Inferring activities from context measurements using Bayesian inference and random utility models

Ricardo Hurtubia
Michel Bierlaire
Gunnar Flötteröd

Fifth Workshop on Discrete Choice Models
August 28, 2009
EPFL

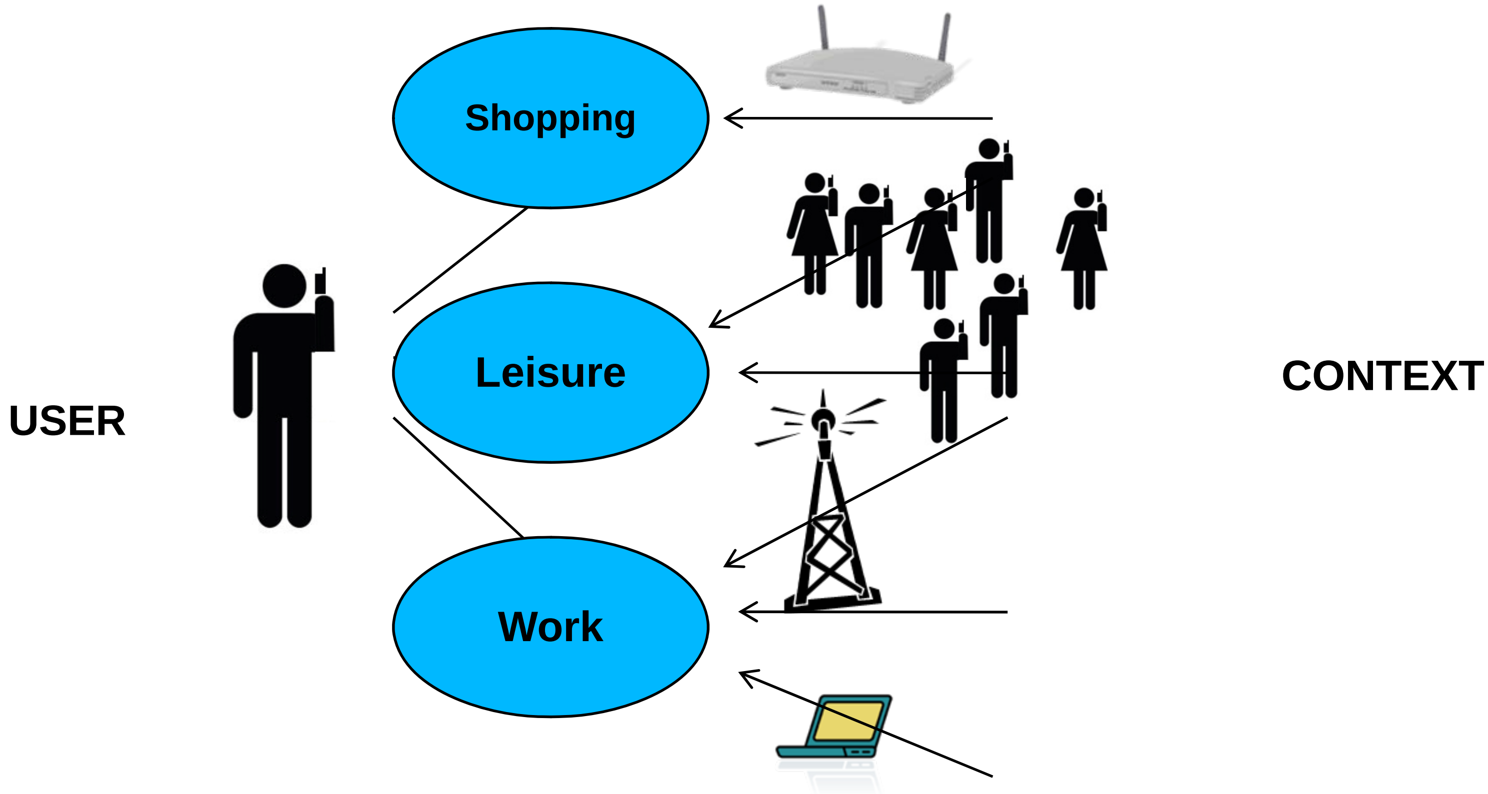
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

1. Motivation
2. Framework
   2.1 Prior model
   2.2 Measurements
   2.3 Likelihood function
3. Inference
4. Results / Case study
5. Conclusions

