ANALYZING TOURS: APPLICATION OF A TRAVELER GROUPING BASED CLUSTER ANALYSIS

Naznin Sultana Daisy, Dalhousie University
Lei Liu, Dalhousie University
Hugh Millward, Saint Mary's University
Proceedings of the 52nd Annual Conference
Canadian Transportation Research Forum

Introduction

The need for mobility is growing with the development of urbanization and its offerings of numerous alternatives for different types of activities. This increasing demand for mobility as well as engagement in different types of activities is making the travel patterns of individuals more complex than ever (Levinson and Kumar, 1995; McGuckin etal., 2005). The level of complexity is higher when people make more than one trip in each tour compared to single-stop round trips (Hensher and Ton, 2000). Tour-based analysis can capture the complexity of trip chains and individual travel behaviour (Rossi and Shiftan, 1997). Based on the purpose of the tour, researchers have analyzed tour complexity mostly for work versus non-work purposes. In comparison to non-work tours, work tours have been analyzed in the literature largely due to their impact on peak-period travel (Wu and Ye, 2008). However, it is evident from previous studies that along with work trips, non-work trips also contribute to the peak period congestion (Purvis, 1994). The comparison of work and non-work tours and their related complexities have been analyzed to some extent (Yun etal., 2011; Ho and Mulley, 2013). Non-work tours can be divided into personal maintenance and discretionary activities. In recent times, participation in these activities has been growing and contributing to traffic congestion in both peak and off-peak periods, in both weekdays and weekends (McGuckin etal., 2005; Yun etal., 2011). Strathman and Dueker (1995) found that a lower percentage of non-work trips are chained with work trips and the percentage is as low as 10% to 20%. This indicates that more than 80% of maintenance and discretionary travel trips were made independently. However, the literature hasn't distinguished the nature of association between tour complexity and accompaniment for both maintenance and discretionary tours. The comparisons are limited to work and non-work tours, and also to the same geographic area and same unimodal choice sets (Lockwood, 2005).

Given the existing research interest in tour complexity, this paper aims to analyze this complexity by clustering travelers according to personal and socio-demographic characteristics. The primary objective of this traveler grouping is to obtain latent clusters based on the personal and socio-demographic similarities of the respondents. Samples with similar personal and socio-demographic attributes, such as age, gender, education level, personal income, and number of household earners, are grouped together.

The grouping of individuals is not new in transportation modeling and travel behaviour analysis. For instance, based on parking payment type (hourly parking payment vs. monthly parking payment), Tsamboulas (2001) divided travelers into two distinct groups to study the travel behavior. The findings of the study suggest that there exist significant differences between travelers from different groups, whereas there are significant similarities within the group. Another study on Chicago (Outwater etal., 2014), divided individuals into two groups based on the work trip and non-work trips, and the study found that there are significant differences between the travel behaviour of work and non-work trips. However, to the best of our knowledge, no studies have applied traveler grouping to tour-based analysis. Thus, the present study utilizes such traveler grouping for tours. The following section provides detailed information of methods utilized in this study.

Methodology

Tour Formation

This study utilized time-use data from the 2010 General Social Surveys (GSS) of Canada. This random stratified survey series started in 1982 and each wave includes a survey of individual household information, personal characteristics, socio-demographic information, and a 24-hour time-diary episode file for each participant. The dataset utilized in this study is based on pooling data from the 2010 wave.

