  $e_i$  whose confidence level of grade is  $H$ . If  $\sum_{n=1}^N \beta_{n,i} = 1$ , then it means that  $S(e_i)$  is a complete evaluation; if  $\sum_{n=1}^N \beta_{n,i} < 1$ , then evaluate  $S(e_i)$  is incomplete; if  $\sum_{n=1}^N \beta_{n,i} = 0$ , it means that the information of  $S(e_i)$  is completely missing.

**3.3.2. Index Combination Algorithm.** First, standardize the weights and make the sum of weights 1, as shown in formula:

$$1 = \sum_{i=1}^L \omega_i, \quad (6)$$

$$\begin{aligned} m_{n,i} &= \omega_i \beta_{n,i} | n = 1, \dots, N; i = 1, \dots, L, \\ m_{H,i} &= 1 - \sum_{n=1}^N m_{n,i} = 1 - \omega_i \sum_{n=1}^N \beta_{n,i} | i = 1, \dots, L. \end{aligned} \quad (7)$$

Among them,  $m_{n,i}$  is the basic probability assignment, and subindex  $e_i$  supports parent indicator  $y$  which is evaluated as degree of  $H_n$ ;  $m_{H,i}$  is the unassigned probability that the index is not assigned after synthesis.  $m_{H,i}$  is split into  $\bar{m}_{H,i}$  and  $\tilde{m}_{H,i}$  two parts, as shown in formulas (8) and (9):

$$\bar{m}_{H,i} = 1 - \omega_i | i = 1, \dots, L, \quad (8)$$

$$\tilde{m}_{H,t} = \omega_i \left( 1 - \sum_{n=1}^N \beta_{n,i} \right) | i = 1, \dots, L. \quad (9)$$

where  $E_I(i) = \{e_1, \dots, e_i\}$  represents a subset of the first  $i$  subindicators;  $m_{n,I(i)}$  for  $E_I(i)$ . All  $I$  indicators support  $E$  evaluation as follows  $H_n$  Degree;  $m_{H,I(i)}$  is the residual probability of all evaluated subindicators in  $E_I(i)$ .

When  $i = 1$ , as shown in formulas (10) and (11):

$$m_{n,I(i)} = m_{n,1}, \quad (10)$$

$$m_{H,I(j)} = m_{H,1}. \quad (11)$$

Then, the coefficients are obtained by using the following iterative formula.  $m_{n,I(L)} \diamond \bar{m}_{H,L(L)}, \tilde{m}_{H,I(L)}$ .

$$\begin{aligned}
K_{I(i+1)} &= \left[ 1 - \sum_{t=1}^N \sum_{\substack{j=1 \\ j \neq t}}^N m_{t,l(i)} m_{j,i+1} \right], \\
m_{n,l(t+1)} &= K_{I(i+1)} \left[ m_{n,l(i)} m_{n,l+1} + m_{H,l(i)} m_{n,l+1} + m_{n,l(t)} m_{H,l+1} \right] n = 1, \dots, N, \\
\tilde{m}_{H,l(i+1)} &= K_{I(i+1)} \left[ \tilde{m}_{H,l(i)} \tilde{m}_{H,i+1} + \bar{m}_{H,l(i)} \tilde{m}_{H,i+1} + \tilde{m}_{H,l(t)} \bar{m}_{H,i+1} \right], \\
\bar{m}_{H,l(i+1)} &= K_{I(i+1)} \bar{m}_{H,l(i)} \bar{m}_{H,i+1}, \\
m_{H,l(i)} &= \tilde{m}_{H,l(i)} + \bar{m}_{H,l(i)} i = 1, \dots, L.
\end{aligned} \tag{12}$$

The confidence of parent index evaluation combination can be calculated by formulas (13) and (14):

$$\beta_n = \frac{m_{n,l(L)}}{1 - \bar{m}_{H,l(L)}} n = 1, \dots, N, \tag{13}$$

$$\beta_H = \frac{\tilde{m}_{H,l(L)}}{1 - \bar{m}_{H,l(L)}}. \tag{14}$$

Finally, the result is shown in formula (15), and the value that  $y$  is rated as  $\beta_n (n = 1, \dots, N)$  in grade  $H_n$  is

$$S(y) = \{(H_n, \beta_n), n = 1, 2, \dots, N\}. \tag{15}$$

**3.3.3. Rank Several Alternatives.** According to the above formulas and steps, the alternative points in each scheme set are evaluated, and the advantages and disadvantages of the alternative schemes are ranked by the utility function theory [1, 14].

