TRIP PURPOSE INFERENCE USING AUTOMATED FARE COLLECTION DATA

Sang Gu Lee and Mark Hickman*

Civil Engineering and Engineering Mechanics
University of Arizona, 1209 E. 2nd Street
P.O. Box 210072, Tucson, AZ 85721-0072

Submitted for Presentation at
4th Transportation Research Board Conference on
Innovations in Travel Modeling, April 30 – May 2, 2012

Revision submitted March 2, 2012

ABSTRACT

In this paper we exploit the extensive farecard transaction data for deriving useful information about transit passenger behavior, namely trip purpose or activity. We show how the farecard data can be used to infer trip purpose and to reveal travel patterns in an urban area. A case study demonstrates the process of trip purpose inference based on farecard data from Metro Transit in the Minneapolis-St. Paul metropolitan area.

Keywords: AFC data, Heuristic rules, Learning algorithms, Trip purpose inference

* Corresponding Author: mhickman@email.arizona.edu; Tel: (520) 626-9420; Fax: (520) 621-2550

1. INTRODUCTION

There is a huge amount of data that is readily available from transit automated fare collection (AFC) systems describing spatial and temporal information of travel patterns. By manipulating and synthesizing these records, we can make more meaningful data sets which give us insight into travel behavior. Assuming that we have such data, one might ask many questions, such as the following:

1. Can the AFC data be utilized in transit service planning procedures?
2. Is it possible to replace the costly conventional origin-destination survey by the use of the AFC data, in conjunction with other readily available data?
3. Can we solve the inherent weakness of the AFC data, in that it contains no information of the passengers' trip purpose and/or socio-economic characteristics?

There have been many efforts to solve the first and second questions. This work can be grouped into two categories: customer behavior analysis (e.g., travel pattern analysis) (*1,2,3,4,5*) and demand forecasting (e.g., origin-destination estimation) (*6,7,8,9,10,11,12*). Bagchi and White (*13*), however, point out that the absence of trip purpose and activity information from the passenger is an intrinsic limitation of the data itself. Trépanier *et al*. (*14*) enrich the AFC data by connecting the data with household travel surveys. In contributing to this line of research, the goal of this study is to explore and implement a potential method of deriving trip purpose from the AFC data.

Using GPS log data from vehicle- and person-trips, Wolf (*15*) proposes to use land use types at the trip end as the primary means to identify trip purpose. In a more recent study, Stopher *et al*. (*16*) collect the address of the respondent's home and work place or school, the two most frequently used grocery stores, and occupation, to enhance their method of deriving trip purpose. Bohte and Maat (*17*) also propose an innovative method that combines GPS logs, GIS technology and an interactive web-based validation application. In contrast, using transit AFC data, our approach to inferring trip purpose is based on the assumption that every transaction is made within a sequential trip chain. This approach is conventionally appealing when farecard data are analyzed. Such an assumption on trip chaining implies that the destination of a trip is also the origin of the following trip.

Compared to the passive GPS log data, the AFC data are more limited in the spatial and temporal dimensions, since a transaction of the AFC is recorded only when the transit passenger swipes his/her farecard, typically when they board a public transit vehicle. For this reason, the following assumptions are added.

- Transit users do not walk a long distance to board at a different stop from the one where they previously alighted.
- Transit users do not use any other modes within their given sequence of daily transit trips.

In addition, there are many conventional travel patterns, especially for transit users. Previous studies describe these travel patterns with different views of point. Kitamura *et al*. (*18*) point out the activity duration and home location play a role in the destination choice decision making. Hannes *et al*. (*19*) reveal individuals' daily activity travel decision processes in general and spatial factors influencing destination choices specifically, using simple if-then conditions. Regularity of people's movement on public transport is observed by McNamara *et al*. (*20*) using data collected from an RFID-enabled subway system. By means of smart card data with time and space resolution (e.g., frequency of stop used), Liu *et al*. (*21*) show the regularity of the mobility patterns, specifically that most people are only active at a limited number of locations. Also, duration is directly related to the activity occurring between trips, and thus can provide us with regular activity patterns in inferring a passenger's trip purpose (*22*). Based on these findings, the following observed patterns can serve as a strong foundation to generate heuristic rules about transit passenger behavior.

- A transit user (e.g., commuter) visits the same location on multiple days in the same week.
- A transit stop used frequently is near to the transit user's work place or home.
- A specific activity has a certain understood time duration.

Bringing together all these aspects can help establish a training set for inferring a transit passenger's trip purpose.

User Information: Bus Passengers' Travel Patterns by Farecard Types

Lee and Hickman (*22*) found that activity and travel patterns differ significantly across the different farecard types, at least comparing Metro Pass[†] (MP) and Stored Value[‡] (SV) card holders with the U-Pass[§] (UP) and College Pass[**] (CP) holders. MP and SV users tend to have very similar travel patterns in terms of temporal and spatial aspects, while UP and CP users show different characteristics. Temporally, for MP and SV users, most transactions are made during the morning and evening peak periods, which seems to be consistent with a conventional commuter travel pattern. Compared to MP and SV users, UP and CP users show not only much less concentration during rush hours, but also a longer distribution of transactions toward the end of the day. The time duration between farecard transactions at origin-destination pairs including, sequentially, in-vehicle, egress, activity, access, and waiting time, varies for different card types as well. For UP and CP users, this duration is evenly and decreasingly distributed over time, while MP and SV users have high peaks between 9 and 11 hours, which may be regarded as a common work-related duration. Spatially, by looking at the boarding stops from the first transaction, many of the MP users live in suburban areas while UP users are concentrated in the University of Minnesota area. This also suggests that the distance between origins and downtown or the University of Minnesota area might be useful as one independent variable to describe commuting or student-oriented travel.

