Sequences, Daily Schedule of Activities, and Fragmentation

Konstadinos G. Goulias, Adam W. Davis, Elizabeth C. McBride
University of California Santa Barbara

Submitted to the 8th Symposium of the European Association for Research in Transportation (hEART 2019), Budapest, September 4-6, 2019.

Fragmentation of activities and travel is defined here as the multiple sequencing of many relatively short and long activities and trips that happen in a personal daily schedule. Information and (tele)Communication Technologies (ICT) that release spatial and temporal constraints combined with disruptive transportation services (e.g., Uber/Lyft) and automation (e.g., self-driving cars) have the potential of added flexibility to reach places and therefore increased fragmentation. In this analysis we employ a new framework and methods of sequence analysis to study daily patterns of activity and travel among different places. We extract summary indicators of daily schedule complexity and variety to then identify significant determinants of fragmentation.

The data used in this analysis comes from the 2012 California Household Travel Survey that collected information at the household and person level, vehicle ownership information, and a one-day place-based travel diary for every respondent. As a test area for this sequence analysis method, only respondents residing in San Luis Obispo and Santa Barbara counties (in the Central Coast region of California) are used. In total for this analysis, we use data from 2,663 persons. The Santa Barbara and San Luis Obispo regions comprise an area with a variety of land use types – ranging from high density/urban to low density/rural. The diary covers from 03:00 on the survey day until 02:59 on the following day. Respondents report every place they go, their travel mode, the top three activities performed at each place, and several other characteristics of the places visited and their travel. Land use surrounding each residence is attached to each household with indicators that are based on a detailed establishments inventory following techniques reported in Chen et al. (2011) and McBride et al. (2017).

The analysis of minute by minute sequences of activity and travel uses as anchors Home, Workplace, School, and Other. A sequence for a person in a day are the 1440 minutes that are labeled as H, W, S, O, and T for travel. In this way we have 5 possible states in 1440 bins. Similarities and differences among these sequences are measured by optimal matching (Abbott & Tsay, 2000) and representative patterns are extracted using hierarchical clustering (Kaufman & Rousseeuw, 2009). Figure 1 shows these six patterns. They show the usual work day and school day. They also show days during which people stay for long periods are home and days during which they run errands (shopping, taking care of personal business, attending civic meetings and so forth). There are, however, two additional patterns that are not found in typical activity-base forecasting models and these are day-patterns for people that travel long distance away from home and days when they come back from long travels.

