Inferring the activities of smartphone users from context measurements using Bayesian inference and random utility models

Ricardo Hurtubia
Michel Bierlaire
Gunnar Flötteröd

Transport and mobility laboratory - EPFL

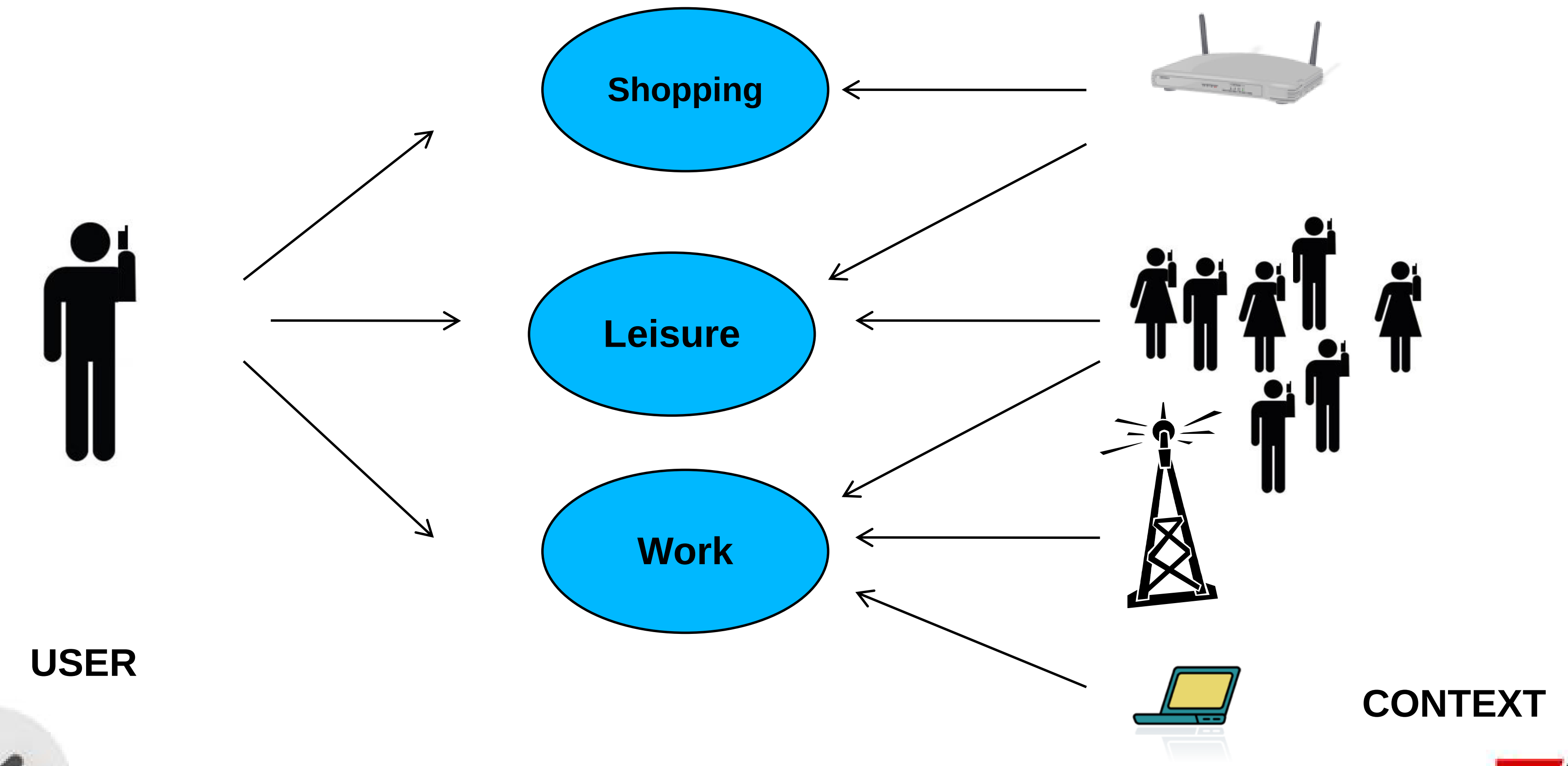
European Transport Conference - October 5, 2009

