Internet Journal of Society for Social Management Systems Vol.12 Issue 2 sms19-6192
ISSN: 2432-552X
ORIGINAL ARTICLE

Analysis on the Change of Travel Interval Pattern Characteristics using Smart Card Data

Hiroaki Nishiuchi[1*], Makoto Chikaraishi[2]

[1] School of Systems Engineering, Kochi University of Technology
185 Tosayamadacho-Miyanokuchi, Kami, Kochi 782-8502, JAPAN

[2] Graduate School for International Development and Cooperation, Hiroshima University
1-5-1 Kagamiyama, Higashi Hiroshima, Hiroshima 739-8529, JAPAN

[*]E-mail: nishiuchi.hiroaki@kochi-tech.ac.jp

Abstract**:** In recent years, an increasingly aging society, a reduction in population, and an outflow of young people to large cities have become common issues in rural cities in Japan. Moreover, the decrease in the number of public transport passengers and the resultant loss of income for public transport companies are serious problems in such cities. However, the role of public transport remains important, because public transport is a mode of transportation that is suitable for an aging society that is experiencing a reduction in population. Therefore, public transport authorities and related organizations require marketing strategies that maintain public transport services for vulnerable users. However, such strategies are always discussed based mainly on information related to income and cost, without consideration for the nature of trips made by public transport passengers. According to interviews with employees of public transport authorities in Kochi City, the outcomes of such discussions always result in plans for service cuts, such as cuts to the number of routes and reductions in tram and bus frequencies. Public transport authorities also said that it is difficult to win back passengers once they stop using public transport after a cut in services. In order to discuss the above issue, understanding the behavior of public transport passengers and the characteristics of public transport usage based on transport data is the first step in gaining fundamental information that is vital to any such discussion in rural cities. The goal of this research is to understand historical change of travel patterns using smart card data. In particular, the present paper focuses on the historical pattern of the change in the public transport use interval. Smart card data, which were collected for almost one year in Kochi City, Japan, is analyzed. Based on the results, there are specific patterns that show that the frequency of public transport use is gradually decreasing.

Keywords: public transport, smart card data, usage interval, time series data

1. Introduction

In recent years, an aging society, a reduction in population, and the migration of young people to large cities are common issues in rural cities of Japan.

Received: 2019.07.31/ Accepted: 2020.05.19

In addition, a decrease in the number of customers and the resultant loss of income for public transport companies are serious problems for rural cities in Japan. However, the role of public transport remains important, as public transport is the most suitable for an aging society experiencing a reduction in population. Since Japan's public transport has generally been operated by private sectors under an independent accounting system, public transport authorities and related organizations need to attract a certain number of users in order to sustain public transport services. If the number is not sufficiently high, the level of service is often reduced. In practice, strategies for improvement of the services are discussed, but the decision is based mainly on information related to income and cost, without consideration for trends of trips made by public transport passengers. Interviews with employees of public transport authorities in Kochi City (conducted by the authors in January 2014) revealed that service cuts had been made to the number of routes and trip frequencies of trams and buses. Public transport authorities also say that it is difficult to win back passengers once they stop using public transport following a cut in services. In order to discuss the above issue, understanding the behavior of public transport passengers and the characteristics of public transport usage is the first step in gaining fundamental information vital for relevant policy.

Monitoring devices, such as sensors that automatically count the number of boarding and alighting passengers have been installed in public transportation in many cities[1)-3)]. Although data from these devices can be used to analyze changes in and variations of the number of passengers, it is not directly possible to analyze the origin-destination (OD) of passengers. Due to this limitation, there has recently been a growing body of literature using data from smart card fare-collection systems in order to directly measure OD between boarding and alighting stops. In particular, many researchers have analyzed the trip patterns and trip regularity of passengers using smart card data in order to understand passenger behavior[4)-12)]. Smart card data have also been analyzed to provide details of passenger behavior for marketing purposes[13),14)]. Such information is fundamental in understanding the passenger usage characteristics of public transport from a marketing strategy point of view and thus is important for the revitalization of public transport in rural cities in Japan. Most marketing strategies are seeking ways to increase demand, but measures to stop the reduction of public transport use have steadily become more important within a declining population.

With the goal of contributing to the revitalization of public transport, Nishiuchi *et al.* analyzed variations in time series data of OD between tram stops (tram OD) in Kochi City using the data from DESUCA, Kochi's public transportation smart card[15)]. Using a state-space model, the OD volume was divided into trend components, daily variation components, and autoregressive components. The results show the temporal rhythms of tram passengers, such as fewer passengers on Mondays and more passengers on Fridays. In addition, Nishiuchi *et al.* analyzed the impact of the reduction in tram services in Kochi, implemented on 1 November 2012[16)]. They developed a model by which to estimate the probability of a change in the number of passengers after 1 November 2012 by applying a Cox hazard model using smart card data. The results showed that OD pairs from a suburban area to a city center could maintain the number of tram users, even after a reduction in the level of service. However, the study was conducted based on an aggregated number of OD passengers and therefore does not show what types of passengers may change their frequency of public transport use, although analyzing smart card data at a disaggregated level can help public transport authorities identify suitable targets for revitalization measures. In particular, it is necessary to identify the types of passengers who are at a higher risk of reducing public transport use in order to avoid a loss of passengers. Moreover, Nishiuchi and Chikaraishi used a Cox hazard model to identify factors affecting the risk of reducing public transport use by using smart card data[17]. There are two objectives: (1) identifying the impact of the reduction in tram services in Kochi implemented on 1 November 2012, and (2) analyzing the trip characteristics of passengers who might reduce their use of public transport.

