Proceedings of the Eastern Asia Society for Transportation Studies, Vol.12,2019

Variability in Behavior Regularities of Bus Users based on Long-Term Smart Card Data Analysis

Pai-Hsien HUNG [a*], Kenji DOI [b], Hiroto INOI [c], Yu-Chun CHANG [d], Monorom RITH [e]

[a,b] *Graduate School of Engineering, Osaka University, Osaka, Japan*
[a] *E-mail: hung.pai.hsien@civil.eng.osaka-u.ac.jp*
[b] *E-mail: doi@civil.eng.osaka-u.ac.jp*
[c] *University of Toyama, 3190 Gofuku, Toyama, Toyama, Japan; E-mail: inoi@sus.u-toyama.ac.jp*
[d] *Department of Transportation Science, National Taiwan Ocean University, Keelung, Taiwan; Email: ycchang.ntou@gmail.com*
[e] *Department of Mechanical Engineering, De La Salle University, 1004, Metro Manila, Philippines; Email: rith_monorom@dlsu.edu.ph*

Abstract: Public transportation services are often modified, especially bus route or stop adjustment, which leads to changes in user behavior. A better understanding of dynamic user behavior is important to provide insights to support public transportation planning. This study utilized the whole year smart card data of monthly user behavior pattern in 2015 and 2016. The behavior patterns are adopted from the previous literature, and three behavior regularity variables are proposed including the number of months, the number of behavior clusters, and the number of behavior changes. We can understand the summary of user behavior regularities governed by the three explanatory variables. The result showed that the changes in the regularity of each user group are very informative for the managers to review the services and pricing policies.

Keywords: User Behavior Regularity, User Behavior Clustering, Smart Card, Bus Service

1. INTRODUCTION

Either the service design of public transportation systems or user behavior generally changes according to changes in time and socioeconomic characteristics. Among various public transport systems, the bus service can run without the fixed rail infrastructure or large terminals. It is one of the highly flexible transportation systems because it does not need complicated construction, buses and stops are the minimum requirements for running the service. Moreover, the time needed for changing is less than one month or shorter. Therefore, the managers have to understand the changes in user behavior to design the service more precisely.

Fortunately, in light of the fast growth of the application of the smart card system, bus operators can get the raw data of each transaction such as time, location, and route when users board or alight the bus, etc. A large dataset not only presents an operating performance via ridership calculation but also shows users' behavior information by using an advanced statistical method (Morency et al., 2006; Morency et al., 2007; Bagchi and White, 2005, Zhong et al., 2015). Accordingly, this study proposes a method to obtain user behavior information quickly via smart card data analysis. Moreover, it goes without saying that users'

* Corresponding author.

behavior information is quite an important foundation that helps managers make better service planning for buses.

1.1 Background

The bus system is widely applied as the mass transit mode in the world in view of its construction cost lower than that of subway or mass rapid transit (MRT) and highly flexible service. The bus service is also a low-cost public transportation system as compared to the railway system, and it is often conducted as a dominant public transport mode or an extension of the railway network. Therefore, the managers have to concern with both long-term planning and short-term reaction for changing the overall traffic features, social environment, or user behavior.

Bus network is a wide network contains a large number of users with various characteristics; the managers must have a large-scale investigation about the understanding of user behavior. Therefore, such tasks usually need lots of budgets and human resource, and most managers use the active responsibility for the user behavior investigation, i.e., questionnaire survey. Nevertheless, the past research finds that users may not answer the questionnaire with their own behavior or preference genuinely, and this leads to a bias in planning (Moody, 2016).

Additionally, the users may adapt their behavior according to the contents of the bus service and socioeconomic characteristics, e.g., becoming a student and income growth. Changing behavior is not like the service design, and most users may have that without anticipation. When users acquire new socioeconomic characteristics, they may know how they should react to the new characteristics in the meantime or earlier. On the other hand, users may have new behavior that reacts to slightly long-term changes in preferences. No matter of user behavior transition, most users can only answer the behavior changes in a simple description, e.g., increasing or decreasing, and they cannot identify what kind of behavior they are. Moreover, the managers create behavior clusters in various professional ways in their questions to the respondents; general users may not understand what the real meaning of the professional terms or sentences of the behavior patterns. The managers still need an interview to understand the users' behavior.

The behavior transition is also an important reference for bus service planning. If the managers can obtain the long-term tendency of behavior transition, they can come up with better planning with consideration of the tendency. Nevertheless, the long-term tendency is an arduous task to obtain. The managers must implement a long-term investigation. Such tasks will cost lots of budgets and human resource, yet they may not produce enough samples still. Therefore, a low-cost methodology that can find out the tendency of user behavior transition will have a contribution for the service design.

1.2 Objectives

Smart card system provides detail information about how users use public transportation systems, and the data are an important reference when the managers make or review their planning. A weekly travel profile can be obtained from the smart card data to understand the user's behavior pattern. However, the supply and demand may change as time goes by, the behavior pattern of the same user may vary in time. Therefore, the managers can make a judgment of user behavior more precisely and planning more efficiently if they understand the user behavior transition beforehand.

This study utilized the smart card data of monthly user behavior pattern in the whole

year, and we can obtain user behavior clusters of each month. From the clustering results, three regularity variables of all users could be obtained such as the number of months, the number of behavior clusters, and the number of behavior changes. The number of months shows how many months the user uses the bus service; the number of clusters is that how many behavior clusters the user has, and the number of behavior changes is that how many times the user changes behavior cluster between former and later months. We can understand the variability of the behavior regularity from these three regularity variables in various user groups, and this will assist managers to make more efficient and precise planning.

The remainder of this paper is organized as follows: Section 2 discusses previous studies about behavior clusters and regularity, Section 3 describes the method used to cluster and compute behavior regularity, Section 4 provides a brief description of the case study in Keelung city, and Section 5 concludes the empirical findings of the study.

2. LITERATURE REVIEW

The application of smart card data is extensively discussed since 2005. The data is simple, but it can be extended to various materials for user behavior research in the field of transportation such as time, location, and service information. Although the bus usage data could be obtained via the smart card data in terms of a number of counts, it is difficult to detect trip purpose. It requires a more advanced method to study deeply on information from a larger amount of smart card data (Bagchi and White, 2005).

In the research field of smart card data, a time interval is quite an important variable. Electrical data could be recorded at any time, and different research objectives may influence the appropriate time interval. If the interval is too short, computational loading will be increased too much, and if the interval is too long, some of the data characteristics may be lost. For bus service, a weekly interval is appropriate to present the users' behaviors. The regularity of departure time can be classified into different behaviors. By using the k-means method to cluster weekly data according to each regularity, we can classify users' behavior and derive users' commuting type (Morency et al., 2006).

The transition of user behavior is also important for a better understanding of the user composition in a service. Over the last two decades, the information technology (IT) has a great development, and most public transportation systems provide a service of high flexibility. Users also change their behavior with a response to service changes. The smart card system can provide a stable and detail data source for planning or analysis taking account of the time variable. Such research assists the managers to have a better understanding of user behavior. For example, the smart card data could be analyzed in a minute or hour scale, and the difference of various user groups or areas could be obtained (Zhong et al., 2016). Another research uses Entropy to estimate the user behavior transition in various days from the smart card data and geographic data. It uses the result to estimate and predict the users' behavior transition (Goulet-Langlois, 2015)

Because the smart card system usually stored all the transaction data for many years, it is possible to find the tendency from the long-term data. For the data of the railway system in Japan, the researchers conducted a survey for the users and collected their smart card usage data for more than 5 years. They grouped the users into 9 patterns among 3 groups, including increasing, decreasing, and stable groups. Here, the usage pattern is a summary of monthly usage count in 65 months. The result of user classification told the tendency of the usage count. We will also use the usage count and weekly patterns to identify the different behavior patterns (Li et al., 2018). Another research in Queensland Australia developed a model to

recognizes automatically the user's weekly profile and detects its long-term changes. The weekly profile is a good way to show the individual behavior pattern, and we can obtain the diversity of the individual behavior that could help for predicting the future pattern (Moon et al., 2018). One research in Singapore utilized the smart card data to present the spatial distribution and their difference in terms of years. The result proposed an approach to analyses the long-term impact of new infrastructures and their evolution dynamics (Sun et al., 2015). The two-year data could use the appearance and disappearance to present the life of card usage, and analyze four behavior patterns, including Intensity change in with-in travel pattern, a subtle change in day-to-day travel pattern, structural change in seasonal travel pattern, and evolution in year-to-year travel pattern. The result examined three levels of changes via two indicators related to mobility and activity location (Chu, 2015).

The variability of travel behavior is another research topic to understand the detail of users' behavior. The variability expresses the user behavior more clearly, and better adjustment could reduce the operating cost (Morency et al., 2007). When an employer meets pressure in the company, the manager is not easy to tell the real situation from an interview or their performance. Obtaining the variability of behaviors of one employer could obtain his/her pressure in the workplace. The study proposed the variability that could show a significant correlation of behavior and physical factor with behavior clustering (Okada et al., 2012). With clustering, the variability of regularity will be more clear and easy to understand. A number of behaviors can represent how much a user is regular or changing. The number of behaviors tells the variability of regularity (Guidotti et al., 2018)

As compared with the previous research, we believe the variability of behavior patterns needs a measurement. The managers could understand and evaluate the effect of the policies via the indicators. The researchers also could enhance the prediction of the behavior pattern with more clear information about the description of the individual's behavior. The operators could understand the user's loyalty with the variation of user behavior (Trépanier and Morency, 2010). We use three behavior regularity variables with a simple calculation to describe the behavior transition regularity. Because the long-term data may contain many unknown variables, and they lead to the estimation more complicated, the simple variables can show the regularity clearly and easy to compare the difference of various user groups or areas.

