WCTR 2013 RIO

SELECTED PROCEEDINGS

THE PATTERN ANALYSIS OF HIGHWAY USE USING LONGITUDINAL DATA OF ELECTRIC TOLL COLLECTION SYSTE

KUNIAKI SASAKI, UNIVERSITY OF YAMANASHI, SASAKI@YAMANASHI.AC.JP, 4-3-11, TAKEDA, KOFU-SHI, YAMANASHI, 400-8511, JAPAN

KOTA NAKAZAWA, CENTRAL NIPPON EXPWAY CO. LTD. K.NAKAZAWA.AA@C-NEXCO.CO.JP, 1-14-13 ON-NA, ATSUGI-SHI, KANAGAWA, 243-0032, JAPAN

TAKASHI YAMAMOTO, NIPPON EXPRESSWAY RESEARCH INSTITUTE COMPANY LIMITED, T.YAMAMOTO.AE@RI-NEXCO.CO.JP, 1-4-1 TADAO, MACHIDA-SHI, TOKYO, 194-8508, JAPAN

HITOSHI IGUCHI, NIPPON EXPRESSWAY RESEARCH INSTITUTE COMPANY LIMITED, H.IGUCHI.AA@RI-NEXCO.CO.JP , 1-4-1 TADAO, MACHIDA-SHI, TOKYO, 194-8508, JAPAN

This is an abridged version of the paper presented at the conference. The full version is being submitted elsewhere. Details on the full paper can be obtained from the author.

ISBN: 978-85-285-0232-9

THE PATTERN ANALYSIS OF HIGHWAY USE USING LONGITUDINAL DATA OF ELECTRIC TOLL COLLECTION SYSTEM

Kuniaki SASAKI, University of Yamanashi, sasaki@yamanashi.ac.jp, 4-3-11, Takeda, Kofu-shi, Yamanashi, 400-8511, Japan

Kota NAKAZAWA, Central Nippon ExpWay Co. Ltd. k.nakazawa.aa@c-nexco.co.jp, 1-14-13 On-na, Atsugi-shi, Kanagawa, 243-0032, Japan

Takashi YAMAMOTO, Nippon Expressway Research Institute Company Limited, t.yamamoto.ae@ri-nexco.co.jp, 1-4-1 Tadao, Machida-shi, Tokyo, 194-8508, Japan

Hitoshi IGUCHI, Nippon Expressway Research Institute Company Limited, h.iguchi.aa@ri-nexco.co.jp , 1-4-1 Tadao, Machida-shi, Tokyo, 194-8508, Japan

ABSTRACT

This purpose of this study is to estimate the trip purposes of expressway users using the longitudinal data of the Electric Toll Collection (ETC) transaction data. The ETC data in Japan can identify each car and is collecting longitudinal records of the cars. We treat the ETC transaction records as the trip record and summarize the data into some indexes by each car. We apply the cluster analysis to the summarized index data and categorized 5 distinctive clusters. We could interpret the trip purposes of the three clusters among the five clusters that occupy 63% of the expressway use. We test the effectiveness of the clustering by applying the clusters to each origin-destination pair of the interchanges in the Chuo Expressway. The clusters show a reasonable component ratio on the weekday and weekend and the direction from the mega city. We conclude that two thirds of the trip purposes of the expressway use can be estimated by analysing the trip records of the expressway.
Keywords: Electric Toll Collection, Expressway usage, Pattern Analysis

INTRODUCTION

More than 87% of the expressway users in Japan use the Electric Toll Collection (ETC) in 2012 (ORSE 2012). Especially, in Tokyo area more than 90% of cars use ETC. Since all of expressways equip this system, most of expressway users have been recorded in the ETC. These records are not for any analysis but simply for recording the transactions for collecting tolls. Accordingly, it does not necessarily include all the essential information for travel analysis such as travel purposes, origins and final destinations of trips and the individual attributes of passengers. Nonetheless, the data would be valuable as it contains all

longitudinal records of car users. That is, the data can be treated as a huge record of the repeated use of expressways. It would be helpful for the traffic management if we can retrieve any information such as the purpose of each trip from the data.

