

Research Article

Metro Train Operation Plan Analysis Based on Station Travel Time Reliability

Ruihua Xu,^{1,2} Fangsheng Wang ^{1,2} and Feng Zhou ^{1,2}

¹Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China

²Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Fangsheng Wang; wangfangsheng@tongji.edu.cn and Feng Zhou; zhoufeng24@tongji.edu.cn

Received 18 July 2020; Accepted 3 April 2021; Published 15 April 2021

Academic Editor: Zhixiang Fang

Copyright © 2021 Ruihua Xu 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.

The train operation plan plays an essential role in metro systems and directly affects transportation organization efficiency and passenger service level. In metro systems, passengers have paid more attention to the travel time reliability (TTR), reflecting the reliability of metro operation management. This article proposes an analysis method of train operation plan based on TTR in the station dimension. First, an automated fare collection (AFC) data-driven framework is established to calculate the station travel time reliability (STTR) and analyze the train operation plan at different periods. The framework structure consists of four steps: AFC data preprocessing, STTR calculation and assignment, clustering algorithm design based on SOM neural network, and train operation plan analysis and optimization. Second, the proposed method is applied to the Beijing metro network as a case study. Several promising results are analyzed that allow the optimization of the existing train operation plan. Our research shows that STTR is a good supplement for the existing metro operation assignment studies, which can help analyze and optimize the train operation plan effectively. This study is also applicable to other metro networks with AFC systems.

1. Introduction

With the ongoing socioeconomic development, urban traffic congestion has become increasingly severe, especially in large cities like Beijing and Shanghai. Metro is playing an increasingly important role in urban public transportation, owing to the outstanding advantages of faster velocity, higher reliability, and larger capacity. With the continuous expansion of the scale of metro networks, passenger demand shows a high-speed growth, while the distribution of passenger demand presents the unbalance characteristics in the time-space dimension. There is an increasingly prominent contradiction between the transportation capacity supply and the passenger flow demand, and it puts forward higher requirements for train operation organizations under networked operating conditions.

As a significant part of the metro operation and management, the train operation plan directly affects transportation organization efficiency and passenger service level. In many large cities, train running intervals are continuously shrinking

during the morning rush hours, while some stations are still highly congested. Passengers have particular travel characteristic, which generally concentrates in individual stations or periods. Restricted by objective conditions of metro network structure, metro transport capacity cannot meet the passenger flow demand at specific locations and periods, resulting in severe partial congestion in the network. The fundamental reason is that the configuration of network transport capacity does not match the distribution of passenger travel needs in the space-time dimension.

Recent research focused on extracting relevant indexes to reflect the train operation plan quality, such as train full-load rate [1–4] and platform congestion degree [5–7]. On one hand, the existing research methods screen the top/bottom ranking sections/stations according to the operation indicators, including section full-load rate and station passenger volume, and essentially sort operation indicators and get the concerned sections or stations, but cannot obtain the potential causes. On the other hand, passengers have

paid more attention to the travel time reliability (TTR) in public transportation, and TTR has been one of the most significant factors affecting transportation services level [8, 9]. In general, passengers will always ride the first train to arrive after they reach the platform unless the train is too crowded. When the transportation capacity cannot meet the passenger demand in some stations and sections, there will be a backlog of passengers waiting in station platforms and, thus, there will be a direct impact on TTR. As for an OD pair in the metro network, TTR typically has two definitions: (1) the probability that passengers can complete a trip within a specified time; (2) the fluctuation degree of the average travel time of passengers.

As the node for passengers to start and finish the journey in the train operation plan, stations are the core of transportation organizations in the metro system. Based on the characteristics of metro network structure and operation management, we propose the definition of *station travel time reliability* (STTR) as the fluctuation degree between the actual time and standard travel time of each OD from this station as the starting station to other stations. Based on the support of big data, STTR analyzes and evaluates the TTR of inbound passenger flow totally to reflect the passenger service level at different stations and periods. Combined with operation experience and travel investigations, the factors that affect the fluctuation of STTR value consist of the following three aspects:

- (i) Passenger flow of the station is excessive.
- (ii) Train running interval of the line is large; that is, the transportation capacity is insufficient.
- (iii) Station location: trains are too crowded when arriving at the station because their capacity has been used in front of this station.

To sum up, the train operation plan analysis should not be limited to the ranking of indicators but also should pay attention to the analysis of potential causes. This study aims to develop a data-driven approach to analyze train operation plans based on the STTR of all stations in the network. The contributions of this article are as follows:

- (1) Based on the AFC data, an STTR measurement model is built to calculate the value of passenger TTR from station dimension and principal component analysis (PCA) is used to process clustering elements.
- (2) Combining the Self Organizing Maps (SOM) neural network, a station clustering framework is established with the STTR values and influence factors to analyze the train operation plan more objectively and comprehensively and explore the specific reasons for low STTR level.
- (3) Apply the proposed approach to the Beijing metro as a case study, and several results are analyzed that inspire the optimization of the existing train operation plan.

2. Literature Review

Numerous studies in the literature related to the train operation plan analysis consist of the following aspects:

operation organization [1–4] and station service level [5–7]. Li et al. [1, 2] constructed an interaction model of trains and passengers and obtained evaluation indexes such as full-load rate, number of passengers, and average waiting time and optimized the train operation plan based on the matching degree of capacity supply and passenger demand. Wang et al. [3] evaluated the adaptability of train operation schemes and passenger demand from three aspects: total adaptability, structural adaptability, and quality adaptability. Lu [4] divided transportation efficiency into three levels: capacity output efficiency, capacity utilization efficiency, and transport demand satisfaction efficiency. Tian [5] used the passenger flow aggregation and congestion as an indicator to measure the service level of the station and as one of the bases for the preparation of the train operation plan. Liu and Chen [6] used the minimization of factors such as the waiting time of passengers at the station as the objective function to establish a multiobjective nonlinear mixed-integer optimization model evaluates and optimizes the line operation plan. Shafahi and Khani [7] considered the minimum transfer waiting time as the goal and combined heuristic algorithms to optimize the transfer of the road network. We find these analysis methods focused on restoring the passenger travel process to extract relevant indexes, such as train full-load rate, platform waiting time. However, in the process of path restoration, the parameters such as the passenger walking time and the train maximum passenger capacity will have a few differences and fluctuations in the space-time dimension. The pattern of empirical values for these parameters will cause vast subjectivity and randomness in evaluation results.

The theory of TTR was first proposed on urban road traffic, and there are several types of research about the TTR analysis in public transportation. Considering travel behavior analysis in the road network, Asakura and Kashiwadani [10] gave a concept of TTR, the probability that passengers can complete the trip within the specified time, and measured TTR of an OD pair in a deteriorated road network [11, 12]. Lam and Xu [13] calculated TTR's value by establishing a traffic flow simulator model and access the reliability of metro systems organization management. Bell and Chirs [14] analyzed travel time change based on sensitivity analysis and described TTR by travel time variance. While some scholars [15, 16] used the buffer time index (BTI) to describe the TTR, BTI is the fluctuation degree between the actual and planned travel time at a specified period. Besides, Lomax et al. [16] defined the unit distance travel time and defined the BTI as the rate between the average travel time and the time of passengers having a 95% chance of arriving at the destination on time.