3. METHODOLOGY OF BEHAVIOR REGULARITY CALCULATION

This section discusses how to cluster bus user behavior from the smart card data and calculate behavior regularity. Monthly behavior cluster is the key concept to conduct the continuous cluster transition.

3.1 Dataset

The smart card data was used in this study, and each raw data contains the boarding and alighting information of single trip by the same card. Data items in each row include Bus ID, Route ID, Card No., boarding time, boarding or alighting bus stop name, and card type. We used the one-month data of each user to conduct the clustering calculation. One single month consists of at least 4 weeks of weekdays and 4 weekends. By summing each hour of one month's usage, the weekly profile will be presented. In order to prevent rare users like one time visitors from influencing the main body of the clustering, those who use less than 4 times per month are grouped as the random user cluster.

Usage count of most bus users' behavior is weekly, including weekday and weekend trip shown in weekly profile. Therefore, we consider there exist 168 (24 hours * 7 days) variables in a week and averaged the frequency in each hour (Pas, 1988; Tarigan et al., 2012; El Mahrsi et al., 2014). Figure 1 shows an example of how the IC card usage raw data transfer into weekly boarding profile. For the same card number (same user), by looking at the boarding time section of IC card usage data, we accumulate each boarding record separately according to its boarding hour so that we can understand the specific frequency per hour. Peak hour characteristics, as well as the difference between weekday and weekend, is now easy to define.

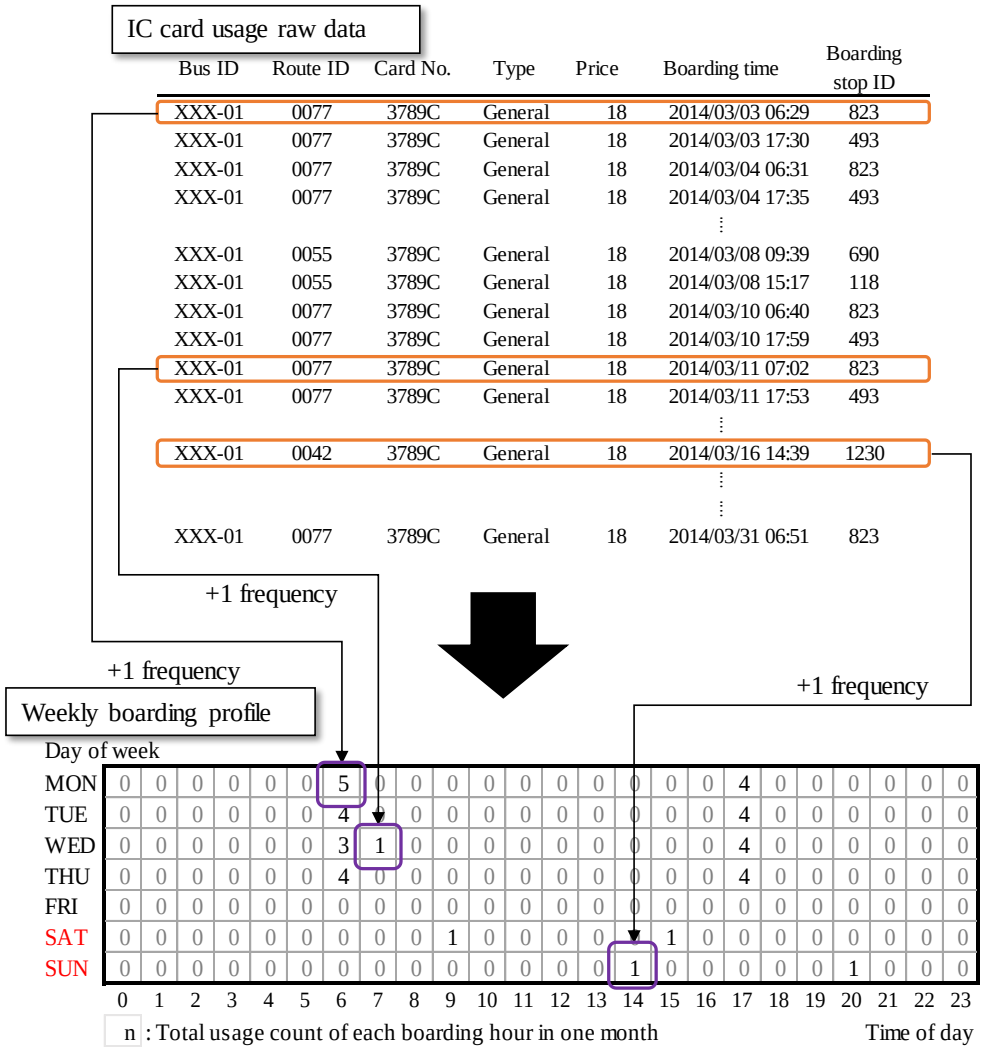


Figure 1. Example of how smart card usage data transfer into weekly boarding profile (Hung et al., 2017)

3.2 Behavior Clustering by Month

The main concept of behavior clustering in this study is to group users with similar behavior and to define bus user behavior decided only by departure time and usage count. For bus service, there are several data columns may not be recorded, i.e. alighting time or the bus stop, because some operators may have different rules about the users using the smart card to pay for the bus service. We use only boarding time and card ID. Other variables were not considered in order to simplify clustering calculation. Although other variables, like weather,

may influence users' decision, users do not change their behavior permanently only by some accidental events. Therefore, we picked up an unaffected month (March 2016) without special events or long holidays to conduct the clustering process. Variables of behavior are departure time and frequency. Therefore, we categorized users' departure at a similar time of the day and near usage count to be in the same cluster.

Figure 2 shows two real examples of users use the bus service in 2015. Both of them use the bus every month, but the behavior patterns (day of using the bus or usage count) are differently observed. For user 1, the minimum usage count for one month is 1, and the maximum one is 96, and the days' distribution (the days the user use the bus service) have no significant pattern. For user 2, the minimum usage count for one month is 6, and the maximum one is 30. User 2 prefers to use the bus service on Tuesday and Thursday, but the patterns vary in different months. According to these two examples, it is necessary to determine the user behavior pattern in each month. Moreover, user behavior regularity is also important for understanding user behavior more precisely.

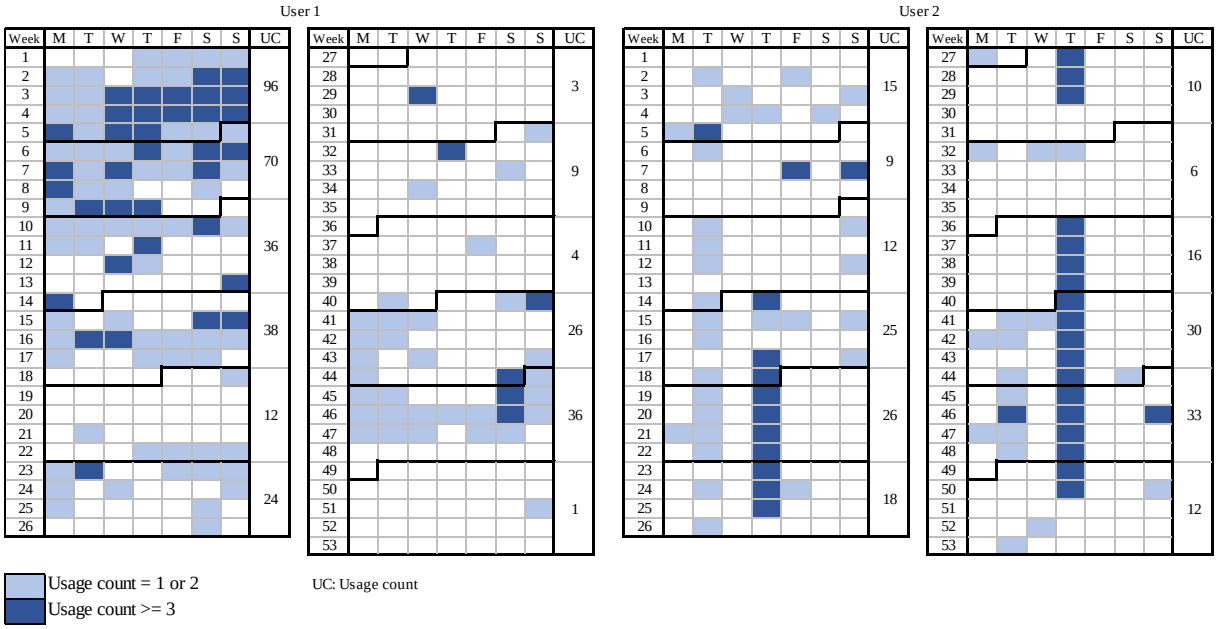


Figure 2. Two examples of the summary of the users use the bus service in 2015

The clustering method used in this study can refer to previous research (Hung et al., 2017). That research uses the EM algorithm (Expectation Maximization Algorithm) to cluster user behavior from users' weekly profile. In the same group, users are with similar behavior pattern, and the users may have a different pattern at a different time. The result shows the user behavior cluster in each month, and the behavior cluster will be null if the user does not use the service in that month. In this study, the null behavior is ignored, and the behavior regularity could be obtained from the user's monthly behavior cluster.

