

ORIGINAL ARTICLE



Construction of Logistics Competitiveness System and Level Measurement of Node Cities of China Railway Express Based on Text Mining

Shiyun Liu¹, Huiying Du¹, Guangda Li^{1,*}, Genxiang Gao¹, Ning Yang²

¹School of Management Science and Engineering, Beijing Information Science & Technology University, Beijing 102206, China

²Shenzhen Topway Video Communication Co., Ltd., Shenzhen 518101, China

*Corresponding Author: Guangda Li

Abstract

This study constructs an initial selection of logistics competitiveness indicators from four dimensions: logistics scale, logistics resources, economic development, and green innovation. Based on the grey relational and redundancy models, the final indicator system is derived. In addition, text mining technology is used to extract high-frequency words from policy texts, and dimension keywords are mined using the Ochiai coefficient. Finally, a composite index algorithm is applied to calculate the weights of dimensions and indicators. The logistics competitiveness levels of the node cities of the China-Europe Railway Express from 2021 to 2023 are calculated, and the influencing factors and urban development potential are explored. This study not only provides a quantitative model for evaluating logistics competitiveness but also demonstrates the key role of data analysis and processing in the logistics field, highlighting the importance of data mining in identifying potential trends and optimizing decision-making.

Keywords: China Railway Express; logistics competitiveness; text mining; indicator system; critic-entropy weighting method

Introduction

China Railway Express is an important land-based international transportation mode connecting China and Europe, which has accumulated more than 50,000 trains and has become a principal way for Chinese enterprises to expand their international markets[1]. However, there are still some challenges in its development, especially the disorganized and unplanned opening plan of the domestic section, which affects transportation efficiency and stability. This is mainly due to the fact that the operation of the China Railway Express depends on the logistics development of the relevant node cities. Evaluating the logistics competitiveness of China Railway Express node cities can help identify service strengths and shortcomings so that targeted measures can be taken to improve efficiency. In conclusion, this study selects 52 China Railway Express node cities, constructs the logistics competitiveness

index system, determines the weights of each index, and calculates the logistics competitiveness level. The logistics development level of the node cities is clarified, with the aim of improving the efficiency and quality of the China Railway Express.

2. Literature Review

The operation of the China Railway Express depends on the logistics development level of the node city. Most of the current studies have borrowed the actual logistics level of the city to measure the logistics development level or operation quality of the nodes in the China Railway Express operation network [2]. Aiming at the research objects of the Belt and Road Initiatives and China Railway Express, scholars mainly construct index systems around the dimensions of logistics scale, transportation

capacity, economic development, informatization, freight security and so on [3,4]. Chen and Li [5] established the competitiveness index of the node cities of the China Railway Express and evaluated the competitiveness of the hub cities of the China Railway Express based on the hierarchical analysis method. Zhu et al. [6] selected the panel data of 18 key provinces in the Belt and Road Initiatives and constructed a competitiveness evaluation index system for the logistics industry from the three dimensions. Dai et al. [7] took the logistics competitiveness of China Railway Express distribution center as the research object, and evaluated the logistics competitiveness of the China-Europe freight train using a multi-index comprehensive evaluation model from three aspects: the environmental support capacity of the regional freight distribution center, the competitive strength of the freight distribution center, and the competitive development capacity of the freight distribution center. Li and Hou [8] conducted cluster analysis and empirical research on the logistics competitiveness of 13 major cities along the Belt and Road Initiatives by using the DAE-WMA method and the social network analysis method. Li [9] conducted cluster analysis and empirical research on the logistics competitiveness of 13 major cities along the the Belt and Road Initiatives by using the in-depth SAE-SOFM model. Characterization and clustering of the logistics competitiveness of 13 core node cities along the northwestern section of the Belt and Road Initiatives.

In logistics competitiveness measurement, unused methods were applied to the study. Yang constructed a comprehensive evaluation model of logistics enterprise competitiveness based on the SEM model. Chen [10] proposed a port logistics function evaluation model based on the entropy weight TOPSIS method. Kang et al. [11] utilized the entropy weight-TODIM method and K-mean clustering analysis to score the logistics competitiveness for assessment and classification. Ma et al. [12] used the entropy value method to construct the evaluation index system of the logistics development level in Chinese provinces and regions. Zhang et al. [13] constructed a DEA evaluation model based on grey cross efficiency by combining grey system theory, cross efficiency method and DEA based on the evaluation index system of China Railway Express operation efficiency. Liu et al. [14] constructed a regional

logistics competitiveness triangle model. Kim et al. [15] used factor analysis to extract the influencing factors of the assessment model and established a location competitiveness assessment model by taking the major port cities in Northeast Asia as an example. Cheng [16] constructed a comprehensive evaluation model of logistics competitiveness with an indicator rating system and entropy-weighted TOPSIS combination scoring model. Sheng [17] used factor analysis to evaluate the competitiveness of logistics of 31 provincial and regional agricultural logistics competitiveness. Liu [18] used factor analysis to conduct a comprehensive evaluation of 18 node cities along the new western land and sea corridor. Chen [19] took the ports along the Belt and Road Initiatives as the research object and used factor analysis to measure their comprehensive logistics competitiveness.

