

Behavioural analysis of railway passengers using smart card data

Y. Asakura¹, T. Iryo¹, Y. Nakajima¹, T. Kusakabe¹, Y. Takagi¹
& M. Kashiwadani²

¹*Department of Civil Engineering, Kobe University, Japan*

²*Department of Civil & Environmental Engineering, Ehime University, Japan*

Abstract

Smart card systems have been spreading globally not only for purchasing general goods and services, but also for fare payment of public transport. When an individual traveller is holding a public transport smart card, the precise times of his/her passing through ticket gates are recorded in the smart card system. The longitudinal smart card data of a large number of passengers have been stored by a public transport company. The aims of this paper are to analyze the behaviour of railway passengers using the smart card data, and to evaluate the effects of train operation policies. The analysis is focused on the before and after comparison of travel choice behaviour of passengers when the railway company changed the train timetable.

The individual passenger's passing times at the entrance and exit ticket gates of origin and destination stations were aggregated for a small discrete time interval in a day. Through analyzing the departure and arrival time distribution and the travel time distribution, it was found that the passengers smoothly adjusted their travel behaviour for the improved train timetable. When the passing times of ticket gates of both origin and destination stations are combined with the train timetable, it becomes possible to identify the train that each traveller was likely on board. Through the before and after comparison, it is shown how the railway passengers have changed their train choosing behaviour for the improved time schedules.

Keywords: smart card, public transport, travel behaviour analysis, travel time savings, passenger demand.



1 Introduction

The railway smart card system was first introduced in Japan in 2000, and has been rapidly expanding in Japan Railway (JR) corporations and other private railway companies. They are, for example, Suica of JR East, ICOCA of JR West and PiTaPa of private railway companies in the Kansai area. The passing time of a smart card holder is recorded every one minute when he/she goes through a ticket gate. The historical record of the railway smart card can provide an accurate large size data set for passengers' movement. For example, the demand fluctuation of railway users can be analyzed for peak-hours and off-peak hours, and the longitudinal trend of user behaviour could be identified.

Those analytical results will be quite useful for railway companies when they recognize the longitudinal and spatial demand characteristics. It also becomes possible to evaluate the detailed effects of transport policies on user's behaviour. For example, the smart card data could be used to find out how the introduction of a new train diagram changes user's behaviour. The data are also attractive for academicians in the field of travel behaviour and transport science as the methodologies for analyzing the longitudinal travel data analysis could be proposed and validated through empirical analysis.

There were a few publications about the utilization of the smart card data for demand analysis, management and planning of public transport. Amongst them, Lehtonen et al. [3] discussed the possibility of using the smart card payment system data in public transport planning and statistics by using the data in Finland, which was one of the leading countries using smart cards. Bagchi and White [1] also considered the potential role of smart card data for travel behaviour analysis. Through the case studies in Southport, Britain, they mentioned the advantages and limitations of the smart card data for public transport analysis. Utsunomiya et al. [7] suggested potential uses of smart card data for transit planning. They studied the frequency and consistency of daily travel patterns and variability of smart card customer behaviours by using 'Chicago data'.

Although public transport smart card transaction data provide continuous and detailed travel information for demand modelling, certain aspects of information are incomplete. Chu and Chapleau [2] recently presented enrichment methods of archived smart card data. They estimated the arrival time of bus runs and identified linked trips using spatial-temporal concepts. By using the smart card automated fare collection system data, Trepanier et al. [6] presented a model to estimate the destination location for each individual boarding a bus with a smart card. Similar attempts were made by Sahin and Altun [5]. They proposed a method to extract ridership and operation information for better transit management by using the data in Istanbul, Turkey. Morency et al. [4] focused on the measurement of transit use variability. They used Canadian transit network data from 277 consecutive days and applied data mining techniques for measuring spatial and temporal variability of transit use for different types of cards.



The aims of this paper are to analyze the behaviour of railway passengers using the smart card data, and to evaluate the effects of train operation policies. The analysis is focused on the before and after comparison of travel choice behaviour of passengers when the railway company has changed the train timetable.

2 Smart card data

Two railway companies in Japan provided the smart card data containing the passing time of a ticket gate of each smart card holder. The number of the smart card holders is approximately 10% or more of the total passengers. The identification information of an individual user is anonymously processed and the privacy of the smart card holder is strictly protected throughout the analysis. The items in one record of the smart card are as follows; year/month/day, the name of departure station, the time (passing time of an entrance ticket gate) of the departure station, the name of arrival station, the time (passing time of an exit ticket gate) of the arrival station.

