

Mining temporal and spatial travel regularity for transit planning

Le Minh Kieu¹, Ashish Bhaskar¹, Edward Chung¹

¹Smart Transport Research Centre, Faculty of Built Environment and Engineering, Queensland University of Technology, Brisbane, Australia

Email for correspondence: leminh.kieu@qut.edu.au

Abstract

Smart Card data from Automated Fare Collection system has been considered as a promising source of information for transit planning. However, literature has been limited to mining travel patterns from transit users and suggesting the potential of using this information. This paper proposes a method for mining spatial regular origins-destinations and temporal habitual travelling time from transit users. These travel regularity are discussed as being useful for transit planning. After reconstructing the travel itineraries, three levels of Density-Based Spatial Clustering of Application with Noise (DBSCAN) have been utilised to retrieve travel regularity of each of each frequent transit users. Analyses of passenger classifications and personal travel time variability estimation are performed as the examples of using travel regularity in transit planning. The methodology introduced in this paper is of interest for transit authorities in planning and managements.

1. Introduction

Policies have been issuing by transit agencies in order to increase the attractiveness of public transportation by offering better services. Understanding travel patterns of individual transit users are helpful for providing personal oriented benefits and information. More importantly, knowledge on individual travel patterns can facilitate strategic transit planning (Utsunomiya et al., 2006). Traditionally, the individual travel pattern is explored by surveying the passengers. However, the traditional travel diaries approach is rather costly and time consuming. More importantly, the results are only valid at the certain survey time and within the surveyed respondents.

Automated fare collection system using Smart card in Brisbane is collecting large quantities of individual travel transactions. The Smart card data facilitates a continuous, multi-day method for exploring the travel pattern of the whole population of card users. Compared to the traditional travel survey approach, the travel pattern analysis using Smart card is much less costly and can be carried out continuously for monitoring the transit system. Moreover, the sample size is also much bigger as all Smart card users can be considered, compared to a limited number of respondents in travel survey approach.

Recently, a noticeable number of studies have been published using Smart card data. Itinerary reconstructions, transferring estimations, origin-destination (OD) assessment and regularity are among the explorations of travel patterns from Smart card data in literature. Authors in the literature have successfully connected individual transactions from Smart card users to reconstruct their itineraries (Bagchi and White, 2004, Chu and Chapleau, 2008, Chu and Chapleau, 2010, Ma et al., 2013, Jang, 2010). Jang (2010) analysed the transfer patterns of passengers in Seoul, Korea. Travel time of trips using subway was concluded as generally faster than the ones using bus. Farzin (Farzin) integrated Bus stop location, Automatic Vehicle Location and Smart card data for an automatic OD matrix construction study in Sao Paulo, Brazil. Chu and Chapleau (2010) explored multi-day travel behaviours of subgroups and individual cardholders in Ontario, Canada. "Anchor points" or the regular boarding locations of each cardholder were identified and visualised. On a recent study, Ma

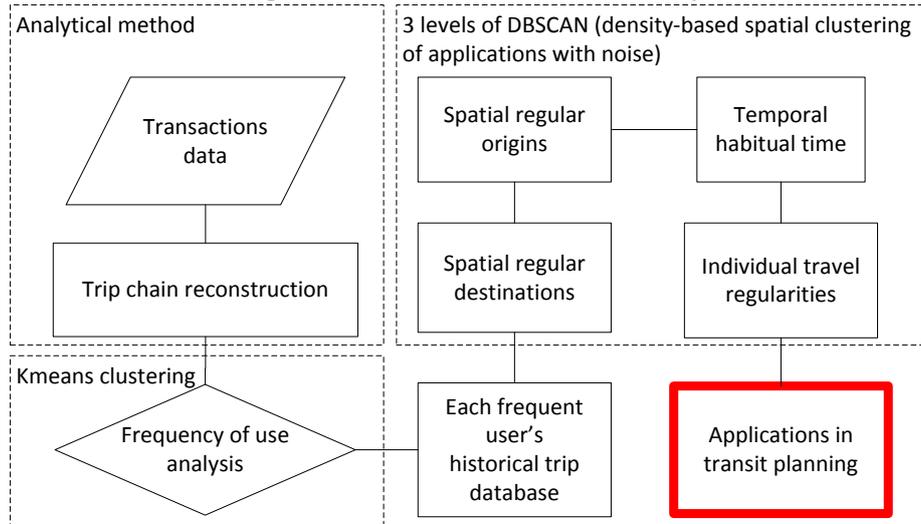
et al. (2013) mined the travel patterns from Smart card data. The authors classified the transit users to 5 levels of regularity based on temporal and spatial travel patterns of users. Existing studies have given some insights into the individual travel patterns. Most of the authors utilise data mining and visualisation techniques to expose interesting information from the Smart card data and suggesting potential usages of the travel patterns. However, the studies which aim for using the knowledge of travel patterns for practical use in strategic transit planning are still limited. Seaborn et al. (2009) suggested that Smart card data could provide useful knowledge and could be the inputs for quantitative cost-benefit models. Utsunomiya et al. (2006) suggested many practical applications of Smart card data analysis including market research, service planning, demand forecasting, pricing and marketing, etc. The unavailability of trip purpose information hinders the use of travel patterns from Smart card data in transit planning. The individual trip purposes could introduce biases to the aggregated analysis of the whole system. For instance, the excessive total travel time and transfer time of an individual might source from poor transit service, or the traveller might intentionally lingered his/her transferring time for personal reasons such as shopping or dining. Therefore, travel patterns alone might not be able to facilitate transit planning. The combination of travel pattern and travel purpose inference is then essential to mine only the needed data for transit planning applications.