To our knowledge, little attention has been paid to introducing TTR to the train operation plan analysis in the station dimension, which is of great significance to the metro operation management. Zhang et al. [17] presented a new unit distance TTR evaluation index and method to assess the Beijing metro network. Li et al. [18] proposed a TTR calculation algorithm to analyze the reliability of transfer time quantitatively. Chen [19] proposed the definition and evaluation method of metro network operation reliability

and established a train operation delay propagation model. Based on the data-driven method [20, 21], this article focuses on calculating the STTR and analyzing and optimizing the train operation plan combined with the clustering algorithm.

3. Data Description

3.1. AFC Data. The study addressed in this article requires passenger travel time data extracted from the automated fare collection (AFC) data. The AFC system has become the primary method of collecting metro fares in many cities throughout the world. AFC system provides a large quantity of passenger flow information, recording passengers' activities with original station ID, destination station ID, tap-in time, and tap-out time. Necessary elements for the model formula are summarized (Table 1).

3.2. Train Diagram. The train diagram illustrates the relationship between space and time for train operation (Figure 1). Necessary elements for the model formula are summarized (Table 2). According to the train diagram data, we can extract each line's running interval at different periods for the model formula.

4. Methodology

As mentioned, the existing analysis methods emphasize screening the top/bottom ranking sections/stations according to the operation indicators, which are essentially the index ranking methods. However, an increasing number of researchers and professionals have identified shortcomings in traditional analysis methods. For example, these indicators may be subject to bias and error in evaluation results. Moreover, the manual methods usually only focus on getting the concerned sections or stations but cannot obtain the potential causes. For these reasons, alternative concepts and methods need to be developed. This article proposes a cluster-driven method for analyzing the train operation plan, consisting of four steps: AFC data preprocessing, STTR measurement model, cluster-based analysis method, and train operation plan optimization.

Step 1: AFC data processing

Input AFC data, calculate the lower and upper bound of each OD pair's travel time thresholds, and remove abnormal records that are not between the lower and upper bound of thresholds.

Step 2: STTR measurement model

Based on the Cumulative Chance Measurement Model (CCMM), calculate the values of STTR (NSTTR and LSTTR) by the actual and standard of travel time, and PCA is used to process clustering elements.

Step 3: Cluster-based analysis method

Introduce SOM neural network to clustering algorithm for station classification and explore the specific reasons for low STTR level.

Step 4: Train operation plan optimization

By combining the level of STTR and influencing factors, including passenger flow, train running intervals, and station location coefficient, analyze stations characteristics of different clusters, and design appropriate optimization measures in train operation plans for low-reliability stations and lines. For the convenience of model formulation, relevant sets and parameters are listed in Table 3.

4.1. AFC Data Processing. In general, passengers' travel time between the same OD will be within a reasonable section. Typically, the threshold of the route travel time is determined by the results of travel surveys. First, obtain the actual travel time set of each OD by extracting each passenger travel time from the network AFC ticket dataset. Passenger travel time is the difference between the passenger's tap-in time and tap-out time in the smart card. Secondly, sort the actual travel time data for each OD pair in ascending order. The lower and upper bound of the travel time threshold of the OD (Station i to Station j) are obtained from the following formulae:

$$\begin{aligned} t_{ij}^{\text{low}} &= t_{\text{ceil}(5\% * M_{ij})}, \\ t_{ij}^{\text{up}} &= \min(t_{ij}^{\text{low}} * (1 + a), t_{ij}^{\text{low}} + U), \end{aligned} \quad (1)$$

where t_{ij}^{low} is the lower bound of the travel time threshold; t_{ij}^{up} is the upper bound of the travel time threshold; $t_{\text{ceil}(5\% * M_{ij})}$ is the actual travel times value for the fifth percent [22]; M_{ij} is the number of passengers; a is the relative threshold coefficient; U is the absolute threshold.

The values of a and U are determined through travel surveys; normally, a is 0.6 and U is 20 minutes [22]. Then, the data with the actual travel time at $[t_{ij}^{\text{low}}, t_{ij}^{\text{up}}]$ are retained, and the noise data are removed for each OD travel time set.

4.2. STTR Measurement Model. The measure indicating TTR includes two types: probability and fluctuation. The former indicates the probability that the passenger could complete the trip within the specified time, and the latter reflects the fluctuation degree between the actual and planned travel time. The study in this article focuses on the quantitative relationship between passenger travel time and train operation plan so that we decide to use the fluctuation indicator as the basis of the model.

As distinct from manual methods, the proposed method integrates multiple indicators (STTR, passenger flow, train running intervals, geographic location, etc.) for cluster analysis and classification of stations. Thus, through the analysis of various categories, we can evaluate the operation effect of the train operation plan of stations and lines. Therefore, we propose a measurement model to calculate the STTR value and analyze the correlation between STTR values with these factors and provide the basis for cluster analysis in the next charter.

Firstly, the lower bound of the travel time threshold (t_{ij}^{low}) is used as the standard travel time (t_{ij}^{stand}) of the OD, and the passengers' TTR of one OD pair (station i to station

TABLE 1: Necessary elements of AFC data.

Date	Original station ID	Tap-in time	Destination station ID	Tap-out time	Smart card no.
20161018	000521	07:15:00	001313	09:19:15	200141995
20161018	000551	10:04:56	000429	10:21:20	200141996
20161018	000535	10:02:32	000559	10:45:23	200141997
.....					

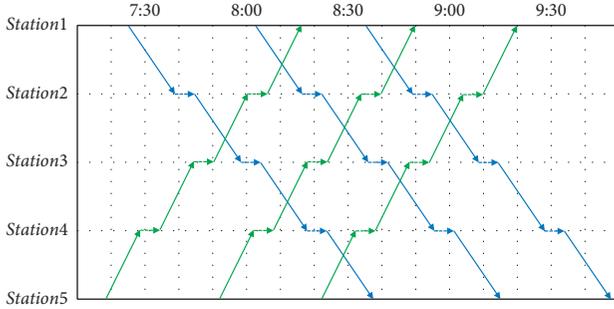


FIGURE 1: Example of actual train diagram.

j) is measured by the average and standard of travel time, as shown in equation (2).

$$TTR_{ij} = \frac{(t_{ij}^{ave} - t_{ij}^{stand})}{t_{ij}^{stand}}, \quad (2)$$

where t_{ij}^{ave} is the average travel time between i and station j .

We divide STTR into Network STTR (NSTTR) and Line STTR (LSTTR). NSTTR is the relationship between this station and all other stations in the network, whereas LSTTR is the relationship between this station and all other stations in the same line. Based on the CCMM presented in TTR studies [23], we measure the STTR (NSTTR, LSTTR) of station i , as shown in equation (3).

$$NSTTR_i = \frac{\sum_{j \in S, j \neq i} (TTR_{ij} * t_{ij}^{stand} * M_{ij})}{\sum_{j \in S} M_{ij}}, \quad (3)$$

$$LSTTR_i = \frac{\sum_{j \in S_i, j \neq i} (TTR_{ij} * t_{ij}^{stand} * M_{ij})}{\sum_{j \in S_i} M_{ij}},$$

where S is the stations set of the metro network and S_i is the stations set of the line which station i belongs to.

