Analysis Of Change Trend Based On K-means And Characteristics Of China's Railway Operation

Wending Jin[1,#], Jinyi Li[2,#] and Zhuohao Fang[3,#,*]

[1]*Faculty of Science, Minzu University Of China, Beijing, China*
[2]*Institute of Artificial Intelligence, Southwest University, Chongqing, China*
[3]*Faculty of information technology, Macau University of Science and Technology University, Macau, China*
*{Corresponding author}*19098533i011016@student.must.edu.mo*
[1] *jwd_16wssy@163.com*
[2] *olivialee79@163.com*
#These authors contributed equally

Abstract

With the advantages of high speed and high efficiency, railway has rapidly become the core of the national transportation system. At the same time, railway industry is also developing vigorously. In this paper, the quantity data of intercity railways, high-speed railways and ordinary trains in 118 cities from 2008 to 2021 are selected as samples, and k-means clustering method is used to conduct clustering analysis and processing on railway train number data, and cities are divided into eight categories. Based on the analysis of specific national policies, the characteristics of each type of city and the quantitative characteristics of different types of trains in the annual clustering results are obtained, and the alluvial map is drawn according to the clustering results, and the transformation law between different types of cities and the number evolution trend of different types of trains are obtained. The results show that: Urban train operation is closely related to the level of administrative management, geographical location and economy of the city.

Keywords: *K-means; High-speed railway; The evolution trend; High-speed operation;*

1.INTRODUCTION

With the introduction of policies such as the "13th Five-Year Plan for Railway Development" in 2017 and "Opinions on Further Improving Railway Planning and Construction" in 2021, high-speed railway construction has entered a golden development period. By the end of 2018, the total length of China's high-speed railways reached 2.9×10^4 km, ranking first in the world in both the length and transportation density of high-speed railways. With the expansion of the high-speed rail network scale, the characteristics of high-speed network and pattern have changed. The emergence of high-speed rail has promoted economic and social development in the region to a certain extent. Therefore, a clear understanding of the spatial structure characteristics of high-speed railway networks and the evolution trend of high-speed railway operation has important guiding significance for future urban spatial development planning. Based on this background, scholars have paid increasing attention to the transportation efficiency, accessibility and network pattern characteristics of high-speed rail in recent years.

In recent years, China's high-speed rail has been incorporated into the national key planning projects. From 2004, the "planning" first proposed to build a "horizontal" high-speed railway grid Bureau, to 2016, the "planning" made changes, that is, the upgrading of high-speed railway network structure from "horizontal" to "eight vertical and eight horizontal", which indicates that China's high-speed railway network structure is more complex. Since the construction of high-speed railway, different scholars have carried out in-depth studies on the regional distribution, functional structure and other characteristics of high-speed railway from different perspectives. The research process found that with the strengthening of high-speed rail construction, the imbalance of the high-speed network pattern also began to highlight the eastern, central and western high-speed rail transport efficiency and showed a trend of decreasing the contact intensity contour iron elements. Therefore the scientific construction of high-speed rail planning layout, the configuration of resources, and economic exchanges between the regional society have a positive role. How to excavate the evolution trend of future high-speed railway

G. Vilas Bhau et al. (Eds.): MSEA 2022, ACSR 101, pp. 439–447, 2023.
https://doi.org/10.2991/978-94-6463-042-8_64

operation according to the current high-speed railway network is the key to laying the foundation for subsequent planning, construction and research, and regional economic exchanges and development. At present, the research on high-speed railway is becoming mature. It is of great practical significance for the research on high-speed railway network to reorganize the research results on the evolution trend of high-speed railway and reflect on them.

In past research, most scholars mainly use complex network theory and social network analysis method to analyze characteristics of high-speed network pattern, and the results derived have a strong practical significance. Although the predecessors' research results with more perfect, there is still a certain problem. Most research results are not based on the long time series are discussed, with some limitations. In addition, the high-speed railway construction is affected by natural geography and other environmental factors, which are still challenging for the study of high-speed railway construction. In view of these problems, based on previous studies, this paper adopts K-means clustering analysis further to discuss the evolution trend of high-speed railway operation. Discussing the aspects of railway cities obtained from clustering is more conducive to summarizing the evolution trend of high-speed railway. On this basis, combined with the transfer results of representative cities in China from 2008 to 2021, the paper can reveal the evolution law of high-speed railway operation in China's high-speed railway construction for a long time series. It provides a scientific basis and correct guidance for further planning of high-speed railway construction in China.

