Investigation of Changes in Passenger Behavior Using Longitudinal Smart Card Data

Rattanaporn Kaewkluengklom[1] · Fumitaka Kurauchi[1] · Takenori Iwamoto[2]

Abstract

To better understand long-term patterns of human mobility, this study examines changes in travel behavior at the individual level based on yearly activity profiles using 3 years of longitudinal smart card data collected in Shizuoka, Japan. We first characterize spatiotemporal patterns of railway usage by k-means clustering, and then investigate variation in cluster membership with time. For among passengers who remained active, regular commuters had similar travel patterns over the study period, whereas infrequent travelers significantly increased their use of the railway system. The evolution of cluster assignment is analyzed and discussed.

Keywords Longitudinal smart card data · K-means clustering · Passenger travel patterns · Public transport

1 Introduction

Deep analysis of passenger travel behavior allows transport authorities and planners to better understand travel supply and demand. Traditional travel demand modeling heavily emphasizes the number of passengers traveling between different origin–destination (OD) pairs, as well as passenger demand at stations. However, overall mobility patterns and habitual behavior are also of great interest, where better understanding of passenger travel patterns could improve decision-making and modeling simulations [1]. Cognizance of changes in passenger habits over time could aid predictions of the demand for public transport. In summary, studying mobility patterns may improve understanding of passenger behavior, such that public transport services can be optimized.

In line with this, this study examined patterns of public transport usage and yearly variability of travel behavior based on 3 years of longitudinal smart card data. We first determined the frequency of smart card transactions per year, where the identity (ID) number of each card was recorded to assess spatiotemporal patterns of public transport usage. Based on these yearly activity profiles, passengers were clustered by mobility pattern. We then investigated variability in passenger cluster membership over sequential 1-year periods to determine whether passengers changed their behavior between years. The outcomes of this study provide a deeper understanding of human mobility patterns and changes therein, which is important for improving travel forecasting models. The results may also be instructive for urban transport planners seeking to identify improvements that would encourage the use of public transport. If a substantial number of passengers change their activities between years, public transport services must consider significant service adjustment, rather than largely maintaining their network with only a few changes.

This paper is organized as follows. The first section discusses the advantages of smart card data analysis. Then, we briefly describe the study area and dataset, and perform cluster analysis of yearly passenger activity patterns. The results of our preliminary analysis of travel behavior are then presented, and variation in cluster membership and behavior over time are discussed. Finally, the study findings are discussed, conclusions are drawn, and suggestions for future work are provided.

✉ Rattanaporn Kaewkluengklom
Rattanaporn.kkm@gmail.com

Fumitaka Kurauchi
kurauchi@gifu-u.ac.jp

Takenori Iwamoto
t.iwamoto@shizutetsu.co.jp

[1] Department of Civil Engineering, Gifu University, 1−1 Yanagido, Gifu 501-1193, Japan

[2] Shizuoka Railway Co., Ltd., 1-1-1 Takajo, Shizuoka 420-0839, Japan

1.1 Smart Card Data Obtained from Automated Fare Collection Systems

Smart card data collected by automated fare collection systems have been widely used recently for studying urban mobility and evaluating the performance of public transport systems [12]. Smart cards enable cash-free payments in public transport systems in many cities around the world. Each card contains an embedded microchip capable of storing and processing data when a passenger enters a subway station [6]. Smart card data not only include the passenger ID, but also the boarding and alighting time and location, fare type, and so on. Such data provide an opportunity to analyze transport use at the aggregate or disaggregate scale from both spatial and temporal perspectives, which can help transport authorities understand the demand and use patterns of public transport networks on a daily basis. Typical boarding and alighting station use patterns, together with land use information, can be analyzed to better understand spatial mobility [15]. Moreover, smart card data enable the study of temporal changes in day-to-day mobility patterns [14], hourly travel patterns [13] and changes in travel behavior over long periods [1, 3, 11].

1.2 Determining Mobility Patterns in a Public Transport System

Many previous studies have used smart card data to examine passenger behavior and public transport use within a network. Various indicators have been used to characterize passenger mobility patterns based on smart card data, such as travel intensity, travel type, timestamp, and location data [3, 15]. Station usage data and weekly public transport use patterns are considered important for determining passenger travel behavior [7]. Zhao et al. [13] analyzed hourly ridership patterns to predict the pattern of station usage according to surrounding land use type. In other studies, public transport usage patterns and fare types were analyzed temporally based on smart card data [1, 3]. The majority of the literature pertaining to passenger travel behavior focuses on temporal or spatial aspects. However, considering both temporal (within- and across days) and spatial aspects of travel behavior could be more instructive.

Cluster analysis has frequently been employed to obtain an overview of passenger behavior. Many clustering algorithms have been developed, including k-means, k-medoid, and density-based spatial clustering, as well as combination unigram/Gaussian mixture models [1, 7, 13]. k-means clustering has been utilized to distinguish among passenger mobility patterns in many studies [5, 11]. Hierarchical two-step clustering of passengers benefits from high computational efficiency [15]. Time series modelling and hierarchical clustering analysis have been used to classify passengers' daily activities

[4]. As a more advanced model, Mohamed et al. [8] applied a combination Poisson mixture/unigram model to classify stations with similar usage patterns into time bins, and passengers based on temporal activity patterns.

1.3 Behavioral Changes Based on Longitudinal Data

In the last 5 years, long-term smart card datasets have become available in many cities, leading to an increase in the number of studies using longitudinal data to assess public transport usage patterns. Investigating changes in mobility patterns can help transport planners improve the accuracy of travel demand models; many such models have been proposed to better understand changes in these patterns over time. Time series models and survival analysis can be applied to many areas of research [10, 15]. Previously, public transport passenger profiles based on 1- and 6-month smart card datasets were derived to detect changes in departure time after the introduction of a discount fare policy, using time series modelling [15]. The retiming elasticity was found to be twice as sensitive in the case of infrequent passengers in the medium term (6 months) versus the short term (1 month). Elsewhere, survival analysis of a 16-week dataset was performed to identify passengers likely to reduce public transport use over time [9]. However, a Cox proportional hazard model of factors associated with passenger travel patterns revealed no significant reduction of public transport usage within a short time period [9]. This suggests that passengers may be unlikely to change their behavior markedly in the event of a reduction in tram services. Further studies with datasets covering a longer period are needed to better understand changes in travel behavior. Previous studies analyzed 1-year smart card datasets in terms of variations in the volumes of travelers and travel patterns [1, 3, 11]. For instance, two studies assessed the evolution of transit use behaviors via multi-week cluster analysis. The results of the two studies were similar: most users showed regular and stable public transport use patterns, even though the patterns of some users were disrupted by public holidays [3, 11]. Briand et al. [1] investigated changes in individual behavior over a 5-year period using a cluster partitioning analysis approach. Although some passengers showed a change in cluster group membership over 1 year, most of these changes were to a cluster with a similar travel pattern. To our knowledge, the timescale (daily, weekly, or yearly) over which travel behavior changes are most evident is still not clear. If passengers find a public transport service satisfactory, they may not change their travel behavior at all. A case study using long-term data would be valuable to clarify this issue.