Particularly, the trip purpose is one of the important information to understand travel behavior because a lot of researches indicated that the value of time would vary with the trip purpose (DeSerpa 1971, Prasetyo et al. 2003, Turouong and Hensher 1985), and thus the elasticity for toll change would vary by trip purpose. However, the trip purpose is generally difficult to obtain without questionnaire survey that is expensive to survey a lot of users. If we can estimate the trip purposes from the ETC transaction data, we have incentive to use it as the proxy of a questionnaire survey, even though the accuracy is relatively low.
Our focus of this study is to estimate the trip purposes of each expressway user by using the ETC transaction records. For the estimation of trip purpose, we introduce an assumption that there is a representative purpose of expressway use for each user, because we will estimate the purpose of the pattern of repeated use of the expressway. We make some indicators of patterns of expressway use in a year for each individual, which describe the feature of individual expressway use in a year. If we can find some clusters which can be related to a general feature of a purpose of the trip, we can assign a trip purpose to the cluster. This type of research would be even more important because the measurement of traffics or travels is developing along with the development of the information and communication technology (ICT) devices, such as SF (Stored Fare) card, the Smart Phone and so on. However, generally such data is not originally designed for travel behavior measurement. Therefore, certain important information for travel behavior analysis may be absent. If we can obtain a certain useful travel information by analyzing such data appropriately, we can analyze travel behavior multifaceted by using the data from ICT devices.

The utilization of ETC to transportation analysis has started at the end of the 20th century. At first, the privacy issue was (Holdener 1997, Ogden 2001) the main topic concerning the ETC, following by the adoption problem of ETC. The penetrating of ETC use of the other car users followed the privacy issue (Chen et al. 2007, Holguin-Veras 2011, Jou et al. 2011) because prevailing of ETC is expected to reduce environmental problems related to congestion at toll gates. Now the utilization of ETC record would become one of the important topics as a transportation data because the study to utilize the transaction data on public transport use for transportation analysis has also developed using smart card records of public transport (Bangch and White 2005, Utsunomiya et al. 2006, Morency et al. 2007, Kusakabe et al. 2012). Though some of the studies using such automatically collecting data analyzed the use pattern of the public transport (e.g. Utsunomiya et al. 2006), few researches tried to estimate trip purposes of the transaction data. The reason of not estimating the trip purposes must be the identity of trip purpose of each individual during the measurement period. It is true that the purpose of the trip would not be always the same during the period. However, it is also not conceivable that trip maker generates trips of the different purpose repeatedly day by day. The most probable case of expressway frequent use is that the purpose of most of the trips is the same, while a few of the trips are not the same purpose. If we can find any characterizing patterns of expressway use and can guess the main purpose of those patterns of expressway use, that can be the first step to understand the purpose of cars on the expressway.

Thus, this research is to categorize expressway use using ETC transaction data and estimate the representative trip purpose of each individual from that category.

DATA AND METHODOLOGY

The data we use is the complete record of 365-day use in 2007 of the expressways around west-north side of Tokyo which is shown as the red lines in Figure-1. Two main expressways that connect between Tokyo and the third largest city Nagoya are the Tomei Expressway and the Chuo Expressway. We chose the 2007 data because of the stability of expressway toll. After 2008, the toll of the expressway has been changed frequently as the temporal social experiments for the purpose of evaluating the effect of toll change. The demand structure of expressway is affected by the toll system because the drivers can choose alternative routes of non-toll road other than the expressways. Therefore, we think analyzing the 2007 data provides more stable behavioral patterns of expressway use.

Figure 1 - the objective expressway network

The number of the transactions of ETC we use is about 170,000,000 records. Among all types of cars from big truck to small car which are recorded in the data, we use only the data of the passenger cars because the purpose of the trucks' trips is obviously business. Each ETC user is assigned a randomized ID that is common to the same registration of ETC transaction. The total number of the unique ID is about 5,000,000 in this dataset. The most frequent use by one ID in the year is 2697 times, and the minimum is one time. The average of the use is about two times in 2007 and the median of that is about eight times. Though this huge amount of transaction records is hard to analyse because of the difference of information amount on each ID, we arrange indicators of the expressway usage of each ID as shown in Table-1, for summarizing information on each individual. Because these indicators are affected by the expressway use through the year, we assume these contain the information about the expressway use pattern of each ID. Therefore, we use these indicators as the data for analysing that pattern.

Table 1 - indicators of individual use of expressway

Indicator	Summary
Number of uses	The number of uses of the expressway in a year
Weekend use	The frequency of expressway use on weekends and holidays
Weekday use	The frequency of expressway uses on weekdays
Day of week	The frequency of expressway use on the each day of the week
Time of day	The number of use by each time of day

The basic characteristics of the data

We show the basic characteristics of the data in this section. Figure-2 shows the distribution of the expressway use on the day of the week. The vertical axis indicates the total amount of the expressway use of this data. The ratio of weekend use is 34%. The trend of the expressway use increases towards the end of the week and decreases towards the middle of the week.