Motivation



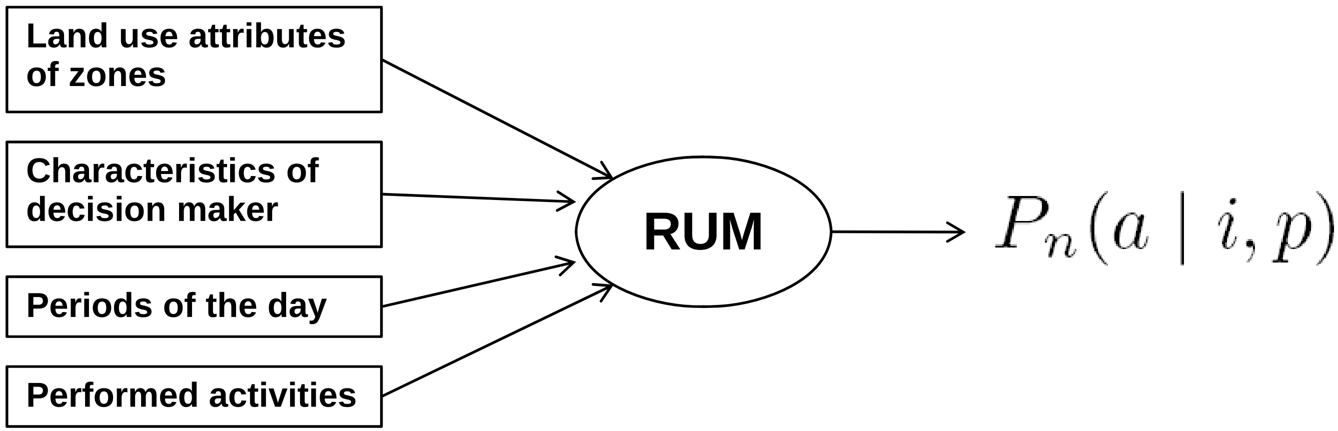
General framework

- Objective: combine general knowledge of population's behavior and individual context variables' measurements into estimates of an individual's activities
- Available data:
  - Reported activities in Swiss Transport Microcensus 2005
  - Measurements from a smartphone for one user over a two-month period
  - Activity survey
- Bayesian inference:

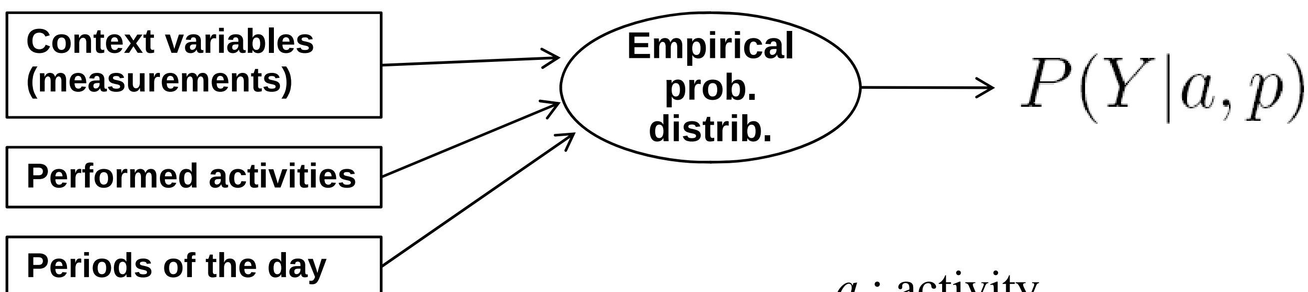
$$P(activity|measurements) \propto \underbrace{P(activity)}_{\text{Prior}} \cdot \underbrace{P(measurements|activity)}_{\text{Likelihood}}$$

General framework

- Prior:



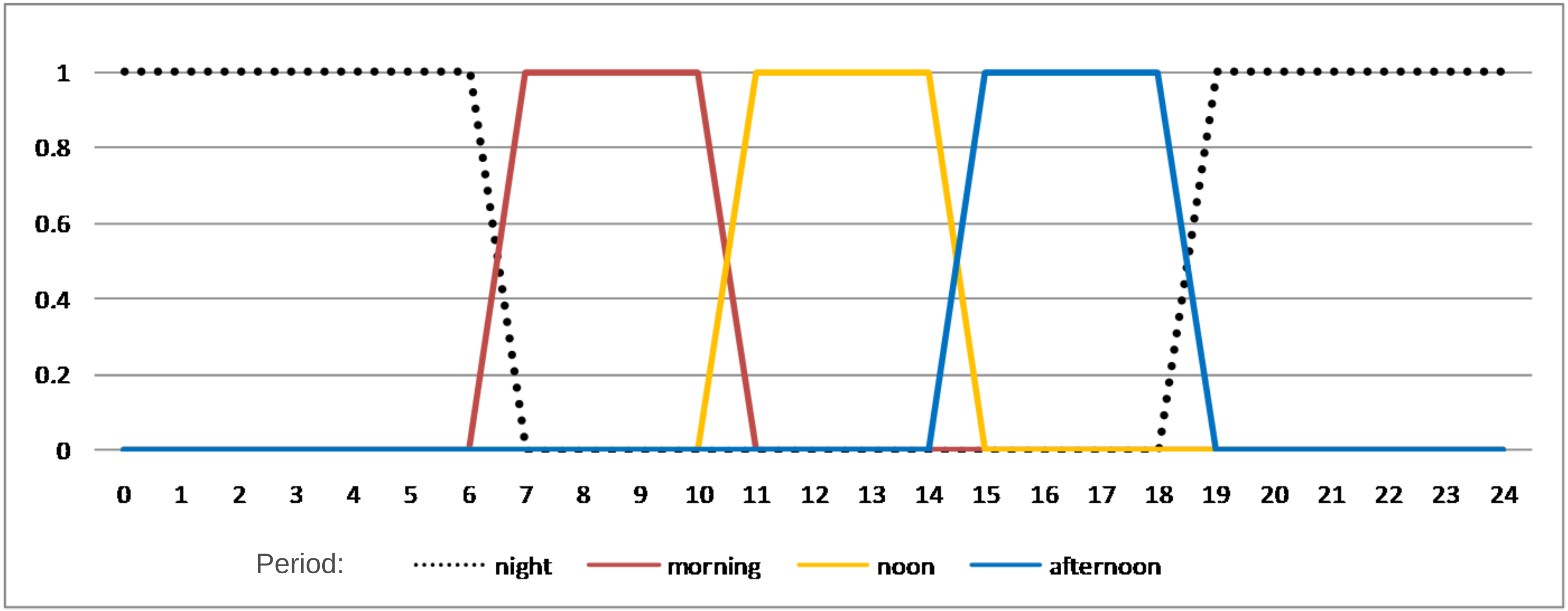
- Likelihood:



a : activity
i : zone
p : period
Y: measurement

Time discretization

- Membership function:



Prior model

- Probability of performing a certain type of activity given a location (zone) and a period of the day
- Structure: Multinomial logit

$$P_n(a \mid i, p) = \frac{exp(U_{na}(z_i, z_n, p))}{\sum_{a'} exp(U_{na'}(z_i, z_n, p))}$$

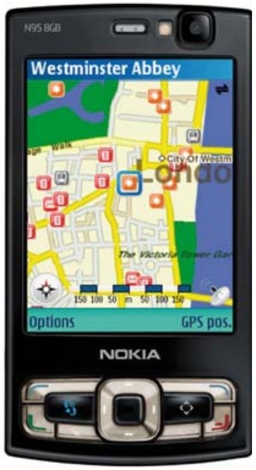
a : type of activity (work, study, leisure, shopping….)
z_i : land use attributes of zone i
z_n : attributes of user n
p : indicator of the period of the day, $p \in \{\text{morning, noon, afternoon, night}\}$

Prior model (Estimations results)

	parameter	work	study	shopping	services	leisure	other
	constant	-	-0.532	2.031	2.311	3.522	0.656
	male	0.713	-	-0.377	-0.278	-	-
n	employed	2.132	-	-	-	-	-
	children	-	-	-	-	-	0.379*
p	morning_dummy	2.720	-	0.887	1.341	-	-
	noon_dummy	1.001	-	-	-	-	-
	industry	0.025	-	-	-	-	-
	commerce	-	-	0.077	-	-	-
	services	0.046	-	-	0.055	0.024	-
	other	0.032	-	-	-	0.053	0.065*
i	retail	-	-	1.074	-	-	-
	long term retail	-	-	0.554	-	-	-
	restaurant	-	-	-	-	0.109	-
	school*age<19	-	1.694	-	-	-	-
	high_educ*student	-	1.328	-	-	-	-
	morning*student	-	6.516	-	-	-	-
	noon*student	-	4.212	-	-	-	-
$p \times n$	morning*age>60	-	-	1.114	-	0.836	-
	afternoon*age<19	-	-	-	-	0.813	-
	afternoon*age>60	-	-	-	-	-0.242	-
	night*age19_25	-	-	-	-	1.683	-

Measurements

- Measurements from a smartphone (Nokia N95)
- Variables:
  - GPS location
  - Nearby networks (LAN,GPRS, cell id)
  - Nearby Bluetooth devices
  - Movement detection (accelerometer)
  - …