The longitudinal analysis approach of Briand et al. [1] was adopted in this paper, i.e., a generative model in which time is represented as a continuous variable, rather than by discrete time bins. As in that study, variability in travel behavior at the individual level was determined based on membership of

Fig. 1 Map showing Shizuoka Prefecture and (inset) the Shizuoka–Shimizu Line (adapted from Wikipedia)

cluster groups devised according to temporal activities from year to year. Rather than considering only the temporal dimension, the spatial dimension of activities was also analyzed in this study. Changes in travel behavior pattern at the yearly scale were determined using 3 years of data. Changes in passenger behavior at the individual level were assessed according to cluster membership over the 3-year period years. The

variability can be understood how passengers change their behavior pattern by time, based on a single user.

2 Methodology

2.1 Case Study and Dataset

We used the local railway network in Shizuoka Prefecture, Japan as a case study. As shown in Fig. 1, Shizuoka Prefecture is located in the Chubu region and is near to many attractive landmarks, such as Mount Fuji. The dataset comprises "LuLuCa" smart card data, provided by the Shizutetsu Group. The 11-km Shizuoka Railway serves over 10 million rail passengers, and over 30 million bus passengers, at 15 stations located in the center of Shizuoka City. At present, more than 60% of these passengers use a smart card. Four types of LuLuCa smart card are available, but the data from only two of these ("LuLuCa PASAR" and "LuLuCa Plus"), which can be used to pay for both rail and bus services, were analyzed herein. The data are largely anonymous, although we also obtained some personalized data for the analysis.
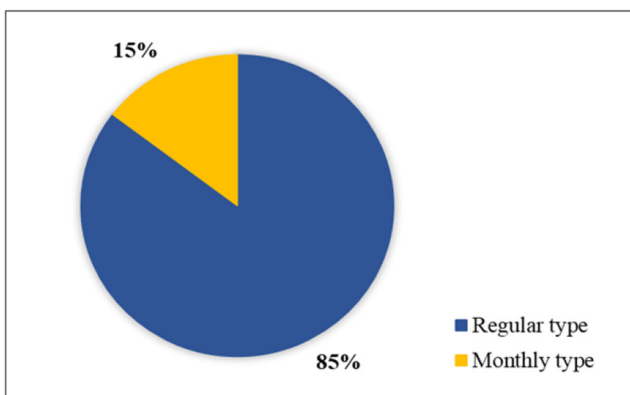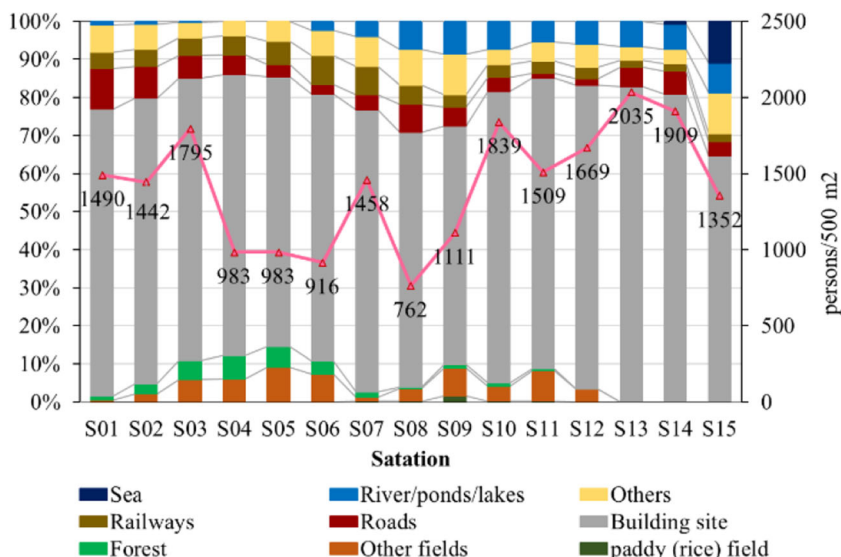

Fig. 2 Proportions of card types

Fig. 3 Land use types around the railway stations

Using a dataset comprising almost 3 years of data (from January 1, 2014 to October 31, 2016), this study aimed to identify the travel patterns of railway passengers along the Shizuoka–Shimizu Line (Fig. 1). The dataset principally comprises boarding and alighting dates and times, and subway station usage and fare data, according to passenger ID. The dataset contains data for 17,422,190 trip transactions made by 220,268 card holders. The data were not filtered prior to the study in terms of whether or not each passenger made at least one transaction during the study period.

To interpret the dataset according to the fare data and card IDs, regular and monthly card types were distinguished (85% and 15% of railway users, respectively; Fig. 2). Land use and population density data within 800 m of the stations were also obtained, as per a previous study [13]. Figure 3 shows the two distinct types of station in terms of nearby land use: those with a high proportion of buildings and low proportion of fields (other than paddy fields) nearby (S1, S2, S7, and S12–15), and those with varying proportions of fields, forest, and "other" land use types (S3–6 and S8–11). The population distribution,

in terms of the density per 500 m², around the stations also suggested two types of station, i.e., stations with high (S1–3, S7, S10–15) and low (S4–6, and S8–9) nearby population densities. Stations with a high proportion of buildings nearby have seen major increases in population density. These additional data were used to assist in the interpretation of the mobility behavior of railway passengers.

2.2 Cluster Analysis and Variability in Travel Behavior over 1 Year

To characterize passenger travel patterns and changes therein within 1 year, we adopted for this analysis k-means clustering, which is a classic clustering approach suitable for analyzing Big Data, via the simplest and fastest distance-based algorithm [11]. We used k-means clustering with the Hartigan–Wong algorithm and Euclidean distances to group railway users with similar spatiotemporal behavior. We then attempted to identify specific characteristics of each group.