Trip Purpose

On-board survey data from Metro Transit in 2005 can be analyzed for travel behavior, transit use, demographic characteristics, and other aspects of travel for existing riders, and can be utilized for travel demand estimation and forecasting. On-board survey results of passengers reveal that approximate 92% of passenger trips involve home-based trips. More specifically, home-based work trips represent about 65% of all trips. This also suggests that work- and university-related purposes (combined total about 78%) are appropriate for many planning applications because transit is most successful at serving these trip purposes. For these reasons, this study mainly focuses on mandatory activities/purposes (work- and school-related trips, and other). Obviously, all the potential trip purposes (e.g., shopping, recreation; or home-based, non-home-based) should be considered in further research.

[†] Metro Pass (MP) cards are only available participating employers and refilled automatically each month.
[‡] Stored Value (SV) cards can store up to $400 and the correct fare will be deducted whenever this card is used.
[§] U-Pass (UP) cards are only available for active University of Minnesota students and valid for one semester.
[**] College Pass (CP) is offered to students enrolled at participating college pass schools, and valid for one semester.

2. DATA DESCRIPTION AND PREPROCESSING

Automated Fare Collection (AFC) data: Go-To Cards in Metro Transit

In the Metro Transit AFC system (introduced in 2004), a transaction is recorded each time a user swipes his/her fare card, typically when they board a bus. Each transaction has basic operational information, like the transaction date/time, the route number, the card type, whether the transaction is an initial boarding or a transfer, and the GPS location (but not the bus stop). Each farecard is assigned a unique serial number, which can be used as the primary way of tracking the movement of an individual passenger through the sequence of transactions during a day.

Parcel-level land use data: MetroGIS

The parcel-level land use data from 2008 includes a standard set of attributes (e.g., address, description of land use, status of tax exemption) based on each tax parcel polygon (*23*).

Google's General Transit Feed Specification

The GTFS is an open format updated by transit agencies in the US and used by Google to incorporate transit information (e.g., routes, schedules) into applications (e.g., Google Maps) (*24*). One of the required data items, the *stops.txt* file, includes the location of individual stops.

Previous work by the authors has shown that the utilization of the combination of the AFC, GTFS, and parcel-level land use data enhances the ability to better understand the following:

- Travel pattern analysis (*22*), illustrating significantly different travel behavior by passengers with different card types;
- Origin-destination estimation (*25*), identifying the locations of certain activities; and,
- A spatial and temporal linkage between transit demand and land use patterns (*26*).

Data Preprocessing

Metro Transit schedules are based on weekdays, Saturdays and Sundays. The period from 3 AM on one day to 3 AM on the next day is used as a "service" day. A set of typical weekdays, specifically Monday to Friday, November 17 through 21, 2008, is selected for this analysis. We have chosen a sample of transit users by different card types in order to better understand their travel behavior at an individual (disaggregate) level. This sample has 213 MP and 93 U-Pass (UP) cardholders who generate 2488 and 1199 transactions, respectively, during the full set of 5 weekdays. Using the information from an initial and any transfer boarding, 1056 and 500 O-D pairs, respectively, are estimated, which can be used to recognize the passengers' activity locations and durations.

While temporal information is very accurate in our data set, spatial information sometimes needs to be inferred. One alternative means of inference is to use an aggregate approach; e.g., a transit stop located within a certain or specific land use environment. For example, downtown Minneapolis stops can be grouped for identification purpose only by the downtown zone, based on Metro Transit's unique fare policy: transfers in the downtown area only cost $0.50. The University of Minnesota area stops can be also grouped to identify a school-related trip. These cases can be identified by their downtown-related activity location.

It is well-known that the effect of geographical distance to the central business district (CBD) is one of the primary determining factors of housing price. From this point of view, we also develop one attribute based on spatial separation, which is the distance from the trip origin to the

CBD (near the University of Minnesota). Finally, one interesting fact is that many express and limited-stop routes are operated on a "pay-as-you-leave" basis when traveling from downtown in the outbound direction (generally during the PM peak period). In contrast, all local routes collect passenger fares at boarding. As a result, route type may also be used to assess the trip purpose.

3. METHODOLOGY

The process of trip purpose inference using farecard data is shown in Figure 1. Assuming that transit users' behavior is observed across an ordinary week (Monday through Friday, without holidays), various attributes can be considered.

- User information (e.g., their selected farecard type);
- Temporal information (e.g., consistency of the time of travel across days, duration between transactions on a given day); and,
- Spatial information (e.g., frequency of transactions at a given location).

By developing behavioral and heuristic rules (e.g., if-then) from this information, a training set for deriving trip purposes can be built. Using this set, a decision tree-based classification technique is conducted with a test set that is used to determine the performance of the model, and then the results will be compared with separate on-board survey data to determine its consistency with known travel patterns.

Suppose that a person uses the transit system mostly for commuting with their unique farecard, and that two transactions are observed on each weekday. For one day, the first transaction is made during the early morning at a specific location on a specific route. The second transaction (last trip of the day and assumed return trip) is made during the late afternoon at another location on the same route. By looking at this trip's regularity and repeatability (e.g., observed across 5 consecutive weekdays), in conjunction with limited user information (e.g., a pass type that is only available from participating employers), we may infer that this trip is work-related, if the first transaction location is near a residential area and the second is within a commercial or office area in the CBD. For this process, we have adopted a simple decision tree structure.