This paper explores the use of individual travel patterns for strategic transit planning using Smart card data in Brisbane, Australia. We propose a new method for travel pattern and travel purpose inference by exposing spatial and temporal regular travel patterns from each cardholder. The origins and destinations that the cardholder usually travels between are defined as “regular OD”, while the regular time of travel for each regular OD is called “habitual time”. The regular OD and habitual time can be generally called “travel regularity”. Mining only the travel regularity could facilitate inferring the trip purposes. For instance, it is fair to assume work/study trips if the traveller starts the trips habitually at around 7:30 AM from a regular origin and arrives a few minutes before 8 AM at a regular destination in the CBD. It is unlikely that the traveller would intentionally delay his/her trips in this case. The possibility of inferring the trip purpose through travel patterns has also been discussed by existing studies in the literature (Chu and Chapleau, 2008, Chu and Chapleau, 2010). In this paper, we focus only on the travel regularity of the frequent users.

The framework of study is illustrated in Figure 1. Using temporal and spatial characteristics of each cardholder’s transactions on a given day, his/her complete trips are reconstructed. The OD, number of transfers, mode and route uses, total time and transfer time are derived for each trip. Using K-means clustering on the number of trips, the frequent users are identified. Analysing each individual frequent user’s trip database, we apply three level of Density-Based Spatial Clustering of Application with Noise (DBSCAN) for finding his/her travel regularity. The information is then used for applications in transit planning.

The paper starts with a short discussion of the potential of using temporal and spatial travel regularity in transit planning. The next section describes the reconstruction process for completing daily itineraries from the cardholders. The travel pattern analysis using K-means clustering and DBSCAN is explained in the following section. Before ending the paper with the conclusion, an analysis of using individual travel patterns on strategic transit planning is discussed.

Figure 1 Generic framework of study



2. Potential of using travel regularity in transit planning

Travel regularity is essential components of individual travel patterns. The potential of using them for strategic transit planning is discussed in this section.

2.1 Classifying the passengers by frequency and regularity of use

Passenger classification is possibly the direct application from the travel regularity. Classification of cardholders by frequency and regularity of use should considerably benefit the operators in providing better personal oriented services and information to transit users. Cardholders can be clustered into infrequent users, frequent user without a regular OD, frequent user with regular ODs but without a habitual time and frequent user with regular ODs and habitual times.

Infrequent users are not loyal users of the system. The information given to them should be marketing, advertisement or special offers in order to attract more usages. Frequent users without a regular OD should be given different transit options (such as information given by a public transport planner) to travel to multiple locations. Frequent user with regular ODs but without a habitual time should be assisted with the predicted travel time and reliability of travel time on different time-of-day. Finally, passenger-oriented real-time information on incidents, service changes can be provided to frequent user with regular ODs and habitual times before their trips.

Smart card data can be used to mine travel regularity from each user. Temporal and spatial travel regularity can facilitate the classification of passengers into the aforementioned 4 types.

2.2 Using travel time regularity in personal travel time variability

In the literature, travel time variability (TTV) is measured as the variance in a vehicle's travel time between two points. Day-to-day TTV receives most of the attention from the existing studies (Noland and Polak, 2002) as it measures the variance of travel time of the same trip on multiple days. It shows how much unreliability or inconsistency a transit trip is. However, the traditional way to measure TTV limits to travel times of a single journey on a single vehicle. Smart card data could extend the definition of TTV by exploring the personal experience of variability on the whole trip between origin and destination location, including transfer time. We define this type of TTV as "personal TTV".

Personal TTV has advantage over traditional TTV as the direct means of unreliability of travel time. For instance, a traveller made a trip between point A and C by firstly took a transit route between A and B, made a transfer at B, and finally boarded the second transit

vehicle to travel between B and C. The traditional TTVs of the two trips A-B and B-C may be not high, which overlooks the fact that the traveller could spend excessive transfer time between the two trips. Poor timetable coordination, late arrival of the first and early arrival of the second vehicle are associated with the unreliability of the total travel time in this case. Thus, personal TTV acts as a complete measure of transit performance by reflecting all sources of inconsistency on a personal trip. It can be used by transit operators for monitoring their system, or by transit users in multimodal trip planner as a utility of the mode or route choice.