Suppose there are two schemes.  $a, b$ ,  $u(H_n)$  is the utility function of grade  $H_n$ ,  $H_{n+1}$  overmatch  $H_n$ , that is  $u(H_{n+1}) > u(H_n)$ .

If all evaluations are complete, then  $\beta_H = 0$ ; formula (16) is used to calculate the expected utility of the parent index, and then advantages and disadvantages of the schemes.  $a, b$  are compared. if and only if  $u(y(a)) > u(yb)$  scheme  $a$  is better than scheme  $b$ :

$$u(y) = \sum_{n=1}^N \beta_n u(H_n). \tag{16}$$

If any subindex evaluation is incomplete, then  $\beta_H > 0$ ,  $[\beta_n, (\beta_n + \beta_H)]$  that is,  $y$  is rated as confidence interval of  $H_n$ . In this case, three parameters are defined to describe the evaluation of  $y$ , namely, minimum utility, maximum utility, and average expected utility, and their calculations are shown in formulas:

$$u_{\max}(y) = \sum_{n=1}^{N-1} \beta_n u(H_n) + (\beta_n + \beta_H) u(H_N), \tag{17}$$

$$u_{\min}(y) = (\beta_1 + \beta_H) u(H_1) + \sum_{n=2}^N \beta_n u(H_n), \tag{18}$$

$$u_{\text{avg}}(y) = \frac{u_{\max}(y) + u_{\min}(y)}{2}. \tag{19}$$

Among them, the utility of  $H_1$  is the lowest, and the utility of  $H_N$  is the highest.

To evaluate of parent index  $y$ , if and only if  $u_{\min}(y(a)) > u_{\max}(y(b))$ , scheme  $a$  is better than scheme  $b$ ; if and only if  $u_{\min}(y(a)) = u_{\min}(y(b))$  and  $u_{\max}(y(a)) = u_{\max}(y(b))$ , scheme  $a$  and scheme  $b$  are similar. Otherwise, the ranking is generated according to the average expected utility.

## 4. Location Analysis of Logistics Distribution Center Based on K-Means

**4.1. Basic Data.** If a retail company wants to choose a local logistics distribution center, after on-the-spot investigation, 18 alternative site selection schemes are preliminarily determined [15, 16]. In order to make the subsequent numerical processing and calculation more concise and easier to understand, the coordinates of the geographical position are transformed into two-dimensional coordinates in rectangular coordinate system, then the rectangular coordinates of the candidate points are obtained, as shown in Table 1.  $(\alpha_i, \beta_i)$  represents the coordinates of the  $i$ th alternative point in MapInfo;  $(a_i, b_i)$  represents the Cartesian coordinate of the  $i$ th alternative point,  $A_i$  represents the alternative point [17].

**4.2. Preliminary Location Based on K-Means.** According to the distance information between candidate points, the candidate points are divided into  $K$  classes,  $2 = K \leq 5$ , and the specific flow is as follows:

- (1) Initialize  $k$  cluster centers randomly.
- (2) Calculate the distance between each data point and the class center, namely  $d(x_i, u_l)$ , and aggregate it into the class closest to the point (principle of nearest neighbor).
- (3) Calculate  $K$  new class centers, and the coordinates of each class center are the average coordinates of all data points in the class; and calculate the measure function value  $f$  at this time, that is, the formula (1) in the core algorithm.

TABLE 1: Location coordinate information of alternative points.

