User Behavior Transition Mapping for Bus Transportation Planning based on Time Series Data Analysis of Travel E-ticket Information

Pai-Hsien HUNG [a], Kenji DOI [b], Hiroto INOI [c]

[a,b,c] *Graduate School of Engineering, Osaka University, Osaka, 565-0871, Japan*
[a] *E-mail: hung.pai.hsien@civil.eng.osaka-u.ac.jp*
[b] *E-mail: doi@civil.eng.osaka-u.ac.jp*
[c] *E-mail: inoi@civil.eng.osaka-u.ac.jp*

Abstract: In the past, bus service planning in metropolitan area was a crucial procedure of bus operators, private organizations, or local governments. One of the important topics of bus service improvement is of course how to understand the actual users' decisions, or strong points with which they can attract the users to their bus, in other words. However, there are only a small minority of users announce their thoughts directly. Therefore, in order to understand user's likes or dislikes, we use a time series data analysis technique to large e-ticket data. In order to obtain user's decisions derived from e-ticket system, we propose this new method to cluster the users via user behavior. User behavior transition map is to be drawn from continuous behavior transition data. A quantitative evaluation of user's decisions will be shown in the transition result, and therefore can target cluster by cluster to realize each need.

Keywords: E-Ticket Data, User Behavior Transition, Bus Service Evaluation

1. INTRODUCTION

Bus service is a highly flexible transportation system. It needs to react instantly to any types of slight modifications, e.g. building constructions, new vehicle type emergence, or the user's socioeconomic change. Once something is modified, the users may change their behavior. Therefore, managers are required to foresee what users will think about those modifications and their decision on the bus service itself, when they would like to apply some new planning upon their service. However, it is time and cost consuming to collect user's responses via traditional methods, such as interviews or questionnaire surveys. Managers usually obtain the information via sampling surveys with high cost, since only a few users give their feedback voluntarily. Moreover, such kind of methods are useless for bus service planning which requires quick reactions and has a large number of users.

Fortunately, as the application of e-ticket systems grows quickly, bus operators at present can get raw data of each transaction such as time, location and route when users board a bus. Large data sets not only present operating performance via ridership calculation but also derives users' behavior information via advanced statistical methods (Morency *et al,* 2006; Morency *et al,* 2007; Bagchi and White, 2005, Zhong *et al,* 2015). Accordingly, this study proposes a method to obtain users' behavior information quickly via e-ticket data analysis. Moreover, it goes without saying that users' behavior information is a very important foundation that helps managers to make better bus service planning decisions.

1.1 Background

Bus service is a widely spread public transportation system, and its construction cost is much lower than that of other transit systems such as subway or MRT (Mass Rapid Transit). It is also a highly flexible service, so it exists in almost all areas. Since bus service only needs vehicles and the bus stops, it only takes a short time (less than one month) to modify its route or stops. In addition, buses can drive not only on the bus exclusive lanes but also on general roads with other private vehicles.

Because of these characteristics, managers can easily change their bus plans at any time. However, due to that flexibility, it is difficult to find the optimal solution. Moreover, bus users' demands change day by day, so the bus service itself needs to be dynamically improved based on the current demand.

In practice, "Professional Judgement" is commonly used to make bus service planning decisions (USF CUTR, 2009; FHWA, 2007; Boyle, 2006). Except for new bus services that need whole area wide transportation planning, bus service planning can be made by using existing OD information. But since the analysis of area wide data consumes quite a lot of time and budget, OD information does not suit for planning within a small area or scale.

When conducting bus service planning, one of the important factors to consider is quick performance evaluation of the entire bus service. Because bus service contains multiple routes and various kinds of users, managers will spend lots of time trying to evaluate various service items. In order to evaluate services quickly, managers prefer using income, ridership, or feedbacks as evaluation indexes. But indexes like income or ridership are macroscopic, and so cannot derive individual user's satisfaction. Therefore, a quick procedure that can present individual user's satisfaction will help managers to evaluate service in a real sense.

1.2 Objectives

Last decade, general e-ticket systems provided user's detailed usage data quickly and completely, and therefore it helped to produce various data for planning (Bagchi and White, 2005). Most of the bus managers or operators calculated the ridership or number of users from e-ticket data to evaluate their performance. But it is impossible to obtain personal behavior transition from ridership without the usage judgment in a consecutive time interval. Behavior transition requires the judgement of users' behavior types in each time interval. Also, bus users only have two decisions, either to use or not to use that certain bus service, considering any kinds of variables. These two decisions will directly impact the ridership of bus service. In other words, the ridership is directly related to the users' decision.

Therefore, this study is to derive the reliable users' behavior transition according to the characteristics above and behavioral transition at consecutive time intervals. Basic behavior types will be considered first by analyzing e-ticket data in a single month. They will be the basis for transition judgement. After that, users' transition results could be calculated via behavior transition of the same users in consecutive time intervals, and drawn on a map that would allow for managers to improve the service easily. Behavior transition of overall systems then can provide reliable information of users' decisions.

This paper organizes as follows; Section 2 discusses previous studies about behavior clusters and bus service planning. Section 3 discusses method to cluster and compute behavior transition. A

case study in Tainan city is shown in Section 4. Section 5 discusses application of behavior transition map. Section 6 is the conclusion part of this study.

2. LITERATURE REVIEW

Application of e-ticket data has been discussed since 2005. E-ticket data is simple but it can be extended to various materials for transportation user behavior research, according to time, location, and service information. However, though usage could be obtained via e-ticket data calculation (e.g. frequency) it is hard to obtain what the purpose of each trip is. It requires more advanced methods to study in depth on information from much larger e-ticket data (Bagchi and White, 2005).

In research fields of e-ticket data, time interval is a very important variable. Electrical data could be recorded 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 describe the users' behaviors. 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 user's commuting type (Morency *et al,* 2006).

Continuous period analysis on smart card data is another method for operators to understand behavioral transition. In London, the data showed that the number of cards remaining active in the following period decreased in similar patterns at various routes (Ortega-Tong, 2013). That situation was also found in Tainan city. When users decided not to use the bus service, it also meant that they changed their behavior due to some reasons. But it is difficult to consider all reasons without conducting a costly survey. Therefore, a simple method to know who would have what behavior is valuable for bus service managers.

Although ridership is an easy way to evaluate performance of bus services, it may sometimes lose important information for managers to realize the real causes. By aggregating data, it is easy to conduct statistical processes or to know tendencies. However, people are not always following common rules to live; any kind of reasons might influence their decisions. (Flyvbjerg *et al,* 2005; Hägerstraan, 1970). Behavior clustering is a clear method to cluster users with similar behavior types to decrease noise or error in finding target users.

Previous studies in travel behavior research used cross-sectional data due to lack of longitudinal data. Some studies use 3 days data to over one month or longer. Recently, many of the studies use big data to consider mobility pattern (Chen *et al,* 2016). Big data of mobile phone records can obtain the human activity spaces and examine patterns of 12 consecutive months. Monthly variation in individual spatial behavior rose 17% by season (Järv *et al,* 2014).

More advanced behavior analysis on weekly data considers that estimating a mixture of unigrams models from trips captured through e-ticket data can retrieve weekly profiles depicting different public transportation demands. Clustering results provides better suggestions based on users' socioeconomic characteristics. Grouping with similar clusters also shows behavior information in different levels. Land use data will enhance the model to estimate the user's socioeconomic characteristics (El Mahrsi *et al,* 2014; Sun *et al,* 2013).