2.3 Deriving dynamic OD matrices of regular transit usages

Estimating OD matrices is considered as essential for transportation planning as it represents the traffic demand (Bert et al., 2008). Smart card data can be utilised for deriving dynamic public transport OD matrices. This problem has been addressed by a number of authors in the literature (Chu and Chapleau, 2010, Farzin, 2008, Trépanier et al., 2007).

Dynamic OD matrices of regular users could bring additional information compared to OD matrices of the whole cardholders' population. First, regular users are the most loyalty users to the service. Loyalty is one of the most important determinants for the growth of any transit system (Foote et al., 2001). For maintaining the customer satisfaction and loyalty of the system, the areas where many regular users are located could be planned with new routes, or additional transit vehicles. The location where most of regular users spend excessive time (low accessibility) to travel to the CBD can be invested with more transit facilities. In a survey conducted in 2006 in Australia, the most commonly reason for non-transit users to go to work/study was unavailability of public transport service at a preferable time (Australian Bureau of Statistics, 2008). Urban planners can also evaluate the transit-oriented development (TOD) through the regularity of transit usages (Dill, 2008). Second, the whole population OD matrices could be biased on the day of survey, for instance due to incidents or special events. Conversely, the OD matrices of regular users could be considered as unchanged and unbiased for a noticeable amount of time. It facilitates transit planners to predict and evaluate the impact of policies to the movements of transit users.

The OD matrices of regular users can be estimated using the location of the first boarding stops and last alighting stops of frequent users.

2.4 Investigate the changes in travel regularity due to changes in the transit system

Policy changes in transit management affect passenger travel regularity or behaviour. For instance increasing peak periods transit fare and reducing off-peak periods fare would encourage more passengers to switch from using public transport during peak to off-peak periods. Investigating travel regularity before and after the policy changes would reveal the changes in passenger behaviour due to the adjustment of the system. The information is essential for policy makers to evaluate the applied management schemes and optimise the new plans.

2.5 Developing a multimodal public transport travel assistance

The last application of travel regularity benefits the transit user. By exploring the different personal TTV and total travel time of his/her own historical trips and the trips of people around his/her area to the same destination, some useful information can be provided. A total travel time prediction model that forecasts the travel time between origin and destination stop can be developed. Another application of the model could be suggesting the best time to board the first transit vehicle in order to reach the destination on time. Using the input as the preferred alighting time at the final destination, the model uses historical regular trip database to predict the total travel time and the first boarding time at origin location. The prediction algorithm could be used in a travel assistance mobile application that supports transit users.

3. Case study and data description

The research is carried out from the data provided by Translink, which is the transit authority of South East Queensland, Australia. The dataset is a compilation of around 10 million transactions made by 996,132 card holders over the bus, city train and ferry network of Brisbane city, Australia from 1st March to 30th June 2012. Each transaction includes both boarding and alighting time when the cardholder gets on and off a transit vehicle, but not transferring activities. Other information includes Card ID, ticket type, route used and boarding/alighting stop. The Card ID is hashed into unique numbers for maintaining privacy of each cardholder. The Table 1 shows the first 5 transactions of an anonymous cardholder on June 2012.

Table 1 Sample transactions of a cardholder on June 2012

Day	Route	Service	Direction	Boarding Time	Alighting Time	Ticket type	Boarding stop	Alighting stop
20-Jun	385	7059	OB	10:38	10:45	SV Adult	Roma St Bus Station Platform 1	'Paddington Central' Latrobe Terrace - 10
5-Jun	333	5869	OB	17:48	17:55	SV Adult	King George Square Station 1D	Royal Brisbane & Women's Hospital PL 2
5-Jun	330	6237	IB	19:55	20:01	SV Adult	Royal Brisbane & Women's Hospital PL 1	Roma St Bus Station Platform 2
5-Jun	385	7059	OB	20:17	20:39	SV Adult	Roma St Bus Station Platform 1	'Hilder Road' Waterworks Road
8-Jun	385	7059	OB	18:20	18:42	SV Adult	King George Square Station 1C	'Hilder Road' Waterworks Road

The text-based information on date, direction, ticket type, boarding and alighting stop is converted in to number-based format for reducing the expensiveness of the large dataset analysis. Unique ID is assigned for each bus/city train/ferry stop. The analysis in this paper focuses only on working days (weekdays excluding public holidays and school holidays) and the trips within the Brisbane area.