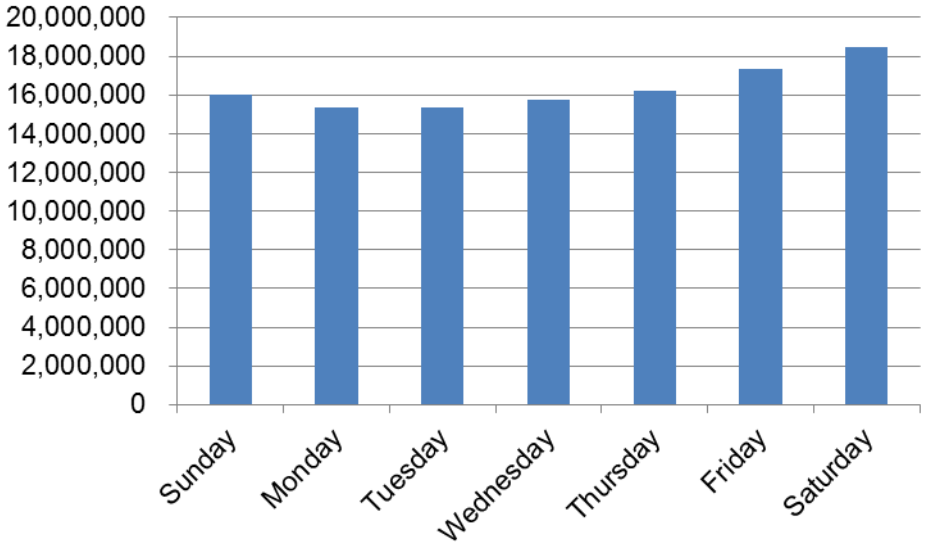

Figure 2 - the number of travel records in the data on the day of week

Figure 3 shows the distribution of the exiting time of cars from the expressways. The distribution of the exiting time shows that the number of cars exiting from the expressway is less at midnight, while there are two peaks, in the morning and the evening. Figure-4 shows a result of the survey about how many trips started for different purposes on each time of day in Iwata City located along the Tomei Expressway. The diversity of the departure time suggests certain characteristics of the trip purpose. That is, the departure time of commuting has keen peak and the return trip to home has also clear peak at evening. We divided the time of a day into four categories to discriminate the use pattern of the expressway shown in Table 2. Similarly, the repeated use of the expressway by an individual would be applicable to the indicator of the trip purpose, because the trips for commuting are necessarily repeated and the trips for business tend to be repeated. That is, the frequency and time of repeated use of the expressway would also be an important source of estimating the trip purpose.

Actually, the purpose of expressway use of an individual is not always the same and would be versatile in the expressway use. Yet, we assumed that the most typical purposes of an individual affect the indicators of expressway use, because it is not conceivable to assume that there are a lot of purposes that are valuable to pay an expensive toll of Japanese expressway on ordinary use. We are going to derive the patterns of the expressway use by applying the indices listed in Table 1 in 2007 data. As mentioned earlier, the numbers of the unique ID and its usage are huge, so that we adopt k-means method to find the patterns from the data. K-means is one of the clustering methods suitable for big data because the algorithm does not require the distance among all elements. However, the analysts have to pre-assume the number of the clusters when calculating them and choose the adequate number of the cluster according to the result of the clustering. Besides, the standardized variables were used, because if the variance of the variables is not the same, it is hard to understand the characteristics of each cluster.

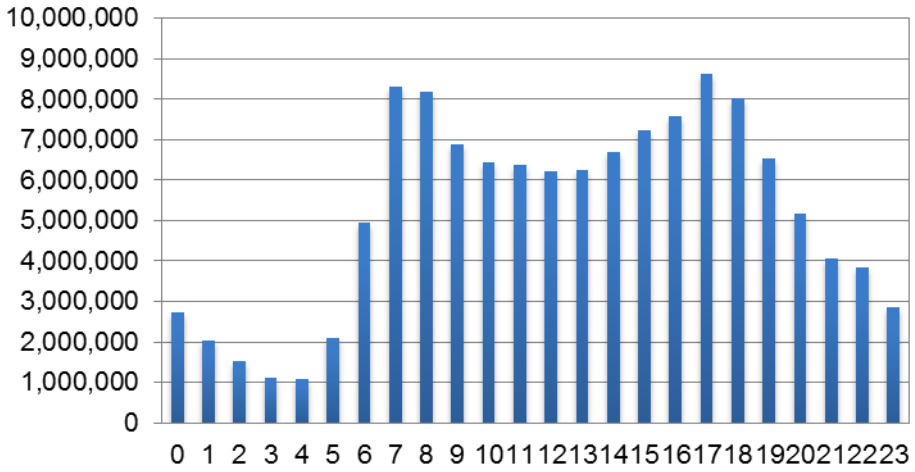

Figure 3 - Distribution of expressway use of the time of day

Table 2 - Time category of expressway use

Period of time	Time	Characteristics
Morning	5~8	Most of Commuters will be finished in this period.
Daytime	9~16	Miscellaneous purposes other than commuting
Evening	17~20	After the peak hour in the evening, that includes the commuters
Others	21-23, 0-4	Midnight, less traffic on expressway