Alternative point	$\alpha$ (ft)	$\beta$ (ft)	$\alpha$ (cm)	$b$ (cm)
$A_1$	1,238.98	-155.1415283	4.5	-1.3
$A_2$	1,198.48	-92.415482	3.992911912	-1.884816128
$A_3$	1,185.28	-233.1626959	3.944976682	-1.954195812
$A_4$	982.1491682	-159.4385181	3.533818915	-1.336859633
$A_5$	756.8486851	-131.8664663	2.851118146	-1.198295832
$A_6$	133.8193184	-111.1861832	2.313915184	-1.931441935
$A_7$	562.4184537	-224.2815433	2.144366112	-1.881483322
$A_8$	688.8581426	-248.4481183	2.463621138	-2.183191988
$A_9$	912.9452856	-335.2591255	3.282165694	-2.811198586
$A_{10}$	886.6215149	-391.3122121	2.822981818	-3.281111489
$A_{11}$	585.485953	-368.1268481	2.12821523	-3.188312823
$A_{12}$	434.9951949	-318.8858182	1.581188161	-2.682888513
$A_{13}$	622.8349332	-489.1128649	2.263988112	-4.111154314
$A_{14}$	628.2282818	-616.5838564	2.289946823	-5.169789498
$A_{15}$	543.844362	-586.9139218	1.976486416	-4.921111536
$A_{16}$	461.3585958	-652.8381538	1.688118843	-5.48394198
$A_{17}$	495.411286	-711.2891413	1.611898812	-5.881142653
$A_{18}$	328.4434383	-742.9456862	1.193888196	-6.229488292

- (4) Stop clustering until the maximum iteration number  $m$  is met or satisfied  $|F_1 - F_2| \leq \epsilon$ . In K-means algorithm,  $M$  is selected as 200, 300, and 400, respectively.

$M$  is the value determined according to the size of the data volume, not the larger the better. When solving practical problems, several values can be selected for comparison. When the clustering results under two values are the same, one of them can be arbitrarily selected as the maximum number of iterations.  $\epsilon$  is a minimum number;  $F_1$  and  $F_2$  represent the measure function value of two iterations.

When  $M$  is 300 and 400, the clustering result is the same, so in this paper, the maximum iteration number  $M$  is 400.

**4.3. Analysis of Results.** The BWACR index is used to analyze the clustering validity of the results with different clustering numbers of  $K$  values. Because the attributes of the sample points, that is, candidate points, are expressed by their two-dimensional position coordinates, the cosine values of the included angle between classes and the cosine values of the included angle within classes are calculated according to formula (4), and finally, the corresponding  $K$  value is obtained as the best clustering number when the average value of cosine value ratio of included angle between classes within classes is the maximum.

The relationship between different values of  $K$  and BWACR is shown in Figure 2.

It can be seen from Figures 4-1 that, when  $2 = K \leq 5$ , the value of BWACR reaches the maximum when  $K = 5$ , and at this time, the balance point between the intraclass dispersion degree and the intra class compactness can be found, so that the two can achieve the best effect, and the alternative points can be divided into five categories, namely:  $M_1 = \{A_1, A_2, A_3, A_4\}$ ,  $M_2 = \{A_5, A_6, A_8\}$ ,  $M_3 = \{A_7, A_9, A_{10}, A_{11}, A_{12}\}$ ,  $M_4 = \{A_{13}, A_{14}, A_{15}\}$ ,  $M_5 = \{A_{16}, A_{17}, A_{18}\}$ .

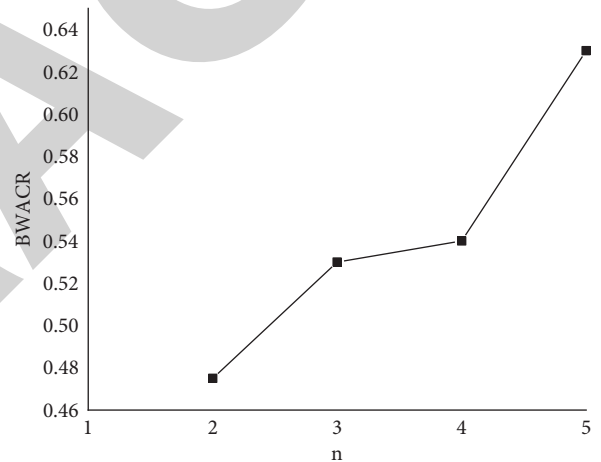


FIGURE 2: Relationship between K and BWACR index value.