4. Reconstruction of travel itineraries

The first step to derive travel regularity is through reconstructing the travel itineraries, or trip chains. The transactions from each cardholder for each given working day are analysed using logical analytical approach. The objective is to combine related transactions into completed trips from origin to destination, including the transfer; and to differentiate these completed trips. The workflow is illustrated on Figure 2.

From the transaction database, each transaction from a specific cardholder for each working day is analysed. A "reconstructing indicator" valued between 0 or 1 is used for decide if the cardholder is starting at an origin location or currently in the middle of a trip. First, we filter out some noises such as uncompleted transactions (e.g. lack of alighting information), long-distance transactions (e.g. trips with origin or destination beyond Brisbane area) and transactions with boarding is the same as alighting location (which is likely a mistaken boarding). Second, a fixed threshold of 60 minutes is used to identify whether the two transactions are connected. This threshold is chosen differently in the literature. It ranged from 30 minutes (Bagchi and White, 2004), 60 minutes (Ma et al., 2013), 90 minutes (Hofmann and O'Mahony, 2005) or a set of thresholds (Seaborn et al., 2009). The 60 minutes is chosen in accordance with Brisbane's public transport threshold for transferring trips (Translink, 2007). Third, if the following transaction's alighting stop is not similar with the identified origin (first boarding stop); the transactions are keeping connected until the last transaction. This step is for segmenting the behaviour of coming back home within the 60 minutes transfer threshold into a new trip. Thus, any completed trip would have a separated

origin and destination. The Card ID along with information on boarding/alighting activities and route choices are recorded. However, the city train/subway routes information is not available in the database. All city train routes are coded as 999 in this study. An example of the completed trip chains is showed in the

Table 2. Here *Trip ID* is the assigned unique identification number for each completed trip. *First Board Time* and *Last Alight Time* measure the timestamp of first boarding at the Origin stop and last alighting at the Destination stop for each *Trip ID*. *Stop ID Sequence* and *Route ID Sequence* show which stops and routes that the passenger used during the studied *Trip ID*.

