Daily Estimation of Passenger Flow in Large and Complicated Urban Railway Network

Shuichi Myojo

Railway Technical Research Institute, Tokyo, Japan

Abstract

Railway passenger flow data including the on-board number of each train, the number of passengers getting on and off at each stop and the number of transferring people in each station can be used for planning facilities, train scheduling and other planning. In Japan, the data of Transportation Census of Urban Cities, renewed every five years, is available to grasp public transport passenger flow in detail. This study aims at estimating metropolitan railway passenger flow using OD (i.e. Origin-Destination) matrix data in order to analyze passenger flow everyday. Fortunately, many railway operators have incorporated automatic ticket gate systems into train stations in large cities, and these machines not only check tickets of passengers, but also store OD matrix data by counting the number of passengers for each OD pair at specific time period. However the OD matrix data do not give any information about the routes of passengers. Since passengers of one OD pair have enormous number of alternatives about their routes and trains in complicated urban railway network, it is not simple to determine their routes. This implies the need to develop route-choice model of passengers. We have developed a method of estimation of passenger flow using OD matrix data gathered from ticket gates in hundreds of stations and timetables of thousands of trains. By analyzing the estimation result, we can obtain the change curves of the numbers of passengers on each train, the number of passengers getting on and off at each stop, the number of transferring passengers in each station, and other relevant factors. These data can be used for short-term demand forecasting and train rescheduling.

1 Introduction

It is very important to reschedule train timetables properly so that they meet passengers’ demand when train traffic is disrupted. Therefore, it is necessary to realize a method to forecast passengers flow in detail after train traffic is disrupted as well as an advanced method to reschedule trains in short time. We expect that it is possible to realize traffic prediction of high hitting ratio by referring to the data of the past days that are similar to the day of prediction in their attributes such as month, day of the week, weather, and other factors. In order to find a similar day, it is necessary to obtain and store detailed passengers flow data everyday.

In Japan, the Ministry of Land, Infrastructure, and Transportation takes the Transportation Census of Urban Cities in order to perceive passenger flow of public transportation in three major metropolitan areas. This census can provide very detailed data of passenger flow because passengers corporate with the survey fill their choice of transportation mode on the census day. However, the census does not meet our needs because the Japanese government takes it only every five years.

On the other hand, rolling stock incorporated recently can measure total weight of on-board passengers and calculate congestion rate. This function will make it possible to obtain all the numbers of passengers on all trains. However, it cannot provide the number of on-board passengers for each destination, because it is impossible to grasp each passenger’s destination by measuring the total weight of passengers only.

We have developed a method to estimate passenger flow based on OD matrix data stored by ticket gates in stations. Although OD data do not provide directly any information about lines and trains whom passengers took, the method speculates their choice of paths including their lines and trains. By analyzing estimated result, it is possible to obtain the number of passengers aboard each train for each destination, the number of passengers boarding and disembarking at each station, the number of transfer passengers at each station and other relevant factors. In addition, the estimated numbers of on-board passengers with a proto-type system that we have developed based on this method have a strong correlation with the number reported by train conductors.
2 OD data during each period

Ticket gate systems utilized by major railway operators in Japan count the number of passengers of each OD (Origin & Destination) pair in each partitioned period of one day.

![Diagram](A station -- B station -- C station -- D station)

<table>
<thead>
<tr>
<th>Origin</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50</td>
<td>30</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>60</td>
<td>40</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>40</td>
<td>30</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>50</td>
<td>60</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: An example of OD data.

OD data has the following advantages:
- It is possible to obtain the data on any date because gate machines work everyday.
- The data is almost complete enumeration because gate machines count almost all passengers.

In spite of such advantages, OD data has a disadvantage that it does not provide directly any information about lines and trains whom passengers took.

3 Proposed method to estimate passenger flow

3.1 Basic idea of the method

The proposed method estimates passenger flow with following steps.
1. Select a pair of OD from OD data gathered from ticket gates.
2. Search all paths that those who travel the OD selected in procedure (1).
3. Calculate the number of passengers on the paths searched in procedure (2) during each period.
4. Store the number of passengers calculated in procedure (3).
5. Repeat procedures (1), (2), (3) and (4) until all pairs have been finished.
6. Calculate the total number of passengers that pass each link.

Figure 2 shows all procedures mentioned above.

