Research on Urban Rail Transit Operation Periods Division Based on Passenger Flow Characteristics Mining

Changfeng Zhu, Liangsheng Zhao, Linna Cheng, Chao Zhang, and Shengyun Xue

***Abstract*—To fully explore the changing patterns of Urban Rail Transit (URT) passenger flow characteristics over different operation periods and to accurately formulate and adjust train operation services to match capacity with demand, this paper introduces a novel approach. An ordered clustering algorithm based on Simulated Annealing (SA) is introduced as an improvement over the traditional optimal segmentation method for dividing URT operation periods. Taking Nanjing Metro Line 3 as a case study, the paper selects multiple passenger flow characteristics related to operation periods, such as passenger flow at different periods of the day, cross-section passenger flow at various periods, and the average value of passenger flow change coefficient of adjacent cross-sections in a single period, etc., conduct an empirical analysis. Indices such as the silhouette coefficient and C-H coefficients are introduced to evaluate the effectiveness of the new algorithm in dividing operation periods. The results indicate that the optimal segmentation method, which considers the characteristics of multiple passenger flows, along with the ordered clustering algorithm based on SA, can better reflect the changing patterns of passenger flow compared to methods that consider only single passenger flow characteristics. These findings can serve as a basis for preparing train operation plans. In addition to the running time, the ordered clustering algorithm based on SA outperforms the optimal segmentation method in terms of loss function, silhouette coefficient, and C-H coefficient. This demonstrates that the ordered clustering algorithm based on SA is more effective at dividing the URT operation periods. This method provides a more practical solution for dividing URT operation periods.**

***Index Terms*—Period Division, Urban Rail Transit (URT), Simulated Annealing (SA), Passenger Flow Characteristics**

Manuscript received May 23, 2024; revised November 4, 2024.

This work was largely supported by the National Natural Science Foundation of China (No.72161024) and “Double-First Class” Major Research Programs, Educational Department of Gansu Province (No.GSSYLXM-04).

Changfeng Zhu is a professor and PhD supervisor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: cfzhu003@163.com).

Liangsheng Zhao is a postgraduate student in the Transportation Planning and Management Program, School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, phone: +86 17393180538, e-mail: zhaoliangsheng007@163.com).

Linna Cheng is a doctoral candidate in the Transportation Planning and Management Program, School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: chengjj @163.com).

Chao Zhang is a doctoral candidate in the Transportation Planning and Management Program, School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 77052635377@qq.com).

Shengyun Xue is a postgraduate student majoring in transportation, School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 215505901@qq.com).

I. INTRODUCTION

WITH the persistent grow of urban scale and the rapid aggregation of the population in large cities, urban road traffic congestion has been exacerbated, making travel increasingly difficult for residents [1][2][3]. Due to its advantages of speed, safety, punctuality, and environmental protection, URT has emerged as a crucial mode of transportation that helps to tackle the daily commuting difficulties experienced by inhabitants in metropolitan areas [4][5][6].

URT operators do not provide precise train schedules for passengers to consult and plan their travel times. To accommodate the travel demands of passengers during different periods, a common practice among URT operators is to divide the day into peak and off-peak periods, adopting varying departure intervals for each. Due to the complexity of transportation organization, it is challenging for URT operators to make quick changes to their operation plans, which often remain the same for several hours during both peak and off-peak times. However, the travel activities of urban residents exhibit a clear clustering effect in the time dimension [7], leading to frequent overcrowding on trains. This makes it difficult for URT operators to consistently provide satisfactory services to passengers [8].

Given that passenger flow exhibits regularity and similarity over relatively short periods, it’s feasible to identify characteristic patterns of URT passenger flow. By doing so, we can finely divide the entire day’s operation time into several distinct periods, thereby understanding the travel habits of passengers [9][10][11]. URT operators can then develop train operation plans and detailed periods-based schedules according to the observed fluctuations in passenger flow. This approach aids in aligning passenger demand with capacity supply, ultimately providing more efficient travel services for passengers.

The concept of dividing traffic operation periods has its roots in the application of signal timing at road traffic intersections. Initially, the use of Data Mining Methods for periods division was proposed in [12], where a clustering algorithm was employed to automatically segment signal control periods. Subsequently, [13] introduced the Genetic Algorithm, utilizing optimization principles to determine the period for each time point. However, this approach risked creating isolated periods, prompting the author to refine the model in [14]. Furthermore, [15] proposed a clustering analysis method to identify data with similar passenger flow characteristics.

Subsequently, the theory of traffic operation periods division has been applied to the division of bus operation periods. [16] counted the passenger flow, road congestion and other indicators during the traffic operation period, and divided the whole operation cycle into different periods based on fuzzy clustering algorithm. [17] divided the entire daily operation time of public transport into different periods using a hierarchical classification algorithm. [18] used an ordered clustering algorithm to division different characteristic periods and optimized the vehicle scheduling problem. [19][20][21] Using the traffic flow of different lanes in a roadway intersection as a characteristic variable, the operation periods were divided by the clustering algorithm. It can be found that the above studies have merged the adjacent periods with similar passenger flows to form a traffic period. These research ideas have laid the foundation for the division of rail transit operation periods.

To the best of our knowledge, Relative to other forms of transportation, the literature on the segmentation of operational periods for rail transit is notably sparse. A common practice among technical staff at most operating enterprises is to divide the operating periods into peak and off-peak times based on practical experience, and then formulate rail transit operation plans [22][23]. In recent years, some scholars have conducted meaningful studies on this topic, such as [24], which used the inbound passenger flow at each time point of the target line station as a descriptive variable and divided the entire daily operation time into four classes and six time periods using the nearest neighbor propagation clustering algorithm. [25] studied the temporal distribution of passenger flow on suburban railway lines. [26] divided the day into different periods according to the passenger flow demand, and operated trains with varying intervals of departure and the number of marshaled vehicles in each period to best match transport capacity and passenger demand in different periods.

The aforementioned literature has studied the signal timing at road traffic intersections and the division of bus and rail transit operation periods using various methods. However, there are still several shortcomings: (1) Most of the division methods can not reflect the timing characteristics of the operation periods, which may lead to the non-adjacent traffic periods being divided into one class, and it is also easy to produce isolated periods. (2) The existing research only uses the passenger flow at different periods of the day as the descriptive variable, and does not consider the impact of multi-passenger flow characteristics on time division, so the change characteristics of the entire line's passenger flow are difficult to be reflected. (3) The time samples selected in the existing literature are too few to reflect the smooth transition between the peak periods and off-peak periods.

Considering the previously identified deficiencies, this paper offers advancements in the following three domains. (1) Based on the traditional optimal segmentation method, this paper proposes a URT operation periods division algorithm based on SA, which has more advantages in accuracy than the traditional method. (2) The existing research only considers the impact of the passenger flow at different periods of the day (single passenger flow characteristics) on the division of the operation periods. This paper selects the multi-passenger flow characteristics such as the passenger flow at different periods of the day, cross-section passenger flow at various periods, and the average value of passenger flow change coefficient of adjacent cross-sections in a single period to divide the operation periods of URT. (3) The effectiveness evaluation indices of the clustering algorithm are introduced to evaluate the performance of the traditional optimal segmentation algorithm and the ordered clustering algorithm based on SA under the conditions of single passenger flow characteristics and multiple passenger flow characteristics.

The rest of this paper is organized according to the following structure: Section II describes the division of URT operation periods. Two methods for dividing the URT operation periods will be introduced in Section III, including the traditional ordered clustering algorithm and the newly proposed operation periods division algorithm based on SA. Section IV examines the passenger flow characteristics of Nanjing Metro Line 3 and presents the results from different algorithms. In Section V, the effectiveness of various methods is verified, and the sensitivity of different parameters is analyzed. The last Section is the conclusion, which summarizes the research results of the case.

II. Problem Description

The passenger flow data of URT has a significant correlation in adjacent periods. Based on this, we can mine the time series data to get the change law of passenger flow characteristics with the operation periods [27][28]. Dividing the operation periods of URT can help operating enterprises adjust and optimize train services in a more planned and systematic manner [29][30].