Figure 2 Travel itineraries reconstruction process

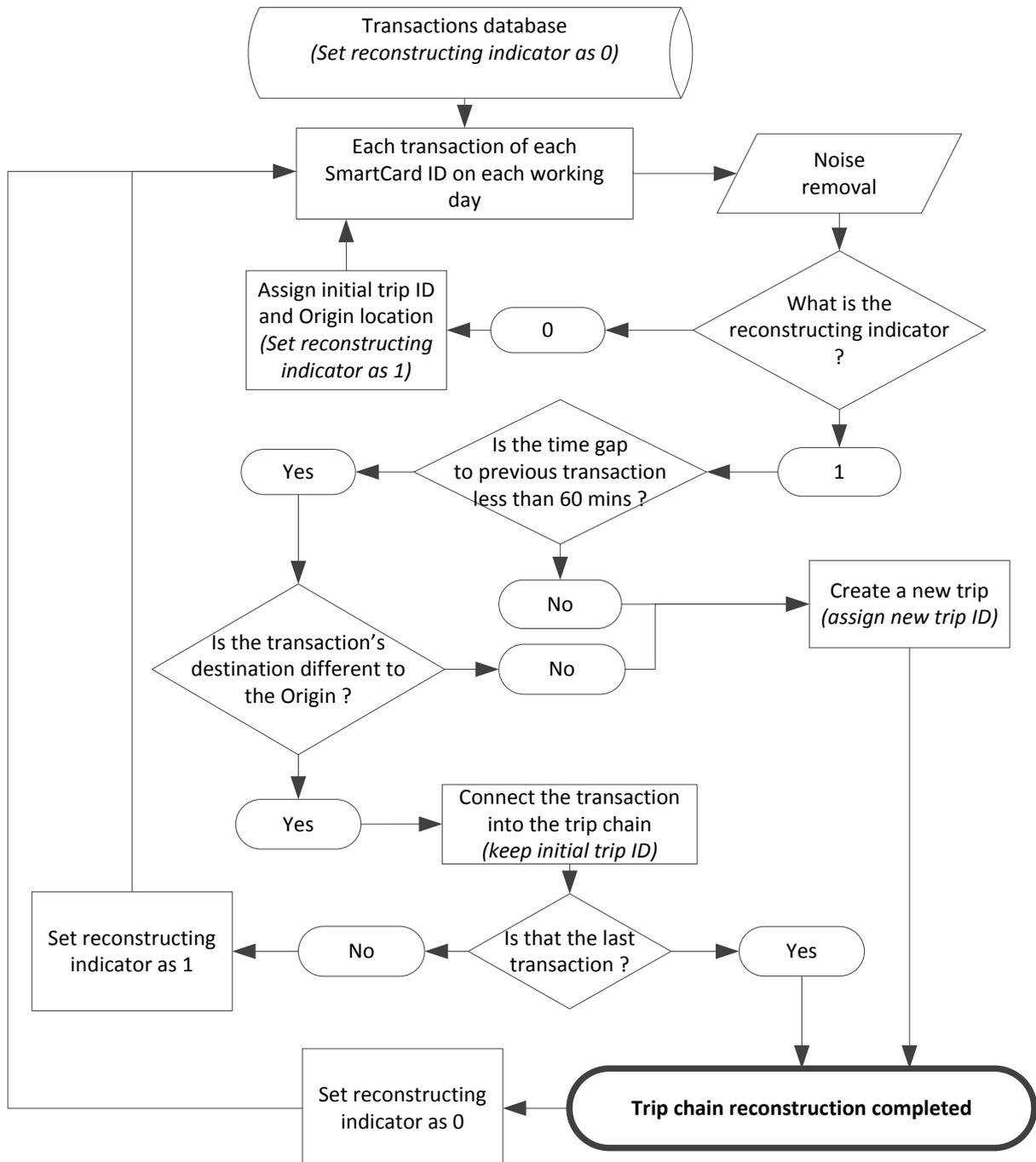


Table 2 Example of completed travel itineraries

Card ID	Day	Trip ID	First Board Time (h)	Last Alight Time (h)	Stop ID Sequence	Total travel time (min)	Total transfer time (min)	Total time (min)	Route ID Sequence
X1	1-4	1697	11:20	11:27	5->4	6.34	0	6.34	999
X2	1-4	1083	9:17	9:27	54371->24653	9.9	0	9.9	726
X3	1-4	1415	14:58	15:46	12861->2452 ->15212	7.16	41.42	48.58	550->562
X4	1-4	1412	14:47	16:17	10730->499->88	57.18	32.7	89.88	690->999

5. Mining spatial and temporal travel regularity

Travel time regularity of frequent users can be mined from the trip itineraries. The first step to be done is cluster the cardholders into frequent users and infrequent users. The second step is to retrieve the spatial and temporal travel regularity from each frequent user's historical trip database.

5.1 Identifying frequent users

The frequent users and infrequent users are clustered according to the number of trips within the dataset. The travel regularity would be retrieved from only the frequent users. Infrequent users are not transit-dependent, i.e. they are likely having other main mean(s) of transport. In this paper we focus on public transport travel regularity (repeated temporal and spatial travel patterns) which is unlikely the patterns of the ones mainly using private transport.

K-means clustering algorithm is applied to the set of each cardholder's transit usages to find a threshold that differentiate between frequent use and infrequent use. K-means is chosen on account for its high classification performance. With the number of cluster equals to 2 (frequent and infrequent users), the threshold has been determined as 53 trips. Since there are 72 working days between March and June 2012, the threshold is equal to taking at least 1 trip/day on at least 75% of the working days. An example of the clustering results is represented in the Table 3. Cluster 1 refers to infrequent users while cluster 2 refers to frequent users.

Table 3 Example of clustering results using K-means with 2 clusters

Card ID	Trips completed	Cluster
X1	5	1
X2	62	2
X3	124	2
X4	86	2
X5	2	1
X6	54	2

Among 996,132 unique Card ID that has been identified in the dataset, 78.1 % of them (769,837 IDs) can be labelled as infrequent users. The remaining 21.9% (226,295 IDs) cardholders are frequent users. They are hereafter the main focus of the study. The analysis of whole population of cardholders would be discussed in future studies.

5.2 Mining spatial regular origins-destinations and temporal habitual travel time regularity

Exploring each frequent user's historical trip database, travel regularity can be exposed. In this paper we adopt the DBSCAN algorithm to cluster repeated spatial and temporal patterns (Ester et al., 1996). DBSCAN is a density-based clustering algorithm which automatically decides the number of clusters from the density distribution of points. The algorithm is resistant to noise, can handle clusters of different shapes and sizes as high density records are grouped with each other. Most importantly, number of clusters and initial cluster centres are not predetermined.

However, it is very important to choose the optimal density reach distance ϵ and the minimum number of points *MinPts*. A point i_c can be considered as a "core point" if it has more than *MinPts* points within the distance of ϵ . A point i_b can be considered as a "border point" if it does not have more than *MinPts* around, but it lies within the range of a core point. A point i_n can be considered as a "noise" it is not a core or a border point. A combination of core points i_c forms a cluster. The value of *MinPts* and ϵ should be carefully determined as it noticeably affected the results of clustering process.

The process of DBSCAN clustering applied in this paper is explained by the help of an example. The Figure 3 illustrates the full historical travel itineraries of an anonymous traveller of the morning trips (alighting timestamps less than 10). Here, the A points represent the first boarding stops, C points represent the last alighting stop where the traveller start/end the trips and B points represent the transfer points. The stops are clustered in this section according to their spatial characteristic – their UTM coordination. The objectives of the clustering are to find repeated travel patterns and filter the noises. The noises are the trips which do not belong to any patterns.