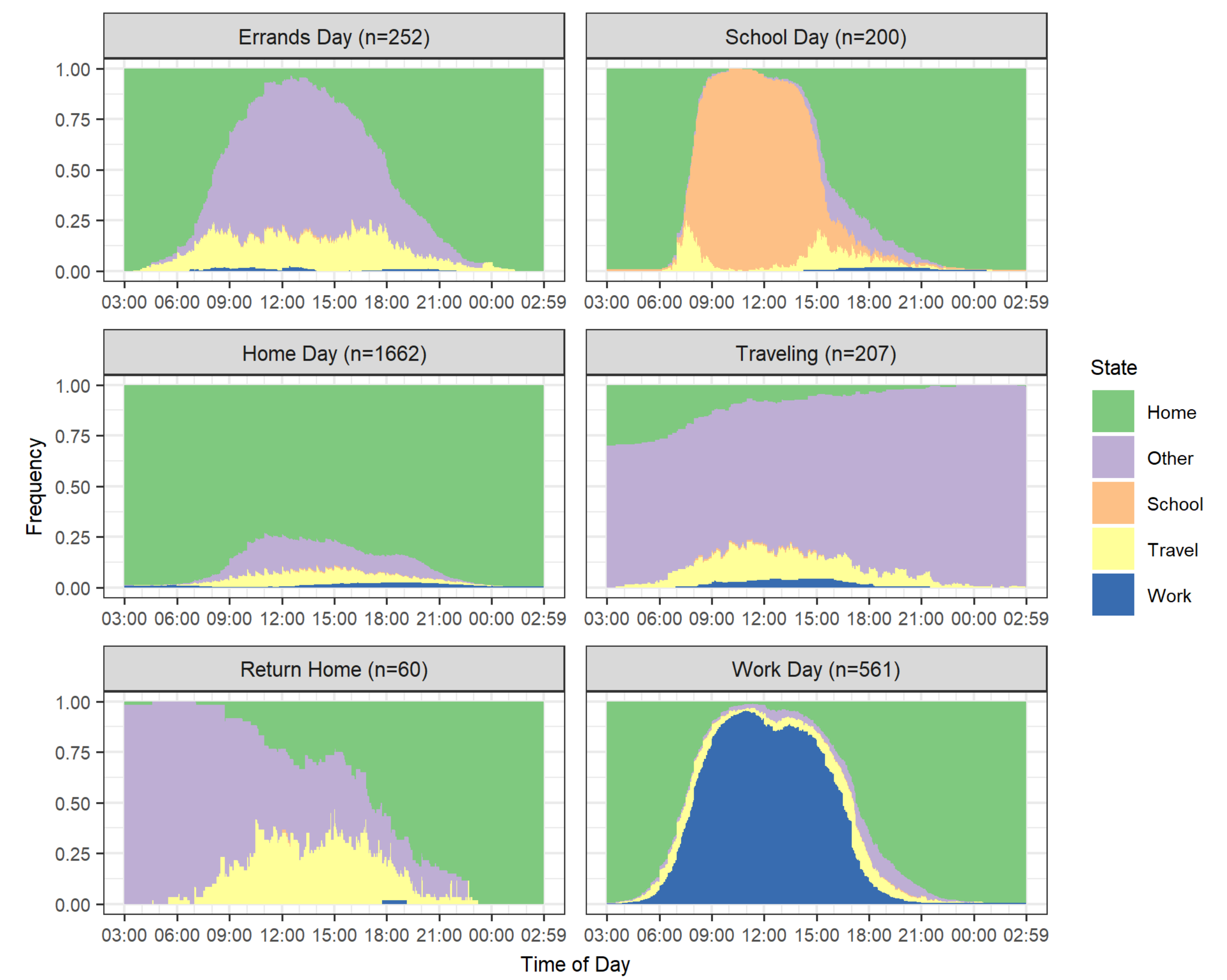

Figure 1 Six Representative Patterns in the Observed Sequences

Analysis of the activity-travel fragmentation uses Entropy and Turbulence and these summary indicators show substantial variability in schedule fragmentation. Entropy is a measurement of "prediction uncertainty".

$$h(x) = h(\pi_i, \ \dots \pi_s) = -\sum_{i=1}^{s} \pi_i \log(\pi_i) \qquad \text{(Eq. 1)}$$

Where x is the sequence, s is number of potential states and π_i is proportion of occurrences of the *i*th state in the considered sequence. The proportion of minutes allocated to each state over the course of the entire day and the number of distinct states that a person visits, drive the value of Entropy. For this measure, the number of state changes and the contiguity (or lack thereof) of states do not matter. It is simply using the proportion of total time spent in each state, regardless of the number of different episodes that time is spread over.

If a person has no state changes during the entire day, for instance if they spend all day at home, their Entropy would be zero. In contrast, someone who moves around a lot will have "high" Entropy. The range of entropy values depends on the number of distinct states. Sequences with more unique states have higher potential maximum Entropy values, and Entropy is at its highest when people spend equal amounts of time in each state. In our study with five distinct states (Home, Travel, Work, School, and Other), the maximum Entropy is 1.61.

The second measure – Turbulence – is a bit more complicated than Entropy in terms of what it uses for its calculations.

$$T(x) = \log_2\left(\phi(x)\frac{s_{t,max}^2(x)+1}{s_t^2(x)+1}\right) \quad \text{(Eq.2)}$$

- x represents the sequence of activities and travel episodes in one person's diary;
- $\phi(x)$ is the number of distinct subsequences in sequence x;
- t_i is duration in each distinct state and is used to compute the mean consecutive time and variance below (i=1,…, number of distinct episodes);
- s_t^2 is variance of the state-duration for the x sequence;
- $s_{t,\,max}^2$ is the maximum value that the variance can take given the total duration of the sequence x

$$s_t, \max = (n-1)(1-\bar{t})^2 \quad \text{(Eq. 3)}$$

- n is length of distinct state sequence
- $\bar{t}$ is mean consecutive time spent in the distinct states

Turbulence uses the number of distinct subsequences in a given sequence; and the number of consecutive time points spent in a given state. Consider a person with a daily sequence H-T-W-T-H meaning the person was at home (H) in the morning, traveled (T) to work (W), and after work traveled (T) back home (H). This sequence would contain the following subsequences: an empty sequence; the full sequence itself; subsequences of the type T-W-T-H, W-T-H, and T-W-T; discontinuous subsequences like T-T-H (which skips the work activity); and single activities H, T, and W. Enumerating all these subsequences yields ($\phi(x) = 27$) possible combinations that respect the precedence of activities in the H-T-W-T-H sequence. For a given sequence of activities (x), the measure of Turbulence is a measure of variability in terms of distinct activities, the order of these activities, and the variance of the durations of these activities in a day. All this makes Turbulence a measure of schedule complexity. Figure 2 shows the two summary indicators.