Another study identifies 11 clusters of users' travel patterns in London, and uses the clustering results to improve travel demand models. It shows the clustering results for two different months to understand the stability of clusters. Stability of cluster shifting can show whether the

users' behaviors are stable or not (Goulet-Langlois *et al,* 2016; Arentze *et al,* 2011; Kamruzzaman *et al,* 2011).

According to above literatures review, there are reasonable and diverse application for obtaining user's behavior from e-ticket data. Travel behavior from e-ticket data can also assist transportation planning. Ma et al. (2016) use DBSACN method to obtain trip chains from smart card data, and stated that the travel pattern and regularity levels are important information for researchers seeking to understand day-to-day urban travel behavior variability and facilitate active-based travel demand model development. Goulet-Langlois et al, 2016 inferes weekly travel behavior patterns from smart card data, and apply the result to derive mode choice of each pattern. Pelletier et al. (2011) reviews some researches about smart card data analysis and its application for transportation planning. Most research can analyze data based on user characterization and classification without personal information of users.

Due to limited resources and the high flexibility of bus service planning, it is important to consider what the service target is. User types are various and each one of them requires appropriate service. This study focuses not only on behavior clustering but also on behavior transition using consecutive time interval e-ticket data. Transition results of the overall system will show the tendency of users behaviors and how the behavior changes. Managers or operators will realize where the target is according to the behavior tendency, and what kind of target will bring improved performance.

3. CALCULATION METHODOLOGY OF BUS USER BEHAVIOR TRANSITION

This section discusses how to cluster bus user behavior from e-ticket data and to compute behavior transition. Weekly profiling is the key concept to conduct the clustering, and the average frequency is set to be one month.

3.1 Dataset

E-ticket data is used in this study, and each raw data contains the boarding and alighting information of a single trip by the same card. Data items in each row includes Bus ID, Route ID, Card No., boarding and alighting time, boarding and alighting location, and card type. We will use each user's data in one month to conduct the clustering calculation. One single month at least includes 4 weeks of weekdays and 4 weekends. By summing each hour of one month's usage, a weekly profile can be determined. In order to prevent the 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 random user cluster.

Frequency of most bus users' behavior are weekly, including weekday and weekend trip shown in weekly profile. Therefore, we consider there exists 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 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 records separately according to its boarding hour so that we can understand the specific frequency per hour. Peak hour characteristics as well as difference between weekday and weekend is now easy to define.

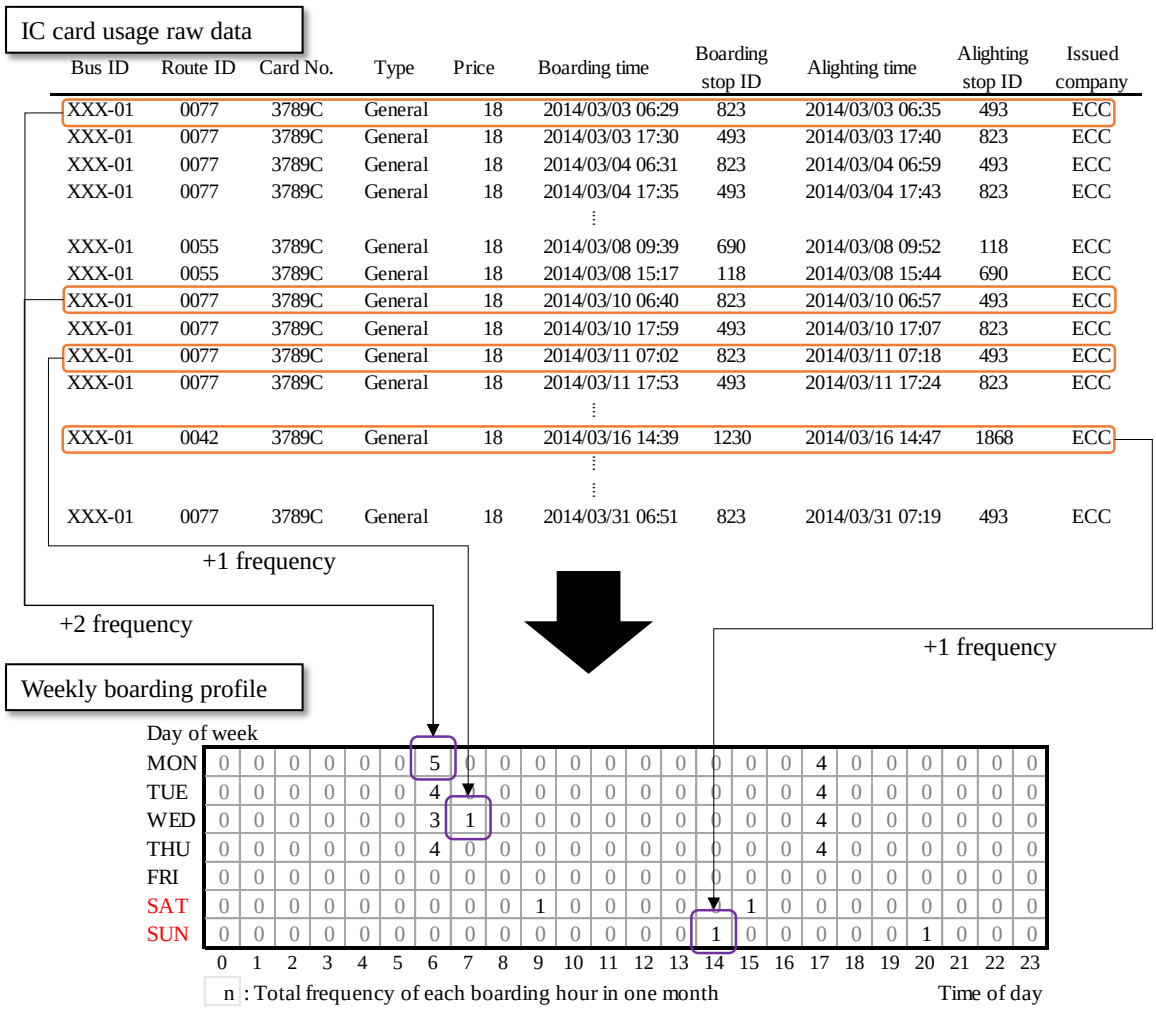


Figure 1. Example on how IC card usage data transferred into weekly boarding profile

3.2 Bus user behavior clustering

The main concept of clustering in this study is to group users with similar behaviors, and to define bus user behaviors only based on departure time and frequency. Other variables are not considered in order to simplify the clustering calculations. Although other variables, like weather, may influence users' decision, users do not change their behavior permanently because of accidental events. Therefore, we picked up a generic month without special events or long holidays to conduct clustering process. Variables of behavior are departure time and frequency. Therefore, we categorized users depart at similar time of the day or near frequency to be in the same cluster.

EM algorithm is used to find maximum likelihood parameters of a statistical model shown in formulation (1). Figure 2 is the diagram of EM. All samples that belong to each cluster is a probability value instead of a boolean value, and all probability of each sample belongs to all cluster are sum to 1. By iterative E-step and M-step in EM, we can get higher likelihood. But EM is a time-consuming algorithm, it needs to be more improved in order to apply in the real world.