Table 1 Variables included in the cluster analysis model

Variable	Definition
Morning trips (peak hours; alighting)	Proportion of trips made each day between 07:00–08:59
Daytime trips (boarding)	Proportion of trips made each day between 09.00–16:59
Evening trips (peak hours; alighting)	Proportion of trips made each day between 17:00–19:59
Nighttime trips (boarding)	Proportion of trips made each day between 20:00–23.59
OD 1	Most used OD pair
OD 2	Second most used OD pair
OD 3	Third most used OD pair
Major stations	Shinshizuoka (S1), Kusanaki (S10), Kitsunegasaki (S13), Sakurabashi (S14) and Shinshimizu (S15)

Fig. 4 Daily distribution of smart card transactions across 1035 days
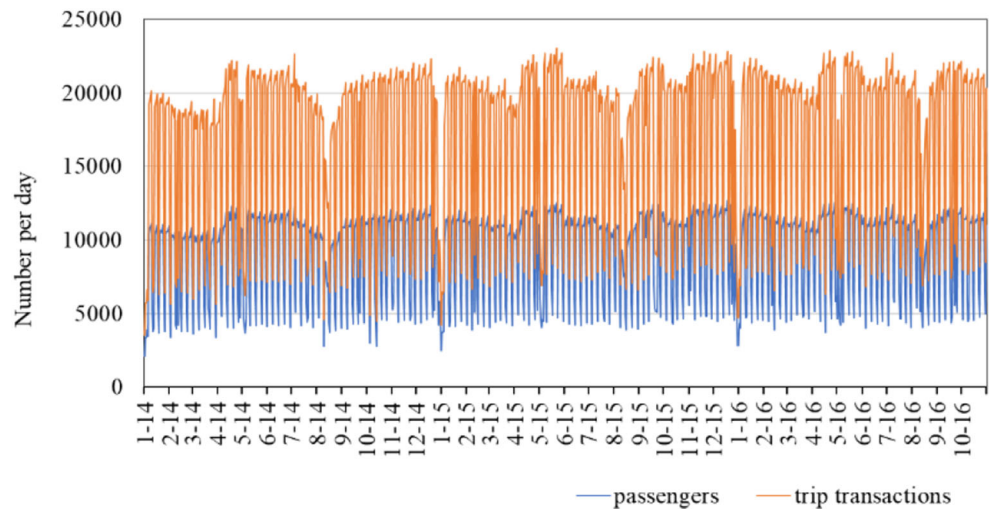
To aid clarity, we converted the individual-level behavioral data into yearly profiles of spatial and temporal activities. The distribution of trips among four daily time periods illustrated how passengers used public transport at specific times of the day. For employees who commute, the proportion of trips during peak hours on weekdays should be higher. The spatial data included relative frequency of travel between different OD pairs and use patterns for the five major stations within the railway network (based on when each passenger "taps" on or off). Habitual OD pair usage data can shed light on the variation (or uniformity) of railway passenger route choices within 1 year. If a passenger primarily commutes to work from home, the use frequencies of the most and second-most used OD pairs should be similar. However, if passengers use many different OD pairs, the rate of use of the second-most used pair should be significantly lower than that of the most used. Additionally, boarding and alighting station usage for a passenger's most-used stations can provide insight into the locations visited, particularly

when land use data are available. Of the 55 original variables, those used in the final model of urban mobility patterns are shown in Table 1.

To evaluate railway user behavior, change over the 3-year period, three separate yearly profiles were derived for each passenger and combined into one dataset for cluster analysis. Changes in cluster membership over the 3-year period were determined based on the card IDs.

3 Results and Discussions

3.1 Preliminary Analysis of Travel Behavior

The 3-year LuLuCa dataset was analyzed to identify travel behavior patterns. Figure 4 shows the number of transactions per day; ~20,000 trips were made by ~10,000 card holders on weekdays (approximately two trips per day per passenger), versus ~6000–10,000 trips on weekends. The number of

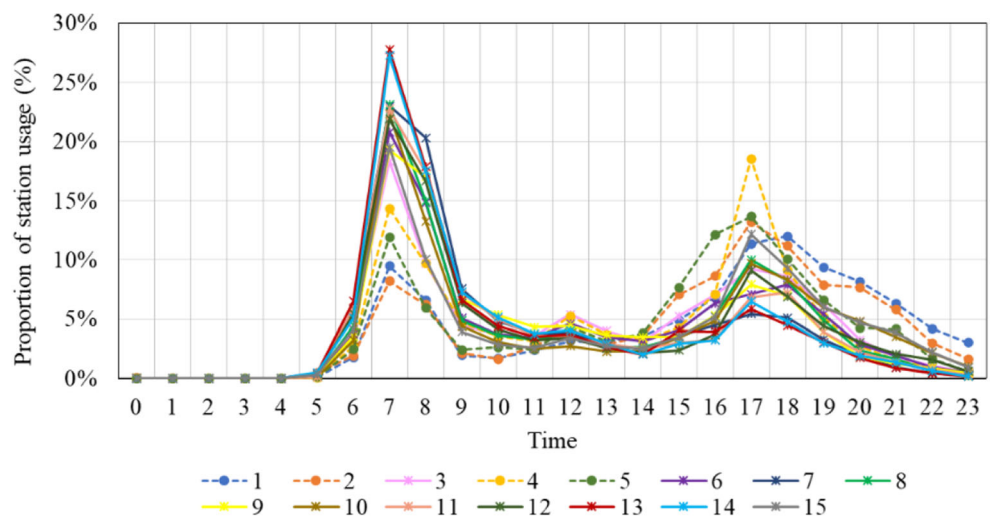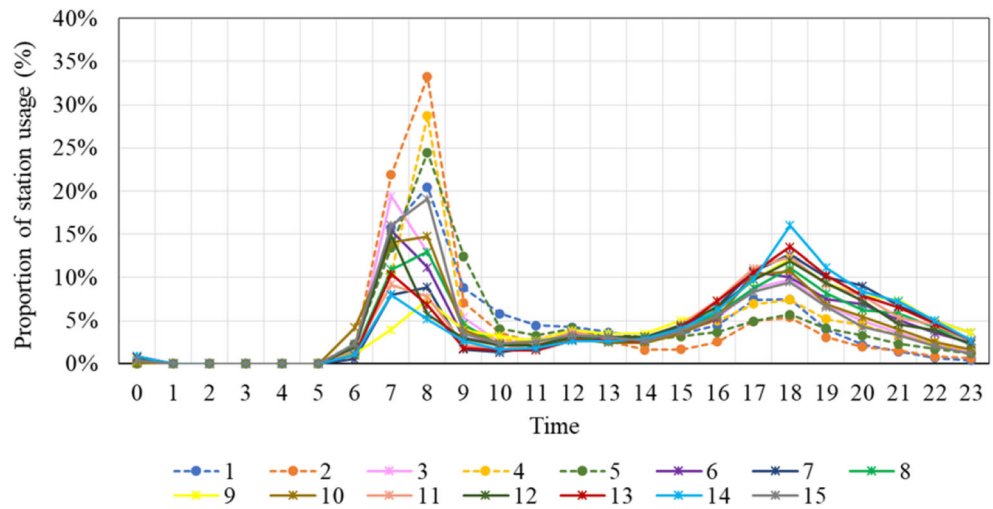**Fig. 5** Hourly distribution of boarding

Fig. 6 Hourly distribution of alighting

transactions decreased markedly during national and school holidays (i.e., in January, April and August).