In summary, the current research results on logistics competitiveness are abundant. However, more scholars quantify the competitiveness level only from the objective and data perspective. Fewer studies have analyzed the competitiveness of the node cities of the China Railway Express from both subjective and objective levels. There is an urgent need to enrich this part of the research. Therefore, this study considers policy factors in all aspects, from the construction of the indicator system to the quantification of the weights of the indicator system, and tries to quantify and analyze the competitiveness of the node cities of the China Railway Express.

3. Methods

3.1. Grey Relation Analysis

The Grey Relation Analysis is suitable for solving interrelationship problems with multiple variable criteria[20]. It is used to identify the evaluation indicators of the logistics competitiveness level of the preliminary selection, eliminate the indicators with less consistency with the evaluation goal, and construct a set of minimum complete and adequate indicators that can reflect the logistics competitiveness status of the node city. The calculation steps are as follows:

3.1.1. Calculation of Correlation Coefficients

On the basis of determining the sequence of reference indicators, the correlation coefficients between the indicators of each evaluation object

and the reference indicators are calculated using the formula [21]

$$\gamma_{ij} = \frac{\min_i \min_j |y_{i0} - y_{ij}| - \rho \max_i \max_j |y_{i0} - y_{ij}|}{|y_{i0} - y_{ij}| - \rho \max_i \max_j |y_{i0} - y_{ij}|}, i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (1)$$

Where y_{i0} is the reference indicator after homogenization of the i th evaluation object, and y_{ij} is the j th indicator after homogenization of the i th evaluation object. The ρ is the discrimination coefficient, which takes the value in the range of (0,1) and usually takes 0.5.

3.1.2. Calculation of correlation

Calculate the mean value of the correlation coefficient between an indicator and a reference indicator, i.e., the correlation. The formula is:

$$\gamma_j = \frac{1}{m} \sum_{i=1}^m \gamma_{ij} \quad (2)$$

3.2. Redundancy test

The redundancy test of the indicator system mainly examines whether there are redundant indicators in the indicator system, i.e., it calculates

the average degree of correlation between the indicators in the whole indicator system. The formula is:

$$RD = \frac{\sum_{i=1}^m \sum_{j=1}^m |r_{ij}| - n}{n^2 - n} \quad (3)$$

Where r_{ij} is the correlation coefficient between indicators and n is the number of indicators. Generally, RD is considered acceptable for redundancy as long as it is not greater than 0.5 [22].

3.3. Critic-entropy weight method

The critic-entropy weight method can realize the complementary advantages of objective assignment methods [23]. Wu et al. [24] found that the standard deviation carries a magnitude, the correlation coefficient can be damaging, and

the indicator conflict is only related to the magnitude of the correlation coefficient, which has nothing to do with the positive or negative value. Referring to the research results of Lu et al. [25], the critical method replaced the standard deviation with the standard deviation coefficient to eliminate the effect of the magnitude. Absolute values were taken for correlation coefficients to eliminate the effect of positive or negative signs. Normalization was used for data preprocessing.

3.3.1. The Critic Method

$$C_j = \frac{\sigma_j}{\bar{x}_j} \sum_{i=1}^m (1 - |r_{ij}|) \quad (4)$$

$$w_j = \frac{C_j}{\sum_{j=1}^n C_j} \quad (5)$$

Where σ_j is the standard deviation of the j th indicator, \bar{x}_j is its mean, r_{ij} is the correlation coefficient between the i th indicator and the j th

indicator, and w_j is the weight of the first indicator computed by the critic method.

3.3.2. The Entropy Weighting Method

$$P_{ij} = \frac{Y_{ij}}{\sum_{i=1}^n Y_{ij}} \quad (6)$$

$$E_j = \frac{1}{\ln(m)} \sum_{i=1}^n P_{ij} \ln \frac{1}{P_{ij}} \quad (7)$$

$$w'_j = \frac{1 - E_j}{\sum_{i=1}^n (1 - E_j)} \quad (8)$$

Where Y_{ij} is the value of the j th evaluated indicator of the i th research object before and after standardization; m is the number of evaluated research objects, and n is the number of evaluated indicators; P_{ij} is the weight of the j th evaluated

indicator of the i th research object; E_j is the information entropy of the j th indicator; and w'_j is the weight calculated by the entropy weight method for the j th indicator.

