CiteSpace-based Logistics Distribution Center Site Selection Research

Li Liu¹, Shiyi Liu²

¹School of Economics and Management, Southwest Petroleum University, Sichuan Chengdu 610500, China
²School of Economics and Management, Southwest Petroleum University, Sichuan Chengdu 610500, China

Abstract: China's logistics industry is developing at a high speed, but the distribution efficiency can't match with it, to reduce the waste of resources and costs on the way of distribution, it is crucial to study and improve the distribution center siting issue. Taking the 494 core journals on the study of logistics distribution center siting in Knowledge Network as the basic data, CiteSpace is used to visualize and analyze the related literature in this research field. Through the keyword clustering analysis, this field can be divided into four directions: the main body of research, the influence factors of site selection, the purpose of site selection, and site selection methods, and the comprehensive analysis shows that the early focus of China's research is on the site selection methods, such as genetic algorithm and center of gravity method, and the future research direction is the joint optimization of the site selection-path and the two-layer planning.

Keywords: Logistics and distribution center site selection; CiteSpace visual analytics; site selection methods.

1. Introduction

In recent years, China's logistics industry has developed rapidly, but the distribution efficiency cannot match it. The problem of destruction of products in the distribution process has caused a huge waste of resources and costs. The factors leading to the loss can be divided into several aspects, such as inappropriate choice of distribution center location, irrational planning of distribution path, cumbersome distribution links, and backward distribution conditions, the most important factor is the inappropriate choice of distribution center location.

In 1909, Alfred Weber [1] for the first time proposed the theory of logistics distribution center location. Distribution center location, that is, there are multiple supply points and multiple demand points in the region, select a certain number and location of the establishment of distribution centers and optimization of distribution centers. The distribution center between the customer and the supplier has the role of the top and bottom. Reasonable site planning can effectively reduce costs promote the harmony of products and services, and to ensure the balance of the entire logistics system. Therefore, how to choose the right logistics distribution center is of research significance.

This paper visually analyzes the core journals in this field in the database from 1998-2023 through bibliometric methods and with the help of the CiteSpace visualization tool. Comprehensive analyses were conducted in terms of the number of articles issued, core research authors and institutions, keyword co-occurrence and clustering, and timeline charts to derive the research hotspots and development trends of this research field in a certain period.

2. Research Methodology and Data Sources

2.1. Research method

To comprehensively explore the research situation of domestic logistics distribution center site selection, this paper adopts the bibliometric analysis method, using Excel and CiteSpace (version 6.1.R6) software to analyze the current status of distribution center site selection research. The combination of the two can visually show the current status of domestic distribution center siting research, research hot spots, and future trends.

2.2. Data Sources

In this paper, "distribution center siting" and "logistics center siting" are used as keywords in China Knowledge Network, and the type of literature selected is "academic journals", and a total of 1,431 articles in Chinese literature from 1998 to 2023 are retrieved. To ensure the comprehensiveness and accuracy of the study, the journal sources were selected as "Beida Core" and "CSSCI", and finally 494 core journals were retrieved, which were exported and used as the research samples.

3. Literature Review

3.1. Literature Review on Distribution Center Site Selection Methods

Site selection methods can be categorized into qualitative and quantitative, and they can be used in combination. Qualitative methods include an expert scoring method and the Delphi method, etc. However, qualitative methods have some defects because they are too subjective in practical application scenarios. Quantitative methods are more objective and include the center of gravity method, mathematical planning method, multi-criteria decision-making method, and so on.

Professor Aiken [2] combined the site selection model with a mathematical planning method, listed the constraints according to the actual situation, established the objective function with the lowest site selection cost, and found the optimal solution. W.J.Stevenso [3] studied the distribution route and warehouse address problem under the uncertainty of demand. Changqing Liu [4] used the center of gravity method to construct a site selection model to finally obtain more reasonable and detailed address coordinates.
In conducting the site selection study, constructed an objective function containing a variety of cost elements and established a site selection model based on the combination of genetic algorithm and MATLAB. Juan Liu [6] et al. used an optimized BP neural network model to learn the distribution center with perfect data and selected the best distribution center site from both qualitative and quantitative perspectives. Xu Xiaoping et al. [7] introduced uncertain variables into the basic wolfpack algorithm, and the simulation shows that the proposed improved wolfpack algorithm can optimize the site selection model.