The data on station usage by hour, i.e., boarding and alighting, are shown in Figs. 5 and 6, respectively for 15 stations. In both figures, there are two clear peaks: one each in the morning and evening. For stations 6–15 S6–15), ridership was higher during the morning peak hours (07:00–09:00) versus evening peak hours (17:00–20:00); the pattern was reversed for S1–5. These patterns can be explained by commuting trips; government institutions and private companies in Japan start and finish work at approximately 08:00–09:00 and 17:00–20:00, respectively. The patterns are consistent with previous studies [2, 13].

The variation of usage of boarding and alighting stations shows that those with a greater number of boarding passengers in the morning had fewer boarding passengers in the afternoon. Similarly, stations with many alighting passengers in the evening had fewer alighting passengers in the morning. This is in accordance with these stations being surrounded by residential areas (S3 and S6–15). By contrast, the stations with more boarding passengers in the evening and more alighting passengers in the morning are located in commercial and business districts (S1, 2, 4 and 5). For S3, 10 and 15, the

proportions of boarding and alighting passengers were both highest in the morning; these stations are in close proximity not only to residential areas, but also to other popular destinations such as universities and interchange stations. The land use and population density data shown in Fig. 3 can help us understand patterns of station usage. For example, stations in residential areas with diverse land use types and high population density have a greater number of boarding passengers during the morning peak period, whereas stations in commercial and employment areas with a high percentage of buildings and lower population density tend to have more alighting passengers during the morning peak period.

3.2 Cluster Analysis of Rail Passengers

For categorizing railway users, the k-means method was used. The elbow method was employed to determine the optimal number of groups, k. Figure 7 plots the variance within groups against the number of clusters. Although the variance largely stabilized after 4 clusters, we used a 10-cluster solution to ensure that the highest possible proportion of variance could be explained, and that the clusters could well-distinguish travel patterns with greater explanatory power. The number of cluster groups can be increased to achieve a more fine-
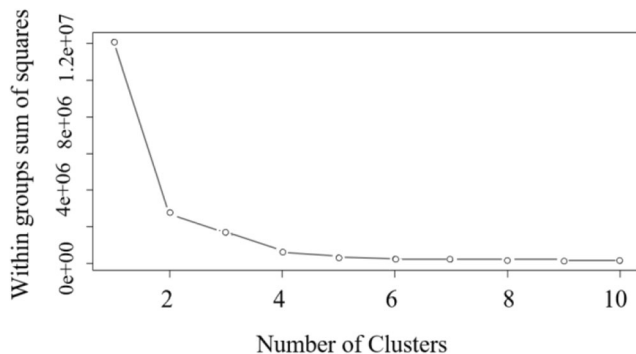

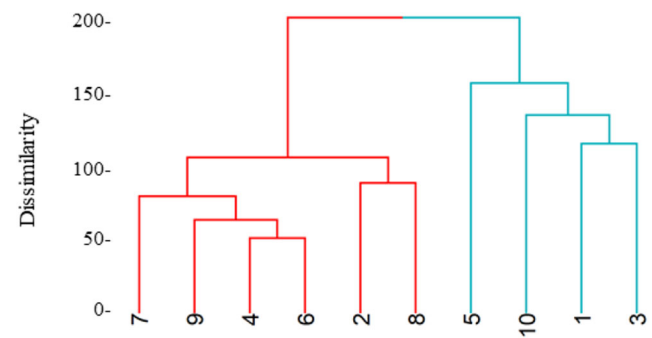**Fig. 7** Plot of within-group variance by number of clusters

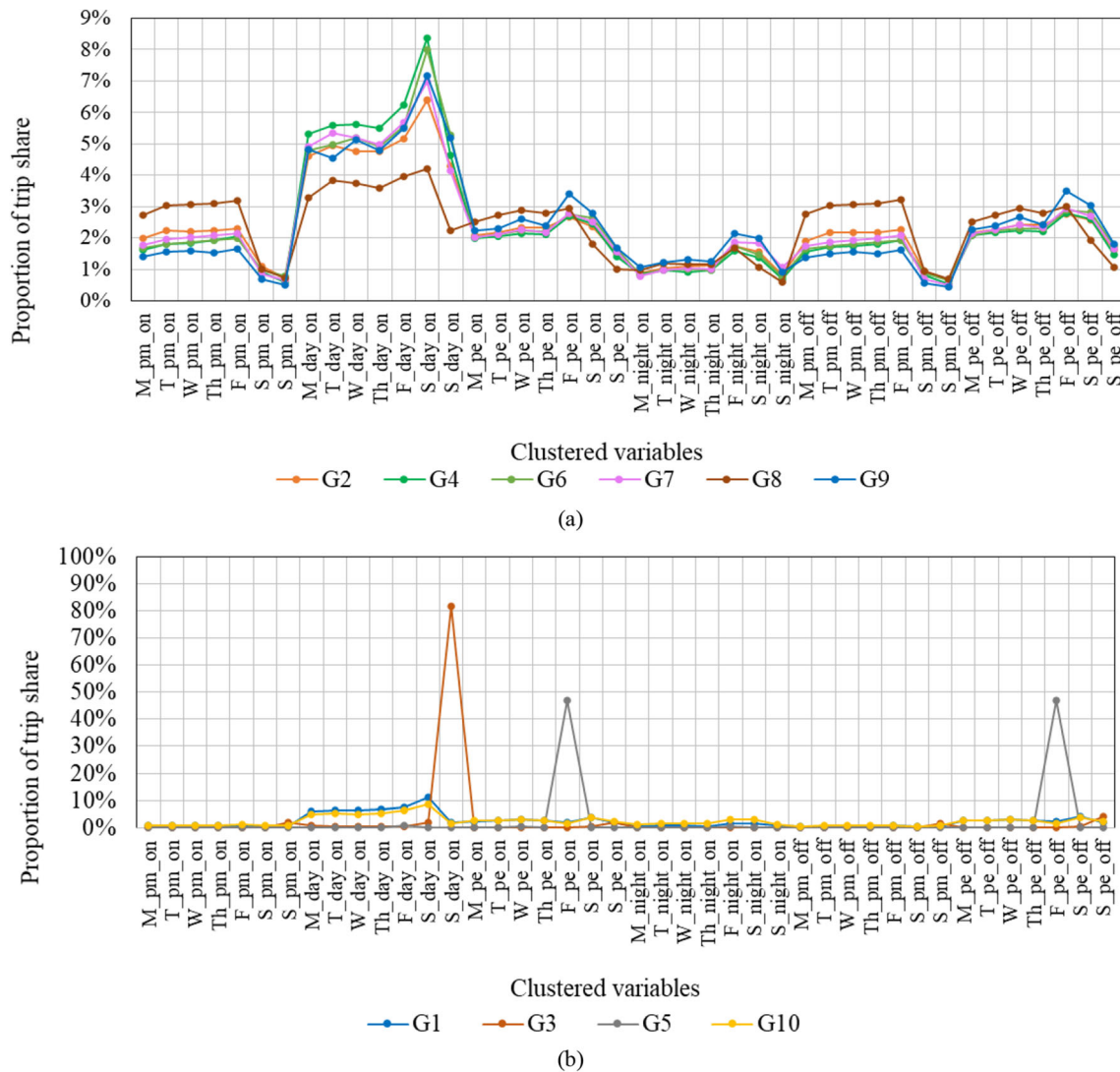**Fig. 8** Hierarchical clustering dendrogram showing the 10 clusters

