Research Article

Two-Level Supply Chain Network Construction in Big Data Environment

Yu Lei, Ruiyuan Ma, Hongshu Ye, and Deng Chen

Hubei Province Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China

Correspondence should be addressed to Deng Chen; dchen@wit.edu.cn

Received 26 January 2022; Revised 23 February 2022; Accepted 3 March 2022; Published 29 March 2022

Copyright © 2022 Yu Lei 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.

Supply chain management has become a critical problem for any company looking to meet targets in terms of economic competitiveness, time, and quality of service especially in an economic context marked by the globalization of trade and the speed of industrial cycles. Selecting the right suppliers and establishing proper ordering and shipping systems are important to reduce unnecessary resource loss and improve the core competitiveness of enterprises. However, most scenarios for supplier selection, order allocation, and transportation solution selection remain largely unknown, particularly in the context of big data. To overcome this problem, this study presents a two-level supply chain network. The ranking and selection of suppliers are achieved by constructing a supplier importance measurement index system and establishing an evaluation model with principal component analysis, while combining the greed idea to establish constraint equations for multiple indicators, order allocation to selected suppliers, and finally, the allocation of transportation solutions considering the loss rate of operators. Finally, a real-world example is employed to demonstrate the efficacy of the algorithm.

1. Introduction

The development of dynamic alliances has been aided by the international competitive climate and the rise of informationization, which has revolutionized the way businesses compete with one another. Each enterprise in the supply chain is a community of interests, and by using complementary advantages and synergy effects, it can generate benefits that would not be possible if the enterprise were independent, as well as improve the enterprise’s fundamental competitiveness [1]. As the starting point of logistics, the beginning point of capital flow, and the endpoint of information flow in the supply chain, the choice of suppliers has been recognized as a key factor for enterprises to maintain their strategic position, because it directly affects the cash flow and profitability of enterprises. That is why choosing the right supplier to establish a strategic partnership and a reasonable distribution ordering scheme as well as the supplier’s transportation scheme system is especially important for the core companies in the supply chain [2].

The industries form one of the most different and unstable sectors within the global economy [3]. It must bring together a diverse set of expert skills across geographically dispersed short-term project environments, as well as deal with rapidly changing demand cycles, project-specific product demands, and fluctuating manufacturing conditions. Over the last 20 years, subcontracting has grown in popularity, further complicating the situation by fragmenting the manufacturing process [4].

Supply chain networks describe the flow of commodities and information by connecting organizations to service the end-customer [5]. There are some challenges in sustainable supply chain management. Previously, principal component analysis (PCA) has been proposed in supplier selection problems, and PCA has been used in the design of supply chain networks [6]. Data analysis approaches have been presented to handle supplier selection and order allocation challenges, ranging from simple weighted scoring to complicated mathematical programming methods [7]. Some studies propose a strategy-based approach using PCA-data envelopment analysis to solve the supplier selection problem when data are interrelated and interdependent. Some other studies used principal component analysis (PCA) methods to decrease the number of supplier selection criteria for
It is critical to consider the types of materials ordered and the allocation of transportation options in supply chain management. The quantity of materials to be supplied by suppliers is governed by the suppliers’ development. Therefore, the manufacturer’s purchasing decisions depend on the quantity of material available from suppliers and the stability of the material supplied. Figure 1 depicts the structure of the supply chain and its decisions. Firstly, we determine the selection criteria and then establish a principal component analysis model according to the evaluation criteria to determine the ranking and selection of suppliers, according to the greedy idea of “choosing the least supplier,” “the most economical,” and “the least loss rate.” The finalized ordering and supply plan is derived by solving the three layers of constraint optimization conditions in turn, to order suppliers to provide the maximum amount of raw materials, and to determine the ordering and supply plan under the condition that the transshipment capacity of the forwarder is certain.

It can be divided into two purposes: determining the ordering scheme and determining the transshipment scheme, with three levels of constraints: least supplier, most economical, and least loss. The constraint planning equation is set out to solve these three conditions one by one, based on the importance of the supplier ranking of the composite index, and priority is given to the supplier with the highest score in the composite score, to meet the requirements of the least supplier and the most economical. After data collection, the loss rate of the forwarder is calculated, and the greedy algorithm is employed according to the principle of matching the supplier with a low loss rate priority and a program is developed to achieve the minimum loss. After comparing the program with the original ordering and shipment program, we found that the capacity could be increased by 82.98% in the best case.