Applying Cox hazard modeling is required in order to set the definition of timing of target phenomena occurrence, such as the survival duration after taking medicine. However, in the abovementioned research, we had difficulty in defining the public transport use stop for each passenger based on smart card data. The authors then defined the public transport use by deciding a threshold value by statistical aggregation from smart card data in the abovementioned study, which did not state with certainty that the definition was correct. Therefore, these phenomena should be clarified through fundamental analysis.

In order to address the issues mentioned previously, the present paper describes the tendency for passengers' daily public transport use to change using smart card data. In particular, the present paper focuses on the time series tendency of the interval of public transport use based on smart card data. The goal of the present paper is to define public transport use stoppage as a phenomenon that can be observed based on smart card data.

2. Study area and DESUCA smart card

The DESUCA smart card became available on January 25, 2009 and can be used on trams and buses in Kochi Prefecture. Kochi Prefecture is located on Shikoku Island in the southern part of Japan. The population in residential areas within 500 m of a tram stop has been estimated to be approximately 154,000 (Kochi Prefecture, 2011)[18]. The tram network in Kochi City consists of an east-west line and a north-south line, which cross at Harimayabashi Tram Stop, where passengers can transfer between the two lines. The smart card payment system is accepted by the Tosa Electric Railway Co., Ltd., Tosaden Dream Service K.K., Kochikenkotsu, Inc. rail services, and buses operating over a wide area in Kochi Prefecture (Kenkohokubukotsu Co., Ltd., and Kochikenkotsu Inc. in Susaki, Suginokawa, and Yusuhara). These cards serve as commuter passes and can be issued anonymously with personalized data recorded on these cards so that the cards can be returned to the owner if lost. Five types of cards are issued, including a card for children up to elementary school age and "Nice Age" cards for users of 65 years and older. Cards available for adults include personalized and non-personalized cards, as well as a special card for disabled passengers.

The data used in the present study are acquired from DESUCA records collected around one year from October 1st, 2014 to September 30th, 2015. In the 365 collection days, the number of weekdays was 246 days and the number of holidays, including Saturdays, Sundays, and national holidays, was 119. The item for data analysis from the smart card data is shown in Table 1, and the number of unique IDs for the entire dataset is 55,784. Figure 3 shows a histogram for the number of appearance days of each passenger in one month. In the present paper, we assumed that one unique ID represents one specific

Internet Journal of Society for Social Management Systems Vol.12 Issue 2 sms19-6192
ISSN: 2432-552X

passenger if we observe the same card ID in our database. Moreover, the proportion of DESUCA card users compared to the total number of passengers, including passengers who paid fares by cash, is approximately 70% according to an interview survey of the public transport authority by the authors.

Figure 1 clearly shows that most passengers traveled one or two days in one month on average. On the other hand, there are groups of passengers who traveled around 20 days in one month. Therefore, a high-frequency passenger group exists in the study area.

3. Analysis of usage interval change

3.1 Outline of the analysis

The purpose of the present paper is to grasp the type of pattern of the interval of public transport use days in one-year time series data that exists, and to identify the insight that public transport authorities should consider in investigating marketing strategies for passengers who at risk of public transport use stoppage. In order to obtain knowledge for this purpose, the point that should be considered is discussed by conducting a cluster analysis to form several groups of public transport passengers.

Table 1 Outline of smart card data

Data Collection Hours	05:00 to 23:00
Variables measured/recorded	Date: Year, Month, Day Card ID: Provided for all DESUCA cards (always the same once given) Type of Card: With/Without Registration, Handicapped Passenger, Child Passenger, Elderly Passenger Time (Boarding): Hour and min when boarding tram Time (Alighting): Hour and min when exiting tram Tram Stops: Name and code of each stop Distance between Tram Stops
Bus/Tram Companies	Tram: Tosa Denki-Testudo Bus: Tosa Denki-Testudo Tosaden Service Kochi-ken Kotsu Kenko Hokubu Kotsu

