High-speed Railway Passenger Flow Analysis Method and Platform Design Based on the Concept of One-Day-One-Plan

Bowen Nie, Lei Nie and Huiling Fu*
School of Traffic and Transportation, Beijing Jiaotong University, China

*Corresponding author email: hlfu@bjtu.edu.cn

Abstract. Based on the concept of One-Day-One-Plan, passenger demand should be considered different every day while timetable should also be adjusted according to passenger flow fluctuation. The analysis of passenger flow provides basis of line planning problem and train timetabling problem. This paper uses cluster analysis to integrate the characteristic of passenger flow data, designs and subdivides the service level of each OD. A data analysis platform connecting with Line Planning Section and Train Timetabling Section is developed to achieve the quick response of the real-time passenger flow fluctuation, ensure the supply-demand match and improve service level.

Keywords: Passenger flow; Data analysis; One-Day-One-Plan; Cluster analysis.

1. Introduction

With the development of economy and the improvement of people's consumption level, high-speed railway has become the first choice of short and medium distance travel for passengers due to its high punctuality, high safety, high speed and high comfort. This further improves the heterogeneity of passenger flow between different lines and different origin-destination pairs (OD) of an individual line. In order to meet the demand and ensure that 'Every passenger can get on the train, while every train doesn’t waste its capacity', the law about how passenger flow change by time should be explored. The concept of One-Day-One-Plan is to adjust service level of each OD, implement different line plans and timetables on different days of a week in order to adapt to the fluctuation of passenger flow. Taking China Railway Shanghai Group Co. as an example, it implemented the Weekday-plan with less trains from Tuesday to Thursday in order to reduce the waste of capacity in off peak periods, and the Weekend-plan with more trains from Friday to Monday in order to increase the train seating capacity in peak periods.

One-Day-One-Plan requires decision makers to conduct more refined analysis and design corresponding train service level according to real-time passenger flow, flexibly establish the line plans and timetables. Therefore, an effective platform and data analysis method should be developed. Exist researches shows different ways to construct passenger flow analyzing platform. Based on SQL Server database, Liyan[1] built a data analysis and chart display system for high-speed railway passenger tickets. Tianru[2] called TRS system to obtain the passenger flow data of Wuhan-Guangzhou high-speed railway extension line, and modeled the function relationship between passenger flow and other attributes. Hettrakul[3] used econometric techniques and investigated heterogeneous behavior of different kinds of passenger. Most of the platforms in given research have complex response processes, and the data processing is not refined enough. Based on the existing research, in order to realize the fine-analysis for passenger flow and quick response to any change, this paper constructs a real-time passenger flow data analysis and processing platform and OD service level design method.
2. Passenger Flow Analysis Method

This section refining the passenger flow data and extracted the heterogeneous passenger flow characteristics of different ODs by cluster analysis.

2.1. Data Integration and Processing

By observing the fluctuation patterns of passenger flow dispersed by 1 hour, 2 hours and 4 hours, it is found that the fluctuation trend of 2-hours-curve is basically similar to 1-hour-curve, but the dimensions is reduced by 50%. In addition, provided that periodic train operation is considered, many high-speed railway lines in China are of long-distance and suitable for adopting 2-hours periodic timetables. Therefore, for each OD, we choose 9 periods of a day: 6.00-8.00, 8.00-10.00 … 22.00-24.00, depicting the fluctuation line of passenger flow.

In addition, considering the OD of non-departure station, the average time for the train from the departure station to the O station is \( t_{\text{travel}} \), then the passenger demand of this OD in a period of \([t_1, t_2]\) is actually satisfied by trains with the departure time in the period of \([t_1 - t_{\text{travel}}, t_2 - t_{\text{travel}}]\). In the design of periodic line plan, the time interval of the train is based on its departure time. Therefore, the division of time periods should be adjusted in the process of passenger flow analysis and service level determination of non-departure ODs.

2.2. Cluster Analysis and Determination of One-Day-One-Plan Model

According to the method proposed above, we have 63 passenger flow data of each OD, as a variable of cluster analysis. We are sure that the information contained in each dimension is mutually exclusive. In order to avoid the reduction of the sensitivity of algorithm caused by the difference of OD passenger flow, each group of data is standardized.

Combined with the previous research precedent, we choose K-means clustering method. The Elbow-principle is used to determine the number of clusters. We hope that the OD with more passengers can be better described, so we test the average value of the distance from the individual OD to the cluster center under different K values, especially the ODs with large passenger flow, and select the clustering result with smaller weighted average value as relatively appropriate clustering results.