This paper uses the home-based tour as the unit of analysis and identifies the accompaniment arrangement for each tour, number of tours made in a day, number of stops made in each tour, and mode choices for each tour. The home-to-home journey, for which origin and destination is home without any occurrence of intermediate home stops, is defined as a home-based tour, comprising a sequence of out-of-home trips (Shiftan, 1998). Tours starting or ending outside the participant's home were eliminated due to interpretability problems. A total of 1,067 tours of 724 individuals were obtained. After identifying the tours, the tour purpose was assigned based on the priority order by Stopher *et al*. (1996), which awards the highest priority to work, followed by education, then personal maintenance (e.g., meals, medical care, grocery shopping, snacking), and then discretionary activity (e.g., social, recreational, religious, voluntary). In addition, the travel mode for each tour was selected based on the longest in-mode travel time. Number of stops and average duration of time spent at each stop was calculated through data mining methods. Since this study attempts to explore the determinants of tour-based mode choice, the accompaniment arrangement was divided into five categories (i.e. solo, partially joint with household member/s, fully joint with household member/s, partially joint with non-household member/s, fully joint with non-household member/s). Solo tours are those where an individual completed all the trip segments alone. Partially joint are those tours where an individual shared some trip segment with someone but remaining trip segments alone. Fully joint tours are those where an individual completed all the trip segments with someone.

Cluster Analysis

Cluster analysis is an effective approach to cluster individuals. More specifically, it is a multivariate statistical method to classify the samples or indexes (Jain etal, 1999; Everitt etal., 2001). In cluster analysis, samples with similar personal and socio-demographic characteristics are clustered together. In this study, Ward's linkage-based hierarchical clustering was utilized to cluster the individuals. In Ward's method, clusters are joined so that the variance within a cluster is minimised. Ward's method states that the distance between two clusters, X and Y, is equal to the increase in the sum-of-squares if they were merged:

$$\Delta(X,Y) = \sum_{i \in X \cup Y} \| \overrightarrow{A_i} - \overrightarrow{m_{X \cup Y}} \|^2 \sum_{i \in X} \| \overrightarrow{A_i} - \overrightarrow{m_X} \|^2 \sum_{i \in Y} \| \overrightarrow{A_i} - \overrightarrow{m_Y} \|^2$$

$$= \frac{n_X n_Y}{n_X + n_Y} \| \overrightarrow{m_X} - \overrightarrow{m_Y} \|^2$$

Where $\vec{m}_j$ indicates the center of cluster j and the number of points in cluster j is represented by n_j. The merging value of combining X and Y is Δ.

The sum-of-squares starts at zero in hierarchical clustering because every point is in its own cluster and gradually grows as the cluster merging happens. Ward's method keeps this growth of sum-of-squares as small as possible. Ward's linkage method of cluster analysis on individual travelers was applied to categorize all travelers into several classes based on their personal and socio-demographic characteristics. Traveler groupings of 724 individuals were identified by utilizing five personal and socio-demographic attributes, which were gender, age, education level, personal annual income, and number of earners in household.

Tour Characteristics

Figure 1 shows the tour characteristics derived from the home-based tours formed from the GSS 2010 dataset. Figure 1(a) represents the percentage of mode shares by the main purpose of the tour. Percentages show that people drive more for work tours (23.99%) and personal maintenance tours (30.74%). On the other hand, travelers choose auto drive, auto passenger, and walk mode for discretionary activities. Only 1.41% of work tours and surprisingly only 1.22% of education tours are done by public transit; the latter figure is more understandable when we recall that the figure refers to tours, not single-stop round trips. From Figure 1(b), it is evident that weekend travel is dominated by maintenance and discretionary related tours whereas weekday travel exhibits a more balanced mix of tours related to all types of activities.

For accompaniment arrangement, it is interesting to note that work tours are mostly partially joint tours with outside household members. It is also interesting to note that non-work tours are predominantly partially or fully joint with outside household (HH) members or HH members, Figure 1(c). Only 19.68% of all tours are solo tours. As expected, Figure 1(d) shows that work tours had the longest duration (523.89 minutes) similar to other existing tour based studies (Paleti etal., 2011), whereas maintenance tours have shorter duration compared to discretionary tours.