The two railway companies changed their train diagram in 2006. The data for analysis covers 24 weekdays, 12 days before, and after the diagram change. The weekend data were not analyzed as there were special events almost every weekend, and that might have some effects on passenger's behaviour. The number of record was about 122,000 in a day, and 730 M byte for 24 days. As the size of the data set was very huge, Relational Database Management System was applied and SQL programme is used for matching and aggregation process.

3 Methodologies

Fig. 1 shows a conceptual map of the gate passing times of a passenger and the train schedule. This map is used for the fundamental aggregation of the smart card data and the estimation of the passengers' flow. The horizontal and the vertical axis of the figure indicate the times in a day for the arrival (destination) and the departure (origin) stations, respectively. A point denoted by *Passenger* indicates the gate passing times of the origin and destination stations of an individual passenger. The train departure time of the origin station immediately after the gate passing time of the individual is denoted by a horizontal solid line in the figure. This is the most probable train that the passenger is on board. The train arrival time of the destination station just before the exit gate passing time of the passenger is also shown by a vertical solid line.

The on-board travel time is simply defined as the difference between the train departure and arrival times. The difference between the entrance gate passing time and the train departure time at the origin station is the waiting time, and that of the exit gate passing time and the train arrival time at the destination station is the exit loss time. The difference of the gate passing times between the origin and the destination stations is equal to the total travel time that is the sum of the waiting time, the on-board travel time and the loss time.



The following sections describe the arrival time distribution of the passengers at an origin station, the total travel time distribution between origin and destination stations, the overlapping analysis of the passengers' gate passing times and the train timetable. Through the before and after comparison of the improvement of train timetable, it is shown how the railway passengers have adjusted their train choosing behaviour.

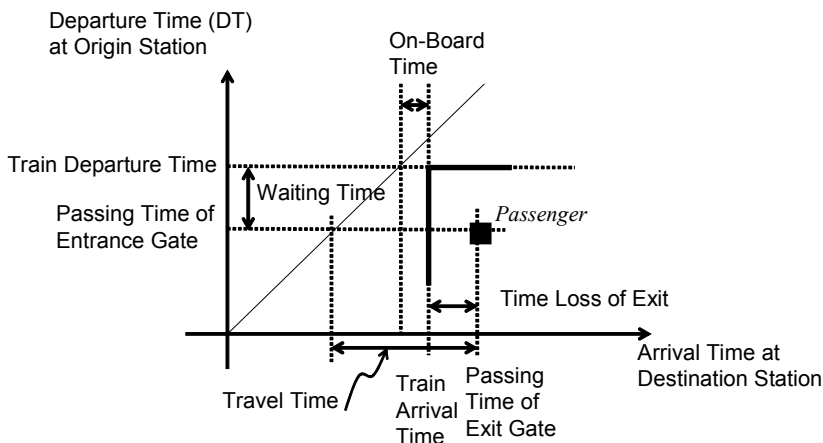


Figure 1: Gate passing times and train timetable.

3.1 Arrival time distribution

The number of passengers arriving at an origin station is aggregated every one minute, and divided by the total number of passengers of the day. The time interval is very short and this distribution shows the time-by-time dynamics of how passengers arrived at the station considering the train timetable.

3.2 Travel time distribution

The total travel time of a passenger is calculated and aggregated between a pair of origin and destination (OD) stations. This shows the OD travel time distribution. Before and after comparison of the timetable improvement, the travel time savings between the OD pair can be estimated.

3.3 Overlapping analysis

As shown in the Fig. 1, the passengers' gate passing times of the same OD pair are plotted with the train departure and arrival times. A passenger who enters the origin station before a train departure time can get on the train. Also, he/she can exit the destination station when he/she takes the train that arrives before the exit gate passing time. A passenger in Fig.1 could get on the train(s) if he/she was plotted in the lower right area surrounded by two solid lines.

Gate passing times of a large number of passengers of the same OD pair are plotted in the same diagram overlapped with the train timetable. The diagram could be separated into multiple rectangular areas that indicate the combination of train(s) departure and arrival times. The number of dots plotted in each area shows the possible number of passengers who may use the combination of train(s). For example, those who were plotted in the area A in Fig. 2 might get on train 2 at the origin station and arrive at the destination station by the same train. Those who were in the area B might get on train 1 and change train 3 on the way to the destination. However, the passengers in the area C have the two possibilities of the train combinations; train 2 only or train 1 and 2. Details of a rule based identification method are now under development and will be presented elsewhere.