Since accident variables can easily be left out, one simple month is selected to compute clustering parameters. And then, these parameters are used onto other months as well to obtain clustering results of each month. The whole year data now consists of 12 months of data cluster. Figure 3 shows this clustering process.

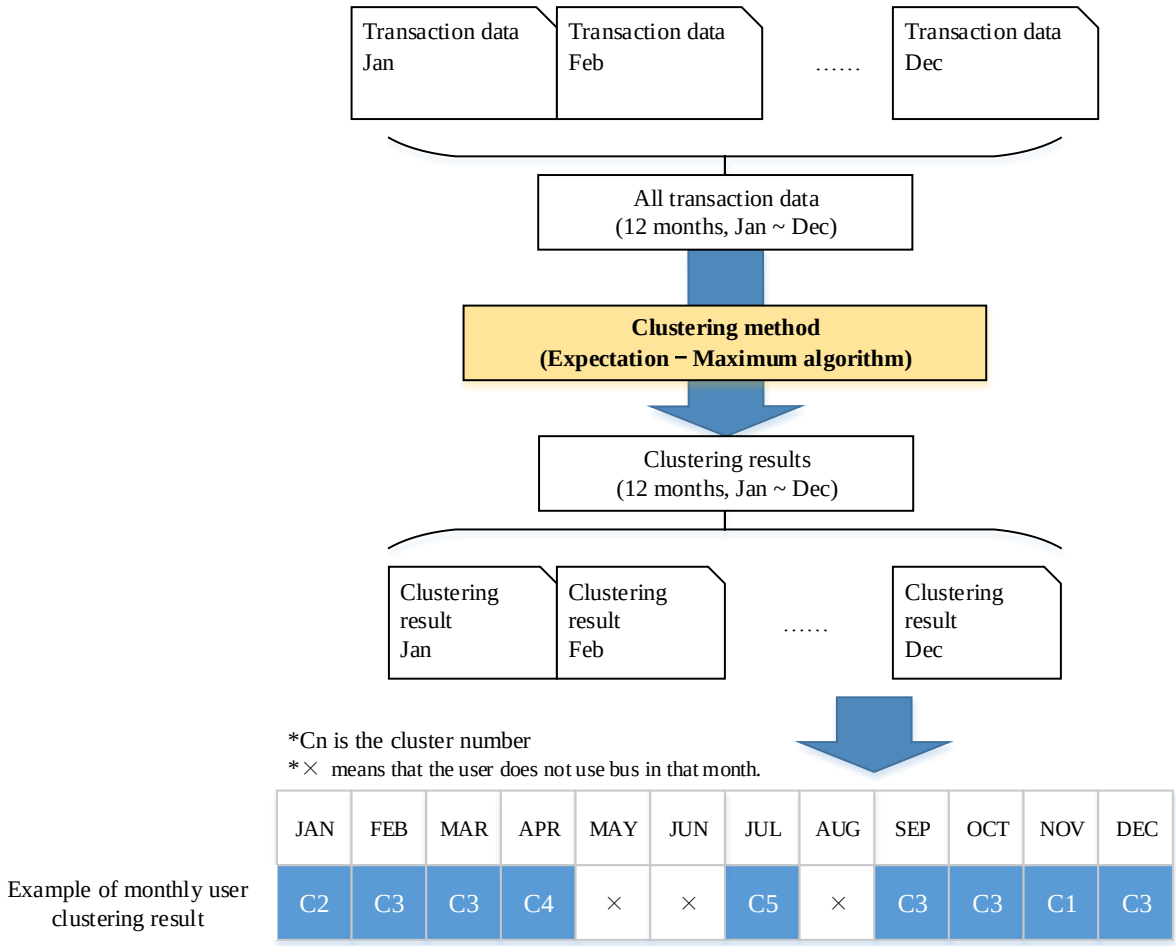


Figure 3. The process of obtaining yearly behavior clustering (Hung et al., 2017)

3.3 Behavior Regularity

The clustering results of previous sections show the behavior clusters in one year of each user who ever uses the bus in that year. This section proposes three behavior regularity variables to describe the regularity of behavior transition. Those three variables show in the following:

1) N_MON: Number of months that the user uses the bus service. It shows that how long the user uses the service and the service will be a major alternative when it is getting bigger whether the usage count is high or not. Zhao et al. (2018) proposed a method to detect the behavior change with consideration of frequency, temporal, and spatial dimension. The time of the individual use the service is a good index to identify the change of behavior pattern. Briand et al. (2017) used the smart card data to represent the year-to-year change of user behavior. The result proposes a good way for planners to understand or simulate the change in user behavior. We use the number of months to represent the more detail change of an individual's behavior. It can also be obtained from general statistical calculation.
2) N_CLU: Number of distinct behavior cluster that the user has. This value shows the difference between user behaviors at a different time. Ferrer i Cancho and Lusseau (2006) shows the number of patterns and their occurrence times could measure the correlation between behavioral events. Agliati et al. (2006) evaluated user-agent interaction with consideration of detection of patterns and pattern complexity. When the value is getting bigger, it means the user is getting more unstable, and it

may show the observable difference of user's demand.

3) N_TNS: Number of behavior transition between the former and latter months. This value tells whether the behavior changed between the former and latter months. Holmqvist et al. (2011) proposed a measurement for evaluating focused versus overview eye movement. In a series of eye movement, they use the transition matrix to find the most frequent transition, and the number of transitions could use for probability calculation, that could apply to enhance the behavior pattern. prediction. Li et al. (2017) also use the number of transitions to enhance the estimation of the driving style. It shows the user behavior is more unstable when the value is getting bigger. Compared with N_CLU, N_TNS emphasizes the behavior change between two months, and the N_CLU emphasize the diversity of the behavior.

Figure 4 shows the process of obtaining the three regularity variables and the user distribution of each two of them. The meanings of the three regularity variables are periodicity (N_MON), diversity (N_CLU), and stability (N_TNS), respectively. All of them can represent different properties of the user behavior regularity. In the example, we found that the three regularity variables could express the variation of user's behavior transition. Here, we do not consider the direction where the behavior changed. Because the direction of the behavior transition has a deep relationship with the behavior patterns, we focus on the variability of behavior transition instead of the variability of behavior patterns in this study. The degree of variability of behavior regularity is vital for the managers to understand their users' behavior.

