Destination Choice Model including panel data using WiFi localization in a pedestrian facility

Loïc Tinguely,
Antonin Danalet, Matthieu de Lapparent & Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

April 16th, 2015

Table of contents

1. Introduction
2. Methodology
3. EPFL case study
   - The EPFL campus
   - Modelling
   - Results
   - Validation
   - Comments
4. Future work

1 Introduction

2 Methodology

3 EPFL case study
- The EPFL campus
- Modelling
- Results
- Validation
- Comments

4 Future work

The context of the project

Increasing transport demand worldwide and especially in Switzerland

- Need to optimize existing and future multimodal transport hubs (e.g., railway stations, airports)
- Using modern technologies (e.g., Wi-Fi localization) to track, model and understand pedestrians behavior
  - Utilize a Bayesian approach to detect pedestrian destination-sequences from Wi-Fi signatures (Danalet *et al.*, 2014).
  - Model the sequential choices of activity (Danalet & Bierlaire, 2015) and destination (here).

What we propose

A general methodology to model pedestrian destination choice using activity episodes sequences from WiFi localization.

- Accounting for panel nature of data.
- Considering field anisotropy

We present an application on the EPFL campus for activity type: eating.

- A catering destination choice model

1 Introduction

2 Methodology

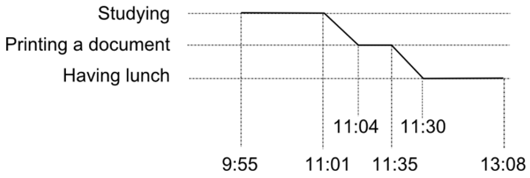
3 EPFL case study
- The EPFL campus
- Modelling
- Results
- Validation
- Comments

4 Future work

An activity episode sequence

The output of the Bayesian approach (Danalet *et al.*, 2014) consists of activity episode sequences.

Nb of observations: 112, Nb of activity episodes: 3, Date: 2012-06-29						
Start_time	End_time	Floor	Name	Type	X coordinate	Y coordinate
09:55:01	11:01:30	1	Library_name	Library	533226.888831	152274.939064
11:04:39	11:30:03	1	Printer_Lib	Printer	533229.919333	152284.564615
11:37:23	13:08:04	1	Self-service_Lib	Restaurant	533197.354323	152223.135494



Each activity type corresponds to several possible destinations.

Use of activity episode sequences

For each destination, three categories of attributes exist: sequence, activity episode and alternative attributes.

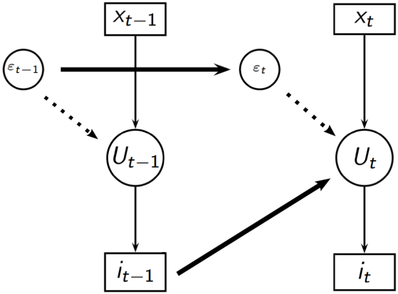
Sequence attributes	Activity episode attributes	Destination attributes
Day of the observation	Activity Type	Capacity
Socio-economic attributes	Start/end times	Price/Quality
Individual specific attributes	Coordinates	Integration
	Floor	Opening hours

The comparison of sequences of a same individual permit to catch the previous choices.

The comparison of successive activity episodes permit to calculate the distance covered (based on a weighted shortest path algorithm).

A dynamic model

The utility function at time t can take into account the choice performed at time $t-1$. It means that the observations and the error terms are not independent anymore.



Dynamic Markov model, Bierlaire (2014)

Wooldridge correction

According to Wooldridge (2002), it is possible to overcome agent effect by defining an unobserved heterogeneity density function c_i:

$$c_i | y_{i,0}, z_i \sim Normal(\alpha_0 + \alpha_1 y_{i,0} + z_i \alpha_2, \sigma_\alpha^2)$$

As a first guess, we consider that $\alpha_2 = 0$:

$$c_i = \alpha_0 y_{i,0} + \sigma_i$$

σ_i is normally distributed and independent of y_{i0}. y_{i0} is the first choice ever made by an individual i.