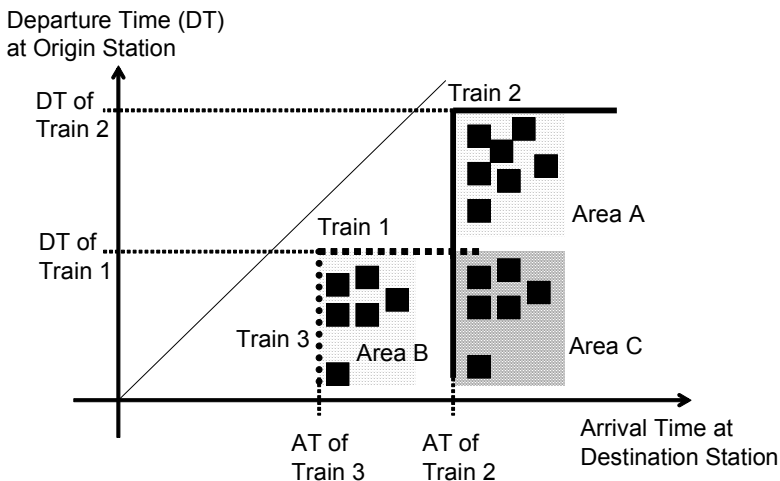


Figure 2: Possible trains on board.

3.4 Effects of timetable improvement

The effects of timetable improvement can be found through before and after comparisons of the overlapping map. If the bunch of passengers moves towards upper left after the timetable improvement, this implies that those passengers achieve late departure and early arrival. The comparison of the overlapping diagrams before and after the timetable improvement will provide useful information to know the effects of the timetable improvement. This could be discussed for an individual passenger. If a passenger plotted at a point denoted by a black square in the Fig. 3 moves to the hatched area 1, he/she obtains the benefit from the timetable improvement. Even if he/she moves to the hatched area 2, this passenger saves travel time by earlier departure. However, if he/she is obliged to move area 3, this passenger has to arrive at the destination lately if he/she departs earlier.

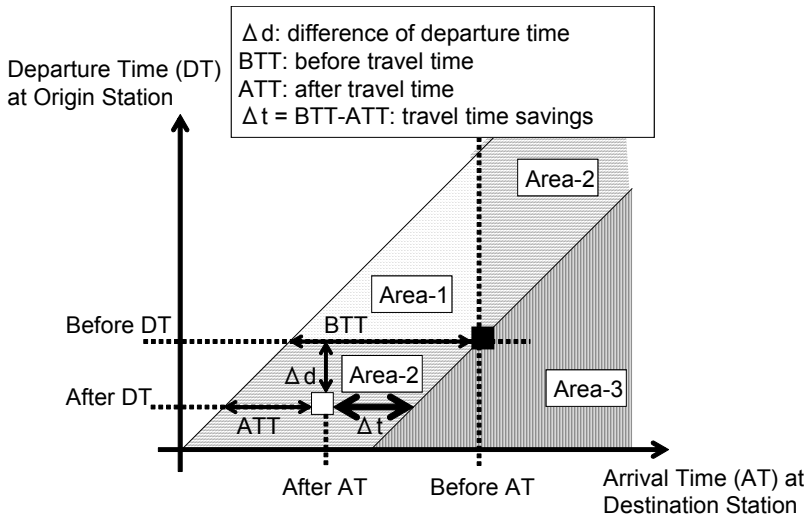


Figure 3: Effects of individual passengers.

4 Empirical analysis

The smart card data related to the station A-1 of one railway company A and stations B-1, B-2, B-3 of the other company B are used for empirical analysis. After the timetable improvement, limited express commuter trains stop at the stations A-1 and B-1.

4.1 Arrival time distribution

The number of passengers passing through the entrance gate is aggregated for every one minute at the station A-1 during a morning peak hour. This is calculated for each 12 days before and after the timetable improvement, respectively. Fig. 4 shows the arrival time distribution of passengers at the station A-1. A white triangle and a black triangle along the horizontal time axis denote the departure times of local and rapid trains before and after the timetable improvement, respectively. A black arrow indicates the departure time of a limited express train.

It is shown that the waves of the passenger's arrival time can be found in accordance with the train departure times. The peak of each wave is about 5 minutes before a train departure time. The peak moves to the departure time of an express train after the improvement. The wave becomes higher in the later time period after the improvement. This indicates that the passengers depart lately as they can use faster trains.