Time Consistency (TC) and Space Consistency (SC)

It is well known that travel patterns vary from day to day, so that a one-day observation is not enough to capture every aspect of people's behavior over time. By looking at whether day-to-day data shows regularity or variability, one can infer changes in trip purposes. Trépanier *et al*. (*27*) estimate individuals' alighting stop finding a similar boarding which occurred in previous days.

The time consistency (TC) at an individual level can be extended in a straightforward way to analyze the travel patterns in day-to-day data. Three temporal components, the first and last/subsequent transaction times and the time between these transactions, can be investigated using typical statistics (e.g., average, variance). In addition to the TC, the spatial dimension of activities also can be analyzed using the frequency of visits to a location. The spatial consistency (SC) from origins to destinations at an individual level can be observed from land uses (e.g., trip generator/attractor) in the vicinity of the most frequently used bus stops.

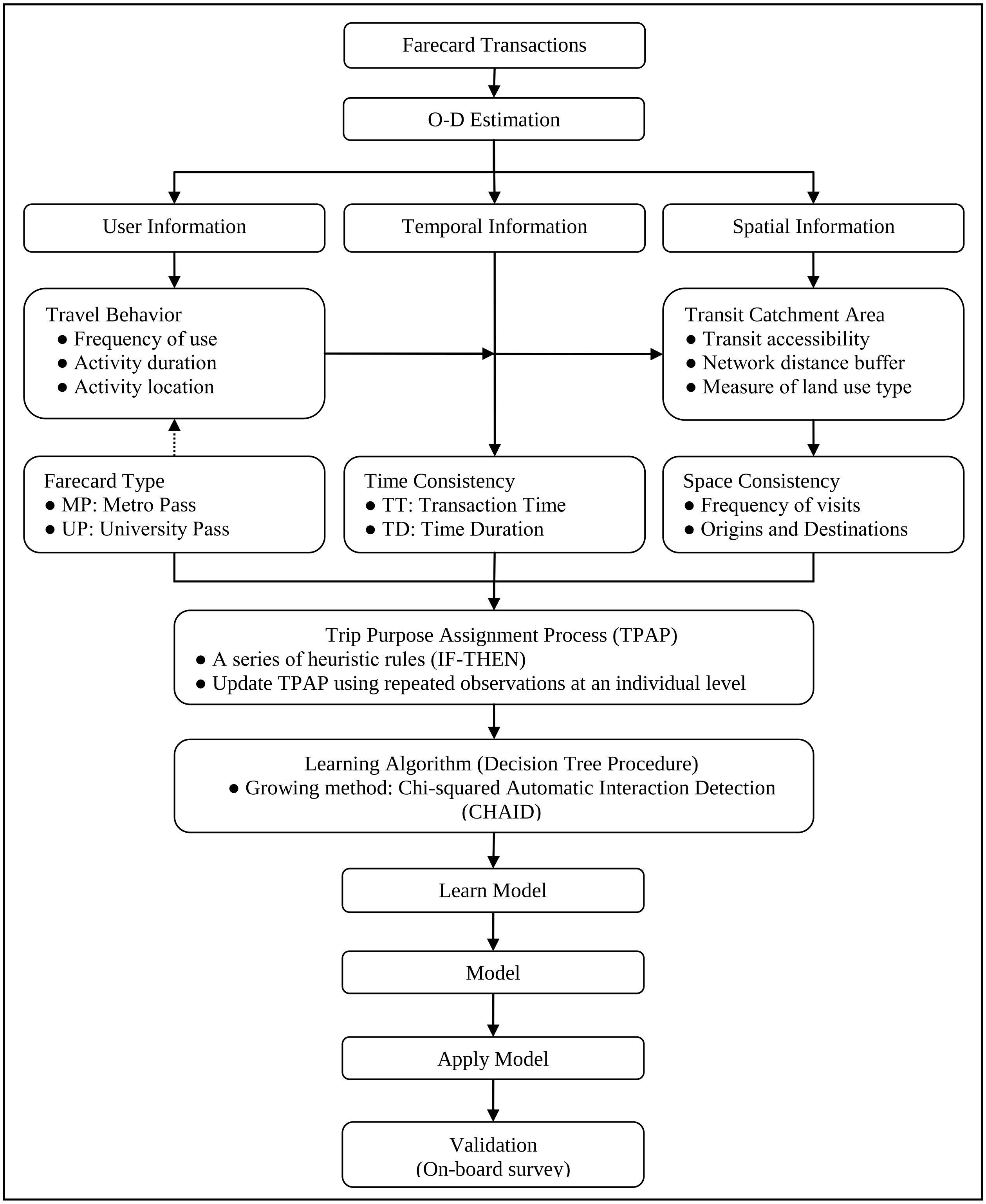


Figure 1 Overall process of trip purpose inference

Trip Purpose Assignment Process (TPAP)