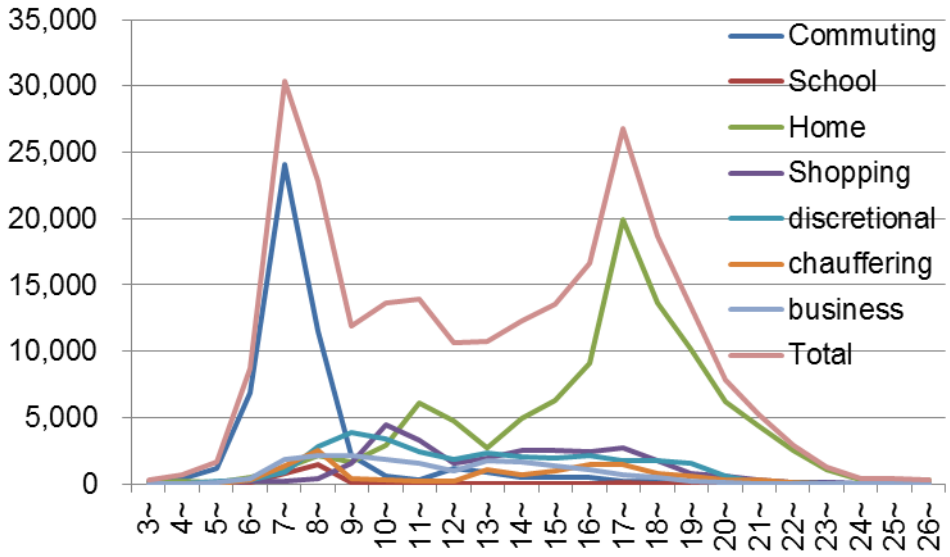

Figure 4 - The trip generation by trip purpose on each time period (Shizuoka Prefecture 2012)

RESULTS AND FINDINGS

We applied the K-MEANS clustering method (Berry and Linoff, 2004) to clustering the about five million individuals by the indices of the expressway use of an individual. The result is shown in Figure 5 and Table 3, Figure 5 indicates the summary of the clusters visually. We chose 5 clusters after we tried 3 to 8 clusters because we thought that the 5 clusters are the most reasonable to discern the patterns of the expressway use. The horizontal axis is the variable of the cluster and the vertical axis is the value of the center of balance on each cluster. From the vertical axis value on each indicator, we can describe the feature of each cluster. The features of Cluster 1 are that the frequency of the use is low, and that weekend daytime use is slightly high. The feature of Cluster 2 is that the frequency of the use is high, especially in the peak hours during the morning and during the evening on weekdays. It is easy to imagine that the major trip purpose of this cluster 2 is to commute. The features of Cluster 3 are that the frequency of use is average during weekdays and the time of day is non-peak daytime. The features of Cluster 4 are that the frequency of use is average on weekends, and that the special feature of this cluster is the time of day of the expressway use, midnight and early morning. The features of Cluster5 are that the frequency of the use is low, so that the characteristics cannot be discerned.

Table 3 summarizes the features of each cluster and shows the ratio in the IDs and the amount of data. Although the Cluster 5 seems dominant due to its 83% occupancy in the IDs, it is not dominant because it occupies only less than 30% of the total amount of the traffic. That is, 17% of ETC users occupy more than 70 % of whole traffic.