2.1. Proposed Method. This study is divided into three comprehensive phases: the goal of the first phase is to develop sustainable supplier selection criteria. In this phase, we have developed our supplier selection criteria by using some literature references, combining the current corporate context and the context of big data. The goal of the second phase is to develop a compliant supplier selection process. The suppliers are assessed and chosen using the criteria established in the first phase. Finally, the third phase’s purpose is to accomplish the order allocation procedure. The order allocation problem is modeled as a mathematical model with several conditional constraints in this step. The research framework of this study is shown in Figure 2. A comprehensive discussion of each phase is provided below.

2.2. Supplier Selection Criteria. The purpose of this phase is to identify the basic criteria for sustainable supplier selection. The purpose of sustainable supplier selection is to evaluate suppliers considering economic, environmental, and social factors. Therefore, the criteria for each factor are selected. Through a comprehensive literature review, we found the most important criteria studied for each of the previous perspectives, and therefore, we extracted a list of potential sustainable supplier selection criteria as given in Table 1, applying the appropriate criteria is the key to achieving a successful supplier selection process. In other words, if an appropriate set of criteria is established, measured, and applied to evaluate supplier performance, an
accurate supplier selection process will emerge. Therefore, it is reasonable to extract an appropriate list of criteria from the initial list of selection criteria previously considered in the literature (Chai et al., [11]), since different industries have different characteristics, which should be considered in their supplier selection process. In other words, although the initial list was developed from the previous literature, the final list was determined by the data related to that business and supplier, avoiding more artificial and objective factors and more applicable to real-world problems in the context of big data.

3. PCA-Based Supplier Selection

In this phase, the best suppliers are selected and ranked using the finalized sustainable supplier selection criteria. We selected the data related to a certain company and its two-tier supply chain network for the last five years, 402 suppliers, and 8 operators based on the relevant data. We processed relevant data by correcting the null with the mean and disregarding the missing values. According to the definition of normal distribution, the probability of being outside $3\delta$ from the mean is $P(|x - \mu|>3\delta) < 0.003$, which is a very small probability event, and by default, we can conclude that the sample with a distance of more than $3\delta$ from the mean does not exist. Therefore, when the sample distance from the mean is greater than $3\delta$, the sample is considered an outlier. An example of the outlier distribution is shown in the box plot in Figure 3.

We selected six indicators including total supply, total order quantity, violation rate, dependence degree, supply satisfaction rate, and cooperation stability degree. The specific definitions of each of these six evaluation indicators are shown in Table 1. The evaluation steps based on PCA are as follows:

$F$ in Table 1 can be expressed as follows.

$$F = \frac{G_i}{D_i}X_i, \quad i = 1, 2, 3 \ldots, 402,$$

$$X_i = n - W_i, \quad i = 1, 2, 3 \ldots, 402, \quad n = 240,$$

where $G_i$ is the sum of the five-year supply quantities of S001, S002, ..., S402, $D_i$ is the sum of five-year orders from S001, S002, ..., S402, $X_i$ shows the total number of five-year orders from suppliers, $n$ is the total number of weeks in five years, i.e. 240, and $w_i$ is the number of orders from suppliers at the end of five years.

Similarly, the $R$ in Table 1 can be expressed as

First, we normalize the difference between order quantity and supply quantity.

$$c_{ij} = a_{ij} - b_{ij},$$

$$D_{ij} = \frac{(c_{ij} - c_{ij}^{\text{max}})}{(c_{ij}^{\text{max}} - c_{ij}^{\text{min}})}.$$  \hfill (2)

Next, we find the mean value of the difference, which indicates the degree of cooperative stability.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - x^-)^2}{N - 1},$$  \hfill (3)

where $a_{ij}$ represents the value of the order quantity of the $i$-th family needed in the $j$-th week, $b_{ij}$ is the value of the supply quantity of the $i$-th family in the $j$-th week, and $i$ represents the number of the supplier. Then, the supplier evaluation index system belonging to our background is constructed, as shown in Figure 4.