Fig. 9 Railway usage by time of day and day of the week for different groups of passengers (G1–G10): (**a**) regular commuters and (**b**) infrequent travelers

grained analysis. Moreover, hierarchical clustering was performed with Ward's minimum variance method; the dendrogram of the 10-cluster solution, showing the means of all indicator variables derived from the k-means clustering, is provided in Fig. 8. The level of dissimilarity shows that the 10 clusters could be divided into two main groups.

The railway usage clusters are plotted in Fig. 9. The clustered variables corresponding to four time periods within the day (peak morning [PM], peak evening [PE], nighttime [Night], and daytime [Daytime]), together with boarding (on) and alighting (off) information, are displayed on the x-axis and - the average share of trips (as a percentage)

Table 2 Attributes of the user groups identified by clustering analysis

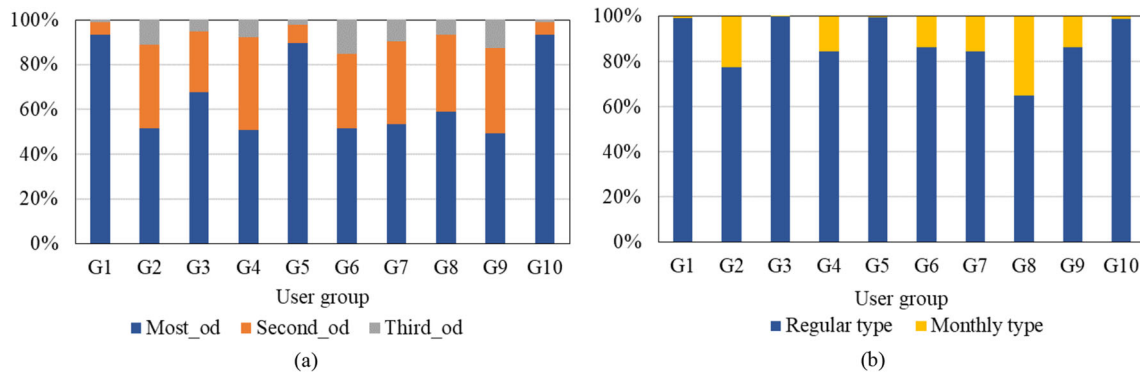| Attribute	User group									
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
Total trips per 3 years	111,393	2,268,620	35,691	5,646,036	5273	3,232,052	2,556,695	2,311,350	1,105,437	149,643
Average trips per year	8	120	4	88	3	89	80	144	78	11
Number of passengers (%)	13,071	18,834	9990	64,005	1911	37,192	31,964	16,042	14,122	13,137
	(5.9%)	(8.6%)	(4.5%)	(29.0%)	(0.9%)	(16.9%)	(14.5%)	(7.3%)	(6.4%)	(6.0%)

Fig. 10 (**a**) Relative use frequency of the three most common origin–destination (OD) pairs and (**b**) proportion of trips by smart card type

corresponding to each x-axis category for each group is displayed on the y-axis. Two main temporal profiles were identified: one for regular commuters, i.e., home-to-work and school commuters (with a high number of total trips per year despite differing boarding times) and one for holiday and "specific activities" travelers (who mostly traveled during holiday periods and in the evening, and used the railway service less frequently overall).

Passenger group 8 (G8), which contained 7% of all card holders, typified the commuting pattern of travel (Fig. 9[a]), with high proportions of morning trips (boarding and alighting) on weekdays and greater overall railway use (a mean of ~144 trips per year per passenger). Similarly, the G2, G4, G6, G7 and G9 travelers can be regarded as commuters, with a mean of ~80 journeys per year. These five commuter groups showed a higher proportion of daytime travel than G8, and lower proportions of morning and evening peak-hour travel. All commuter groups were characterized by frequent trips on Saturdays; some of these passengers might need to make trips outside of working hours to complete other activities.

The G1, G3, G5 and G10 passenger groups were classified as infrequent travelers (Fig. 9[b]). Although G1 and G10 showed some travel activity on weekdays, the average number of such trips was relatively low (~10 times per year); these passengers may use public transport only for specific and irregular activities. The inclusion of

passengers with a small number of trips per year in these groups can be explained by the lack of a minimum number of trips threshold in this study. The attributes of each cluster group derived from the clustering analysis are summarized in Table 2.

The OD distribution by group is shown in Fig. 10(a). Infrequent travelers tended to use one particular OD pair relatively frequently, which accords with their propensity to travel by rail only for specific purposes. This contrasts with regular commuters, for whom the use rate of their most-used OD pair was similar to that of their second-most used OD pair, different from their home-to-work travel behavior. Zou et al. [15] reported similar results, i.e., the rate at which low-frequency travelers used one OD pair was disproportionately high, whereas regular commuters showed greater variety in OD pairs. Furthermore, the card type distribution varied among cluster groups, as shown in Fig. 10(b); regular commuters were more likely to use monthly cards, which provide a discount for frequent travel, while infrequent travelers tended to use regular cards.