Fig.1 illustrates the schematic representation of the URT time series data. D weekdays with similar attributes are selected, and the whole day operation time of the URT is divided into N time units on average. The passenger flow data throughout the day for the URT forms a time series, the time series of the day d can be denoted as $T_d = \{t_1^d, t_2^d, \cdots, t_N^d\}$. The passenger flow data of the day D are averaged according to the time unit, which is used to describe the passenger flow characteristics of each time unit during the whole day operation time of URT. Then the whole day time series used for the division of URT operation periods can be expressed as $T = \{t_1, t_2, \cdots, t_N\}$, where

$$t_n = \frac{1}{D}\sum_{d=1}^{D} t_n^d \; . \tag{1}$$

Obviously, using the method described above, we can obtain several sets of time series data, these time series data are in accordance with the order of time, and the data contain the sequence of passenger flow reflects the change law of URT passenger flow on the target line in a day.

According to the similarity of passenger flow data characteristics of each time unit, the time units with the highest similarity are divided into the same class, and each class is an operation period. The different operation periods after clustering can be expressed as $\{t_{i1}, t_{i1+1}, \cdots, t_{i2-1}\}$, $\{t_{i2}, t_{i2+1}, \cdots, t_{i3-1}\}, \cdots, \{t_{iK}, t_{iK+1}, \cdots, t_N\}$, where $1 = i1 < i2 < \cdots < iK \leqslant N$. The consecutive time periods segmented by the above method can be used as a basis for train operation planning.

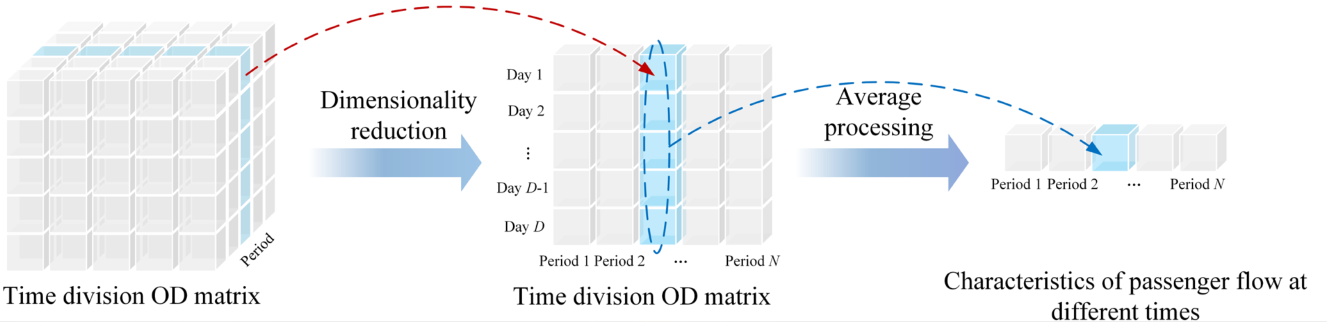

Fig.1. Schematic diagram of URT time series

III. Methodology

In view of the problem of URT operation periods division, this paper first presents the ordered clustering algorithm commonly used by researchers. For the deficiencies of the ordered clustering algorithm in the face of large samples, uses the idea of optimization and proposes an ordered clustering algorithm based on the SA. We take these two approaches to divide the operation periods of URT.

A. Ordered Clustering Algorithm

Two key issues need to be considered in the division of the URT operation periods: first, sample size, which is related to the selection of time unit length. Second, the sequence of period samples requires that the original order cannot be changed after the samples are divided. Fisher ordered sample clustering algorithm, also known as the optimal segmentation method, can take time units as samples without changing the order of samples and take multiple groups of passenger flow characteristic data as sample attributes. According to the similarity of passenger flow characteristics within the class, the time units with substantial similarity are classified into one class, and the total deviation and minimum of each class are used as the judgment criteria for optimal division to achieve the automatic division of ordered samples. The steps of the algorithm consist of the following four parts.

(1) Define class diameter

Suppose the time series sample of an operation period G (class G) is $\{t_s, t_{s+1}, \cdots, t_l\}$, where t_s is the row vector $\{t_s, t_{s+1}, \cdots, t_l\}$, and the sum of squares of intra-group deviations is defined as the diameter of the period G, which can be calculated by (2).

$$D(s,l)=\sum_{r=s}^{l}(x_r-\bar{x}_j)^{\mathrm{T}}(x_r-\bar{x}_j) \tag{2}$$

Where, $D(s,l)$ is the class diameter of G period. x_r is the attribute value of the sample in the period G. $\bar{x}_j$ is the mean value of the attribute j in period G.

(2) Define classification loss function

$b(N,K)$ is a way of division that divides the time series T into K periods (classes), and the serial number is used to represent the elements in the class. The divided operation periods can be expressed as $\{i_1, i_1+1, \cdots, i_2-1\}$, $\{i_2, i_2+1, \cdots, i_3-1\}, \cdots, \{i_K, i_K+1, \cdots, N\}$, where the division point $1=i_1<i_2<\cdots<i_{K+1}-1=N$.

The loss function is represented by the sum of the squared deviations, denoted as $L[b(N,K)]$, for each class within $b(N,K)$. Among them,

$$L[b(N,K)]=\sum_{k=1}^{K}D(i_k,i_{k+1}-1) \tag{3}$$

Where, k is the $k-$th period, and $k=\{1,2,\cdots,K\}$.

If N and K are known, when $L[b(N,K)]$ is the minimum, $b(N,K)$ is the optimal segmentation method under the given classification number K.

(3) Recursion formula

The principle of the optimal segmentation method is a class of dynamic programming problems based on the fact that the first $K-1$ segments are the optimal segmentation, and the recursive formulas for the loss function are shown in (4) and (5).

$$L[b(N,2)]=\min_{2\leqslant i\leqslant N}\{D(1,i-1)+D(i,N)\} \tag{4}$$

$$L[b(N,K)]=\min_{K\leqslant i\leqslant N}\{L[b(i-1,K-1)]+D(i,N)\} \tag{5}$$

(4) Determine the optimal number of clusters

This paper quantifies the operation process by referring to the "elbow method" to define the inflection point in the loss function trend chart as the optimal cluster number. It converts the problem of finding the inflection point into the problem of solving the slope change rate of adjacent periods, which can be calculated by (6).

$$L(K)=\left|\frac{L_d(K)-L_d(K-1)}{L_d(K)-L_d(K+1)}\right| \tag{6}$$

Where, $L(K)$ is the rate of change of slope in adjacent periods. $L_d(K)$ is the deviation slope of adjacent segmentation times, and the specific calculation method is as in (7).

$$L_d(K)=L[b(N,K+1)]-L[b(N,K)] \tag{7}$$

When $L(K)$ reaches the maximum, the corresponding K is the optimal clustering number of the problem.

The optimal segmentation method is based on the principle of dynamic programming, which categorizes the samples sequentially. However, existing studies have indicated that with large sample sizes, the method is prone to converging on local optimal solutions. This issue can be addressed through the application of optimization ideas.

B. Ordered Clustering Algorithm Based on SA

Based on the principle of optimal segmentation, the URT operation periods division problem can be represented by the optimization problem shown in equation 8.

$$\arg\min f=\sum_{k=1}^{K}D(i_k,i_{k+1}-1) \tag{8}$$

The optimization problem is to find the optimal way to divide an ordered sample with a sample size of n into k classes. The objective is to maximize the similarity of the passenger flow characteristics of the samples in the divided periods. To prevent the occurrence of operation periods that are too short, we introduce a constraint that each operation period must contain at least 2-time units.

To prevent getting stuck in a suboptimal solution, the SA algorithm accepts suboptimal solutions according to a certain probability. At high temperatures, the acceptance probability for both optimal and suboptimal solutions is high, allowing the SA to search globally for the optimal solution. As the temperature decreases, the probability of accepting suboptimal solutions diminishes, while the optimal solutions are accepted with increasing probability, leading the SA to gradually converge on the optimal solution.

Since the SA algorithm can effectively update the solution of discrete variables, it offers a wide search scope in the early stage and gradually narrows it down in the later stage, making it well-suited to the problem at hand. Therefore, this paper designs a corresponding SA algorithm to address the issue, and the ordered clustering algorithm based on SA is constituted by the following five steps.

