A Study on Passengers’ Train Choice Model in Urban Railways

Noriko FUKASAWA  
Munenori SHIBATA
Transport Planning and Marketing Laboratory, Signalling and Transport Information Technology Division

This study aims to develop a model of passenger behavior when choosing trains from among various types of train service such as local and rapid services in urban railways, in order to evaluate passenger flows. This paper describes an online survey method for obtaining passenger activity data using departure information display images. A disaggregate demand model was developed for train choice behavior based on the survey data which were collected using the method. The developed model is put into practice by means of an application named “Train-Choice Model Viewer”, which can calculate the time series data of passengers’ volume at stations and in trains. It makes it possible to simulate the congestion level on each train and in each station when the time schedule is input.

Keywords: train choice behavior, urban railways, passengers’ flow, departure information displays

1. Introduction

1.1 Background and objectives of this study

Urban railway systems provide various types of train services (such as rapid, special rapid and local services) on a single line to shorten the travel time of selected trains by making trains pass some stations without stopping. Thanks to such services, passengers can choose among different types of train depending on the purpose of travel and preferences, namely, rapid or special rapid trains if they want to reach their destinations more quickly, and local trains if they prefer to reach their destinations without changing trains. At present, railway companies do not ascertain in much detail which trains passengers want to choose for each purpose, except in the case of reserved-seat passengers. In order to provide smoother passenger transport services, however, it is important to identify the factors that affect passenger choice of trains, and to ascertain accurately passenger behavior in the form of boarding rates of each train.

With the development of the transport network, numerous studies have attempted to model passenger choice of route, in order to forecast passenger demand when building new rail lines, and to estimate distribution of passengers on competitive lines [1]. However, there have been few studies on modeling passenger behavior in choosing which trains to use on the same line. Accordingly, this study attempts to identify the factors affecting the decision-making process involved in choosing a train and to develop a model of passenger behavior when selecting trains, using passenger behavior data about which kinds of passenger chose which types of trains, and under which circumstances.

A major railway line in a major urban area was selected to conduct a case study. This rail line has double tracks in each direction. Three types of train running at various speeds use the line, i.e. type A, B and C. The type A is the fastest train which has the least stops among all types. The type C is the slowest train which stops all of the stations in the line. The type B is the 2nd fastest train which has the 2nd most stops. It has a complex train operation schedule featuring numerous train types and frequent arrivals and departures. For this reason, it is thought that identifying the factors that affect passenger train selection behavior and ascertaining the actual conditions of train choice behavior for such a railway line will facilitate smoother urban transport services. Since the previous study [2] already developed a model for route selection with regard to passenger behavior involving detours and waiting times when train schedules are disrupted, this type of passenger behavior will not be considered in this study. Instead, this paper concentrates on passenger behavior when trains run on time or with a maximum delay of 15 minutes.

1.2 Considerations for modeling train choice

In general, demand analysis and demand forecasting for urban transport is achieved through application of disaggregate demand models such as disaggregate logit models. A disaggregate demand forecasting method is based on random utility theory, and is a method for mathematically modeling passenger choice behavior based on the assumption that passengers recognize the multiple available choices as alternatives and rationally choose routes and means of transportation through comparison and respective service levels, including time required and fares. Since it is thought that when a passenger wants to use rail transportation, he or she will choose an faster train or a slower train by comparing the service level of each, and decide the route, namely their preferred train, train choice was considered in this study to be an issue of route selection. Therefore for this study it was decided that a disaggregate logit model would be applied. Aggregating individual behavior is a central part of demand forecasting when using a disaggregate demand model. Consequently in order to estimate passenger flows using the model, data was supplemented by estimating total numbers of passengers on the route according to forecasts made on the basis of the number of passengers passing between stations in each time period obtained from automatic ticket checking data (OD data).
2. Survey of the actual circumstances of passenger train selection

2.1 Factors affecting train selection

Choosing the factors that define the utility of each type of train is one important process in modeling train selection. When treating the choice of multiple train types on a single line as a problem of route selection, one of the generally important explanatory variables in route selection, namely cost or rail fare, remains unchanged, and other appropriate explanatory variables for train selection have to be extracted. A survey was therefore conducted to identify other factors affecting train choice.