![Diagram](1) Select one OD pair
   (2) Search all paths that passengers of the OD pair might take
   (3) Calculate the number of passengers on each path searched in the above procedure
   (4) Store the number of passengers on each path
   Finish all OD pairs
   (6) Calculate the number of passengers on each link

Figure 2: Procedures to estimate passenger flow.
3.2 Procedure to search paths

As described in Chapter 2, it is difficult to identify the trains that passengers of each OD pair took. Therefore, by referring to train timetables, the proposed method searches paths that passengers could take. In this method, we define a path as a series of trains that passengers may take to get to their destinations. In other words, a path taken by one train is not the same as the paths taken by other trains though they are the same railway line(s). For example, eight paths shown by thick arrows in Figure 3 start from (A) station and arrive at (D) station between 15:00 and 15:30.

![Figure 3: An example of paths](image)

3.3 Calculation of the number of passengers on the searched paths

In this procedure of the estimation method, the number of passengers of each period of the selected OD pair is distributed among searched paths, which have been picked up by the search procedure described in Section 3.2. If the data shown in Figure 1 is in the period from 15:00 to 15:30, the 40 passengers who departed (A) station and arrived at (D) station might choose a path from the seven alternatives shown by thick arrows in Figure 3. The method allocates the 40 passengers among the seven paths considering the attribute of each path.

3.4 Calculation of the number on each link

In the method, a link represents a movement of passengers such as travelling from a station to the next stop on a train, transferring, or other actions. A path consists of one link or more, and some links are contained by one path or more. In Figure 4, two paths from (A) station to (D) station, shown by dot-lined arrows, share a link shown by the thick line. In addition, other paths of other OD pairs also contain the same link. By summing up the number of passengers of all paths that share the link, it is possible to obtain the number of passengers on the link. The estimation method repeats this procedure on all links. That is, it is possible to grasp the number of passengers on each train at any location. Moreover, it is possible to categorize on-board passengers by their destination.
4 Various data obtained from the estimated result

4.1 A proto-type system

Based on the method, we have developed the simplest proto-type system that searches shorter time paths and allocates the number of passengers equally. We have applied this system to estimate passenger flow in a commercial railway network in the following condition:

- Number of stations: 392
- Number of trains: 4,785
- Number of OD pairs: 30,604

The system can estimate within an hour with PC whose CPU is 3.6GHz Pentium® 4 processor.

4.2 Number of on-board passengers

The result of estimation provides the change of number of on-board passengers of the train. Figure 5 shows an example of a histogram of the estimated number of passengers of a commercial train. The shaded bars represent the number of passengers on board while the train is running, and white bars represent the number of passengers while it is at stops.

![Figure 5: Numbers of on-board passengers.](image-url)
4.3 Number of passengers boarding and alighting at each stop

By comparing the number of passengers before a stop and at the stop, it is possible to obtain the number of passengers who alighted at the stop. By comparing that before departure and after departure, it is also possible to obtain the number of passengers who boarded at the stop. Moreover, it is possible to categorize the passengers who alighted at a stop into two groups (i.e. group of passengers who transferred to other trains and group of passenger for whom the stop was their final destination). It is also possible to categorize passengers who boarded.

Figure 6 shows an example of a train. Bars of positive value mean the number of passengers who boarded train, and those of negative value the number of alighting passengers. It is very difficult to obtain such data by only counting the number of on-board passengers.

4.4 Number of on-board passengers for each destination

We can also categorize the passengers based on the difference of their destinations as is shown in Figure 7.

![Figure 6: Number of passengers boarding and disembarking at each stop.](image)

![Figure 7: Number of on-board passengers for each destination.](image)
Figure 7 shows the fact that most of passengers get off the train at (J) station. Therefore, the railway operator might consider using two trains with different capacity instead of a single train. Moreover, such a figure may help to find out such a train that should run longer.

5 Verification of estimated result

By comparing the reported number of passengers by train conductors, we have verified the number of on-board passengers estimated by the proto-type system. Train conductors of nearly a half of all trains report the maximum number aboard between two or more stations by visual counting. Figure 8 and 9 show the comparative results. Figure 8 shows the result of comparison on a line where only local trains run. Figure 9 shows that on another line where rapid trains make connections with local trains. On both lines, especially on the line of local trains only, correlation coefficient is high.

6 Conclusion

It is possible to estimate the passenger flow and to obtain detailed data containing the number of boarding passengers and disembarking passengers, the number of on-board passengers for each destination using OD data during each period stored in ticket gates. Moreover, the number of on-board passengers estimated by the simplest proto-type system has a very close correlation with reported number of passengers by train conductors. We believe that this method will enable to realize daily estimation of passenger flow in large and complicated urban railway network.

References