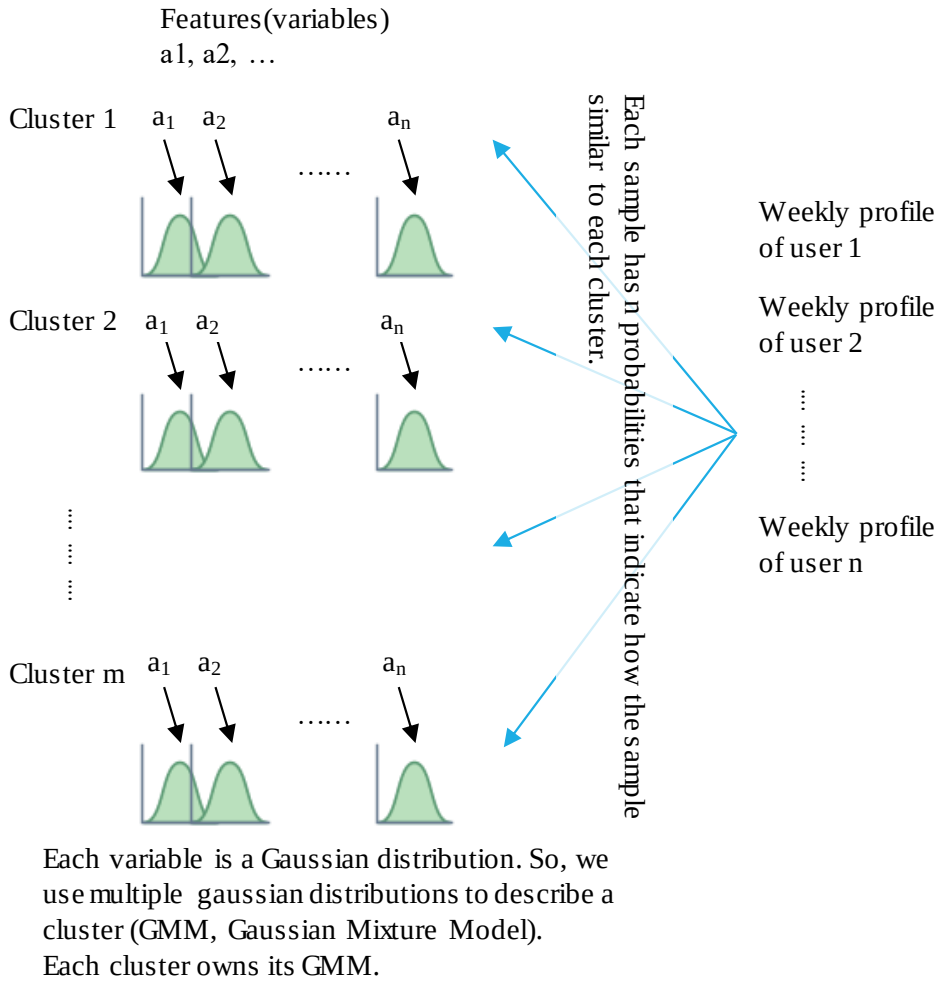


Figure 2. EM algorithm and clustering problem diagram

$$p(X \mid \lambda) = \prod p(x_i \mid \lambda) \tag{1}$$

where,

X : all weekly profile of all users

x_i : weekly profile of user i

λ : parameters.

The EM algorithm is used to cluster users of previous simple month data, and it is also appropriate for clustering. It is an unsupervised algorithm, like k-means. The advantages of EM are that it is simple, robust and easy to implement. It also has a better explanation for missing data (Dempster *et al,* 1977; El Mahrsi *et al,* 2014).

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

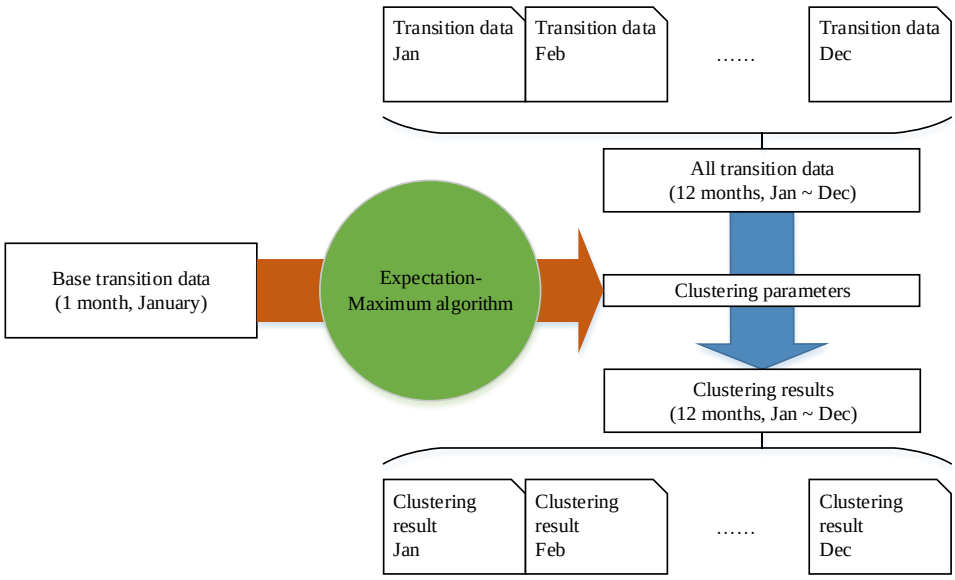


Figure 3. Clustering process for whole year data

3.3 Behavior transition calculation

In case of computing behavior transition, clusters should be ordered for better understanding of the cluster's key characteristics. In this study, regularity is used to evaluate users' tendency to use bus service. As a regular user, more regularity means more frequency, like students or commuters. Departure time is a significant index that indicates the regularity. Besides, commuters usually take bus at peak hours, e.g. 7:00-9:00 and 17:00-19:00. So, the regularity value is split into morning peak and afternoon peak, having 12:00 in the middle. Cluster regularity is considered as formulation (2), regularity value is the sum of standard deviation of morning and afternoon peak for all trips in the same cluster. The random cluster whose users use less than 4 times per month is forced to be the most random cluster.

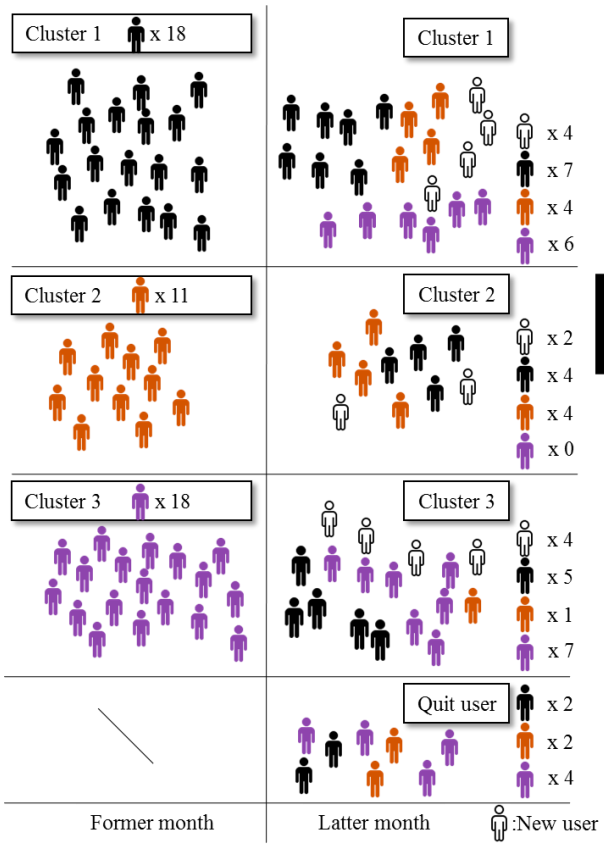
$$\text{Cluster regularity} = S_{MP} + S_{AP} \tag{2}$$

where,

S_{MP} : Standard deviation of boarding hour in morning peak (0-11) of all users' boarding in each cluster.