Determine the clustering results of each OD on a line, we observe the fluctuation patterns of each cluster, and classify the consecutive days with similar fluctuation patterns of passenger flow into the same plan. If it is similar from Tuesday to Thursday, it will be collectively classified as Weekday-plan; if it is similar on Saturday and Sunday, it will be collectively classified as Weekend-plan. Finally, determine the One-Day-One-Plan model for the line, such as \(1 + 3 + 1 + 1 + 1\), which is shown in Figure 1 (Monday-plan, Weekday-plan, Friday-plan, Saturday-plan, Sunday-plan, etc.).

![Figure 1. Determine the One-Day-One-Plan Mode.](image)

2.3. Clustering Analysis of OD Characteristics and Passenger Flow Components

The analysis of the passenger flow component is helpful to determine the train service level for different types of passenger flow, such as daily commuting passenger flow, weekly commuting passenger flow, business passenger flow, tourist flow, etc. As the data doesn’t directly contain the component
information, some OD features can be clustered to make analysis. The following features are mainly considered:

2.3.1. The fluctuation of passenger flow. OD of different components, has different fluctuation pattern, which is shown in Figure 2. The above results have been used to cluster the wave patterns.

![Figure 2. Relationship Between Different Components and Different Fluctuation Pattern.](image)

2.3.2. Station attributes. Such as station level, city level, etc. As an example, the business passenger flow and commuting passenger flow in ODs between big cities account for a relatively high proportion. The tourist passenger flow in ODs between tourist cities also accounts for a relatively high proportion.

2.3.3. Relation between O and D station. Such as mileage, economic differences, political level differences, etc. For example, commuting passenger flow between a capital city and small cities surrounding it accounts for a high proportion.

2.3.4. Time. Different periods of a day bring different components. For example, the morning and evening peak of working day is dominated by commuting passenger flow, and it at noon accounts for a relatively low proportion.

| Table 1. Judgement Conditions and Adjustment Range of Each Component |
|---------------------------------------------------------------|
| Component          | Suitable conditions      | Adjustment | Enable conditions | Adjustment | Unfavorable conditions | Adjustment |
| Daily Commuting Flow (DCF) | Between large and small city, short millage | +10% | Short millage | No change | Small cities, long millage | -10% |
| Weekly Commuting Flow (WCF) | Between large and small/medium city, medium millage | +10% | Medium millage | No change | Small cities | -10% |
| Business Flow (BF) | Large cities, high GDP level | +10% | Large cities | No change | Small cities | -10% |
| Tourist Flow (TF) | Big cities or Tourist cities | +10% | Tourist cities | No change | Without tourist cities | Change to 0% |
2.3.5. Existing service level. The limited transportation capacity may restrain the actual transportation demand. According to the results of cluster analysis, we can qualitatively judge which kind of passenger flow component the ODs of each cluster is more suitable for. As an example, when we find a fluctuation type that has an obvious early peak, which matches the judgment characteristics of daily commuting passenger flow, we then consider this OD with a high proportion of commuting passenger flow. According to the questionnaire and other methods, the macro proportion of various passenger flow components on the line can be obtained. Combined with the analysis, we determine whether an OD is suitable or unfavorable for the existence of different components, then fine-tune the proportion of these components. Qualitative judgment conditions and adjustment range of 4 kinds of passenger flow components are shown in table 1.

3. Service Level Analysis Method
The Service Level Design of each OD is to mainly determine the train service frequency, which refers to the number of trains stopping at both origin station and destination station within a certain period of time. Under the concept of One-Day-One-Plan, in order to achieve refined passenger flow matching, it is necessary to determine the service frequency of each OD in each week and each period according to the existing timetables and passenger flow. For OD with a large number of passengers, we can divide the total service frequency for different passenger flow components. There are also Commuter trains or Tourist trains may be operated.

3.1. Determine the Relationship between Daily Passenger Flow and Daily Service Frequency
As we have the average daily passenger flow and the train service frequency of the existing line plans and timetables, we can generate a scatter diagram and obtain a fitting curve, then the corresponding relationship between them is determined. In the existing line plan in China, not all the ODs have the appropriate service levels which fit the passenger flow demand well. The train service frequencies of some ODs between small and medium stations are too high while some of ODs between large stations are too low. The corresponding relationship can determine the appropriate service level for each OD according to its passenger flow demand.
In practical operation, it can be assumed that the train service level of OD with the largest passenger flow of an individual line, or between major stations, is reasonable. Based on these ODs, some ODs with unreasonable service level, less passenger flow and less regular service frequency can be eliminated in order to reduce the interference to fitting calculation.

