VARIABILITY OF PASSENGER TRAVEL PATTERNS OBSERVED USING SMART CARD DATA IN JAPAN

Toshiyuki Yamamoto, Shasha Liu and Toshiyuki Nakamura

Introduction

Regularity and variability in human mobility are considered important factors for better understanding of urban dynamics. From the viewpoint of public transport operators, it is difficult to adjust the service to the demand; however, the operation costs can be reduced via better adjustment of the operation, and it provides a better experience to passengers. Smart card data enable observation of demand using a much larger sample size than traditional survey methods such as paper-and-pencil questionnaire surveys. One of the advantages of using smart card data as travel behaviour survey data is the ability to obtain information from the same respondent continuously without causing any burden to the respondent. Thus, the regularity and variability of travel behaviour can be relatively easily observed using smart card data. Smart card data are used in this study to investigate the regularity and variability of passenger travel patterns in space and time. Many studies have investigated the regularity and variability of travel patterns based on smart card data (Morency et al., 2007; Ma et al., 2013; Zhong et al., 2016; Zhao et al., 2017; Manley et al., 2018; Ouyang et al., 2018; Espinoza et al., 2018). However, the individual sociodemographic characteristics of the users, such as age and gender, are not usually considered because such information is not automatically attached to the smart card data. The fare type was used by Morency et al. (2007) to analyze travel patterns. Chandesris and Nazem (2018) stated that smart card data including sociodemographic characteristics are very interesting to analyze travel patterns. Therefore, in this chapter, smart card data obtained in Shizuoka Prefecture, Japan, are used, which contain not only conventional transaction information but also sociodemographic information, including age and gender. Using this unique dataset, the regularity and variability of travel patterns are investigated in this study, considering the sociodemographic differences as well as fare types.

Data

The smart card transaction data were collected in Shizuoka Prefecture, Japan, from April 1, 2018 to March 31, 2019. These data include the card ID, tap-in date and time, tap-out date
and time, boarding and alighting stop names, public transport line names, and travel costs. This smart card is named the LuLuCa card, which is issued by a public transport operator, the Shizutetsu Group. The urban public transport (PT) system in Shizuoka City and neighbourhood cities is mainly operated by this operator, including one railway line with 15 stations and bus networks with approximately 1600 bus stops. More than 90% of public transport passengers pay for their trips with registered LuLuCa cards in this area. Commuter pass holders can use PT without per-usage fees within the predefined period; however, this is usually only within the segment between one origin and one destination, which are typically the nearest PT stations to the home and the workplace of the commuter. Moreover, the operator offers a monthly pass specific to the elderly aged 65 years and over, which can be used for any segment. Detailed explanations on the LuLuCa card can be found in Nakamura et al. (2017) and Li et al. (2018).

Records with missing information were discarded, and trip legs within a trip were connected, assuming the transfer time to be within 30 minutes to identify the origin and destination of the trip. The average daily boarding ridership is 64,492, and the variations in the average daily boarding ridership across months are shown in Figure 37.1. This figure shows the daily boarding ridership is smaller in January and August, as they are holiday seasons in Japan. In addition, as Figure 37.2 shows, weekends have a smaller average daily boarding ridership than weekdays. Within the weekdays, Monday has a slightly smaller average daily boarding ridership than that of other days. This is because there are more holidays on Mondays because of the ‘Happy Monday’ system in Japan, which moves a number of public holidays to Mondays, creating three-day weekends for those with five-day work weeks.

![Figure 37.1 Variation of average daily boarding ridership across months](image-url)
Methodology

Considering that information pertaining to commuters rarely using PT is insufficient to reveal travel patterns, these users were first filtered out so as to recognise individual travel patterns in the subsequent analysis. In this study, PT users taking PT eight or more times over one month and having at least three months of smart card transaction records were selected. After filtering, 19,426,660 trip records for 76,257 users in the Shizuoka PT system from April 1, 2018 to March 31, 2019 were obtained.

Many PT users have trips during relatively fixed time periods over multiple days. In other words, they have regular temporal travel patterns. To recognise temporal patterns, previous studies usually break down time into discrete time bins (e.g., hourly intervals) or predefined time windows (e.g., morning peak, midday and evening peak). However, these time decompositions may not be appropriate for everyone, since PT users have different travel habits. Besides, the temporal patterns could not be recognised well if the boarding time distributes around the border of two time slots. To address this problem, the boarding time is first represented as minutes from midnight, for example, 4:30 was represented as 270 and 17:00 as 1020. Then, the density-based spatial clustering of applications with noise (DBSCAN) algorithm was used to extract the temporal travel patterns. Unlike most non-hierarchical clustering algorithms, the DBSCAN algorithm does not require the number of clusters to be defined, and it is robust to outliers given that it can recognise noise. Two key parameters need to be specified in the DBSCAN algorithm, including the distance threshold $Eps$ and the minimum number of points $MinPts$. $Eps$ defines how close points should be to each other to be considered a part of a cluster. If a point falls within the $Eps$ distance, this point will be included in an existing cluster. $MinPts$ specifies the minimum number of points in each cluster. For example, if $MinPts$ is set to six, then there are at least six points in each cluster; otherwise, points in this cluster are considered noise (Ma et al., 2013). Based on the preliminary analysis, the $Eps$ was set to 30 minutes and $MinPts$ to 24.