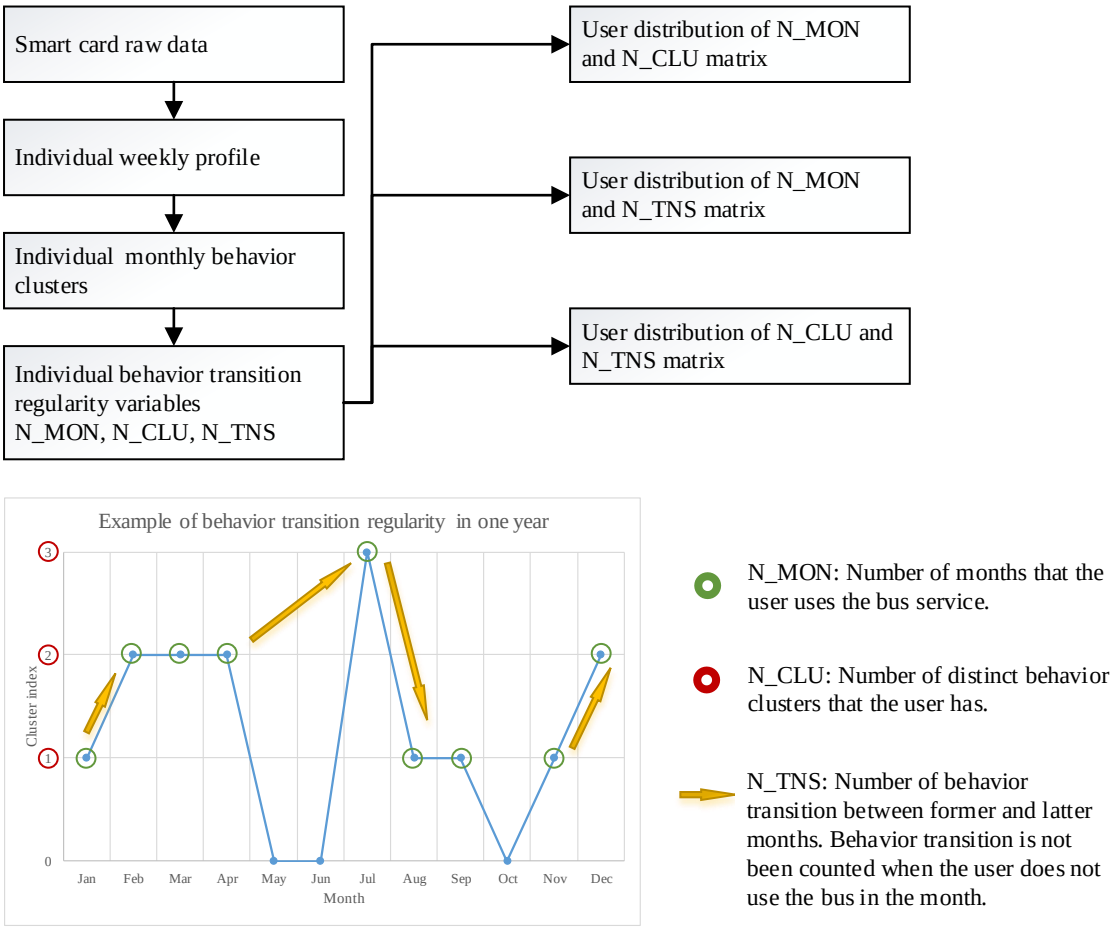


Figure 4. The process of obtaining the three regularity variables

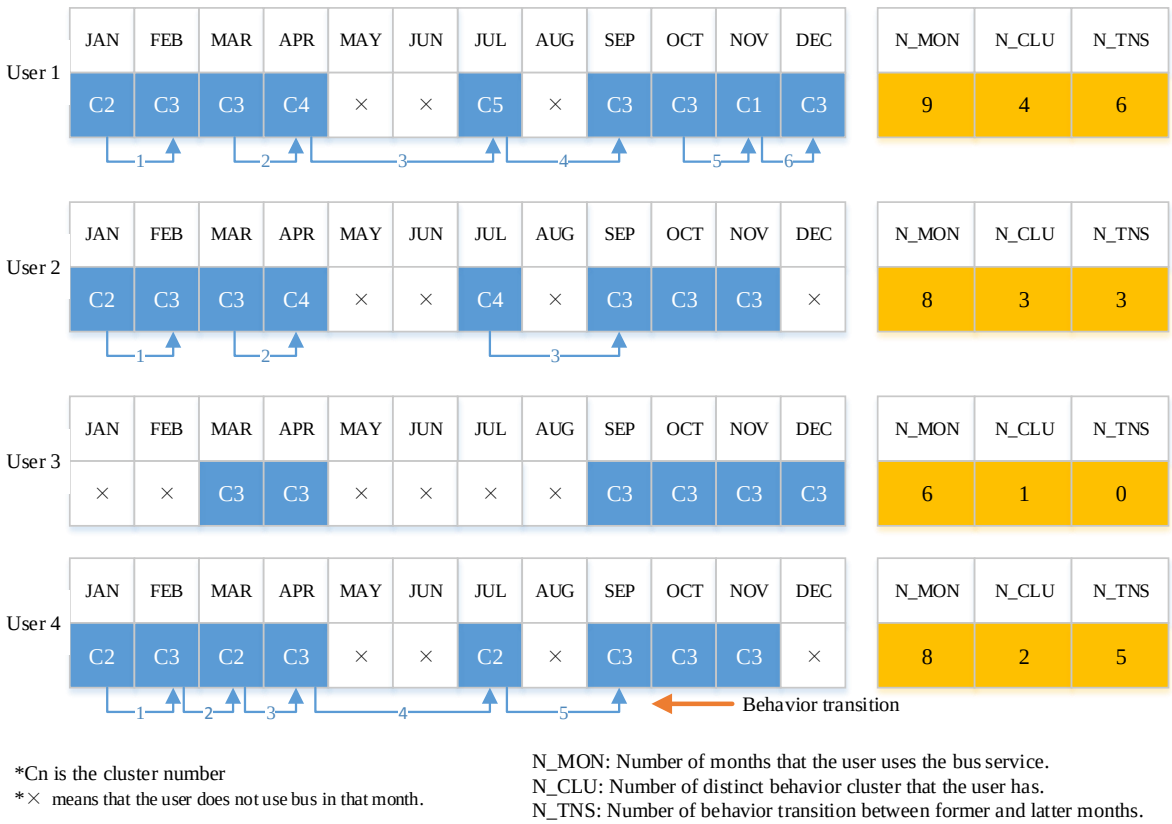
Those regularity variables of one user will satisfy the following conditions.

1) $1 \leq N\_MON \leq 12$

2) $1 \leq N\_CLU \leq N\_MON$

3) $0 \leq N\_TNS \leq N\_MON - 1$

Figure 5 shows an example of the regularity calculation of four samples. In this example, user 1 uses bus service for nine months of one year, and the N_TNS is high because the C3 is not continuous. N_TNS of User 2 is lower because the C4 only interrupt C3 once. Both N_CLU and N_TNS of User 3 are lower because User 3 has only one behavior cluster. User 4 changes behavior cluster frequently even N_CLU value is 2.



*Cn is the cluster number

*× means that the user does not use bus in that month.

N_MON: Number of months that the user uses the bus service.
N_CLU: Number of distinct behavior cluster that the user has.
N_TNS: Number of behavior transition between former and latter months.

Figure 5. Example of behavior transition regularities calculation

4. CASE STUDY OF USER BEHAVIOR REGULARITY

This section uses the data of Keelung city as an example to represent the regularity of bus user behavior via the process in Section 3. The process transforms the smart card data into the variability of regularity in order to understand the tendency of user behavior transition.

4.1 Socioeconomic Characteristics of Keelung

Keelung city locates at the northern part of Taiwan, shown in Figure 6. In Keelung city, 95% of areas are filled with hill and mountain area, and most of the citizens use the railway and public transport to commute to other counties. The land area is 132.76 km2, wherein six bus companies operate on 56 city bus routes. There is a major port for cargo and cruise in

Keelung, and the number of cruise tourists reached 1.4 million in 2016. Keelung city has a few flat areas, one railway service, and various bus services. The mode share of the city bus in Keelung is over 12.4% that is the second highest in Taiwan, after Taipei city. The population is over 370,000, and over 15% are elderly people who are over 65 years old. People cannot park or drive private vehicles properly because of the narrow road space. Visitors and tourists, as a result, prefer to use public transport. Moreover, the port terminal is close to the train station, and the cruise tourists may use train or bus service to transfer.

