MODELING ACTIVITY SCHEDULING BEHAVIOR OF INDIVIDUALS FOR TRAVEL DEMAND MODELS

Mohammad Hesam Hafezi, Dalhousie University
Lei Liu, Dalhousie University
Hugh Millward, Saint Mary's University
Proceedings of the 52nd Annual Conference
Canadian Transportation Research Forum

Introduction

During the last decade or so in North America, activity-based travel demand modeling has become more popular than four-step and tour/trip travel demand models among urban and transport modelers. The disaggregation feature of activity-based models can help to capture different aspect of travel behavior complexity, such as flexibility in individual's activity schedules, and associations between trips and activity participation. To date, researchers and practitioners have employed different approaches for development of activity-based models. A number of important specifications which affect the average level and variability of such models are prediction accuracy, reproducibility, computational time, large scale operation capability, and performance at the household level.

The current paper presents cutting-edge methods and progress in developing a new comprehensive pattern recognition modeling framework for use in activity-based travel demand modeling. The framework leverages activity data to derive clusters of homogeneous daily activity patterns, in order to infer the scheduling behavior of individuals. In this paper, numerous new machine learning techniques are employed in the pattern recognition modeling framework. The pattern recognition model is applied to data from the large Halifax STAR household travel diary survey. The proposed modeling framework has much higher reproducibility and shorter computational time compared to other alternative modeling frameworks. Furthermore, the proposed modeling framework can be applied to any applications that contain a group of linked sequences, such as day-to-day variations in transit usage.

Data

This study uses time-diary and GPS geo-coordinate data from the Space-Time Activity Research (STAR) survey data. The STAR survey represents the world's largest deployment of global positioning system (GPS) technology for a household activity survey. A brief explanation follows, and full descriptions of the survey design and the socio-demographic characteristics of respondents can be found in TURP, 2008; Millward and Spinney, 2011; Millward et al. 2013; Spinney and Millward, 2011.

The Halifax STAR project collected fully geo-referenced 2-day (i.e. 48-h) time diary data from primary respondents within 1,971 randomly selected households in Halifax Regional Municipality (HRM), between April 2007 and May 2008, with an overall response rate of 21%. Primary respondents were randomly selected within each household, and were aged 15 years or older. They carried a GPS data logger (Hewlett Packard iPAQ hw6955) for a 48-hours reporting period, recorded a daily ''activity log'' during that period, and finished a computer-assisted telephone interview (CATI) time-diary survey the day after the two-day reporting period had ended. The respondents' descriptions of their out-of-home activities were prompted and validated by the GPS data. The original 188 activity types were aggregated into 9 activity categories as follow: (1) home chores, (2) home leisure, (3) night sleep, (4) workplace, (5) shopping & services, (6) school/college, (7) organizational/hobbies, (8) entertainment, and, (9) sports.

Methods

In this paper, we propose an approach that scales up existing activity generation and activity scheduling methods using a new pattern recognition framework. Based on the fact that homogeneous clusters may increase the accuracy of inferring the scheduling behavior of individuals in activity-based modeling, our approach leverages activity data to derive clusters of homogeneous daily activity patterns. Our modeling framework is straightforward and minimizes exogenous errors. In the following section, we present only a concise overview of the method, and interested readers are referred to Hafezi et al. 2017 for more explanation.

First, we identify individuals with homogeneous activity patterns and group them into clusters. For this purpose, we initialize both cluster number and cluster centroids using a dynamic subtractive clustering algorithm (equations 1 and 2). Next, the Fuzzy C-Means (FCM) clustering technique is employed as a fully unsupervised machine learning algorithm (equation 3). The outcomes of this phase include several homogeneous clusters containing critical information such as activity type, start time, duration probability distribution, end time, and sequential arrangement of activities. The multiple sequence alignment method is then executed to identify the sets of representative activity patterns (equation 4). The outcomes of this phase can be used as a target for optimization of individual daily activity patterns. Lastly, numerous decision trees are grown up to discover the inter-dependencies among different attributes such as socio-demographic variables, start time, duration, sequences, travel mode and location in each identified cluster. The authors of this paper are currently undertaking a development of the MLN algorithm to harmonize these decision trees.