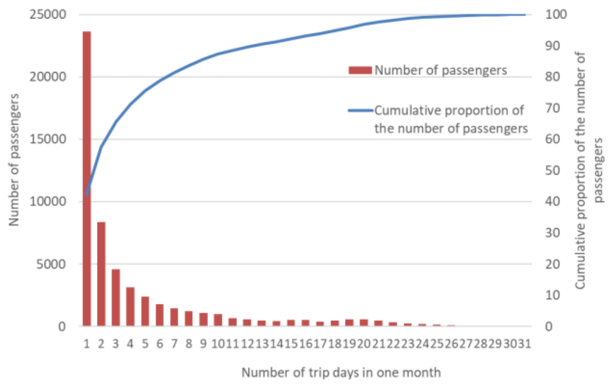

Figure 1 Distribution of the number of trips per month

ID	2014/10/1	2014/10/2	2014/10/3	2014/10/4	2014/10/5	2014/10/6	2014/10/7	2014/10/8	2014/10/9
6051	1	1	0	1	1	0	0	0	0
6052	0	1	1	1	1	0	0	1	0
6053	0	0	0	0	0	1	1	1	0
6054	0	0	0	0	0	0	0	0	0
6055	0	0	0	0	0	0	0	0	0
6056	0	0	0	0	0	0	0	0	0
6057	1	1	1	0	0	0	0	1	1
6058	1	1	0	1	1	0	0	0	1
6059	1	1	0	0	0	0	0	1	0
6060	0	1	0	0	0	0	0	0	1
6061	0	0	0	1	1	1	1	0	1
6062	0	0	0	0	0	0	0	0	0
6063	0	0	0	0	0	0	0	0	0

1 (red): use, 0 (white): not use

Figure 2 Dataset of public transport use days

The interval of public transport use day in this analysis is aggregated by a simple rule to determine whether a target passenger appears (travels) on thatday during the data collection period. The present study grasps the change in the interval of the travel days for each passenger. This aggregation is applied to all passengers (55,784 IDs). Figure 2 shows an example of data aggregation. The horizontal axis indicates the data collection period, and the vertical axis shows the card ID. A 1 in a red cell indicates that a passenger appears on the day, whereas a 0 in a white cell indicates that the passenger does not appear. In the present paper, the characteristics of time series change in each cell are shown visually. However, it is impossible to directly visualize this change for all passengers. Therefore, non-hieratical cluster analysis is conducted in order to simplify the tendency of the time series characteristics of the interval of public transport use days. The data of appearance of each day (0 or 1 in each cell) are used to cluster the passengers. Thus, 365 data for each passenger are used as input data in the clustering analysis. Therefore, non-hieratical cluster analysis is used in the present study. The present paper discusses whether there are passenger groups that have specific characteristics in the pattern of the interval, based on the results of clustering analysis. In particular, the present paper assumes five patterns of the change in the public transport use day interval, such as 1) low-frequency passengers who appeared randomly, 2) high-frequency passengers who appeared continually, 3) passengers who suddenly stopped or started the use of public transport, 4) passengers whose interval gradually became narrower day by day, and 5) opposite characteristics with pattern 4). Therefore, as the hypothesis of the present research, public transport authorities should grasp and investigate public transport use promotion or other marketing strategies for passengers belonging to group 1) or group 4) by keeping or narrowing their public transport use day interval.

3.2 Number of clusters for non-hierarchal cluster analysis

Analysts have to decide the number of clusters in case non-hierarchal clustering analysis is applied. Most studies made this decision subjectively for the convenience of the analysts. On the other hand, there are several methods by which to decide the number of cluster quantitatively. The present study also decides quantitatively the number of clusters by applying the elbow method for k-means cluster analysis and can decide the number of clusters. The elbow method is used to investigate k-means clustering on the dataset for each range of the number of clusters. The elbow method then calculates the value of the sum of squared errors (SSE) and draws a line graph of each SSE value for each number of clusters set by the analysts.

Figure 3 shows the SSE values for each number of clusters by calculating collected DESUCA data, as shown at Figure 3. When the number of clusters is decoded by the elbow method, the change of the SSE value is the focus. In the case of Figure 4, the gap of the SSE value becomes smaller and smaller by increasing the number of clusters. In particular, the gap of the SSE value appears smaller when the number of clusters is five. In the elbow method, the number of clusters should be five in our DESUCA dataset. Therefore, in the next section, the characteristics of the pattern of public transport use day interval on each cluster are discussed.

3.2 Results of the analysis

Table 2 describes the number of passengers who have been classified into each group. Clustering analysis was conducted by SPSS statistics. The present study sets the number of clusters to five. Therefore, 55,748 passengers are divided into each cluster in five groups. From Table 2, the highest number of classified passengers is in cluster 4 by 38,171 IDs, the proportion of which is 68%. The

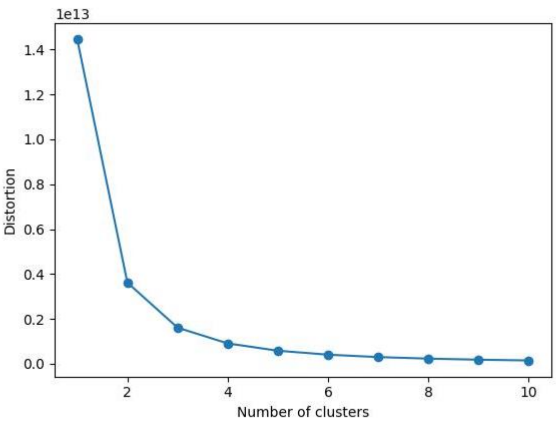

Figure 3 Sum of squared errors for each number of clusters

Table 2. Number of passengers in each cluster

Cluster	1	2	3	4	5
Number of passengers	2,207	7,854	5,081	38,171	2,470

second largest number of passengers (7,854 IDs) belongs to cluster 2, and the third largest number of passengers belongs to cluster 3 (5,081 IDs). The number of passengers that belong to cluster 5 is 2,470 IDs, and the number of passengers that belong to cluster 1 is 2,207 IDs.