Step 1: Initialization of SA' parameters

Maximum number of iterations $maxgen$ =500

Initial temperature T =1000

Annealing rate α =0.95

Disturbance times L =10000

Step 2: Generate the initial solution

Randomly generate $K-1$ integers within the range $[1, n-1]$ as initial split points (denoted as x_0), arrange the logarithmic column x_0 in ascending order, and calculate the corresponding objective function value y_0 .

Step 3: Update solution and objective function value

Apply random disturbance to the initial segmentation point x_0 to generate a new solution x_1 , calculate the corresponding objective function value y_1, and $\vartriangle y$ can be calculated by (9).

$$\vartriangle y = y_1 - y_0 \tag{9}$$

(1) When $\vartriangle y \leq 0$, accept the new solution y_1 as the current optimal solution, assign x_1 to x_0 , and assign y_1 to y_0 .

(2) When $\vartriangle y > 0$, judge whether to accept the optimal solution according to probability $e^{-\vartriangle y/\alpha T}$.

Step 4: Repeat the disturbance and acceptance process for L times at temperature T .

Step 5: Check whether the number of iterations of SA meets the maximum number of iterations. If so, terminate the SA. Otherwise, cool down slowly and return to **Step 3**.

Fig.2 illustrates the schematic of the SA.

Although we have conducted noise reduction and dimensionality reduction on passenger flow data, due to the large scale of the data, the search time of the SA algorithm is longer than we expected, and the specific running time will be described in Section V. The selection of the initial solution of SA algorithm is crucial to the convergence time and the quality of the solution. We discovered that an effective strategy to decrease the algorithm's running time involves manually setting an initial solution by dividing a sensible operation period, taking into account passenger flow variations throughout the day.

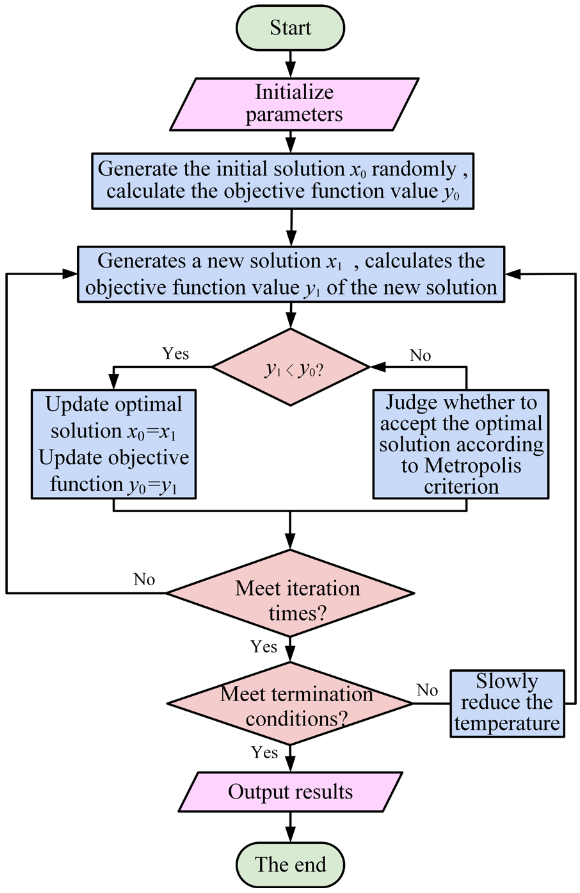

Fig.2. Flowchart of the ordered clustering algorithm based on SA

IV. CASE STUDY

The validity of the URT operation period division method proposed in this paper needs to be verified, we use the AFC data from Nanjing Metro Line 3 for five consecutive working days as the primary dataset for validation. Upon analyzing the existing passenger flow data, it is determined that when the time unit length is set to 15 minutes, the passenger flow characteristics of the line can be fully preserved, ensuring a smooth transition between peak and off-peak periods. The operational hours of Nanjing Metro Line 3 are from 06:00 to 24:00, and the entire day's operation time is divided into 72-time units, each separated by an interval of 15 minutes.

According to the time unit, the passenger flow data is segmented to obtain the OD matrix of passenger flow in different time units. The passenger flow characteristics, such as passenger flow at different periods of the day, cross-section passenger flow at various periods, the average value of passenger flow change coefficient of adjacent cross-sections in a single period are calculated. The URT operation periods are divided by two algorithms, considering the characteristics of a single passenger flow and multiple passenger flows. The specific process is shown in Fig.3.

A. Passenger Flow Characteristics of Nanjing Metro Line 3

Analysis of actual operation data from URT reveals that passenger flow characteristics are influenced by various factors. These factors can be categorized into temporal and spatial factors. Recognizing the deficiency in existing studies, which often fail to comprehensively consider the characteristics of multiple passenger flows, this paper takes into account both the temporal and spatial characteristics of passenger flows when dividing URT operation periods.

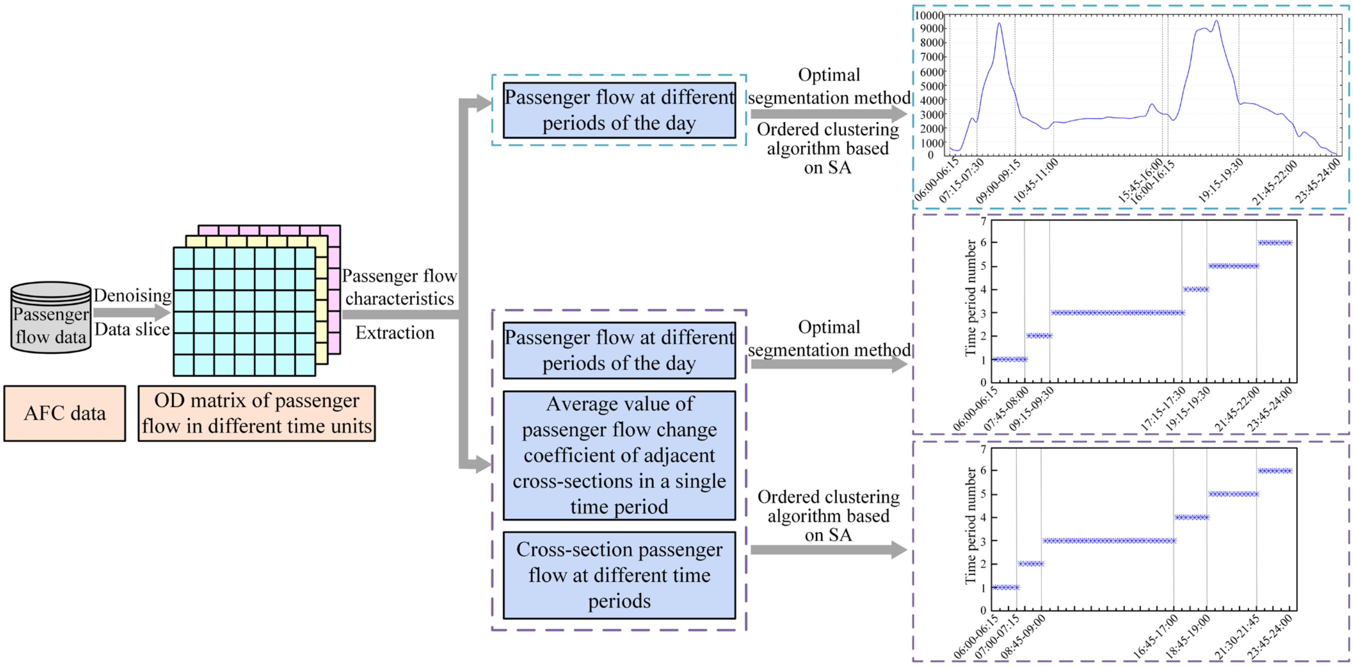

Fig.3. Overall process of case analysis

A.1 Passenger Flow at Different Periods of the Day

For a URT line, passenger flow at different periods of the day Q_i is the sum of the inbound passenger flow of all stations and the passenger flow into the line during the whole day's operation time. Passenger flow at different periods of the day reflects the passenger flow of the whole line over a unit period. It is an essential indicator for the division of URT operation time. Passenger flow of Nanjing Metro Line 3 at different times of the day is shown in Fig.4.

The 15-minute passenger flow of the whole line fluctuates between 156 and 9561, indicating significant variability in passenger flow across different periods. This variability results in operation periods of differing lengths. Specifically, the operation periods with high passenger flow contain fewer time units, while low passenger flow contain more time units.