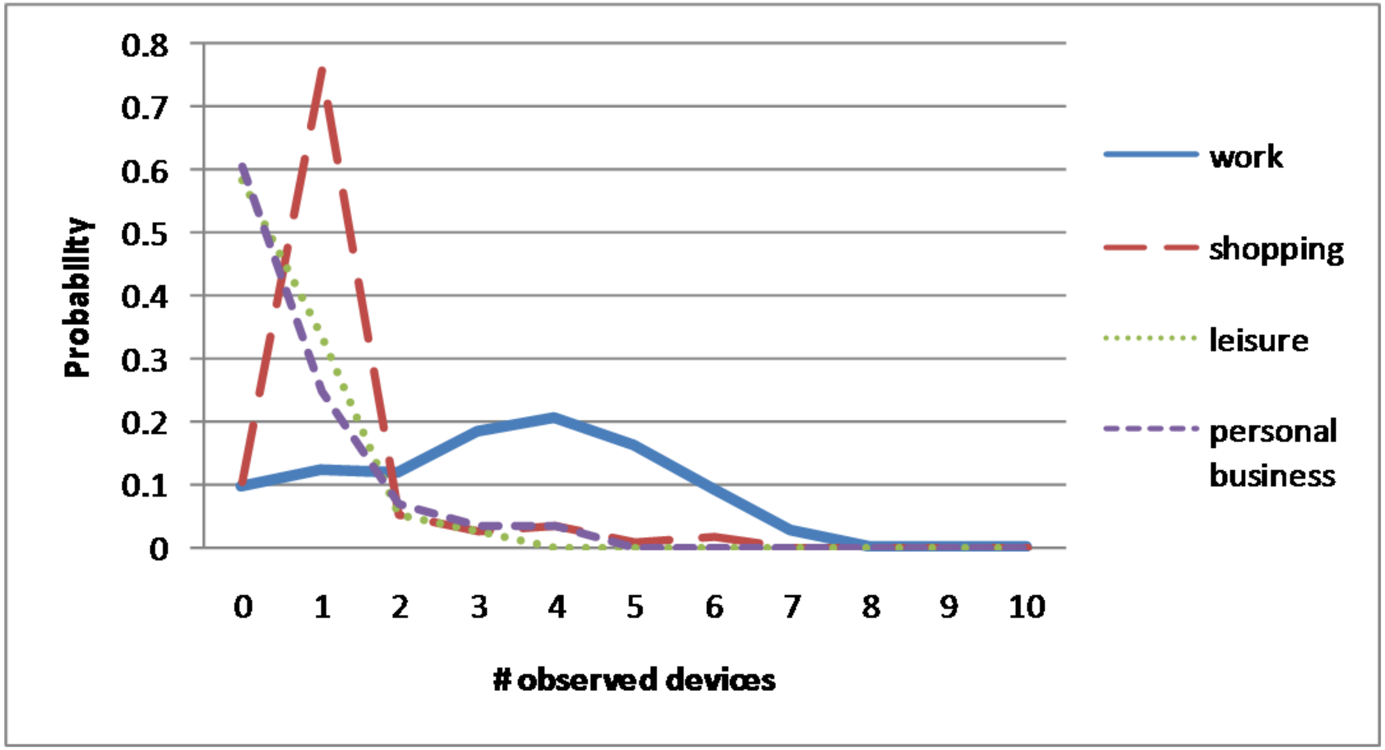
- One respondent:
  - Two months measuring context variables
  - Answering daily activity survey
    - Location
    - Time
    - Type of performed activity
    - Transport mode

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

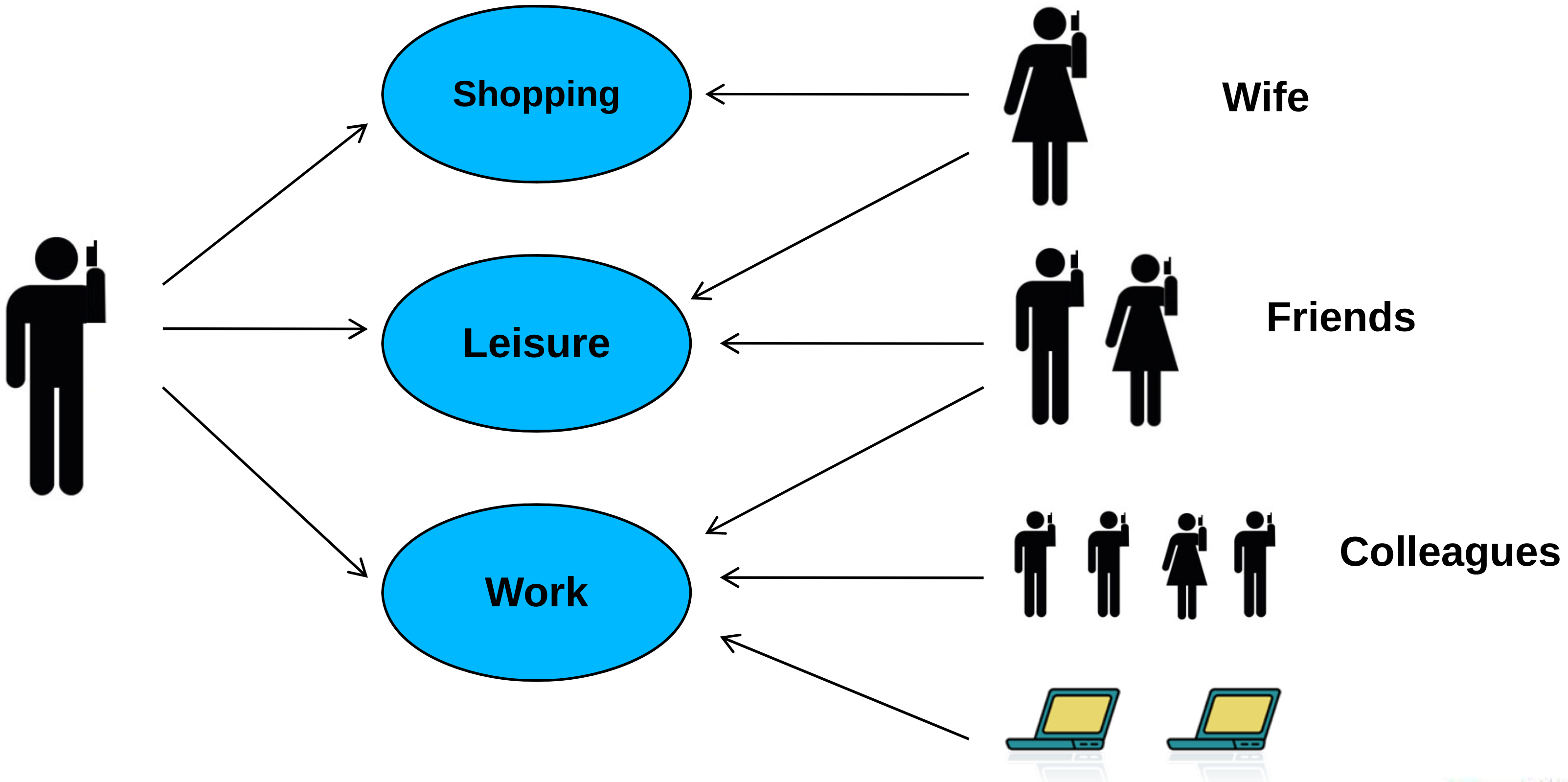
Measurements (aggregated statistics)