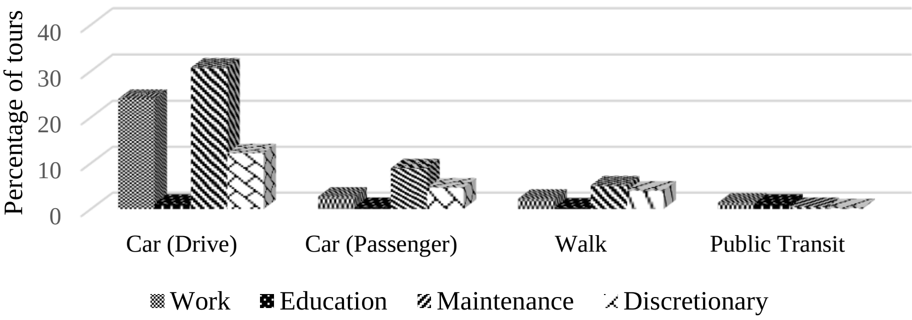

Figure 1(a): Modal share by the purpose of tour

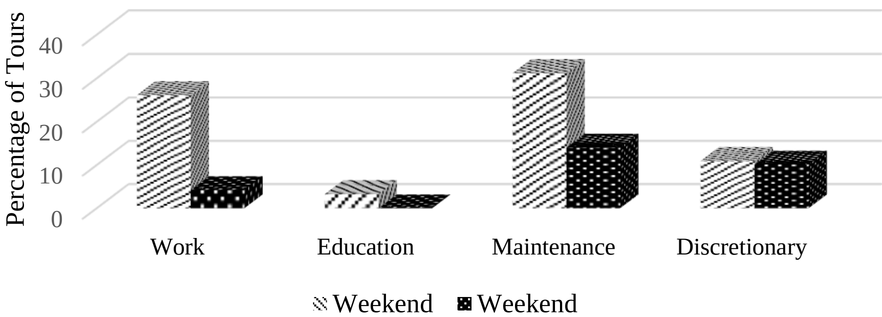

Figure 1(b): Weekday and Weekend tour purposes

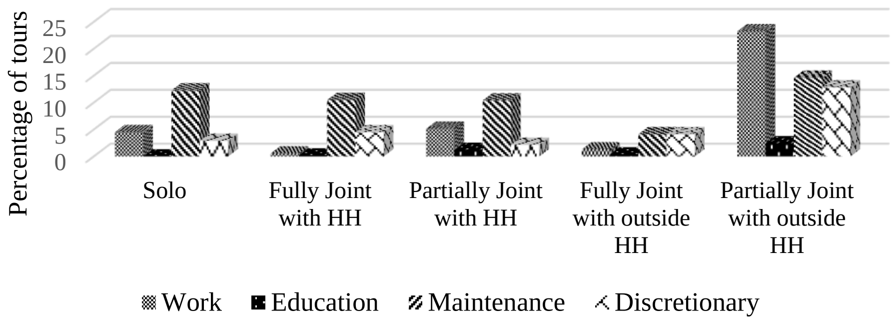

Figure 1(c): Accompaniment arrangement of tours

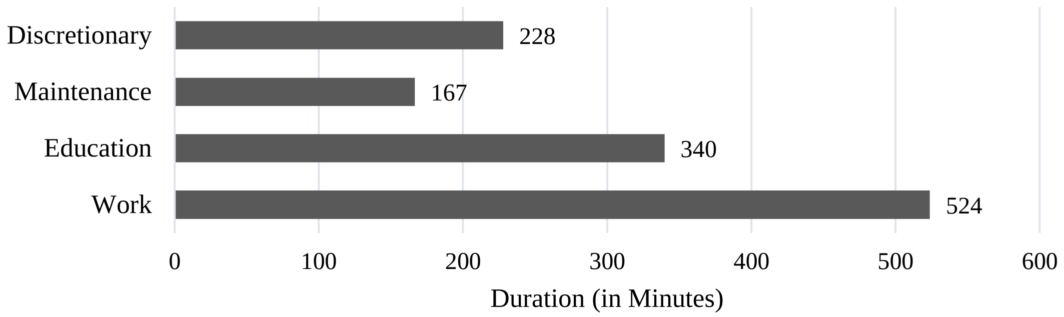

Figure 1(d): Average duration of different types of tours (in Min)