Through cluster analysis, we draw the following conclusions: 1. There is a strong correspondence between the overall train operation and the urban administrative level. 2. The number of trips is concentrated in the economically developed urban areas in the East, and the train density in the East and middle is higher than that in the northeast and West. 3. Economic factors are the fundamental driving force for the generation of high-speed rail passenger flow.

2.LITERATURE

With the deepening of high-speed railway construction, eastern, central and western China have their own regional and local high-speed railway network, and China's high-speed railway network is gradually becoming mature. Current researches on high-speed railway mainly focus on the pattern characteristics of high-speed railway network ([1][4][5][6][7]), the evolution trend of high-speed railway network pattern [2], spatial characteristics of high-speed railway network [4][8], etc.

2.1. Study On The Pattern And Characteristics Of High-Speed Railway Network

2.1.1. Research Based On Complex Network Theory

In the research process, based on previous studies, Wang Leihui [6] found that due to the limitations of methods and data, most studies failed to comprehensively evaluate the balance of high-speed railway network on the whole, and the evaluation indicators were relatively single. In this regard, Wang Leihui [6], based on complex network theory, took HSR cities as nodes most representative nodes were selected to establish the evaluation index of equilibrium, and evaluated the development of HSR pattern by combining the offshore-sharing model and point-line-plane hierarchical analysis. The research results show that the high-speed railway network is characterized by "strong vertical connection and weak horizontal connection," and the structure of high-speed railway develops from hierarchical to "flat".

2.1.2. Research Based On Social Network Analysis

With the in-depth research on the high-speed railway network, Zhao Yinghui, Chu Nanchen and otherscholars [1][2][7]

summarized previous experience, constructed a high-speed railway operation frequency matrix, and studied the structural characteristics of a high-speed railway network in different urban areas agglomerations by using social network analysis method. The structure and characteristics of high-speed railway network are systematically analyzed through network density and network centrality, core edge structure and condensed subgroup, and small-world effect. The research results show that China's high-speed rail network is relatively loose as a whole, important high-speed rail networks show obvious "corridor effect", and the "small-world effect" within urban high-speed rail network is prominent [1][2]. Using social network analysis, Zhong Yexi [8] also pointed out three regional organizational forms of high-speed railway network, which will change from single core to network with the improvement of high-speed railway network.

2.2. Study On Spatial Efficiency, Linkage Pattern And Supply-Demand Relationship Of High-Speed Railway Lines

Jiang Haibing [4] analyzed the strength of high-speed railway connection in China through traditional gravity model and GIS spatial analysis technology in his article "Spatial Linkage Pattern of Urban High-speed Railway Passenger Transport in China" published in 2018. The results show that the distribution of high-speed railway

passenger volume is related to the city level. The passenger volume decreases with the increase of time distance and greatly influences within 4h along the high-speed railway. The connection strength of high-speed rail can be divided into six levels, and the central regional city forms a "hub-spoke" structure. Early studies discussed the evolution of the high-speed rail network from accessibility [8]. Jiang Haibing [3] studied the space efficiency and supply-demand relationship of China's high-speed rail lines from the perspective of accessibility pointed out daily as a high-speed rail link accessibility evaluation index, and introduced the population-weighted accessibility. This further portrayed the rail line space evolution characteristics of accessibility, relatively comprehensive high-speed rail in China space efficiency, and the relationship between supply and demand [3].

2.3. Summary

The evolution trend of high-speed railway operation is explored by using K-means clustering analysis method. Based on existing studies, it can be seen that China's high-speed railway network pattern has highlighted the characteristics of imbalance, and there are significant differences in the distribution of network characteristics in the eastern, central and western regions. On the main trunk lines of high-speed railway, the "corridor effect" is prominent; The spatial efficiency of high-speed rail and the relationship between supply and demand are also restricted by factors such as urban development level and population scale. However, most studies based on social network analysis theory are limited to static high-speed railway network structure, and there are few studies on long time series, which has certain limitations.

3.THE RESEARCH METHODS

The evolution trend of high-speed railway operation is explored by using K-means clustering analysis method.

3.1. Data Processing

3.1.1. Data Description

Taking all the flights of 31 provinces in China from 2008 to 2021 as the research object, the number of different series of trains flowing through the city is counted. There are 11 series of trains, including A series (x_1), C series (x_2), D series (x_3), G series (x_4), K series (x_5), L series (x_6), N series (x_7), S series (x_8), T series (x_9), Z series (x_{10}) and ordinary express trains (x_{11}).