S_{AP} : Standard deviation of boarding hour in afternoon peak (12-23) of all users' boarding in each cluster.

After regularity sorting, behavior transition in adjacent months is conducted by computing the shifting rate from the former cluster to the latter cluster. The shifting rate is the ratio indicating the change in e-tickets usage from former clusters to latter clusters. There are two special clusters in each of the former and latter clusters. One is the "New" cluster, which means that the e- tickets are only used in the latter month; the other is the "Quit" cluster, which means that the cards are only used in the former month. Figure 4 is a behavior transition result example, showing how to calculate the transition ratio from a user clustering result.



Behavior transition ratio table

Former to latter		Latter clusters (number of user)				
		Cluster 1	Cluster 2	Cluster 3	Quit	Total
Former clusters	Cluster 1	7	4	5	2	18
	Cluster 2	4	4	1	2	11
	Cluster 3	6	0	7	4	17
	New	4	2	4	-	10

Former to latter		Latter clusters (Shifting rate(%))				
		Cluster 1	Cluster 2	Cluster 3	Quit	Total
Former clusters	Cluster 1	38.9%	22.2%	27.8%	11.1%	100.0%
	Cluster 2	36.4%	36.4%	9.1%	18.2%	100.0%
	Cluster 3	35.3%	0.0%	41.2%	23.5%	100.0%
	New	40.0%	20.0%	40.0%	-	100.0%

Figure 4. Procedure of behavior transition ratio computation in adjacent months

3.4 Behavior transition map drawing

Now just ignore the ratio of those who remain in the same cluster both in former and latter months, since this data simply shows no shift at all, and then concentrate on the data of each of the remaining latter clusters. Among the remaining data, the one with the maximum shifting rate is the cluster whose users tend to shift. This figure also expresses the desired shift level of users. Connecting all former clusters to the maximum shifting rate of the remaining latter clusters by using a directional link, we will obtain a transition map. Figure 5 is an example which shows this connection on the map. Not only the latter cluster with maximum shifting rate, but also the latter clusters with second or less top shifting rate could be considered in transition map. With this map, managers can easily understand the tendencies of all users in the service, including New and Quit users. Moreover, since each cluster indicates a specific behavior, managers can easily know who the target users are.

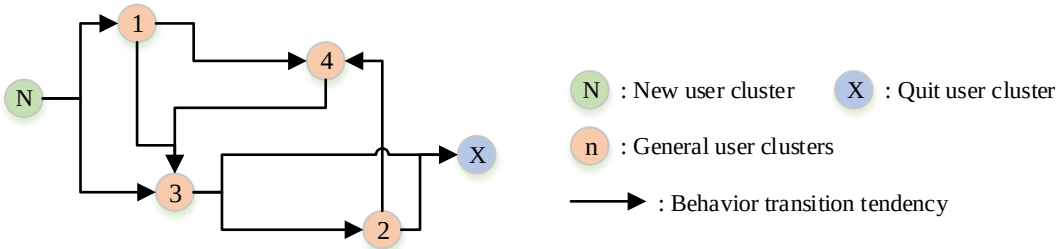


Figure 5. Example of behavior transition map at consecutive months

The transition map consists of behavior clusters and their own transition tendency. In this map, each cluster can be a target to observe in the evaluation procedure. In order to find the significant problem in several possible tendency paths, a Primary Tendency Path (PTP) between two clusters is required. As stated above, all clusters are cluster results from user behavior (weekly profile) and be sorted by regularity of departure time. The ordered clusters mean that we use departure time regularity as an index to organize clusters. New and Quit users are two critical clusters draw managers' attention. Once a manger chooses a target cluster to observe, the paths will provide useful information to identify problems with meaningful tendency, and therefore the PTP could be the most critical issue.

4. CASE STUDY OF BUS USER BEHAVIOR TRANSITION

This section uses data of Tainan city as an example to examine bus user behavior transition via the process in Section 3. The process transforms e-ticket data into behavior transition tendencies in order to understand bus service performance and target of improvement.

4.1 Socioeconomic characteristics and bus network of Tainan

Tainan locates at southern part of Taiwan, shown in Figure 6. From December 2010, Tainan County and Tainan City have been merged into Tainan special municipality. Its population amounts to 1,885,199 (2015), its area is 2191.7 km^2, and its bus service contains 105 routes and 3 companies. The bus network is also shown in Figure 7. The downtown area is located in the south west. From 2012 December, an e-ticket system has been applied to all bus services. In June of 2013, the DOT of Tainan started to operate in six main lines after re-planning the original bus services. Although the total ridership is continuously growing, the city bus mode share is still halted at 1%.

The downtown area locates at the southeast corner, and northern part is minor downtown. The selected bus operator provides service at northern rural and downtown area. The E-ticket data of the selected bus operator in 2014 is used in this study, and the amount of the operator's data is 867,328 rows of data.

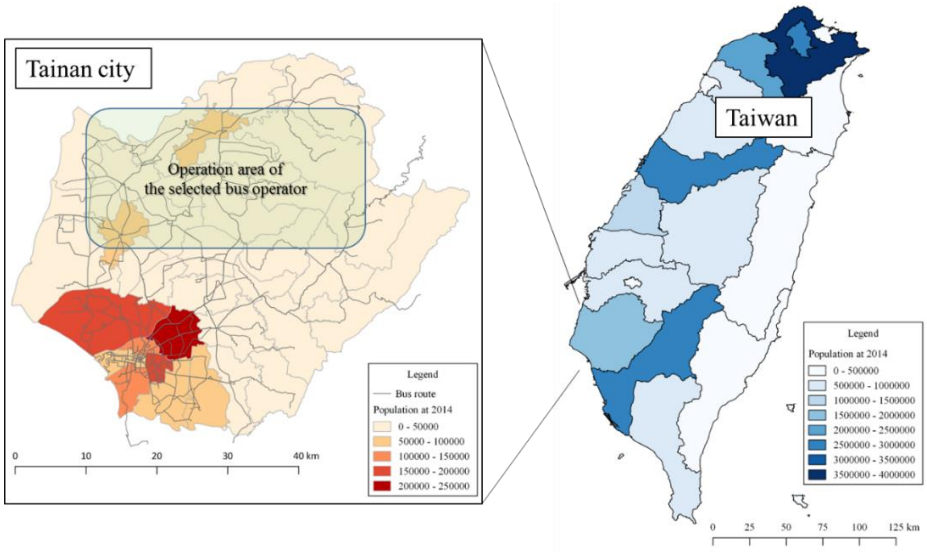


Figure 6. Population and bus network of Tainan city

4.2 A case study of bus user behavior clustering

For the selected bus operator, there were 12,002 e-tickets used in 2014 January, and 8,691 (72.4%) of them are random users who use less than 4 times in that month. The other 3,311 e-tickets are users who use more than 3 times and therefore are being clustered via the EM algorithm. We used the 'mclust' module in R software to solve the clustering problem. Figure 8 shows the clustering result. The BIC (Bayesian information criterion) value is a criterion for model selection among the finite set of models. For the EM algorithm, larger BIC values provide strong evidence for the model and the associated number of clusters. In Figure 7, the best set of model and cluster number is cluster 7 on model VII (BIC: -5115.31). VII model means that the statistical model is spherical in shape distribution and variable in volume in each cluster.

