# A Pedestrian Destination-Chain Choice Model from Bayesian Estimation of Pedestrian Activities using Sensors Data

**Antonin Danalet***
**Bilal Farooq**
**Michel Bierlaire**
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne, Switzerland

## 1 Introduction

In recent years, urban growth and its pressure on urban infrastructure created interest in pedestrian behavior. Pedestrian modeling is emerging as a tool for designing new infrastructures and optimizing the use of current ones. Data collections and development of modeling approaches are vital in estimating the demand for these infrastructures.

Given sensor traces, we are interested in developing a dynamic model that can predict the destination chain of an individual in public facilities.

Cost and privacy often avoid from installing high precision sensors such as cameras covering an entire pedestrian infrastructure (e.g., airport or a railway station). We are interested in sensors with full coverage of the facility, in order to estimate the overall demand. Such data exist, but with low precision (e.g., traces from WiFi infrastructures).

As a first step, we developed a methodology to collect activity-episodes sequences from scarce data, directly modeling the imprecision in the measure. It generates several candidate lists of activity-episodes sequences associated with a corresponding likelihood. Then, we use this output as observed choice in a dynamic destination-chain choice model in order to forecast the total demand in the facility. We make the assumption that pedestrian travel demand is derived from the demand for activities. The main variable we study is not a trip, or even a tour, but a sequence of activity-episodes ("pattern"). Our approach is integrated system of choice models, based on the concept of utility maximization and inspired by Bowman (1998). Travel outcomes are part of an activity scheduling decision.

*antonin.danalet@epfl.ch

## 2 Bayesian Method for Measuring Pedestrian Activities

Our methodology to collect activity-episodes sequences from digital traces from communication networks merges measured localization and pedestrian map information to compute the likelihood that a given activity-episodes sequence has generated the observed traces. Results are presented in the case of WiFi traces in Table 1, but could be used with other data collection techniques, such as traces from cellular networks or Bluetooth tracking.

| **Model** | | | **Truth** | | | $\Delta x$ |
|---|---|---|---|---|---|---|
| *Time spent* | *Floor* | *Location* | *Time spent* | *Floor* | *Location* | *(in m.)* |
| $U$(8:40,8:40) - $U$(10:38,10:38) | 3 | Office | 8:32-10:30 | 1 | Classroom | 61 |
| $U$(10:45,10:45) - $U$(11:51,11.51) | 3 | Office | Until 11:47 | 3 | Author's office | 7 |
| $U$(12:04,12:04) - $U$(12:47,12.53) | 2 | Classroom | From 11:55 | 1 | Restaurant | 0 |
| $U$(13.02,13.03) - $U$(13:03,13.44) | 3 | Office | Around 13:00 | 3 | Author's office | 7 |
| $U$(13.06,13.47) - $U$(13:53,14:02) | 2 | Restaurant | Around 14:00 | 2 | Cafeteria | 0 |
| $U$(13.55,14.04) - $U$(19:40,19.44) | 3 | Office | Until around 19:45 | 3 | Author's office | 9 |

Table 1: Comparison between the most likely output of the model and the truth as reported by one author. $\Delta x$ represents the distance (in meters) in the same horizontal plane.

The input of this probabilistic method consists in time stamps and localization data. We define a measurement as $\hat{s} = (\hat{x}, \hat{t})$, where $\hat{x}$ is the position of the measurement and $\hat{t}$ the timestamp of the signal. For a given individual, we assume a chronologically ordered sequence $(\hat{s}_1, ..., \hat{s}_n)$, which is abbreviated as $\hat{s}_{1:n}$, where $n$ is the total number of measurements.

We define an activity episode $a = (x, t^-, t^+)$, where $x$ is the location of the activity episode, and $t^-$ and $t^+$ are the activity episode start and end times. The output of the probabilistic method consists in a set of candidate activity-episodes sequences $(a_1, ..., a_m)$, which is abbreviated as $a_{1:m}$, where $m$ is the total number of activity episodes. Each candidate list of activities $a_{1:m}$ is associated with the likelihood of being the true one.