We use $X_i$ to denote the metrics in Figure 5. Then, standardize the process by converting each indicator $a_{ij}$ into a
where \( \mu_i \) and \( s_j \) denote the sample mean and sample standard deviation of the \( j^{th} \) indicator.

\[
\ddot{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j}, \quad i = 1, 2\ldots, 402, \quad j = 1, 2\ldots, 6, \quad (4)
\]

\[
\ddot{x}_{ij} = \frac{x_{ij} - \mu_j}{s_j}, \quad j = 1, 2\ldots, 6. \quad (5)
\]
We calculate the correlation coefficient matrix \( R \), \( R = ( r_{ij} )_{5 \times 5} \)

\[
r_{ij} = \frac{\sum_{k=1}^{402} a_{ik} a_{kj}}{402 - 1}, i, j = 1, 2, 3 \ldots 6,
\]

where \( r_{ij} \) is the correlation coefficient between the \( i \)-th indicator and the \( j \)-th indicator. Next, we calculated the eigenvalues and eigenvectors and calculated the eigenvalues of the correlation coefficient matrix \( \lambda_j \) (\( j = 1, 2, \ldots, 6 \)) are calculated, which is the cumulative contribution of the principal components \( y_1, y_2, \ldots, y_6 \). When \( a_p \) is close to 1 (\( a_p = 0.85, 0.90, 0.95 \)), the top \( p \) indicator variables \( y_1, y_2, \ldots, y_p \) are selected as the \( p \) principal components instead of the original 6 indicator variables, so that a comprehensive analysis of the \( p \) principal components can be performed. The standard deviation, contribution rate, and the cumulative contribution rate of all principal components were finally obtained as shown in Table 2.

\[
b_j = \frac{6\lambda_j}{\sum_{k=1}^{6} \lambda_k}, j = 1, 2, 3 \ldots 6,
\]

\[
a_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{6} \lambda_k}.
\]

Principal component analysis was performed for the six indicators, and the standard deviation of principal components, contribution rate of each principal component, and cumulative contribution rate are shown in Table 2. Scatter plots of principal components and the orientation of the original coordinates under the principal components are shown in Figure 6, and the loading matrices corresponding to the above eigenvalues are shown in Table 3.

According to Figure 5: gravel plot and cumulative contribution rate, the cumulative contribution rate of the first four principal components selected has reached 97.6\%, the other two principal components are discarded to achieve the purpose of dimensionality reduction, and finally, the top 10 with better performance are obtained, as shown in Table 4.

The KMO test [12, 13] is used to check the correlation and bias correlation between variables, and the value is between (0, 1). The closer the KMO statistic is to 1, the stronger the correlation between variables, and the weaker is the bias correlation, the better the effect of PCA. In the
actual analysis, when the KMO statistic is below 0.5, it is not suitable to apply the PCA method at this time, and consideration should be given to redesigning the variable structure or using other statistical analysis methods. From the test result $\text{MSA} = 0.53$ (i.e., KMO statistic > 0.5) if the variables are independent of each other, the common factor cannot be extracted from them, and the factor analysis cannot be applied. The Bartlett spherical test [14] determines that if the correlation array is a unit array, the independent factor analysis of each variable is invalid. The results of the $R$-linguistic test [15] showed that $p = 3.181605e^{-94}$ (i.e., $p$ value < 0.05), which means that the criteria were met and the data were spherically distributed and the variables were independent of each other to some extent. In summary, the model passed the KMO test and Bartlett sphere test for the method.

**Table 2:** The standard deviation of principal components, the contribution rate of each principal component, and the cumulative contribution rate.