$$K_n = \sum_{m=1}^{z} e^{-\frac{4}{u_r^2}\|c_n - c_m\|^2} \tag{1}$$

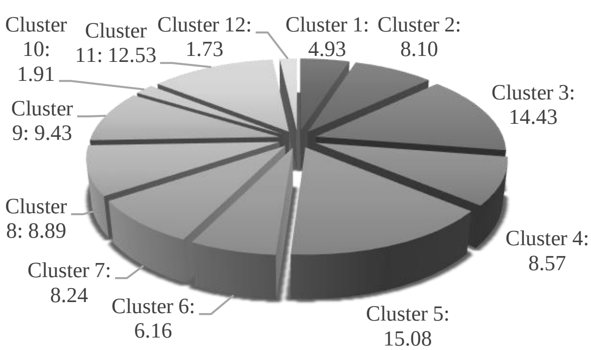
$$\beta_{nh} = e^{-\frac{4}{u_w^2}\|c_n - c_h\|^2} \tag{2}$$

$$f_{nh} = \frac{\left(\frac{1}{\left\|c_n - f_f\right\|}\right)^{\frac{1}{(\tau-1)}}}{\sum_{h=1}^{n_f}\left(\frac{1}{\left\|c_n - f_f\right\|}\right)^{\frac{1}{(\zeta-1)}}} \tag{3}$$

$$K_{n,m} = max\begin{cases} K_{n-1,m-1} + 1 & if\ e_n = e_m \\ K_{n-1,m-1} - \rho & if\ e_n \neq e_m \\ K_{n-1,m} - \sigma & \\ K_{n,m-1} - \sigma & \end{cases} \tag{4}$$

Results

The proposed modeling framework was applied to 2,778 activity days drawn from the large Halifax STAR survey data. Individuals with homogeneous activity patterns were identified through the machine learning process and grouped into twelve clusters, as shown in Figure 1. In general, the algorithm identified six clusters with heterogeneous activity patterns for worker groups. For the non-worker groups, the algorithm recognized four clusters with dissimilar activity patterns. Lastly, two single clusters were identified for students and for people who mostly spent their time at home.



Cluster #1: extended work-day workers
Cluster #2: non-worker
Cluster #3: 8-4 workers
Cluster #4: non-worker
Cluster #5: stay-at-homes
Cluster #6: shorter work-day workers
Cluster #7: 7-3 workers
Cluster #8: non-worker, morning shopping
Cluster #9: non-worker, afternoon shopping
Cluster #10: evening workers
Cluster #11: 9-5 workers
Cluster #12: students

Figure 1. Twelve Identified Clusters (with percentages of respondents).