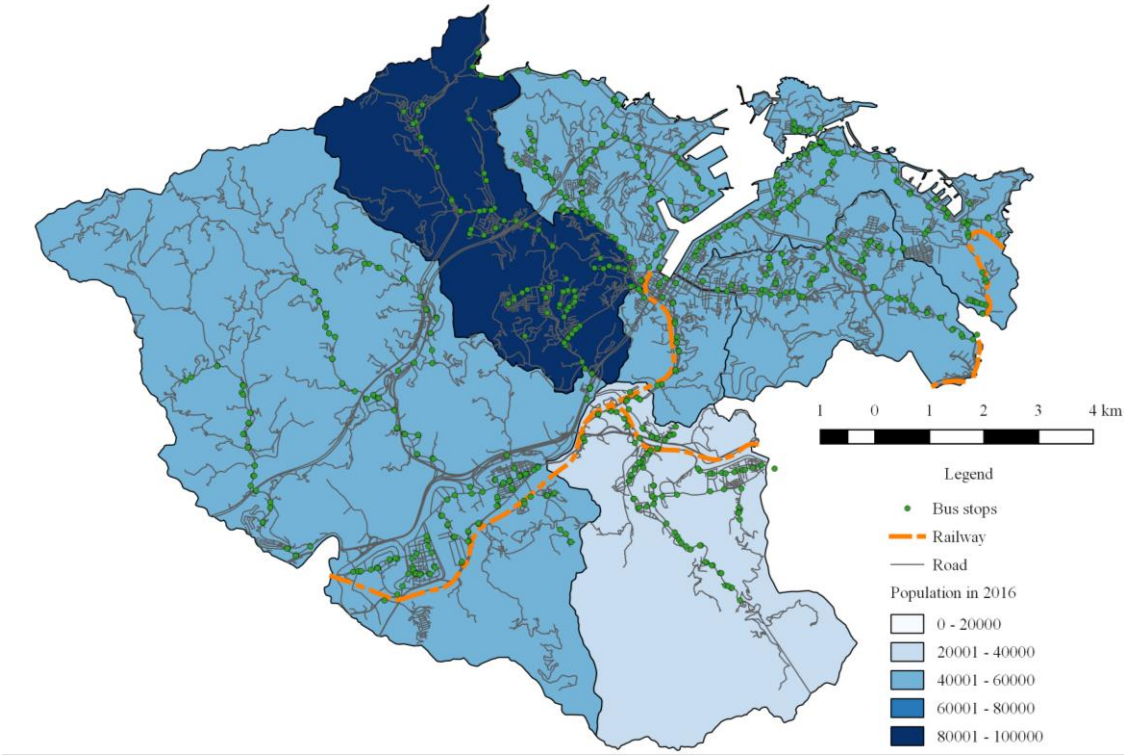
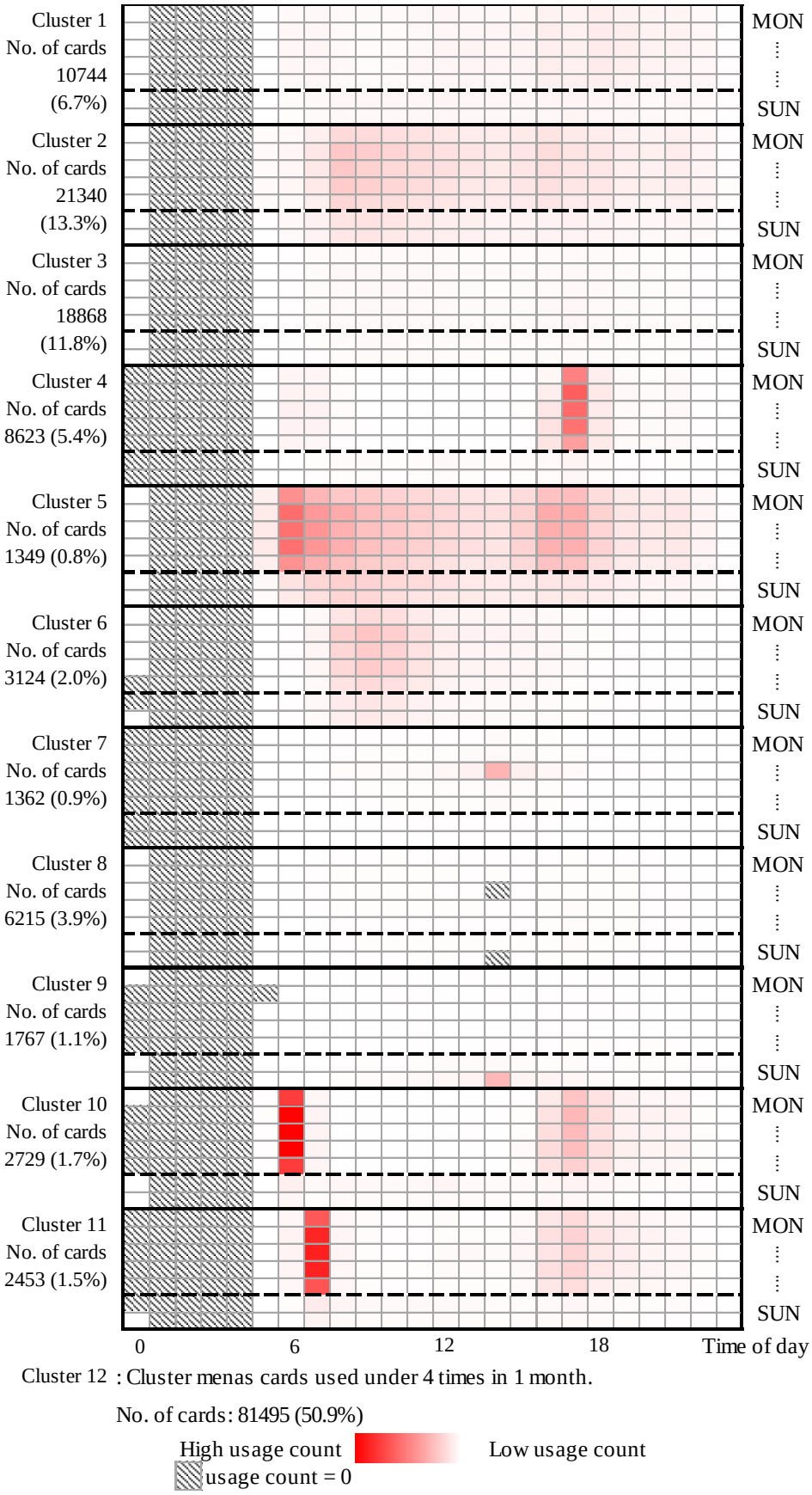


Figure 6. Keelung city map and city bus stop location

4.2 Variability of Behavior Regularity Variables

This study selects city bus service in Keelung city as the study area. There were 856,997 smart cards used in 2015 and 2016 and 35,832,859 transactions. We use the data in March 2016 for obtaining behavior clustering parameters. The EM (Expectation-Maximum) algorithm clusters the data into 11 behavior clusters and the 12th cluster is the users who use bus less than 4 times in that month (Hung et al., 2017). The behavior clustering result shows in Figure 7. There are total 160,069 cards used during that month. Users belong to cluster 4, 5, 10, 11 observably use the bus service at morning or afternoon peak hours. Users belong to cluster 1, 2, and 3 are users tend to use the service at a stochastic time. This result can understand the users' behavior according to the pattern of the using time and usage count. Then, using the cluster parameters to determine the behavior cluster of each user in each month via their weekly profiles, and the cluster of a month will be null if the user does not use the bus in it. After determining all behavior clusters of all users, those three regularity variables of each user could be calculated via the methodology stated in section 3.



Cluster 12 : Cluster menas cards used under 4 times in 1 month.

No. of cards: 81495 (50.9%)

High usage count — Low usage count

usage count = 0

Figure 7. Behavior patterns of all clusters in the selected bus operator (March 2016)

Because the students are one of the major groups of the bus users, and they do not have regular classes in summer and winter for a long-term vocation; therefore they change the behavior certainly. Therefore, the data for behavior regularity analysis should contain at least 12 months to represent the complete regularity. The regularity may have a bias if the period of data is shorter than 12 months.

Besides the regularity in a specific time interval, the tendency of the behavior regularity transition is also an essential reference for the bus service planning. The tendency can assist the managers to understand the influence of the bus policies to the users in the area. When the period of data for regularity analysis is one year, it may need more than 10 years of data to represent the tendency effectively. However, it is not only hard to acquire the long-term data, but it is also more complicated if the other socio-economic variables are considered for the data analysis. Therefore, this study used the same concept with respect to the moving average. We split the data of 24 months into 13 groups. Each group is the data for a length of one year. All groups are of a continuous time period with the first and last month changing. For example, group 1 is from 2015.01 to 2015.12, group 2 is from 2015.02 to 2016.01, and the last one is from 2016.01 to 2016.12. This process can produce data in 13 groups to represent the tendency.

Figures 8 through 10 show the tendency of the three regularity variables and their changes. As apparent from Figure 8, the proportion of users decreases with respect to an increase in N_MON, besides the proportion of N_MON = 12 is about twice the proportion of N_MON = 9 ~ 11. The distribution shows the users have a slight tendency for shifting to a periodic user. The distribution of N_MON = 2 ~ 5 is decreasing, and the first half of N_MON = 6 ~ 8 is increasing, while the last half is decreasing. This information can assist the managers in understanding the distribution of bus user in terms of time length. An increase in N_MON = 12 means the users have an intention to use the bus service.

As illustrated in Figure 9, the user distribution of N_CLU represents the diversity of the behavior cluster. There are over 97% of users have four behavior clusters or less. It means the user' behavior is not so complicated. However, the user distribution of N_CLU > 4 are increasing, and it means the bus service satisfy users' various demand.

The N_TNS shown in Figure 10 is similar to N_CLU, the tendency of N_TNS ≥ 5 are increasing. It means the users' behavior is getting more unstable. The user distribution of N_TNS = 0 noticeably increases with respect to N_CLU = 1, and it means an increase in the number of visitors or tourists.

If the managers only read one single period data without the information in Figures 8 through 10, they can only obtain the proportions of users for each variable. However, this study uses data in a continuous time period to represent the user behavior tendency. This information is expected to assist the managers to understand the user behavior and its tendency more precisely and evaluate the influence of the current bus policy. These results provide evidence that the change tendency of the proportion of users is not a simple straight line. Instead, the tendency is curve-shaped, and it could be different according to various values of each variable. When the researchers want to make the prediction or estimation of the long-term behavior pattern, these results would enhance the model with consideration of detail information of the known proportion changed. Because the tendency comes up from the rolling yearly data by month, the difference of various months or seasons could be eliminated. It could tell the more general tendency of the real situation.