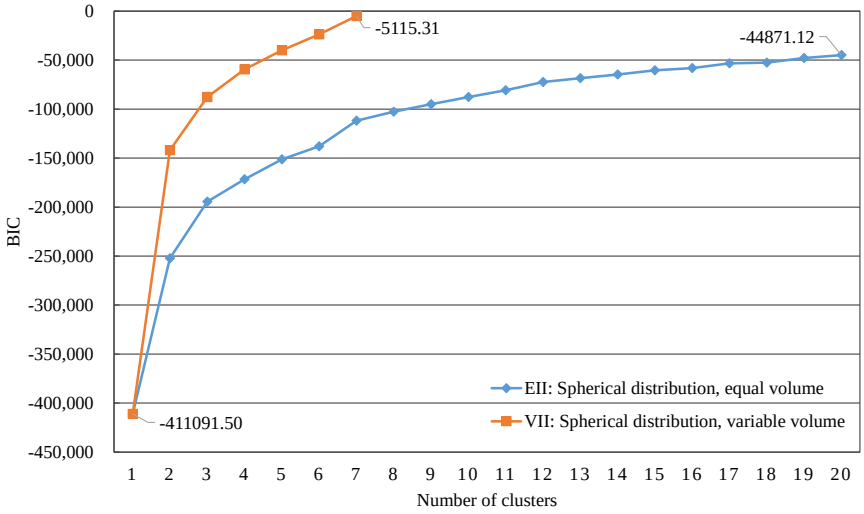


Figure 7. Result of EM method from data of the selected bus operator in January 2014

After clustering and regularity sorting, there comes 7 usage patterns and 1 rare cluster listed in Figure 8. Cluster 1 is the most regular users', who use the bus around 6:00 and 17:00 and should be standard commuters; the average frequency shows that they use buses almost every weekday. Cluster 2 is the regular users who use the bus around 7:00, and there exists only half frequency at the afternoon peak. Cluster 3 is the users who use the bus at the morning peak only. Cluster 4 is the opposite; they use the bus only at the afternoon peak. Cluster 5 is the users who use the bus not only at the morning and afternoon peaks but also at some adjacent several hours of both peaks, like college students or employees. Cluster 6, 7 are random users who use the bus at random departure times. The difference between these two is that cluster 7 has a higher frequency at weekends.

The results show regularity sorting clearly and closer to the real world. The cluster regularity of clusters 1 ~ 4 are less than 3 (hours), so they are considered as a regular group. The cluster regularity of clusters 5 ~ 8 are over 3, so they are considered as a random group.

4.3 Bus user behavior transition

Data in 12 months could be clustered via the clustering parameters in section 4.2, and get 11 behavior transition tables at adjacent months such as Table 1 and Table 2. They are examples of

bus user transition from January to February and from June to July. In general, most clusters have a maximum ratio as a preservation rate, which amounts to 70.2% in the regular group, but for clusters 7 and 8 this is not the case. Cluster 7 has only a 21.6% shifting rate to itself, which means users in this cluster change quite often. In the case of the bus service, this result is a warning for managers telling that these users may easily stop using their buses.