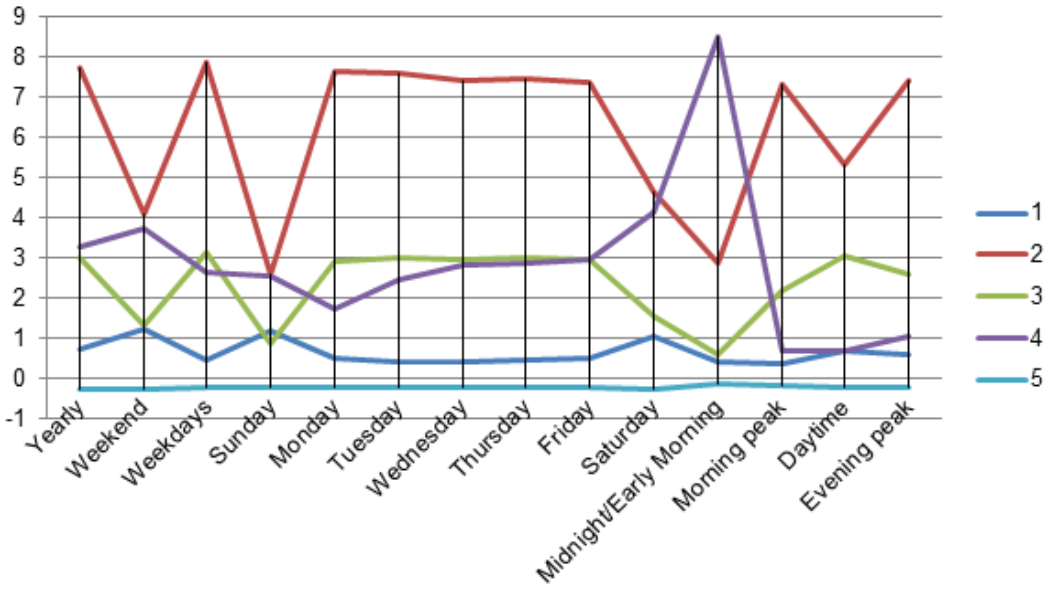

Figure 5 - The center of balance of the final clusters

Table 3 - The features of each cluster and the component ratio on ID and record

Cluster	**1**	**2**	**3**	**4**	**5**
Frequency of use	Low	High	Middle	Middle	Few
Frequent use Day of week	Weekend	Weekday	Weekday	Weekend	-
Time of Day for Use	Daytime	Morning/evening	Daytime	Midnight/ early morning	-
Name of Pattern	**Less Frequency use**	**Commuting**	**Daytime middle use**	**Weekend off-peak**	**Rarely use**
Component ratio on all ID	12%	1%	3%	1%	83%
Component ratio on all records	31%	14%	20%	6%	29%

Application of the result of clustering to Chuo expressway

To see the effectiveness of the proposed clustering, we applied the clustering to the Chuo Expressway (its literal meaning is Central Expressway) users. Though the clustering was constructed from the whole data including two other expressways in Figure 1, we apply the result only to the Chuo Expressway to see its transferability and applicability. The Chuo Expressway is one of the main expressways, connecting Tokyo megalopolis and Nagoya megalopolis. The Chuo Expressway goes through the central mountainous area of Japan and has a branch in the resort area of Mount Fuji. Each toll gate of the Chuo Expressway mostly has some characteristics such as for leisure or for business, because the toll gates of Chuo Expressway are mostly located in the resort areas or the urbanized areas. Therefore, we can investigate the effectiveness of the proposed clustering by applying the result of the clustering to each OD pair. As stated above, because some of the interchanges in the Chuo

Expressway have specific characteristics, we expect some of the interchanges are characterized by a superior cluster. The Figures (Figure 6, Figure 7, Figure 8 and Figure 9) show the component ratio of each cluster on each OD pair. We fix one of the terminals of the trip at the west gate of Tokyo (WGT), because there is a toll gate on the roadway and it is the possible to capture all the cars from and to Tokyo. Thus, the result shows the features of traffics going between the west of Tokyo and the center of Tokyo. Figures (Figure 6, Figure 7, Figure 8 and Figure 9) show the component ratio of each cluster as well as the distance from WGT and the traffic volume between WGT and some other major interchanges. The traffic volume is denoted by the orange bar chart. Each component ratio is shown by line graphs. The left vertical axis denotes the ratio of the cluster and the right vertical axis denotes the traffic volume (PCU/per year). These figures show the specific characteristics of the passenger cars between Tokyo megalopolis and the areas around each interchange along the Chuo Expressway. The component ratio of each cluster is not different between going to Tokyo and going out of Tokyo, that can be confirmed by the difference between Figure-6 and Figure-8 and the difference between Figure-7 and Figure-9. Meanwhile, the difference between weekdays and weekend, which is the difference between Figure-6 and Figure-7 and the difference between Figure-8 and Figure-9, is significant even in the same interchange pair. In general, the component ratio of the trip purpose is different between the weekends and the weekdays, which is supported by other national surveys. This is one of the supporting evidences of indicating a difference of typical purpose of each cluster in this study because the clusters in this study can describe the differences between weekend and weekdays.

On weekdays out of Tokyo (Figure 6), the component ratio of Cluster 2 (commuting) is around 30% at the nearby interchanges. But the number of such users decreases along with the distance from WGT. In the interchanges that are more than 100km from WGT, the ratio of commuting (Cluster 2) is rare, while the ratio of less frequency users (Cluster 1 and Cluster 5) increases. The Interchange 8, 9, 10 and 11 are in the Kofu area where the population is about 500,000 and is the major area along the Chuo Expressway. The ratio of Cluster3 whose trip purpose is estimated mainly as a business is relatively high in this area because some manufacturing industries are located in this area. The Interchange 7 located in the resort area collects the users who belong to the rare use cluster (Cluster 5), even on weekdays.