4.2 Travel time distribution

Travel times were calculated for those who passed the entrance ticket gate of the station B-1 during 7:00 am and 9:00 am and arrived at the exit gate of the station

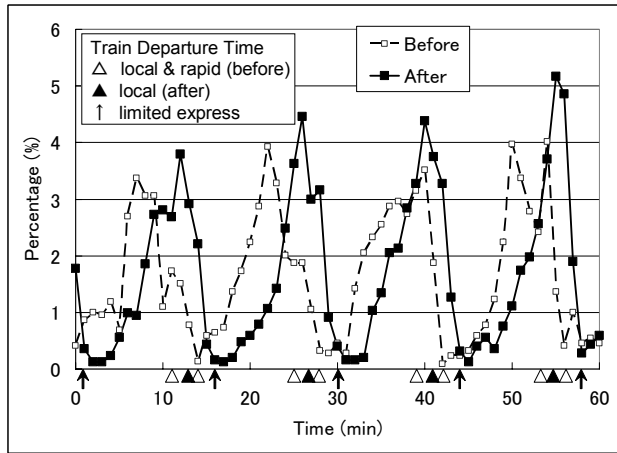


Figure 4: Arrival time distribution.

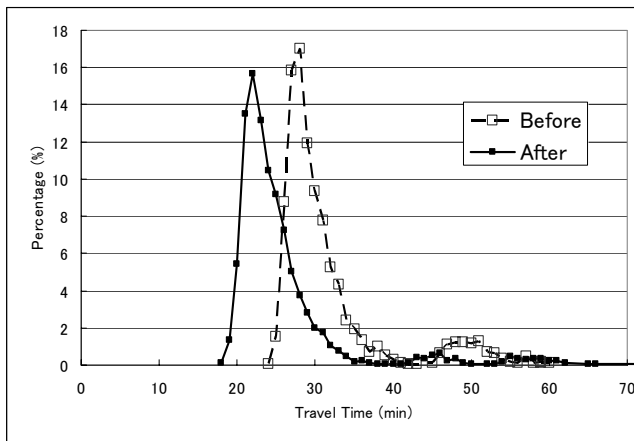


Figure 5: Travel time distribution.

B-2 before 9:00 am. Then, they were aggregated and the travel time distribution (Fig. 5) was obtained between the station OD pair B-1 to B2.

It is found that the distribution moves to the left after the timetable improvement. This means that the OD travel time decreases as limited express trains stop at the station B-1. The number of passengers at station B-1 also increases by 11% after the improvement. This is probably because by the new users from adjacent stations, the new smart card holders and the demand shift from the shoulder hours before 7:00 am and after 9:00 am.

4.3 Departure and arrival times of individual passengers

The effects of the timetable improvement can be found clearly in the behavioural changes of individual passengers. Fig. 6 depicts the departure and arrival times of passengers before and after the improvement. Each dot in the figure denotes one day of the 12 days. When multiple days are plotted at the same point, the colour of the dot becomes deeper.

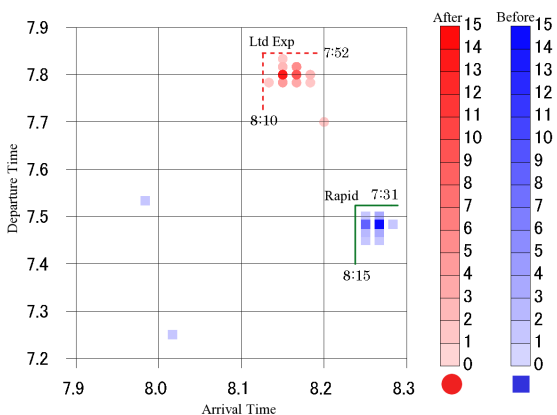


Figure 6: Departure and arrival times of a passenger.

Let us consider a passenger who had used a rapid train before the improvement which took 45 minutes to the destination. He/she adjusted the behaviour immediately after the introduction of the new timetable. On-Board travel time was reduced to 20 minutes by the limited express. This passenger has achieved 20 minutes of late departure and 5 minutes of early arrival.

4.4 Overlapping analysis

Local trains and rapid trains stopped at station B-1 before the timetable improvement. It was possible to take either local or rapid trains and then change to a limited express train at station B-3 located on the way to station B-2. Fig. 7 shows the scatter-gram of the departure and arrival times of passengers who travelled from the origin station B-1 to the destination station B-2. A darker dot indicates the larger number of passengers with the same departure and arrival times. It is found that there are two types of passengers; one changed to the limited express train at station B-3 and the other used the same train from station B-1 to station B-2. The ratio of those two groups was approximately 70 to 30.