Three models

The choice of the alternative d at time t performed by i is rewritten as:

$$y_{d,i,t} = \beta z_{d,i,t} + \rho y_{i,t-1} + \alpha_0 y_{i,0} + \sigma_i + u_{i,t}$$

σ_i is a time-invariant unobserved effect and $u_{i,t}$ is an error term that is iid over time and individuals. We consider three models:

Static model	Dynamic strict exogenous model	Dynamic with agent effect model
$\rho = 0$ $\alpha_0 = 0$ $\sigma_i = 0$	$\rho \neq 0$ $\alpha_0 = 0$ $\sigma_i = 0$	$\rho \neq 0$ $\alpha_0 \neq 0$ $\sigma_i \neq 0$

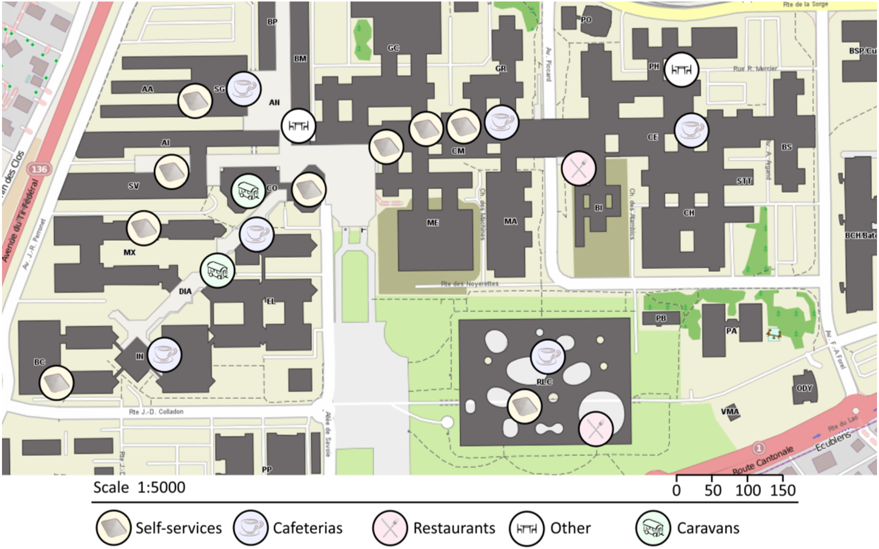
1 Introduction

2 Methodology

3 EPFL case study
- The EPFL campus
- Modelling
- Results
- Validation
- Comments

4 Future work

A map of the EPFL campus



Some facts about the eating establishments

- 9 self-services + 6 cafeterias + 2 caravans + 2 restaurants + 2 others = 21 alternatives.
- Availability of services (e.g., microwaves, sandwiches, drinks. . .), capacities and prices are similar between destinations of a same type
- The quality (food, cleanness, service) and consumers habits are regularly measured via paper-and-pencil and Internet surveys.
- Crossing the campus on foot takes between 10 and 15 minutes.

Some facts about the activity episodes

- There are 2008 visits of eating establishments during a period of 3 months performed by 192 individuals (students and employees).
- 40% of the visits are made during the lunch period (between 11:30AM and 2PM).
- In average, students and employees walk 175 meters to reach an eating establishment
- Individuals have habits since they usually visit a same destination several times.

Modelling (1)

We develop a linear in parameters Multinomial Logit Model.

$$P(d|D) = \frac{e^{\mu V_{dn}}}{\sum_{j=1}^{D} e^{\mu V_{jn}}}$$

Parameter	Variable	Variable description	Time period
ASC_d	1	-	
$\beta_{DIST\_LUNCH_{TYPE}}$	*lunch_distance*	distance from the previous activity episode 0 otherwise	lunch
$\beta_{DIST\_MORNING}$	*morning_distance*	distance from the previous activity episode 0 otherwise	morning
$\beta_{DIST\_AFTERNOON}$	*afternoon_distance*	distance from the previous activity episode 0 otherwise	afternoon
$\beta_{NO\_DISTANCE\_AV}$	*distance_not_av*	1 if no distance is available 0 otherwise	
$\beta_{EVALUATION_{TYPE}}$	*evaluation_survey*	quality evaluation on a [1;6] scale 0 otherwise	lunch
$\beta_{PRICE\_STUDENT}$	*price_min_student*	price for the cheapest hot meal if student 0 otherwise	lunch
$\beta_{PRICE\_EMPLOYEE}$	*price_min_employee*	price for the cheapest hot meal if employee 0 otherwise	lunch

