

Optimization Design of Passenger Flow in Rail Transit Station in Shanghai Based on Data Mining

Gang Chen

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China

Keywords: rail transit station, passenger flow, optimization, traffic load, anti-resistance

Abstract. According to a report about Shanghai Metro, Metro Line 2 traffic are increasing month by month trend, but capacity has become saturated. Subway passenger flow is a comprehensive passenger and passenger transfer station. In addition to increasing capacity, by ground transportation shunting the transfer passenger flow diverted to other lines of the site. Then large passenger flow of No. 2 line should be relieving the problem solution. By analyzing sample data about 10 working days from April 6, 2015 (Monday) to April 17 (Friday), it can be found that most of the passengers of Lujiazui Station are from the subway station in the northern part of Shanghai. This paper puts forward the optimization scheme -- a new semi-circle line. There are 7 sites from Shanghai West Railway Station of the 11th line to Lujiazui Station. of Line 2. The optimization scheme is compared with actual planning of Shanghai Metro. Construction of new semi-circle line has basic agreement with the Shanghai Metro Planning line 18 and line 21 in 2020-2025. From empirical point of view, this paper proves rationality of optimization scheme.

Introduction

In developed countries, the proportion of public transport (including rail transport) in the whole city is lower than that of the developed countries. Road traffic on the ground is almost no long time congestion. Private car penetration rate is higher, so residents tend to travel in their own car.^{[1][2]} Therefore, foreign research direction is about passengers ride public traffic behavior theory analysis, hoping to have conclusion of the study to keep passengers of public transport, improve the passengers of public transport loyalty and satisfaction^{[3][4][5]}.

According to a report about Shanghai Metro, Metro Line 2 traffic are increasing month by month trend, but capacity has become saturated. Subway passenger flow is a comprehensive passenger and passenger transfer station. In addition to increasing capacity, by ground transportation shunting the transfer passenger flow diverted to other lines of the site^[6]. Then large passenger flow of No. 2 line should be relieving the problem solution.

Data Preprocessing

The original data sample is all transaction records of Shanghai public transport card in April 2015. SPSS Statistics IBM 19 and SAS 9.4 are used to preprocess the raw data. Data samples show that working day of the largest inbound traffic over the weekend's largest inbound traffic. Therefore, working day of the subway operating pressure is greater. Research working day passenger flow optimization program has more practical significance. Therefore, we choice of April 6, 2015 (Monday) to April 17, 2015 (Friday) between the ten working days of Shanghai subway traffic data to analyze.

Filter out the Subway Card Transaction Data. Because the original data set contains a subway, bus, taxi and ferry all the use of official card transaction records. SPSS cannot screen Chinese characters. Therefore, through the variable re encoding selected types of consumer for "subway" data, will be "subway" was coded as 1, "bus" 2, "for rent" 3, "ferry" 4 and delete buses, taxis and ferries data, delete code for 2,3,4 data.

Filter out the Early Peak of the Data, and Remove Useless Variables. Peak passenger flow occurs in the morning from 7:00 to 9:00 and 17:00 to 19:00. This part of the passenger flow is commuting commuter passenger flow. Therefore, the problem of early peak for optimization, the evening peak of the problem can be solved with the same approach. So screen out the 7:00 to 9:00 of the subway card data, and delete a few useless variables.

Data Cleaning, Data Mining Before the Need to Remove Non Normal Data. A complete subway ride process, including the station and the station. Therefore, the normal data refers to the same transaction number in each day of the transaction records should be in pairs appear. If a bus card number only pit or only the station records, the record is not complete, it should be deleted. Otherwise it will have an impact on the results of data mining. In this paper, we use SAS to filter the number of the bus number which is even number.

Data Analysis

It needs to find a huge problem in the number of line 2 passenger stations, that is, the destination of the huge passenger flow. Therefore, it needs to screen out all the subway line 2 out of the station records.

Pick out the Station Records Data. Site on line 2 should be numbered. Such as Xu Jing station of No. 2 line encoding for 1, Hongqiao train station line 2 encoding 2. And so on. The amount of field for the record amount of 0 indicates that the passengers into the subway brake machine. Amount is not equal to 0 indicates that the passengers out of the subway gate machine. Select the amount of data does not equal to 0 of the data. It is the record of the station. In Metro Line 2 in the morning peak period the station passenger flow peak is compiled as a site for 14 -- Lujiazui Station and other passenger flow larger sites is the people's Square, Nanjing West Road, Zhangjiang hi tech park, Jing'an Temple. If destination is to optimize Lujiazui station passengers travel routes, then it will be able to reduce the operating pressure Metro Line 2. Lujiazui station is a special site. Although it is not a transfer station, but it is a set of office, business, shopping, tourism as one of the comprehensive site, so whether it is working days or holidays or weekends, it has a lot of traffic. People's Square, Nanjing West Road, Jingan Temple and Nanjing road and other sites is a transfer site, for the passenger flow and the original passenger flow is bound to result in the site's large passenger flow. If we can put forward a optimization scheme, the shunt to the Lujiazui Station passenger flow, change the number of passenger flow path, which is neither in line No. 2 site (such as the East Nanjing Road station, people's Square Station, Century Avenue Station) transfer and direct access to the Lujiazui Station, Metro Line 2 operating pressure can be reduced a lot.