Next, the characteristics of each cluster are clarified. Figures 4 through 8 shows the time series change data of public transport use day interval for each cluster for 400 IDs. The horizontal axis in these figures indicates the period of data collection, which is from 1 October 2014 to 30 September 2015, and the vertical axis indicates 400 IDs. In these figures, a passenger traveled on a certain day if the cell is red and not on a day if the cell is white. The characteristics of each cluster are shown visually in the following.

Cluster 1, which has 2,207 IDs, represents the passengers who started to frequently use public transport from April. Moreover, some passengers started to use public transport from the middle of the period, before April. Therefore, this cluster can be considered as having new customers and high-frequency public transport passengers. Public transport authorities may increase the number of such passengers. This cluster can be considered to be made up of passengers who suddenly start to use public transport (group 3 in Section 3.1).

Cluster 2, which has 7,854 IDs, represents passengers who use public transport most days. In addition, cluster 3, which has 5,081 IDs, represents passengers who frequently appear most days other than the passengers in cluster 2. The interval of public transport use days is shorter than that for the passengers in cluster 2. Therefore, passengers who are divided into clusters 2 and 3 can be described as moderate- or high-frequency passengers. Public transport authorities may taking care them to keep their interval to try not decrease public transport use of those passengers. In addition, clusters 2 and 3 can be considered as having high-frequency passengers (group 2 in Section 3.1).

On the other hand, cluster 4, which is the largest group by 38,171 IDs, has a different tendency. The passengers in this cluster traveled frequently around October, but the interval of public transport use days (red) gradually decreased and disappeared after April. Public transport authorities should consider investigating a specific promotion scheme for this type of passenger before passengers stop using public transport. In addition, cluster 4 can be distinguished as passengers who suddenly stop using public transport (group 3 in Section 3.1). However, this group may include random passengers. Therefore, additional and detailed clustering should be considered.

In cluster 5, most of the passengers suddenly stop using public transport. The timing of the stoppage is around the end of March. In Japan, the fiscal year starts from April. These passengers may change their lifestyles at the end of March. Actually, there is no way to keep passengers who change their home address, for example, passengers who change their home address to enroll in school or start working at a company outside of Kochi prefecture. However, the above tendency does not hold for all of the passengers in this group. Therefore, a more detailed analysis under the point of view of spatial and temporal trip patterns is recommended as further tasks to determine the difference in their behavior characteristics between the period before April and the period after April. In addition, this cluster is categorized into a group of passengers who suddenly stop using public transport (group 3 in Section 3.1).

Internet Journal of Society for Social Management Systems Vol.12 Issue 2 sms19-6192
ISSN: 2432-552X

Based on the analysis results described above, passengers who has the trip pattern such as variations in public transport use days interval is possible to exist. Therefore, the present paper shows the possibility of identifying the types of passengers who may stop using public transport by focusing on the change of public transport use day interval.

However, the analysis conducted in the present paper is only used for the dataset of passengers' appearance in smart card data. Therefore, a more detailed consideration of the dataset, such as consideration of the pattern of used tram and bus stops, the timing of their travel, and the day of the week, is necessary.

4. Conclusions

The present paper analyzed the change of the interval of public transport use days for each public transport passenger using smart card DESUCA data, which is collected in Kochi City. In particular, the present study focuses on the travel behavior of public transport use stop. Passengers have been classified by k-means cluster analysis based on the dataset of the public transport use days' interval for each passenger. The analysis results described the existence of passengers who are low-frequency users, and the interval becomes wider and wider as time passes. In order to maintain public transport services in rural cities in Japan, it is important to monitor the daily trip change of these passengers and to investigate individual public transport use promotions. On the other hand, the reason why the interval of some passengers becomes narrower as time passes should be identified. The characteristics of the passengers who have a sudden change