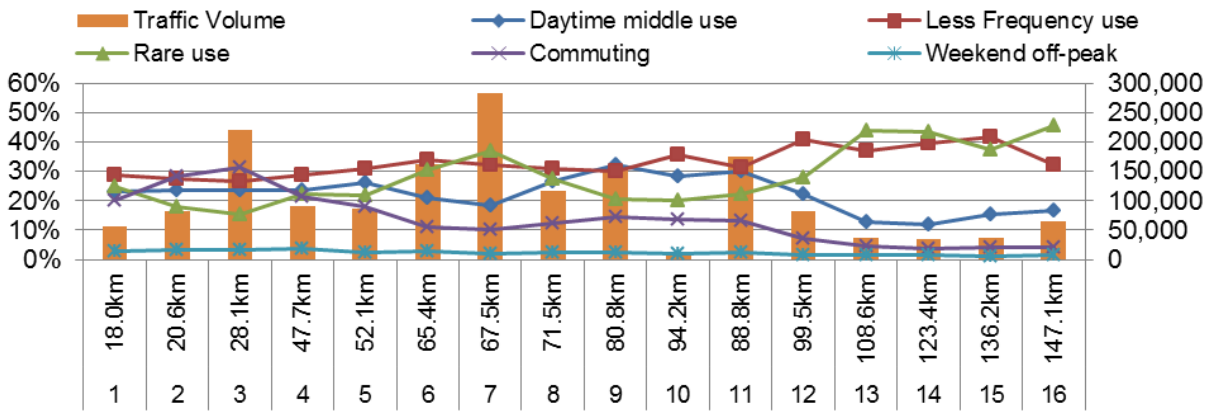

Figure 6 - The traffic volume and component ratio of the clusters on each interchange from WGT (Weekdays)

On weekends, among the traffic going out of WGT, the estimated ratio of the commuting cluster is less than 10% at most interchanges, while it is more than 30% in the nearby interchanges on weekdays. The Interchange 7 is in the resort area and collects more cars which belong to the Cluster 1 and the Cluster 5. These clusters mean "less frequent use" and "rare use". The most typical purpose of both clusters is estimated as leisure from these results, though the other miscellaneous purposes would be included in this cluster. In nearby interchanges, the ratio of "rarely use (Cluster 5)" is smaller than other interchanges. This result supports our estimation of the typical purpose of this cluster as leisure.

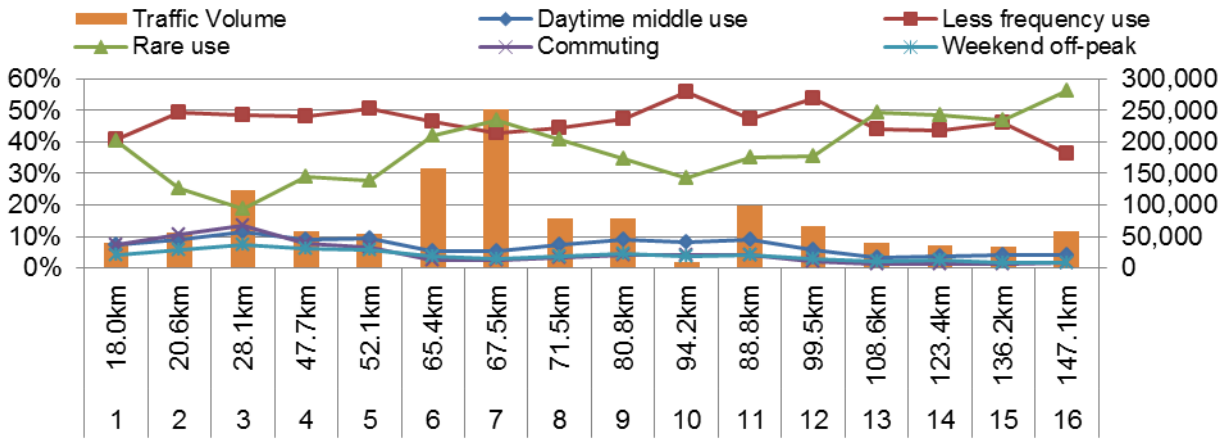

Figure 7 - The traffic volume and component ratio of the clusters on each interchange from WGT (Weekends)

The clustering of traffic bound for WGT is shown in Figure-8 and Figure-9. Because the Interchange 1 is open only from one side, the traffic of Interchange 1 does not exist in these Figures. Both on weekdays and on weekends, the tendency of clustering is almost the same to that of the traffic of opposite direction. The cluster 1 and cluster 5 is the dominating cluster on most of the interchanges. The other three clusters have no more than 35% of the traffic on all interchanges. The sum of the ratio of Cluster 1 and Cluster 5 is more than 90%, especially on the interchanges in the resort area. The main difference between weekdays and weekend is the presence of Cluster 2 and Cluster 3. These clusters would be related to business activities. This suggests that the typical purposes of Cluster2 and Cluster3 are commuting and business, respectively.