FIGURE 1 Tour Characteristics of Home-Based Tours

Discussion of Results

Traveler Grouping

The cluster analysis was conducted for the traveler grouping of the individuals based on their personal and socio-demographic attributes, including age, gender, education level, income, household size, and number of HH earners. The Ward's linkage based clustering technique was employed for clustering the individuals. Table 1 illustrates the cluster results of the 724 respondents. Four clusters were obtained and labeled as Cluster 1, 2, 3, and 4, with proportions of 18.45%, 31.47%, 23.12% and 26.96%, respectively. Personal attributes utilized in the clustering technique were represented by nominal or ordinal category indexes, which are reported in the last row of Table 2. For example, the index values of female and male were 0 and 1, respectively.

From Table 1, it was revealed that the average indexes of personal attributes of the four clusters varied visibly from each other, which indicates that the personal and socio-demographic characteristics are very different. For example, the gender indexes of cluster 1, 2, and 4 are 0.41, 0.29, and 0.33 respectively, indicating that all these clusters are female dominated. On the other hand, cluster 3 has an index value of 0.65 indicating that this cluster is male dominated. It was also found that the age index of cluster 1 is much higher than the index value of the other three clusters, indicating a higher number of aged travelers in cluster 1. The income index value for cluster 3 indicates that the travelers of this cluster have higher income level compared to the other three clusters. Cluster 2's age index indicates that a higher number of young travelers belong to this cluster. Cluster 4 has the lowest personal income index and highest education level index and highest number of earners index, indicating that this cluster comprises travelers with higher educational level and more than one earner per household.

TABLE 1 The Cluster Analysis Results of the Individuals

Cluster Label		Cluster 1	Cluster 2	Cluster 3	Cluster 4
Size		18.45	31.47	23.12	26.96
Average index values of Personal and HH attributes	Gender	0.41	0.29	0.65	0.33
	Age	5.57	2.82	4.14	3.82
	Education level	2.76	2.14	2.03	3.12
	Income Level	6.66	6.62	10.05	1.62
	Household Size	1.58	2.78	2.56	2.63
	Number of earners	0.52	0.83	0.81	1.08
Age	No of HH earners	Income Level	Education Level	Household Size	Gender
15Yr >=75Yr	None Three	$0 $100k	PhD School	One Six	F M
1 7	0 3	1 12	1 5	1 6	0 1

Table 2 provides the cluster averages of travel time, number of tours per day, average duration per stop, and average number of stops per tour. From Table 2, it is found that there are some clear differences among the four clusters in terms of tour attributes. For example, travelers from cluster three travel longer (45.53 mins), with higher number of stops (3.54 times) and longer average stop duration (127.10 mins). On the other hand, cluster one consists of travelers who had shortest average travel time (39.02 mins), shortest stop duration (98.76 mins), and lowest average number of stops (2.69 times).

TABLE 2 Cluster Descriptive Statistics of Explanatory Variables

Cluster	Avg. Travel Time (in Min)	Avg. Number of tours	Avg. duration of stops (in Min)	Avg. Number of stops
Cluster 1	39.02	1.61	98.76	2.69
Cluster 2	42.08	1.74	125.04	2.90
Cluster 3	45.53	1.70	127.10	3.54
Cluster 4	40.71	1.66	103.70	2.75

Conclusion

This paper presented cluster-based analysis for the home-based tours. The cluster-based approach can be utilized to capture differences between homogenous clusters of travelers in terms of mode choices and their influential factors. The immediate future work includes predicting modal shares for comparing the accuracy of the model. Since this study used public-use micro data, one limitation of this study is that it cannot incorporate the effects of built-environment variables into the models. The next step will include incorporating the built-environment variables into the tour-mode choice modeling using the Halifax Space-Time Activity Research (STAR) dataset. From the cluster analysis, it has been found that there exist significant differences among the four travelers' clusters in terms of tour attributes. This explains the need for incorporating the heterogeneity of travelers' grouping in the travel demand forecasting model. An improvement of this work can be achieved by modeling the household interactions for joint tours, with special attention to chauffeuring trips. This study offers a clustering approach to model the travelers in separate internally homogenous clusters, so that mode-choice modeling can reflect the latent heterogeneity among the travelers.