A revealed preference (RP) survey was conducted among passengers who regularly use the major railway line featured in this study to gather practical insight about which trains they use. The questionnaire asked participants about the type of train they had taken most recently and the reason for choosing that type of train. Table 1 gives an outline of the survey.

Figure 1 summarizes the results of which trains were actually taken and the factors influencing that choice for respondents who were able to choose any train type. The results show that a high percentage of respondents chose the answer “Agree” for reasons related to arrival time, journey time, number of stops, need to change trains, and congestion on board the train: “Wanted to arrive quickly,” “Few stops,” “Shorter journey time,” “Can reach destination without changing trains,” and “Train not crowded.” On the other hand, reasons related to the train itself and facilities on board, such as “Comfortable seating,” “Comfortable ride,” “Women’s-only car available,” and “Restrooms available,” were not considered very important. Based on these results, the following factors were adopted to explain the utility function of type of train; time-related factors such as arrival time, journey time, and waiting time, and factors relating to inconvenience of rail travel such as number of stops and number of train changes, and congestion levels.

2.2 Method of obtaining train choice data using departure information display images

Data about people’s movements is often obtained using person-trip surveys such as urban transportation censuses. Most surveys of this kind ask questions about transportation choices on a particular day: what kind of people travel between which places, for what purpose and with which means of transport. When developing a disaggregate demand model such as a disaggregate logit model, however, it is preferable to obtain data not only about choices actually made (i.e., actual routes, means of transportation, and trains used) but also about the alternatives not selected (i.e. routes, transport means, and trains that are available but not used). For the present study in particular, i.e. choice of a single train from among multiple options, the available alternatives will vary with the time of train arrival, and choices may differ due to minor differences in circumstances, such as the waiting time until the next couple of trains and the stations at which each of those trains will stop. Detailed data about these surrounding circumstances therefore needs to be collected to understand what could have conditioned the choice made, in addition to trip data relating to the choice ultimately made.

Accordingly, respondents were asked to photograph the departure information displays at the ticket gate or on
the platform when they arrived at their departure station, allowing trip data to be collected. The trains shown on the departure information displays represent the choices available to passengers at that time, and the different train types shown are the service levels that are to be compared. It is possible to ascertain simultaneously both actual choices made and the service level of each other alternative by surveying the factors available to passengers such as the arrival time and number of transfers up to their destination, which trains they chose, and which trains they did not choose. It has been highlighted however that RP surveys on past experience of subjects give rise to issues of measurability when asking about the past, since in some cases the subjects have only vague recollections of service levels, particularly for the options not chosen. This survey method therefore can remedy these issues because of the captured images of the conditions in which the choices were made. Past surveys also suffered from inconsistent information about services, since they relied passenger input, whereas this new approach which uses images of the train information displays makes it possible to corroborate what respondents say in the survey, increasing the reliability of the trip data collected.

This method can be considered to be a new method for surveying personal railway-trip data which optimizes the advantages of on-line surveys, which have become common because of advancing information and communications technology. The advantages include the spread of mobile and smartphones equipped with cameras, which have made it easy to take photos at anytime and anywhere and to easily send and receive image files over the Internet.

2.3 Survey of actual conditions of train choice

An on-line RP survey was conducted for this study of the actual travelling practices of passengers who used the line. Participants were asked about all the trains they took, including departure and destination stations, the last time they had used the line (in the last week or so), the reasons why they chose the trains they did, and the alternative trains they could have taken but didn’t. Since the objective of this survey was to ascertain how people actually decided which train to take, it was limited to users of stations and times during the day when all three travel options were available. Table 2 summarizes the survey.