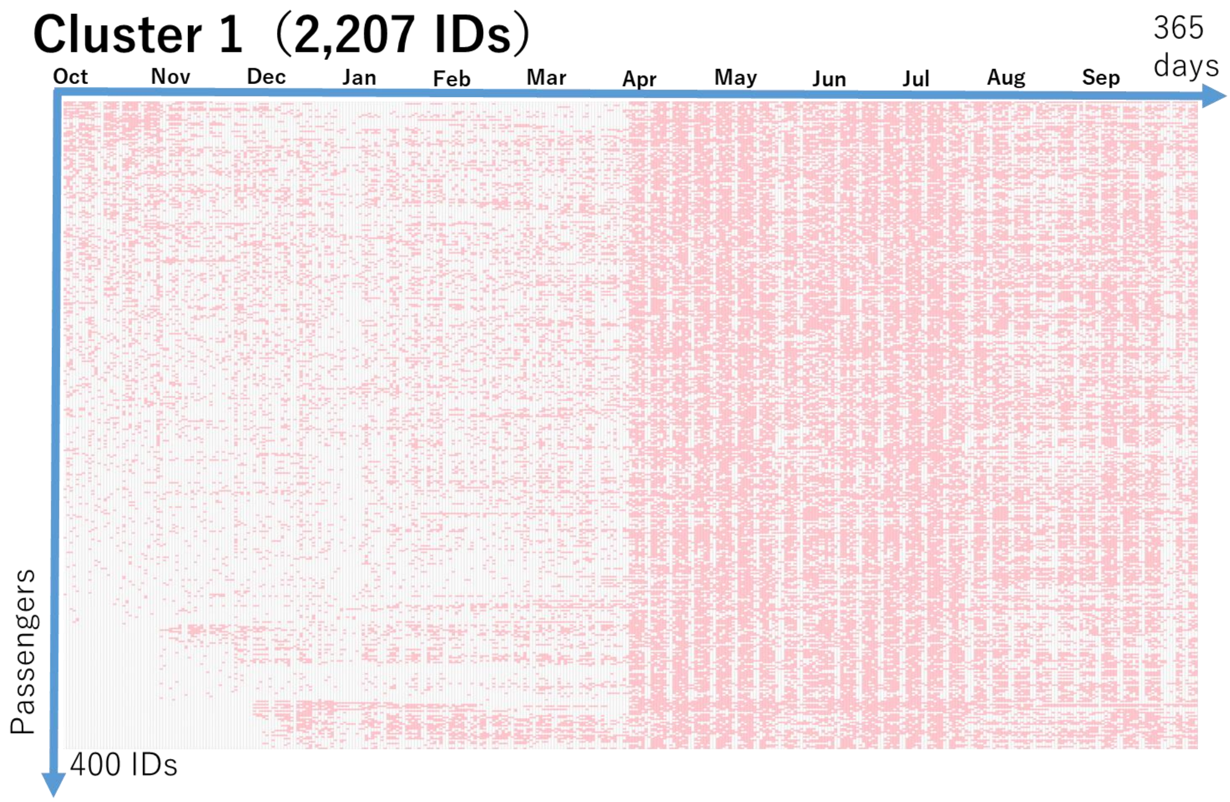

Figure 4 Visualization of public transport use days of cluster 1

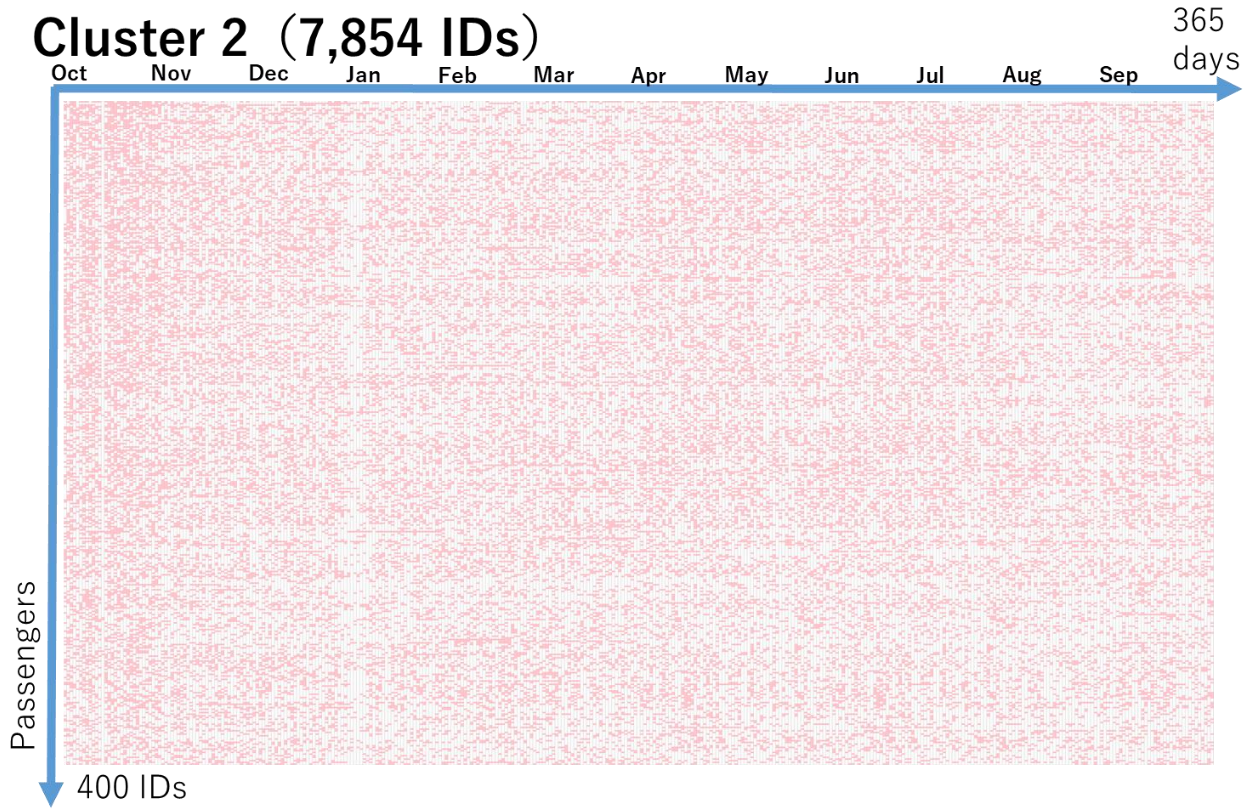

Figure 5 Visualization of public transport use days of cluster 2

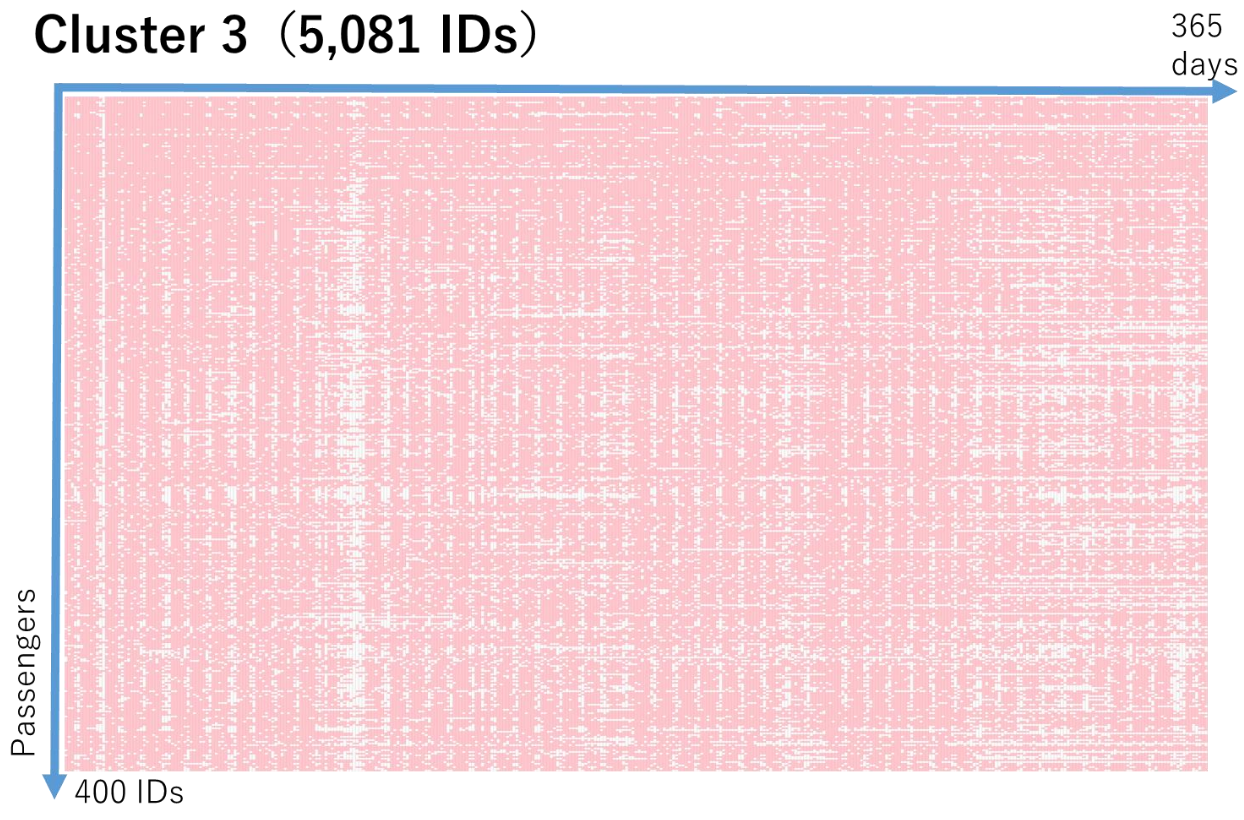

Figure 6 Visualization of public transport use days of cluster 3

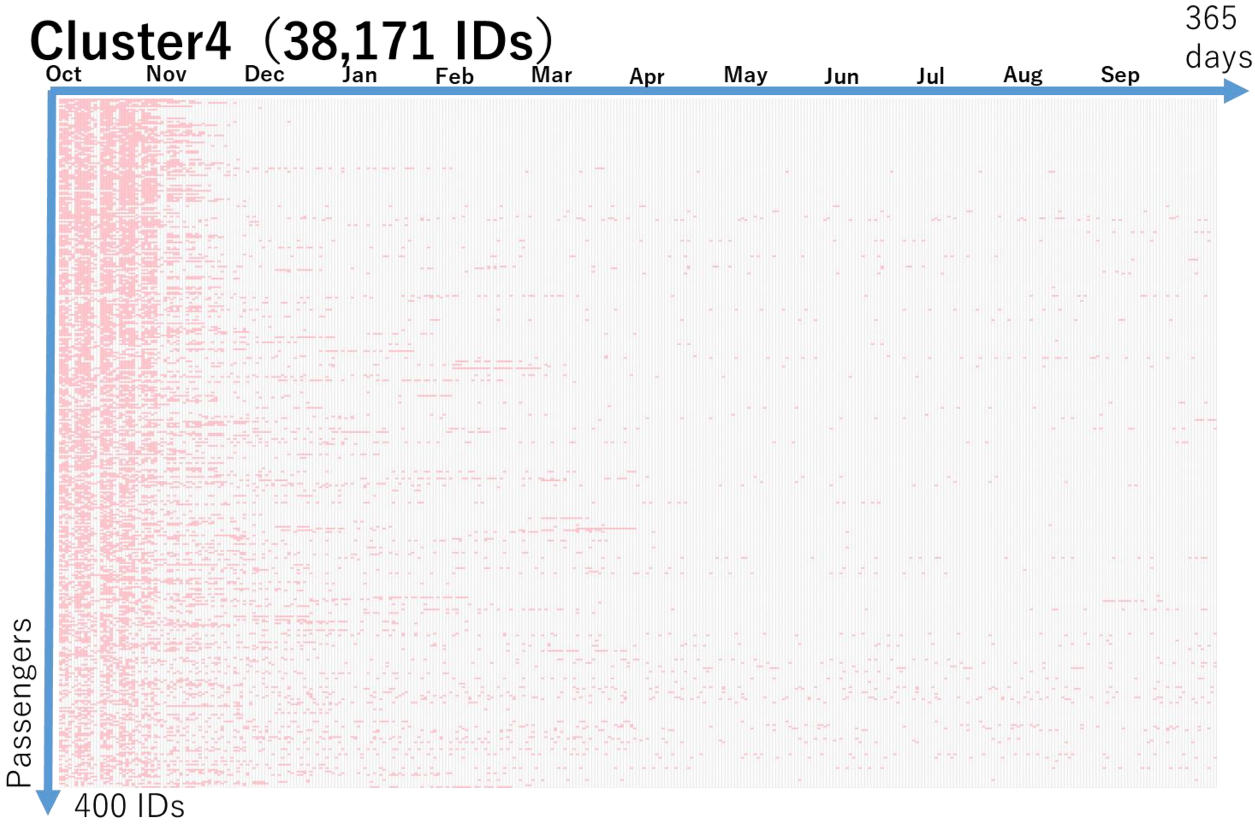

Figure 7 Visualization of public transport use days of cluster 4

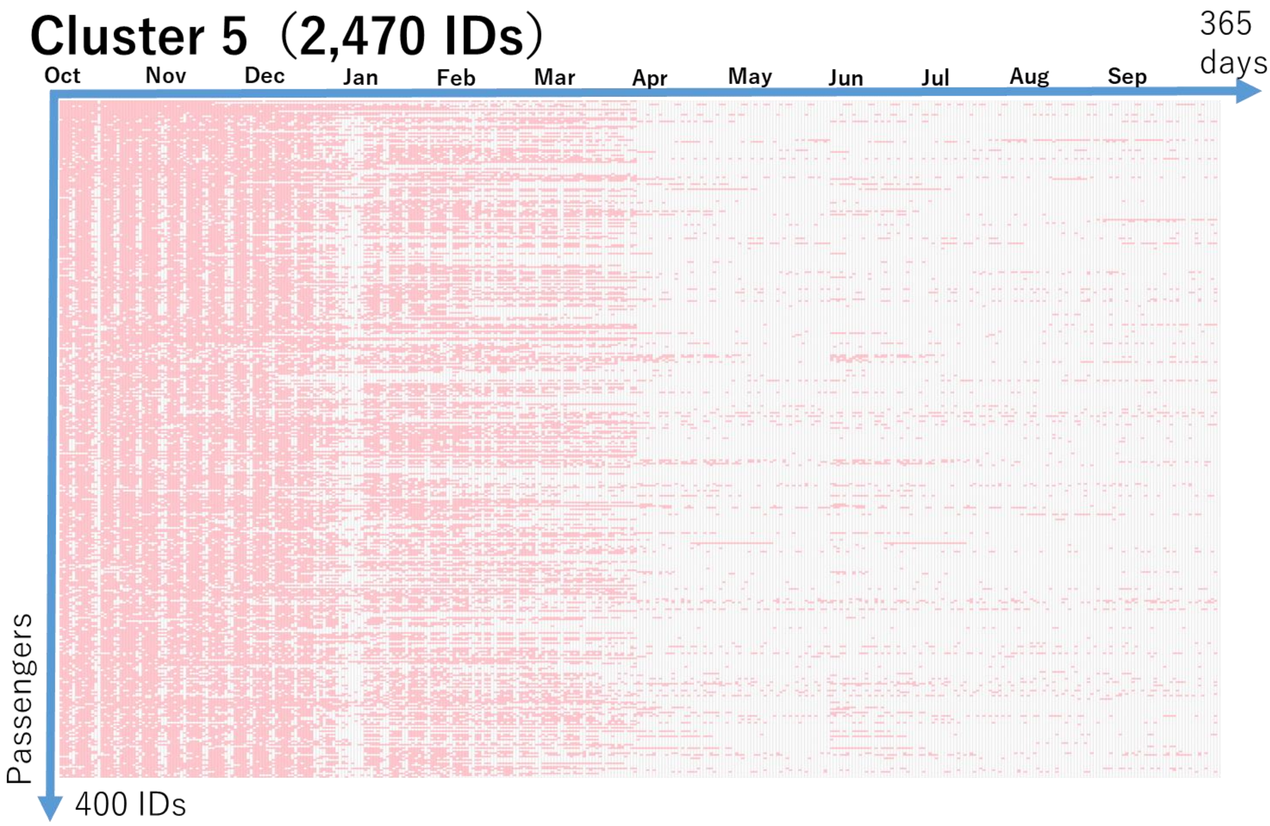

Figure 8 Visualization of public transport use days of cluster 5

ISSN: 2432-552X

in usage frequency should be analyzed. Moreover, these tendencies are considered to be strongly related to passenger attributes. Therefore, detailed analysis for these passengers is also recommended using information from smart card data, such as the type of the card, the day of the week, and the time of day.

In the future, a detection methodology should be developed for the change of public transport use stop using not only smart card data but also other related passenger information. The reason why passengers stop using public transportation in rural cities in Japan should be clarified. Public transport marketing in a rural city should be investigated from both academic and practical points of view.