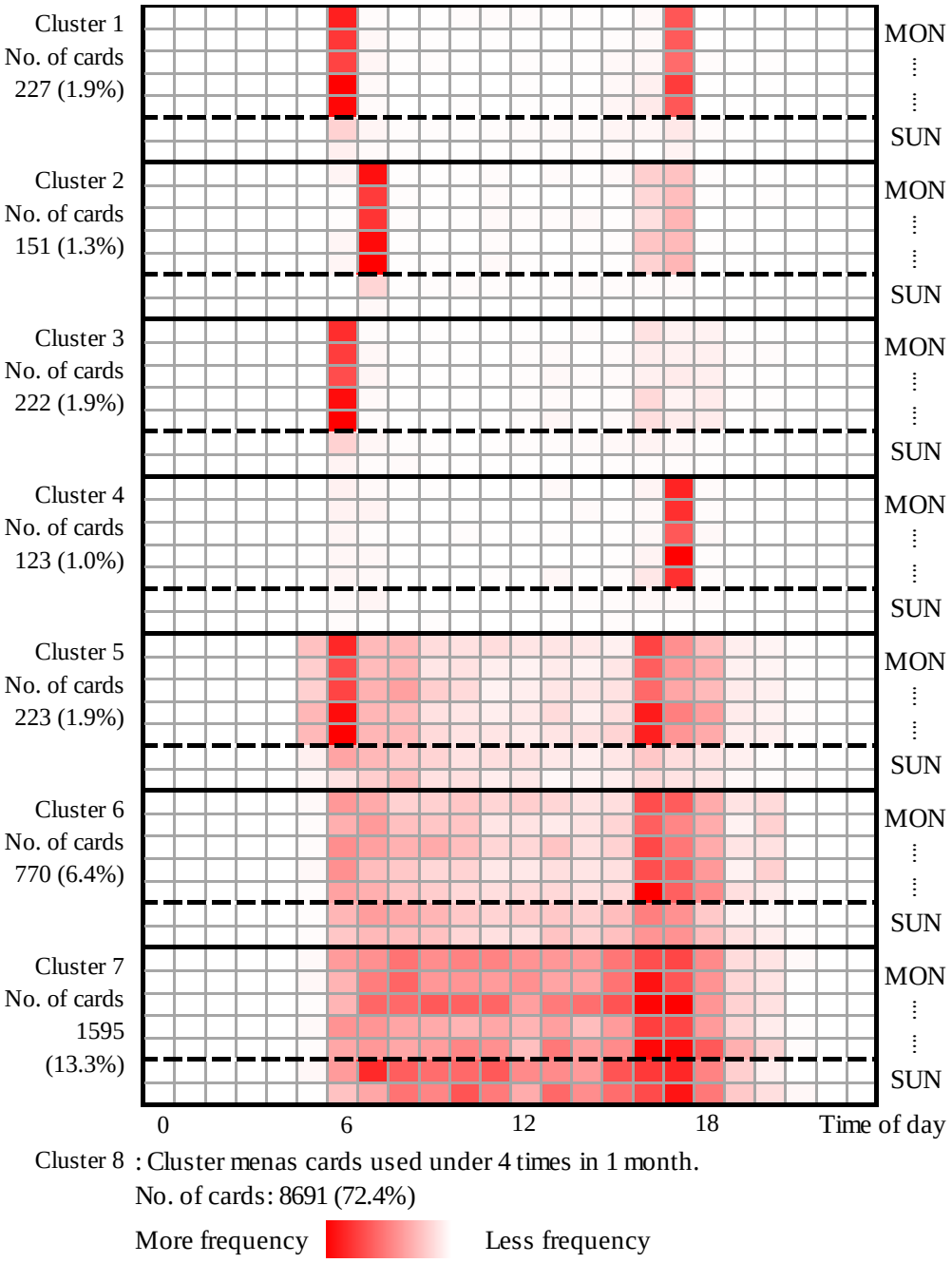


Figure 8: Behavior patterns of all clusters in the selected bus operator

Table 1. Behavior transition result at 2014 January to February

2014.01 to 2014.02		Latter clusters (after transition)									
		1	2	3	4	5	6	7	8	Quit	Total
Former clusters	1: Regularly, both MP and AP	68.7%	0.4%	3.5%	4.0%	1.8%	4.0%	3.5%	4.8%	9.3%	100.0%
	2: Regularly, MP and half f. in AP	0.7%	70.2%	0.7%	2.6%	6.0%	7.9%	6.0%	2.6%	3.3%	100.0%
	3: Regularly, MP only.	8.6%	0.9%	59.9%	0.9%	5.4%	9.5%	2.7%	3.2%	9.0%	100.0%
	4: Regularly, AP only	5.7%	1.6%	0.0%	56.1%	0.0%	8.9%	7.3%	10.6%	9.8%	100.0%
	5: Randomly in peak hour.	6.3%	3.1%	5.4%	2.7%	44.8%	22.0%	4.0%	4.0%	7.6%	100.0%
	6: Randomly	0.8%	1.2%	1.2%	3.8%	2.5%	28.1%	23.0%	24.2%	15.5%	100.0%
	7: Randomly with higher f..	0.3%	0.5%	0.5%	0.8%	0.4%	8.6%	21.6%	33.2%	34.2%	100.0%
	8: Randomly and less than 4 times	0.1%	0.1%	0.1%	0.2%	0.1%	1.1%	5.7%	30.1%	62.6%	100.0%
	New	0.2%	0.1%	0.1%	0.2%	0.1%	1.3%	7.3%	90.7%	0.0%	100.0%
	Total	1.1%	0.7%	0.9%	0.8%	0.8%	3.1%	8.1%	56.5%	28.2%	100.0%

MP: Morning peak; AP: Afternoon peak; f.: Frequency

Table 2 shows transition from June to July. The difference is that the cluster preservation rates of the regular group decreases from 56.1%-70.2% to 7.3%-33.3%. This is during the summer vacation period, which shows that many users in the regular group are students, and they stopped or reduced the frequency of using the bus services during their summer vacation.

Table 2. Behavior transition result at 2014 June to July

2014.06 to 2014.07		Latter clusters (after transition)									
		1	2	3	4	5	6	7	8	Quit	Total
Former clusters	1: Regularly, both MP and AP	14.0%	2.5%	18.6%	1.3%	12.3%	13.6%	11.4%	10.2%	16.1%	100.0%
	2: Regularly, MP and half f. in AP	0.0%	33.3%	0.8%	0.0%	4.8%	16.7%	9.5%	8.7%	26.2%	100.0%
	3: Regularly, MP only.	2.8%	1.2%	25.8%	0.8%	8.1%	15.3%	12.1%	11.3%	22.6%	100.0%
	4: Regularly, AP only	1.7%	1.7%	0.6%	7.3%	7.9%	13.5%	17.4%	20.2%	29.8%	100.0%
	5: Randomly in peak hour.	1.3%	2.7%	6.1%	0.5%	32.7%	18.4%	14.1%	9.0%	15.2%	100.0%
	6: Randomly	0.5%	1.3%	1.3%	1.1%	6.8%	23.9%	16.7%	20.9%	27.5%	100.0%
	7: Randomly with higher f..	0.1%	0.5%	0.4%	0.4%	2.2%	11.0%	15.7%	26.2%	43.4%	100.0%
	8: Randomly and less than 4 times	0.1%	0.2%	0.2%	0.2%	0.3%	2.7%	6.5%	24.7%	65.2%	100.0%
	New	0.2%	0.5%	0.6%	0.3%	0.7%	3.6%	10.9%	83.2%	0.0%	100.0%
	Total	0.3%	0.7%	1.0%	0.4%	1.7%	5.4%	9.7%	45.7%	35.1%	100.0%

MP: Morning peak; AP: Afternoon peak; r.: Ridership

4.4 Bus user behavior transition map

For the data of the selected bus operator in 2014, there should be 11 transition results like in Table 1. Table 3 comes from the maximum shifting rate of the latter clusters via all transition results of all neighboring months. Clusters of the random group in Table 3 all tend to shift into more random clusters or the “Quit” cluster. Clusters in the regular group (Cluster 1 ~ 4) have 3 or more different latter clusters in Table 3. It means that the latter clusters vary from month to month. Managers can select the suited Connected Level (CL: number of latter clusters chosen to show) to connect all

clusters. In this study, as in Figure 9, we show the latter clusters with CL = 2 to draw the transition map.

Clusters 3 and 6 in table 3 are the only two clusters which tended to be more regular, and others tended to be more random or eventually become Quit clusters. This tendency shows that users prefer shifting to more random clusters. However, managers cannot consider that from data based only on ridership, since the total ridership may slowly grow. Users quitting the bus service could be the cause preventing the city bus mode share from growing.

Table 3. Behavior transition result with maximum shifting rate of whole year 2014

	Clusters	Transition with maximum shifting rate expect itself										
		Jan to Feb	Feb to Mar	Mar to Apr	Apr to May	May to Jun	Jun to Jul	Jul to Aug	Aug to Sep	Sep to Oct	Oct to Nov	Nov to Dec
Former clusters	1: Both MP and AP	Quit	5	5	5	6	3	6	3	5	5	5
	2: MP and half f. in AP	6	5	5	5	6	Quit	6	8	5	5	5
	3: MP only.	6	5	5	5	Quit	Quit	6	1	1	6	5
	4: AP only	8	5	5	8	7	Quit	6	Quit	5	5	5
	5: Random in peak hour	6	6	6	6	6	6	6	6	6	6	6
	6: Randomly	8	5	8	8	7	Quit	7	8	7	8	8
	7: Higher f. in weekend	Quit	Quit	Quit	8	8	Quit	Quit	Quit	Quit	Quit	Quit
	8: Less than 4 times	Quit	Quit	Quit	Quit	Quit	Quit	Quit	Quit	Quit	Quit	Quit
	New	8	8	8	8	8	8	8	8	8	8	8

The latter cluster shows most often in all year of each former cluster.
The latter cluster show second often in all year of each former cluster.
n The latter cluster number with first highest probability evaluate from former cluster except former cluster itself.
MP: Morning peak; AP: Afternoon peak; f.: Frequency