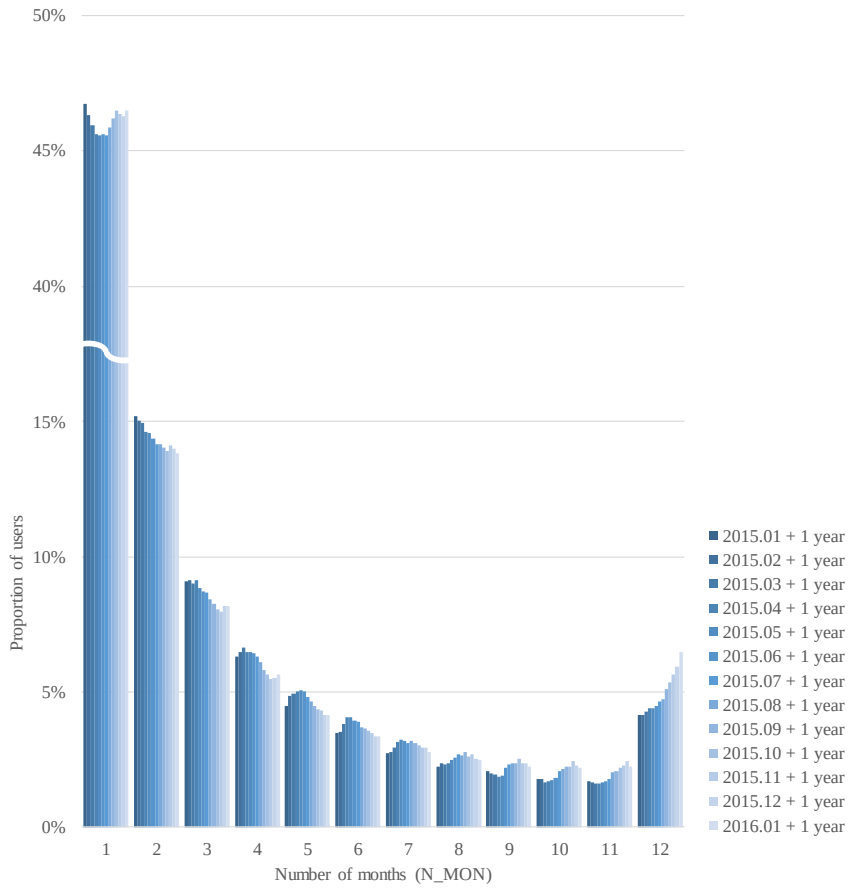


Figure 8. The tendency of user distribution changes in terms of N_MON

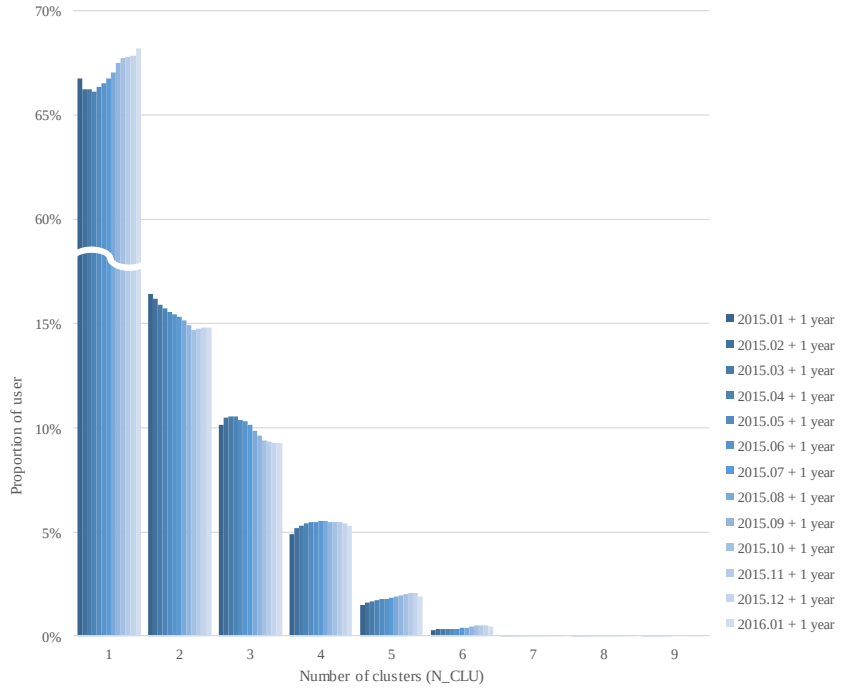


Figure 9. The tendency of user distribution changes in terms of N_CLU

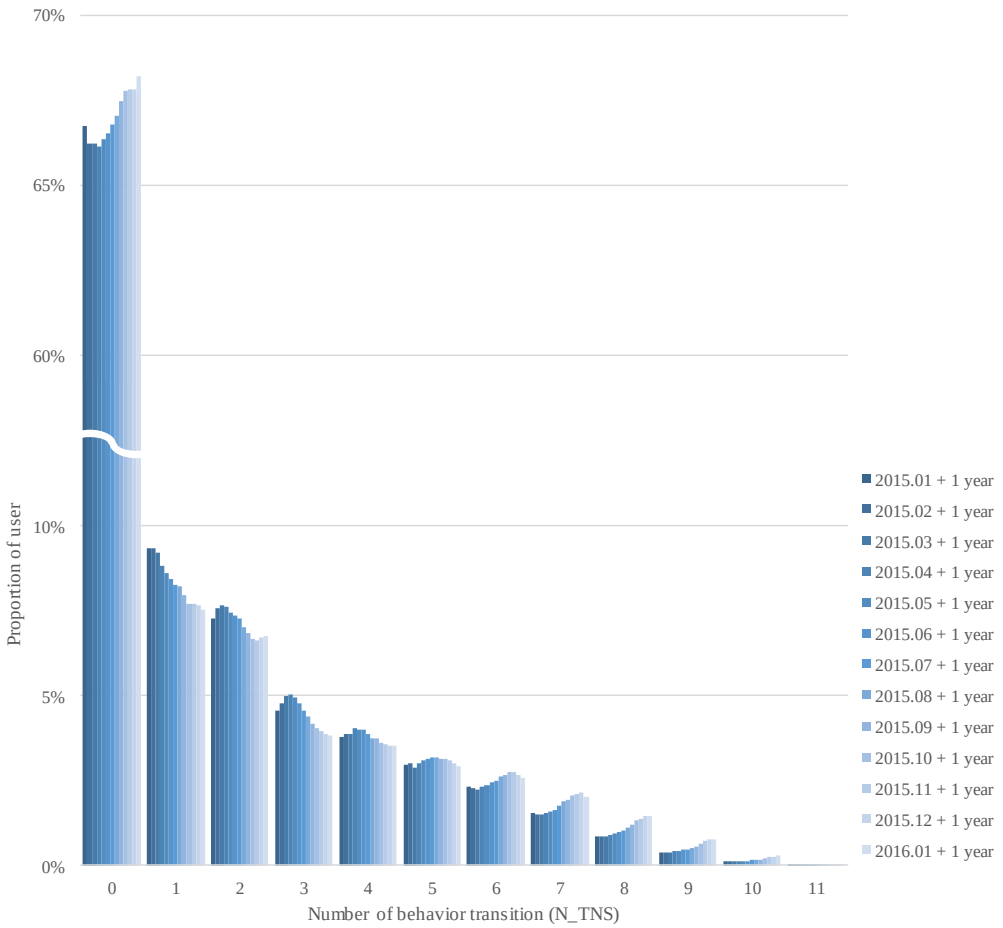


Figure 10. The tendency of user distribution changes in terms of N_TNS

4.3 User Behavior Regularity

In light of a relationship among the three regularity variables, Figures 11 through 13 show the relationship diagram of each two variables of them. Figure 11 shows the user distribution between N_MON and N_CLU. There are 13 diagrams, and the period of data starts from Jan 2015 ~ Dec 2015 to Jan 2016 ~ Dec 2016. The relationship between these two variables can assist the managers in evaluating diversity and periodicity. If a user's N_CLU is low, it means that a countermeasure may be more suitable to a user. If a user's N_MON is high, it means that the effect of a countermeasure may exist longer than lower N_MON.

From each diagram illustrated in Figure 11, the highest user distribution is at both N_MON and N_CLU equal to 1. Most users with this set of values are visitors, and the users rarely use the bus. Generally speaking, the number of this behavior has a high proportion of users. There are about 10% of users with N_MON = 2 ~ 4, but most of their N_CLU = 1 ~ 2. The users with N_MON = 5 ~ 11 are spread more, and most of their N_CLU = 2 ~ 3. The proportion of users with N_MON = 12 is higher than N_CLU = 5 ~ 11, and most of them have N_CLU = 3 ~ 4. It shows the average period of individual behavior cluster (the users with N_MON = 12) is about 3 to 4 months.

From the tendency of all diagrams in Figure 11, it shows that the user distribution at N_MON = 12 and N_CLU = 3 ~ 4 increased noticeably (1.4% → 2.0%). It means the proportion of the users, who use the bus every month, is increasing and most N_CLU = 3 ~ 4. In addition, the proportions of N_CLU = 2 ~ 3 is decreasing. Proportions of both N_MON

and N_CLU = 3 ~ 5 are increasing slightly.

Figure 12 shows the correlation between N_MON and N_TNS. Those two variables represent the stability of user behavior transition. When the N_TNS is getting higher, the user will not keep the same behavior pattern in a time period, and who might be a non-commuter. Also, the long-term pricing policy does not appreciate the users with high N_TNS. The pricing policy should consider the difference between each behavior pattern to appreciate them. Evident from Figure 12, the highest proportion of user is at N_MON = 1 and N_TNS = 0. Most users with N_MON = 12 are with N_TNS = 6 ~ 7, it means they may change the behavior pattern every two months on average. The users with N_TNS = 1 have a special tendency. When N_MON ≥ 4, the proportion of users of N_TNS = 1 is lower than N_TNS = 0 or 2. The N_TNS = 1 means the users will change their behavior once and will not change back or to others.

Figure 13 shows the correlation between N_CLU and N_TNS. These two variables represent the tendency of the behavior transition that could be diversity or stability of user behavior transition. In this case, the users have the intention to lower stability.

According to Figure 11, we can understand that the users in cell N_MON = 12 and N_CLU = 3 observably increase. In this cell, the proportion of charity card users is about 50%. This result could provide evidence of the effect of the policy which is “free charge for the elderly user in Keelung city”. The city government announced this welfare policy from 2007 and got lots of positive feedback from elderly citizens. This result provides a quantified index to evaluate the effect of the policy. According to Figure 13, the change tendency heads to higher N_CLS and N_TLS, these two values indicate the users tend to use the service more time and much diversity. With this result, the service could satisfy the users’ demand in various usage pattern. The tendency could represent a quantified value for evaluating the effect of the policies.