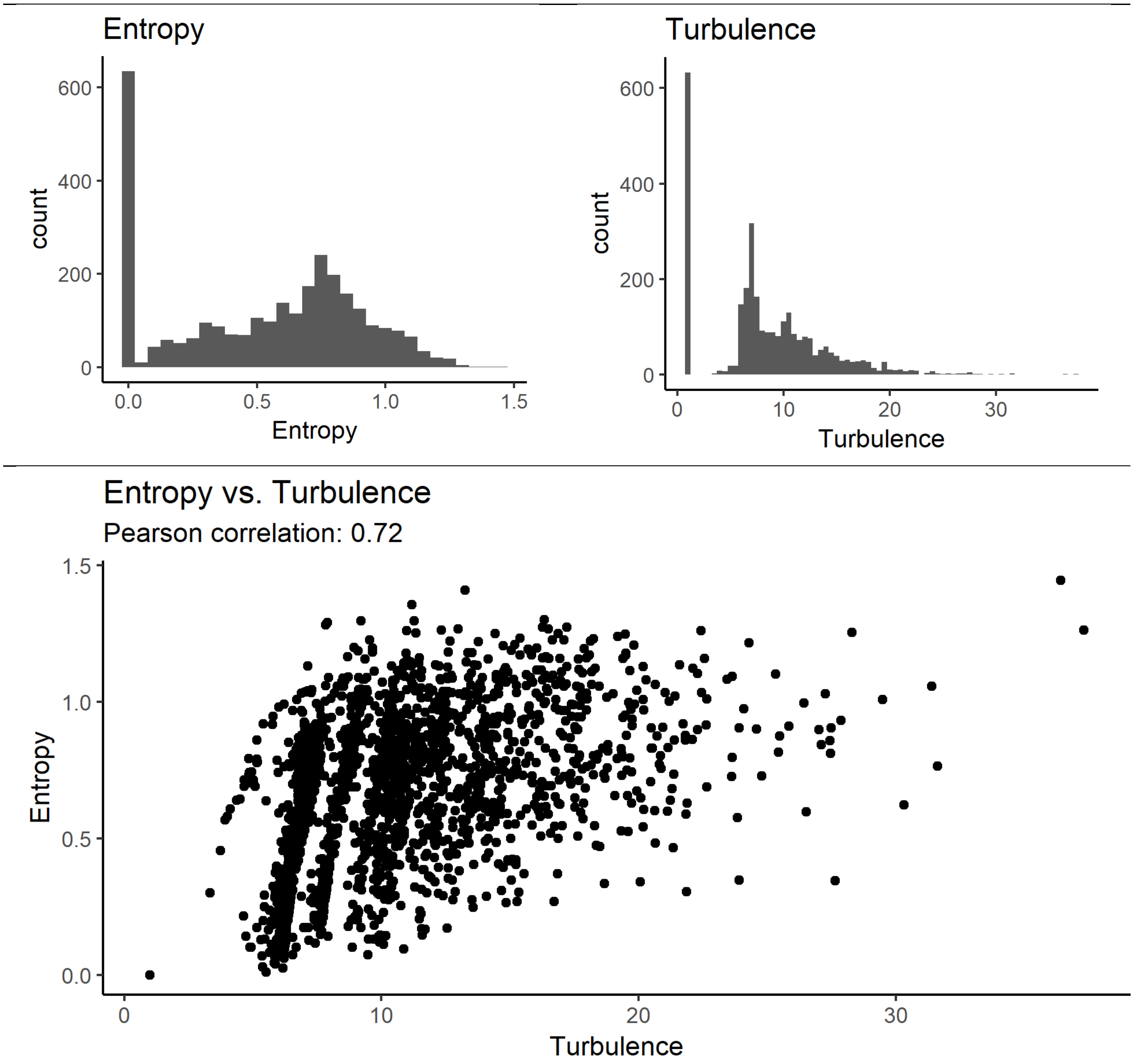

FIGURE 2 Entropy Versus Turbulence

The top portion of Figure 2 shows we have a large concentration of observations at value zero for entropy and the value of 1 for Turbulence. These are indicative of the people that stayed home all day. We applied Tobit regression to both indicators and used as explanatory variables person, household, and place of residence characteristics. We found that people aged 25 to 65 have the most fragmented schedules, especially when they have children over age 4 in the household. Escorting and joint participation in activities with children is a clear motivation for this. We also find significant inhibition of fragmentation when incomes are below poverty. People that live in urban and suburban environments tend to have more fragmented schedules. Key to all analyses here is the day of the week. Sunday is the day with the least fragmented schedules and each day of the week has a different composition of activities, travel, and durations. At the conference we will provide the results of more in-depth analysis of daily patterns and explain in more detail the use of locations that are labeled as other in this summary.

References

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological methods & research*, *29*(1), 3-33.

Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S. Y., Lei, T., ... & Hu, H. H. (2011). Development of indicators of opportunity-based accessibility. *Transportation Research Record*, *2255*(1), 58-68.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

McBride, E. C., A. W. Davis, J. H. Lee, and K. G. Goulias. Incorporating Land Use into Methods of Synthetic Population Generation and of Transfer of Behavioral Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2017. https://doi.org/10.3141/2668-02.