Secondly, according to the previous analysis in this article, the influencing factors for the STTR level include passenger flow, train running intervals, and station location, so that we use the three influencing factors and the values of NSTTR and LSTTR as clustering elements:

- (i) *Passenger flow*: the inbound passenger flow of the station, that is, the total OD passenger flow with the station as the departure station during this period;
- (ii) *Train running intervals*: the train operation plan running intervals of the line where the station locates during this period;
- (iii) *Station location coefficient*: analyze the geographic location of all stations in the network, and extract

the central station, and set its station location coefficient as 0, and the station location coefficients of other stations are determined by the OD standard travel time from it to the central station;

- (iv) NSTTR;
- (v) LSTTR.

These five elements are the add-in values in this model. The first three are derived from the passenger flow statistics system, train operation plans, and geographical statistics data. Moreover, the calculations of NSTTR and LSTTR are directly related to AFC data. Traditional evaluation methods focus on the calculation and simple ranking of indicators, but the specific causes of the station or line's poor indicators are not enough.

Thirdly, we use PCA to reduce the clustering elements' dimensional reduction and analyze the correlation between STTR values with these factors. As a multivariate statistical method based on orthogonal transformation, PCA indexes multiple related variables of the research object into a few unrelated variables and retains feature vectors with significant contributions [24]. These unrelated comprehensive variables include most information provided by the original variables, thereby achieving dimensionality reduction. Specific steps are as follows:

Step 1: Normalization

Scale clustering elements to a normal distribution with a mean of 0 and a variance of 1.

Step 2: Correlation coefficient matrix

Compose the normalized clustering elements into a 5-dimensional random vector:

$$X = (x_1, x_2, x_3, x_4, x_5), \quad (4)$$

where the covariance of x_m and x_n is the correlation coefficient of them, namely,

$$p_{m,n} = Cov(x_m, x_n). \quad (5)$$

The correlation between x_m and x_n is

- (i) positive correlation, when $p_{m,n} > 0$;
- (ii) negative correlation, when $p_{m,n} < 0$;
- (iii) irrelevant, when $p_{m,n} = 0$.

The larger the absolute value of $p_{m,n}$, the stronger the linear correlation of x_m and x_n . Finally, obtain the correlation coefficient matrix $D(X)$ of X .

Step 3: Principal components extraction

Extract the feature root of $D(X)$ and convert it to the corresponding standard feature vector μ_k , which is the

TABLE 2: Necessary elements of train diagram.

Date	Train ID	Destination station ID	Stop station ID	Stop order	Arrival time	Departure time
20161018	0429	000941	000933	5	07:12:15	07:12:45
20161018	0431	000941	000933	5	17:15:20	17:15:50
20161018	0103	000521	000545	13	10:45:23	10:45:53
.....						

TABLE 3: Relevant sets and parameters in model formulation.

Sets/parameters	Definition
M_{ij}	Passengers number of OD (station i to station j)
a	The relative threshold coefficient
U	The absolute threshold
t_{ij}^{low}	Travel time threshold lower bound of OD (station i to station j)
t_{ij}^{up}	Travel time threshold up bound of OD (station i to station j)
t_{ij}^{stand}	The standard travel time value of OD (station i to station j)
$t_{ij}^{\text{ceil}(5\%*M_{ij})}$	Actual travel times value for the fifth percent
t_{ij}^{ave}	Travel time average value of OD (station i to station j)
TTR_{ij}	Travel time reliability value of OD (station i to station j)
S	Stations set of the metro network
S_i	Stations set of the line which station i belongs to
NSTTR_i	The weighted average TTR value between station i and all other stations in the network
LSTTR_i	The weighted average TTR value between station i and all other stations in the same line

contribution rate of the main component Z_k , and sequentially extract Z_1, Z_2, \dots, Z_r . Moreover, the cumulative contribution rate of these principal components reaches the specified threshold, which is 70% generally.

Therefore, previous approaches focus on index calculation and ranking, screens the top/bottom ranking sections/stations according to the operation indicators. Compared with the traditional measurement methods about TTR, the model proposed in this article has the following differences: (1) the evaluation index (STTR) in this model is directly calculated by AFC data and does not need path restoration; (2) because of the metro network complexity and OD quantity diversity, this model calculates TTR values from station dimension and divides them into two levels of network and line; (3) by integrating analysis with other factors, the model analyzes the correlation between STTR values with these factors and provides data support for cluster analysis method and train operation plan optimization in the next charters.

4.3. Cluster-Based Analysis Method. In this section, we will identify the stations with low reliability and provide some suggestions for improving the STTR level. Clustering analysis is commonly used to categorize large amounts of data. Considering different clusters tend to show distinct differences in the clustering analysis results, and the abnormal points can help distinguish the potential outliers in the data. In this article, by analyzing the different parts determined by cluster analysis, we can identify low-reliability metro stations and propose optimization suggestions in the train operation plan.

We use the SOM neural network to categorize and analyze stations. Based on the values of principal

components, the stations with higher similarity are in the same group, and the attributes of these stations are considered to be the same. As an unsupervised learning neural network, SOM [25] has strong self-organization characteristics and only an input layer-competitive layer (Figure 2).

Compared with the K-means clustering algorithm, the advantages of the SOM neural network include the following:

- (i) Not affected by the initialization of the cluster centroid
- (ii) Improving the processing ability of nonlinear data
- (iii) Reducing the influence of noise data

The SOM neural network cluster algorithm includes the following parts.

4.3.1. Determine the Number of Clusters. We use Silhouette Coefficient (SC) method to determine the number of cluster groups, that is, the number of station categories. The SC method combines the clustering degree of Cohesion and Separation. The Cohesion refers to the average distance between the sample point i and all other elements in the same cluster, denoted as $a(i)$. The Separation means the average distance between the sample point i and the points in the other cluster, traversing other clusters to obtain the minimum value, denoted as $b(i)$; the cluster is the neighbor cluster of i . The sample point i contour coefficient is

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

The larger the average of all stations' contour coefficient, the better the number of clusters.

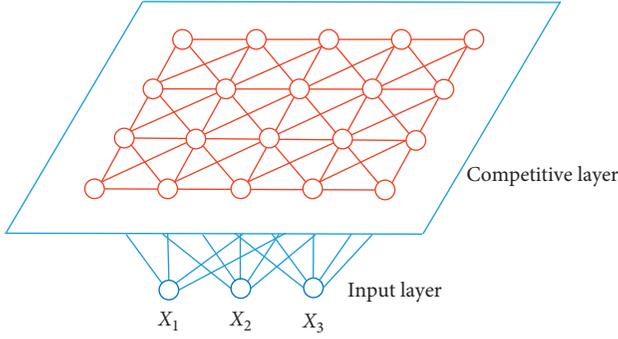


FIGURE 2: The structure of SOM neural network.

4.3.2. *SOM Network Initialization.* Import the principal component data of each station into the input layer of the SOM neural network. The data format is

$$U_k = [u_1^k, u_2^k, \dots, u_N^k]^T, \quad (7)$$

where $k = 1, 2, \dots, r$, r is the number of principal components, and N is the number of stations.

We construct the initial neuron network of the competition layer, and the weight vector expression of the neuron node j and the input layer data is

$$W_j = [w_{j,1}, w_{j,2}, \dots, w_{j,s}]^T, \quad (8)$$

where $j = 1, 2, \dots, M$, M is the number of neuron nodes in the competition layer.