That is to say, the best location number of logistics distribution center is 5.

**4.4. Comprehensive Evaluation of Location of Logistics Distribution Center.** Through the analysis above,  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$ , and  $M_5$  are the sets of candidate schemes, respectively, and the D-S reasoning method is used to select the best candidate point from each candidate scheme as the location of the logistics distribution center. According to the comprehensive index system, weight, attribute, and evaluation range of project's location, the quantitative attribute is integrated into the former attribute according to the principle of equivalence by using the D-S evidence method. According to the difference between the evaluation levels of basic attributes and subattributes, and on the basis of the relationship of mapping transformation, the utility evaluation values of each option to be selected are obtained.

According to the above process and formulas (3)–(19) in the D-S evidence algorithm, the comprehensive utility

TABLE 2: Comprehensive utility evaluation and ranking results of alternative points under each scheme set.

Scheme collection	Alternative point	Comprehensive evaluation of utility value	Sort	Best solution
$M_1$	$A_1$	0.5521	4	$A_3$
	$A_2$	0.5951	3	
	$A_3$	0.6892	1	
	$A_4$	0.6113	2	
$M_2$	$A_5$	0.5862	2	$A_8$
	$A_6$	0.5784	3	
	$A_8$	0.6228	1	
$M_3$	$A_7$	0.5556	5	$A_{10}$
	$A_9$	0.5924	3	
	$A_{10}$	0.6465	1	
	$A_{11}$	0.6440	2	
$M_4$	$A_{12}$	0.5621	4	$A_{15}$
	$A_{13}$	0.6153	2	
	$A_{14}$	0.6106	3	
$M_5$	$A_{15}$	0.6402	1	$A_{18}$
	$A_{16}$	0.6712	2	
	$A_{17}$	0.6051	3	
	$A_{18}$	0.6805	1	

TABLE 3: Utility evaluation and ranking results of secondary indicators in scheme  $M_1$ .

Scheme collection	$M_1$				
Parent indicator	Best solution	$A_1$	$A_2$	$A_3$	$A_4$
Operating environment	$A_3$	0.4748	0.4328	0.5788	0.4589
Sort		2	4	1	3
Traffic condition	$A_3$	0.4728	0.5476	0.6964	0.5928
Sort		4	3	1	2
Natural environment	$A_3$	0.5146	0.5484	0.5932	0.5148
Sort		3	2	1	3
Cost	$A_4$	0.7277	0.8681	0.7401	0.9022
Sort		4	2	3	1
Information quality	$A_3$	0.6021	0.6126	0.6732	0.5733
Sort		3	2	1	4
Resource conditions	—	0.7192	0.7192	0.7192	0.7192
Sort		1	1	1	1
Legal policy	—	0.5	0.5	0.5	0.5
Sort		1	1	1	1

evaluation and ranking results of each alternative point under each scheme set are shown in Table 2.

It can be seen from Table 2 that in each scheme set of  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$ , and  $M_5$ . The comprehensive utility values of  $A_3, A_8, A_{10}, A_{15}, A_{18}$  are the highest, which are 0.6892, 0.6228, 0.6465, 0.6402, and 0.6806, respectively. Specifically,  $A_3$  has the largest value, which is the best scheme.

The ranking result of utility value of each alternative point in each scheme according to a certain index attribute can be obtained. The utility evaluation and ranking result of secondary index in scheme  $M_1$  are shown in Table 3.

From Table 3, it can be seen that the operating environment, traffic conditions, and natural environment of  $A_3$

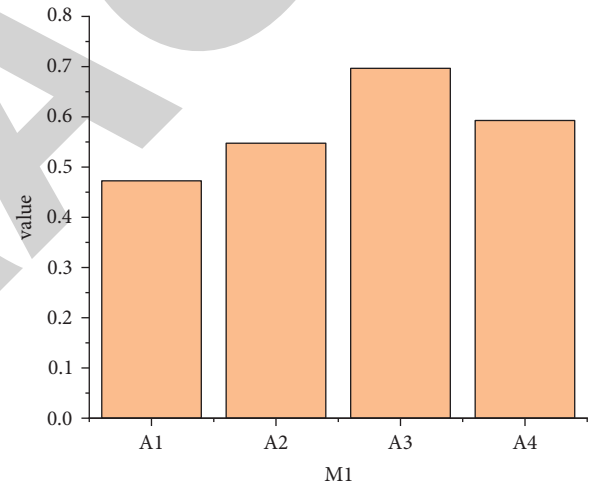


FIGURE 3: Utility values of various schemes in  $M_1$  under traffic condition attributes.

are the best compared with other locations in the scheme set, and the cost of  $A_4$  is lower than that of other locations, while the resource conditions and legal policies of each location in the scheme set are the same.

In addition, we can also get the ranking of schemes under a certain secondary index, for example, the ranking results of  $A_1, A_2, A_3$ , and  $A_4$  are 2, 4, 1, and 3 under the traffic condition index. As in the scheme slave  $M_1$ , under the traffic condition, the ranking results of utility values of  $A_1, A_2, A_3, A_4$  are shown in Figure 3, which are 0.4728, 0.5476, 0.6964, and 0.5928, respectively. It can be seen that if only the traffic condition attributes are considered, the ranking result of utility values from high to low in  $M_1$  scheme set is as follows  $A_3, A_4, A_2, A_1$ , and the optimal location is  $A_3$ .

Therefore, the best location in logistics distribution can be obtained by the K-means algorithm.



## 5. Conclusion

Distribution center is an important part of logistics and plays an important role in the logistics system. The location of distribution center has great influence on logistics cost and transit time. Based on the K-means clustering algorithm, the location of logistics distribution center is studied, and the evaluation system of location based on traffic conditions, laws and policies, resource conditions, business environment, natural environment, cost and information quality is constructed. On this basis, K-means clustering algorithm and D-S evidence method are used to determine the location of logistics distribution centers scientifically and efficiently. Through the combination of the two methods, the uncertainty in the decision-making process can be fully considered, and the results obtained are closer to the reality. At the same time, the method can also effectively solve the problem of multiattribute evaluation. When the weights of factors affecting site selection in different periods of real life are different, the new comprehensive attribute evaluation can be obtained by directly changing the weight of attributes at any time. Therefore, the location of logistics distribution center based on the K-means clustering algorithm is practical, scientific, and effective.

## Data Availability

The dataset can be accessed upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was funded by 1. 2020 Science and Technology Youth Project “Research on the Development Model of Wanzhou Rural Eco-industrialization under the Rural Revitalization Strategy” of Chongqing Municipal Education Commission, the project number is KJQN202003507. 2. 2020 Science and Technology Research Project “Research on the Path of Local Higher Vocational Colleges Serving Rural Revitalization - Taking Chongqing Three Gorges Vocational College as an Example” of Chongqing Three Gorges Vocational College, the project number is cqsx202020. 3. 2020 Higher Education and Teaching Reform Research Project “The Exploration and Practice of “Course Certificate Integration” Model of Higher Vocational Logistics Management Major in Chongqing under the “1 + X” Certificate System” of Chongqing Municipal Education Commission, the project number is 203625. 4. The Scientific Research Planning project “Research on the Exam Mode of “Course, Competition and Certificate Integration” of Higher Vocational Logistics Management Major in Southwest China under the “1 + X” Certificate System” of Chongqing Higher Vocational Technical Education Research Association, the project number is GY201014.