We apply two levels of DBSCAN. On Figure 3, the number $[X,Y,Z]$ shows the results of DBSCAN clustering. The value of X represents the number of trips from that particular location. The value of Y refers to the assigned cluster on the first level, while the value of Z (if any) refers to the assigned cluster on the second level of DBSCAN. If the value of either Y or Z is -1, it means that the point is considered as a noise on the respective clustering level.

- The first level analyses the whole travel itineraries of the passenger. It clusters the last alighting points (C points) into regular cluster(s) and filters the noises. It is important to notice that for underlying the repeated patterns, each trip's last alighting stop is considered as a point in the database. For instance, the point $C[42, 1]$ on the Figure 3 is considered as 42 points at the same coordination. The regular destinations are clustered in this step.
- The second level examines the database of first boarding points of each cluster from the first level. In other words, we explore each value of Y . The first boarding points are clustered into their own clusters and noises are filtered. For instance on Figure 3, three patterns $A[32, 1, 1]$, $A[8, 1, 2]$, $A[2, 1, -1]$ are from the cluster 1 of Y . Their values of Z show that they are clustered into two clusters, and the last pattern $A[2, 1, -1]$ is considered as noises. The regular origins are identified.

The DBSCAN is applied on both levels with ϵ is set as 1, and *MinPts* is set as 8. The value of ϵ denotes that the passengers can walk to any stop within a 1 km Euclidean distance. Keeping ϵ constant, we analysed the clustering results of different value of *MinPts* and the value of 8 is chosen. This is to ensure that the clustering algorithm will be effective on both noise filtering and pattern clustering. The value of *MinPts* also denotes that among 16 weeks of the dataset, the cardholder should follow the clustered travel regularity by at least 1 trip/week on 50% of the weeks. The combination of regular origin-destination is called Regular OD.

Another level of DBSCAN is performed from the database of each value of Z for finding repeated time patterns for each regular trip. For instance, we analysed 32 trips with $Y=1$ and

$Z=1$ that started from point $A[32,1,1]$. We apply 1-variable DBSCAN to cluster the first boarding timestamps (in hour) of the trips in the database. The reason is that the cardholder can decide when he/she leaves the house but not the arrival time at the final destination stop. DBSCAN is applied with ϵ equals to 0.25 (15 minutes) to take into account the variability in the transit vehicle arriving times and the value $MinPts$ equals to 3. The habitual boarding time is explored in this level of DBSCAN. The Figure 4 sums up the workflow of the three levels of applied DBSCAN.

Figure 3 Travel regularity of a traveller in the database

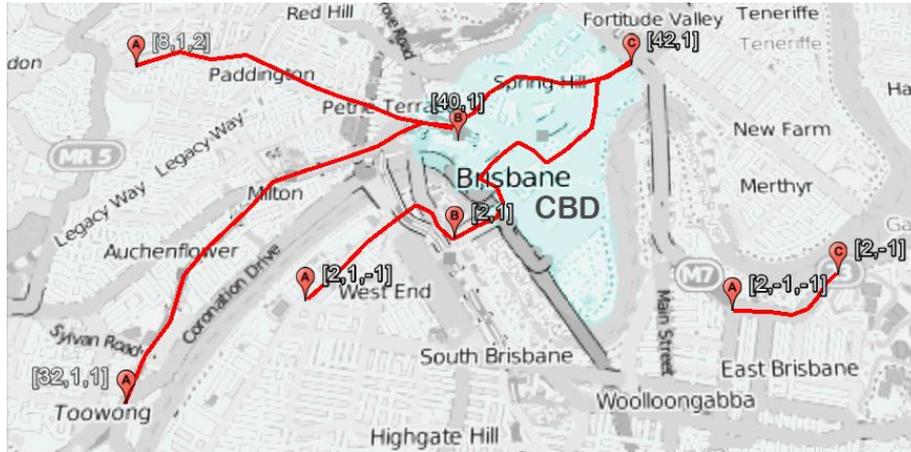
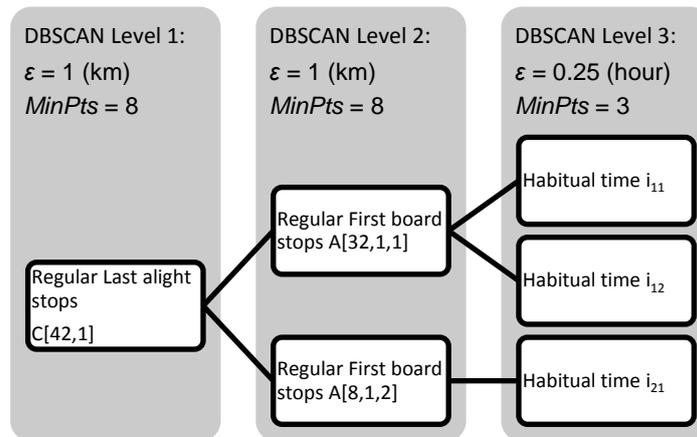


Figure 4 Three levels of DBSCAN workflow



The Table 4 shows an example of 3 cardholders' travel regularity.