Because some economically underdeveloped areas can not see its evolution trend and law, this paper selects 118 cities in the first, second and third-tier for statistical data analysis according to the "2020 ranking of China's urban commercial charm".The following table shows the cities represented by different regions of China selected in this data:

Table 1: Cities represented by different regions in China

Region	Some representative cities
Northeast China	Harbin, Changchun, Shenyang, Dalian, etc
North China	Beijing, Tianjin, Shijiazhuang, Cangzhou, etc
Central China	Wuhan, Changsha, Zhengzhou, Luoyang, etc
East China	Shanghai, Hangzhou, Suzhou, Nanjing, etc
South China	Guangzhou, Shenzhen, Dongguan, Foshan, etc
Northwest China	Yinchuan, Xi'an, Urumqi, Lanzhou, etc
Southwest China	Chengdu, Chongqing, Guilin, Guiyang, etc

3.1.2. Data Preprocessing

The number of trains in each year is divided into three categories: intercity railway (y_1), high-speed railway (y_2) and ordinary train(y_3), so according to the formula :

$$y_1 = x_2 \tag{1}$$

$$y_2 = x_3 + x_4 \tag{2}$$

$$y_3 = x_1 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} \tag{3}$$

to calculate the number of various trains each year.

The data must be preprocessed before K-means clustering analysis. There is too much difference in the number of trains in different series in the data in this paper. If we analyze it directly at this time, the clustering result may be biased to the side with larger data, which is inconsistent with the actual situation, so it needs to be standardized, ordering

$$y_i^* = \frac{y_i - \overline{y_i}}{\sqrt{\sigma_{ii}}} \tag{4}$$

for data analysis.

3.2. Research Method Description

3.2.1. Basic Thought

Cluster analysis is a conventional and unsupervised pattern recognition method. Without the interference and restriction of prior knowledge, cluster analysis can divide the data in the sample into several categories that can

reflect the internal structure of the data, so as to obtain the original information belonging to the data, and finally cluster the observed samples into the same category to complete the clustering process. A complete clustering process must include feature selection, nearest neighbor measure, clustering criteria, result verification, result judgment and other steps.

K-means clustering minimizes the sum of squares of the distance between all vectors in the data set and the center of the data set. The algorithm uses the sum of squares of errors as the clustering criterion. It optimizes the clustering results through repeated iterations, the minimum sum of squares of the distances from all samples to their respective category centers.

3.2.2. Algorithm Flow

Suppose the input sample is $S = a_1, a_2, \dots, a_n$,the algorithm flow is as follows:

Step 1: Select k samples（$u_1, u_2, \dots, u_k$）as the initial condensation point.

Step 2: The algorithm flow is as follows: for each sample x_i. Mark it as the distance from the center of the category:

$$label_i = argmin||a_i - u_i||(1 \le j \le k) \quad (5)$$

Step 3: Update each category center iteration to the mean of all samples belonging to the category:

$$u_j = \frac{i}{|C_j|}\sum_{i \in c_j} x_i \quad (6)$$

Step 4: Repeat step 2 and step 3 until the final change of the category center is less than a certain threshold.

The data set is analyzed by K-means clustering, and the relevant flow chart is as follows:

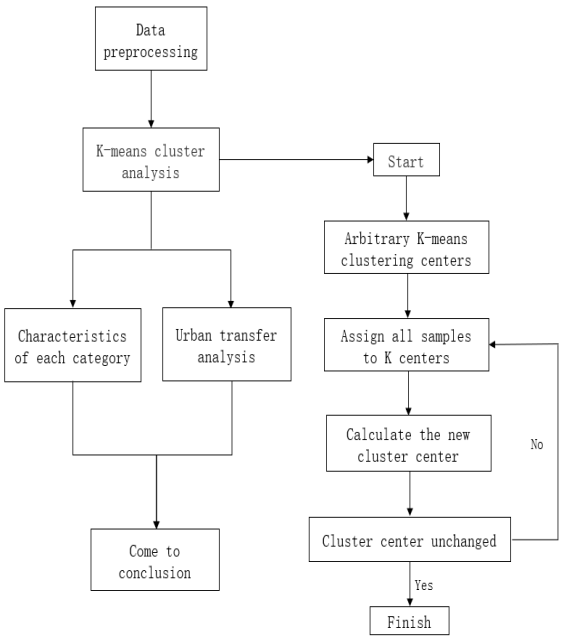

Figure 1: Article flow chart

4.THE RESEARCH ANALYSIS

4.1. Characteristics Of Railway Cities

Exploratory experiments were carried out on the data set, and the K value was determined as 8. Table 2 below presents the relevant statistical information from 2008 to 2021:

Table 2: Relevant statistical information of some years

Year	Statistical Data		
2008	High-speed railway		
	Minimum	Mean	Maximum
	0	7	149(Guangzhou)
	Ordinary train		
	Minimum	Mean	Maximum
	0	60	559(Beijing)
2009	High-speed railway		
	Minimum	Mean	Maximum
	0	14	225(Guangzhou)
	Ordinary train		
	Minimum	Mean	Maximum
	0	60	262(Beijing)

...	...	...	...
		Intercity railway	
	Minimum	Mean	Maximum
	0	38	414(Guangzhou)
		High-speed railway	
2020	Minimum	Mean	Maximum
	0	260	1130(Shanghai)
		Ordinary train	
	Minimum	Mean	Maximum
	0	113	372(Beijing)
		Intercity railway	
	Minimum	Mean	Maximum
	0	23	317(Guangzhou)
		High-speed railway	
2021	Minimum	Mean	Maximum
	0	264	1226(Nanjing)
		Ordinary train	
	Minimum	Mean	Maximum
	0	139	449(Beijing)

The clustering results and the quantitative characteristics of each series of trains are shown in Table 3:

Table 3: Final clustering results

Category	Intercity Railway	High-speed Rail	Ordinary Train
Category Ⅰ	non-existent	Maximum quantity	Maximum quantity
Category Ⅱ	Maximum quantity	Very large quantity	Very large quantity
Category Ⅲ	Some exist	Higher quantity	Higher quantity
Category Ⅳ	Some exist	Medium quantity	Higher quantity
Category Ⅴ	Some exist	Medium quantity	Medium quantity
Category Ⅵ	Some exist	Minority	Medium quantity
Category Ⅶ	Some exist	Minority	Minority
Category Ⅷ	Some exist	Very few quantity	Very few quantity

The first type of cities don't have intercity railways, but have the largest number of high-speed railways and ordinary trains. When there were no high-speed trains and only ordinary trains in 2008 and 2009, the representative cities were Guangzhou, Shenzhen, Shanghai and other first-tier cities in the developed Pearl River Delta and Yangtze River Delta. At this time, it already had the highest number of ordinary trains in China. After the opening of high-speed railway in 2010, the representative city is Guangzhou, with 283 high-speed railways. In 2021, the first-class cities have the highest number of high-speed railways, including 1188 high-speed railways in Shanghai and 1226 high-speed railways in Guangzhou. The first type of cities are also at the highest number level for normal trains.

The second type of city has the largest number of intercity railways. There are 317 intercity railways in Guangzhou, which reduces the number of high-speed railways and ordinary trains compared with the first type of cities. The representative cities are Nanjing, Shenzhen and other developed Pearl River Delta and Yangtze River Delta regions. There were 50-80 high-speed trains in 2008 and 2009, and about 1000 high-speed trains in recent years. The number of high-speed trains and ordinary trains is second only to the first category.

Some cities in the third category have intercity railways, with a high number of high-speed railways and ordinary trains. The representative cities are other first-tier cities like Beijing, Hangzhou, and Suzhou. Some developed cities have high intercity railways, and all cities have a high number of high-speed railways and ordinary trains. In 2008 and 2009, except for 81 bullet trains in Beijing, other cities have 40 bullet trains and about 100 ordinary trains. There will be 600 ~ 800 high-speed trains and 200 ~ 400 ordinary trains with economic development in 2021.

The same part of the fourth category cities has intercity railways, with a medium number of high-speed railways and ordinary trains. Compared with the third category, the number of high-speed railways decreases and ordinary trains increases. The representative cities are other first-tier cities such as Tianjin and Xi'an and developed second-tier cities like Shijiazhuang. In 2008, only some cities had 10 ~ 30 bullet trains and 80 ~ 140 ordinary trains. In 2021, there will be 250 ~ 550 high-speed trains and 250 ~ 450 ordinary trains.