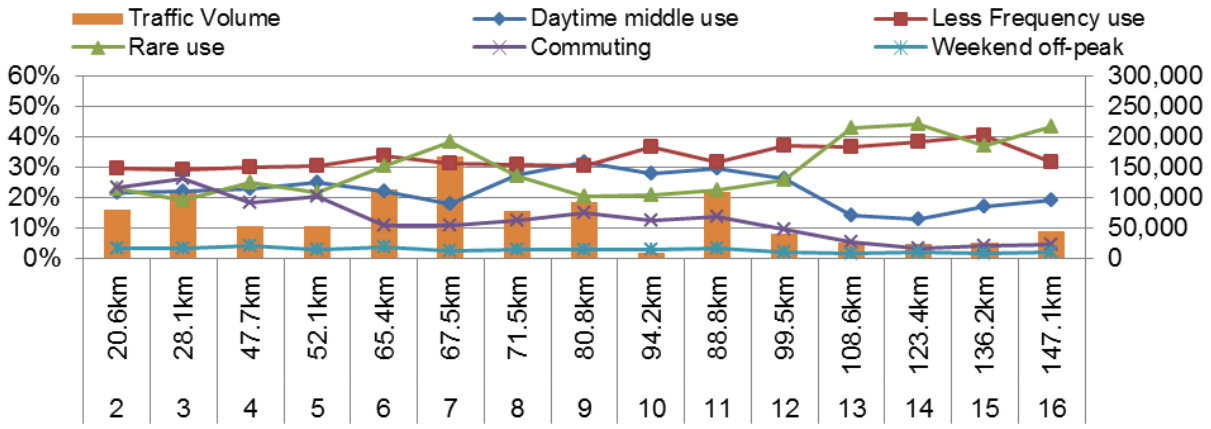

Figure 8 - The traffic volume and component ratio of the clusters on each interchange to WGT (Weekdays)

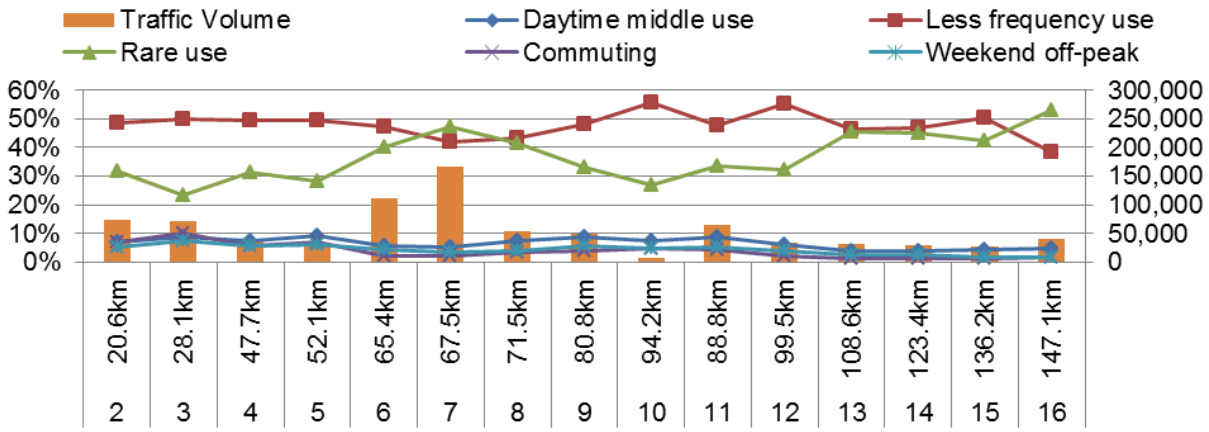

Figure 9 - The traffic volume and component ratio of the clusters on each interchange to WGT (Weekends)

Our examination of discriminating the pattern of the expressway use of the indexes formed in terms of the frequency at the time of day and on the day of the week demonstrates the consistent result with the expected ratio both on weekdays and on weekends. Cluster 2 corresponds roughly to the commuting pattern indicated in the survey at Iwata City, though we cannot prove the transferability of Iwata City Case and Chuo expressway use. The characteristics of Cluster 3 are similar with the business trips. The ratio of these two clusters occupies 34% of the expressway use. Though the ETC transaction record does not have the trip purpose information, those who typically use the expressway for commuting can be discriminated by the one-year record due to the obvious commuting feature. In addition, those who typically use the expressway for business trip were also discriminated by the pattern of the ETC transaction records for a year. The trip purpose is essential to analyze travel behaviors since the value of time or some important indexes of travel behaviors are affected by the trip purpose. Moreover, the rare use cluster (Cluster5) must be leisure trips because the interchange 7 is in the most popular sightseeing area in Japan and collects more users belonging to the Cluster5. It would be supporting evidence that another questionnaire survey in the area around interchange 7 showed that the frequency of visiting this area for leisure is almost one or two (Suzuki et al. 2008). We conclude that the data on the one-year expressway use offers sufficient information about the typical trip purposes of two thirds of expressway use in Japan.