- Aprox 8700 measurements
- Distribution of number of detected devices:



Measurements (disaggregate)

Frequent Bluetooth devices: some devices are mostly observed when performing certain types of activities

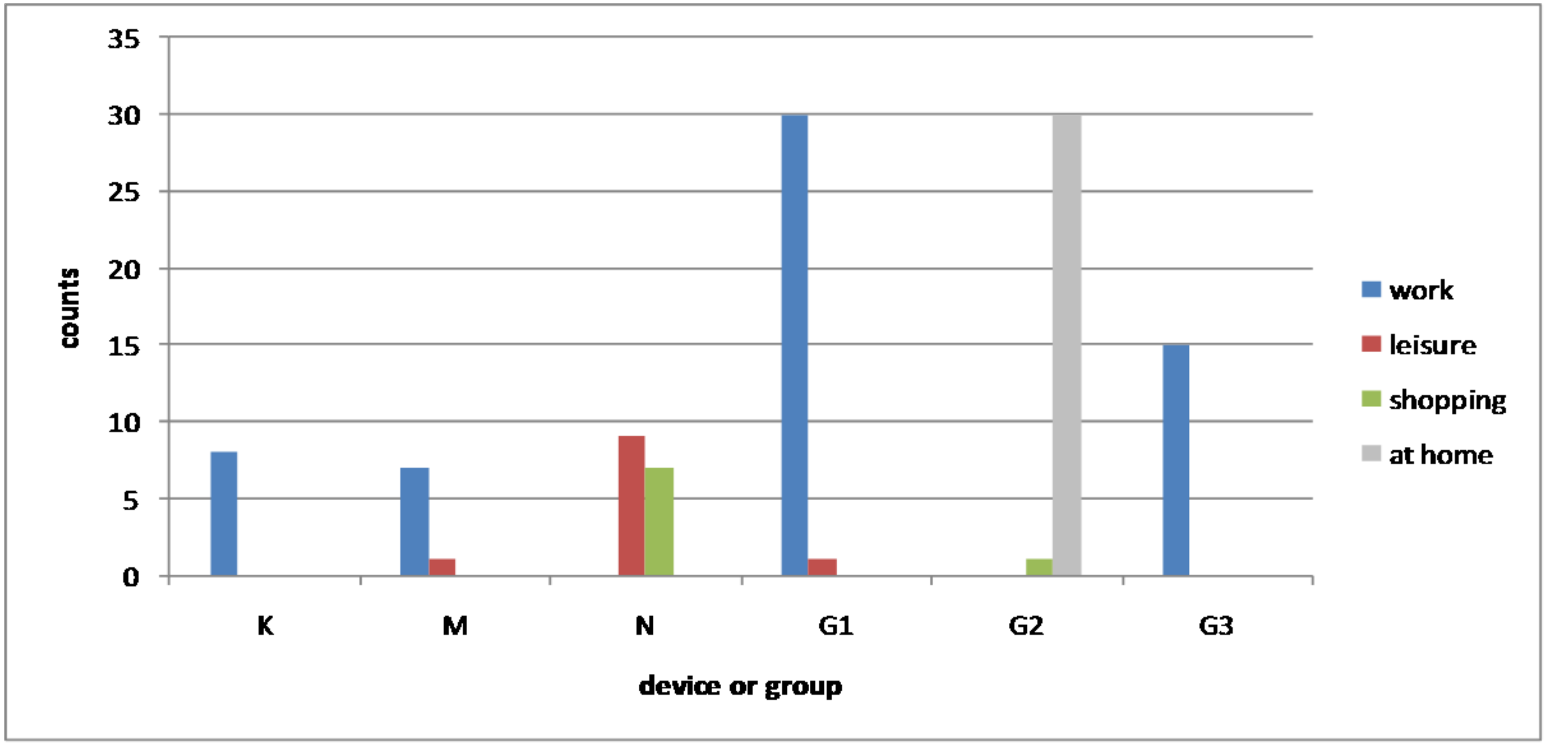


TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Measurements

- 14 independent devices appear more than 4 times
- Grouped according to activity-type correlation