Typically, the passenger flow throughout the day on weekdays exhibits distinct double-peak characteristics — namely, a morning peak and an evening peak. Staff from URT operation companies can manually segment the operation periods based on these fluctuations. However, this manual division method often overlooks the temporal changes in cross-sectional passenger flow, resulting in a relatively crude division that fails to reflect the spatial characteristics of passenger flow.

A.2 Average Value of Passenger Flow Change Coefficient of Adjacent Cross-sections in a Single Period

The shift in the magnitude of traffic is continuous in the time dimension. Specifically, the changes in passenger flow at the same cross-section between adjacent periods are relatively similar. Additionally, the passenger flow of adjacent cross-sections within the same period also exhibits similarities. To investigate the correlation of cross-sectional passenger flow differences in adjacent periods, the cumulative distribution function (CDF) curves for the differences in cross-sectional passenger flow on Nanjing Metro Line 3 during the intervals of 8:00-8:15, 8:15-8:30, 8:30-8:45, and 8:45-9:00 were plotted. The CDF curve of cross-section passenger flow difference in adjacent periods is shown in Fig.5.

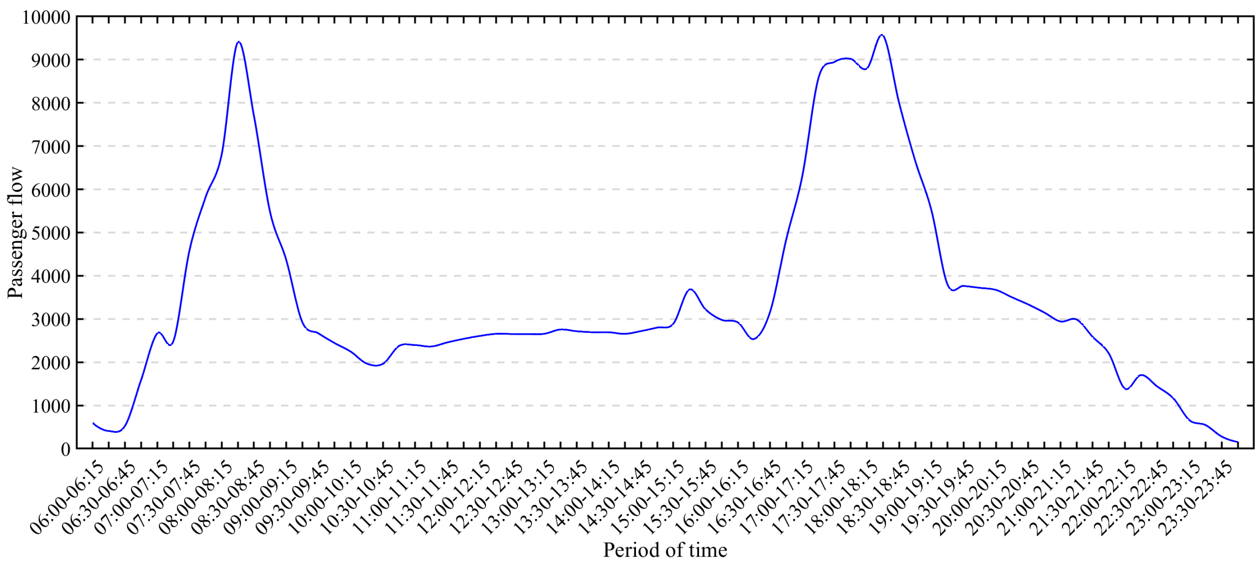

Fig.4. Passenger flow of Nanjing Metro Line 3 at different times of the day

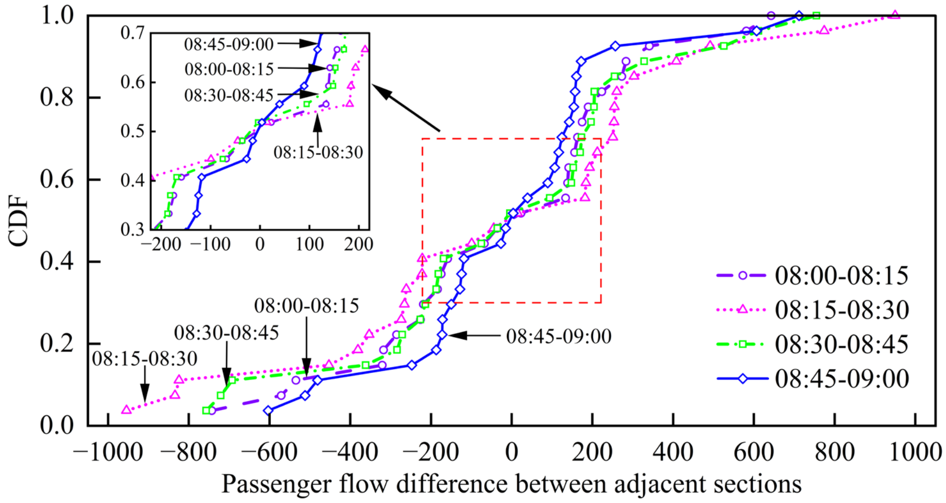

Fig.5. CDF curve between adjacent time periods

As shown in Fig.5, the change trends of the four CDF curves are similar, indicating that the characteristics of cross-sectional passenger flow are correlated within continuous periods. Therefore, the difference in cross-sectional passenger flow between adjacent periods can be utilized to divide URT operation periods.

The passenger flow change coefficient of adjacent cross-sections in a single period is used to describe the passenger flow characteristics of URT in the time dimension, and its mean value reflects the difference in passenger flow between adjacent cross-sections in a specific period. The average passenger flow change coefficient of adjacent cross-sections in a single period can be calculated by (10).

$$\theta_i = \frac{1}{M}\sum_{m=1}^{M}\left|\frac{Q_m^i - Q_{m-1}^i}{Q_{\max}}\right| \tag{10}$$

Where, Q_m^i is the passenger flow of the m cross-section in period i. Q_{m-1}^i is the passenger flow of cross-section m-1 in the period i. $Q_{\max}$ is the maximum cross-section passenger flow in one direction. M is the number of cross-sections.

It has been observed that the average value of passenger flow change coefficient of adjacent cross-sections in a single period has similar characteristics in adjacent periods. To illustrate this feature, the average value of the passenger flow change coefficient of adjacent cross-sections in different periods of the target line is counted with hours as a single period, and the results are shown in Fig.6.

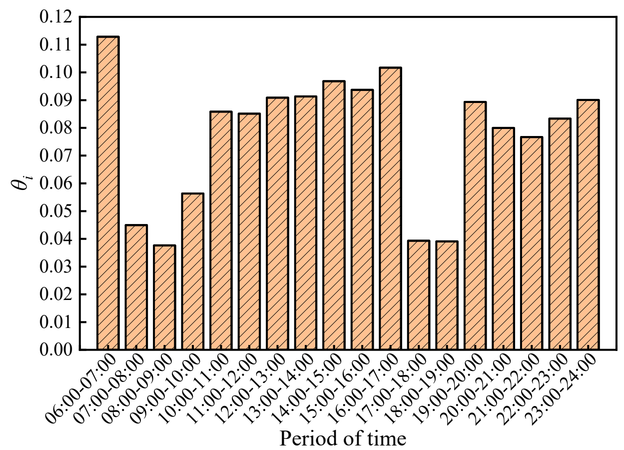

Fig.6. Average value of passenger flow change coefficient of adjacent cross-sections in different periods

The average value of passenger flow change coefficient of adjacent cross-sections at different times of the day and the whole-day passenger flow of URT show opposite trends. The values in the morning and evening peak hours are smaller, indicating that in the morning and evening peak hours of URT, the passenger flow in adjacent cross-sections is closer and the passenger flow in each cross-section is more concentrated. During the off-peak hours, the volatility of the cross-section passenger flow with time is larger.

A.3 Cross-section Passenger Flow at Different Periods

The passenger flow across various time intervals within a cross-section pertains to the traffic passing through each segment of the URT system. Analyzing the section passenger flow across different periods for URT provides insights into the traffic volume at each line segment's cross-section throughout various timeframes. The cross-section passenger flow at different periods of the target line is shown in Fig.7.