This likelihood takes into account the inaccuracy in the WiFi traces, and also the capacities of the possible destinations. The activity-episodes sequence probability $P(a_{1:m}|\hat{s}_{1:n})$:

$$P(a_{1:m}|\hat{s}_{1:n}) \quad = \quad \frac{P(\hat{s}_{1:n}|a_{1:m}) \cdot P(a_{1:m})}{\sum_{a \in C} P(a_{1:m})} \tag{1}$$

where $P(\hat{s}_{1:n}|a_{1:m})$ is the measurement likelihood, $P(a_{1:m})$ is the prior knowledge we have about the activity-episodes sequence based on the capacities of the different locations, and $C$ is the set of all candidate activity-episodes sequences.

## 3 Data collection: a campus

We conduct an experiment on the campus of our university. As a localization tool, we use Cisco Context Aware Mobility API with the Cisco Mobility Services Engine (MSE). It localizes people using their smartphone through the traces in the WiFi access points. For the capacity function, we used information about number of seats in restaurants and restaurant opening hours, and number of students registered for each period of time and each classroom.

200 students from 6 different classes and 300 employees, all randomly selected at EPFL, were tracked for one day. The identity has been anonymized.

## 4 Next step

At hEART, we will present a model of the pedestrian activity choice based on the results of the Bayesian approach using the same case study.

Pedestrian destination choices are made sequentially: first destination pattern choice (i.e., the destination like "going to class" or "going for lunch" without any notion of destination or time), second the times of day choice models when activity episodes start and end, and finally, the location choice: once the pedestrian knows he wants to go for lunch and at 1pm, he can choose the restaurant. The top level decision is influenced by the lower model through expected maximum utility variables, and conversely lower level choice is conditioned by the top level decision.

The challenges are the choice set generation of possible activity locations, the measurement equations for observations, the dynamic of the system and the specification of the model.

We consider first a universal choice set for the possible locations. It includes all 5387 locations on campus. This does not model the consideration set and is not consistent with behavior, but it guarantees that no important location for the decision maker is omitted. In case this is not tractable, we will also explore importance sampling. A subset of locations are sampled for the model estimation depending on time. Then, sample bias must be corrected in the model specification to obtain consistent estimates.

An important deviation from the standard choice models is that our observations here are not deterministic. Our choice set contains several alternatives with their probabilities of being the correct one depending on data. Similarly to Frejinger (2008) in the case of route choice, we need to build a latent class model in order to estimate a location choice model. The likelihood function used in this case is:

$$P_n(\hat{s}_{1:J}|x_{1:I})P_n(x_{1:I}|\mathcal{C}_n,\beta) \tag{2}$$

where $P_n(\hat{s}_{1:J}|x_{1_I})$ is the measurement equation, giving the probability of observing signal

$\hat{s}_{1:J}$ if the actual location sequence is $x_{1:I}$, and $P_n(x_{1:I}|\mathcal{C}_n, \beta)$ is the location choice model giving the probability that individual $n$ select location sequence $x_{1:J}$ within choice set $\mathcal{C}_n$. This model depends on unknown parameters $\beta$ which must be estimated.

Location choice is a multi-stage decision process. We assume that the $i$-th location is determined only by the previous location and time and is independent from previous activity-episodes. We can thus decompose the likelihood function in Eq. (2):

$$\prod_{j_1=1}^{j_{I_1}} P_n(\hat{s}_{j_1}|x_1) P_n(x_1|\mathcal{C}_n, \beta) \prod_{i=2}^{I} \prod_{j_i=1}^{j_{I_i}} P_n(\hat{s}_{j_i}|x_i) P_n(x_i|\mathcal{C}_n, \beta, x_{i-1}, t^-_{i-1}, t^+_{i-1})$$

where $j_i$ are the indexes of signals corresponding to $i$-th location.

At hEART, we will present specification and results for the pedestrian location choice model, and validate our model.

## References


Bowman, J. L. (1998). *The Day Activity Schedule Approach to Travel Demand Analysis*, PhD thesis, Massachusetts Institute of Technology.

Frejinger, E. (2008). *Route choice analysis*, PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne.