Multi-level Logistics Network Node Siting Model Based on K-Means

Jie Liu[1], Shuang Tian[2], Qingqing Wang[3], Chenguang Zhang[1](✉), Min Ning[1], and Changlong Li[1]

[1] Beijing Aerocim Technology Corporation Limited Ltd., Beijing, China
z126314032@126.com
[2] Beijing Institute of Control and Electronic Technology, Beijing, China
[3] Beijing Aerospace ChangFeng Co., Ltd., Beijing, China

Abstract. Artificial intelligence plays a pivotal role in global logistics and supply chain management. AI enables real-time optimization for routes and minimizes wastage while increasing delivery efficiency. Companies can leverage AI in logistics to optimize resources and be more efficient. Currently, the node siting of logistics transportation networks faces two challenges. (1) The unreasonable selection of upstream and downstream nodes on the logistics transportation network can result in the wastage of transportation costs. (2) The siting of nodes in logistics transportation networks just takes the effect of geography into account and ignores other aspects that may have an impact on transportation efficiency. To tackle these challenges, we design a multi-level logistics network node siting model based on K-Means. Firstly, we adopt a **space first and then attribute** clustering strategy to cluster according to the spatial coordinates of each community to obtain the initial position of each distribution node, and optimize the siting based on the weight of express delivery in each community. Secondly, we determine the location of transit nodes and large distribution centers in turn according to the siting of distribution nodes. Taking Erdao District, Changchun City as an example, we finally build a logistics transportation network composed of 3 large distribution centers, 11 transit nodes and 42 distribution nodes using this model. This model can effectively solve the problem of node siting in multi-level logistics networks and provide a reference for future siting planning of logistics enterprises .

Keywords: Artificial Intelligence · K-Means · Weighted Average · Silhouette Coefficient

1 Introduction

The rapid growth in demand for express delivery due to the continuous development of e-commerce has led to an increase in logistics transportation network nodes and logistics transportation chains [1]. However, manually selecting the location of transportation nodes is no longer sufficient to meet current logistics needs. Therefore, artificial intelligence technology is integrated into modern logistics systems to achieve intelligent siting.

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 437–444, 2023.
https://doi.org/10.2991/978-94-6463-222-4_47

This integration can minimize the length of logistics transportation routes, reduce logistics transportation costs, and improve the operational efficiency of logistics systems. The integration of artificial intelligence technology into modern logistics systems has become an important research topic in recent years due to its potential to improve the efficiency and effectiveness of logistics systems [2, 3].

Nowadays, the siting process of logistics transportation network nodes requires addressing two problems. The first problem is to achieve optimal siting from the overall architecture of the logistics transportation network. Logistics transportation network nodes in a city can be roughly divided into three levels: the upstream network consists of large distribution centers, the midstream network consists of transit nodes and the downstream network consists of distribution nodes. Express delivery from several transit nodes is managed by a large distribution center. As a result, the location of big distribution centers is influenced by the dispersion of transit nodes. The same is true for the connections between communities, distribution nodes, and transit nodes. The second problem is that the siting of distribution nodes is influenced by two factors: spatial factors and attribute factors. The location coordinates of communities are spatial data, and the express delivery volume of communities is attribute data. Transportation expenses will rise if the distribution node is far away from each community. Also, if a community as a whole has a high express delivery volume, the courier may deliver it in many batches. Therefore, it is an important challenge to design an algorithm for global optimization of multi-level logistics transportation network siting under the condition of comprehensive consideration of space factors and attribute factors.

Many existing methods have been applied to solve the siting problem of logistics transportation network nodes. They can be roughly divided into two categories: heuristic algorithms and comprehensive decision-making models. Wang X et al. used genetic algorithms to optimize the location of the original logistics storage center [4]. Huang Y et al. designed a siting model for logistics centers based on a particle swarm optimization algorithm [5]. The above heuristic algorithms have high time complexity. Yazdani M et al. designed a comprehensive decision-making model to determine the siting of logistics centers [6]. Özmen M et al. designed a multi-criteria decision-making method to solve the problem of logistics center location in real life [7]. However, heuristic algorithms and comprehensive decision-making models have not taken into account the effect of express delivery volume on logistics efficiency or the unreasonable site choice of upstream and downstream nodes.