3.3.3. The comprehensive evaluation index

$$w = \frac{w_j + w'_j}{2} \quad (9)$$

Where assuming that the two weighting methods are of equal status, w is the weight of the j th indicator calculated using the combined weighting method [26].

4. Evaluation Index System Construction of Logistics Competitiveness Level

4.1. Construction of the Initial Indicator System

This study adopts the combination of "top-down" and "bottom-up" methods to construct the evaluation index system [27]. Top-down means that based on the principle of quantifiable and operable and relying on the development foundation of the current situation of the "14th Five-Year Plan" for Modern Logistics

Development, logistics scale, logistics resources, economic development, and green innovation are identified as the four dimensions of the indicator system. Bottom-up means that the disclosure of China Railway Express operation information is incomplete and the availability of data is difficult to meet the needs of academic research. Therefore, representative, open and transparent external data are selected, and the data indicators are screened to correspond to the dimensional layers. In summary, the preliminary indicator system contains 4 dimensions, and 36 indicators are screened to be aggregated to the corresponding dimension layer, as shown in Table 1.

Table 1 Preliminary selection of the evaluation index system

| Characteristics | Index | Unit | Index Attribute |
|-------------------------|--|------------------------|-----------------|
| Logistics scale A | Road mileage | kilometers | + |
| | Road freight | 10,000 tons | + |
| | Railroad freight | 10,000 tons | + |
| | Cargo and mail throughput | ton | + |
| | Road cargo turnover | billion ton kilometers | + |
| | Number of CR Express boxes | box | + |
| | Return proportion | % | + |
| | Number of CR Express routes | columns | + |
| | Total postal operations | billions (yuan) | + |
| Logistics resource B | City Logistics Hub Rating | / | + |
| | Value added to the logistics industry | billions (yuan) | + |
| | The growth rate of value added in the logistics industry | % | + |

| | | | |
|------------------------|--|-----------------------------------|---|
| | Number of logistics company | / | + |
| | Financial expenditures in the logistics industry | billions (yuan) | + |
| | Civilian car ownership | 10,000 vehicles | + |
| | The ratio of logistics and warehousing land to urban land area | % | + |
| | Land area for road transportation | hectares | + |
| | Land area for railroad transportation | hectares | + |
| Economic development C | GDP per capita | yuan | + |
| | GDP growth rate | % | + |
| | Total retail sales of consumer goods | billions (yuan) | + |
| | urbanization rate | % | + |
| | Growth rate of resident population | % | + |
| | Per capita disposable income of urban residents | yuan | + |
| | Value added of primary industry/GDP | % | + |
| | Value added of the secondary industry/GDP | % | + |
| | Value added of the tertiary industry/GDP | % | + |
| | Import amount | billions (yuan) | + |
| | Export amount | billions (yuan) | + |
| Green Innovation D | Number of patent applications | parts | + |
| | Students enrolled in general higher education | 10,000 people | + |
| | Number of R&D personnel | people | + |
| | Number of Internet broadband access subscribers | 10,000 households | + |
| | Number of cell phone subscribers | 10,000 households | + |
| | Energy consumption per unit of GDP | tons of standard coal/10,000 yuan | - |
| | Total energy consumption | 10,000 tons of standard coal | - |

4.2. Indicator data sources

The scope of this study is the 14th Five-Year Plan period. At present, more than 90 cities in China are involved in the operation of China Railway Express. The 52 node cities are selected as the research object based on the number of open lines in 2021-2023. The data for measuring the level of logistics competitiveness come from the 2021-2023 China Urban Statistical Yearbook, China Regional Statistical Yearbook, China Energy Statistical Yearbook, prefecture and city yearbooks, prefecture and city statistical bulletins, and the General Administration of Customs. Due to some missing indicator data for some cities,

this study uses the approximate substitution of the province-wide data divided by the GDP ratio.

4.3 Amendments to the indicator system

4.3.1 Initial screening based on Grey Relation Analysis

On the basis of the preliminary selection of the indicator system, the grey correlation of each indicator is calculated, and its calculation results and ranking are shown in Table 2. According to the trade-off between the actual significance of each evaluation index and the number of indexes, about 20% of the indexes with small correlations are eliminated [22]. In this study, a total of 7 indicators are eliminated sequentially according to the calculation results.