The problem of distribution center location is very complicated and is affected by the enterprise's own business philosophy and business characteristics, and also by the overall market demand, regional spatial distribution, government policies, and other external environmental factors. Scholars at home and abroad have conducted a large number of studies from different perspectives and proposed different methods, but none of them is universal and needs to be analyzed according to the specific situation. The author believes that in future research work, the use of heuristic algorithms, and the combination of multi-criteria decision-making methods and simulation methods is the general trend. The comprehensive use of a variety of means can be realized to learn from each other, thus improving the accuracy and reliability of site selection.

3.2. Research review on the direction of distribution center site selection

Researchers at home and abroad have conducted detailed studies on the positioning of different main products in distribution centers, and the harvest is quite fruitful. Di Weimin [8] et al. took into account the perishable characteristics of agricultural products, combined with a hybrid genetic algorithm to construct a perishable agricultural product site selection model with limited distribution capacity. Yuan Qun et al. [9] constructed an optimization model for the location of cold chain logistics distribution centers by considering the cost, time window constraints, cargo loss rate, and other factors, and designed a hybrid genetic algorithm model. Wang Haoxiang et al. [10] explored the elements constituting the cost in-depth and constructed a two-layer cost model about the site selection of cold chain logistics centers for pharmaceuticals. Ioannis Giannikos [11] established a multi-objective planning model for the site selection and transportation of hazardous materials based on the total cost, risk, and fairness of risk-taking in the site selection and transportation.

Establishing a siting method based on different attributes of distribution centers is very relevant and can improve the science and feasibility of the siting plan. Combined with the different main products of the distribution center, comprehensive and meticulous all-around site selection consideration, is not only conducive to the improvement of the site selection problem but also conducive to a more in-depth study of the field.

4. Visual analysis of logistics distribution center location research

4.1. Analysis of the amount of literature issued

The amount of literature issued indicates that the research topic in the year research heat and attention degree, the more the literature, the more active the academic community about the research. According to the literature data, the literature distribution map is drawn, and the specific situation is shown in Figure 1.

As can be seen from Figure 1, the domestic research in the field of logistics and distribution site selection started in 1998 and then showed an upward trend until 2013 when it reached a maximum of 46 articles, and then began to show a downward trend. At the beginning of the twentieth century, the state fully encourages domestic and foreign manufacturers to jointly invest in the construction of logistics enterprise infrastructure and to improve the logistics and distribution facilities. In this context, experts and scholars have invested in the study of distribution center siting problems. Foreign research on the location of distribution centers first began in the early 1960s, compared with foreign countries, China's research in this field started late, and the initial progress was slow.

4.2. Analysis of core research authors and institutions

Compared with other non-core research, core authors and institutions are more adept at grasping the research trends, hotspots, and directions in this field. This paper uses the cluster analysis function of CiteSpace to jointly analyze the authors and institutions in the field of distribution center siting research, setting the time slice as "8 years", as shown in Figure 2.

![Figure 1. Distribution of distribution center siting research literature, 1998-2023](image)

![Figure 2. Visual Analytics Map of Core Study Authors and Institutions](image)
A total of 195 nodes and 227 connecting lines were obtained, with a density of 0.012. Therefore, it can be seen that the co-occurrence network density of nodes between authors and institutions in this field is higher, the cooperation relationship between authors or institutions is closer, and a certain scale of cooperation team has been formed and has a greater influence in the field. In terms of the size of the cooperative network, the largest cooperative network is formed by Southwest Jiaotong University, Beijing Jiaotong University, etc.; in terms of the location of the cooperative institutions, it is easier for the institutions in the same region to form a cooperative network, for example, in Hubei Province, a cooperative network has been formed by Huazhong University of Science and Technology, Wuhan University of Science and Technology and so on. In addition to these large cooperative networks, there are many cases of individual research or research by separate institutions. It is difficult to cover all aspects of research by separate individuals and institutions, and it is not possible to do resource integration, so only cooperation between individuals and institutions can complement each other's strengths and give full play to the maximum energy.