4.3.3. *Competitive Learning in SOM Network.* SOM neural network adopts the method of competitive learning during the training process. Each input data point finds a node that matches it best in the competitive layer, called its activation neuron (WN). Then, use the stochastic gradient descent method to update the parameters of the active node and the data points it covers. The competitive learning process includes the following steps:

Step 1: The initialization parameters of the competition layer nodes have the same parameter dimensions as the input layer data dimensions.

Step 2: According to the Euclidean distance, match point i of the input layer to the nearest node(WN) in the competitive layer:

$$\|u_i - \text{WN}\| = \min \left\{ \|u_i - w_j\| \right\}, \quad j = 1, 2, \dots, M. \quad (9)$$

Step 3: Set WN as the center, the connection weights between other neurons in the neighborhood of the competition layer and the input layer neurons are modified:

$$w_j(t+1) = w_j(t) + h_{cj}(t)[X_j(t) - w_j(t)], \quad (10)$$

$$j = 1, 2, \dots, M, \quad j \neq c,$$

where t is the number of iterations, $w_j(t)$ is the connection weight of the node j and the input layer at the

moment t , $X_j(t)$ is the input sample vector of the node j at the moment t , and $h_{cj}(t)$ is the neighborhood kernel function of WN at the moment t , namely:

$$h_{cj}(t) = \exp\left(-\frac{d_{cj}^2(t)}{2 * \delta^2(t)}\right), \quad (11)$$

where $d_{cj}^2(t)$ is the lateral distance between neuron j and WN and $\delta(t)$ is the amount of network width at the moment t , that is:

$$\delta(t) = \delta(0) * \exp\left(-\frac{n * \log \delta(0)}{1000}\right), \quad (12)$$

where $\delta(0)$ is set as the radius of the initial grid.

Step 4: Update the node parameters until the feature map gradually converges. The neurons in the SOM competition layer continually iterate and cluster simultaneously to divide the stations covered by each neuron into groups.

4.4. *Train Operation Plan Optimization.* After these steps, we obtain several station clusters. Combining the values of NSTTR and LSTTR, we can analyze the characteristics of all clusters and identify the stations with low reliability. By combining passenger flow analysis, train running interval, and station location, we could put forward several suggestions for train operation plans from these three aspects.

Take Line X as an example; as shown in Figure 3 there is only a long routing with the train running interval being 4 minutes in the train operation plan of Line X.

And the measures we could apply include the following:

- (1) Minify train running interval. It is the most convenient method to improve the STTR level of all stations by increasing the number of trains per hour, as shown in Figure 4.
- (2) Adopt the long-short routing operation mode. As shown in Figure 5, if the stations with low-reliability concentrate in a certain section (Station C to Station B), a short routing can be introduced.
- (3) Fare incentives or congestion alerts.

When the train operation plan's transport capacity is close to saturation, we can adopt some other measures to encourage passengers to choose other routes to the destination station. In metro systems, fare incentives are emerging as a method to manage peak-hour congestion, including two strategies: a time-based fare incentive strategy (TBFIS) and a route-based fare incentive strategy (RBFIS). With the development of science and technology, passengers can be reminded of the congestion in some stations and sections in real-time through mobile apps or large screens in the stations to switch paths in time.

5. Case Study on Beijing Metro

5.1. *The Network and Existing Analysis Method.* In this section, the quality of methodology will be illustrated using a

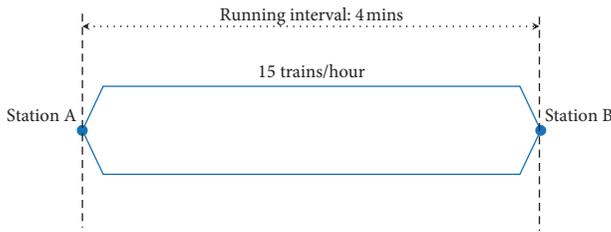


FIGURE 3: Original train operation plan of Line X.

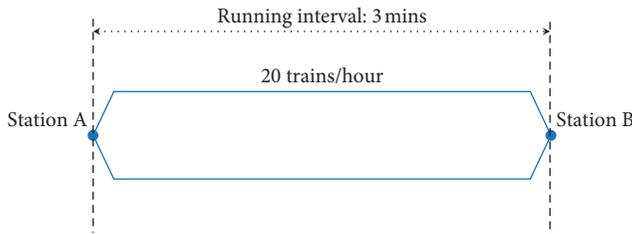


FIGURE 4: Minify train running interval.

real case of the Beijing metro system. In 2016, its network consisted of 18 lines and 326 stations (Figure 6), and there were more than 6,000,000 daily trips on average.

Over the past few years, the Beijing metro system has developed rapidly and is now one of the world's largest. According to the aggregation and dissipation of passenger flow, we divide the area inside the red frame into urban districts. Divide the area outside the red frame into suburban districts. In this article, we use a total of 39,453,138 AFC records on weekdays, calculate all the results by C#. Net and PL/SQL database programming. According to tap-in time in AFC data, we divide the study period into three parts: (1) morning peak periods: 6:00–10:00; (2) off-peak periods: 10:00–16:00; (3) evening peak periods: 16:00–20:00.

The train operation plan analysis method currently employed by the Beijing metro system is the index ranking method. This method screens the top/bottom ranking sections/stations according to the operation indicators, including section full-load rate and station passenger volume. The existing method is essentially to sort operation indicators and get the concerned sections or stations.

5.2. General Analysis. First, we calculate the values of NSTTR and LSTTR of each station at different periods based on the STTR measurement model. The values of NSTTR and LSTTR reflect the STTR level. Figure 7 shows the visualization of NSTTR and LSTTR values of all stations at different periods; the darker the color, the bigger the NSTTR/LSTTR value, that is, the lower the STTR level.

The values of NSTTR and LSTTR present a significant difference in the space-time dimension. In the morning peak, the urban stations' NSTTR and LSTTR values are small, while the suburban stations' values are larger, but there is the opposite in the evening peak. Furthermore, there is a relative balance in the off-peak periods.

Then, we select six lines with huge passenger demand, the NSTTR and LSTTR values are shown in Figure 8, and the

stations at the red frame in the horizontal axis are urban stations. The results show that the LSTTR value of most stations is slightly lower than the NSTTR value, and their changing trend is nearly consistent in one line. In the morning peak, the values of NSTTR and LSTTR in urban stations are smaller; that is, the STTR level of urban stations is higher than suburban stations generally.

5.3. Clustering Analysis. Based on the general analysis, we choose the morning peak as the study period for clustering analysis. First, we obtain each station's passenger demand from the AFC data and calculate the running interval of each line (Table 4) from the train diagram data. According to the Beijing metro network's geographical location analysis, we regard Tiananmen West Station as the central station of the network, as the red pentagram in Figure 6. Then, we calculate the location coefficient of all stations.

5.3.1. Clustering Elements Processing. We use SPSS statistical software for the PCA on cluster elements, and the correlation coefficient matrix of cluster elements is shown in Table 5. There are positive correlations between the NSTTR, LSTTR and passenger flow, train running intervals, and station location coefficient.

The principal components with the top two rankings are extracted, and the cumulative variance is 76.8%. As shown in the principal component matrix (Table 6), the principal component PC1 represents mainly passenger flow, train running intervals, and location coefficient, while the principal component PC2 represents mainly NSTTR and LSTTR.