To address this challenge, we design a multi-level logistics network node siting model based on K-Means. From the perspective of the overall architecture, we use a **bottom-up** approach to build the logistics transportation network layer by layer. From the perspective of multiple influence factors, we adopt a space first and then attribute clustering strategy based on K-Means [8, 9].

2 Symbol Description

In this paper, the symbol description is shown in Table 1. The logistics transit nodes in a region are generally divided into three levels: large distribution centers, transit nodes, and distribution nodes. Considering the geographical location of each community in the

Table 1. Symbol Description

Serial Number	Symbol	Symbol Description
1	n	The number of communities
2	k	The number of distribution nodes
3	k_{max}	The maximum number of distribution nodes
4	k_{min}	The minimum number of distribution nodes
5	k_{best}	The optimal number of distribution nodes
6	num_A	The population of distribution node A
7	p_i	The population of the community i
8	w_i	The population weight of the community i
9	$(\overline{x}, \overline{y})$	The optimized coordinates of distribution node A
10	x_i	The horizontal coordinate of the community i
11	y_i	The vertical coordinate of the community i

region and the volume of express delivery, a logistics transportation network is planned for the region and determined the siting of each large distribution center, transit node, and distribution node.

3 Method and Experiment

As mentioned in Chapter 1, the siting of the distribution nodes is determined first based on spatial data. Subsequently, we further optimize the siting of the distribution nodes based on attribute data. According to this strategy, the location of the transit node and large distribution center are determined in turn. The model flow chart is shown in Fig. 1.

3.1 Locate Distribution Nodes

To make the location of distribution nodes more reasonable, we utilize the clustering strategy of space first and attribute second to determine the location of distribution

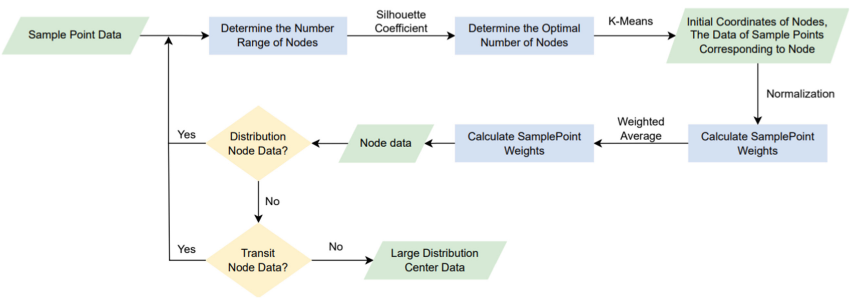

Fig. 1. Model flow chart

nodes. Firstly, the K-Means clustering algorithm is used to determine the initial location of distribution nodes and the corresponding distribution nodes of each community. Then, the location of distribution nodes is further optimized according to the logistics quantity weight of each community.

Find the Initial Location of the Distribution Node.
Due to the significant impact of the number of distribution nodes (k) on the clustering effect of the K-Means algorithm, we first determine the range of k and then evaluate the clustering effect of each k using the **silhouette coefficient** method to determine the most reasonable number of distribution nodes [10]. Then we use the K-Means algorithm to cluster and obtain the initial position of each distribution node and the corresponding distribution nodes of each community.

a. Identify the k Range: According to the dataset, there are n communities in a region. Based on the principle that each distribution node is responsible for 3 to 5 communities, the maximum value k_{max} and the minimum value k_{min} can be determined according to formula (1). To evaluate as many k as possible, we use the principle of rounding down for k_{max} and rounding up for k_{min}.

$$\begin{cases} k_{max} = \frac{n}{3} \\ k_{min} = \frac{n}{5} \end{cases} \tag{1}$$

b. Identify the Ideal k: We calculate the silhouette coefficient for each k and select the k with the largest silhouette coefficient as the k_{best}, which represents the most reasonable number of distribution nodes.
c. Identify the Initial Location of Distribution Node: To obtain the coordinate of each cluster center, which represents the coordinate of each distribution node, we use each community as a sample point and k_{best} as the number of clusters for K-Means clustering. We also obtain the cluster label number corresponding to each sample point.