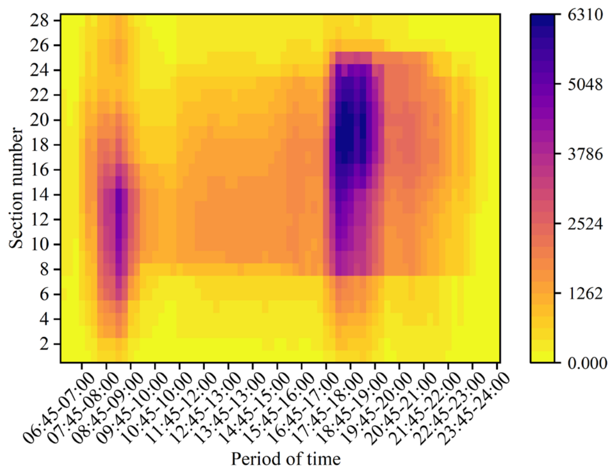

Fig.7. Cross-section passenger flow of the target line

The morning peak hour of the target line on weekdays is about 7:00-9:00, and the evening peak hour is about 17:00-19:00, and the duration of the evening peak hour is significantly longer than that of the morning peak hour, indicating that compared with the morning peak hour, travelers exhibit increased flexibility in selecting their departure times, and more sensitive to the time of the morning peak hour, which is in line with the travel habits of passengers on weekdays. By observing the passenger flow intensity of each cross-section, it can be found that there is a separation of work and residence in Nanjing Metro Line 3.

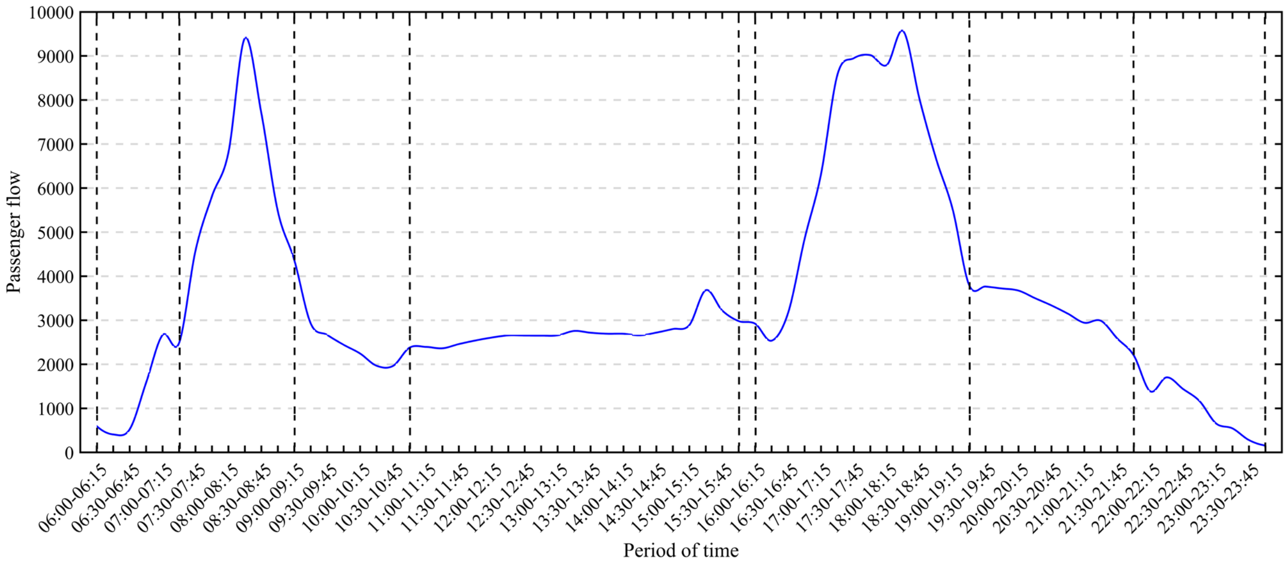

Fig.8. Division results of operation period at different periods of the day

B. Result Analysis

B.1 Result Analysis of Single Passenger Flow Characteristics

When dividing the operation periods of the target line only considering passenger flow at different periods of the day, the calculation results show that the optimal segmentation method and the ordered clustering algorithm based on SA have the same division results. The specific division results are shown in Fig.8. The eight-time periods are 06:00-07:30, 07:30-09:15, 09:15-11:00, 11:00-16:00, 16:00-16:15, 16:15-19:30, 19:30-22:00, and 22:00-24:00.

It can be observed from Fig.8 that both algorithms accurately identify the peak and off-peak periods of passenger flow, and the division results reflect the pattern of passenger flow changes over time in different periods. However, the time intervals after division are too discrete, and the duration of some operation periods is too short (e.g., 16:00-16:15), which is not conducive to formulating a train operation plan. It is evident that when considering only the characteristics of single passenger flow, the results of the operation periods division lack strong guidance for actual operations. Consequently, it is imperative to account for the influence of multiple passenger flow attributes.

B.2 Result Analysis of Multi-Passenger Flow Characteristics

Select multiple passenger flow characteristics. The operation periods are divided based on the optimal segmentation method, and the changing trend of the class loss function is shown in Fig.9.

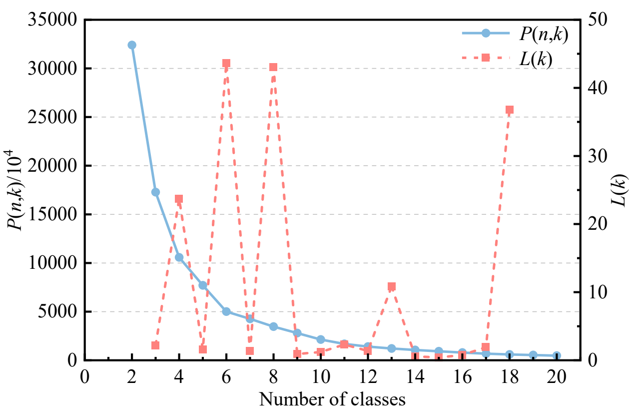

Fig.9. Variation trend of the class loss function

k =6 marks the inflection point in the variation trend of the loss function from steep to gentle. At this time, the change rate of the slope of adjacent periods reaches the maximum value of 43.6, so the best number of classes is 6. The results of the operation periods divided based on the optimal segmentation method are listed on the left side of Tab.1.

TABLE I
DIVISION RESULTS OF URT OPERATION PERIODS CONSIDERING MULTI-PASSENGER FLOW CHARACTERISTICS

Optimal segmentation method			Ordered sample clustering based on SA		
Period number	Operation period	Period name	Period number	Operation period	Period name
1	06:00-08:00	Morning off-peak period	1	06:00-07:15	Morning off-peak period
2	08:00-09:30	Morning peak period	2	07:15-9:00	Morning peak period
3	09:30-17:30	Midday off-peak period	3	09:00-17:00	Midday off-peak period
4	17:30-19:30	Evening peak period	4	17:00-19:00	Evening peak period
5	19:30-22:00	Evening off-peak period	5	19:00-21:45	Evening off-peak period
6	22:00-24:00	Night off-peak period	6	21:45-24:00	Night off-peak period

It can be seen from Tab.1 that the six time periods of Nanjing Metro Line 3 divided by the optimal segmentation method are 06:00-8:00, 08:00-9:30, 09:30-17:30, 17:30-19:30, 19:30-22:00 and 22:00-24:00.

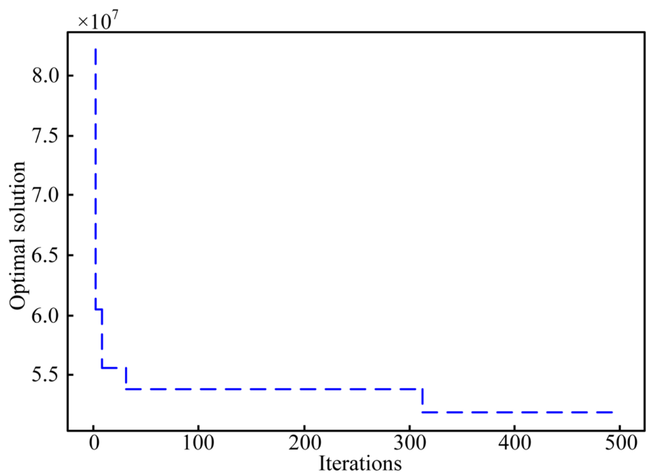

Fig.10. Algorithm iteration curve

Similarly, under the characteristics of multiple passenger flows, the URT operation periods are divided by the ordered clustering algorithm based on SA. Fig.10 shows the iterative curve of the ordered clustering algorithm based on SA for the division of the URT operation periods. After 314 iterations, the SA converges to the optimal solution (division point), which is the division point [0,5,12,44,52,63,72]. This solution divides the operation periods into six classes. The division results of the operation periods are listed on the right side of Tab.1, and the visualized division results are shown in Fig.11.