Regarding spatial travel patterns, the frequency of use of the five stations used most often by each group is shown in Fig. 11. Boarding and alighting at S1 and alighting at S15 was typical of the G1 and G10 cluster groups, who we regrouped as "travelers with specific purposes".

User group	Description
G8	**Regular commuters (workers)**
	High total number of trips (many during peak hours), few trips on Sundays, high number of passengers, high proportion of monthly smart card users
G2, G4, G6, G7, G9	**Regular commuters**
	High total number of trips (many during the daytime) on weekdays and Saturdays, few trips on Sundays, high number of passengers
G1, G3, G5, G10	**Holiday/Specific activities travelers**
	Low total number of trips; especially few trips on weekdays, many more trips on Sundays and Fridays (particularly during evening peak hours), mostly using the same origin–destination pairs

Table 3 Characteristics of passenger groups

Fig. 11 Proportion of trips by
station for the 10 passenger
groups
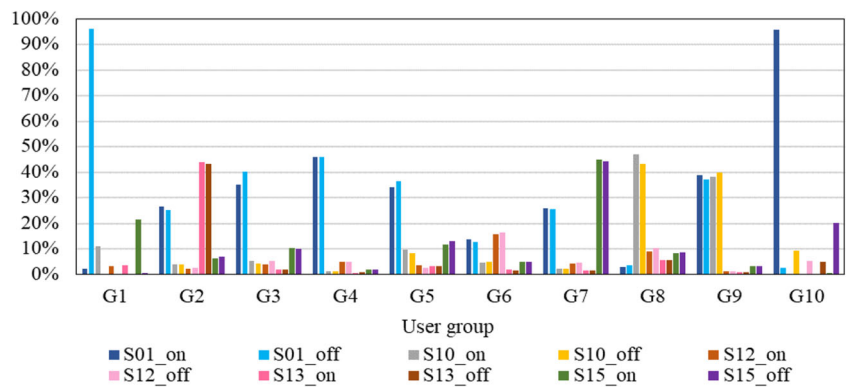
Variations in boarding and alighting stations were observed among the regular commuter groups. S1 and S15 showed typical patterns for stations in the middle of commercial areas that provide connections to other railway lines. The relative proportions of boarding and alighting for S10, S12 and S13, which are within residential areas, varied among the passenger groups. S10 was used frequently by G8 and G9; this station is located in an area of housing and is near to various employers and a private university, corresponding to the large proportion of trips on weekdays. The land use data also helps explain the relationship between the boarding and alighting patterns for each station and passengers' travel patterns over time, as shown in Figs. 5 and 6. The characteristics of the passenger groups are summarized in Table 3.

3.3 Changes in Travel Behavior over Time

This section presents the results of our investigation of the year-to-year variation in passenger behavior, based cluster membership. Figure 12 shows the changes in cluster membership from 2014 to 2016; the proportion of users in each group (denoted by different colors) and the year (from left to right, 2014 to 2016) are shown. In addition to the 10 passenger groups described previously, there is a "nin" group of passengers with cards that have not yet been activated: if a card ID is not activated in the current year but is active in the following year, this card ID is identified as "nin". There is also an "out" group, comprising passengers with cards that have not been used to make any payment (owned by individuals no longer traveling within the railway system).

Fig. 12 Proportions of passenger
groups by year
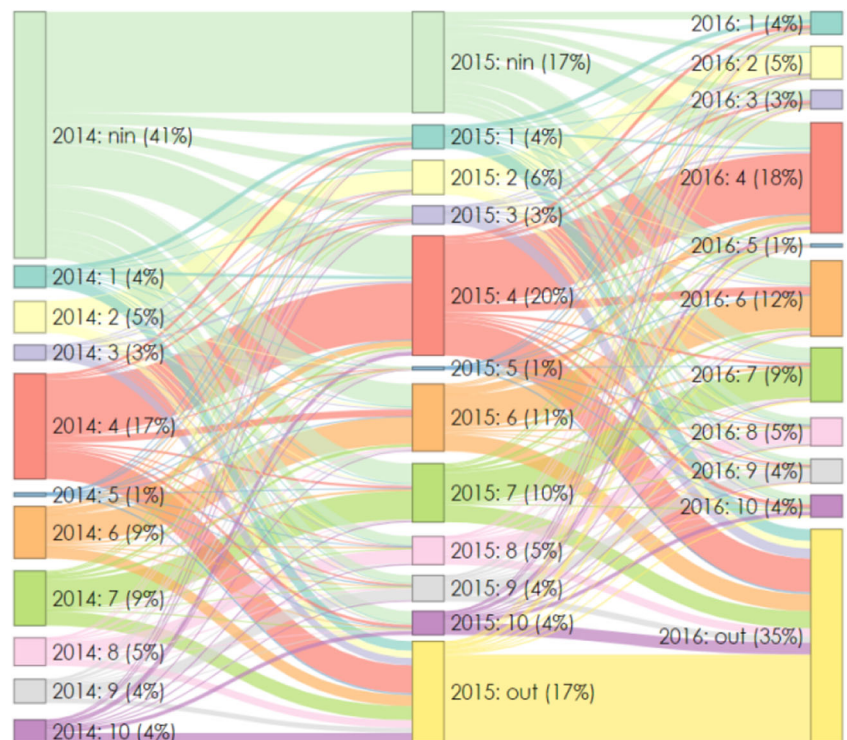
The results reveal that a large proportion of cards become inactive after 1 or 2 years (indicated by membership of the "nin" or "out" group). In total, only 36,203 passengers (16%) used the same card over the 3-year period; many travelers may have changed to another type of card (84% changed card type).

Most travelers remained in the same group over the 3-year period. The mean rates of change in cluster membership over the 3-year study period were calculated; the results are presented in Table 4. All cluster groups are distributed horizontally by candidate group; each row of the table displays the probability that card users changed their group to a new one, including the "out" group, and the elements sum to 100%. In every group, a large proportion of cards "left", i.e., transitioned into the "out" group, particularly in G1, G3, G5, and G10. However, a limitation of the smart card data analyzed in this study is that each card ID is tied to a credit card with an expiration date; because of passenger privacy, we could not access information after the expiration date, so the "out" group is not considered in the rest of this paper.