In this survey, as shown in Fig. 2, the respondents were asked to upload images of the departure information displays and data was collected relating to passenger cognition about what respondents themselves thought or predicted when choosing their train, in terms of order of arrival at the destination station, differences in time of ar-

Table 2 Survey outline of passengers’ behavior for train choice

| Survey period       | From February 25 to March 11, 2011 |
|---------------------|-----------------------------------|
| Survey respondents  | Users of a major rail line in a metropolitan area (at least once a week) |
| Survey method       | On-line survey                     |
| Number of respondents | Total 775                         |

Fig. 2 Train choice survey using pictures of departure information display
rival, level of congestion, and stations where they would get off the train if they took it, not only for the train they actually rode, but also all trains shown on the departure information displays.

2.4 Results of survey on actual conditions affecting train choice

Figure 3 summarizes the results of the train choice survey. The number of passengers against their boarding times show largely the same distribution as that ascertained previously using automatic ticket checking data and other sources; however, since the subjects of the survey were limited to station users and periods during which all train types were available, the passengers in the early morning and late night were excluded. 74% of passengers took the train either to commute to and from school or work, or for some other non-leisure related business. Those in a hurry accounted for 61% of passengers, while the remaining 39% of passengers were not pressed for time when taking the train. A look at distances reveals that 64% of passengers travelled less than 30 km, while 58% of passengers took trains to travel sections in which the fastest train stopped twice at most. It was also clear that passengers who took a single trains, without changing, accounted for 84% of the subjects, while less than 1% of the passengers had more than two changes. Furthermore, the trains chosen at the departure station were distributed fairly evenly among the three types. Asked about reasons for choosing the trains, a high percentage of respondents gave time-related reasons, namely “Arriving at just the right time” and “Shorter journey time,” as well as the reason “No need to change trains.” Furthermore, 40% of respondents favored the statement “Train not crowded.”

3. Developing a train choice model

3.1 Applying a disaggregate logit model

Although all three types of trains were being operated on the line being investigated for this study, for simplicity a route selection model was devised in which only one route was chosen out of a possible two: one on an faster train and the other on a slower train. The definition of faster train and slower train determined by distance travelled as follows: When traveling a short distance, a faster train was defined as the type A trains and type B trains, while a slower train was defined as the type C trains. When traveling a long distance, a faster train was defined as the type A trains, and a slower train was defined as type B trains and type C trains. This is because when traveling a short distance there is little difference between the type A train and type B train in terms of transit time and number of stops, while when traveling a long distance there are considerable differences in transit time and other properties between the type A and type B train, and because very few passengers chose to ride type C trains over long distances. We defined a long distance and a short distance by number of stops of the type A train; in case of two or fewer stops, defined as a short distance and in case of having three or more stops defined as a long distance. A sample of 421 subjects meeting the above conditions was taken and used to construct a model.

For the explanatory variable based on number of transfers, an estimate number of transfers was derived from individual response data collected at origin and destination stations together with answers gathered through the form shown in Fig. 2 to the question about where they would have alighted, were they to have taken a particular train. Train delays were found by comparing existing data on scheduled times and expected arrival time.

As noted in Chapter 1, a disaggregate logit model was applied to develop a train choice model. Since this model was used to model individual choice behavior mathematically based on random utility theory, the probability $P_i$ that an individual would choose the choice 1 out of the two choices 1 and 2 was defined in the manner shown in (1), and (2) shows the utility function $V_m$ for the route $m$.

$$P_i = \frac{e^{V_i}}{e^{V_i} + e^{V_j}}$$  \hspace{1cm} (1)
\[ V_n = \sum_k \beta_{kn} \cdot X_{kn} + \alpha_n \]  

(2)

where \( X_{kn} \): explanatory variable for item k, \( \alpha_n \): constant term, \( \beta_{kn} \): individual parameters,