3.2. Subdivision of Daily Service Frequency
After the reasonable daily service frequency of each OD has been obtained, it should be subdivided into the service frequency for each period. In order to guarantee the regularity of timetable, service frequency of each period should not be completely different. Combined with the operating experience, these periods can be divided into ‘peak period’ and ‘non-peak period’. For example, the peak period may usually be 8.00-10.00 in the morning and 4.00-8.00 in the evening. As for an OD, the service level in all the peak periods and all the non-peak periods of a day are the same, while it may be different in peak period of different days, in order to fit the periodic fluctuation of passenger flow.
For ODs with high service level, the service frequency can be further subdivided according to the results of passenger flow component analysis, such as allocating a certain proportion of train dedicated service commuting passenger flow in each morning. The subdivision may still be helpful to differential pricing of the railway train products.

4. Case Study
Taking Shanghai-Nanjing Intercity High Speed Railway as an example, using data from July 14 to July 21 in 2019. The platform is based on the environment of Intel Core i5-7300hq CPU and 20G memory. The data fields are shown in the table 2.
Table 2. Fields.

| Field Name      | Type  | Field Name     | Type  |
|-----------------|-------|----------------|-------|
| BOARDDATE       | Char  | PSGNUM         | Int   |
| TRAINNO         | Char  | SEATTYPE       | Char  |
| ORI STATION     | Char  | MILE           | Int   |
| END STATION     | Char  |                |       |

The design diagram of this platform is shown below in Figure 3.

Cluster of the fluctuation patterns for each OD is shown in Figure 4. It can be seen that for most of the ODs, there are two peaks in one day. The morning peak is always among 7 am and 10 am, while the evening peak is always among 4 pm and 8 pm. Because of the commuting passengers, on Friday and Sunday, the evening peaks are always higher than other days (Cluster 1, Cluster 5, Cluster 7, Cluster 9). And the morning peaks on Monday and Saturday are higher than other days (Cluster 2, Cluster 5, Cluster 8, Cluster 9).

Also, for most of the ODs, there are some common differences between the fluctuation patterns of passenger flow on Friday, Saturday, Sunday, and Monday, while similar from Tuesday to Thursday. Therefore, it is determined that the One-Day-One-Plan mode of this railway line is “1 + 3 + 1 + 1 + 1”, which is interpreted as Monday Plan + Weekday Plan + Friday Plan + Saturday Plan + Sunday Plan.

We analyze the passenger flow component and choose the daily average passenger flow, urban population, GDP, OD level, political connection, mileage, train service frequency of the existing line plan, travel time, tourist city (equals 2 means both origin station and destination station are tourist city while 0 means none) as cluster variables. The clustering results are shown in table 3. ODs in cluster 1 have the largest daily average passenger flow and contains most of the important ODs between big stations, having the highest OD level. And for the large cities are always tourist cities, the cluster 1’s ODs also have the largest number of tourist city. The ODs in cluster 4 have the highest service frequency in the exist line plan, but their passenger flow are lower than cluster 1. In fact, these ODs are mostly between Changzhou, Wuxi and Suzhou. Further analysis shows that these ODs’ service frequencies are higher than necessary. Cluster 6 have the highest political connection and mileage, and it contains ODs from small stations located in Shanghai (such as Shanghai West Station and Anting North Station) to small stations located in Nanjing.
According to the characteristics of each cluster, we determine whether it is conducive to the generation of different passenger flow components, and fine tune different OD passenger flow components. Taking Nanjing-Shanghai HSR line as an example, the fine-tuning process and results are shown in table 4.

| Cluster | Daily average passenger flow | Urban population | GDP per person | OD level | Political connection | Mileage | Service frequency | Travel time | Tourist city |
|---------|-------------------------------|------------------|----------------|---------|---------------------|---------|------------------|-------------|-------------|
| 1       | 4982.1                        | 1371.0           | 3.5            | 7.3     | 2.7                 | 159.8   | 49.5             | 1.0         | 1.8         |
| 2       | 59.1                          | 812.1            | 5.2            | 5.0     | 0.5                 | 31.1    | 5.1              | 0.2         | 1.2         |
| 3       | 63.4                          | 208.6            | 6.0            | 2.1     | 1.1                 | 79.7    | 6.1              | 0.5         | 0.0         |
| 4       | 2030.4                        | 901.3            | 5.1            | 4.6     | 1.2                 | 90.2    | **64.3**         | 0.6         | 1.0         |
| 5       | 513.2                         | 808.5            | 3.9            | 5.3     | 2.8                 | 88.9    | 7.7              | 0.5         | 1.0         |
| 6       | 892.8                         | 897.9            | 0.9            | 6.8     | **3.6**             | 196.7   | 4.4              | 1.2         | 1.4         |
| 7       | 66.6                          | 441.7            | 4.5            | 3.9     | 2.2                 | 4.4     | 1.2              | 0.8         |             |