5.3.2. Clustering Results. First, we use the Contour Coefficient Method [26] to determine the optimal clustering number as 4. Then, we use MATLAB software for cluster analysis. The initial network of the SOM neural network is a 6×6 neuron network. During the competitive learning process, each neuron updates its position and stations connected to it. Then, the categories of stations in the network are classified. As shown in Figure 9, the final positions of all neurons are in the red network, and the black points are the station points. The number of stations covered by each neuron is shown in Figure 10.

Finally, the station clustering results and the distribution in the network are shown in Figures 11 and 12, where the same color points are the stations of the same group, and the size of the station shape in Figure 11 is proportional to the passenger volume.

Based on the above analysis, we analyze each station's group characteristics.

Cluster 1. In Cluster 1, the values of PC1 and PC2 are low; that is, the levels of STTR and influence factors are high. These stations are distributed mainly in the urban districts, their passenger demand stress is weak, and transport capacity supply is high. These stations do not need to improve the train operation plan.

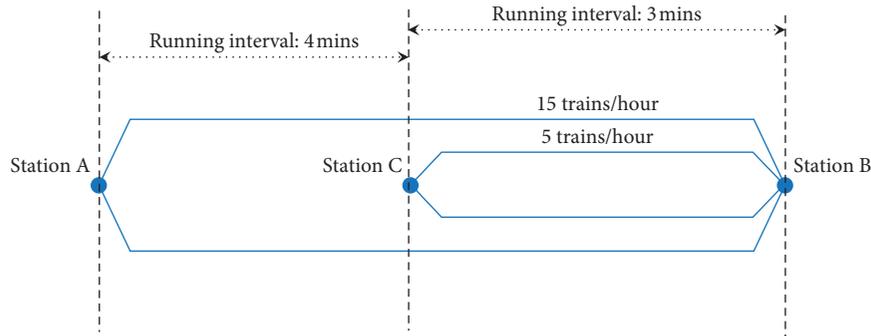


FIGURE 5: Long-short routing operation mode.

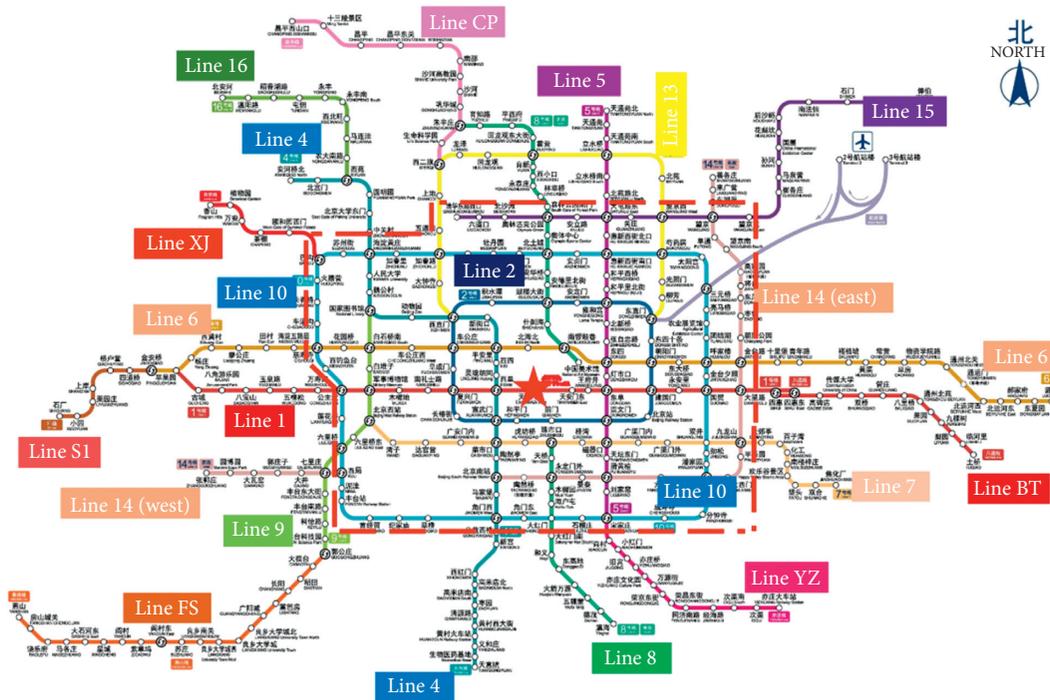


FIGURE 6: Beijing metro network.

Cluster 2. In Cluster 2, the values of PC1 are low, while PC2 is high; that is, the STTR levels are high, but influence factors are weak. There is a high matching of transport capacity and passenger demand, but a weak level of STTR in these stations. The representative lines and stations are Line YZ and Line FS (East). Regard these stations as potential stations that need attention.

Cluster 3. In Cluster 3, the values of PC1 are high, while PC2 is low; that is, the STTR levels are weak, but influence factors are high. The representative lines and stations are Line4 (South), Line9 (South), and Line BT.

Cluster 4. In Cluster 4, the values of PC1 and PC2 are high; that is, the levels of STTR and influence factors are weak. The representative lines and stations are Line CP(North), Line8(North), Line5(North), Line15 (Northeast), and Line14 (West).

5.4. Train Operation Plan Analysis. Considering the distribution of each cluster station in the metro network, in this charter, we focus on the lines and stations in Cluster 3 and Cluster 4 and divide the potential causes of weak STTR into the following three aspects:

- (1) Passenger flow stress is strong.
There is intense passenger flow stress in Line5 (North), Line8 (North), Line6 (East), and Line BT. These lines and stations may require minifngyi train running intervals. Furthermore, introduce the additional custom buses to divert commuter passenger flow.
- (2) Train running interval is large.
There is a weak STTR level in Line4(North) because the train running intervals in these stations are 240 seconds, while those of other stations of Line4 are 120 seconds; thus, Line4(North) requires