Optimization the Location of Distribution Node.
Since the number of logistics is correlated with the number of people, we optimize the distribution node location coordinates using the number of people to replace the number of logistics. We take the distribution node A as an example, and the optimization process is as follows:

a. Get the Population Weights of Each Community: According to the previous clustering results, we can obtain all the communities that A is responsible for. The total population of all communities under the control of distribution node A is marked num_A, the population of the i-th community is marked p_i, and the population weight of the i-th community is marked w_i. We use the **normalization** approach to calculate the population weight of each community according to formula (2).

$$w_i = \frac{p_i}{num\_A} \tag{2}$$

b. Optimize Coordinates: We mark the optimized coordinates of A as $(\overline{x}, \overline{y})$, the horizontal coordinate of the i-th community as x_i and the vertical coordinate of the i-th community as y_i. The concept of the **weighted average** is employed to determine the optimum coordinates of distribution node A in accordance with formula (3), which is based on the principle that the more populated communities are closer to the distribution node.

$$\begin{cases} \overline{x} = \sum w_i \times x_i \\ \overline{y} = \sum w_i \times y_i \end{cases} \tag{3}$$

Finally, we get the optimized coordinate for 42 distribution nodes.

3.2 Locate Transit Nodes

Similar to Sect. 3.1, we use the optimized distribution nodes obtained in Sect. 3.1 as sample points. We determine the maximum and minimum value of transit nodes according to the principle that each transit node governs distribution nodes for 3 to 5. Then, we use the silhouette coefficient method to determine the most reasonable number of transit nodes and obtain the initial positions of each transit node through K-Means algorithm.

We believe that *num_A* is the population of distribution node A, and we use the weighted average concept to get the optimal position coordinates of each transit node.

Finally, we locate 11 transit nodes.

3.3 Locate Large Distribution Centers

Similarly, by using each transit node as a sample point, we are able to determine the locations of 3 large distribution centers.

For the Erdao District in Changchun City, we eventually construct a logistics and transportation network that consists of 3 large distribution centers, 11 transit nodes, and 42 distribution nodes. The location coordinates of the transit nodes and large distribution centers are shown in Table 2, and the clustering effects before and after optimization are respectively shown in Fig. 2 and Fig. 3.

Table 2. The location coordinates of the transit nodes and large distribution centers

Node	Horizontal Coordinate	Vertical Coordinate
Transit Node	68.0293	48.0211
	70.1441	49.3498
	69.4946	49.4156
	68.7049	47.8808
	69.3716	47.2551
	68.4021	42.4935
	69.0036	44.1143
	71.9341	57.9490
	69.2881	53.5063
	69.3432	54.5834
	68.9224	53.0179
Large Distribution Center	69.2329	48.1607
	71.9341	57.9490
	69.2057	53.6877