Modelling (2)

Parameter	Variable	Variable description	Time period
β_{TAP_BEER}	*beer_av*	1 if tap beer is available 0 otherwise	after lunch
β_{DINNER}	*dinner_av*	1 if dinner is available 0 otherwise	dinner
$\beta_{CAPACITY\_TERRACE}$	*capacity_terrace*	outside number of seats if the weather is good 0 otherwise	lunch
$\beta_{CAPACITY\_INSIDE}$	*capacity_inside*	inside number of seats 0 otherwise	lunch
$\rho_{PREVIOUS\_CHOICE}$	*previous_choice*	1 if the destination was the previous destination 0 otherwise	lunch
α_{FIRST_CHOICE}	*first_choice*	1 if the destination was the first destination 0 otherwise	lunch
σ_d	1	-	

Obviously more parameters were tested but not kept in the model (because they were not significant or did not make sense).

Results

Results are similar for all three models.

	Static model		Dynamic strict		Dynamic agent effect	
Parameters	Value	*t*-test	Value	*t*-test	Value	*t*-test
$\beta_{DIST\_LUNCH\_CAFET}$	-0.00703	-16.69	-0.00633	-14.82	-0.00396	-7.96
$\beta_{DIST\_LUNCH\_REST}$	-0.00276	-2.18	-0.00256	-2.01	-0.00163	-0.98
$\beta_{DIST\_LUNCH\_SELF}$	-0.00646	-19.99	-0.00579	-17.38	-0.00382	-9.96
$\beta_{DIST\_MORNING}$	-0.00379	-5.97	-0.00396	-6.17	-0.00244	-3.15
$\beta_{DIST\_AFTERNOON}$	-0.000606	-1.31	-0.00103	-2.19	-0.000785	-1.32
$\beta_{NO\_DISTANCE\_AV}$	-4.89	-13.84	-4.5	-12.92	-3.26	-8.13
$\beta_{EVALUATION\_CAFET}$	1.79	9.98	1.76	9.54	1.99	8.6
$\beta_{EVALUATION\_SELF}$	1.88	9.66	1.84	9.19	2.07	8.14
$\beta_{PRICE\_STUDENT}$	-0.0681	-2.07	-0.0579	-1.73	-0.0613	-1.23
$\beta_{PRICE\_EMPLOYEE}$	-0.00537	-0.18	0.000374	0.01	0.00183	0.04
β_{TAP_BEER}	0.669	3.62	0.6	3.24	0.801	3.07
β_{DINNER}	0.943	3.35	0.986	3.5	0.474	1.31
$\beta_{CAPACITY\_TERRACE}$	0.00162	1.84	0.00148	1.65	0.00234	2.17
$\beta_{CAPACITY\_INSIDE}$	0.00277	1.29	0.00309	1.43	0.00604	2.26
$\rho_{PREVIOUS\_CHOICE}$	0	0	1.76	17.12	0.373	2.85
α_{FIRST_CHOICE}	0	0	0	0	2.21	17.8
$\mathcal{L}(0)$	-5035.429		-5035.429		-5035.429	
$\mathcal{L}(\hat{\beta})$	-3238.926		-3104.999		-2328.958	
ρ^2	0.36		0.38		0.54	

Comparison of the models

We compare the three models:

- Accounting for panel nature of data and correcting for agent effect issue increase the fit with the data and decrease the t-test of actual determinants.