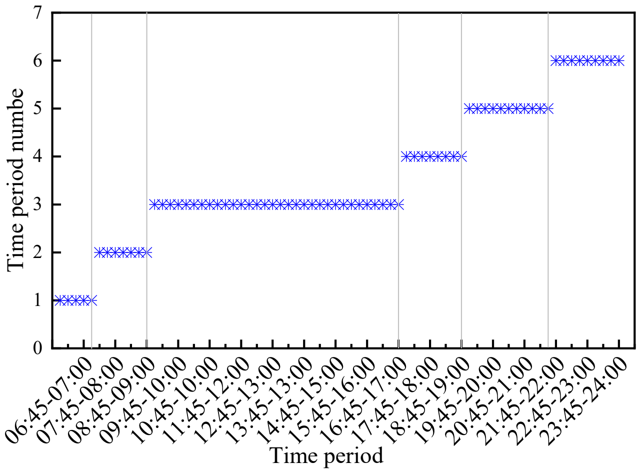

Fig.11. Ordered clustering results based on SA.

From Tab.1 and Fig.11, it is evident that the ordered clustering algorithm based on SA divides Nanjing Metro Line 3 into six periods: 06:00-07:15, 07:15-9:00, 09:00-17:00, 17:00-19:00, 19:00-21:45, 21:45-24:00.

By comparing the division results of URT operation periods under the characteristics of single-passenger flow and multi-passenger flow, it can be found that. The number of operation periods divided by multi-passenger flow characteristics is small, and the length of operation periods is relatively balanced, which can clearly distinguish between peak periods and off-peak periods, and has a high degree of consistency with the actual operation data, which is convenient for making train plans.

However, it is difficult to judge the advantages and disadvantages of the optimal segmentation method and the ordered clustering algorithm based on SA for the division of operation periods only from the division results, and further evaluation is needed for specific evaluation indicators.

V. Discussion and Analysis

A. Algorithm Evaluation

The clustering results need to be evaluated by the effectiveness index to determine the clustering effect and the effectiveness of the clustering method. The classification performance of different clustering methods is tested by selecting the running time, loss function value, silhouette coefficient, and C-H coefficient.

(1) Running time

We compared the running time of the two algorithms in the same runtime environment, programmed using Python on the PyCharm 2019 platform. We used seconds as the unit of measurement to determine the time cost of both algorithms. Fig.12 shows the running times of the two algorithms.

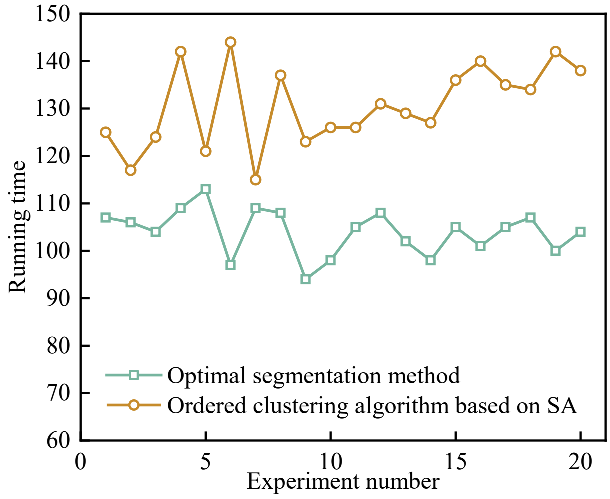

Fig.12. Comparison of running time

The running time of the optimal segmentation method is significantly lower than that of the ordered clustering algorithm based on SA. In addition, the ordered clustering algorithm based on SA exhibits greater fluctuations in running time across different runs, which is attributed to the random search strategy inherent in heuristic algorithms.

(2) Loss function value

In the optimal segmentation method, the sum of the squares of various deviations is utilized to represent the loss function value. This value is used to assess the effectiveness of the segmentation method, the smaller the value, the better the segmentation method performs. The ordered clustering algorithm based on SA aims to minimize the sum of squared deviations within each class, striving to make the characteristics of passenger flow within the same class more similar. Fig.13 shows the loss function values from 20 experimental trials for both methods.

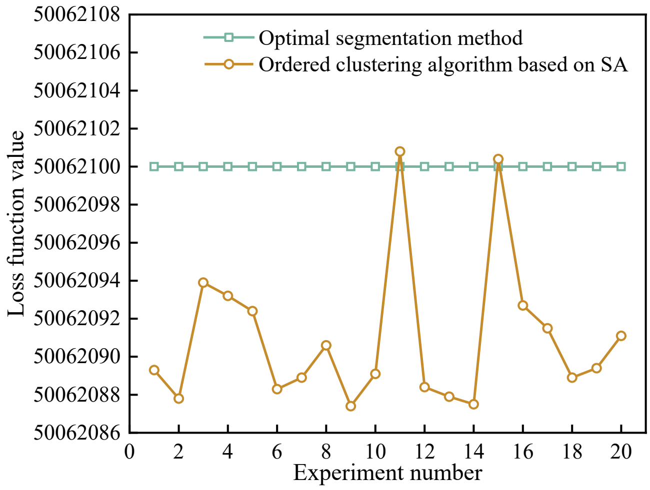

Fig.13. Comparison of loss function value

The loss function value of the optimal segmentation method is 50062100, and the result does not change with the number of experiments. The optimal objective value of the ordered clustering algorithm based on SA fluctuates to a certain extent, and most of the results are significantly better than the optimal segmentation method, except that some experimental data are inferior to the optimal segmentation method.

(3) Silhouette coefficient

The silhouette coefficient S evaluates the clustering effect of different algorithms by the average distance between any sample and other samples in the class and the average distance between any sample and all samples in the adjacent class. The larger the silhouette coefficient value, the better the clustering effect. Silhouette coefficient can be calculated by (11).

$$S=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{11}$$

Where, $a(i)$ is the average distance from the sample i to other samples in the class. $b(i)$ is the average distance from the sample i to all samples in the other classes.

(4) C-H coefficient

The C-H coefficient evaluates the clustering effect by assessing the variance between classes and within classes. The larger the C-H coefficient, the better the classification effect of the corresponding clustering method. The formula used for the calculation is provided in Equation (12).

$$CH=\frac{tr\left[\sum_{q=1}^{k} m_q(c_q-c)(c_q-c)^T\right]}{tr[\sum_{q=1}^{k}\sum_{x\in c_q}(x-c_q)(x-c_q)^T]}\times\frac{N-K}{N-1} \tag{12}$$

Where, m_q is the number of samples in class q. c_q is the center point of the class q. c_p is the center point of all samples.

Compare the results of the optimal segmentation method and the ordered clustering algorithm based on SA and evaluate the performance of the two algorithms. The results are shown in Tab.2.

As shown in Tab.2, although the average running time of the optimal segmentation method is less than that of the ordered clustering algorithm based on SA, the difference is not significant. The ordered clustering algorithm based on SA has demonstrated improvements in various aspects, including the loss function value, silhouette coefficient, and C-H coefficient. These improvements indicate that the ordered clustering algorithm based on SA is effective for dividing URT operation periods and offers clear advantages in terms of accuracy.

To sum up, compared with only considering the passenger flow at different periods of the day, the URT operation periods division method considering the characteristics of multiple passenger flows can better reflect the change rule of the URT passenger flow of the whole day, which is more in line with the actual operation of URT, and can serve as a fundamental element for designing train service plans. By comparing the results of the optimal segmentation method and the ordered clustering algorithm based on SA, it can be found that the ordered clustering algorithm based on SA avoids the results from falling into a local optimum, although the search time for the optimal solution is longer when facing large samples. It can also be seen from the objective function value and clustering result evaluation parameters that the ordered clustering algorithm based on SA has a better classification effect.