In order to build a series of rules, a cluster analysis is conducted using the Statistical Package for the Social Sciences (SPSS) (*28*). This analysis first focuses on the temporal activity of the individual users. Three temporal components, the first and last/subsequent transaction times and duration between them, are selected as attributes from the O-D pairs. As an example of Agard *et al*. (*1, 4*), a split in 4 clusters with a *k-mean* is conducted. After producing 4 clusters, the number of cases in each cluster by farecard type is also investigated. Two of the clusters have easily interpretable travel behaviors with UP users, consistent with previous research (*22*), while cluster 3 clearly shows travel is made during the peak hours with a certain amount of time by the majority of MP users. The 2000 Travel Behavior Inventory (TBI) conducted by the Metropolitan Council in Minneapolis/St. Paul also shows a similar distribution of the duration of work and university/college activities. These all suggest that the temporal regularity of MP users can be incorporated into the determination of the most likely trip purposes. More specifically, in the first split for MP users, two child nodes are classified based on the time duration (i.e., more than 9 hours). The second classification criterion is whether the first transaction occurs during the AM peak period (6-9 am). The third one, the last temporal characteristic, is whether the last or subsequent transaction time is made during the PM peak period (3-6:30 pm). The last classification criteria are composed of spatial characteristics, especially destination or activity location (i.e., downtown, the University of Minnesota area, and other). Since UP users tend to cluster spatially around the university area, the best splitting begins with the activity location. The detailed TPAP including updating procedure is shown in Figure 2.

Updating TPAP based on Time and Space Consistency

To create a decision tree efficiently based on the training set, it is crucial to have the learning algorithm be as accurate as possible. As shown in Table 1(a), one of the O-D pairs for an MP user is assigned to OTHER (shaded in the table) in the TPAP, since this O-D is not classified satisfactorily by the initial rules. However, a closer look shows that this O-D is more likely to be related to work, in terms of its temporal and spatial precision. More specifically, the SC supports this O-D, although OTHER is the trip purpose assigned, because this O-D has exactly the same destination (Stop ID 40168) as in previous days. Now, the initial assignment technique (TPAP1) can be updated, providing a better trip purpose inference. In this study, 4 possible sets of rules are examined for the trip purpose assignment. Each set of rules provides information to check how well the existing rule-based assignment structure generalizes to a larger population. The following methods are employed.

1. TPAP1: the initial assignment process, discerned by looking at repeated observations of MP users across all five weekdays. Both TC and SC must be satisfied on all 5 days.
2. TPAP2: At least 4 weekday activities are assigned to be work-related by the TC and SC checks in TPAP1; or, all 5 weekday activities satisfy the *temporal* conditions in TPAP1, but location information is not available (e.g., due to “pay-as-you-leave” farecard transactions).
3. TPAP3: At least 3 weekday activities are assigned to be the work-related by the TC and SC checks in TPAP1; or, all 5 weekday activities satisfy the *temporal* conditions in TPAP1, but location information is not available.
4. TPAP4: the remaining records may also updated based on the availability of repeated observations. Specifically, the similarity of trip distances between origins and destinations across days, and the proximity of transaction times to the AM and PM peak for common locations, are considered.