Dig out the Passengers in Lujiazui Station as Destination. In this paper, we dig out number of passengers who leave off metro and find out where these passengers get on metro. Then we can speculate these passengers' ride route. We can propose optimization program according to these route.

Using SPSS, in data preprocessing file – subway we can found data about passengers who get off metro in the morning peak period in Metro Line 2, Lujiazui Station and stored in the file, Lujiazui Station. Then we use proc SQL statements and find amount for 0 characteristic which shows the passengers to get off in Lujiazui Station. We can find that the large flow of passengers who get off Lujiazui Station in Metro Line 2 can be divided into two categories. The first part of passengers is from other sites of metro line 2. The second part of passengers is from sites of other metro lines. Combined with Shanghai rail transit network map, analysis of the geographical characteristics of the second part of passengers who get off Lujiazui Station in Metro Line 2, we can find that as passengers in the northern region of Shanghai, if you want to get off Lujiazui station, you must take line 7 or line 1 to the center of the Shanghai, such as Jingan Temple and people's Square, in order to transfer line 2 to Lujiazui station. Because there is no transfer site as rail transit site in northern Shanghai passengers who want to go to Lujiazui station they have to go to Nanjing road station or people's Square station to transfer line 2 then go to Lujiazui station. These passengers not only increased the metro line 2 passenger traffic, but also increased the city's downtown transfer site traffic. From table 1, we can see

that passengers from other the transfer sites in the morning rush hour to Lujiazui Station of line 2 and passenger flow is quite large. If we can make this part passenger flow is not transfer in these sites, we can greatly reduce the passenger flow of transfer sites.

Table 1 Passenger from traffic station to Lujiazui station

station	Passenger from traffic station to Lujiazui station
Jiangsu Road	7084
Jingan Temple	11299
Longyang Road	3043
East Nanjing Road	13809
Nanjing West Road	2090
People square	36425
Century Avenue	22675
Zhongshan Park	1803

Optimization Scheme

In order to achieve the goal of dispersing passenger flow, the site of new subway line needs to make the length of the line to the Lujiazui station is shortest length of the line. It needs to consider the distance and passenger flow between each site to Lujiazui station. We take line 7 as an example. Site selected steps are as follows.

First, Calculate the distance between the site denoted as d_i ($i=1,2,3... 10$). For example: passenger flow from Changping Road station to Lujiazui Station is least. Its distance is 0 meters. Changshou Road station is 831 meters away from Changping road station, Zhenping Road station from the Changping Road station is 1846 meters. and so on. Calculate the percentage of passenger flow from stations to the station Lujiazui, recorded as p_i ($i=1,2,3... 10$). Results are shown in Table 2.

Table 2 station traffic and distance of Metro Line 7

station	Guest flow/ person	Percentage of passenger flow (PI)	Distance (DI) / M
Changping Road	740	0.054492	0
Changshou Road	1449	0.106701	831
Zhenping Road	1294	0.095287	1846
Langao Road	766	0.056406	3433
New Village Road	2038	0.150074	4277
Dahua 3 Road Station	2357	0.173564	5392
Xingzhi Road	1969	0.144993	6683
Field Road	775	0.057069	8952
Shanghai Univer station	1132	0.083358	12578
Gucun Park	1060	0.078056	16978

Set up new transfer station distance Changping Road station distance is d_x . Formula is as

$$\text{Min} \sum_{i=1}^{10} |d_x - d_i| \times p_i \quad (1)$$

$$s.t \ d_x \leq d_{10}$$

Using Excel programming solution, the process and results are shown in Table 3.


According to the result of the solver, the final choice on the 7th line station is from the starting point (Changping Road Station) 5392.99994 meters of the station. It is Dahua 3 Road Station.

Similarly we can have several stations of other lines. The final results are Table 4.

Subway line planning should avoid inward concave "V" line. Adjust the above site as Table 5.