G1, G3, G5, and G10 showed generally similar travel patterns (all are classified as having infrequent travelers), although some of their members transitioned to G4, which showed a very different pattern (i.e., regular commuters). The rate of card transfer to G4 is notable, particularly for G3 and G5 (11% and 10%, respectively). This suggests that infrequent travelers with cards that remain active are likely to change their travel habits, i.e., to rely more on the railway service, as confirmed by the data shown in Fig. 13. In more detail, Fig. 13(a − d) show the proportions of passengers in G1, G3, G5, and G10 who shifted to G4 in 2015 or 2016. G3 and G5 passengers who shifted to G4 reduced their travel activity on Sundays and

increased it on weekdays, as well as their total number of trips, compared to those who remained in G3 and G5.

Also, it can be seen that some regular commuters (G2, G4, and G6–9) changed to another group with a similar pattern. For example, significant proportions of the cards in G2 and G4 changed to G6, which had a similar travel pattern, in that its members take trips often during the daytime on weekdays, while a few changed to infrequent travel groups (G1, G3, G5, and G10). Therefore, regular commuter behavior was stable over time, whereas infrequent travelers tended to increase their use of the railway system. The changes in travel behavior are similar to those reported in Zou et al. [15], who found that low-frequency travelers had similarly stable public transport use patterns.

Despite changes in the travel behaviors of some passengers, the total number of passengers in each cluster remained stable as shown in Fig. 12. These longitudinal data on passenger travel habits should aid transport planners. Although this study found no significant changes in the travel behavior patterns of regular commuters, passengers who traveled mostly for specific activities increased their use of public transport over time. These findings contrast with those of Briand et al. [1], Deschaintres et al. [3] and Viallard et al. [11], who found that public transport use patterns were stable over multi-week and 5-year time frames, with public holidays only slightly affecting the patterns [3, 11]. Despite the present study not assessing the effect of an interruption of transport service operations on public transport usage, as was done in Nishiuchi et al. [9], our results tend to support their suggestion that data analysis over a 1-year period is useful for monitoring changes in passengers' public transport usage patterns.

4 Conclusions

In this paper, a cluster analysis integrating spatiotemporal activity profiles and OD route usage frequency data over time was performed, to better understand variability in railway passenger travel habits over time. k means clustering was used to group railway passengers based on behavior patterns, and to reveal changes in cluster membership from year to year.

The results showed that most trips were made by regular commuters traveling between home and the workplace. Regular commuters mainly traveled during weekday peak times, but also on weekends to pursue leisure activities. The specific activities group, who traveled less frequently by rail overall, showed the opposite weekday

Table 4 Rates of change in cluster group membership over the 3-year study period (%)

	1	2	3	4	5	6	7	8	9	10	out
1	**19.4**	1.4	3.1	**12.7**	0.8	3.4	5.2	1.4	2.7	2.7	**47.1**
2	1.2	**61.9**	0.9	3.7	0.1	**4.1**	3.2	1.5	0.7	1.2	**21.4**
3	3.9	1.4	**11.1**	**14.1**	0.4	5.7	5.3	1.4	1.9	3.3	**51.4**
4	2.8	0.8	2.3	**52.2**	0.3	**6.8**	2.2	0.6	1.6	3.0	**27.5**
5	5.5	2.0	1.9	**10.8**	**7.2**	5.2	5.2	1.8	3.0	4.9	**52.5**
6	1.4	1.8	1.7	**11.4**	0.3	**50.6**	3.9	2.5	1.6	1.4	**23.4**
7	2.4	1.9	1.9	4.6	0.4	**5.0**	**51.5**	1.5	0.9	2.4	**27.6**
8	1.2	1.9	0.9	2.3	0.4	**6.5**	2.6	**50.4**	5.4	0.7	**27.8**
9	3.1	0.7	1.5	**7.7**	0.6	5.4	2.1	6.0	**47.0**	3.1	**22.8**
10	2.8	2.1	2.4	**14.6**	0.9	4.4	5.6	1.0	2.9	**16.7**	**46.6**

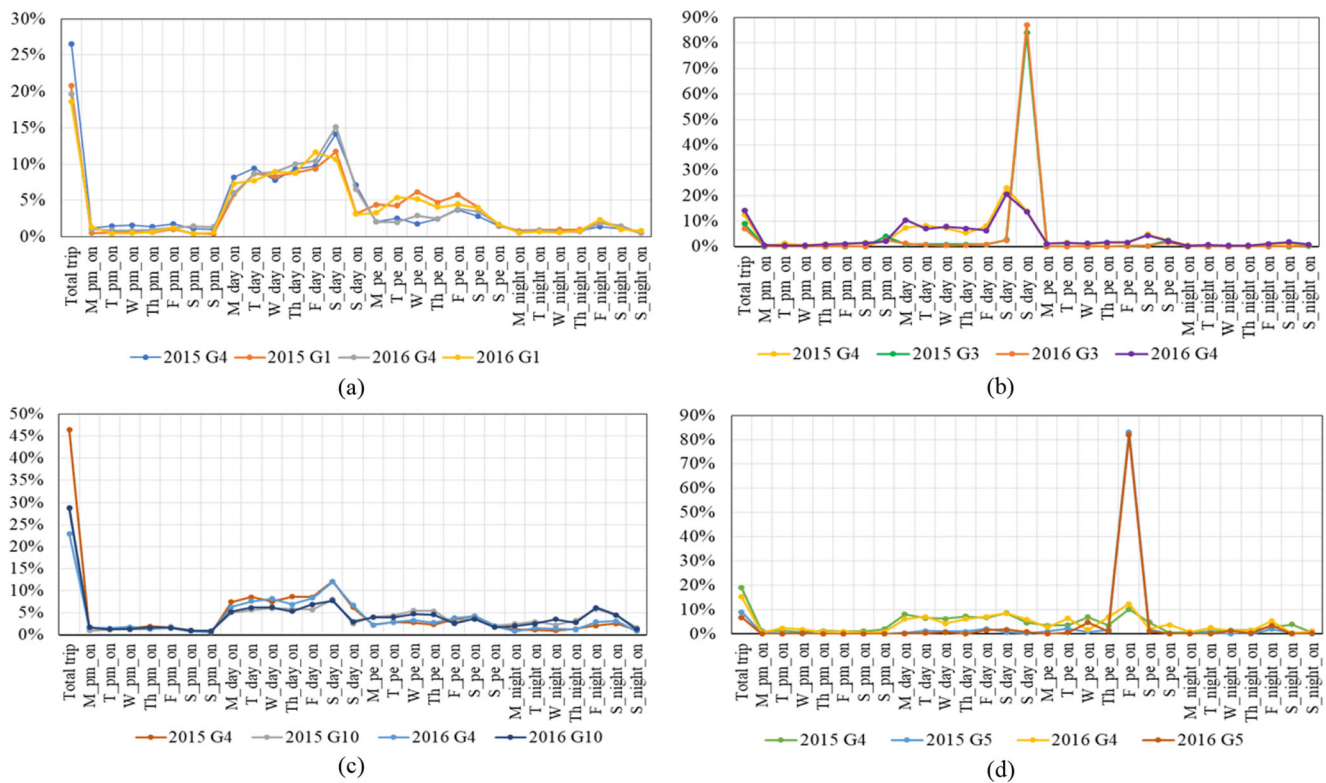Bold text indicates the highest probabilities, respectively.