After introducing the new timetable, limited express trains, faster than rapid trains, were scheduled to stop at the station B-1. Fig. 8 shows the scatter-gram of the departure and arrival times of the same passengers after the timetable improvement. Those who travelled from station B-1 to station B-2 have adjusted themselves to take the new express trains at station B-1. However, there were some passengers who still remained to use local or rapid trains. Those passengers

may prefer to avoid congestion of express trains even if they could save travel times. When Fig. 7 and Fig. 8 are carefully compared, it is recognized that the scattered areas of passengers have moved to the upper left directions of the diagram. This indicates that the departure times at the origin station B-1 becomes later, and the arrival times at the destination station B-2 becomes earlier. After the timetable improvement, the passengers travelling between stations B-1 to B-2 have saved travel times as well as expanding time margins at both trip ends. Figure 7.

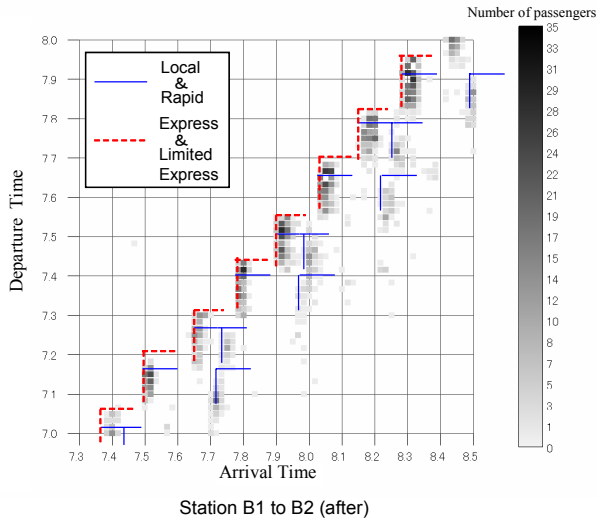


Figure 7: Scatter-gram for new timetable.

5 Conclusion

In this paper, the behaviour of railway passengers has been analyzed by using the smart card data. The accurate origin departure time and the destination arrival time of passengers were the key factors of estimating the distributions of travel times and arrival times. When the time based information was combined with train timetable, it could be possible to estimate how each passenger used the trains between any pair of origin and destination stations.

Using the actual smart card data, it was investigated how passengers adjusted their behaviour after the railway companies had changed the train timetable. The new timetable has brought the large benefit of travel time savings between some pairs of origin and destination stations. It was found that there were two types of passengers with different preferences of trains. The effects of the new timetable may be separately discussed for those different user groups.

The smart card data has a great potential to understand the longitudinal effects of transport policies on users' behaviour. At this moment, the detailed longitudinal analysis has not yet been completed. However, the preliminary

analysis presented in this paper suggests that the smart card data could be analyzed carefully to estimate the various aspects of travel behaviour of individual passenger. The results will be presented elsewhere in the near future.

References

- [1] Bagchi, M. & White, P. R., The Potential of Public Transport Smart Card Data, *Transport Policy*, **12(5)**, pp. 464–472, 2005.
- [2] Chu, K.K.A. & Chapleau, R., Enriching Archived Smart Card Transaction Data for Transit Demand Modeling, Transportation Research Board 87th Annual Meeting 2007 Paper #08-0596, 2008.
- [3] Lehtonen, M., Rosenberg, M., Rasanen, J. & Sirkia, A., Utilization of the Smart Card Payment System (SCPS) Data in Public Transport Planning and Statistics, 9th World Congress on Intelligent Transport Systems, ITS America, 2002.
- [4] Morency, C., Trepanier, M. & Agard, B., Measuring Transit Use Variability with Smart-Card Data, *Transport Policy*, **14(3)**, pp. 193–203, 2007.
- [5] Sahin, I. & Altun, Y. B., Potential Uses of Electronic Fare Payment Records for Public Transit Agencies, *ITE Journal*, **77(12)**, pp. 22–27, 2007.
- [6] Trepanier, M., Tranchant, N. & Chapleau, R., Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System, *Journal of Intelligent Transportation Systems*, **11(1)**, pp. 1–14, 2007.
- [7] Utsunomiya, M., Attanucci, J. & Wilson, N., Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, *Transportation Research Record*, **1971**, pp. 119–126, 2006.