Table 4 Example of travel regularity

Card Id	Reg. OD	Per. Reg. Trip	O	D	Habit. time ID	Per. Habit Trip	Ave. First Board Time	Mean Last Alight time	Number of Trips	Route sequence
X1	1	43.02	51985	52105	1	100.00	7.91	8.38	37	420
X1	2	31.40	58734	51985	1	100.00	15.23	15.58	27	458
X2	4	20.44	43642	1882	1	64.20	8.05	9.35	8	999->370
X2	4	31.67	8	1882	1	18.42	8.64	9.58	14	999->370
X2	4	31.67	8	1882	2	36.84	6.63	7.81	18	999->370

X3	2	9.52	1878	2890 2	1	41.67	17.1 1	18.33	5	543->141 ->139
X3	3	10.32	1878	2888 9	1	61.54	17.2 6	18.71	8	542->141- >169

Cardholder ID 3370606 travelling on two regular ODs. The first regular OD proportioned for 43.02% of his total trips while the second one proportioned for 32.4%. The cardholder is a very habitual traveller in terms of time. On each regular OD, 100% of the trips can be clustered into the same cluster, which means he/she always board a transit vehicle around the habitual time. Route 420 is used for morning and route 458 is used for afternoon trips. The Table 4 shows that travel regularity of each cardholder can be identified using the proposed method.

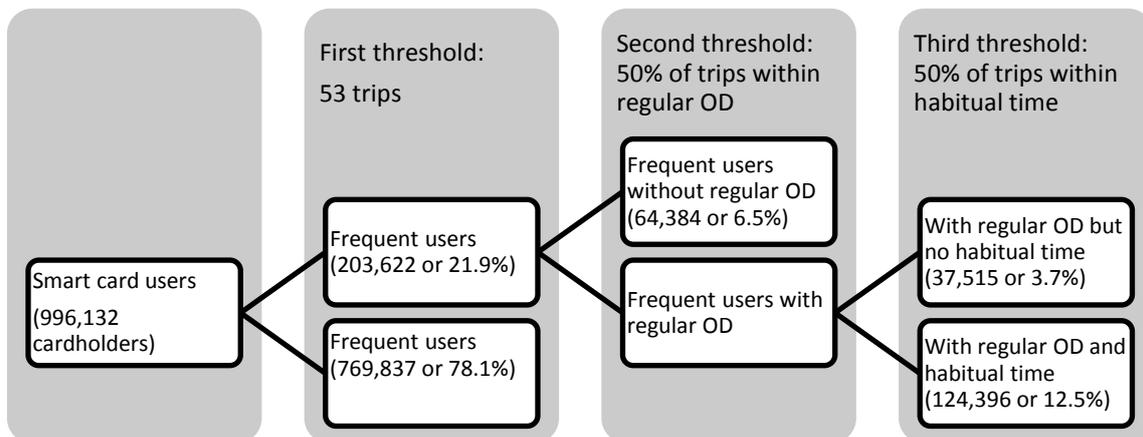
6. Application in transit planning

Once the travel regularity is found, the information can be used for transit planning. The section 2 describes some possible applications of the travel regularity. This section explores the first two applications on classification of passengers and personal travel time variability estimation.

6.1 Passenger classification

The classification of passengers into 4 types has been discussed in the section 2 of this paper. With identified travel regularity of passengers, the classification is relatively straightforward if a threshold is chosen. In this paper, except the first threshold of 53 trips (decided by the K-means clustering method), we choose 50% as the threshold. The detail clustering using a hierarchical tree structure is illustrated in Figure 5. As could be seen from Figure 5, by analysing each passenger’s historical trip database, the type of passengers can be identified.

Figure 5 Framework and results of transit passengers classification



6.2 Personal travel time variability estimation

The personal variability of each passenger can be explored by analysing only the travel regularity, i.e. the clustered trips with regular ODs and habitual time that we observed from the section 5 of this paper. Analysing each clusters and calculating the Coefficient of Variation (CV), we can find the personal travel time variability of each user’s trips. CV measures the variation as a percentage of the mean, or the ratio of the SD to the mean.

$$CV = \frac{SD_{TT}}{\overline{TT}} \times 100$$

Where: CV = Coefficient of variation (%)

SD = Standard deviation of travel times

\overline{TT} = Mean of travel times

CV is chosen as a meaningful comparison between two or more magnitude of variations. The Table 5 shows an example of travel time variability calculation for each cardholder's regular trips during the morning period. The information is in accordance with the example in Table 4. The personal travel time variability should be the most trustworthy performance indicator, as the personal experience of the travellers are recorded. Examining the travel time variability of each passenger, the transit authorities can find routes and services with unreliability.