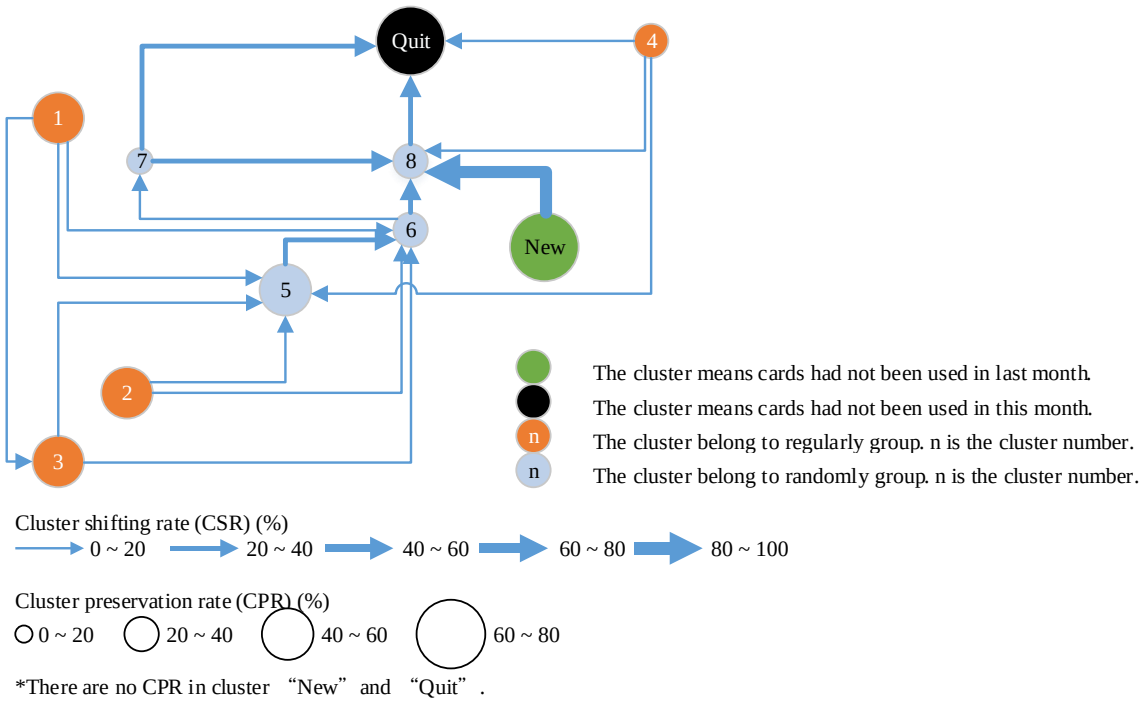


Figure 9. Behavior transition map of the selected bus operator in 2014

5. DECISION SUPPORT BASED ON BEHAVIOR TRANSITION MAP

Based on the clustering results in Section 4, e-ticket data of the whole year could obtain transition result of bus user behaviors via user behavior clustering. We can draw the transition map via connection of latter clusters with higher shifting rates. CL is the number of latter clusters chosen to show in the map. Since CL is the number of latter clusters that we choose, the higher the CL is, the more precisely the tendency can be shown. In other words, if we would like to consider precise tendency in addition to the highest one, we will need to use a higher CL. Therefore, CL can be set differently according to the managers' planning objectives.

5.1 Transition map for decision support

Figure 10 is a transition map with CL = 2. It shows the transition tendency of all clusters and difference between various shifting rates via link width. In addition, we draw PTP (Primary tendency path) which shows the path with highest tendency from one cluster to another, and clusters in the path are behaviors they tend to shift. The Critical Path method is considered to determine the PTP.

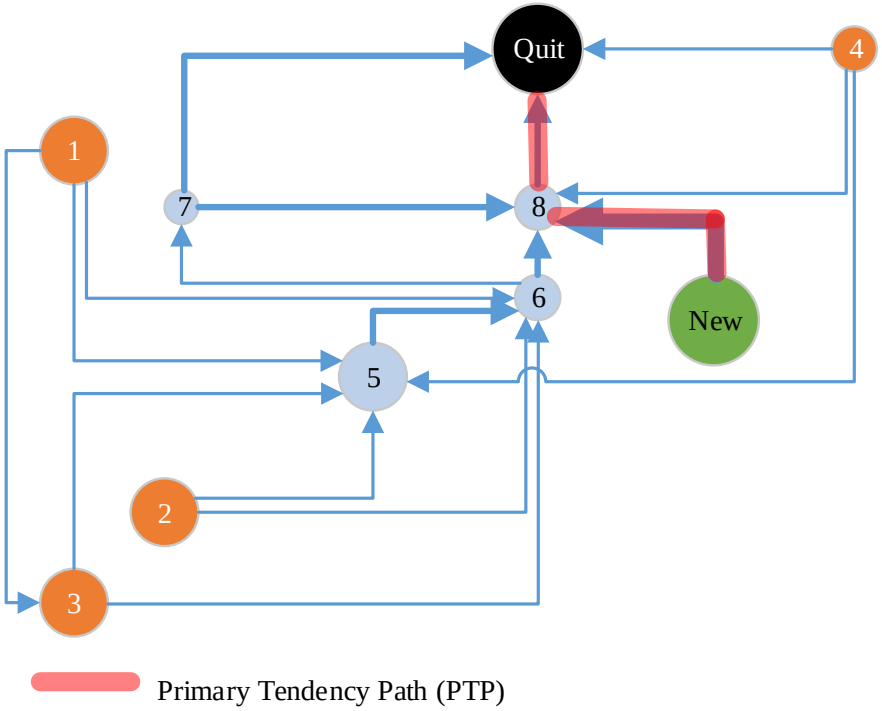


Figure 10. Primary Tendency Path (PTP) on behavior transition map

Once the transition map is drawn, managers can obtain information below from the map in support of their bus service planning.

1) Finding out the target user cluster

When managers plan bus services, the most important first step is to decide the types of user behavior they will serve. Then they can conduct the appropriate planning for specific behaviors.

However, users will change their behavior according to service content or other conditions. Understanding users' behavior transition in advance will allow for countermeasures to be conducted earlier and decrease the number of users that stop using the bus service. In figure 10, clusters 4, 7, and 8 have tendency for the "Quit" cluster directly, and their frequencies of morning peak hours are relatively small. On the other hand, the frequencies of the morning peak hours of clusters 1, 2, 3, and 5 are large, and they do not shift to the "Quit" cluster directly. This means that most users taking buses during morning peak hours will shift to other clusters before shifting to "Quit". Managers can improve service levels for the morning peak or understand the exact reasons why users stop using buses to keep users in their service territory.

According to Figure 10, if managers want to find the PTP from the new user to the regular user group, it cannot be found since the CL is too small and cannot show the desired tendency. This means that the probability of the new users transit into the regular user group is quite low; we must increase CL to present more precise tendency in order to show the path of the new users transit into the regular group. When the CL is higher, the tendency is more significant.

2) Understanding how the user will change their behavior

First the "Quit" cluster will be set as an end, the "New" cluster as a start, and average shifting rate as a link cost. The PTP of the "Quit" cluster is shown in Figure 10. In the PTP, cluster 8 is the only cluster between "New" and "Quit". Most new users tend to use the bus service as cluster 8 and stop using it in the near future. Since most users in cluster 8 are those who use the bus rarely, about 60% of them will stop using the bus service within the next month. It also means that about 60% of the users in cluster 8 will be going to move onto different clusters in upcoming months. This tendency shows that occasional users tend to stop using buses within a month. According to this result, managers need to improve bus service for cluster 8's demand in order to decrease the tendency to "Quit".

3) Evaluating the level of service of a bus route

Since ridership is a significant index for performance evaluation, the level of service of a bus route is often evaluated via ridership of respective routes or the whole service. However, the ridership will be influenced by frequency of each user, and therefore, it is important to know how users will change their behavior according to their disposition toward the bus service. Once the tendency is obtained, managers can evaluate the bus service via users' decisions, without the need for a costly preference survey. Instead of conventional questionnaire survey, this research performs a two-stage procedure to assist bus service planning. First is to grasp the overall tendency of transition; second is in-depth analysis on users' behavior change.

5.2 Applying transition results on transportation planning

This section presents an example of shaping an in-depth analysis according to user behavior transition result. Table 4 is the top 10 ridership and their bus stops of regular group's each transition result, during the period from 2014.01 to 2014.02. There are five schools with in this table as follows; Lioujia junior high school (LJ JHS), Pei-Men A.I. School (PM AIS), Houbi senior high school (HB SHS), Liouying junior high school (LY JHS), and Baihe junior high school (BH JHS). All of them have rather higher ridership in one month, but are derived from different transition results. Managers can not only find out which one of the bus stops adjacent to schools in whole

Tainan needs to be mentioned efficiently, but also compare them with other information to assist bus service planning.