The same part of the fifth category of cities has intercity railways, including 207 intercity railways in Chengdu, with a medium number of high-speed railways and a small number of ordinary trains. The representative cities are the provincial capitals of remote cities such as Chengdu, Chongqing and Harbin. Compared with the fourth category, the number of high-speed railways is the same, and the number of ordinary trains is reduced to about 100.

The sixth category of cities is the most developed second and third-tier cities such as Jiaxing and Dalian. Only Changchun has 130 intercity high-speed trains. Compared with the fifth category, the number of high-speed trains is reduced, and the number of ordinary trains is the same. There are a few high-speed railways and ordinary trains in the seventh category of cities, and the representative cities are Lanzhou, Putian and other medium-sized second and third-tier cities. The eighth category of cities are Yantai, Taizhou and most of the other third-tier cities, with only a small number of high-speed trains and ordinary trains.

4.2. Urban Transfer Analysis

In the first category of cities, in early 2008, the representative cities were Guangzhou and Shenzhen in the Pearl River Delta. In the middle of 2013, the representative cities increased to Shanghai, and gradually developed into the Yangtze River Delta in 2020 and 2021. In early 2008, the representative cities of the second type of cities were Shanghai and most of the Yangtze River Delta, and gradually developed into the Pearl River Delta in 2020 and 2021. Since 2008, due to the large-scale withdrawal of foreign capital, the economic development of the Pearl River Delta has experienced a cold current. However, the World Expo was held in Shanghai in 2010, and the economic operation quality of the Yangtze River Delta has been steadily improved. Therefore, the first type of representative cities have been gradually transferred to the Yangtze River Delta, and the second type of representative cities have been gradually transferred to the Pearl River Delta.

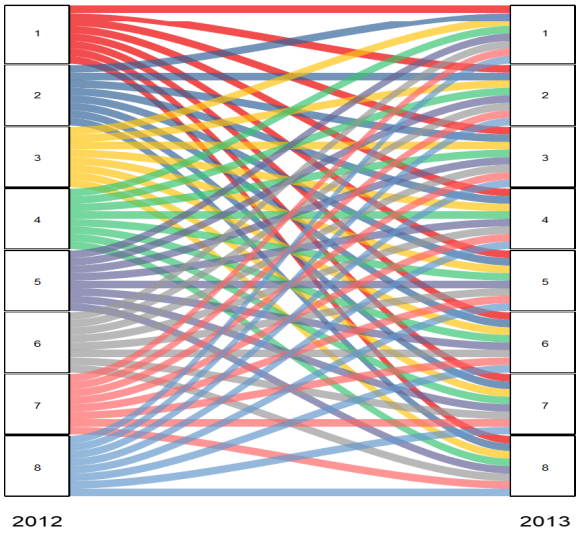

Figure 2: Alluvial change from 2012 to 2013

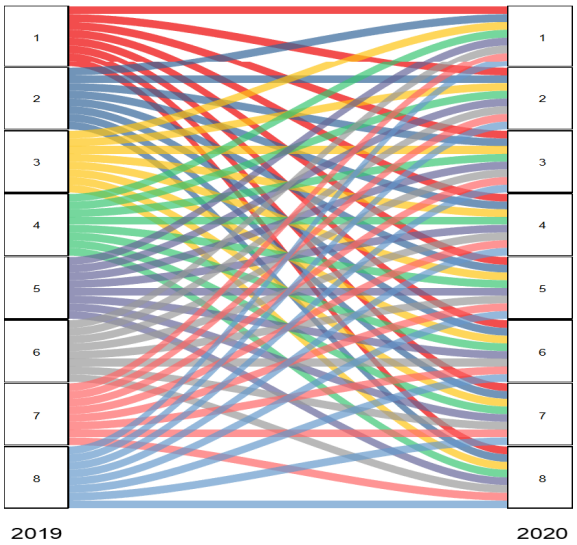

Figure 3: Alluvial change from 2019 to 2020

In the early years of 2008 and 2009, Tianjin was in the fifth category. In the middle of 2014, it was in the same category as Shanghai. In 2015, it was in the same category as Shenzhen and Wuhan, and fell back to the fifth category in 2021. In the medium-term years, Tianjin has the leading role of some high-tech enterprises such as Tianjin port, eco-city, and Binhai, and the railway industry has developed vigorously. However, Tianjin's high-tech and chemical industries have gradually shifted to Beijing in recent years. Because it is close to Beijing, and there are many Bohai Sea ports, Tianjin has reduced many advantages. Therefore, the economic development of Tianjin has gradually declined in recent years. As one of China's four municipalities directly under the central government, Tianjin ranked more than 20 cities in the province in 2020 and 2021.