CONCLUSION

The purpose of this study is to analyze patterns of the expressway use using the one-year ETC transaction record. The patterns of the individual passengers are analyzed by the indexes of the summarized expressway use, such as the number of the records and the frequency on the day of the week and the time of the day. The clustering method we use is the K-means method that is designed for analyzing big data. This study also attempted to categorize the pattern of expressway use by using one-year transaction record and to estimate the representative trip purposes of the cluster. The pattern of the individuals was analyzed by the indexes of summarized uses of expressway, such as the frequency of

expressway use and the frequency on the day of the week and time of the day. Especially, we regarded some of the clusters as the trip purpose based cluster. This categorization allows us to understand the differences of the drivers' trip purposes in each day and each section of an expressway. This information is effective when we need to manage traffic volume on the expressway, because some traffic conditions as the value of time vary depending on the travel purposes. We know that the purpose of expressway use is not always the same within the ETC user, however, if we can get knowledge of the ratio of the users who have the typical purpose of the certain section of the expressway, that would be useful when the expressway companies make plans for the demand management strategy.
We conclude that the longitudinal data of expressway use give us the important information about the expressway users' travel purposes. We are also developing a state-dependence type demand forecasting model (Sasaki et al. 2012) and we are going to apply this categorization for understanding the difference of sensitivity to the level of service. However, the weak point of this analysis is to apply clustering to an individual transaction record. A trip purpose of one person is not always the same. The result of our study discriminates the individuals who use the expressway for almost the same purpose as commuting and business. Our future work is to develop a method of estimating the trip purpose one by one by focusing on the variety of the trips on the expressways.

ACKNOWLEDGEMENTS:

This study was supported by three expressway companies in Japan. They provided not only data for analyzing but useful comments. We would like to express our deepest gratitude to the companies.

REFERENCES

Bagchi, M. and P. R. White (2005). The potential of public transport smart card data, Transp. Policy 12(5), 464–472.

Berry, M. J. and G. S. Linoff (2004). Data mining techniques: For marketing, sales, and customer relationship management, 2nd Ed., Wiley Computer Publishing, New York.

DeSerpa, A. C. (1971). A theory of economics of time, Economic J., 81, 828-846.

ORSE (Organization for Road System Enhancement) Japan Home Page, http://www.go-the ETC.jp/english/system/index.html (in Japanese) accessed 31. Oct. 2012.

Holdener, D. J. (1997). Electronic toll collection information: Is personal privacy protected? Proc. of the Seventh Annual Meeting, ITS America, Washington DC.

Morency C, M. Trepanier and B. (2007). Agard, Measuring transit use variability with smart-card data, Transp. Policy, 14 (3),193–203.

Ogden, K. W. (2001). Privacy issues in electronic toll collection, Transp. Res. C, 9, (2), 123-134.

Prasetyo, L., Fukuda, D. Yoshino, H. and Yai, T. (2003). Analysis of Travel Time Saving benefit by Understanding Individual Needs and Value of Activity Time: Case Study of Tokyo and Jakarta, Transp. Res. Record, 1854, 12-21.

Sasaki, K., Nakazawa, K., Yamamoto, T. and Kawasaki, T. (2012) Analysis of Traffic Volume Variation by Bayesian State-Space Model, Traffic Eng., 47(2), 27-32 (in Japanese).

Shizuoka Prefecture, Report on the Person Trip Survey of Saito Urban area, http://www.pref.shizuoka.jp/kensetsu/ke-510a/other.html accessed May 08, 2013

Suzuki, S., T. Tokiwa and K. Sasaki (2008). The validity of the simple questionnaire survey for sightseeing behaviour, Proc. of Infrastructure Planning, 37(4), (in Japanese).

Turouong, T. P. and D. A. Hensher (1985). Measurement of Travel Tie Values and Opportunity Cost from a Discrete Choice Model, Economic J., 95, 438-451.

Utsunomiya M., Attanucci, J., and Wilson, N. (2006) Potential uses of transit smart card registration and transaction data to improve transit planning. Transp. Res. Record, 1971, 119–126.