	Static model	Dynamic strict exogenous	Dynamic with agent effect
$\mathcal{L}(\hat{\beta})$	-3238.926	-3104.999	-2328.958
Number of parameters	34	35	57

Likelihood ratio test

Static VS Strict: $-2(-3238.926 + 3104.999) = 266 > 3.84$

Strict VS Agent effect: $-2(-3104.999 + 2328.958) = 1552 > 33.92$

Validation

We calibrate and simulate the models on two distinct samples

	Observed market shares		Static estimate		Dynamic estimate	
	NB	%	NB	%	NB	%
Cafeteria Cafe Le Klee	0	0%	1	0.2%	1	0.2%
Self-service BC	24	6.4%	29	7.7%	29	7.8%
Other BM	17	4.5%	9	2.4%	9	2.5%
Cafeteria ELA	28	7.5%	22	6%	23	6.2%
Cafeteria INM	3	0.8%	2	0.6%	2	0.6%
Cafeteria MX	15	4%	17	4.5%	17	4.5%
Other PH	15	4%	16	4.3%	16	4.3%
Cafeteria L'Arcadie	12	3.2%	7	2%	8	2.2%
Self-service L'Atlantide	28	7.5%	29	7.8%	29	7.7%
Restaurant Le Copernic	1	0.3%	1	0.3%	1	0.3%
Self-service Le Corbusier	14	3.7%	15	3.9%	14	3.6%
Cafeteria Le Giacometti	39	10.4%	34	9%	34	9%
Self-service Le Parmentier	29	7.8%	26	7%	27	7.2%
Self-service Le Vinci	0	0%	1	0.2%	1	0.2%
Self-service L'Esplanade	70	18.7%	80	21.3%	78	20.9%
Self-service L'Ornithorynque	25	6.7%	19	5.2%	21	5.7%
Caravan Pizza	14	3.7%	15	4.1%	16	4.3%
Caravan Kebab	12	3.2%	13	3.4%	12	3.3%
Cafeteria Satellite	21	5.6%	28	7.6%	28	7.5%
Self-service Le Hodler	6	1.6%	8	2.2%	7	1.9%
Restaurant Table de Vallotton	1	0.3%	1	0.4%	1	0.4%

Results are accurate for both approaches

Comments

It is possible to develop a destination choice model for pedestrians from activity episode sequences.

The distance and the previous choice are highly significant parameters.

- Correcting agent effect issue with Wooldridge approach improves the model
- One needs to specify the time interval between activity episode sequences

1 Introduction

2 Methodology

3 EPFL case study
- The EPFL campus
- Modelling
- Results
- Validation
- Comments

4 Future work

Future work

Clearly define the time interval between activity episode sequences.

- Develop daily models (e.g., one for Mondays, Tuesdays...)
- Propose a disaggregated (over time and individuals) validation

Improve the detection of Points Of Interest.

- Use the data collected from the pedestrian counters to improve the measure of attractivity

Account for more than one candidate of activity episode sequence.

Use the developed methodology in the context of mutlimodal transport facilities.

STRC | 15th Swiss Transport Research Conference
Monte Verità / Ascona, April 15 – 17, 2015

Thank you for your attention

Now, your questions

References

Antonin Danalet, Bilal Farooq, Michel Bierlaire (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures, *Transportation Research Part C: Emering technologies*

Ben-Akiva, M. and Bierlaire, M. (2003). Discrete choice analysis, in R. Hall (ed.)

Kim, J. and J. Hespanha (2003) Discrete approximations to continuous shortest-path: application to minimum-risk path planning for groups of UAVs, 42nd IEEE International Conference on Decision and Control, 2, ISSN 0191-2216

Matthieu de Lapparent and Michel Bierlaire (2014). Mathematical modeling of behavior , Mathematical modelling of behavior

Pirotte, A (1996) Estimation de relations de long terme sur donnes panel: nouveaux rsultats, *Economie & Prvision*, 126, 143-161, ISSN 0249-4744

Wooldridge J. M. (2002), Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, Michigan State University