<table>
<thead>
<tr>
<th>Comp</th>
<th>Comp1</th>
<th>Comp2</th>
<th>Comp3</th>
<th>Comp4</th>
<th>Comp5</th>
<th>Comp6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>1.7383</td>
<td>1.2684</td>
<td>0.8278</td>
<td>0.7354</td>
<td>0.3346</td>
<td>0.1770</td>
</tr>
<tr>
<td>Proportion of variance</td>
<td>0.5036</td>
<td>0.2682</td>
<td>0.1142</td>
<td>0.0901</td>
<td>0.0187</td>
<td>0.0052</td>
</tr>
<tr>
<td>Cumulative proportion</td>
<td>0.5034</td>
<td>0.7718</td>
<td>0.8860</td>
<td>0.9761</td>
<td>0.9948</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Figure 5:** Gravel plot of principal components.

**Figure 6:** Scatter plot of principal components and orientation of original coordinates under principal components.
Table 3: Loading matrix in principal component analysis standard deviation of principal components, the contribution rate of each principal component, and the cumulative contribution rate.

<table>
<thead>
<tr>
<th>Comp1</th>
<th>Comp2</th>
<th>Comp3</th>
<th>Comp4</th>
<th>Comp5</th>
<th>Comp6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fill</td>
<td>0.3539</td>
<td>0.5761</td>
<td>0.1970</td>
<td>0.1066</td>
<td>0.6980</td>
</tr>
<tr>
<td>Order</td>
<td>0.4931</td>
<td>-0.3396</td>
<td>-0.0771</td>
<td>-0.3229</td>
<td>0.1762</td>
</tr>
<tr>
<td>Supply</td>
<td>0.4937</td>
<td>-0.2666</td>
<td>-0.3103</td>
<td>-0.3562</td>
<td>0.0398</td>
</tr>
<tr>
<td>Default</td>
<td>-0.3609</td>
<td>-0.5361</td>
<td>-0.2603</td>
<td>0.3060</td>
<td>0.6464</td>
</tr>
<tr>
<td>Relay</td>
<td>-0.4303</td>
<td>-0.0456</td>
<td>0.4247</td>
<td>-0.7531</td>
<td>0.2425</td>
</tr>
<tr>
<td>Col_var</td>
<td>-0.2691</td>
<td>-0.4383</td>
<td>0.7815</td>
<td>0.3110</td>
<td>-0.0592</td>
</tr>
</tbody>
</table>

Table 4: Overall evaluation scores of the top 10 suppliers.

<table>
<thead>
<tr>
<th>Comp1</th>
<th>Comp2</th>
<th>Comp3</th>
<th>Comp4</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.4323</td>
<td>-1.3903</td>
<td>-0.8673</td>
<td>0.6194</td>
</tr>
<tr>
<td>2</td>
<td>-0.0936</td>
<td>0.6517</td>
<td>-0.0582</td>
<td>0.1951</td>
</tr>
<tr>
<td>3</td>
<td>1.5425</td>
<td>0.8459</td>
<td>0.0276</td>
<td>1.2900</td>
</tr>
<tr>
<td>4</td>
<td>-1.3873</td>
<td>-1.5715</td>
<td>-0.8917</td>
<td>0.8627</td>
</tr>
<tr>
<td>5</td>
<td>0.4236</td>
<td>1.5023</td>
<td>-0.4621</td>
<td>-0.0939</td>
</tr>
<tr>
<td>6</td>
<td>-1.2414</td>
<td>-0.8716</td>
<td>-0.2386</td>
<td>-0.1461</td>
</tr>
<tr>
<td>7</td>
<td>1.5591</td>
<td>2.3290</td>
<td>-1.3681</td>
<td>1.2700</td>
</tr>
<tr>
<td>8</td>
<td>-1.2147</td>
<td>-0.8772</td>
<td>-1.972</td>
<td>-0.1338</td>
</tr>
<tr>
<td>9</td>
<td>-1.4806</td>
<td>-1.5257</td>
<td>-0.6564</td>
<td>0.4282</td>
</tr>
<tr>
<td>10</td>
<td>-1.2218</td>
<td>-1.0597</td>
<td>-0.6881</td>
<td>0.4457</td>
</tr>
</tbody>
</table>