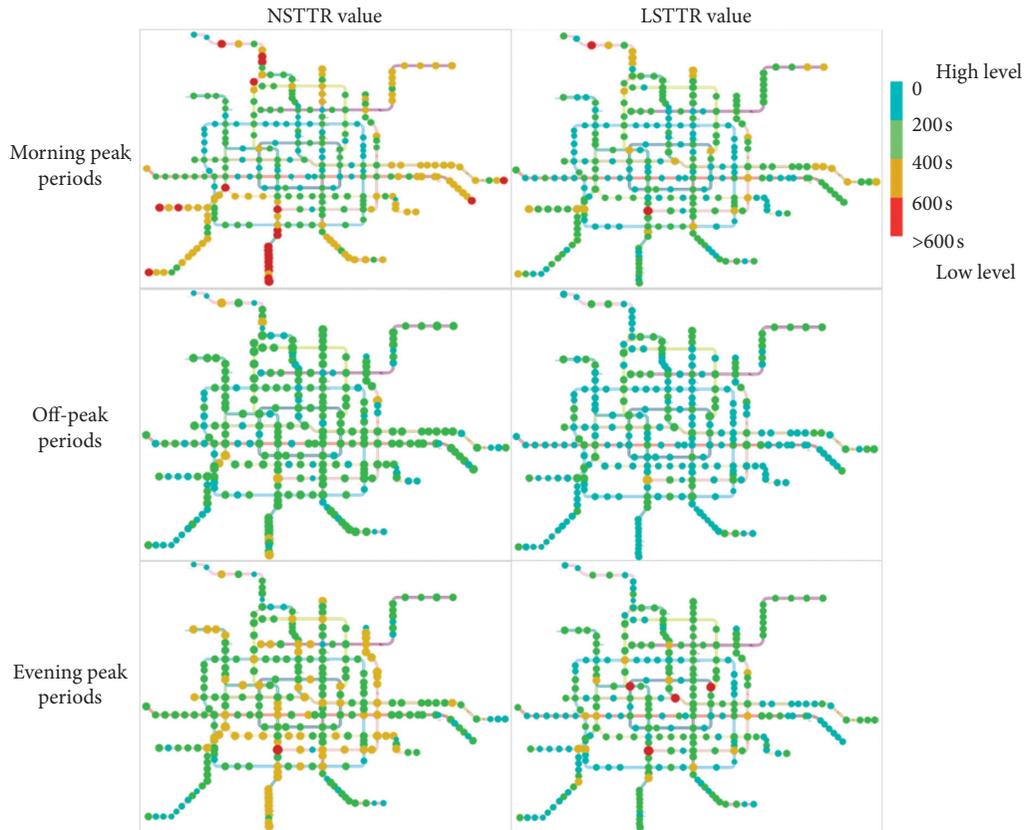


FIGURE 7: The visualization of NSTTR and LSTTR.

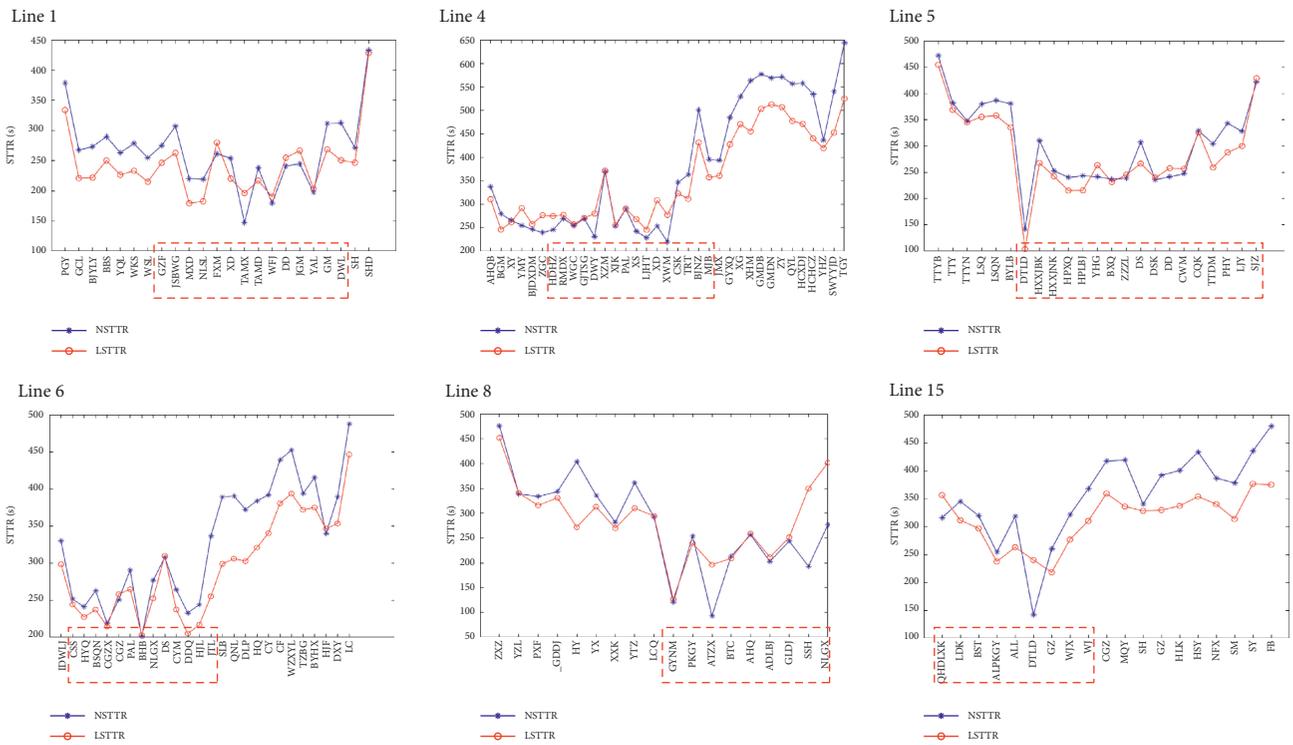


FIGURE 8: The values of NSTTR and LSTTR of six lines.

TABLE 4: Train running interval of lines in the morning peak.

Line name	Running interval(s)	Line name	Running interval(s)	Line name	Running interval(s)
Line1	120	Line6	257	Line14-East	327
Line2	129	Line7	200	Line15	327
Line3	129	Line8	212	Line CP	360
Line4-main	120	Line9	133	Line FS	360
Line4-DX	240	Line13	180	Line YZ	360
Line5	180	Line14-west	514	Line BT	180

TABLE 5: Correlation coefficient matrix.

	Passenger flow	Train running intervals	Location coefficient	NSTTR	LSTTR
Passenger flow	1	-0.21	-0.03	0.16	0.23
Train running intervals	-0.21	1	0.48	0.18	0.22
Location coefficient	-0.03	0.48	1	0.23	0.19
NSTTR	0.16	0.18	0.23	1	0.63
LSTTR	0.23	0.22	0.19	0.63	1

TABLE 6: Principal component matrix of clustering elements processing.

	Passenger flow	Train running interval	Location coefficient	NSTTR	LSTTR
PC1	0.6221	0.53	0.4002	0.2749	0.3104
PC2	0.1621	0.2746	0.216	0.62	0.6835

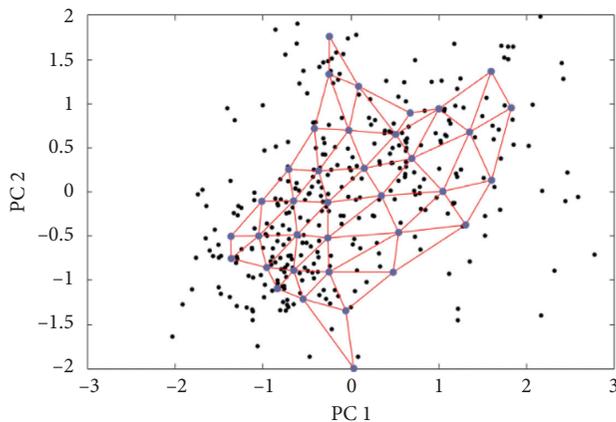


FIGURE 9: SOM neurons final locations.

minifying train running intervals. Similar lines and stations need to minify train running intervals including Line CP, Line14 (West), and Line15, and it is also appropriate to adopt the long and short routing operation mode for them.

(3) Station location problems

As shown in Figure 13, Line FS is located in the southwest of Beijing and has only a transfer station (GGZ), which connects with Line9. GGZ is a terminal station of the two lines, which means the passengers of Line FS who want to go to urban districts must pass through Line9.

According to the AFC data, in the morning peak, the proportion of transfer passengers in the Line FS's total

passenger flow is 82.7%. Thus, the STTR level of Line9 is weak, while the passenger flow stress is weak, and the transport capacity supply is adequate in Line9.