Table 2. Grey Relation Analysis results

| Index | Relational degree | Rank | Index | Relational degree | Rank |
|--|-------------------|------|--|-------------------|------|
| cargo and mail throughput | 0.968 | 1 | Number of CR Express routes | 0.867 | 19 |
| Total postal operations | 0.934 | 2 | GDP per capita | 0.867 | 20 |
| Financial expenditures in the logistics industry | 0.925 | 3 | Land area for road transportation | 0.867 | 21 |
| Number of logistics company | 0.925 | 4 | Value added of the tertiary industry/GDP | 0.866 | 22 |
| Import amount | 0.921 | 5 | Road mileage | 0.864 | 23 |
| Number of patent applications | 0.917 | 6 | urbanization rate | 0.864 | 24 |
| Number of R&D personnel | 0.911 | 7 | Per capita disposable income of urban residents | 0.863 | 25 |
| Value added to the logistics industry | 0.907 | 8 | Number of CR Express boxes | 0.861 | 26 |
| Export amount | 0.902 | 9 | Land area for railroad transportation | 0.861 | 27 |
| Number of cell phone subscribers | 0.901 | 10 | The ratio of logistics and warehousing land to urban land area | 0.858 | 28 |
| Total retail sales of consumer goods | 0.9 | 11 | The growth rate of value added in the logistics industry | 0.858 | 29 |
| Number of Internet broadband access subscribers | 0.888 | 12 | GDP growth rate | 0.857 | 30 |
| Civilian car ownership | 0.884 | 13 | Value added of the secondary industry/GDP | 0.854 | 31 |
| Total energy consumption | 0.884 | 14 | Growth rate of resident population | 0.849 | 32 |
| Students enrolled in general higher education | 0.883 | 15 | Energy consumption per unit of GDP | 0.847 | 33 |
| City Logistics Hub Rating | 0.881 | 16 | Return proportion | 0.845 | 34 |
| Road freight | 0.881 | 17 | Railroad freight | 0.837 | 35 |
| Road cargo turnover | 0.87 | 18 | Value added of primary industry/GDP | 0.83 | 36 |

4.3.2 Redundancy test

According to the eliminated indicator system, calculate the correlation coefficient of the indicator system. Substituting the correlation coefficient into the redundancy formula, the redundancy degree of the indicator system is calculated to be 0.401, which is less than the critical value of 0.5. Therefore, the redundancy degree of the constructed indicator system can be accepted.

4.3.3 Determination of the final indicator system

The final logistics competitiveness index system of this study consists of four dimensions. The logistics scale dimension contains 7 indicators. The logistics resources dimension contains 9 indicators. The economic development dimension contains 7 indicators. The green innovation dimension contains 6 indicators.

5. Determination of dimension weights based on text mining

This study determines the weights of the dimensions in the logistics competitiveness index system of China Railway Express node cities based on text mining. By analyzing the policy text, the weights are determined based on the actual policy content and the government's intention rather than relying solely on subjective judgments or the views of external experts. The transformation process from policy text to indicator weights, which is a more objective and data-driven approach, ensures that the indicator system is not only scientific and reasonable but also closely corresponds to the government's policy objectives and strategies.

5.1 Text data sources and high-frequency word extraction

The selection of textual material was the basis for conducting the text-mining study [28]. The time period for the release of the policy text ranges from 2021 to 2022, which matches the period of this study. At the same time, the subjects of the released texts are highly representative. Therefore, these policy documents are suitable for constructing datasets for text mining. All of the above policy contents are converted into text format, with a total length of 79,495 thousand words. The specific content is shown in Table 3.

Table 3. Relevant Policy Documents on the Development of China Railway Express

| Number | Time of publication | Name of policy | Agency |
|--------|---------------------|--|--|
| 1 | October 29, 2021 | A plan to develop green transportation during the 14th Five-Year Plan period (2021-2025) | Ministry of Transport of the People's Republic of China |
| 2 | December 25, 2021 | Work Plan for Promoting the Development of Multimodal Transport and Optimizing and Adjusting the Transport Structure (2021-2025) | General Office of the State Council of the People's Republic of China |
| 3 | January 18, 2022 | A plan outlining major targets for transportation network development in the 14th Five-Year Plan period (2021-2025) | State Council of the People's Republic of China |
| 4 | August 18, 2022 | China Railway Express Development Report (2021) | Office of the Leading Group for Promoting the Construction of the Belt and Road Initiative, China State Railway Group Co., Ltd |
| 5 | December 15, 2022 | A plan to develop modern logistics during the 14th Five-Year Plan period (2021-2025) | General Office of the State Council of the People's Republic of China |

High-frequency word extraction is an essential and critical technique in text analysis, which can be divided into the following steps: data collection, cleaning text, removing noise information from text, word segmentation processing, removing deactivated words, word frequency statistics, sorting and filtering, and so on. In this study, GooSeeker was used to perform lexical processing and word frequency statistics of the text. The lexical properties were filtered, and only nouns and verbs were retained. Then, the top 300 high-frequency words were extracted for subsequent research. Words such as "logistics",

"development", "construction", etc., appear more frequently, indicating that they are the theme and focus of the dataset.