**Figure 37.2** Variation of average daily boarding ridership across days of the week
Apart from travel regularity in time, many users repeatedly visit the same origin–destination (OD) pairs over multiple days. For each trip, there is an origin and a destination represented by physical PT stations. However, there may be more than one PT station around a user’s origin or destination. Besides, some users may get off the bus before reaching the final bus stop when traffic congestion is heavy. Therefore, it is necessary to group some PT stations as one origin or one destination before extracting spatial travel patterns. Based on the distribution of the distance between PT stations, if the distance between PT stations is shorter than 750 m, these PT stations are considered common PT stations. Afterwards, the OD pair information (including OD pair names, number of trips between OD pairs) of each user was extracted to represent individual spatial patterns.

In addition, travel pattern variability exists from person to person and from day to day. To investigate travel pattern variability, an entropy measure was developed based on the study of Song et al. (2010), characterising the heterogeneity of spatiotemporal patterns as follows:

$$S_k = - \sum_{X_{\eta} \in D_k} P(X_{\eta}) \log_2 P(X_{\eta})$$  
\text{Equation 37.1}$$

where \( S_k \) is the entropy for PT user \( k \), \( D_k = \{X_1, X_2, \ldots, X_L\} \) represents the set of daily travel patterns of user \( k \), \( X_\eta \) denotes the time-ordered sequence of OD pairs at which the user was observed at each hourly interval, and \( P(X_\eta) \) is the probability of finding a particular time-ordered sequence \( X_\eta \). \( S_k \) represents not only the visiting frequency of distinct locations at different time but also the order in which these locations were visited. A higher \( S_k \) indicates that the spatiotemporal pattern is more diverse and more diffuse.

In summary, the regular travel time slots of each PT user were extracted using DBSCAN algorithm, characterising the individual temporal patterns. Considering the spatial relationship between the neighbouring stations, the OD information of each user was recognised, representing the individual spatial patterns. Individual spatiotemporal travel characteristics in public transport network can then be explored through combining the temporal and spatial patterns. Moreover, entropy was used to measure the variability in spatiotemporal travel patterns of heterogeneous population.

**Results**

**Temporal patterns**

Based on the clustering results of temporal patterns, the irregular trips that were recognised as noise (19.10%) were separated from the regular trips (80.90%) so as to calculate the average hourly trips. Figure 37.3 shows the distribution of regular and irregular hourly trips. In this figure, trips made by infrequent PT users were also included in the irregular trips. It was found that there were more irregular trips during the periods 10:00–15:00 and 21:00–23:00, suggesting that irregular trips are more likely to be made in non-peak hours. Besides, the proportion of irregular hourly trips during weekends was usually larger than that on weekdays, indicating that trips have a higher probability of being irregular during weekends.

The number of temporal clusters for each user were also calculated. For 16.69% of PT users, their travel records were recognised as noise, indicating that their boarding time follows a highly irregular distribution. For 44.73% of the users, two temporal clusters were obtained, accounting for the largest proportion of PT users. These users mostly commute using PT.
Variability of passenger travel patterns

Only the temporal patterns of 19.41% of the users could be classified into three or more clusters, indicating that their travel time is relatively diffuse. Regarding the sociodemographic differences, the proportions of users with two and three clusters were higher for males, while females constituted larger proportions in the other numbers of clusters; this finding suggests that the temporal patterns of females may be more diverse. As for the cluster number variability with age, users aged 65 years and over made up a large proportion when the number of clusters was zero (that is to say, all the records were recognised as noise), indicating that the elderly probably have irregular and diverse temporal patterns. For the number of clusters equal to two, the proportions of users aged 12 years and below and those aged between 25 and 64 years were relatively greater. The reason may be that these users commute using PT at fixed time slots.

Spatial patterns

Based on the identification of spatial travel patterns for each user, the number of OD pairs per user were found to be mainly within 15 (70.81%); moreover, only 5.75% of users had more than 30 active OD pairs in a year, indicating that users usually have limited active OD pairs. The average proportion of trips with different OD pairs to the total trips were calculated, with the results shown in Figure 37.4. Most of the trips were made between the top two OD pairs (76.97%), in line with the results of Zhao et al. (2017). This may be because most users have limited active locations, for example, between their homes and workplaces. In addition, the proportion of trips involving the top two OD pairs was slightly higher for males than that for females, suggesting that males have more concentrated active locations compared to females. As for the trip distribution variability with age, the proportion of trips with the top two OD pairs was the highest for users aged 12 years and below (91.05%); it was the lowest for users aged 65 years and over (64.06%), indicating that the elderly have more diverse activity spaces.