Table 4. Top 10 ridership and their bus stop of each transition results in regular group

Time	Transition from 2014.01 to 2014.02											
Former	Regular group (RG)											
Latter	Regular group (RG)				Random group (RN)				Quit (QG)			
Top	Bus Stop Name	Type	Rds.	%	Bus Stop Name	Type	Rds.	%	Bus Stop Name	Type	Rds.	%
1	Singying bus station	Terminal	661	14.3	Singying bus station	Terminal	329	24.4	Singying bus station	Terminal	114	23.3
2	Baihe bus station	Terminal	566	12.3	Houbi senior high school	School	246	18.2	Singying cultrural center	Scenic Area	57	11.6
3	Tainan city farmers' assn.	Residence	545	11.8	Baihe bus station	Terminal	154	11.4	Baihe police station	Governmen t	56	11.4
4	Syuejia	Terminal	515	11.2	Tainan city farmers' assn.	Residence	150	11.1	Baihe junior high school	School	47	9.6
5	Jiali park	Scenic Area	445	9.7	Sanmin Rd. intersection	Residence	101	7.5	Sanmin Rd. intersection	Residence	45	9.2
6	Lioujia junior high school	School	442	9.6	Lioujia	Terminal	87	6.4	Baihe bus station	Terminal	43	8.8
7	Sanmin Rd. intersection	Residence	441	9.6	Liouying junior high school	School	76	5.6	Dingjhang-duanshu	Residence	38	7.8
8	Pei-Men A.I. School	School	353	7.7	Yanshuei	Terminal	72	5.3	Tainan city farmers' assn.	Residence	32	6.5
9	Singying cultrural center	Scenic Area	351	7.6	Syuejia district office	Governmen t	69	5.1	Syuejia	Terminal	29	5.9
10	Gasoline station	Residence	288	6.3	Syuejia	Terminal	66	4.9	Jhiazihgang	Residence	29	5.9
	Total		4607	100	Total		1350	100	Total		490	100
	Ridership of all bus stops		11987	-	Ridership of all bus stops		2875	-	Ridership of all bus stops		1047	-

*Rds.: Ridership; RG: Regular group (Cluster 1 ~ 4); RN: Random group: (Cluster 5 ~ 8); QG: Quit

Table 5 is the comparison of those five schools in table 4. BH JHS is the critical bus stop since users who alight here have higher tendency of quitting bus service. According to the comparison between demand and frequency per day, it is clear that BH JHS has higher demand and lower frequency. Usually students and faculties must go to school at the morning peak, therefore having lower frequency in the morning peak will induce users to be less satisfied, then leads users up to quit using bus service. If operators enhance the frequency of BH JHS bus stop usage, they will then have opportunity to retain more users within their service. Moreover, by conducting more detailed analysis, e.g. departure time or location analysis, operators can obtain much better frequency for BH JHS.

Table 5. Comparison of five schools in top 10 ridership of each group transition

Bus stop	Lioujia junior high school (LJ JHS)	Pei-Men A.I. School (PM AIS)	Houbi senior high school (HB SHS)	Liouying junior high school (LY JHS)	Baihe junior high school (BH JHS)
Group transition	RG → RG	RG → RG	RG → RN	RG → RN	RG → Q
Bus routes	Y1, Y2, Y3	BL2, BL3, BR	Y9, Y10	Y3	Y10, Y13, Y14
Total frequency / day	20	49	26	5	7
Total frequency (morning peak) / day	5	8	4	1	2
No. of students	491	1844	221	78	556
No. of faculties	67	200	48	20	70

6. CONCLUSIONS

This study proposed a simple and efficient procedure to obtain user behavior transition from e-ticket data. By applying the EM method, bus users were clustered into several groups, and their behavior transition between each adjacent month was estimated. A transition map was drawn by connecting pairs of clusters with higher shifting rates. The map showed transition tendency and a quantitative index of bus users in each cluster. Generally speaking, if tendency is getting more regular, it means the bus service is getting more suitable to users. If tendency is getting more random, there should be an improving plan to change the tendency according to service or cluster of users. The quantitative transition rates also help managers to evaluate and plan the bus service more efficiently.

Using PTP (Primary Tendency Path) in a transition map is important for managers to determine the transition tendency of some clusters. Section 5 showed an example of PTP from the Quit cluster. PTP helps managers to know which cluster will become a Quit cluster. Moreover, managers can make improvements and allocate resources to prevent users from becoming part of the Quit cluster.

The method developed in this study is different from previous ones that assumed all ridership to be the same. Actually, ridership is a composition of diverse user types which changes with time. Resource allocation could be more appropriate with the understanding of who the target is and of the quantitative transition rates. In addition, the conventional methods help conducting simple statistical calculation to obtain similar transition results, but it prerequires managers to know all of the users' behavior types and which parts of bus service need to improve. Actually, users' behavior type may vary at different period and change time by time. There are also many diverse and potential problems in whole bus service. This research can bring out behavior transition information from overall bus transaction data and allow managers to find the significant problems in whole bus service via an efficient procedure. In a future study, more variables, like routes or ODs, should be considered to enhance the interpretability of clustering results.

ACKNOWLEDGEMENTS

The authors would like to express special thanks to the Department of Transportation in Tainan, Taiwan for providing e-ticket data and GIS map data of Tainan.

REFERENCES

Arentze, T. A., Ettema, D., Timmermans, H. J. P. (2011). Estimating a model of dynamic activity generation based on one-day observations: Method and results. *Transportation Research Part B: Methodological*, 45(2), 447–460.

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

Boyle, D. (2006). *Fixed-Route Transit Ridership Forecasting and Service Planning Methods*. Washington, D.C.: Transportation Research Board.

Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285–299.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, 39(1), 1–38.

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. In *The 3rd International Workshop on Urban Computing (UrbComp 2014)*. New York.

FHWA. (2007). *The Transportation Planning Process: Key Issues A Briefing Book for Transportation Decisionmakers, Officials, and Staff*. Retrieved from http://www.planning.dot.gov/documents/briefingbook/bbook.htm

Flyvbjerg, B., Skamris Holm, M. K., Buhl, S. L. (2005). How (In)accurate Are Demand Forecasts in Public works. *Journal of the American Planning Association*, 71(2), 131–146.

Goulet-Langlois, G., Koutsopoulos, H. N., Zhao, J. (2016). Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies*, 64, 1–16.

Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, 66(1), 1–6.

Järv, O., Ahas, R., Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies*, 38, 122–135.

Kamruzzaman, M., Hine, J., Gunay, B., Blair, N. (2011). Using GIS to visualise and evaluate student travel behaviour. *Journal of Transport Geography*, 19(1), 13–32.

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.

Ortega-Tong, M. A. (2013). *Classification of London's Public Transport Users Using Smart Card Data*. MIT.

Pas, E. I. (1988). Weekly travel-activity behavior. *Transportation*, 15(1–2), 89–109.

Pelletier, M.-P., Trépanier, M., Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557–568.

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.

USF Center for Urban Transportation Research. (2009). *Best practices in transit service planning*. Retrieved from http://www.nctr.usf.edu/pdf/77720.pdf

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

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.