5.2 Dimensional keyword mining based on Ochiai coefficient

5.2.1 Co-word analysis

In this study, co-word analysis is used to determine the similarity between all high-frequency words and the keywords selected based on the names of the dimensions (which are referred to as the "core words" of the study) and to construct a one-to-many relationship between the dimensions and the keywords, so as to make the

keyword-based weighting more scientific. Co-word analysis is based on the assumption that if two or more words appear frequently together in a document set, then there may be a relationship between these words, or they may be related to a specific topic or concept. First, a symmetric co-word matrix is formed by pairing high-frequency

keywords two by two and counting the number of times they co-occur in the sample documents. Second, the co-word matrix is transformed into a similarity matrix using the Ochiai similarity coefficient [29]. The calculation formula is as follows:

$$\text{Ochiai correlation coefficient} = \frac{\text{Co-occurrence times of word A and B}}{\sqrt{\text{the frequency of word A}} \times \sqrt{\text{the frequency of word B}}} \quad (10)$$

5.2.2 Calculation of the dissimilarity matrix

Since too many values of 0 in the similarity matrix may lead to a decrease in the accuracy of the statistical analysis, it is necessary to obtain the dissimilarity matrix by calculating 1 minus each element of the similarity matrix. In the dissimilarity matrix, the closer the value to 0 means the more similar the two keywords are, and vice versa means the more significant the difference.

5.2.3 Cohesive Hierarchical Cluster Analysis

Coalescence hierarchy analysis using the dissimilarity matrix allows further understanding of the distance or similarity between different elements of textual data, thus revealing the structure and relational hierarchy of the data. The clustering algorithm is executed by starting with each data point as a separate cluster and gradually merging the most similar (or least dissimilar) clusters until all data points are merged into one cluster or a specific stopping condition is reached [30]. In this study, cohesive hierarchical clustering was used, and the Python visualization of the clustering results is shown in Figure 1.

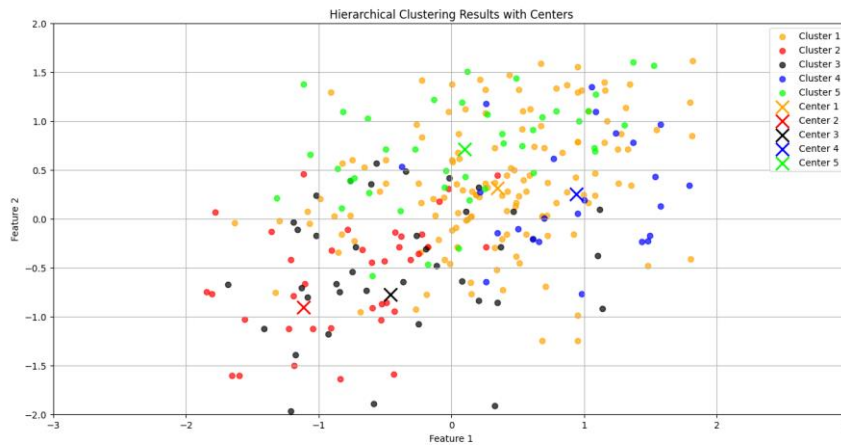


Figure 1. Results of cohesive hierarchical cluster analysis

5.2.4 Keyword Filtering

The following principles were used for selection. First, it belongs to the same category as the core word of this dimension in the clustering analysis results. Second, the similarity with the core words of this dimension was in the top 10% of all high-frequency words. Third, the similarity with the

core words of all other dimensions is lower than the similarity with the core words of this dimension. Fourth, there are clear connotations and meanings. The dimension keywords are shown in Table 4. Their textual word frequency data contain information about the government's attention to the dimension to which they belong.

Table 4. Keyword filtering results for each dimension

| Number | Dimension | Keyword | |
|--------|-----------|-----------|---------------|
| | | Core word | Extended word |
| | | | |

| | | | |
|---|----------------------|-------------------|--|
| 1 | Logistics scale | scale | freight, node, port, expand, realize, hub, railroad, logistics park, guide |
| 2 | Logistics resources | resource | development, mode, service, function, encouragement, acceleration, enhancement, construction, strengthening, facilities, level, coordination, promotion, advancement, improvement, integration |
| 3 | Economic development | economics | industry, industry chain, play, high quality, supply, elements, space, global, integration, depth, adaptation, modernization, demand |
| 4 | Green innovation | green, innovative | innovation, promote, model, cultivate, enterprise, explore, logistics, efficiency, support, traffic, transportation, in-depth, transport, transformation |

5.3 Calculation of dimension weights

Based on the logistics competitiveness evaluation index system, taking the keyword word frequency as the data source, combined with the entropy weighting method to carry out the weighting calculation of the four dimensions of logistics scale, logistics resources, economic development and green innovation [31], the results are shown in Table 5. The performance of each dimension is analyzed. The information entropy value of each dimension is above 0.6, indicating that these

dimensions have certain complexity and uncertainty. Logistics scale has the best performance in terms of information utility and weight, while logistics resources have the lowest utility value, and further optimization or integration of resources is needed to improve their utility. In addition, more investment in green innovation can be made to enhance its role in the logistics system. Balance and optimization among the dimensions will be the key to improving overall logistics efficiency.