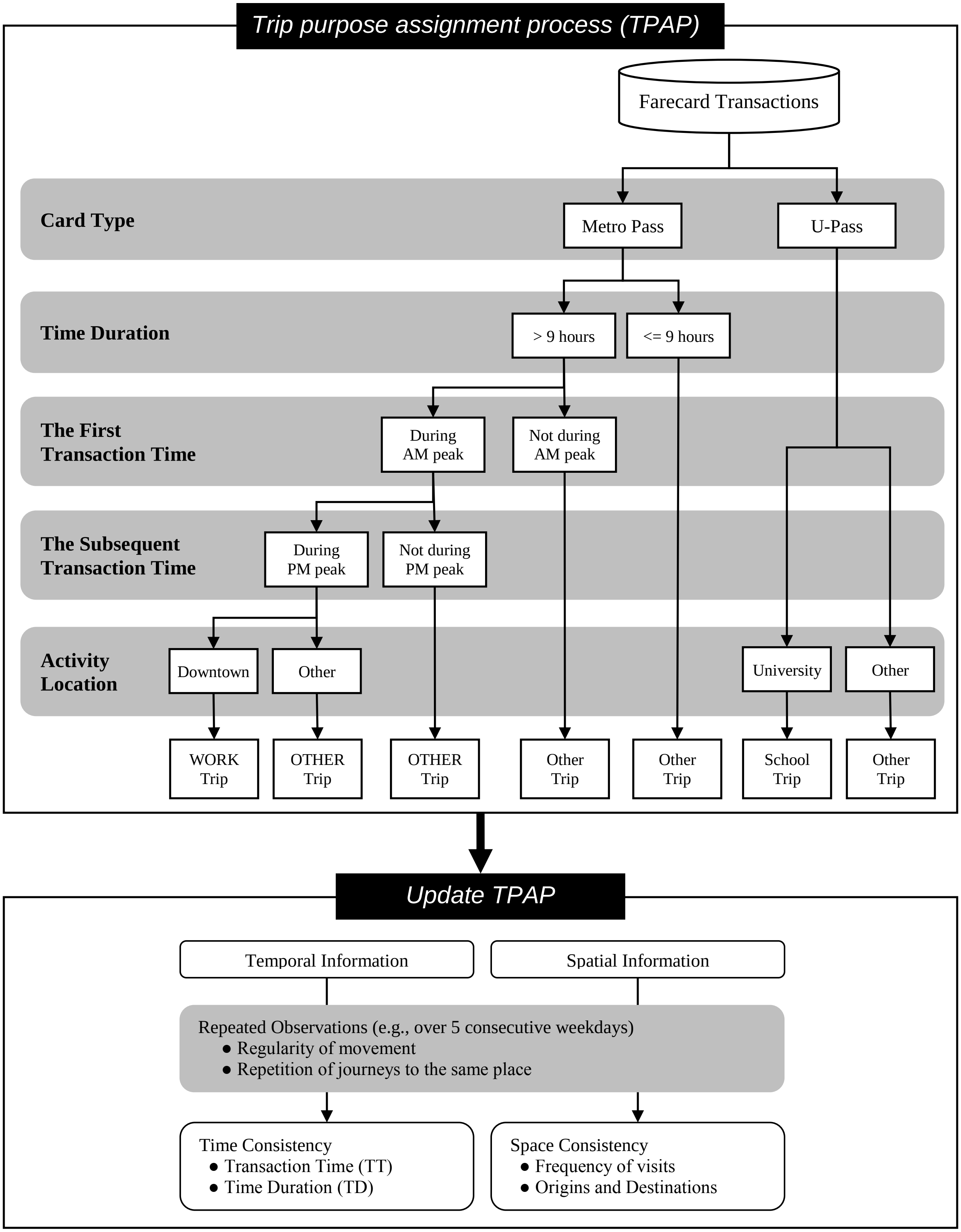


Figure 2 Rule-based decision trees for trip purpose assignment process

Table 1 Enhancement of learning algorithm using time and space consistency

(a) O-D pairs estimation and TPAP

CARD TYPE	TRANSACTION DATE	TRANSACTION TIME	DURATION	ROUTE NUMBER	USE TYPE	BOARDING STOP ID	TRIP PURPOSE ASSIGNED
MP	11/17/2008	7:32:12		6	9	1317	WORK
	11/17/2008	17:14:35	9:42:23	6	9	40168	
	11/18/2008	7:33:44		6	9	1317	WORK
	11/18/2008	16:51:24	9:17:40	6	9	40168	
	11/19/2008	7:29:21		6	9	1317	WORK
	11/19/2008	17:21:47	9:52:26	6	9	40168	
	11/20/2008	7:24:44		6	9	1317	WORK
	11/20/2008	17:12:31	9:47:47	6	9	40168	
	11/21/2008	7:35:51		6	9	1317	OTHER
	11/21/2008	16:21:36	8:45:45	6	9	40168	
UP	11/17/2008	9:32:26		6	9	16128	SCHOOL
	11/17/2008	12:27:27	2:55:01	6	9	40278	
	11/18/2008	12:50:02		6	9	16130	SCHOOL
	11/18/2008	16:13:05	3:23:03	2	9	40278	
	11/18/2008	18:38:02	2:24:57	6	1	16112	Transfer
	11/19/2008	12:10:26		6	9	16128	SCHOOL
	11/19/2008	16:35:30	4:25:04	6	9	40278	
	11/20/2008	12:50:55		6	9	16128	SCHOOL
	11/20/2008	15:49:58	2:59:03	2	9	40278	
	11/20/2008	17:36:57	1:46:59	6	1	16112	Transfer
	11/21/2008	10:08:28		6	9	16128	SCHOOL
	11/21/2008	15:29:55	5:21:27	6	9	40278	
	11/21/2008	18:17:35	9:22:33	587	9	41096	

Transactions are aggregated to O-D

Check the time and location

(b) Time and space consistencies

			MP	UP
User Information	Farecard type		MP	UP
	Transaction frequency per day		2.0	2.4
Time Consistency	First transaction	AVERAGE	7:31:10	11:30:27
		Earliest (1)	7:24:44	9:32:26
		Latest (2)	7:35:51	12:50:55
		(2) - (1)	0:11:07	3:18:29
		VARIANCE	0:00:01	0:06:05
	Last/ Subsequent transaction	AVERAGE	17:00:23	15:19:11
		Earliest (1)	16:21:36	12:27:27
		Latest (2)	17:21:47	16:35:30
		(2) - (1)	1:00:11	4:08:03
		VARIANCE	0:00:25	0:13:12
	Time Duration (including the length of time spent)	11/17/2008	9:42:23	2:55:01
		11/18/2008	9:17:40	3:23:03
		11/19/2008	9:52:26	4:25:04
		11/20/2008	9:47:47	2:59:03
		11/21/2008	8:45:45	5:21:27
		AVERAGE	9:29:12	3:48:44
		MIN	8:45:45	2:55:01
		MAX	9:52:26	5:21:27
		MAX-MIN	1:06:41	2:26:26
		VARIANCE	0:00:32	0:02:46
Space Consistency	Origin	Frequency of visits	5	5
		Catchment type	Residential	Residential
	Destination (activity location)	Frequency of visits	5	5
		Catchment	CBD	U of M

Decision Tree Classification

Classification is an important learning algorithm in a data mining problem. Due to the fact that it is simple and easy to understand, the Decision Tree procedure is the most popular model, and can be used for identification of homogeneous groups with high or low risk. In addition, it makes it easy to construct rules for making predictions about individual cases (*29*). It also provides validation tools for exploratory classification analysis. For this, a split-sample validation (70% for the training sample and 30% for the test sample) is selected. This gives 1067-1132 observations for the training data set, and 424-489 observations for the testing data set. The 70/30 split was maintained, but the number of successful observations varies based on the TPAP case.

SPSS provides several tree-growing methods, including Chi-squared Automatic Interaction Detection (CHAID). At each step, CHAID chooses the independent (predictor) variable that has the strongest interaction with the dependent variable. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable (*29*).