Acknowledgements
We would like to thank Tosaden-kotsu Co. Ltd. for providing the dataset used in this research. The present study was supported by JSPS KAKENHI Grant Number 17K06605.

References

1) Tamin, O.Z., 1997. Application of transport demand models for inter-regional vehicle movements in West-Java (Indonesia). *Journal of the Eastern Asia Society for Transportation Studies*, 2(3): 903-916.

2) Kita, H., and Tsukioka, S., 2005. OD-Trips estimation of bus passengers based on getting-on-and-off volume data. *Infrastructure Planning Review*, 22:145-152 (*in Japanese*).

3) Zhang, C., and Teng, J., 2013. Bus dwell time estimation and prediction: A study case in Shanghai-China. *Procedia - Social and Behavioral Sciences*, 96: 1329-1340.

4) Chapleau, R., and Chu, K.K.A., 2007. Modeling transit travel patterns from location-stamped smart card data using a disaggregate approach. *Proceedings of World Conference on Transportation Research*, 11. CD-ROM.

5) Morency, C., Trépanier, M., and Agard, B., 2007. Measuring transit use variability with smart-card data. *Transport Policy*, 14:193-203.

6) Pelletier, M.P., Trépanier, M., and Morency, C., 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19: 557-568.

7) Devillaine, F., Munizaga, M., and Trépanier, M., 2012. Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2276:48-55.

8) Asakura, Y., Iryo, T., Nakajima, Y., and Kusakabe, T., 2012. Estimation of behavioural change of railway passengers using smart card data. *Public Transport*, 4(1):1-16.

9) Nishiuchi, H., King, J., and Todoroki, T., 2013. Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data. *International Journal of ITS Research*, 11(1): 1-10.

10) Ma, X., Wu, Y. J., Wang, Y., Chen. F., and Liu, J., 2013. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, 36: 1-12.

11) Kieu, L. M., Bhaskar, A., and Chung, E., 2014. Transit passenger segmentation using travel regularity mined from smart card transactions data. *Proceedings of the 93rd Annual Meeting of*

the Transportation Research Board, USB.

12) Kusakabe, T., and Asakura, Y., 2014. Behavioral data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies*, 46:179-191.

13) Seaborn, C., Attanucci, J., and Wilson, N., 2009. Analyzing multimodal public transport journeys in London with smart card fare payment data. Transportation Research Record: *Journal of the Transportation Research Board*, 2121, 119–126.

14) Páez, A., Trépanier, M., and Morency C., 2011. Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. *Transportation Research Part A*, 45:640-652.

15) Nishiuchi, H., Kobayashi, Y., Todoroki, T., and Kawasaki T., 2014. A study of variability of the public transportation users using state space model: A case study of tosa electric railway. *Proceedings of the Japan Society of Traffic Engineers*, 34: 497-500 (*in Japanese*).

16) Nishiuchi, H., Todoroki, T., and Kawasaki T., 2015. A Study on Change of Number of Tram Passenger by Survival Time Analysis –Case Study on Tosa-Electronic Railway Company-. *Journal of the Japan Society of Transportation Economics*,58:113-120 (*in Japanese*).

17) Nishiuchi, H., and Chikaraishi, M., 2018, Identifying Passengers Who Are at Risk of Reducing Public Transport Use: A Survival Time Analysis Using Smart Card Data, *Transportation Research Procedia*, 34: 291-298.

18) Public Transport Division in Kochi Prefectural Government, 2011. Business report of measures of verification, investigation and cost-benefit analysis related to public transportation. Kochi Prefectural Government (*in Japanese*).