Table 4. Fine-tuning Process of Nanjing-Shanghai OD.
### Results

|                | Commuting flow | Business flow | Tourist flow | Others |
|----------------|----------------|---------------|--------------|--------|
|                | 34%            | 40%           | 15%          | 10%    |
|                | 34%            | 40%           | 15%          | 10%    |
|                | 29%            | 46%           | 15%          | 10%    |
|                | 29%            | 46%           | 15%          | 10%    |
|                | 29%            | 46%           | 15%          | 10%    |
|                | 33%            | 43%           | 14%          | 10%    |
|                | 33%            | 43%           | 14%          | 10%    |
|                | 30%            | 42%           | 16%          | 11%    |
|                | 30%            | 42%           | 16%          | 11%    |

Then we analyze the relationship between the daily average train service frequency and the daily average passenger flow, and use a straight line passing through the origin point and a logarithmic function curve tangent to it to fit the function. The above analysis results will be put directly into the platform and will not change in the short term, however, if the passenger flow on the line rises obviously, data fitting is still needed again. The train service frequency correction results corresponding to the passenger flow data are shown in Figure 5.

![Service Frequency Correction Results.](image)

**Figure 5. Service Frequency Correction Results.**

The daily service frequency is divided into peak period and non-peak period as shown in table 5. If an OD has a service frequency of 7 during non-peak period, and ‘+3’ in its peak period, then the peak period frequency should be 10. The platform reflects the variation of real-time passenger flow on the change of service frequency in different weeks, peak periods and non-peak periods, so as to achieve flexible supply and demand matching.

### Table 5. Service Frequency subdivision.

| Odname (abbreviations for stations are used) | Service frequency of non-peak period | Train service frequency of peak period |
|---------------------------------------------|--------------------------------------|---------------------------------------|
|                                             | 6.-8.                  | 8.-10.                  | 10.-12.                 | 12.-14.                 | 14.-16.                 | 16.-18.                 | 18.-20.                 | 20.-22.                 | 22.-24.                 |
| SH-SZ                                       | 7                      | 0                       | +3                      | 0                       | 0                       | +3                      | +3                      | 0                       | 0                       |
| SZ-SH                                       | 7                      | 0                       | +1                      | 0                       | 0                       | +1                      | +1                      | 0                       | 0                       |
| SHHQ-SZ                                     | 6                      | 0                       | +1                      | 0                       | 0                       | +1                      | +1                      | 0                       | 0                       |
| SZ-SHHQ                                     | 6                      | 0                       | +1                      | 0                       | 0                       | +1                      | +1                      | 0                       | 0                       |
| NJ-SH                                       | 5                      | 0                       | +2                      | 0                       | 0                       | +2                      | +2                      | 0                       | 0                       |
| SH-NJ                                       | 5                      | 0                       | +2                      | +2                      | 0                       | +2                      | +2                      | 0                       | 0                       |

### 5. Conclusion

In order to make the programming of line plan and timetable more efficient, achieve a better matching of supply and demand, this paper constructs a data analysis and processing platform to realize the timely response to the change of passenger flow, generates the service level demand, prepares the corresponding line plans and timetables, and adjusts the transportation supply flexibly. In the future research, we will develop the line plan and timetable calculating module, fully achieve the automation
of data acquisition analysis modeling optimization, further strengthen the ability of supply and demand matching, and provide passengers with a higher level of train service.

Acknowledgement
The authors would like to thank referees and the editor for very helpful and constructive comments throughout the reviewing process. This work was supported by the National Key Research and Development Plan of China under grant (2016YFE0201700, 2018YFB1201402).

References
[1] Li-yan HUANG. Research on Passenger Flow Analysis of High Speed Railway Based on Big-Data[D]. Southwest Jiaotong University, 2017.
[2] Tian-ru BI. Study on Distribution Law of Time about Passenger Flow in Wuhan-Guangzhou High-speed Railway[D]. Southwest Jiaotong University, 2018.
[3] Hetrakul P, Cirillo C. Accommodating taste heterogeneity in railway passenger choice models based on internet booking data[J]. Journal of Choice Modelling, 2013, 6:1-16.