Table 5. Calculation results of the entropy weighting method for each dimension

| Number | Dimension | Information entropy | Information utility value |
|--------|----------------------|---------------------|---------------------------|
| 1 | Logistics scale | 0.6184 | 0.3816 |
| 2 | Logistics resources | 0.737 | 0.263 |
| 3 | Economic development | 0.7242 | 0.2758 |
| 4 | Green Innovation | 0.6934 | 0.3066 |

6. Empirical analysis

As shown in Table 6, the logistics competitiveness scores range from 0.0273 (Tieling) to 0.6127 (Shanghai), showing apparent differences. Analyzed from the perspective of scores, all node cities can be divided into three classes. Firstly, the high-scoring cities are characterized as follows. Shanghai As China's economic and financial centre, Shanghai excels in logistics competitiveness with its well-developed ports, modern logistics system and superior geographic location. Its comprehensive transportation network and international trade capabilities make it an important node for the China Railway Express. Beijing and Chongqing both have advantages in terms of policy support and market demand. As a political centre, Beijing has a solid economic foundation, while Chongqing is an

important logistics node in the southwest due to the uniqueness of its geographical location. Second, medium-scoring cities are characterized as follows. Cities such as Shenzhen, Guangzhou and Chengdu, with their high-tech industries and large consumer markets, have relatively high scores but still have room for improvement in the construction and operational efficiency of international logistics corridors. Finally, low-scoring cities are characterized as follows. Cities such as Tieling and Yingtan are relatively weak in terms of infrastructure, economic vitality, and policy environment, which limits the competitiveness of logistics development. The government should strengthen infrastructure development and improve policy support to enhance its position in the national logistics network.

Table 6 Average Logistics Competitiveness of Node Cities

| Rank | City | Competitiveness | Rank | City | Competitiveness | Rank | City | Competitiveness |
|------|--------------|-----------------|------|-----------|-----------------|------|-------------|-----------------|
| 1 | Shanghai | 0.6127 | 19 | Dongguan | 0.1751 | 37 | Ganzhou | 0.0975 |
| 2 | Beijing | 0.4804 | 20 | Shenyang | 0.1622 | 38 | Hohhot | 0.0884 |
| 3 | Chongqing | 0.4775 | 21 | Dalian | 0.1530 | 39 | Liuzhou | 0.0821 |
| 4 | Shenzhen | 0.4637 | 22 | Harbin | 0.1530 | 40 | Zibo | 0.0768 |
| 5 | Guangzhou | 0.4080 | 23 | Jinhua | 0.1421 | 41 | Lianyungang | 0.0704 |
| 6 | Chengdu | 0.3445 | 24 | Linyi | 0.1410 | 42 | Chifeng | 0.0669 |
| 7 | Suzhou | 0.3123 | 25 | Changchun | 0.1406 | 43 | Rizhao | 0.0639 |
| 8 | Xian | 0.2811 | 26 | Tangshan | 0.1382 | 44 | Ulanqab | 0.0614 |
| 9 | Wuhan | 0.2696 | 27 | Guiyang | 0.1372 | 45 | Luzhou | 0.0605 |
| 10 | Zhengzhou | 0.2690 | 28 | Hengyang | 0.1352 | 46 | Dezhou | 0.0599 |
| 11 | Tianjin | 0.2565 | 29 | Xuzhou | 0.1321 | 47 | Mianyang | 0.0547 |
| 12 | Nanjing | 0.2550 | 30 | Weifang | 0.1302 | 48 | Deyang | 0.0506 |
| 13 | Qingdao | 0.2315 | 31 | Yantai | 0.1269 | 49 | Nanping | 0.0429 |
| 14 | Changsha | 0.2245 | 32 | Urumqi | 0.1248 | 50 | Bayannur | 0.0391 |
| 15 | Shijiazhuang | 0.1903 | 33 | Quanzhou | 0.1220 | 51 | Yingtian | 0.0368 |
| 16 | Hefei | 0.1877 | 34 | Taiyuan | 0.1199 | 52 | Tieling | 0.0273 |
| 17 | Ji'nan | 0.1851 | 35 | Nanchang | 0.1192 | | | |
| 18 | Xiamen | 0.1825 | 36 | Yiwu | 0.1114 | | | |

According to the ranking of scores, Shanghai ranks first, showing its strong advantages in logistics scale, logistics resources, economic development and green innovation. The scores of the top 10 cities are much higher than those of the bottom 40 cities, showing a high concentration and imbalance of market resources. This phenomenon reflects the differences in development levels among cities and the imbalance of resource allocation. From the perspective of geographical distribution, the scores of eastern coastal cities are generally higher than those of inland cities, showing the imbalance of regional economic development. This shows that while promoting the development of China Railway Express, policymakers need to pay attention to the logistics infrastructure and economic development of inland areas.

