Authors Antonin Danalet, Michel Bierlaire

Title A new approach to activity choice modeling for passively collected data

Track General Papers

Director Mark Wardman

Abstract

1 Introduction

Travel demand models are necessary to forecast the effects of changes in the transportation system, such as new infrastructures or new policies. Since the 1970s, the activity-based approach assumes that travel demand is derived from the demand for activities. We extend this approach to pedestrian infrastructure and to passively collected data (e.g., WiFi traces).

Our goal is to develop a model of choice of activity sequences adapted for sensor data. Compared to the existing literature, challenges include the measurement un- certainty of sensor data, the absence of a structure of tours based on a daily start and end point (i.e., home), and the absence of a clear primary activity (i.e., work).

2 Methodology

We presented last year at hEART an algorithm to generate activity-episode se- quences from WiFi traces (Danalet et al., 2013). We present here a modeling frame- work for this output.

By discretizing time, we define an “activity network” (Figure 1). It contains all possible activity patterns and represents the universal choice set. Each node represents the performance by an individual of an activity type k for a unit of time τ. Activity paths are a representation of an activity pattern in the activity network. Depending on the imprecision of the measure, one activity pattern could be represented by several different activity paths (Figure 1).

[Figure 1: Illustration: an activity network for a train station, with ambiguity in start and end times.]

2.1 Choice set generation

In path choice context, the universal choice set is big and not tractable. We propose to use recent developments in the sampling of alternatives to generate a choice set. Flötteröd and Bierlaire (2013) use a Metropolis-Hastings algorithm to sample paths according to an arbitrary distribution, avoiding complete enumeration. This approach offers the opportunity to define the sampling probability for the whole path and not only by link. We define a weight function based on length and attractivity measure: $\delta(\Gamma) = -\mu_v \cdot \sum_{v\in\Gamma} \delta_v(v) - \delta_\Gamma(\Gamma)$, where $\delta(\Gamma)$ is the target weight for path Γ, μ_v is the weight of the sum in the equation, $\delta(v)$ is the weight of node v, i.e., the attractivity of node v as defined in Danalet et al. (2013), and $\delta(\Gamma)$ is the non- link-additive weight of path Γ. It expresses a penalty for path lengths different from the observed ones. Figure 2 shows an example of choice set generation based on the example of Figure 1.

[Figure 2: 20 generated paths with constraint on the link to e and constraint about the length with $\mu_v = 1$. Attractivity is 5 for waiting for the train, 1 for having a coffee and 4 for buying a ticket.]

2.2 Activity path choice modeling framework for WiFi traces

The data from Danalet et al. (2013) are not directly related to activity types and durations. Each individual measurements $\hat{m}_{1:J}$ may correspond to different activity paths $A_{1:T}$ in the activity network. In order to associate the measurements to the activity network in a probabilistic manner, we propose the following choice probability (Bierlaire and Frejinger, 2008; Chen, 2013):

[Equation 1]
where $P(\hat{m}_{1:J}|A_{1:T})$ is the measurement likelihood that measurement $\hat{m}_{1:J}$ has been generated by an individual performing activity path $A_{1:T}$ (see Danalet et al. (2013) for a complete definition) and $P(A_{1:T}|U;\beta)$ is the activity choice model for choice set U.
To be operationalized, the activity choice model must correct for the sampling of alternatives and for the correlation structure of a route choice.

2.2.1 Sampling of alternatives

We assume the choice set to be the universal choice set. The sampling of alternatives requires the deterministic part of the utility to be corrected in order to estimate unbiased parameters. According to Frejinger et al. (2009), a sampling correction term must be added:
[Equation 2]
where $k_{\Gamma n}$ is the number of times activity path Γ is drawn in C_n and q(j) is the sampling probability.
The sampling probability q(Γ) is available using the unnormalized target weights δ(Γ) and does not require full enumeration for normalization (it cancels out in the logit formulation).

2.2.2 Activity path size

Adapting the traditional path size logit model (Ben-Akiva and Bierlaire, 1999), we correct the utility for the correlation related to overlapping segments of paths. Assuming that all K activity types are available for each time interval and that the distance length can be replaced by time length, we define the activity path size APS_Γ for path Γ as:
[Equations 3 and 4]

2.2.3 Path utility

The path utility is the sum of the utilities of individual nodes $A_{k,\tau}$:
[Equation 5]
The utility $V(A_{k,\tau})$ of a node $A_{k,\tau}$ represents the individual utility from allocating time to a certain activity type. The utility includes both the satiation effect and the time of day preference. Satiation effect represents the diminishing marginal utility with duration, $\eta_k \ln(t_k)$, where t_k the duration of activity type k and η_k a satiation parameter for activity type k (Ettema et al., 2007). The time-of-day utility depends on both the activity type k and the time interval τ. It can be generally expressed as $\beta_{k,\tau} I_{k,\tau}$, where $I_{k,\tau}$ is a dummy variable and $\beta_{k,\tau}$ the corresponding parameter. In practice, some β's might be statistically equal. The path utility is:
[Equation 6]
Finally, the deterministic part of the utility correcting for both the sampling of alternatives and the correlation due to the physical overlapping of paths in C_{in} is:
[Equation 7]

3 Case study and contribution

We will present an estimation of the model on the campus passive data and on airport interview data. Figure 3 shows a typical choice by one employee on campus. The measurements are most of the time unambiguous, but on the last activity episode, there are two possible activity types: lab or office.
[Figure 3: Activity pattern for one employee's device on May 23, 2012. The x-axis represents the time of the day. The colors/patterns represent the different categories of the points of interest. The y-axis is the probability to be the correct point of interest.]

Our methodology contributes to the understanding of activity choice in pedestrian facilities, such as transportation hubs. It adapts the traditional approaches by not assuming a tour-based structure. It also assumes no priority in the different activity types in the modeling framework.
Moreover, this approach is adapted to passively collected data (e.g., WiFi traces). It takes into account the precision of the measurement. Passive data are usually cheap and are already available in a lot of pedestrian facilities.