ACKNOWLEDGEMENT

The authors would like to acknowledge the funding support of the Nova Scotia Entrance Scholarship (NSGS) program. The authors would also like to acknowledge the Natural Sciences and Engineering Research Council (NSERC) for their contribution in supporting the research. We would also like to thank the Dalhousie Transportation and Environmental Simulation Studies (TESS) group members for their valuable suggestions. The authors would also like to thank Dr. Ahsan Habib, for his helpful suggestions in the beginning stage of this research.

REFERENCES

Everitt, B. S., S. Landau, and M. Leese. Cluster Analysis (4th edition). London, Arnold Press, 2001.

Hensher, D.A., and T. T. Ton. A Comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E*, Vol.36, No.3, 2000, pp.152-172.

Ho, C., and C. Mulley. Tour-based mode choice of joint household travel patterns on weekend and weekday. *Transportation*, Vol. 40, 2013, pp. 789–811.

Jain, A. K., M.N. Murty, and P.J. Flynn. Data clustering: A review. *ACM Computing Surveys*, Vol.31, No.3, 1999, pp.264-323.

Levinson, D., and A. Kumar. Activity travel and the allocation of time. *Journal of the American Planning Association*, Vol.61, No. 4, 1995, pp. 458-470.

Lockwood, A., Srinivasan, S., Bhat, C.: Exploratory analysis of weekend activity patterns in the San Francisco Bay Area, California. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.1926, 2005, pp.70–78.

McGuckin, N., J. Zmud, and Y. Nakamoto. Trip-chaining trends in the United States: understanding travel behavior for policy making. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1917, 2005, pp. 199-204.

Outwater, M., B. Sana, N. Ferdous, B. Doodford, and J. Lobb. characteristics of premium transit services that affect choice of mode. Transit Cooperative Research Program (TCRP) by Transportation Research Board (TRB), 2014, Washington DC.

Paleti, R., R.M. Pendyala, C. R. Bhat, and K.C. Konduri. A joint tour-based model of tour complexity, passenger accompaniment, vehicle type choice, and tour length. Presented in the 91st Annual Meeting of Transportation Research Board (TRB), Washington DC, January 22-26, 2011.

Purvis, C. L. Changes in regional travel characteristics and travel time budgets in the San Francisco bay area: 1960-1990. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 73, 1994, pp.99-109.

Rossi, T. F., and Y. Shiftan. Tour-based travel demand modeling in the U.S. *Proceedings of Transportation Systems*, Elsevier, New York, 1997, pp. 381-386.

Shiftan, Y. Practical approach to model trip chaining. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.1645, 1998, pp.17–23.

Stopher, P.R., D.T. Hartgen, and Y. Li. Smart: Simulation model for activities, resources and travel. *Transportation*, Vol.23, 1996, pp.293–312.

Strathman, J.G., and K. J. Dueker. Understanding trip chaining. Special reports on trip and vehicle attributes. U.S. Department of Transportation, 1995, pp. 1-27. URL: https://www.pdx.edu/sites/www.pdx.edu.cus/files/PR077.pdf

Tsamboulas, D. A. Parking fare thresholds: a policy tool. *Transport Policy*, Vol.8, No.2, 2001, pp.115–124.

Wu, Z., and X. Ye. Joint modeling analysis of trip-chaining behavior on round-trip commute in the context of Xiamen, China. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2076, 2008, pp. 62-69.

Yun, M., D. Van Herick, and P. Mokhtarian. Non-work travel behavior changes during a temporary freeway closure: The Fix I-5 Project in Sacramento, California. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2231, 2011, pp.1-9.