In addition, in 2020, China supports node cities with superior location conditions, good infrastructure foundation, and potential for standardized operation to carry out the construction of demonstration projects for China Railway Express assembly centers. Zhengzhou, Chongqing, Chengdu, Xi'an, and Urumqi became the first batch of China Railway Express assembly center cities. Except for Urumqi, the rankings of the other four cities are all in the top ten. This

shows to a certain extent that the indicator system constructed in this study is reasonable. For Urumqi, because the overall economic scale and market demand are relatively small, the ranking is low and the logistics competitiveness is relatively weak. But it is still a medium-scoring city because Urumqi, as a core node city of the Silk Road Economic Belt, connects the two major economic circles of Asia and Europe and has unique hub location advantages.

7. Conclusions and Suggestions

7.1. Conclusions

This study fully considers the policy factors and constructs the evaluation index system of logistics competitiveness of China Railway Express node cities from the perspective of combining subjectivity and objectivity. Based on text mining, the level of logistics competitiveness is analyzed empirically, revealing the differences between different node cities in terms of logistics scale, logistics resources, economic development and green innovation. The results show that Shanghai ranks first in logistics competitiveness, thanks to its strong economic foundation, modernized logistics system and good geographical location. High-scoring cities such as Beijing and Chongqing also show clear advantages, while

low-scoring cities such as Tieling and Yingtan need more infrastructure and policy support. Overall, the logistics competitiveness of eastern coastal cities is significantly higher than that of inland cities, reflecting the unevenness of regional economic development.

7.2 Suggestion

Based on the above conclusions and analysis, this study puts forward the following suggestions. First, the government should increase investment in logistics infrastructure in low-scoring cities to enhance their logistics service capacity in order to strengthen their position in the national logistics network. Second, targeted policies should be formulated to encourage logistics development in inland areas, including financial support, tax incentives and market access facilitation. Third, logistics enterprises should be encouraged to invest more in green innovation and enhance environment-friendly logistics services to meet the challenges of sustainable development. Fourthly, with the "Belt and Road" initiative, we will strengthen ties with the Asia-Europe Economic Circle, promote the construction of international logistics corridors for inland cities, and enhance their competitiveness. The aim is to better promote the logistics development of each node city, narrow the competitiveness gap between regions, and realize the high-quality development of the China Railway Express.

References

- Choi, K.-S. The current status and challenges of China Railway Express (CRE) as a key sustainability policy component of the Belt and Road Initiative. *Sustainability* 2021, 13(9), 5017–5017.
- Zhang, J.; Zhang, Y.; Yuan, Q. Measurement and spatial connection of logistics economic development of China Railway Express node cities. *Business and Economic Research* 2021, (01), 106–109.
- Gan, W.; Yao, W.; Huang, S. Evaluation of green logistics efficiency in Jiangxi Province based on three-stage DEA from the perspective of high-quality development. *Sustainability* 2022, 14(2), 797–797.
- Sun, W.; Zhao, L.; Wang, C.; Li, D.; Xue, J. Selection of consolidation centers for China Railway Express. *International Journal of Logistics Research and Applications* 2019, 23(5), 417–442.
- Chen, C.; Li, Q. Competitiveness of China Railway Express node cities based on AHP. *Lecture Notes on Data Engineering and Communications Technologies* 2022, 186–193.
- Zhu, F.; Tan, B.; Wang, T. Evaluation of the competitiveness of regional logistics industry under the perspective of ecological niche—Taking the key provinces of the "Belt and Road" as an example. *Economy and Management* 2019, 33(01), 30–35.
- Dai, Y.; Li, Q.; Song, H. Research on the optimal layout of domestic cargo distribution centers of China Railway Express train under the "Belt and Road" initiative. *Journal of Chongqing University of Technology (Social Science)* 2018, 32(10), 66–75.
- Li, N.; Hou, X. DAE-WMA optimization algorithm for city logistics competitiveness analysis. *Computer Engineering and Application* 2019, 55(16), 246–254.
- Li, N. Evaluation of logistics development level in core cities of "Belt and Road". *Statistics and Decision Making* 2022, 38(04), 184–188.
- Yaxu, Y. Comprehensive evaluation of logistics enterprise competitiveness based on SEM model. *Journal of Intelligent & Fuzzy Systems* 2021, 40(4), 6469–6479.
- Chen, Z. Port logistics function evaluation model based on entropy weight TOPSIS method. *Discrete Dynamics in Nature and Society* 2022, 2022(1), 5006900.
- Kang, X.; Pang, H. City selection for fresh produce e-commerce's market entry strategy: Based on the perspective of urban logistics competitiveness. *Transportation Research Interdisciplinary Perspectives* 2022, 13, 100537.
- Ma, W.; Cao, X.; Li, J. Impact of logistics development level on international trade in China: A provincial analysis. *Sustainability* 2021, 13(4), 2107.
- Zhang, J.; Chang, J.; Lin, P.; et al. Operation efficiency evaluation of the China-Europe freight train based on grey cross-efficiency DEA. *Scientific Programming* 2020, 2020(1), 8843733.
- Liu, G.; Zhu, J.; Liu, X. Comprehensive evaluation of regional logistics

