Estimation Method of Trip Purpose from Smart Card Data: Data Fusion Approach

Takahiko KUSAKABE *

Department of Civil Engineering,

Tokyo Institute of Technology, Tokyo, Japan

Yasuo ASAKURA

Department of Civil Engineering,

Tokyo Institute of Technology, Tokyo, Japan

***** Email: t.kusakabe@plan.cv.titech.ac.jp

1 Introduction

Smart card systems are installed to the fare collection systems of public transport. Smart card systems can collect the data continuously along the fare collection as well as they can collect the ID (identification) information of each traveller's smart card. IDs enable us to track individual travellers. Tracking of travellers enable us to analyse individual travellers' trip frequency, their boarding sections, and their trip sequences. Therefore, travel patterns and their variability over long-term periods can be analysed. However smart card data are fragmentary in behavioural analysis. This is because the smart card data are not collected with the explicit goal of behavioural analysis; rather, they are collected for fare collection. These data do not include the travellers' origins, destinations or trip purposes. The previous studies (for example [1]) attempted to detect travellers' behavioural patterns. However, the activities of travellers, which influences on these patterns, should be subjectively guessed by analysts.

Conventional methods for observation of behaviour can collect the data items for behavioural analysis. For example, person trip survey data observes whole trips in a day, including trip purposes, travel modes, actual origins, and destinations. However, it is difficult to collect day-to-day data for long time periods via a person trip survey, because of the cost, processing load, and privacy protection of each respondent. Although smart card data provide only fragmentary information on travellers' behaviour, they can provide it continuously, which is difficult to achieve with a survey. On the other hand, person trip survey can observe detailed information about travel behaviour though they cannot observe it during long-term period continuously. If the advantages of smart card data can be incorporated with person trip survey data, railway administrators would get well understanding of the changes in travellers' behaviour in the long-term period. It would be motive for conducting detailed surveys and improving their operation.

The aim of this study is to develop a data mining methodology to estimate the behavioural contexts of trips and to find their changes observed in smart card data. In order to estimate the behavioural contexts from smart card data, data fusion methodology using person trip survey data is developed. Person trip survey data is used for interpretation of the behavioural contexts.

2 Data Fusion Methodology

In order to understand trip contexts from smart card data, this study estimate the trip purpose of each trip generated by smart card users. In the data fusion method, smart card data and person trip data are tied up with the common data items of smart card data and person trip data. In order to tie up these data, this study employs Bayesian probability. Distribution of trip frequency of each trip purpose, which represents time series characteristics, is estimated by Bayesian interface.

The common data items of smart card data and person trip data are only "Time of day and places of boarding and alighting train." Although the common data items are limited, the trip purpose seems to influence the alighting time and the interval until the next boarding following an alighting at the station. This study refers to this interval's duration as the "staying time." Smart card data and person trip survey data can respectively calculate the alighting time and the staying time. The staying time would include the duration of traveller's activities, travel time between the station and their destinations, and travel time between their other destinations. If a traveller has a single activity near the station after, the staying time nearly represents the duration of this activity.

Let T be the alighting time and let D be the staying time. Both smart card system and person trip survey can observe T and D. However trip purpose A, which are commuting to work, commuting to school, leisure, business, or return trip, can be observed by only person trip data. On the other hand, frequency of trip k made by each traveller in a month can only be observed by smart card data. Let $p_p(A|T,D)$ be share of trips of purpose A when the trips are made by travellers who pass through the gate at time T and have staying time D. This share is only derived from person trip data because A is only observed by person trip data. The share of smart card users whose trip purpose A is estimated by using $p_p(A|T,D)$ in place of the one derived from smart card users. This is calculated by

$$p(A) = p_p(A|T,D)p_s(T,D) \quad (1)$$

where $p_s(T,D)$ is the share of trips calculated from smart card data whose attributes are T and D.