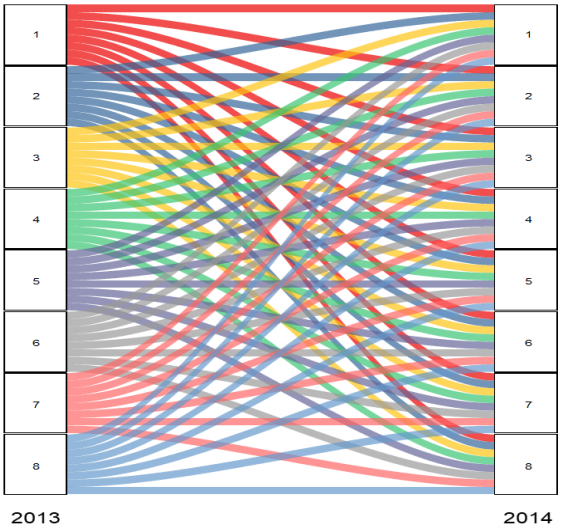

Figure 4: Alluvial change from 2013 to 2014

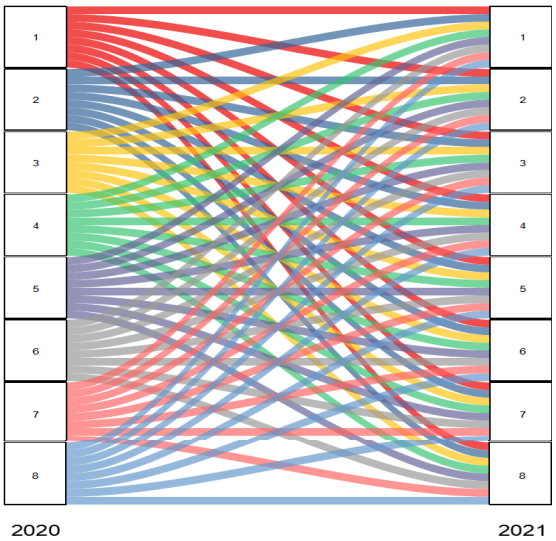

Figure 5: Alluvial change from 2020 to 2021

Harbin and Shenyang, the two capital regions of Northeast China, were in the same category as Shanghai and Beijing in 2008 and 2009. With time development, Shenyang is in the fourth category, and Harbin is in the sixth category in 2020 and 2021. Shenyang and Harbin were once heavy industrial centers in Northeast China and even the whole country. They are also famous cultural cities with a long history. In recent years, the economy of Northeast China has been declining continuously. Even under the plan of revitalizing northeast China, there is still no major reversal. Therefore, the economy has also been falling continuously in recent years. Accordingly, transportation and economy are inseparable, and the most direct manifestation is the reduction of the number of trains.