3.2 Results of estimating parameters

Table 3 shows the results of explanatory variables and parameters to compare faster trains and slower trains estimated using the maximum likelihood method. The likelihood ratio, an indicator of the degree to which a model fits a set of data, is 0.358 for the short-distance model and 0.522 for the long-distance model. Since both exceed 0.2, indicating a good likelihood ratio, the models can be considered precise enough. Explanatory variables not chosen because their parameters were not statistically significant, are indicated with dashes ("-"). With respect to the effect on train choice, regardless of whether the distance travelled was long or short, total required time and congestion of the trains had statistically significant negative effects, while number of transfers had a negative effect only for short distances but no obvious effect for long distances. Another statistically significant negative effect, platform congestion for short distances, was also identified. While selected by many passengers in the survey, the effect of other factors such as number of stops, ride time, waiting time, travel purpose and level of time pressure, could not be identified.

The accuracy of the model was tested using 303 RP survey data responses regarding train choice collected a year later for the same line. The results were relatively strong, 71.0 % for short distances and 77.8 % for long distances. This indicates that the developed model remains applicable over time. The survey used to collect the data is summarized in Table 4.

| Table 4 Outline of survey used to collect verification data for train choice model |
|----------------------------------------|------------------|------------------|
| Survey period | From February 13 to 21, 2012 |
| Survey respondents | Users of a major rail line in a metropolitan area (at least once a week) |
| Survey method | On-line survey |
| Number of respondents | Total 303 |

4. Estimation of number of train passengers and number of passengers waiting in the station using the train choice model

Changing the operating time of a single train by one or two minutes in an urban area where timetables run a tight schedule, significantly influences the level of congestion on a line. For this reason, when revising train schedules it is

| Table 3 Estimation results |
|-----------------------------|
| Utility function: \( V \) |
| | Short distance | Long distance |
| | parameter: \( \beta \) (t-value) | parameter: \( \beta \) (t-value) |
| Total required time (min) | -0.180 (-5.540***) | -0.185 (-4.955***) |
| Total ride time (min) | - | - |
| Total waiting time (min) | - | - |
| Waiting time at origin station (min) | - | - |
| Number of stops | - | - |
| Number of transfers | -1.968 (-3.364***) | - |
| Congestion level of train \(^1\) | -1.121 (-5.519***) | -1.417 (-4.862***) |
| Congestion level of platform \(^2\) | -0.364 (-3.463***) | -0.223 (-1.537) |
| Age | - | - |
| Sex | - | - |
| Purpose of travel \(^3\) | - | - |
| Time pressure \(^4\) | - | - |
| Outbound / inbound \(^5\) | -0.571 (-1.753*) | -0.883 (-1.842*) |
| Likelihood ratio | 0.358 | 0.522 |
| Hit ratio | 78.9% | 84.8% |

***: p<0.01, **: p<0.05, *: p<0.1

\(^1\): Congestion level of train (1: empty / 2: vacant seats / 3: somewhat crowded, 4: crowded)
\(^2\): Congestion level of platform (1: some passengers waiting for a train, 2: a number of queues of passengers waiting for a train, 3: many queues of passengers at all train door positions, 4: passengers crowded on platform)
\(^3\): Purpose of travel (1: business / 0: private)
\(^4\): Time pressure (1: very pressed for time, 2: pressed for time, 3: somewhat pressed for time, 4: not in a hurry)
\(^5\): Outbound / inbound: 1: outward leg, 0: inward leg
necessary to ensure that proposals match actual passenger demand. Proposed revisions need to be prepared with great care. Consequently, an application was developed entitled “Train-Choice Model Viewer” capable of simultaneously evaluating train and station congestion, as a means to employ the train choice model developed in this study. This application calculates time series of passenger numbers at all the stations and on all the trains on the subject line, based on train schedule data and the developed train choice model. It was assumed that passenger volumes passing between stations in each time period were obtained from the automatic ticket checking data (OD data) referred to in Chapter 1, and that passengers would show the same decision-making behavior. This makes it possible to overcome the issue of aggregation in a disaggregate model.