Distribution of the trip frequency of each trip purpose is estimated by using the Bayesian inference. Let $p_s(k,T,D)$ be the share of trips that are made by travellers passing through the gate at time T who have trip frequency k and staying time D. This share is only derived from smart card data because k is only observed by smart card data. A joint probability of trips made by travellers who have frequency k and purpose A is derived by

$$p(k,A) = p_p(A|T,D)p_s(k,T,D) \quad (2)$$

Then the distribution of the frequency is calculated as the posterior distribution $p(k|A)$ using person trip survey data and smart card data. This is described as

$$p(k|A) = \frac{p(k,A)}{p(A)} = \frac{p_p(A|T,D)p_s(k,T,D)}{p(A)} \quad (3)$$

3 Empirical Analyses

Empirical analysis is conducted with actual smart card data obtained from a Japanese railway company in the Osaka area, which is the second largest metropolitan area in Japan. The data period is weekdays during 20 months, from October 2007 to May 2009. The person trip survey data is obtained in the Osaka region in 2002, called as “4th Keihanshin Metropolitan Area Person Trip Survey”. The trips alighting at the station are used for the estimation.

In order to verify the estimation method, the trip purpose of hypothetical smart card data that are created from the person trip survey data is estimated by using conditional probability $p_p(A|T,D)$. With hypothetical smart card data, we can confirm whether the estimation results are correct because they are created from the person trip survey data. Figure 1 shows the result of verification. The result shows that about 80% trips are correctly estimated in total. The trips about commuting and return trips are successfully estimated. However results show the difficulty to estimate about leisure trips, commuting to school trips, and business trips.

Figure 2 shows the day-to-day changes of composition of trip purpose that is estimated from Eq.(1). Several different characteristics of changes are found in different trip purpose. For example, commuting trips and leisure trips are relatively stable. Large decline of commuting trips are found around summer and year-end holiday season. The large variations of return trips are found during summer season. This may be affected by return trips of the baseball games from the other station.

Figure 4 and 5 show the distributions of frequency of trips about commuting and leisure trips respectively that are estimated by Eq.(3). The frequency of trips made by a large number of commuting travellers is as much as the number of weekdays of month. Most of leisure trip travellers are observed once in each month. Frequency of commuting trips is decreased in holiday season, namely August and December. These results show that time series changes of trip purpose that is difficult to obtain from person trip survey data.

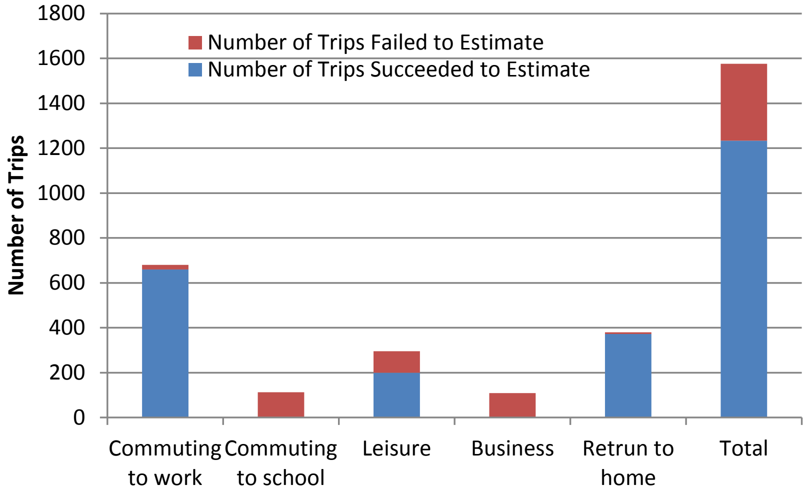

Figure 1 Verification of estimation with hypothetical smart card data created from person trip survey data