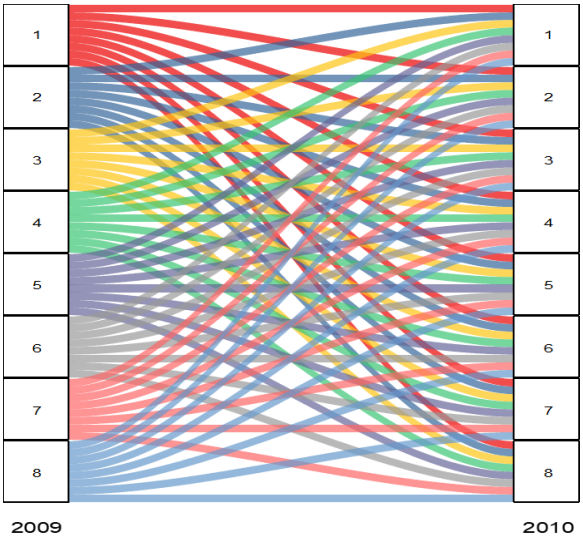

Figure 6: Alluvial change from 2009 to 2010

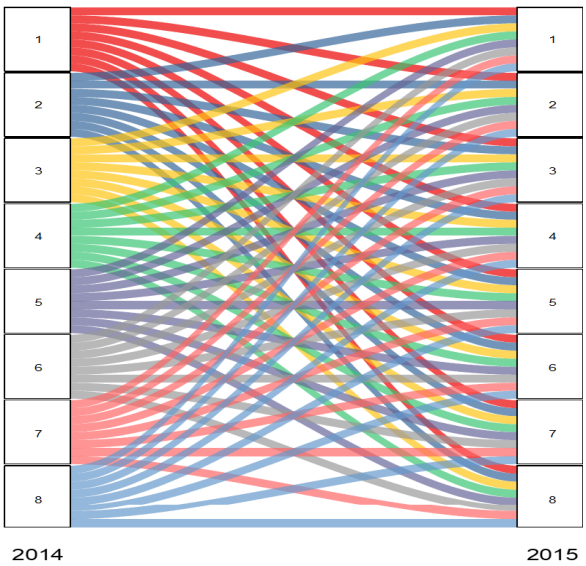

Figure 7: Alluvial change from 2014 to 2015

5.CONCLUSIONS

With the continuous development of economy and the necessity of people's travel, the number of high-speed, intercity, and Ordinary trains has been increasing from 2008 to 2021. Based on K-means clustering, the conclusions are as follows:

1. There is a strong corresponding relationship between the overall train operation and the urban administrative level. From the general situation in 2008-2021, the provincial capital cities have more intercity railways, high-speed railways and normal trains than other cities in a province. Moreover, among the cities along the high-speed railway, the number of high-grade cities passing far exceeds that of low-grade cities.

2. The number of trips is concentrated in the eastern urban areas with a developed economy, and the train number density in the East and middle is greater than that

in the Northeast and West. According to the clustering results, the developed urban areas in the East are concentrated in the first three categories, the central region in the fifth category, the northeast region in the sixth category, and the western region in the seventh and eighth categories.

3. Economic factors are the fundamental driving force for the generation of high-speed rail passenger flow. Although China's high-speed rail network has gradually covered all cities, it can be seen that the Yangtze River Delta, Pearl River Delta and Beijing Tianjin Hebei are hot spots of high-speed rail, which have a significant position in the overall railway in the country. Moreover, these three urban agglomerations also have a significant economic position in the whole country.

By analyzing the characteristics of high-speed railway cities and representative cities in China, this paper discusses the evolution trend of train operation in China. Because the research content is affected by many factors, on the basis of this paper, combined with K-means clustering algorithm for follow-up research, it is helpful to further explore the development trend of high-speed railway. In addition, the evolution trend of high-speed rail operation is affected by the natural geographical environment. Although the cluster analysis of high-speed rail cities is carried out in this paper, there are still deficiencies in the analysis, which is also a problem that needs to be paid attention to in the follow-up research.

REFERENCES

[1] Chu Nanchen, Zhang Pingyu, Jiang Bo. (2018). Spatial characteristics of China's high-speed railway network from the perspective of daily high-speed railway flow.J. Geographical Research,37:2193-2205.

[2] Chu Nanchen, Zhang Pingyu, Li He, Jiang Bo. (2019).Pattern characteristics and evolution trend of intercity high-speed rail network in Northeast China. J. Geoscience, 39:761-769.

[3] Jiang Haibing. (2019). Research on spatial efficiency and supply-demand relationship of China's high-speed rail lines from the perspective of accessibility. J. Resources and environment in the Yangtze River Basin, 28: 2299-2308.

[4] Jiang Haibing, Qi Yi, Li Chuanwu. (2018).Spatial connection pattern of urban high-speed railway passenger transport in China. J. Economic geography, 38:26-33.

[5] Wei Sheng, Xu Jiangang, Ma Haitao. (2019). Structural characteristics and formation mechanism of high-speed railway network in Yangtze River Delta. J. Resources and environment in the Yangtze River Basin, 28:739-746.

[6] Wang Liehui, Xia Wei, Ning Yuemin.(2017). Analysis on the unbalanced distribution pattern of High-speed railway cities in China--Based on the perspective of comparison with ordinary railway.J. Urban Development Studies,24:68-78.

[7] Zhao Yinghui, Chu Nanchen, Guo Jingpeng. (2017). Structure and characteristics of high-speed railway network in Three urban agglomerations in China.J.Economic Geography,37:68-73.

[8] Zhong Yexi, Guo Weidong. (2020). Structural characteristics and organizational model of China's high-speed rail network .J. Geographical science, 40:79-88.