Table 3 Program solving process and results

pi	di	dx	ABS(dx-di)*pi
5.45%	0		0.00
10.67%	831		88.67
9.53%	1846		175.90
5.64%	3433		193.64
15.01%	4277		641.86
17.36%	5392		935.86
14.50%	6683		968.99
5.71%	8952		510.88
8.34%	12578		1048.48
7.81%	16978		1325.23
		0	5889.514



pi	di	dx	ABS(dx-di)*pi
5.45%	0		293.82032
10.67%	831		486.66340
9.53%	1846		337.88836
5.64%	3433		110.50029
15.01%	4277		167.33210
17.36%	5392		0.00001
14.50%	6683		187.18550
5.71%	8952		203.16642
8.34%	12578		599.00973
7.81%	16978		904.35641
		5392.99994	3290.96

Table 4 Preliminary determination of new semi-circle transfer station

Line	Station	Line	Station
Line 1	Peng Pu Village	Line 8	Quyong Road
Line 3	Jiangwan town	Line 10	Jiangwan Stadium
Line 7	Dahua 3 Road Station	Line 11	Shanghai west railway station

Table 5 Determination of new semi-circle transfer station

Line	Station	Line	Station
Line 1	Wenshui Road	Line 8	Quyong Road Station
Line 3	Chifeng Road Station	Line 10	Post and Telecommunications Village Station
Line 7	Xingzhi Road Station	Line 11	Shanghai west railway station

Feasibility Demonstration of Expected Demand for Optimization

Demonstration Method. In this paper, the expected demand for the new semi-circle is transformed into the shortest path problem and solved by the Dijkstra algorithm. First of all, it is needed to demonstrate that the new semi-circle line meets the needs of passengers.

According to the principle of Dijkstra algorithm, if the sequence of {vs, v1,... , vt} is the shortest path from vs to vt, then the sequence {vs, v1,... , vt-1} must be the shortest path from vs to vt-1. We set up a comprehensive index called "traffic resistance ", denoted by Ri (i=1,2... N) as weights on path arcs. In the comprehensive consideration of the three factors of the travel time, travel costs and the degree of congestion, Ri formula is as follows:

$$R_i = (C_i + \lambda \cdot T_i) \cdot S_i \tag{2}$$

Ci is the cost. Its unit is yuan. λ is Time value, It can be available annual per capita disposable income of the average annual per capita working hours, Its unit is yuan / minute; Ti is travel time, Its unit is minutes. Si indicates the congestion coefficient of the train.

Demonstration Process. First, we should calculate the weights on each arc. That is, Ri. Ci can be calculated in accordance with the mileage valuation that is 3 yuan less than 6 km. It would be increase of 1 yuan per 10 k after 6 km. According to the Shanghai Municipal Bureau of statistics report shows that in 2015 the per capita disposable income of Shanghai is 49867 yuan, while the annual per capita working hours of 1950 hours. According to the calculated value of time value λ is 0.4262 yuan / minute Calculating formula is:

$$\lambda = (49867 / (1950 * 60)) = 0.4262. \tag{3}$$

Ti is ratio of distance and speed of the train between two sites.

Si is the ratio of the early peak of every metro line accounted for the total passenger flow. The sample data set is ten working days from April 6, 2015 to April 17, 2015, The Si of the new

semi-circle line is based on the average value of the passenger flow rate of all the lines involved in the new link. It is the average number of Line 1, line 2, line 3, line 7, line 8, line 10 and line 11.

Results are shown in Table 6.

Table 6 Percentage of passenger traffic (peak data from April 6, 2015 to April 17, 2015).

Line number	No.1	No.2	No.3	No.4	No.5	No.6	No.7
Percentage of passenger flow	14.7%	15.3%	9.4%	6.8%	1.9%	5.1%	8.6%
Line number	No.8	No.9	No.10	No.11	No.12	No.13	No.16
Percentage of passenger flow	10.4%	10.0%	6.5%	6.5%	2.1%	1.9%	0.9%

According to formula 2 weight value of each arc can be got in shortest path graph. Results are shown in Figure 1, table 7 and table 8.