Table 2 Scenarios for the TPAP model

	Description	Measurement level	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Independent variable	TPAP 1, 2, 3, and 4	Nominal	X	X	X	X	X	X
Dependent variable	CardType	Nominal	X	X	X	X	X	X
	TransTimeFirst	Scale	X	X	X	X	X	X
	TransTimeNext	Scale	X	X	X	X	X	X
	DistOriginToCBD	Scale			X	X	X	X
	RouteFirstType	Nominal					X	X
	RouteNextType	Nominal					X	X
Data set		-	Training	Testing	Training	Testing	Training	Testing

Using 6 different scenarios (Table 2), statistical analyses can be performed based on the risk and the resulting classification table of each model. Risk, a measure of the tree's predictive accuracy, is composed of a risk estimate and its standard error. For categorical (nominal or ordinal) dependent variables, the risk estimate is the proportion of cases incorrectly classified after adjustment for prior probabilities and misclassification costs. In a similar way, the classification table shows the number of cases classified correctly and incorrectly for each category of the dependent variable.

4. RESULTS

Trip Purpose Assignment Process (TPAP)

The TPAP assigns a trip purpose to each daily activity by developing conditions based on travel patterns. TPAP can be updated by the repeated observations of farecard transactions in time and space, as noted for TPAP1, TPAP2, TPAP3, and TPAP4. At each level of assignment, other trip or missing values are re-examined with the previous assignment, in conjunction with time and space consistencies. For TPAP1, TPAP2, and TPAP3, more interest is focused on individuals who have been assigned at least 3 work-related activities. At the final level (TPAP4), the time and space consistencies are even more relaxed.

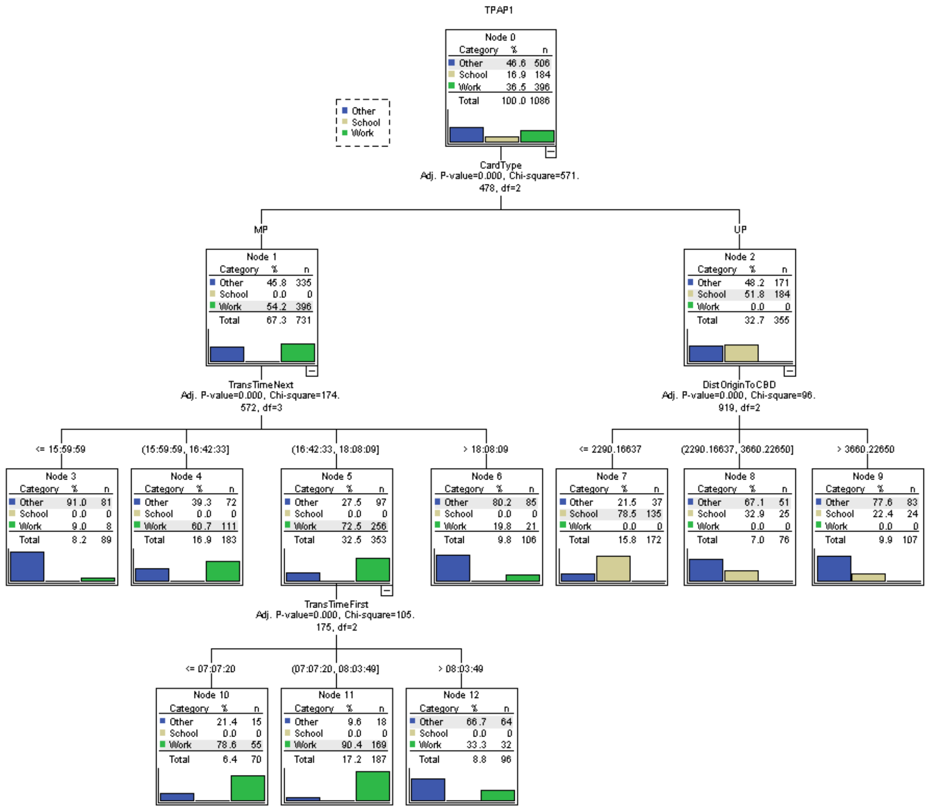
Decision Tree Classification

Results of decision tree classification can be investigated among scenarios and each level of TPAP. Overall, the best result appears at TPAP2, which means repeated observations can capture and update the trip purpose of MP users with low risk (lowest fraction of mis-classified observations). This is shown in Table 3 below. Since TPAP4 is dominated by work-related activities with more relaxed treatment of time and space consistencies, the result of TPAP4 may be biased.