Measurements

- Definitions:

$$j \in \{\text{K,M,N,G1,G2,G3}\}$$

All devices or groups (j) are assumed to be independent

State of all devices $$Y = (y_j)$$

where

$$y_j = \begin{cases} 1 & \text{if device } j \text{ is observed} \\ 0 & \text{if not} \end{cases}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood

- Empirical probability of observing a device given the activity type and period of the day:

$$P(y_j = 1 \mid a, p) = \frac{N_{jap} + \varepsilon_a \cdot \alpha}{N_{ap} + \alpha}$$

where:

– N_{ap}: number of activities type a performed during period p

– N_{jap} : number of activities type a, performed during p, where device j was detected

– ε_a : expected probability of observing any device while performing activity type a

– α : weight of "uninformed prior knowledge"

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Likelihood

- Probability of measurements given the activity type and period of the day:

$$P(Y|a,p) = \prod_j \left(P(y_j = 1|a,p) \cdot y_j + P(y_j = 0|a,p) \cdot (1 - y_j) \right)$$

Probability of observing device *j*

Probability of not observing device *j*

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Inference

- We update the prior using the likelihood of the Bluetooth devices' measurements

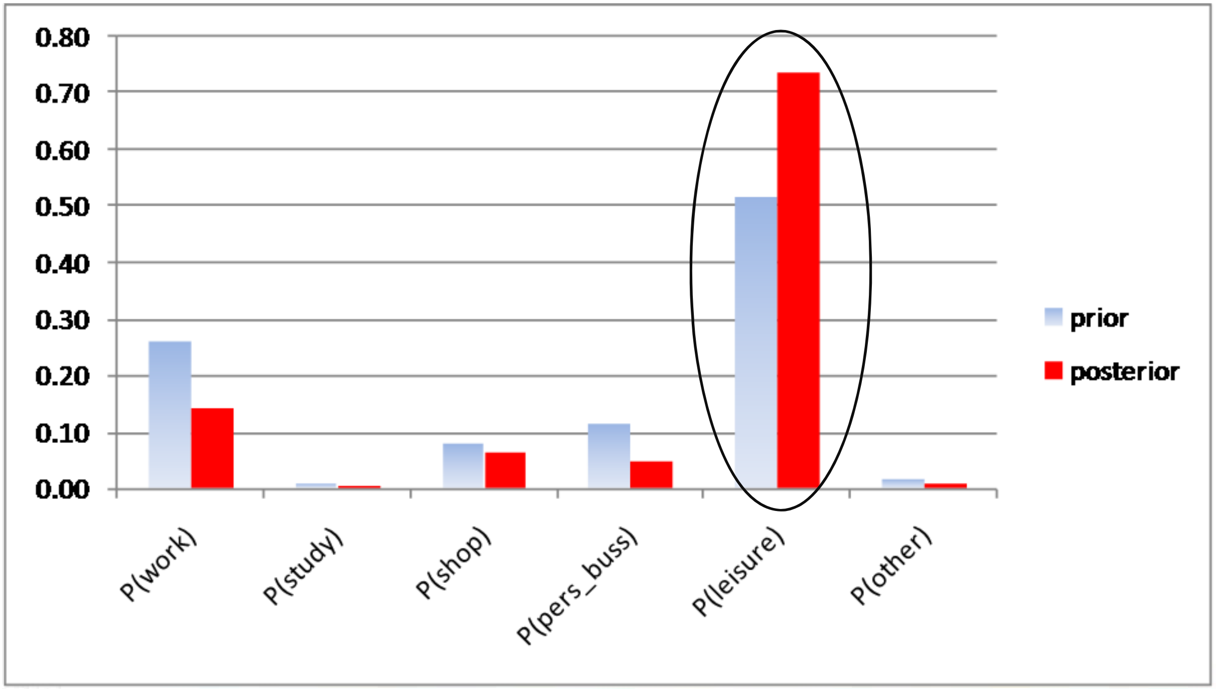
$$P(a|Y,i,p) = \frac{P(Y|a,p)\cdot P(a|i,p)}{P(Y|i,p)}$$

where:

$$P(Y|i,p) = \sum_{a'} \left(P(Y|a',p)\cdot P(a'|i,p)\right)$$

Case study

- A particular event
  - Leisure activity performed at work location during afternoon/night
  - Detection of devices:
    - K (observed only at work)
    - G1 (frequent at work, also observed at leisure)
    - N (frequent at shopping and leisure, never observed at work)