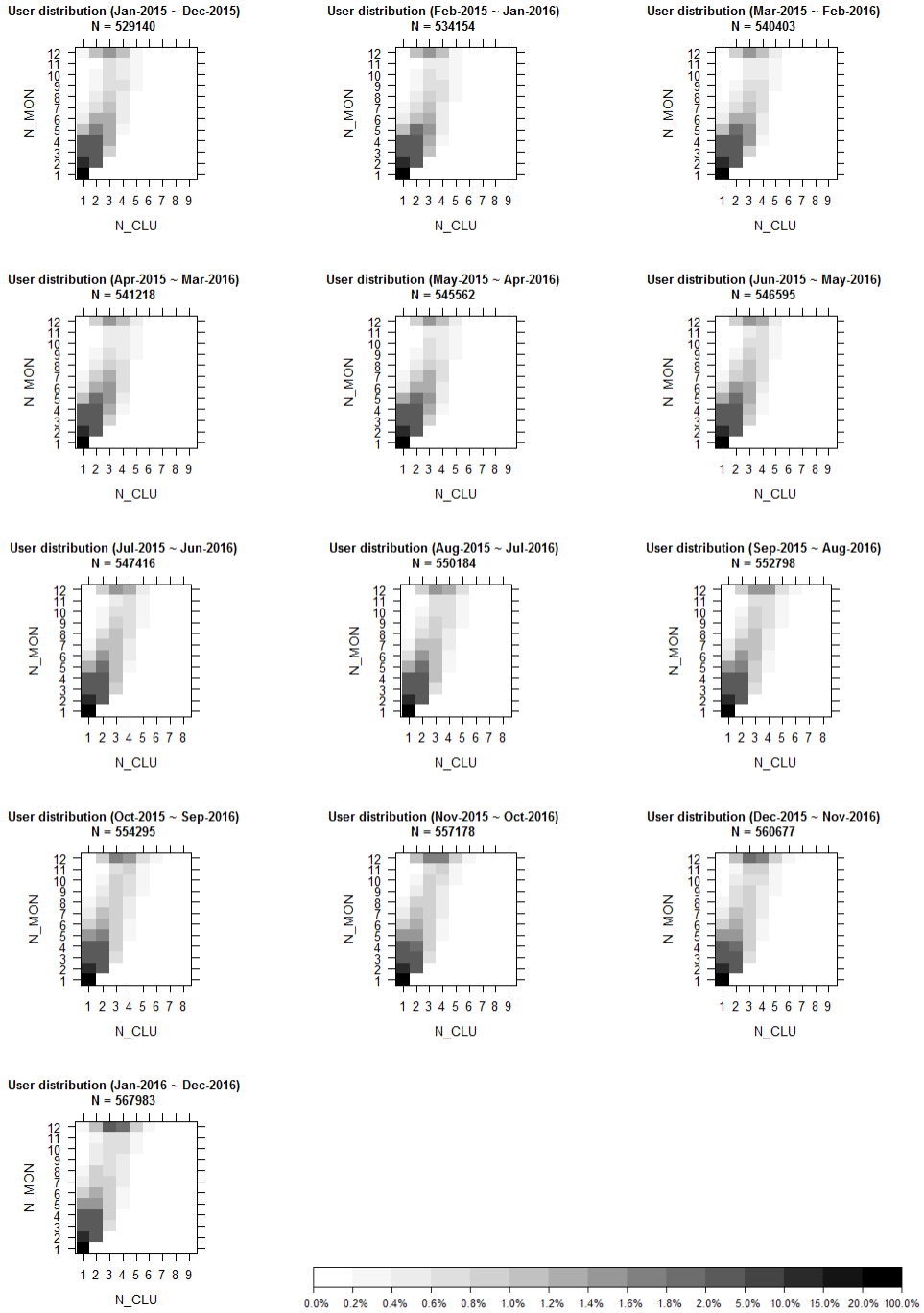


Figure 11. Correlation of user distribution among regularity variables, N_MON and N_CLU

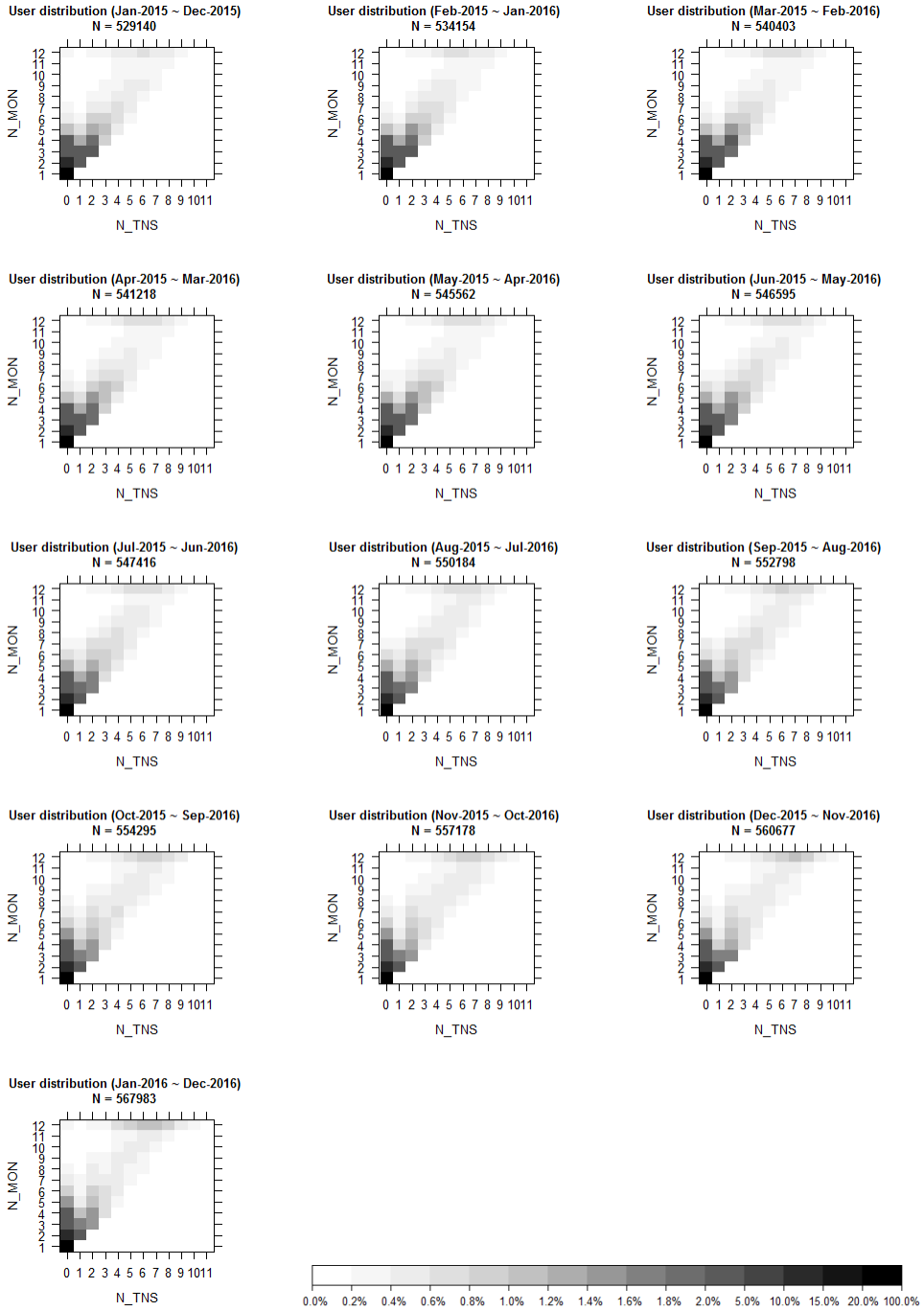


Figure 12. Correlation of user distribution among regularity variables, N_MON and N_TNS