Table 1. List of organizations with $\geq 6$ publications

<table>
<thead>
<tr>
<th>Rankings</th>
<th>Organizations</th>
<th>Volume of Publications</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Southwest Jiaotong University</td>
<td>12</td>
<td>2011</td>
</tr>
<tr>
<td>2</td>
<td>Huazhong University of Science and Technology</td>
<td>8</td>
<td>2008</td>
</tr>
<tr>
<td>3</td>
<td>Shanghai Institute of Technology</td>
<td>8</td>
<td>2006</td>
</tr>
<tr>
<td>4</td>
<td>Beijing Jiaotong University</td>
<td>8</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>Southeast University</td>
<td>8</td>
<td>2004</td>
</tr>
<tr>
<td>6</td>
<td>University of Science and Technology for National Defense</td>
<td>6</td>
<td>2008</td>
</tr>
<tr>
<td>7</td>
<td>Central South University</td>
<td>6</td>
<td>2004</td>
</tr>
</tbody>
</table>

Table 2. List of authors with $\geq 3$ publications

<table>
<thead>
<tr>
<th>Rankings</th>
<th>Authors</th>
<th>Volume of Publications</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wang Fei</td>
<td>5</td>
<td>2010</td>
</tr>
<tr>
<td>2</td>
<td>Zhang Yi</td>
<td>4</td>
<td>2006</td>
</tr>
<tr>
<td>3</td>
<td>Wang Xiaobo</td>
<td>4</td>
<td>2006</td>
</tr>
<tr>
<td>4</td>
<td>Gao Zhenhua</td>
<td>4</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>Chen Senfa</td>
<td>4</td>
<td>2005</td>
</tr>
<tr>
<td>6</td>
<td>Yang Maosheng</td>
<td>3</td>
<td>2007</td>
</tr>
<tr>
<td>7</td>
<td>Mao Zedong</td>
<td>3</td>
<td>2009</td>
</tr>
</tbody>
</table>

4.3. Keyword co-occurrence analysis

Keywords are the condensation of the core content and ideas of the articles. According to CiteSpace, the keyword co-occurrence map shown in Figure 3 was obtained, which generated a total of 494 nodes and 743 connecting lines. As shown in Fig. 3, the hot words in this field are "site selection", "genetic algorithm", "center of gravity" and so on.

Table 3. Statistical table of high-frequency keywords (frequency $>20$)

<table>
<thead>
<tr>
<th>Arrange in order</th>
<th>Keywords</th>
<th>Centrality</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Site Selection</td>
<td>0.35</td>
<td>158</td>
</tr>
<tr>
<td>2</td>
<td>Distribution centers</td>
<td>0.39</td>
<td>96</td>
</tr>
<tr>
<td>3</td>
<td>Genetic algorithm</td>
<td>0.33</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>Logistics center</td>
<td>0.17</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>Site Selection Model</td>
<td>0.2</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>Logistics</td>
<td>0.36</td>
<td>29</td>
</tr>
</tbody>
</table>

Based on the keyword highlighting, it is possible to observe the degree of change of a keyword in different years, thus showing the turning point of the study. As can be seen in Figure 4, 17 keywords such as "site selection problem" and "distribution center" are shown. From the perspective of highlighting intensity, the "siting model" has the highest highlighting intensity (3.84), which is the research frontier in the field of distribution center siting research, and the lowest highlighting intensity is the "particle swarm" (1.7), which is probably due to the limitations of the particle swarm algorithm, which is unable to deal with discrete optimization problems well. Possibly because of the limitations of the particle swarm algorithm: it cannot better deal with discrete optimization problems and is easy to fall into the local optimum; from the point of view of the highlighting time, the highlighting of "logistics" lasts the longest, for 8 years, and the highlighting time is the lowest for the "siting problem" and "particle swarm", both of which have been highlighted for a long time. The lowest are "site selection problem" and "particle swarm", both of which have only 1
year of highlighting time.

From 2000 to 2010, people pay attention to "e-commerce", "site selection model", etc. The reason may be that under the rapid development of e-commerce, enterprises realize the importance of correct site selection for logistics and distribution.