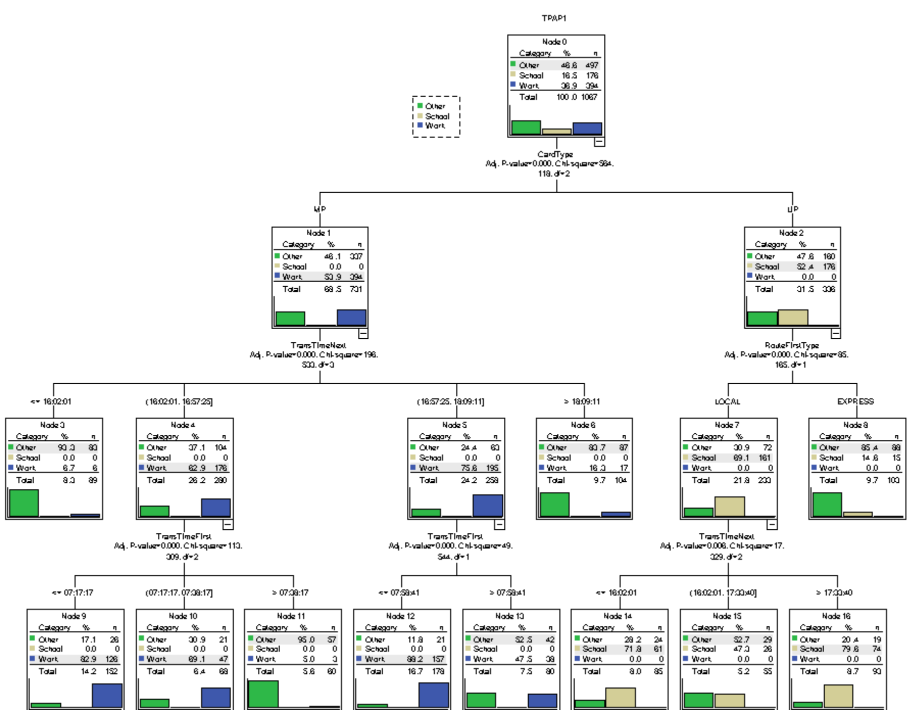
Table 3 Results of Risk

		Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6	
	Sample	Est.	Std. Error	Est.	Std. Error	Est.	Std. Error	Est.	Std. Error	Est.	Std. Error	Est.	Std. Error
TPAP1	Training	0.251	0.011	0.244	0.013	0.204	0.010	0.232	0.013	0.193	0.010	0.202	0.012
	Test	-	-	0.250	0.021	-	-	0.251	0.020	-	-	0.225	0.019
TPAP2	Training	0.254	0.011	0.263	0.013	0.208	0.010	0.208	0.012	0.197	0.010	0.188	0.012
	Test	-	-	0.271	0.021	-	-	0.198	0.019	-	-	0.234	0.019
TPAP3	Training	0.267	0.011	0.269	0.014	0.220	0.011	0.246	0.013	0.210	0.010	0.211	0.012
	Test	-	-	0.284	0.021	-	-	0.245	0.020	-	-	0.223	0.019
TPAP4	Training	0.151	0.009	0.152	0.011	0.104	0.008	0.097	0.009	0.093	0.007	0.099	0.009
	Test	-	-	0.144	0.017	-	-	0.104	0.014	-	-	0.116	0.015

The classification tree provides a graphical representation to help interpret the classification results. For most scenarios for MP users, the temporal characteristic of *TransTimeNext*, meaning the return trip time in the PM peak, is considered the primary determining factor of whether an activity is work-related. Figures 3 shows the decision tree output of TPAP1 using 4 and 6 independent variables, respectively. For the training data set, the former decision tree has 13 nodes, including 9 terminal nodes, and the latter decision tree has 17 nodes, including 11 terminal nodes. The tree maps of MP users remain the same in terms of classification and depth. This suggests that the most useful independent variables for work-related activities of MP users are temporal characteristics, especially for *TransTimeNext* (the last transaction time, typically in the PM peak period). However, the tree maps of UP users are more sensitive to the number of attributes. More specifically, it appears to have high probability of school-related trip if *DistOriginToCBD* is less than 2,290 meters, without route type information, or if *RouteFirstType* (the first route taken by the passenger) is a local route. In other words, trips of UP users who live near the CBD (near the University of Minnesota) or take a local route for the first transaction, can be recognized as a school-related trip.



(a) 4 independent variables



(b) 6 independent variables

Figure 3 Decision tree output of TPAP1

5. CONCLUSIONS

This study addresses how it is possible to successfully infer passengers' trip purpose and activity with limited resources using the automated fare collection (AFC) data. It is more focused on useful features of the farecard transaction data, such as user characteristics and spatial and temporal information, which contribute a great deal to the development of heuristic rules for inferring trip purpose.

Although the farecard transaction data do not contain important information about the passengers and their trip purposes, inferences can be made through the trip purpose assignment process. Rather than looking at travel patterns at any particular point in time or particular location, we can look at the ways in which they have changed over time at an individual level. In order to do this, it is necessary to consider repeated observations in temporal and spatial dimension for better understanding of travel patterns. The results of different scenarios have demonstrated the benefit of a more accurate learning algorithm. After building the decision tree, the proposed model can be used to predict trip purposes from other transaction data.

This study proposes a practical use of the farecard transaction data for deriving useful information about transit passenger behavior, and provides preliminary findings using the AFC data. Our approach only deals with a small set of transit users. The results, however, seem to be enough for researchers to open a new promising line of investigation. Making sense of transit users' trip purpose from the farecard transaction data is still unexplored territory. Through the proposed methodology, in conjunction with O-D pairs, we gain the ability to combine an unbiased source of transit demand data with an inferred trip purpose or activity, which is difficult to capture by traditional surveys. More classification processes, interpretation and validation effort (e.g., validation with on-board survey data) still needs to be done. We are currently exploring other learning algorithms and further evaluation of the learning effectiveness will be performed.