Fig. 13 Travel patterns for selected groups in 2015 and 2016: comparison of G4 with (a) G1, (b) G3, (c) G5, and (d) G1

versus weekend usage pattern. Passenger mobility patterns were also analyzed by station. Applying cluster analysis to railway use data should be instructive for public transport authorities, who may need to adjust rail services and provide other means of public transport (i.e., buses) at stations with the highest passenger demand (i.e., those near residential and commercial areas). In addition, individuals who strongly favor private vehicles should be encouraged to use public transport more frequently. Transport networks adjust their services to ensure better coverage of the busiest OD pairs. Considering both spatial and temporal aspects of travel may provide a better understanding of public transport demand by location. Future analyses of smart card data should consider other forms of public transport, such as bus services, to provide greater insight into human mobility in Shizuoka City and other cities.

The cluster membership analysis revealed that changes therein typically involved only one or two clusters. The regular commuters showed stable cluster membership patterns, either staying within the same cluster or moving to a similar one, whereas infrequent travelers significantly increased their public transport use over time. Despite some changes in individual passenger cluster memberships from year to year, the number of passengers in each

cluster remained stable. This result could facilitate decision-making by travel authorities as it pertains to re-designing a network in responses to changes in passenger travel behavior patterns. Services for regular commuters should be adjusted to cover any increase in the number of trips made by infrequent travelers.

There were some limitations to this study. First, although railway usage was analyzed by time of day and day of the week, changes in by season were not considered. The academic year in Japan starts in April and ends in March, and travel behavior of a significant number of graduating students would likely show seasonal variability. Analyzing the data according to the academic year, or by 6- versus 12-month periods, could be instructive. The time series analysis based on the finest temporal scale allowing for meaningful data aggregation could refine our analytical model of long-term variations in mobility patterns. Moreover, the smart card data used in this period may be biased, and when the penetration ratio gets higher, the study should check whether the mobility pattern is stable or not. The largest limitations of this study relates to those passengers whose smart cards were no longer active but reactivated in the next year; this is of interest and would merit further study if the data can be obtained in a way that respects passengers' privacy.

References

1. Briand, A.S., Côme, E., Trépanier, M., Oukhellou, L.: Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Trans. Res. Part C: Emerg. Technol. **79**, 274–289 (2017)
2. Chen, Z., Fan, W.: Extracting bus transit boarding stop information using smart card transaction data. J. Modern Trans. **26**(3), 209–219 (2018)
3. Deschaintres, E., Morency, C., Trépanier, M.: Analyzing transit user behavior with 51 weeks of smart card data. Transp. Res. Rec. **2673**(6), 33–45 (2019)
4. He, L., Agard, B., Trépanier, M.: A classification of public transport users with smart card data based on time series distance metrics and a hierarchical clustering method. Transportmetrica A: Trans. Sci. **16**, 56–75 (2018)
5. Kim, M.K., Kim, S.P., Heo, J., Sohn, H.G.: Ridership patterns at subway stations of Seoul capital area and characteristics of station influence area. KSCE J. Civ. Eng. **21**(3), 964–975 (2017)
6. Kurauchi, F., Schmöcker, J.D.: Public transport planning with smart card data, vol. 17, pp. 114–130. CRC Press, New York (2017)
7. Mohamed, K., Côme, E., Baro, J., Oukhellou, L.: Understanding Passenger Patterns in Public Transport through Smart Card and Socioeconomic Data. UrbComp, Seattle (2014)
8. Mohamed, K., Côme, E., Oukhellou, L., Verleysen, M.: Clustering smart card data for urban mobility analysis. IEEE Trans. Intell. Transp. Syst. **18**(3), 712–728 (2016)
9. Nishiuchi, H., Chikaraishi, M.: Identifying passengers who are at risk of reducing public transport use: a survival time analysis using smart card data. Trans. Res. Procedia. **34**, 291–298 (2018)
10. Pelletier, M.P., Trépanier, M., Morency, C.: Smart card data use in public transport: A literature review. Trans. Res. Part C: Emer. Technol. **19**(4), 557–568 (2011)
11. Viallard, A., Trépanier, M., Morency, C.: Assessing the evolution of transit user behavior from smart card data. Transp. Res. Rec. **2673**(4), 184–194 (2019)
12. Welch, T.F., Widita, A.: Big data in public transportation: a review of sources and methods. Transp. Rev. **39**(6), 795–818 (2019)
13. Zhao, X., Wu, Y.P., Ren, G., Ji, K., Qian, W.W.: Clustering analysis of ridership patterns at subway stations: a case in Nanjing, China. J. Urban Plann. Dev. **145**(2), 04019005 (2019)
14. Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G.: Variability in regularity: mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PLoS One. **11**(2), e0149222 (2016)
15. Zou, Q., Yao, X., Zhao, P., Wang, Z., Yang, T.: Measuring Retiming Responses of Passengers to a Prepeak Discount Fare by Tracing Smart Card Data: a Practical Experiment in the Beijing Subway. J. Adv. Trans. (2019)

Rattanaporn Kaewkluengklom obtained a master's degree in civil engineering from Khon Kaen University, Thailand. She is currently pursuing a PhD in civil engineering at Gifu University, Japan, based on smart card data as it relates to transport networks.

Fumitaka Kurauchi is a Professor in the Faculty of Engineering at Gifu University, Japan. His research interests include travel behavior based on dynamic traffic information, resilient transport network design and analysis, and traffic management.

Takenori Iwamoto is a manager at Shizuoka Railway Co. Ltd., Japan. He used smart card data for his PhD thesis, concerned demand forecasting in the context of marketing for Shizuoka Railway Co. Ltd.