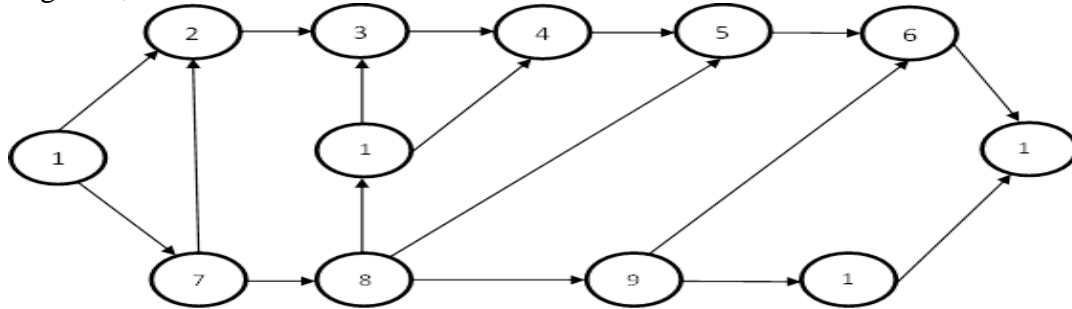


Fig. 1 abstract of the new semi-circle line

Table 7 number and corresponding station name

Site number	Station name	Site number	Station name
No.1	Lujiazui Station	No.7	Nanjing East Road Station
No.2	Post and Telecommunications Village Station	No.8	People's Square Station
No.3	Quyong Road Station	No.9	Jing'an Temple Station
No.4	Chifeng Road Station	No.10	Jiangsu Road
No.5	Wenshui Road	No.11	Hongkou Football Stadium Station
No.6	Xingzhi Road Station	No.12	shanghai west railway station

Table 8 weight of each station

Point	End Point	Weight Ri	Point	End Point	Weight Ri
No.1	No.2	0.527709	No.7	No.8	0.117375
No.1	No.7	0.635063	No.8	No.11	0.449733
No.2	No.3	0.0626	No.8	No.9	0.156501
No.3	No.4	0.065209	No.8	No.5	0.898817
No.4	No.5	0.317188	No.9	No.6	0.553328
No.5	No.6	0.169542	No.9	No.10	0.319282
No.6	No.12	0.202147	No.10	No.12	0.249327
No.7	No.2	0.162063	No.11	No.3	0.10638
			No.11	No.4	0.066104

By using Excel programming, the solution of the shortest path is obtained. The formula for the objective function is:

$$\text{SUMPRODUCT}(V284:V300, X284: X300) \tag{3}$$

The shortest path nodes from V1 to V12 are: V1, V2, V3, V4, V5, V6, and V12. This path is just like the new semi-circle line. From the Shanghai West Railway Station of 11th line to, Lujiazui Station of line 2, there are Xingzhi Road station of 7th line, wenshuilu stations of 1st line, Chifeng Road station of 3rd line, Quyong Road Station of line No. 8 and Telecommunications Village Station No. 10 Line. There are seven sites. If from Lujiazui station to other stations on this new semi-circle it would be found that you can save time, save money, and the car will not be so congested than from other stations in line 2 transfers to other subway lines t. So we can prove that the new line is to meet the needs of passengers. Therefore, the programmer proposed in this paper is feasible in theory.

Summary

Based on the results of deep mining on the data of Shanghai public transport card in April 2015, we found that the number of working days in the morning rush hour in Lujiazui stations biggest. Therefore, we selected Lujiazui station as the output site to study. We analyze the location and transfer station of station passenger in Lujiazui station, and give the optimization scheme of the new semi-circle line according to the results of the analysis. That is Xingzhi Road station on Route 7 lines, wenshuilu station on the 1st line, Chifeng Road station on the line 3, Quyang Road station on the 8th line and on Dian Xin Cun station on the 10th line. We prove the feasibility of the optimization scheme based on the expected demand, and the optimization scheme is compared with the actual planning of Shanghai metro, the optimization scheme of this paper has certain rationality.

Acknowledgement

Miss Qi-fei Lu made a great contribution to data processing as student guiding by professor Chen's.

References

- [1]Ka Kee Alfred Chu, “Two-year Worth of Smart Card Transaction Data – Extracting Longitudinal Observations for the Understanding of Travel Behaviour”, *Transportation Research Procedia*, 2015, Volume 11, P365~380
- [2] Takahiko Kusakabe, Yasuo Asakura, “Behavioural data mining of transit smart card data: A data fusion approach”, *Transportation Research Part C: Emerging Technologies*, September 2014, Volume 46, P179~191
- [3]Timothy Spurra, Alfred Chua, Robert Chapleaub, Daniel Piché, “A Smart Card Transaction “Travel Diary” to Assess the Accuracy of the Montréal Household Travel Survey”, *Transportation Research Procedia*, 2015, Volume 11, P350~364
- [4] Hiroaki Nishiuchia, Tomoyuki Todorokib, Yusuke Kishib, “A Fundamental Study on Evaluation of Public Transport Transfer Nodes by Data Envelop Analysis Approach Using Smart Card Data”, *Transportation Research Procedia*, 2015, Volume 6, P391~401
- [5] Chen Gang, Decision-making model of business continuity management. *Lecture Notes in Electrical Engineering*, 140, pp. 285–289, 2012.
- [6]Chen Gang, Risk evaluation of business continuity management by using green technology. *Communications in Computer and Information Science*, 113, pp. 86–92, 2010