- competitiveness considering multiple reference points and dynamic index improved analytic hierarchy process. *Symmetry* 2020, 12(5), 847.
16. Kim, S. H.; Lee, K. H.; Kang, D. W. Analytic hierarchy process modelling of location competitiveness for a regional logistics distribution center serving northeast Asia. *Journal of Korea Trade* 2020, 24(3), 20–36.
 17. Cheng, W. Evaluation of logistics competitiveness of provinces and cities along the New Western Land-Sea Corridor based on entropy weight and TOPSIS method. *J. Manag. Decis. Sci.* 2020, 3, 23–33.
 18. Sheng, Z. The evaluation of regional logistics competitiveness of agricultural products in China and spatial analysis on the differences. *Pakistan Journal of Statistics* 2014, 30(5).
 19. Liu, H. Evaluation of logistics competitiveness of node cities along the New Land-Sea Corridor in Western China—Based on factor analysis and cluster analysis. *International Journal of Frontiers in Sociology* 2023, 5(3).
 20. Kuo, Y.; Yang, T.; Huang, G. W. The use of grey relational analysis in solving multiple attribute decision-making problems. *Computers & Industrial Engineering* 2008, 55(1), 80–93.
 21. Sarraf, F.; Nejad, S. H. Improving performance evaluation based on balanced scorecard with grey relational analysis and data envelopment analysis approaches: Case study in water and wastewater companies. *Evaluation and Program Planning* 2019, 79, 101762–101762.
 22. Jiang, G.; Zhu, T. Construction and empirical evidence of evaluation index system for rural resilience in China. *Statistics and Decision Making* 2024, 40(07), 75–80.
 23. Zafar, S.; Alamgir, Z.; Rehman, M. H. An effective blockchain evaluation system based on entropy-CRITIC weight method and MCDM techniques. *Peer-to-Peer Networking and Applications* 2021, 14(5), 3110–3123.
 24. Wu, Z.; Guan, J.; He, J. An empirical study on the calculation of minimum wage standard—Dynamic combination calculation based on objective weight of CRITIC-Entropy Weight Method. *Modern Econ Sci* 2019, 41(03), 103–117.
 25. Lu, H.; Zhao, Y.; Zhou, X.; Wei, Z. Selection of agricultural machinery based on improved CRITIC-Entropy weight and GRA-TOPSIS method. *Processes* 2022, 10(2), 266–266.
 26. Li, L.; Mo, R. A comprehensive decision-making approach based on hierarchical attribute model for information fusion algorithms' performance evaluation. *Mathematical Problems in Engineering* 2014, 2014(1), 124156.
 27. Liu, S.; Du, H. Research on site selection of China Railway Express assembly center. *Proc. SPIE* 2023, 12989, 1298913.
 28. Wang, Y.; Song, C.; Wang, X.; Gao, P. A comprehensive index algorithm coupling text mining and high-dimensional ranking weighting method and an application case of provincial development evaluation in China. *Geography Research* 2023, 42(10), 2556–2573.
 29. Liu, L.; Zhang, L.; Ye, P.; et al. User needs mining based on topic analysis of online reviews. *Tehnički Vjesnik* 2019, 26(1), 230–235.
 30. Mushtaq, N.; Masood, N.; Khattak, J. A.; et al. Health risk assessment and source identification of groundwater arsenic contamination using agglomerative hierarchical cluster analysis in selected sites from upper Eastern parts of Punjab province, Pakistan. *Human and Ecological Risk Assessment: An International Journal* 2021, 27(4), 999–1018.
 31. Xinfu, Y.; Huimiao, Y.; Shengran, W. Construction and application of enterprise water security evaluation index system in the context of sustainable development. *Water Resources Economy* 2024, 1–8.