From 2010 to 2020, we are concerned about the "optimization model", "carbon emission", etc. During this period, the state puts forward the green development concept of "green water and green mountains are golden silver mountains". The concept of green development. Scholars, when facing the problem of site selection, not only consider the cost and benefit but also start to consider environmental issues such as carbon emission.

Since 2020, we have started to pay attention to the issues of "path optimization" and "two-layer planning", and the combination of site selection and path can better achieve the goal of maximizing benefits. "Path optimization" and "two-tier planning" have been hot research topics in the past five years, and are likely to be the future development trend in this field.

4.4. Keyword Cluster Analysis

Cluster analysis of keywords using CiteSpace software can be used to understand whether there is any commonality in the research hotspots in the field. Two important indicators are generated after clustering, namely, modular value (Q-value) and meaning contour value (S-value) to assess the quality of the effect of clustering. Q-value>0.3 indicates significant community structure, and S-value>0.7 indicates that the clusters are highly concentrated, at which time the model is of practical research significance. The parameters in the clustering model of this paper were Q=0.8645 and S=0.9644, which indicated that the clustering effect was good and reliable. As shown in Figure 6, a total of 20 clusters were obtained. The 20 clusters were analyzed according to these four directions as follows:

(1) Research subject. Site selection has an impact on distribution mode, distance, timeliness economic efficiency, etc. In site selection optimization, the main body of the distribution center should be ensured, fully combining the different directions of the military, medicine, cold chain and emergency response, etc., and using a variety of algorithms such as multi-criteria decision-making methods and heuristic algorithms to select the optimal location of the center.

(2) Site selection influencing factors. The site selection influencing factors can be divided into the following three: the first is economic factors, including upfront land costs, construction costs, transportation costs, and labor costs. The second is the transportation factor, which mainly considers the accessibility of transportation; finally, there are social factors, and the selected site should have a wide range of services, complete public infrastructure, and a high degree of environmental protection.

(3) Purpose of site selection. Site selection can better plan the logistics distribution network so that the dispersed mode of transportation becomes centralized. Distribution centers can promote the coordination between the nodes of logistics, and improve resource integration and resource utilization efficiency.

(4) Site selection methods. Methods can be broadly divided into qualitative and quantitative methods. The most widely used qualitative analysis methods are the Delphi method and expert scoring method; quantitative methods are divided into two types of continuous-discrete.

4.5. Analysis of Distribution Center Siting Research History

Analyzing the evolution path not only enables researchers to make clearer the development background and direction of the research topic but also predicts the research trend of the topic. The time view function of CiteSpace can visualize and analyze this evolution path.
The cluster labels are the vertical axis and the time years of the keywords are the horizontal axis to get the time mapping of keyword clusters for the location of logistics and distribution centers. In this paper, the first 15 clusters are intercepted to show the development of keywords in each cluster.

The keywords with larger nodes are concentrated from 1998 to 2005, and they are: "distribution center", "site selection", "genetic algorithm" and so on. Both the "genetic algorithm" and "center of gravity method" are research methods on site selection, so it can be seen that the early research focuses on the study of site selection methods.

The high-frequency keywords from 2005 to 2015 are: "e-commerce", "reverse logistics", "logistics park" and so on. The high-frequency keywords are relatively evenly distributed in each category, indicating that distribution center seating has developed at a high speed in all directions during this period.

In 2015-2023, only three keywords, "site allocation", "path optimization" and "site path", appear in three clusters, indicating that in recent years, the field has This indicates that the field has been devoted to the direction of joint site-path optimization in recent years.

5. Conclusion

Whether the distribution center site selection is reasonable or not determines whether the logistics system can give full play to its functions, and the study of distribution center site selection is of great significance to both enterprises and society. In this paper, we used CiteSpace to visualize and analyze the literature on distribution center sitting in the Knowledge Network database, and comprehensively analyzed the research hotspots and the development process in the field in the past two decades in terms of the number of articles, core research authors and institutions, keyword co-occurrence and clustering, and timeline charts, etc. For the sitting method, the combination of multiple sitting methods is a major trend. For the site selection method, the combination of multiple site selection methods is a general trend, and future research on site selection will no longer be limited to a single site selection but will combine site selection with path optimization, to better reduce logistics costs. In addition, future research on site selection will also make full use of big data to improve the scientific feasibility of site selection.

References