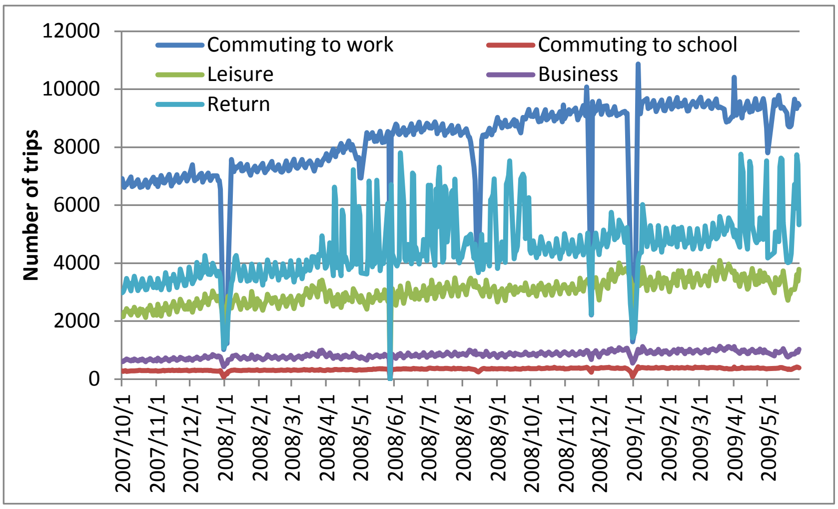

Figure 2 Day-to-day changes of composition of trip purpose of smart card users

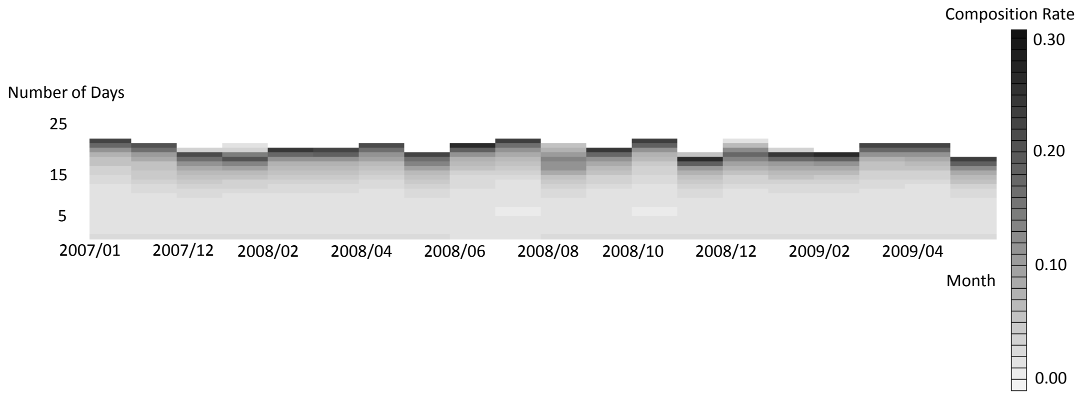

Figure 3 Frequency of commuting trips of smart card users

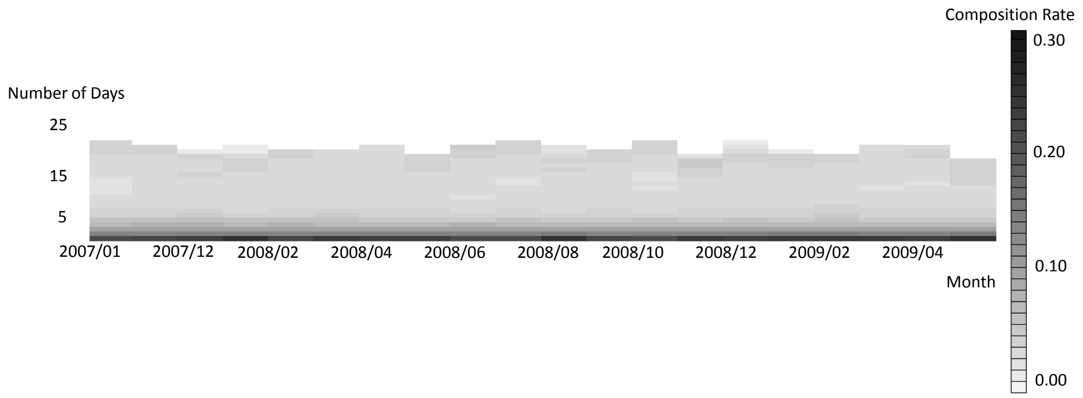

Figure 4 Frequency of leisure trips of smart card users

4 Conclusions

This study focuses on developing the datamining methodology to analyse day-to-day trip contexts using combination of secondary data, which are smart card data and person trip survey data. The advantages of smart card data are incorporated with the advantages of person trip survey data by the developed method. The model of developed method using the Bayesian inference estimates the trip purpose of trips in the smart card data.

As the results of the analysis, the methodology can illustrate the travellers who cause the change of the demand by seeing the share of the trip purpose and the relationship between the trip frequency and the estimated trip purpose. These results are difficult to obtain when person trip survey data and smart card data are separately used. Knowing the changes of trip purpose enables us to guess the cause of changes of demand and it would provide better understanding to railway administrators when they improve their operation.

References

[1] Morency, C., Trépanier, M. and Agard, B.: Measuring Transit Use Variability with Smart-card Data, Transport Policy, Vol.14 (3), pp.193–203, 2007.