B. Parameter Analysis

The size of passenger flow and model parameters have a certain impact on the division results of the URT operation periods. We analyze the effects of parameter changes on the division results of the passenger flow periods by adjusting the passenger flow at different periods of the day, the average value of passenger flow change coefficient of adjacent cross-sections in a single period, cross-section passenger flow at different periods and the annealing rate of SA.

When the passenger flow at different periods of the day, cross-section passenger flow at various periods, and the average value of passenger flow change coefficient of adjacent cross-sections in a single period are taken as 0.1 times, 0.5 times, 1.0 times, 2.0 times, and 3.0 times the actual value of the case(denoted as β), the optimal objective values for various parameters are presented in Fig.14.

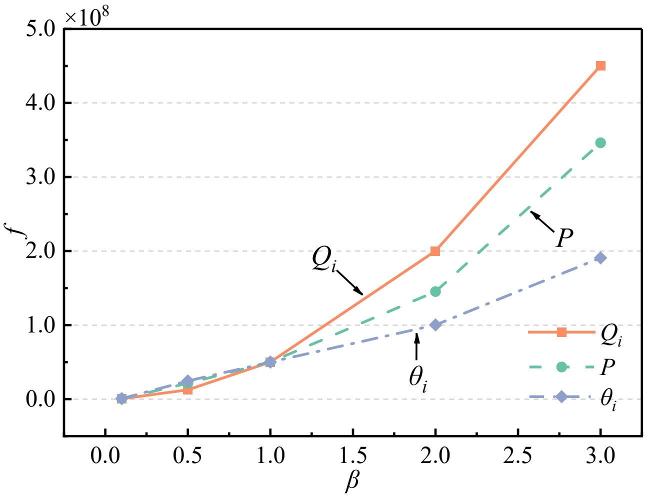

Fig.14. Optimal objective value for various parameters

From Fig.14, in general, the optimal solution increases with the increase of passenger flow, and its growth rate also increases, indicating that the optimal solution is directly related to the value of corresponding passenger flow characteristics.

However, when the passenger flow sizes of the three passenger flow characteristics are changed respectively, it can be found by comparing the results of the division of the operation periods that the division results of the periods will not change with the change of the passenger flow, which shows that the division results of the operation periods of URT have nothing to do with the size of the passenger flow at different periods of the day, but only with the similarity law between the passenger flow data. This rule can also be seen in Fig.15. This aligns with the real-world scenario.

TABLE II
ALGORITHM EVALUATION RESULTS

Algorithm	Running time	Loss function value (Optimal objective value)	Silhouette coefficient S	C-H coefficient
Optimal segmentation method	104	50062100	0.12	2452
Ordered clustering algorithm based on SA	131	50062087.4	0.15	2494

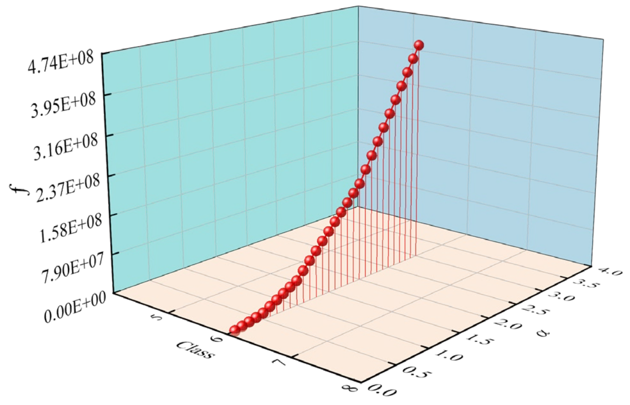

Fig.15. Optimal objective value under different parameters

By observing Fig.14, it can also be found that the growth rate of the optimal solution is the fastest when passenger flow at different periods of the day Q_i increases, indicating that the passenger flow at different periods of the day Q_i is the main influencing factor for the division of URT operation periods, the secondary influencing factor is cross-section passenger flow at various periods P, and the average value of passenger flow change coefficient of adjacent cross-sections in a single period θ_i has little impact on the division of URT operation periods.

The annealing rate of SA determines the cooling rate of the algorithm, typically ranging from 0.6 to 0.99 [31]. By adjusting the annealing rate of SA to 0.6, 0.7, 0.8, 0.9, 0.95, and 0.99, the influence of algorithm parameters on the division results of the URT operation periods was observed. The results of the number of periods divisions and the variation of the optimal objective value with the annealing rate α are illustrated in Fig.16.

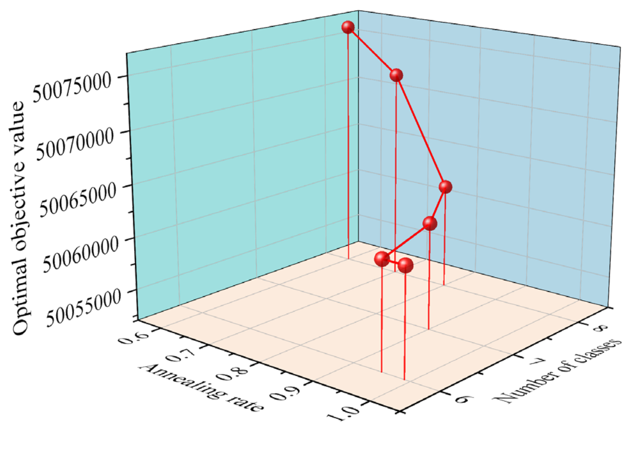

Fig.16. Effect of annealing rate

It is easy to see from Fig.16 that the value of the annealing rate impacts the number of operation periods and the optimal solution. When $\alpha \geqslant 0.9$, the optimal solution is consistent with the optimal solution in the case and remains unchanged. When $\alpha = 0.9$, the URT operation periods are divided into seven classes, and when $\alpha = 0.95$ and $\alpha = 0.99$, the URT operation periods are divided into six classes. Combining some periods when $\alpha = 0.9$ means the same division results as $\alpha = 0.95$ and $\alpha = 0.99$ can be obtained. It can be seen that the value of the annealing rate for the URT operation periods should be as high as possible so that all solutions in the whole solution space can be searched.

VI. Conclusion

The division of the URT operation periods is an essential basis for operating enterprises to formulate and adjust train plans and improve operation services. This paper summarizes the existing research, introduces the optimal segmentation method in the ordered clustering method, and proposes a new ordered clustering algorithm based on SA for the division of the URT operation periods. The proposed algorithm is verified by the AFC data of Nanjing Metro Line 3 for five consecutive working days by selecting the passenger flow at different periods of the day, cross-section passenger flow at various periods, and the average value of passenger flow change coefficient of adjacent cross-sections in a single period. The conclusions of this paper consist of the following four.

(1) When dividing the operation periods of the target line only considering the characteristics of single passenger flow (passenger flow at different periods of the day), the division results of the operation periods using the optimal segmentation method and the ordered clustering algorithm based on SA are the same. Still, the matching between the division results and the actual demand is poor.

(2) The multi-passenger flow characteristics that integrate time and space can reflect the impact of multiple factors on the fluctuation of passenger flow throughout the day. The optimal segmentation method and the ordered clustering algorithm based on SA divide the operation time into six periods. The peak and off-peak periods are obvious, and the division results are in line with the actual operation situation.

(3) By analyzing the segmentation results of the optimal segmentation method and the ordered clustering algorithm based on SA, it is found that the ordered clustering algorithm based on SA has a longer search time, the difference in running time is 27 seconds, but the optimal target value is improved by 12.6, the silhouette coefficient by 25%, and the C-H coefficient by 1.7%. These enhancements suggest that the ordered clustering algorithm based on SA, when combined with multiple passenger flow characteristics as proposed in this paper, offers greater advantages for dividing URT operation periods.

(4) By analyzing the sensitivity of parameters such as passenger flow and annealing rate of SA, it can be found that the passenger flow of passenger flow characteristics does not affect the division results of specific periods but only the value of the objective function in the division results. To improve the accuracy of URT operation time division results as much as possible, the value of the annealing rate should be larger, such as above 0.9.

This study has conducted detailed research on the division of URT operation periods, and the contents to be further studied are reflected in the following aspects. This paper only considers the relationship between the change in passenger flow characteristics on weekdays and the operation periods, and does not consider the division of the URT operation periods during holidays, especially when the passenger flow continues to be large. The division of the operating periods for URT is a prerequisite for developing train operation schedules and operating plans based on demand and capacity matching. The next step in this paper will be to study train operation schemes for URT based on the division of operating periods.