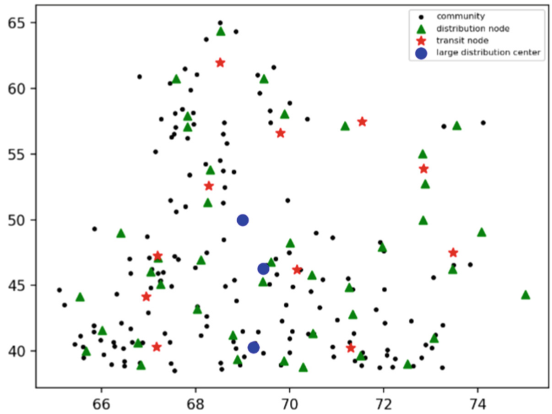

Fig. 2. The clustering effects before optimization

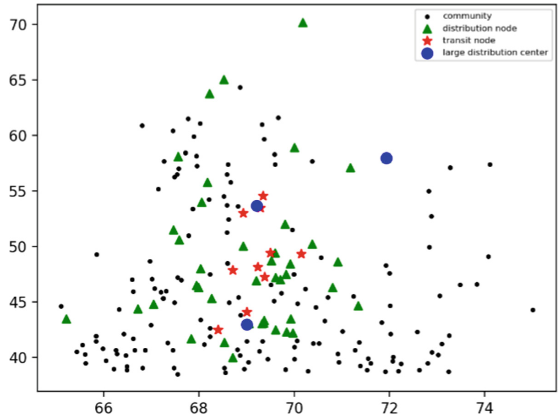

Fig. 3. The clustering effects after optimization

4 Conclusion and Future Work

For multi-level logistics transportation networks, we develop a K-Means-based node siting model that builds a multi-level logistics transportation network with integrated consideration of geographic location and express volume. Our model can be applied to other similar siting problems with appropriate adjustment. However, our model still has some limitations. For instance, our model only estimates the number range of upstream nodes based on the principle that each upstream node governs 3 to 5 downstream nodes. But this constraint is absent when clustering is performed, which leads to a situation where one upstream node may govern 6 to 7 downstream nodes. Our research can help with site selection challenges in the commerce, logistics, and emergency disaster assistance sectors.

References

1. Kayikci Y (2019) E-Commerce in logistics and supply chain management. IGI Global: 1015-1026. doi:https://doi.org/10.4018/978-1-5225-7362-3.ch076
2. Zhou X. (2021) The Probes into the Innovation Direction of Modern Logistics Management Mode in Artificial Intelligence Era. CISAI, IEEE, 2021: 855-859. doi: https://doi.org/10.1109/CISAI54367.2021.00173.
3. Zhao J, Xie F. (2020) Cognitive and artificial intelligence system for logistics industry. International Journal of Innovative Computing and Applications, 11(2-3): 84-88. doi:10.1504/ijica.2020.107118
4. Wang X, Xia Q. (2020) Optimization of logistics warehouse location based on genetic algorithm. Cyber Security Intelligence and Analytics: 745-752. doi:https://doi.org/10.1007/978-3-030-15235-2_102
5. Huang Y, Wang X, Li T, et al. (2022) The Location Selection of Logistics Center in City Based on Particle Swarm Optimization. doi:https://doi.org/10.21203/rs.3.rs-1964001/v1
6. Yazdani M, Chatterjee P, Pamucar D, et al. (2020) Development of an integrated decision making model for location selection of logistics centers in the Spanish autonomous communities. Expert Systems with Applications, 148: 113208. doi:https://doi.org/10.1016/j.eswa.2020.113208

7. Özmen M, Aydoğan E K. (2020) Robust multi-criteria decision making methodology for real life logistics center location problem. Artificial Intelligence Review, 53: 725-751. doi:https://doi.org/10.1007/s10462-019-09763-y
8. Dwivedi S, Bhaiya L K P. (2019) A Systematic Review on K-Means Clustering Techniques.
9. Zhu J, Sun Y, Chen L, Zhou W, Meng Y W. (2017) A Spatial Clustering Method Based on Uneven Distribution of Non-spatial Attributes—Identifying City Commercial Center. Geomatics and Information Science of Wuhan University, 42(12): 1696-1702. doi:10.13203/j.whugis20150590
10. Řezanková H. (2018) Different approaches to the silhouette coefficient calculation in cluster evaluation.21st International Scientific Conference AMSE Applications of Mathematics and Statistics in Economics: 1–10.