

# Research on the Development of Rail Transit Based on Cluster Analysis

Guoling Lang<sup>1,\*</sup>

<sup>1</sup>China Urban Construction Design & Research Institute Co. LTD, Beijing, China

**Abstract.** This article analyzes the current development status of rail transit in various cities in China. Most cities are still dominated by subways. Then the relevant data is processed to obtain the linear relationship between transfer stations and rail mileage, rail stations and summarize the domestic rail Establish quantitative standards and conclude that rail transit has not yet become a transportation network. Then use the five indicators of line mileage, number of stations, number of interchange stations, average daily passenger volume, and passenger transport intensity to cluster the domestic urban rail transit using the method of cluster analysis, combining spatial Euclidean distance and compactness and other indicators Analysis, the classification results obtained have certain reference value.

## 1 Introduction

In 2021, a total of 45 cities in mainland China currently have rail transit, with an operating mileage of nearly 8,000 km. The newly added rail transit mileage in 2020 is about 1,250 km. Our country has devoted a lot of effort to rail transit, invested a huge amount of money, and a large number of talents poured into the construction of rail transit. Among the rail transit operating lines in my country, 79% are subway stations, 2.73% light rail, 1.23% monorail, 10.1% urban fast rail, 6.09% modern tram, 0.72 maglev, and APM the ratio is 0.13%. Overall, the subway is still the main force in the development of my country's rail transit.

## 2 Analysis on the development characteristics of urban rail transit

The unevenness of the track development between cities comes from the influence of multiple angles. The main reason is that the population size and economic level of the cities are quite different. However, the growth rate of urban rail in various cities as a whole shows a certain regularity. According to the data collected The data of the domestic urban rail transit is fitted, and the first letter of each city is simplified to represent the name of the city. Selected 40 representative cities in China for analysis to explore the relationship between the various attributes of rail transit, including the relationship between line mileage, number of stations, number of interchange stations, average daily passenger volume, and passenger intensity.

### 2.1. The relationship between line mileage and the number of stations

The two cities with the longest rail transit mileage in China are Beijing and Shanghai, with mileages of 771.8km and 809.9km, respectively. The two cities also have the largest number of rail stations. According to the data fitting situation, the number of rail stations and the mileage The relationship is linear growth, the accuracy of the fitting is as high as 0.95, and the slope of the curve is close to 0.5. This indicates that there are 0.5 stations for every 1km increase in rail transit mileage, that is, the average distance between domestic rail stations is about 2km, which is also consistent with National construction standards.

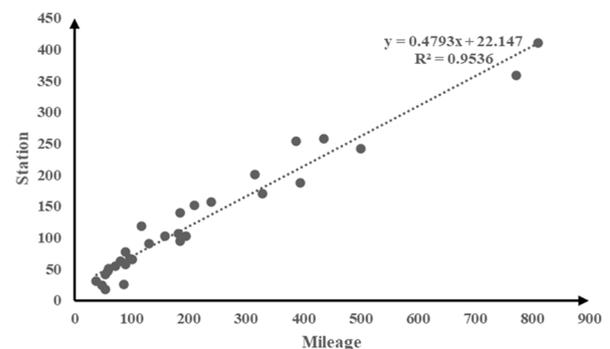


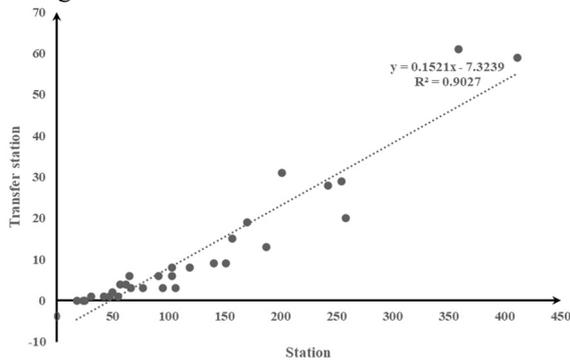
Figure 1 Linear relationship between mileage and station

### 2.2 The relationship between the number of stations and transfer stations

The transfer station refers to the station where the line crosses, including two lines, three lines, four lines, etc., that is, the stations that can be transferred. With the

\* Corresponding author: 874083329@qq.com

number of stations, the transfer station also shows a linear growth trend, and the fitting accuracy reaches 0.9. From the data situation, it can be interpreted that for every 10 additional stations, 1.5 stations will become transfer stations. From the perspective of the rail network, the rail transit of most cities in China has not yet formed a network structure. When the urban network structure continues In the process of formation, the number of transfer stations will increase faster, and the possibility of crossing between rail lines in the network will be greater. Compared with the track scale without a network structure, the network structure will be that of most cities in China. Will enter a new stage.