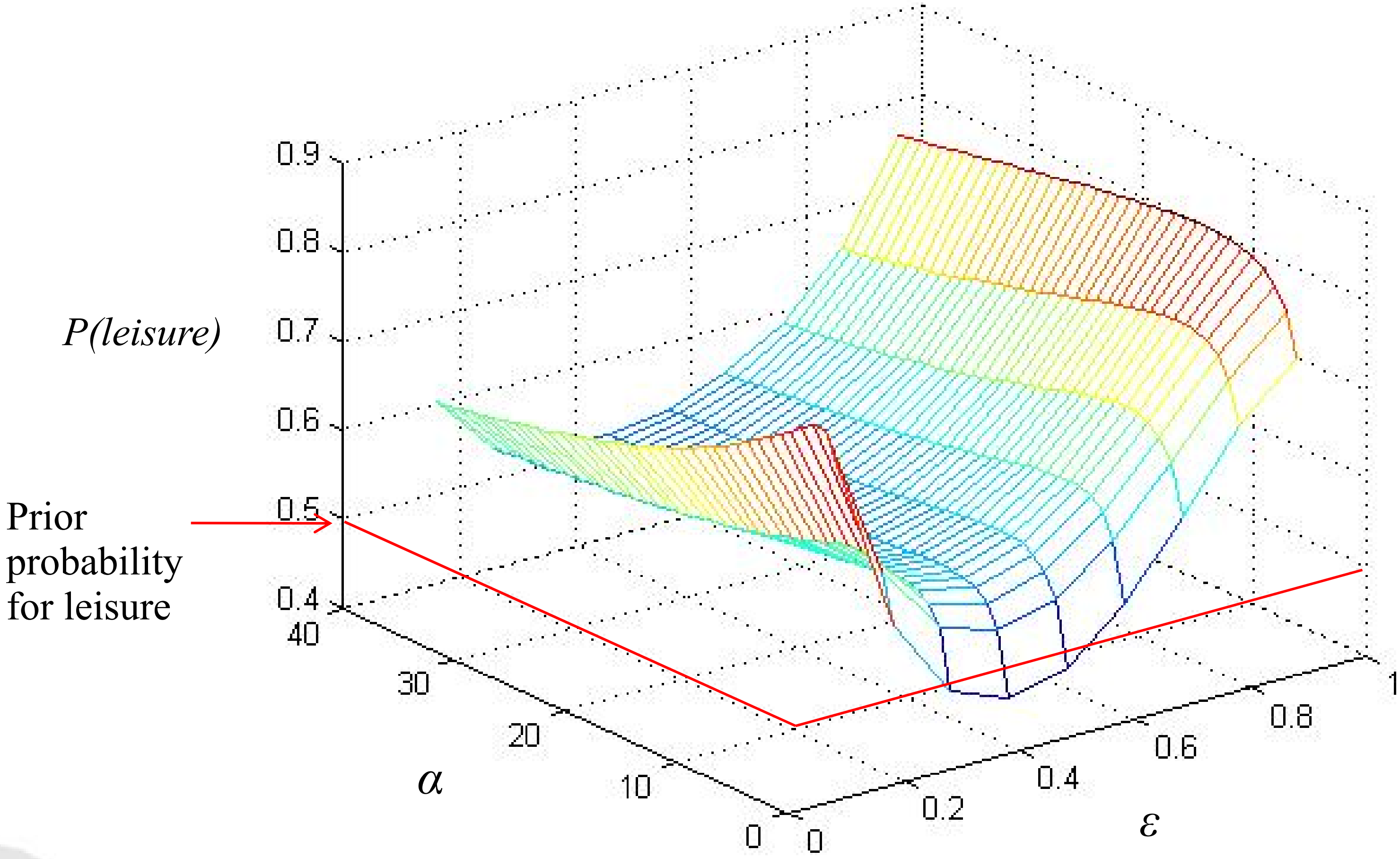
$\varepsilon = 0.01$
$\alpha = 10$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