Minify the running intervals of Line9 (South) appropriately to increase the transport capacity and reduce the impact from Line FS to Line 9. At the same time, Line9 is affected by the transfer passenger flow in the west section of Line 14 at QLZ and intensify the passenger flow pressure of Line9. Specific strategies, including fare incentives or congestion alerts, can be used to encourage more passengers from Line14 (West) to choose to transfer to Line 10 at Station XJ instead of Line9 at Station QLZ so that it can reduce the transportation pressure of Line9.

5.5. Comparison and Analysis. Take the existing analysis method using station passenger volume to compare it with the proposed approach in this article. Figure 14 shows the top 20 stations with inbound passenger volume. The X-axis represents the station name; Y-axis represents the inbound passenger volume; the color represents the station's cluster in Figure 11. It can be seen that most of these stations belong to Cluster 4 (red) and Cluster 2 (green).

In the proposed approach, the TTR and other influencing factors are integrated and analyzed by SOM neural network; these stations' passenger service level reflected by STTR values varies greatly and is divided into different clusters. Moreover, according to its influencing factors, analyze the optimization measures that need to be taken. In the existing method, all of the top 20 stations are considered in the train operation plan, so the comprehensive rating is insufficient and cannot explain why the station service level is low.

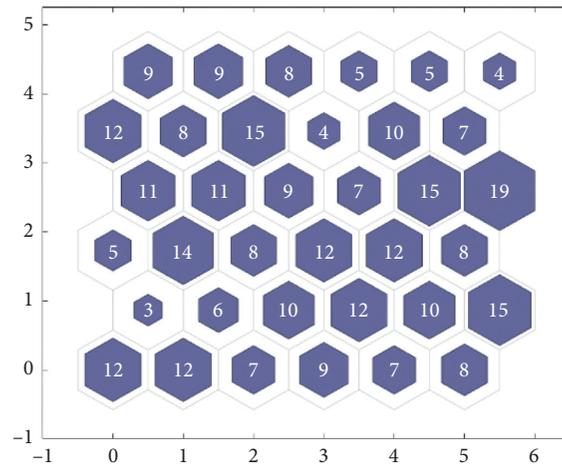


FIGURE 10: The number of stations covered by SOM neurons.

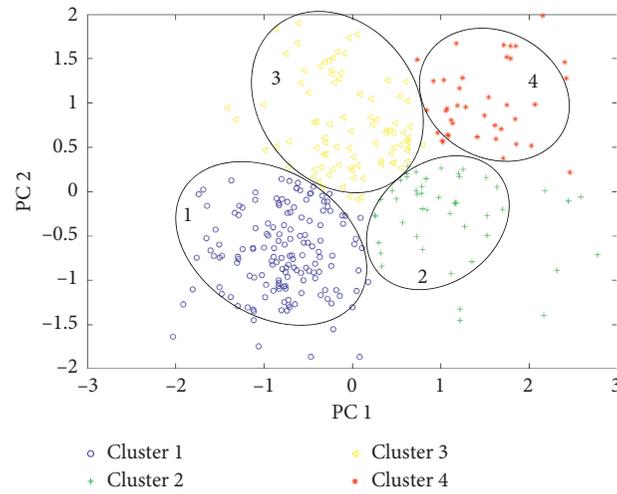


FIGURE 11: Stations clustering renderings.

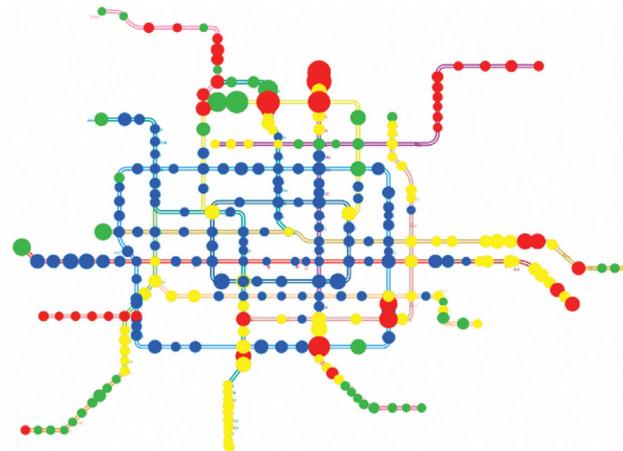


FIGURE 12: Distribution of various cluster stations.

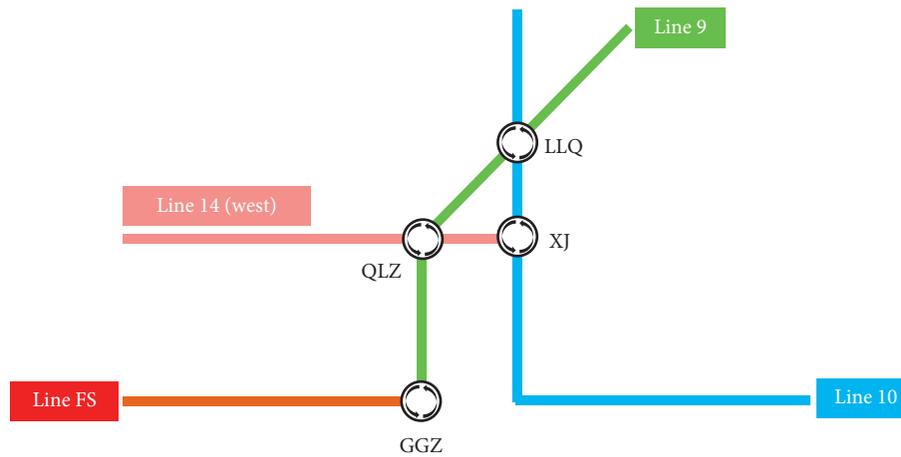


FIGURE 13: The lines connected with Line9.

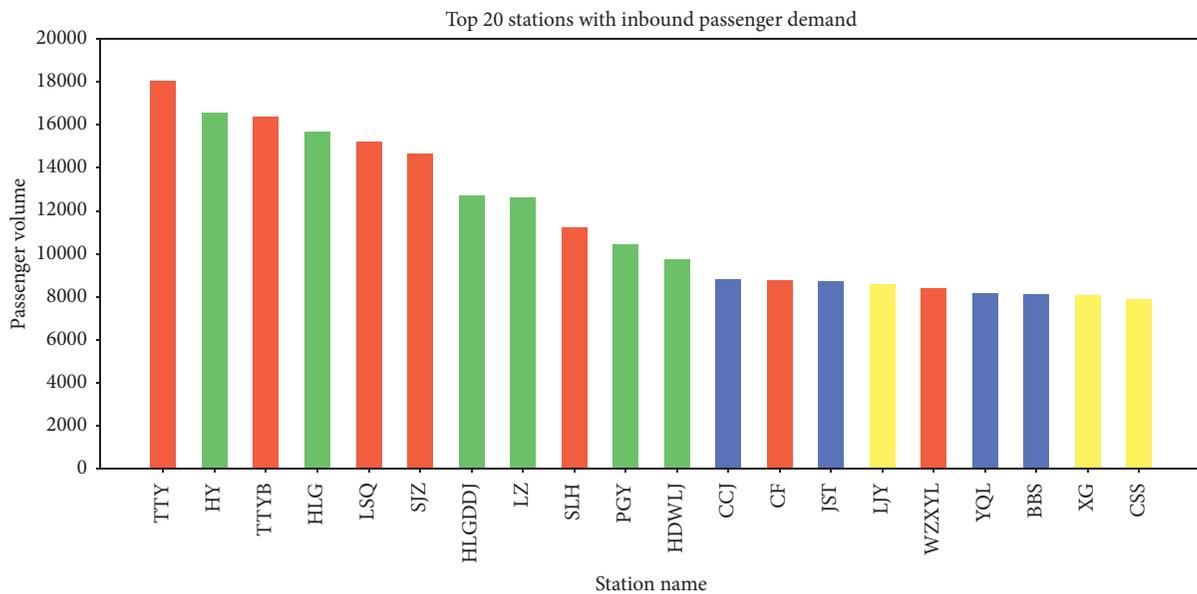


FIGURE 14: Top 20 stations with inbound passenger demand.

6. Conclusion

This article contributes a method for analyzing the train operation plan based on the STTR in the metro system:

- (1) STTR, which is the fluctuation degree between the actual time and standard travel time of each OD from this station as the starting station to other stations, is calculated by AFC data.
- (2) The clustering algorithm based on SOM neural network is efficient in classifying stations and identifying the potential causes of weak STTR level. SOM is an unsupervised learning neural network with strong self-organization and visualization characteristics.
- (3) Taking the Beijing metro network as an example, the framework is applied, and the results are given and discussed in detail. Besides, several suggestions are put forward to optimize the train operation plan.

The application case of the Beijing metro network shows that the proposed method can be used to analyze the train operation plan effectively. It is also applicable to other metro networks with AFC systems. A possible future research direction is to expand the methodology framework to the reliability of transfer time in the time-space dimension. More efforts are also necessary to adopt diversified measures for optimizing operation management, such as asymmetric operation plans.

Data Availability

The AFC data used to support the findings of this study were supplied by Beijing Metro Co., Ltd., under license and so cannot be made freely available. Requests for access to these data should be sent to Mr. Wang, 15221628818@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R & D Program of China (2018YFB1201402). The project was funded by the Ministry of Science and Technology. The authors are grateful for this support.

References

- [1] S. Li, R. Xu, and Z. Jiang, "Evaluation method for matching degree between train diagram capacity and passenger demand for urban rail transit," *China Railway Science*, vol. 38, no. 3, pp. 137–144, 2017.
- [2] S. Li, R. Xu, and K. Han, "Demand-oriented train services optimization for a congested urban rail line: integrating short turning and heterogeneous headways," *Transportmetrica A Transport Science*, vol. 15, no. 2, pp. 1459–1486, 2019.
- [3] W. Wang, L. Cheng, T. Chen, and S. Ni, "A research on adaptability evaluation of train operation plan and passenger flow," *Railway Transport and Economy*, vol. 41, no. 7, pp. 65–71, 2019.
- [4] F. Lu, *Theory of Transport Efficiency of Urban Rail Transit Network*, Beijing Jiaotong University, Beijing, China, 2016.
- [5] S. Tian, *Study on Optimization of Train Routing Planning of Urban Rail Transit Considering Congestion Degree of Transfer Station*, Beijing Jiaotong University, Beijing, China, 2019.
- [6] Y. Liu and D. Chen, "An optimization of long and short routing train plan of urban rail transit," *Urban Rail Transit*, vol. 41, no. 02, pp. 117–122, 2019.
- [7] Y. Shafahi and A. Khani, "A practical model for transfer optimization in a transit network: model formulations and solutions," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 6, pp. 377–389, 2010.
- [8] L. Bos, D. Ettema, and E. Molin, "Modeling effect of travel time uncertainty and traffic information on use of park-and-ride facilities," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1898, no. 1, pp. 37–44, 2004.
- [9] A. J. Pel, N. H. Bel, and M. Pieters, "Including passengers' response to crowding in the Dutch national train passenger assignment model," *Transportation Research Part A: Policy and Practice*, vol. 66, pp. 111–126, 2014.
- [10] Y. Asakura and M. Kashiwadani, "Road network reliability caused by daily fluctuation of traffic flow," in *Proceedings of the 19th PTRC Summer Annual Meeting, Proceeding Seminar G*, pp. 73–84, Brighton, England, 1991.
- [11] Y. Asakura, "Reliability measure of an origin and destination pair in a deteriorated road network with variable flow," in *Proceeding of 4th Meeting of the EURO Working Group in Transportation*, pp. 75–77, Mayaguez, Puerto Rico, April 1996.
- [12] Y. Asakura, M. Kashiwadani, and K. I. Fujiwara, "Functional hierarchy of a road network and its relations to time reliability," *Proceedings of JSCE*, vol. 583, no. 583, pp. 51–60, 2010.
- [13] W. H. K. Lam and G. Xu, "A traffic flow simulator for network reliability assessment," *Journal of Advanced Transportation*, vol. 33, no. 2, pp. 159–182, 1999.
- [14] E. Bell and C. Chirs, *Reliability of Transport Networks*, Research Studies Press Ltd., Biddleswade, England, 2000.
- [15] S. Dai, C. Zhu, and Y. Chen, "Research on time reliability of urban public transit," *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, vol. 5, pp. 869–871, 2008.
- [16] T. Lomax, D. Schrank, S. Turner et al., *Selecting Travel Reliability Measure*, Texas Transportation Institute, College Station, TX, USA, 2003.
- [17] W. Zhang, P. Zhao, and X. Yao, "Reliability evaluation of Beijing metro rate travel time," *Shandong Science*, vol. 26, no. 6, pp. 77–81, 2013.
- [18] W. Li, J. Zhou, F. Zhou, and R. Xu, "Calculation of travel time reliability of park-and-ride based on structural reliability algorithm," *Journal of Southeast University(Natural Science Edition)*, vol. 46, no. 1, pp. 226–230, 2016.
- [19] J. Chen, *Research on Operation Reliability of Urban Rail Transit Network*, Tongji University, Shanghai, China, 2007.
- [20] P. V. S. Rao, P. K. Sikdar, K. V. K. Rao, and S. L. Dhingra, "Another insight into artificial neural networks through behavioural analysis of access mode choice," *Computers Environment & Urban Systems*, vol. 22, no. 5, pp. 485–496, 1998.
- [21] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, *Neural Network Based Temporal Feature Models for Short-Term Railway Passenger Demand Forecasting*, Pergamon Press, Inc., Oxford, England, 2009.
- [22] W. Zhu, W. L. Fan, A. M. Wahaballa, and J. Wei, "Calibrating travel time thresholds with cluster analysis and Afc data for passenger reasonable route generation on an urban rail transit network," *Transportation*, vol. 47, 2020.
- [23] M. Wachs and T. G. Kumagai, "Physical accessibility as a social indicator," *Socio-Economic Planning Sciences*, vol. 7, no. 5, pp. 437–456, 1973.
- [24] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, 1901.
- [25] T. Kohonen, "Self-organizing maps, in Springer Series in Information Sciences," vol. 30, Springer-Verlag, Berlin, Germany, 3rd edition, 2001.
- [26] R. J. Peter, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational & Applied Mathematics*, vol. 20, 1999.