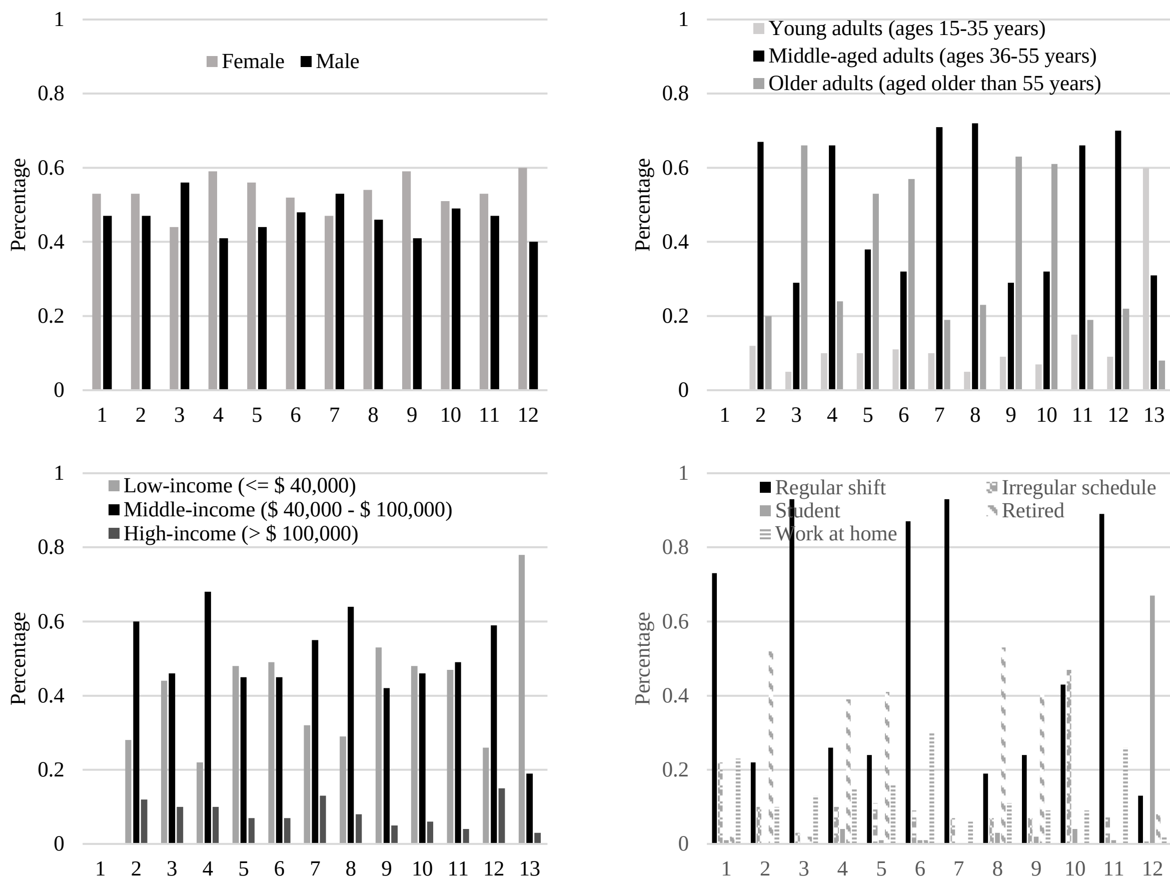


Figure 2. Analysis of Clustered Data

Figure 2 presents an analysis of clustered data. The major portion of workers with regular shifts (8-4 a.m. and 7-3 a.m.) consisted of older and middle-aged males with middle-income level. Non-regular shift workers mostly consisted of older and middle-aged females with middle or low income level. The majority of people who mostly stayed at home were old-aged females with low income level. Lastly, the students group consisted of young adults aged between 15 and 35 years old belonged to the low income level.

Conclusion

This paper contributes to the literature on activity-based models. We deviate from previous studies by introducing a precise, reproducible and fast computational modeling framework for leveraging activity data to derive clusters of homogeneous daily activity patterns and infer the scheduling behavior of individuals. Using the data from the large Halifax STAR travel diary survey, we modeled the 2-day in-home and out-of-home time-use activity patterns of individuals. The algorithm rapidly converged and resulted in twelve clusters of individuals, each with homogeneous activity patterns. Each cluster contains a variety of temporal and spatial information. For instance, we are able to achieve a unique time interval for start time and duration for every activity. Locations and travel distances also can be analyzed, and the association between socio-demographic attributes and activity patterns can be more precisely explored. We are currently working on the development of the MLN algorithm to establish a hybrid framework to predict activity patterns of individuals. The results of this paper are expected to be implemented within the activity-based travel demand model for Halifax, Nova Scotia.

Acknowledgments

This research was supported, in part, by grants from Dalhousie University. Their support is gratefully acknowledged. The authors would also like to thank 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. Data for this research were provided by the Halifax STAR Project, supported through the Atlantic Innovation Fund from the Atlantic Canada Opportunities Agency, Project No.181930.

References

Hafezi, M. H., Liu. L. and, H. Millward (2017). Identification of representative time-use activity patterns using fuzzy c-means clustering. *Transportation Research Record: Journal of the Transportation Research Board*. In press.

Millward, H. and J. E. Spinney. (2011). Time use, travel behavior, and the rural–urban continuum: results from the Halifax STAR project. *Journal of Transport Geography*. 19(1), pp. 51-58.

Millward, H., J. E. Spinney, and D. Scott. (2013). Active-transport walking behavior: destinations, durations, distances. *Journal of Transport Geography*. 28(1), pp. 101-110.

Spinney, J. E. and H. Millward. (2011). Weather impacts on leisure activities in Halifax, Nova Scotia. *International Journal of Biometeorology*. 55(2), pp. 133-145.

TURP. (2008). *TURP (Time Use Research Program)*. Retrieved from Halifax regional space time activity research (STAR) survey: A GPS-assisted household time-use survey, survey methods. Halifax: Saint Mary’s University.