4. Allocation of Orders

When ordering materials, we considered selecting the supplier with the least and most economical solution, according to the score sorting and the cost of materials given different weights, comprehensive sorting priority to choose the supplier with the highest score. To meet the original material inventory to maintain not less than to meet the production capacity and remaining inventory as the target in the first week of supplier selection. While selecting suppliers, we used the difference between two weeks’ capacity and remaining inventory as the target to select suppliers, and to reduce storage costs, we selected suppliers with larger supply in turn after selecting suppliers that just meet the target. Since different suppliers provide different types of materials, i.e., A, B, and C, the enterprise production of each cubic meter of a material demand for A, B, and C is different; so, it is necessary to convert the supplier’s supply into the number of production capacity that the enterprise can produce, we used \( X_i^{\text{avg}} \) to represent the capacity, it can provide for the enterprise to complete, and it was computed using equation (9):

\[
X_i^{\text{avg}} = \frac{X_i}{m_i},
\]

where \( m_i \) represents the cubic meters of material consumed by the \( i \)-th supplier corresponding to its material per cubic meter. Here, the unit prices of A, B, and C materials provided by the enterprise are different, and the ratio is 1.2:1.1:1, we used \( X_i^{\text{price}} \) to represent the cost required, and price represents the cost of each material. It is computed as

\[
X_i^{\text{price}} = X_i^{\text{avg}} \times \text{Price}_i.
\]

To facilitate subsequent calculations, we used the sigmoid function for normalization as

\[
X_i^{\text{price,normalization}} = \frac{1}{1 + e^{-\lambda X_i^{\text{price}}}}.
\]

To consider the economy, we used \( X_i^{\text{score}} \) to represent its score index and then used this index to rank, according to the ranking results to give priority to the suppliers with high scores, and finally the least suppliers and the related ordering scheme through the constraint planning equation.

\[
X_i^{\text{score}} = \frac{Z_i}{X_i^{\text{price,normalization}}},
\]

\[
\begin{align*}
\sigma(j + 1) &= \sum_{i=0}^{n} \sum_{j=0}^{m} w_{ij} \times p_j + \sigma(j) - k, \\
\sigma(0) &= 0, \quad \sigma(i) = 0, i = 0, \\
\sigma(j) &\geq 0, 1 \leq j \leq m, 0 \leq i \leq n.
\end{align*}
\]

5. Determination of the Transfer Scheme Based on Loss Visualization

The average cargo loss rate of eight forwarders is shown in Figure 7. Observing the graphs, we know that the loss rates of different forwarding businesses have large differences, which are tested to be inconsistent with the common distribution characteristics in statistics, and then the transit loss rates of the eight forwarders (T1, T2…T8) are made to go to zero mean as a measure of the loss rates of the eight forwarders. Based on the ordering scheme, the company with a large supply volume is given priority to match the forwarder with a small transit loss rate, one is to reduce the loss, the other assumes that the producer is a rational individual, and such a choice is also consistent with the win-win cooperation concept of the large volume company in the actual situation. For the usual case, a supplier by a transshipment, but if the supply volume is greater than the transport capacity, such a choice is also consistent with the win-win cooperation concept of the large volume company in the actual situation. After processing, we get the average loss rate of eight forwarders, as shown in Table 5. Since the average loss rate is small, we first selected the number of suppliers according to the capacity that can be provided without considering the loss, according to the abovementioned selection of the identified suppliers and the corresponding supplier order allocation, and finally, by calculating the selection of the solution in the order. We get the capacity in this order allocation and the transportation solution can increase up to 82.98%.
6. Conclusion

In an economic setting characterized by the globalization of trade and the pace of industrial cycles, supply chain management has become a vital concern for every organization aiming to meet targets in terms of economic competitiveness, time, and service quality. To prevent avoidable resource loss and improve the core competitiveness of enterprises, selecting the correct suppliers and implementing proper ordering and shipping systems are critical. However, most scenarios for supplier selection, order allocation, and transportation solution selection, particularly in the context of big data, are mostly unexplored. A two-level supply chain network is presented in this study. Construction of a supplier importance measurement index system and establishment of an evaluation model with principal component analysis, while combining the greed idea to establish constraint equations for multiple indicators, order allocation to selected suppliers, and finally the allocation of transportation solutions considering the loss rate of operators, were all used to rank and select suppliers. Finally, a real-world example is implemented to prove the algorithm’s effectiveness.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References