REFERENCES

[1] C. Zhu, X. Yang, Z. Wang, J. Fang, J. Wang, L. Cheng, "Optimization for the train plan with flexible train composition considering carbon Emission," *Engineering Letters*, vol. 31, no. 2, pp. 562-573, 2023.
[2] H. W. Wu, J. Zhen, J. Zhang, "Urban rail transit operation safety evaluation based on an improved CRITIC method and cloud model," *Journal of Rail Transport Planning & Management*, vol. 16, Article ID. 100206, 2020.
[3] M. Nold, and F. Corman, "Dynamic train unit coupling and decoupling at cruising speed: Systematic classification, operational potentials, and research agenda," *Journal of Rail Transport Planning & Management*, vol. 18, Article ID. 100241, 2021.
[4] K. He, X. Cheng, S. Zhang, H. Yang, Y. Yao, M. Peng, et al., "Critical roof opening longitudinal length for complete smoke exhaustion in subway tunnel fires," *International Journal of Thermal Sciences*, vol. 133, pp. 55-61, 2018.
[5] Y. Xing, J. Lu, S. Chen, S. Dissanayake, "Vulnerability analysis of urban rail transit based on complex network theory: A case study of Shanghai Metro," *Public Transport*, vol. 9, pp. 501-525, 2017.
[6] D. Pan, L. T. Zhao, Q. Luo, C. S. Zhang, Z. J Chen, "Study on the performance improvement of urban rail transit system," *Energy*, vol. 161, pp. 1154-1171, 2018.
[7] Y. Chen, Y. Hu, L. Lai, " Demography-oriented urban spatial matching of service facilities: Case study of Changchun, China," *Land*, vol. 11, no. 10, Article ID. 1660, 2022.
[8] F. A. Awad, D. J Graham, L. AitBihiOuali, R. Singh, "Performance of urban rail transit: a review of measures and interdependencies," *Transport Reviews*, vol. 43, no.4, pp. 698-725, 2023.
[9] D. Zheng, Y. Wang, P. Z. Tang, Y. P. Wu, "Application of data mining in the forecasting of railway passenger flow," *Advanced Materials Research*, vol. 834, pp. 958-961, 2014.
[10] M. Tang, J. Cao, D. Gong, G. Xue, "Unsupervised learning-based exploration of urban rail transit passenger flow and travel pattern mining," *International Journal of Computers Communications & Control*, vol. 19, no. 2, 2024.
[11] Z. Jiang, Y. Tang, J. Gu, Z. Zhang, W. Liu, "Discovering periodic frequent travel patterns of individual metro passengers considering different time granularities and station attributes," *International Journal of Transportation Science and Technology*, vol. 14, pp. 12-26, 2024.
[12] B. L. Smith, W. T. Scherer, T. A. Hauser, "Data-mining tools for the support of signal-timing plan development," *Transportation Research Record*, vol. 1768, no.1, pp. 141-147, 2001.
[13] B. Park, D. H. Lee, I. Yun, "Enhancement of time of day based traffic signal control. In SMC'03 Conference Proceedings," 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483), IEEE, Vol. 4, pp. 3619-3624, 2003.
[14] B. Park, and J. Lee, "Enhancement of time of day based traffic signal control. In SMC'03 Conference Proceedings," 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance," *KSCE Journal of Civil Engineering*, vol. 12, pp. 37-44, 2008.
[15] J. Patnaik, S. Chien, A. Bladikas, "Using data mining techniques on APC data to develop effective bus scheduling plans," *Journal of Systemics, Cybernetics and Informatics*, vol. 4, no. 1, pp. 86-90, 2006.
[16] W. Mo, Y. G. Liu, L. Y. Ge, X. Geng, "Optimization of bus departure headway based on fuzzy clustering of operation period considering congestion and passenger flow," *In 2019 5th International Conference on Transportation Information and Safety (ICTIS)*, IEEE, pp. 963-968, 2019.
[17] M. Salicrú, C. Fleurent, J. M. Armengol, "Timetable-based operation in urban transport: Run-time optimisation and improvements in the operating process," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 8, pp. 721-740, 2011.
[18] Z. C. Xu, P. J. HE., J. Teng, L. P. Liu, "Transit vehicles intelligent scheduling optimization based on the division of characteristic periods," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 1502-1512, 2013.
[19] W. Z. Jin, P. Li, W. T. Tao, "Time-of-day interval partition method for bus schedule based on multisource data and fleet-time cost optimization," *China Journal of Highway and Transport*, vol. 32, no. 2, pp. 143-154, 2019.
[20] W. J. Li, F. Sun, X. Y. Li, D. F. Dong, "Time-of-day breakpoints for traffic signal control using dynamic recurrence order clustering," *Journal of Zhejiang University(Engineering Science)*, vol. 52, no. 6, pp. 1150-1156, 2018.
[21] Y. D. Shen, T. H. Zhang, J. Xu, "Homogeneous bus running time bands analysis based on K-means algorithms," *Journal of Transportation Systems Engineering and Information Technology*, vol. 14, no. 2, pp. 87-93, 2014.
[22] D. Y. Chen, D. W. Chen, S. X. Jiang, N. Xu, "Division of metro operation periods based on feature clustering of passenger flow," *Computer Systems & Applications*, vol 30, no. 3, pp. 256-261, 2021.
[23] R. Tang, L. De Donato, N. Besinović, F. Flammini, R. Goverde, Z. Lin, et al., "A literature review of artificial intelligence applications in railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 140, Article ID. 103679, 2022.
[24] W. X. Wang, M. Xiao, L. N. Cheng, Y. S. Du, S. Q. Ni, "Classification of subway operation intervals based on affinity propagation cluster," *Operations Research and Management Science*, vol. 27, no. 12, pp. 187-192, 2018.
[25] X. Ding, H. Hu, Z. Liu, H. Zhu, "The aggregation mechanism mining of passengers' flow with period distribution based on suburban rail lines," *Discrete Dynamics in Nature and Society*, 2017.
[26] H. Niu and M. Zhang, "An optimization to schedule train operations with phase-regular framework for intercity rail lines," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID. 549374, 2012.
[27] S. Sikdar, N. Ganguly, A. Mukherjee, "Time series analysis of temporal networks," *The European Physical Journal B*, vol. 89, pp. 1-11, 2016.
[28] H. Cui, L. Zhou, Y. Li, B. Kang, "Belief entropy-of-entropy and its application in the cardiac interbeat interval time series analysis," *Chaos, Solitons & Fractals*, vol. 155, Article ID. 111736, 2022.
[29] C. Zheng, S. Y. Phang, "Urban rail transit PPPs: Lessons from east asian cities," *Transportation Research Part A: Policy and Practice*, vol. 105, pp. 106-122, 2017.
[30] Z. F. Wang, C. Zhu, J. H. Fang, "Optimization of traveler intermodal transport routes considering carbon emissions," *Engineering Letters*, vol. 31, no. 3, pp. 1255-1264, 2023.
[31] E. Joelianto, B. Rahmat, P. I. Siregar, E. Juliastuti, S. Bandong, "Performance of optimized CSK, DFT, and LOT for video-based container tracking system using simulated annealing," *Engineering Letters*, vol. 30, no. 3, pp. 1050-1064, 2022.

Changfeng Zhu was born in Gansu, China , in 1972. In 2003 and 2013, he obtained master's and doctor's degrees in the Planning and Management of Traffic and Transportation in Lanzhou Jiaotong University. He is now a professor and doctoral supervisor of the School of Traffic and Transportation of Lanzhou Jiaotong University. His main research direction is transportation planning and management. He has made fruitful research achievements, presided over a number of national and local level fund projects, and published more than 50 papers that were retrieved by databases EI and SCI.

Liangsheng Zhao was born in Gansu, China, in 1996. He obtained his bachelor's degree in Traffic and Transportation from Lanzhou Jiaotong University, Lanzhou, China, in the year 2021. He is currently pursuing his master's degree in Traffic and Transportation (the Planning and Management of Traffic and Transportation) in Lanzhou Jiaotong University. His research interests include multi-level rail transit planning and management.