## References

- [1] S. Kadry, G. Alferov, and V. Fedorov, “D-star algorithm modification,” *International Journal of Online and Bio-medical Engineering (iJOE)*, vol. 16, no. 08, pp. 169–172, 2020.
- [2] X. Li and K. Zhou, “Multi-objective cold chain logistic distribution center location based on carbon emission,” *Environmental Science and Pollution Research*, vol. 28, no. 25, pp. 32396–32404, 2021.
- [3] P. Liu and Y. Li, “Multiattribute decision method for comprehensive logistics distribution center location selection based on 2-dimensional linguistic information,” *Information Sciences*, vol. 538, no. prepublish, pp. 231–234, 2020.
- [4] L. Lu and J. Qin, “Multi-regional logistics center location algorithm based on improved K-means clustering,” *Computer system application*, vol. 28, no. 8, pp. 251–255, 2019.
- [5] G. Musolino, C. Rindone, A. Polimeni, and A. Vitetta, “Planning urban distribution center location with variable restocking demand scenarios: general methodology and testing in a medium-size town,” *Transport Policy*, vol. 80, no. C, p. 345, 2018.
- [6] M. Özmen and E. Kızılkaya Aydoğan, “Robust multi-criteria decision making methodology for real life logistics center location problem,” *Artificial Intelligence Review: An International Science and Engineering Journal*, vol. 53, no. 12, pp. 331–335, 2020.
- [7] R. Rajesh Sharma and A. Sungheetha, “Analysis of influencing factors of logistics center location based on comprehensive quality training,” *Indian Journal of Public Health Research & Development*, vol. 4, no. 12, pp. 289–293, 2018.
- [8] H. Lei and S. Li, “An improved algorithm of D-S evidence fusion,” *Journal of Physics: Conference Series*, vol. 1871, no. 1, pp. 237–240, 2021.
- [9] Bienvenido-Huertas David, M.-G. David, J. Carretero-Ayuso Manuel, and E. Rodríguez-Jiménez Carlos, “Climate classification for new and restored buildings in Andalusia: analysing the current regulation and a new approach based on K-means,” *Journal of Building Engineering*, vol. 43, pp. 145–148, 2021.
- [10] M. Mieczysłńska and I. Czarnowski, “K-means clustering for SAT-AIS data analysis,” *WMU Journal of Maritime Affairs*, vol. 20, no. 3, pp. 164–168, 2021.
- [11] U. S. Mukminin, B. Irawanto, B. Surarso, and Farikhin, “Fuzzy time series based on frequency density-based partitioning and K-means clustering for forecasting exchange rate,” *Journal of Physics: Conference Series*, vol. 1943, no. 1, pp. 645–678, 2021.
- [12] A. Misganaw, Y. Noh, C. Seo, D. Kim, and I. Lee, “Developing a ship collision risk index estimation model based on Dempster-Shafer theory,” *Applied Ocean Research*, vol. 113, p. 1162, 2021.
- [13] M. Jabi, M. Pedersoli, A. Mitiche, and I. B. Ayed, “Deep clustering: on the link between discriminative models and K-means,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1887–1896, 2021.
- [14] S. Wu, S. Yang, and X. Du, “A model for evaluation of surrounding rock stability based on D-S evidence theory and error-eliminating theory,” *Bulletin of Engineering Geology and the Environment*, pp. 489–496, 2021.
- [15] W. Cao, Z. Zhang, and C. Liu, “Unsupervised discriminative feature learning via finding a clustering-friendly embedding

- space,” *Pattern Recognition*, vol. 129, Article ID 108768, 2022.
- [16] Y. Zhai, S. Lu, and Q. Ye, “Ad-cluster: augmented discriminative clustering for domain adaptive person re-identification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9021–9030, Seattle, WA, USA, June 2020.
- [17] B. Lyu, W. Wu, and Z. Hu, “A novel bidirectional clustering algorithm based on local density,” *Scientific Reports*, vol. 11, no. 1, p. 14214, 2021.

RETRACTED