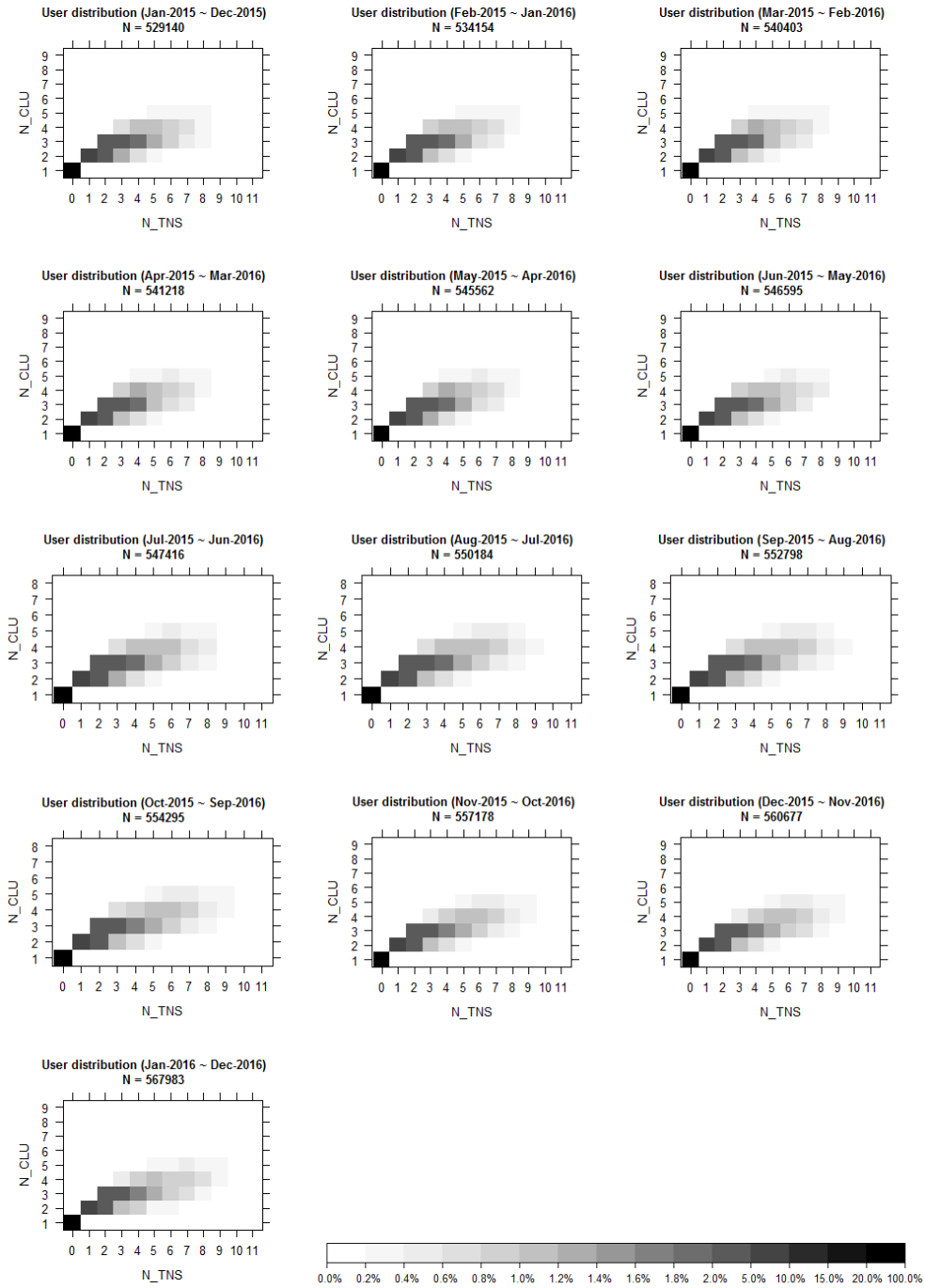


Figure 13. Correlation of user distribution among regularity variables, N_CLU and N_TNS

5. CONCLUSIONS

This study proposed three regularity variables to describe the variability of the user behavior regularity. Those variables can assist the analysts in understanding the periodicity, diversity, and stability of the user behavior. Because the bus service range is widely spread and the users are varied, and it will need a large-scale investigation to obtain the user behavior and their tendency. This study clusters user behavior via smart card data first and obtains the regularity by calculating the three regularity variables. After that, the behavior transition

tendency could be obtained from continuous and step-wise data. This methodology can find the latest tendency of behavior transition with a low budget and less time-consuming. The regularity result is important for bus service planning and evaluation.

This methodology also represents an approach to understand the tendency of behavior patterns by quantified values in detail. Most approaches for estimation and prediction of long-term behavior pattern need the understanding of them. The result of this research provides an efficient way to obtain more detail characteristics, and it also matches the real situation of the case study.

About evaluation, the managers could observe the tendency of behavior transition to evaluate the effect of the previous policy. About planning, the managers could make better pricing or discount policy via the periodicity and diversity of user behavior. The result also can locate users who may be influenced more precisely. This study uses the data in the whole city as the case study. It could analyze the specific bus route or area to find more detail of user behavior further. For each user, it could analyze the tendency of behavior transition of each user further. The empirical results are helpful to make a more precise decision.

Further effort is required to apply the advanced inferential statistics to extract more information from the smart card data, e.g., panel data analysis, structural equation model.

ACKNOWLEDGMENTS

The authors would like to express special thanks to the Bus Management Office, Keelung City, Taiwan, for providing smart card data of Keelung city bus.

REFERENCES

Agliati, A., Mantovani, F., Realdon, O., Confalonieri, L., & Vescovo, A. (2006). Multimodal temporal patterns for the analysis of user's involvement in affective interaction with virtual agents. From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience: Festschrift in Honor of Luigi Anolli., (c), 223–236.

Bagchi, M., & White, P. R. (2005). The potential of public transport smart card data. Transport Policy, 12(5), 464–474.

Briand, A. S., Côme, E., Trépanier, M., & Oukhellou, L. (2017). Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Transportation Research Part C: Emerging Technologies, 79, 274–289.

Chu, K. K. A. (2015). Two-year worth of smart card transaction data - Extracting longitudinal observations for the understanding of travel behaviour. Transportation Research Procedia, 11, 365–380.

El Mahrsi, M. K., Côme, E., Baro, J., & Oukhellou, L. (2014). Understanding Passenger Patterns in Public Transit Through Smart Card and Socioeconomic Data: A Case Study in Rennes, France. The 3rd International Workshop on Urban Computing (UrbComp 2014). New York.

Ferrer i Cancho, R., & Lusseau, D. (2006). Long-term correlations in the surface behavior of dolphins. Europhysics Letters, 74(6), 1095–1101.

Goulet-Langlois, G. (2015). Exploring Regularity and Structure in Travel Behavior Using Smart Card Data. Massachusetts Institute of Technology.

Guidotti, R., Gabrielli, L., Monreale, A., Pedreschi, D., & Giannotti, F. (2018).

Discovering temporal regularities in retail customers' shopping behavior. EPJ Data Science, 7(1), 6.
Holmqvist, K., Andrà, C., Lindström, P., Arzarello, F., Ferrara, F., Robutti, O., & Sabena, C. (2011). A method for quantifying focused versus overview behavior in AOI sequences. Behavior Research Methods, 43(4), 987–998.
Hung, P.-H., Doi, K., & Inoi, H. (2017). User Behavior Transition Mapping for Bus Transportation Planning based on Time Series Data Analysis of Travel E-ticket Information. Journal of the Eastern Asia Society for Transportation Studies, 12, 738–756.
Li, G., Li, S. E., Cheng, B., & Green, P. (2017). Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities. Transportation Research Part C: Emerging Technologies, 74, 113–125.
Li, Y. T., Iwamoto, T., Schmöcker, J. D., Nakamura, T., & Uno, N. (2018). Analyzing long-term travel behaviour: A comparison of smart card data and graphical usage patterns. Transportation Research Procedia, 32, 34–43.
Moody, J., Goulet-Langlois, G., Alexander, L., Campbell, J., Zhao, J. (2016). Measuring Explicit and Implicit Social Status Bias in Car vs. Bus Mode Choice, 95th Transportation Research Board Annual Meeting.
Moon, C. Y., Kim, J., & Hickman, M. (2018). A Machine Learning Approach for Detecting Long-Term Changes in the Weekly Trip Patterns of Public Transport Passengers. Conference on Advanced Systems in Public Transport and TransitData.
Morency, C., Trepanier, M., & Agard, B. (2006). Analysing the Variability of Transit Users Behaviour with Smart Card Data. 2006 IEEE Intelligent Transportation Systems Conference, 44–49.
Morency, C., Trépanier, M., & Agard, B. (2007). Measuring transit use variability with smart-card data. Transport Policy, 14(3), 193–203.
Okada, S., Sato, Y., Kamiya, Y., Yamada, K., & Nitta, K. (2012). Analysis of the correlation between the regularity of work behavior and stress indices based on longitudinal behavioral data. Proceedings of the 14th ACM International Conference on Multimodal Interaction - ICMI '12, 425.
Pas, E. I. (1988). Weekly travel-activity behavior. Transportation, 15(1–2), 89–109.
Sun, L., Axhausen, K. W., Lee, D.-H., & Huang, X. (2013). Understanding metropolitan patterns of daily encounters. Proceedings of the National Academy of Sciences, 110(34), 13774–13779.
Tarigan, A. K. M., Fujii, S., & Kitamura, R. (2012). Intrapersonal variability in leisure activity-travel patterns: the case of one-worker and two-worker households. Transportation Letters: The International Journal of Transportation Research, 4(1), 1–13.
Trépanier, M., & Morency, C. (2010). Assessing transit loyalty with smart card data. 12th World Conference for Transportation Research, 1–16.
Zhao, Z., Koutsopoulos, H. N., & Zhao, J. (2018). Detecting pattern changes in individual travel behavior: A Bayesian approach. Transportation Research Part B: Methodological, 112, 73–88.
Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., & Schmitt, G. (2016). Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PLoS ONE, 11(2), 1–17.
Zhong, C., Manley, E., Müller Arisona, S., Batty, M., & Schmitt, G. (2015). Measuring variability of mobility patterns from multiday smart-card data. Journal of Computational Science, 9, 125–130.