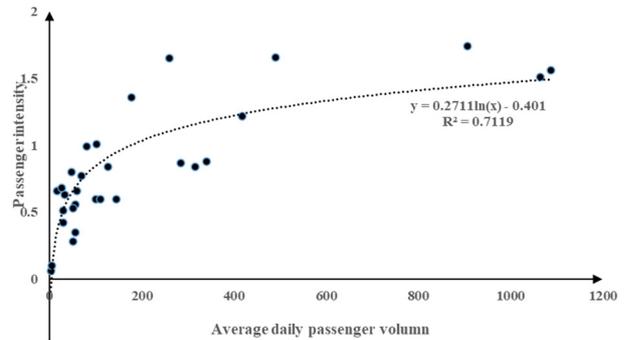


**Figure 2** Linear relationship between stations and transfer stations

### 3 Research on the relationship between rail passenger flow and stations

#### 3.1 The relationship between daily average passenger flow and passenger flow intensity

The average daily passenger flow is an important indicator of the scale of rail transit. The daily average passenger volume of each line is greatly affected by the planning and layout of the surrounding land of the line, and the gap in this indicator is also large. For example, the current daily average passenger transport of Chongqing Line 3 The volume can reach about 800,000, while the passenger flow of Line 4 is only tens of thousands. There is still a large gap in the average daily passenger volume of different lines in the same city. Passenger transport intensity refers to an indicator of the efficiency of rail transit transportation capacity. If the passenger transport intensity is greater, it means that the city's rail transit network layout is better and the passenger flow efficiency is higher. This section mainly wants to explore whether there is a certain correlation between the two, and use curve fitting to analyze and process the two. The results are shown in the figure:

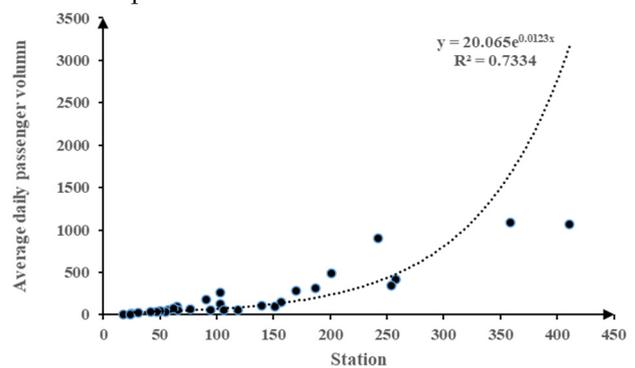


**Figure 3** The relationship between daily average passenger flow and passenger flow intensity

The curve basically conforms to the logarithmic function. The surface passenger flow intensity does have such a logarithmic relationship with the passenger volume within a certain range, and the fitting accuracy reaches about 0.7.

#### 3.2 The relationship between the number of stations and the average daily passenger volume

Through the curve fitting of the number of stations and the average daily passenger volume, it is found that the two have an exponential relationship, and the fitting accuracy reaches about 0.7. The passenger volume has shown a curve growth with the increase of the number of stations, but the network has not yet been established in the initial stage This index relationship is not obvious when the number of stations is between 100 and 400. This relationship is more obvious when the number of stations is between 100 and 400. The index relationship is not obvious in the interval between less than 100 and greater than 400. Therefore, other relationships may exist. The construction of the planned network situation and the route direction have a greater impact, which is worthy of further in-depth discussion.



**Figure 4** The relationship between the number of stations and the average daily passenger volume

### 4 Cluster analysis processing

Clustering analysis is to group the data according to the information found in the data to describe the objects and their relationships. The purpose is that the objects in the group are similar to each other, but the objects in different

groups are not related. The greater the similarity within the group, the greater the gap between the groups, the better the clustering effect. In other words, the goal of clustering is to obtain a higher degree of similarity within clusters and a lower degree of similarity between clusters, so that the distance between clusters is as large as possible, and the distance between the samples in the cluster and the cluster center is as small as possible.

#### 4.1 Clustering process

The first is to prepare the data, including a certain standardization of the data; the second is the screening of characteristics, the most obvious characteristics are selected as the main objects of the research, and they are saved; then is to carry out the selection of these characteristic objects. Categorize, summarize the characteristics of such objects, then perform clustering processing, select metrics to perform cluster analysis on such objects, and finally evaluate the conclusions after cluster analysis. There are many methods of cluster analysis, choosing the right method can make the result closer to the actual situation.

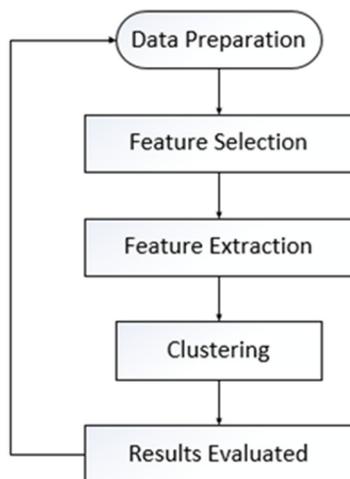


Figure 5 Cluster analysis process

#### 4.2 Cluster analysis to analyze the characteristics of urban rail transit

Different methods of cluster analysis will have a great impact on the results. Common examples in cluster analysis include Euclidean distance, Manhattan distance, Chebyshev distance, and Minkowski distance. Compared with the related algorithms shown, this article uses DBSCN to process the data and adopts the Euclidean distance.

Table1 Comparison of different types of algorithms

Name of algorithm	Type of algorithm	Data type	Anti-jamming	Algorithmic efficiency
ROCK	Hierarchical clustering	Hybrid	Very high	Noraml

BIRCH	Hierarchical clustering	Numerical	Lower	Very high
CURE	Hierarchical clustering	Numerical	Very high	Higher

Table2 Various types of algorithms

Name of algorithm	Type of algorithm	Data type	Anti-jamming	Algorithmic efficiency
CLARANS	Partition clustering	Numerical	Higher	Lower
DENCLUE	Density clustering	Numerical	Noraml	Very high
DBSAN	Density clustering	Numerical	Higher	Noraml

Euclidean distance is the simplest and most intuitive method. The calculation formula is as follows:

$$dict_{ed} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

Minkowski distance is a measure of Euclidean space.

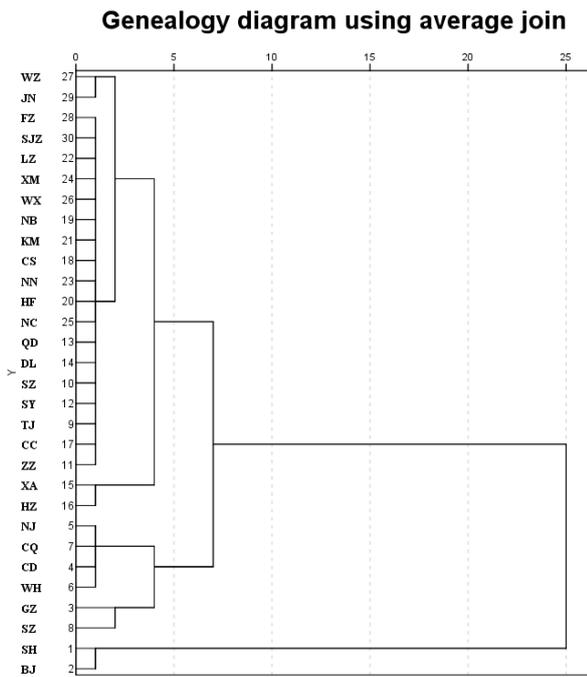
$$dict_{mind} = \sqrt[p]{\sum_{k=1}^m |x_{ik} - x_{jk}|^p}$$

Standardize the data, using Z-score standardization:  $Z_f = X_f - mean_f / S_f$ , using the compactness index to describe the distance from the sample point to the cluster center point.

$$CP = \frac{1}{k} \sum_{i=1}^k CP_i$$

$$DP = \frac{2}{k^2 - k} \sum_{i=1}^k \sum_{j=1}^k ||c_i - c_j||$$

Select five indicators of line mileage, number of stations, number of interchange stations, average daily passenger volume, and passenger intensity, and perform a cluster analysis of rail transit in domestic cities. The results are as follows:



**Figure 6** Cluster pedigree map of rail cities

It can be seen from the pedigree map that rail transit cities can be roughly classified into five categories, which are the five stages of rail transit development. The first stage is Beijing and Shanghai, where rail transit is the most developed, and the second stage is Shenzhen, which is more developed. But it still cannot reach the level of Beijing and Shanghai. The third stage is Chongqing, Chengdu is relatively developed, the fourth stage is Xi'an, etc., which are generally developed, and the fifth stage is Harbin, etc. It is still in its infancy and the facilities are not perfect.

## 5 Conclusion

Through combing the status quo of the development of urban rail transit in China, curve fitting method is used to obtain the relationship between rail station and line mileage, transfer station, and cluster analysis method is used to obtain the classification of urban rail transit. The construction of urban rail transit provides a certain reference and gives some thoughts to the network of rail transit.

## References

1. Guo Ge. Cluster analysis and application of Beijing urban rail transit stations based on passenger flow characteristics and POI data[D]. Beijing Jiaotong University, 2020.
2. Xu Wei, Du Juan Research on the relationship between urban rail transit and regional economic development based on cluster analysis[J]. Value Engineering,2018,37(03):10-12.
3. Shao Yingyu, Ding Baiqun. Classification of Metro Stations Based on Cluster Analysis—Taking

Harbin Metro Line 1 as an Example[J].Forest Engineering,2015,31(03):106-111+116.

4. Chen Jun. Research on the relationship between urban rail transit and regional economic development based on cluster analysis[J].Science and Technology Economic Market,2018(12):45-46+49.
5. Qiao Ke, Zhao Peng, Yao Xiangming, Wen Jiaying. Cluster analysis of urban rail transit nodes[J]. Logistics Technology, 2014,33(05):92-95+256.
6. Gong Tongwei, Yun Yingxia. A method of function identification of urban rail transit station area based on factor analysis and cluster analysis[J].Statistics and Decision,2020,36(05):177-180.
7. Xu Wei, Zheng Changjiang, Ma Genghua, Li Rui, Deng Pingxin. Research on Urban Rail Transit Station Classification Based on k-means Clustering[J]. Journal of Guizhou University (Natural Science Edition), 2018,35(06):106- 111.