6. REFERENCES

1. Agard, B., Morency, C., and Trépanier, M. 2006. Mining public transport user behavior from smartcard data. In: 12th IFAC Symposium on Information Control Problems in Manufacturing – INCOM 2006, Saint-Etienne, France.
2. Morency, C., Trépanier, M, and B. Agard. 2006. Analysing the Variability of Transit Users Behaviour with Smart Card Data. The 9th International IEEE Conference on Intelligent Transportation Systems – ITSC 2006, Toronto, Canada, September 17-20.
3. Morency, C., Trépanier, M, and B. Agard. 2007. Measuring Transit Use Variability with Smart-Card Data. Transport Policy, Volume 14, Issue 3, pp.193-203.
4. Agard, B., Morency, C., and Trépanier, M. 2008. Chapter: Mining Smart Card Data from an Urban Transit Network, Encyclopedia of Data Warehouse and Mining - 2nd Edition, in John Wang (ed.), Information Science Reference.
5. Seaborn, C., Wilson, N.H., and Attanucci, J. 2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. In Transportation Research Record: Journal of the Transportation Research Board, No. 2121, Transportation Research Board of the National Academies, Washington, D.C., pp. 55-62.
6. Barry, J. J., Newhouser, R.,., Rahbee, A., and Sayeda, S. 2002. Origin and Destination Estimation in New York City with Automated Fare System Data. In Transportation

Research Record: Journal of the Transportation Research Board, No. 1817, Transportation Research Board of the National Academies, Washington, D.C., pp. 183-187.

7. Zhao, J., Rahbee, A., and Wilson, N.H. 2007. Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. Computer-Aided Civil and Infrastructure Engineering. Vol. 22, No. 5, pp. 376-387.
8. Trépanier, M., Tranchant, N., and Chapleau, R. 2007. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. Journal of Intelligent Transportation Systems: Technology, Planning and Operations, 11(1), pp.1-14.
9. Chu, K. K. A., and Chapleau, R.. 2008. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. In Transportation Research Record: Journal of the Transportation Research Board, No. 2063, Transportation Research Board of the National Academies, Washington, D.C., pp. 63–72.
10. Barry, J. J., R. Freimer, and H. Slavin. 2009. Use of Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. In Transportation Research Record: Journal of the Transportation Research Board, No. 2112, Transportation Research Board of the National Academies, Washington, D.C., pp. 53-61.
11. Reddy, A., Lu, A., Kumar, S., Bashmakov, V., and Rudenko, S. 2009. Application of Entry-Only Automated Fare Collection (AFC) System Data to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles. In Transportation Research Record: Journal of the Transportation Research Board, No. 2110, Transportation Research Board of the National Academies, Washington, D.C., pp. 128-136.
12. Wang, W., Attanucci, J., and Wilson, N.H. 2011. Study of Bus Passenger Origin-Destination and Travel Behavior Using Automated Data Collection Systems in London. Presented at 90th TRB Annual Meeting, Washington, D.C.
13. Bagchi, M., and White, P.R. 2005. The potential of public transport smart card data, Transport Policy, 12, p. 464-474.
14. Trépanier, M., Morency, C., and Blanchette, C. 2009. Enhancing household travel surveys using smart card data. 88th Annual Meeting of the Transportation Research Board, Washington, D.C.
15. Wolf, J. 2000. Using GPS data loggers to replace travel diaries in the collection of travel data, Dissertation, Georgia Institute of Technology, Atlanta, GA.
16. Stopher, P., FitzGerald, C., and Zhang, J. 2008. Search for a global positioning system device to measure person travel. Transportation Research Part C, 16, pp. 350-369.
17. Bohte, W. and Maat, K. 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C, 17, pp. 285-297.
18. Kitamura R, Chen C, Narayanan R. 1998. Traveler destination choice behavior: effects of time of day, activity duration, and home location. In Transportation Research Record: Journal of the Transportation Research Board, No. 1645. Transportation Research Board of the National Academies, Washington, D.C., pp. 76-81.
19. Hannes, E., Janssens, D., and Wets, G. 2008. Destination choice in daily activity travel: mental map’s repertoire. In Transportation Research Record: Journal of the Transportation Research Board, No. 2054. Transportation Research Board of the National Academies, Washington, D.C., pp. 20-27.
20. McNamara, L., Mascolo, C., and Carpa, L. 2008. Media sharing based on colocation prediction in urban transport. Proceedings of the 14th ACM international conference on Mobile computing and networking. San Francisco, CA.

21. Liu, L., Hou, A., Biderman, A., Ratti, C., and Chen, J. 2009. Understanding individual and collective mobility patterns from smart card records: a case study in Shenzhen. Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO.
22. Lee, S. G., and Hickman, M. D. 2011. Travel pattern analysis using smart card data of regular users. Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, D.C.
23. MetroGIS 2010. http://www.metrogis.org/. Accessed December, 2010.
24. General Transit Feed Specification (GTFS) Data Exchange 2010. http://www.gtfs-data-exchange.com/. Accessed December, 2010.
25. Lee, S. G., Khani, A., Tong, D., and Hickman, M. D. 2011. Aggregate level origin-destination estimation using smart card data and transit schedule. In Proceedings of the 2011 TRF Annual Forum, Proceedings of the 52nd Annual Forum. Long Beach, CA.
26. Lee, S. G., Hickman, M. D, and Tong, D. 2011. Development of a Temporal and Spatial Linkage between Transit Demand and Land Use Patterns. Proceedings of the World Symposium on Transport and Land Use Research, Whistler, BC.
27. Trépanier, M., Tranchant, N., and Chapleau, R. 2007. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. Journal of Intelligent Transportation Systems: Technology, Planning and Operations, 11(1), pp.1-14.
28. SPSS Inc., 2008. SPSS Statistics Base 17.0 User’s Guide.
29. SPSS Inc., 2009. PASW Decision Trees 18.