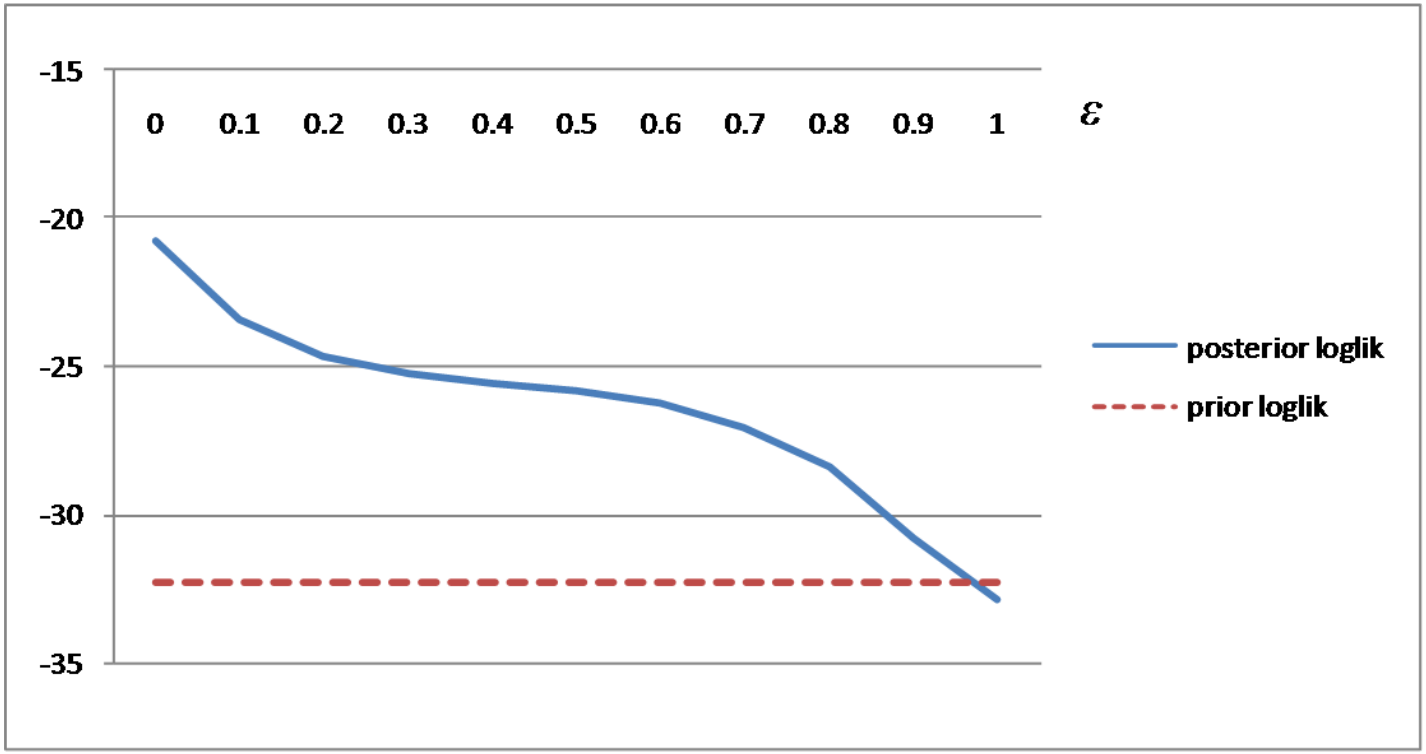
Case study

- Sensibility to α and ε.



Case study

- If we assume a high value for epsilon, the aggregate fit of the posterior distribution deteriorates



$$\sum_{k}^{K} \log\left(\sum_{a} P(a) \cdot 1_{ak}\right)$$

Conclusions and further work

- Inclusion of likelihood improves the probability distributions
- Bluetooth measurements are useful to infer activity type
- More data is required to build general models
- Link between devices (or other variables) and activities
  ➔ additional information to replace survey

Thank you

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Correlation of devices

correl	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	1	G1	G1	G1	G1	G1			G1					
B	0.73	1	G1	G1	G1	G1			G1					
C	0.79	0.78	1	G1	G1	G1			G1					
D	0.81	0.80	0.80	1	G1	G1			G1					
E	0.70	0.68	0.68	0.71	1	G1			G1					
F	0.73	0.59	0.65	0.79	0.60	1			G1					
G	-0.27	-0.25	-0.25	-0.25	-0.23	-0.23	1			G2				
H	0.51	0.61	0.48	0.57	0.40	0.49	-0.19	1				G3		
I	0.58	0.68	0.68	0.70	0.54	0.42	-0.19	0.13	1					
J	-0.26	-0.25	-0.25	-0.24	-0.22	-0.22	0.96	-0.18	-0.18	1				
K	0.41	0.52	0.52	0.54	0.48	0.40	-0.13	0.49	0.29	-0.13	1			
L	0.50	0.52	0.44	0.54	0.39	0.50	-0.13	0.70	0.08	-0.13	0.59	1		
M	0.41	0.44	0.35	0.45	0.30	0.31	-0.13	0.18	0.39	-0.13	0.32	0.18	1	
N	-0.50	-0.47	-0.47	-0.46	-0.43	-0.37	0.54	-0.35	-0.35	0.52	-0.25	-0.25	-0.17	1.00

$$correl(j, j^*) = \frac{\sum \left(y_j - \overline{y_j}\right)\left(y_{j^*} - \overline{y_{j^*}}\right)}{\sqrt{\sum \left(y_j - \overline{y_j}\right)^2 \sum \left(y_{j^*} - \overline{y_{j^*}}\right)^2}}$$

BACK