More specifically, by using OD data for each time period, it was possible to calculate the probability that an individual passenger at a specific station would take a faster or a slower train, based on the utility of each train to that passenger, as evaluated with the train choice model. Figure 4 shows the results output screen. The trains on the simplified track layout diagram, color-coded by their types, run according to schedule at the times shown at the top of the screen. The horizontal rectangular figure above each train shows its level of congestion. When a train arrives at a station, its level of congestion changes in accordance with the number of passengers getting on or off at that station. Also, the vertical bar graphs below the track layout diagram indicate the number of passengers waiting at each station. As time passes, the number of passengers waiting increases in accordance with OD data. When a train departs, this number decreases by the number of passengers found to have boarded the train in accordance with the probability of choosing that train calculated using the train choice model.

In addition, time-series data for the number of passengers boarding each train and the number of passengers waiting at each station can be exported to an external file. Figure 5 shows an example of the time-series being calculated for the number of passengers waiting at a station. In this example, while the number peaks slightly during the morning rush hour, greater congestion occurs at 6:00 pm and after. 7:00 pm is the most congested time, and the station remains congested until the last train of the night. Figure 6 shows an example of the number of passengers on a train between the respective stations. This train ran during the evening rush hour, with numerous passengers boarding at Station S, a central station in an urban area. While the number of passengers gradually decreased after leaving Station S, congestion increased again after reaching Station Y. In this way, the Train-Choice Model Viewer can be used to ascertain the characteristics of congestion at individual stations and on individual trains.

This Train-Choice Model Viewer also can be used to
project in advance in which trains, at what stations, and at what times extreme congestion can be expected, when inputting a new train schedule. In addition, it can be used to forecast how a delay of several minutes may affect congestion in stations and on trains, by inputting actual schedule data.

5. Conclusion

This study models passenger decision-making behavior when choosing between trains, such as rapid and local trains, running on an urban network, a topic which to date has only rarely been investigated. A new method was devised in the course of this research, for collecting data regarding train choice, employing photos of departure information displays taken by subjects when they arrive at their station, to ascertain the actual circumstances in which passenger choose which train to take. This is a new method for surveying personal rail-trip data which leverages the advantages of on-line surveys, which in turn have become increasingly common thanks to progress in information and communications technology. It makes it possible to collect highly reliable data about the choices made and about why other options were disregarded. Using the data collected through this method, a train choice model was designed, applying a disaggregate logit model, for passengers travelling both short and long distances. This made it possible to generate a likelihood ratio, an indicator for the model, which showed a sufficiently good fit. High accuracy of 78.9% and 84.8% was achieved, and applicability of the model over time was also verified by comparing results obtained from data collected about a year later.

Due to the nature of the data collection method employed in this study, the models do not include data about such passengers waiting for trains which do not appear on the departure information displays. The study is scheduled to continue, and data will be collected to further refine the models by incorporating passengers with differing behavioral patterns, and be able to ascertain collective flows of rail passengers.

References

[1] Shibata, M., Okuda, D., and Suzuki, T., “Demand Estimation Model of Inter-Regional Railway Passengers Considering Multi-Mode Route Alternatives,” RTRI Report, Vol. 28, No. 4, pp. 47-52, 2014 (in Japanese).
[2] Muto, M., "Development of a Method to Predict Passenger Numbers upon Resumption of Train," RTRI Report, Vol. 22, No. 6, pp. 17-22, 2008 (in Japanese).

Authors

Noriko FUKASAWA
Senior Chief Researcher, Laboratory Head, Transport Planning and Marketing Laboratory, Signalling and Transport Information Technology Division
Research Areas: Travel Behavior Analysis, Travel Data Utilization

Munenori SHIBATA, Dr. Eng.
Senior Researcher, Transport Planning and Marketing Laboratory, Signalling and Transport Information Technology Division
Research Areas: Travel Demand Forecast