Figure 37.3  Hourly trip distribution by regular and irregular trips across time of day
Spatiotemporal patterns

Combining the temporal patterns and spatial patterns, spatiotemporal patterns were extracted for each user. Then, the ratio of the spatiotemporally regular trips to the total trips for these PT users was calculated. The results showed that males had a higher proportion of regular trips than females, suggesting that the trips undertaken by males follow a more regular spatiotemporal distribution. Besides, users aged 65 years and over showed a lower proportion of regular trips, indicating that the elderly tend to have a less regular spatiotemporal pattern. Furthermore, compared to non-pass holders, the proportion of regular trips was significantly higher for pass holders.

The entropy that characterises the heterogeneity of spatiotemporal patterns was also calculated. The average entropy was 4.01. Figure 37.5 shows the variation in the average entropy by age and ticket types. The users aged 12 years and below had the smallest entropy, suggesting the most regular spatiotemporal travel patterns. In contrast, it showed the highest entropy for users aged 65 years and over. Namely, the elderly have diverse spatiotemporal travel patterns. It was also found that the variability of the spatiotemporal travel patterns increased with age until the early 20s and decreased with age until 65 years. Then, the variability increased again after retirement at 65 years. As for the pass holders, the users aged 19 to 64 years showed smaller entropy than non-pass holders, and the entropy for older users was smaller than that for younger users. Therefore, it can be inferred that the regular commuters became more regular, as indicated by the variations in spatiotemporal travel patterns with age. Among the pass holders and non-pass holders, users aged 65 and over showed different characteristics from younger users; pass holders exhibited higher entropy than non-pass holders, implying that pass holders have more diverse spatiotemporal travel patterns than non-pass holders. This trend is because of the specific monthly pass for the elderly aged 65 years and over. As explained previously, the pass can be used for any segment, as opposed to the commuter pass applicable only within the

![Figure 37.4 Trip distribution of ordered OD pairs](image-url)
Variability of passenger travel patterns

A conclusion could be that the introduction of such a monthly pass encouraged the elderly to engage in diverse spatiotemporal travel patterns, although the endogeneity should be considered where only those who have diverse spatiotemporal travel patterns buy the monthly pass.

Conclusions

Regularity and variability of passenger travel patterns in terms of trip timing and destinations were investigated using smart card data in Japan. The smart card data collected in Shizuoka Prefecture, Japan, for one year were used in this study, which unusually also include sociodemographic characteristics such as age and gender in addition to the conventional smart card transaction information. The temporal travel patterns were extracted using the DBSCAN algorithm, while the spatial travel patterns were extracted considering the neighbouring stops potentially used for starting and ending the same trip. A spatiotemporal entropy measure was also used for characterising the variability of travel patterns.

For temporal patterns, approximately 80% of the trips were identified as regular trips; however, irregular trips accounted for a greater proportion in non-peak hours than in peak hours. Moreover, 45% of users were observed to have two temporal clusters, while only 19% had more than two clusters, and 17% had no temporal clusters. As for spatial travel patterns, 71% of public transport passengers were active between 15 or fewer OD pairs in a year. Moreover, 77% of the trips involved the top two OD pairs. These results suggest that the travel patterns of the frequent public transport users are mostly regular both in time and space. The results of spatiotemporal travel pattern analysis further suggested the variations in the sociodemographic characteristics and ticket types. Males and young users aged 12 years and below were found to have more regular spatiotemporal travel patterns, while users aged 65 years and over had more diverse spatiotemporal travel patterns. In addition, commuter pass holders were unsurprisingly
found to show more regular spatiotemporal travel patterns than non-pass holders because they have fixed workplaces; moreover, the commuter pass is only applicable to the segment between the home and workplace. In contrast, elderly monthly pass holders aged 65 years and over showed more diverse spatiotemporal travel patterns than non-pass holders because this monthly pass is applicable to any segment. Hence, these results imply the possibility that introducing such a monthly pass encourages the elderly to engage in diverse spatiotemporal travel patterns, although the endogeneity should be considered. Besides, mobility is closely associated with the wellbeing and health of the elderly. Such a monthly pass can attract more elderly people to travel by public transport, which will improve the mobility of the elderly and support the older population in dependent living and wellbeing.

Utilising the smart card fare transaction system, a more flexible fare system can be developed by including age-specific monthly passes and commuter passes with multiple destinations. The smart card data can be utilised to determine the feasibility of developing new fare systems and monitor the effects of implementing such new fare systems on spatiotemporal travel patterns.

Travel patterns are also affected by other factors than the operation of public transit services such as new fare systems and timetables. The COVID-19 pandemic has significantly affected the passenger travel patterns, and the smart card data are suitable to understand the effects not only on the number of passengers but also the spatiotemporal changes in the travel patterns at the individual level. Moreover, smart card data with sociodemographic characteristics such as those used in the analysis of this chapter enable investigation of the vulnerable segments in the public transit users and help to develop appropriate measures specific to each segment.

Acknowledgements

The authors sincerely thank the Shizutetsu Group for providing the smart card data.

References