Table 5 Example of personal travel time variability calculation

Card Id	Regular OD	O	D	Habit. time ID	Number of Trips	Ave. Number Transactions	Ave. Transfer time	Ave. Total time	TTV (%)
X1	1	51985	52105	1	37	1.00	0.00	20.61	17.20
X1	2	58734	51985	1	27	1.00	0.00	28.22	13.25
X2	4	43642	1882	1	8	2.00	2.14	77.95	8.12
X2	4	8	1882	1	14	2.00	1.87	56.71	5.27
X2	4	8	1882	2	18	2.00	6.09	70.32	5.69
X3	2	1878	28902	1	5	3.00	6.41	73.30	12.58
X3	3	1878	28889	1	8	3.00	19.94	87.27	12.47

7. Conclusion

Literature on using Smart card data in strategic transit planning is limited to information provision. Most of the existing papers retrieved information of the travel patterns from the cardholders and suggested the potential of using these patterns. This paper proposes a methodology of mining travel regularity and discusses several applications of this information in transit planning.

Travel regularity is spatial regular ODs and temporal habitual boarding time of the frequent transit users. First, individual transactions of each cardholder on each day are combined to reconstruct travel itineraries. The reconstruction process follows a logical structural approach with a 60 minutes threshold to identify transfer activities. Second, the passengers are clustered into frequent and infrequent users according to the number of trips taken during the study period using K-means clustering algorithm. Third, three levels of DBSCAN have been performed to explore the regularity of each cardholder. The data mining result shows that the repeated patterns can be observed. Finally, the travel regularity is used for two quick analyses of passenger classifications and personal travel time variability estimation.

This paper is the initial step of using travel regularity in practical transit planning. Further extensions of the study are in progress, where travel purpose inference is improved by land-use information and more planning applications are introduced using the travel regularity.

8. Reference

- Australian Bureau of Statistics 2008. Public transport use for work and study. Assessed at <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Chapter10102008> on 10:07PM 13rd June 2013.
- Bagchi, M. & White, P. What role for smart-card data from bus systems? Proceedings of the Institution of Civil Engineers. Municipal engineer, 2004. Institution of Civil Engineers, 39-46.
- Bert, E., Chung, E. & Dumont, A.-G. Urban dynamic origin-destination matrices estimation. 15th World Congress on Intelligent Transport Systems, 16 - 20 November, 2008 United States of America, New York, New York City.: ITS Congress Association.

- Chu, K. & Chapleau, R. 2008. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2063, 63-72.
- Chu, K. K. A. & Chapleau, R. 2010. Augmenting Transit Trip Characterization and Travel Behavior Comprehension. *Transportation Research Record: Journal of the Transportation Research Board*, 2183, 29-40.
- Dill, J. 2008. Transit use at transit-oriented developments in Portland, Oregon, area. *Transportation Research Record: Journal of the Transportation Research Board*, 2063, 159-167.
- Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Second international conference on knowledge discovery and data mining, 1996. Amer Assn for Artificial, 226.
- Farzin, J. M. 2008. Constructing an automated bus origin-destination matrix using farecard and global positioning system data in Sao Paulo, Brazil. *Transportation Research Record: Journal of the Transportation Research Board*, 2072, 30-37.
- Foote, P., Stuart, D. & Elmore-Yalch, R. 2001. Exploring Customer Loyalty as a Transit Performance Measure. *Transportation Research Record: Journal of the Transportation Research Board*, 1753, 93-101.
- Hofmann, M. & O'mahony, M. Transfer journey identification and analyses from electronic fare collection data. Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, 13-15 Sept. 2005. 34-39.
- Jang, W. 2010. Travel time and transfer analysis using transit smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2144, 142-149.
- Ma, X., Wu, J. X., Wang, J., Chen, F. & Liu, J. Mining Smart Card Data for Transit Riders' Travel Patterns Transportation Research Board 92th Annual Meeting 2013 Washinton D.C.
- Noland, R. B. & Polak, J. W. 2002. Travel time variability: a review of theoretical and empirical issues. *Transport Reviews*, 22, 39-54.
- Seaborn, C., Attanucci, J. & Wilson, N. 2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2121, 55-62.
- Translink 2007. How to use your go card on the TransLink network: TransLink go card user guide (part 1 of 2).
- Trépanier, M., Tranchant, N. & Chapleau, R. 2007. Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*, 11, 1-14.
- Utsunomiya, M., Attanucci, J. & Wilson, N. 2006. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning. *Transportation Research Record: Journal of the Transportation Research Board*, 1971, 119-126.