Outline

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

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
  - Land use data
  - Measurements from a smartphone for one user over a two-month period
  - Activity survey
- Bayesian inference:

$$P(activity|measurements) \propto P(activity) \cdot P(measurements|activity)$$

Prior — $P(activity)$

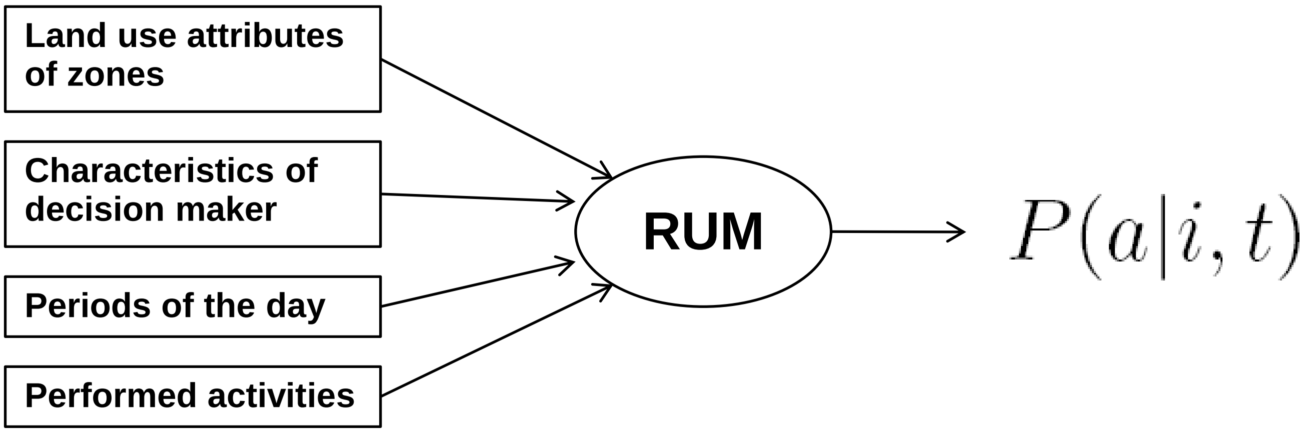
Likelihood — $P(measurements|activity)$

TRANSP-OR

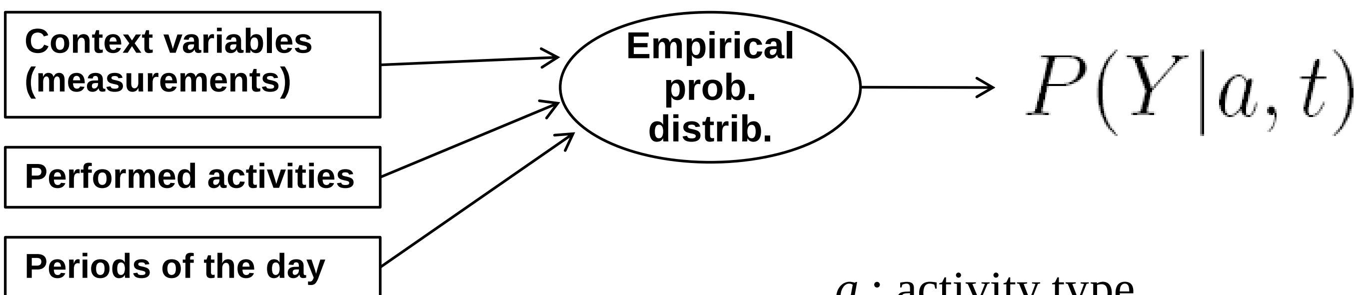
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

General framework

- Prior:



- Likelihood:



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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Prior model

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

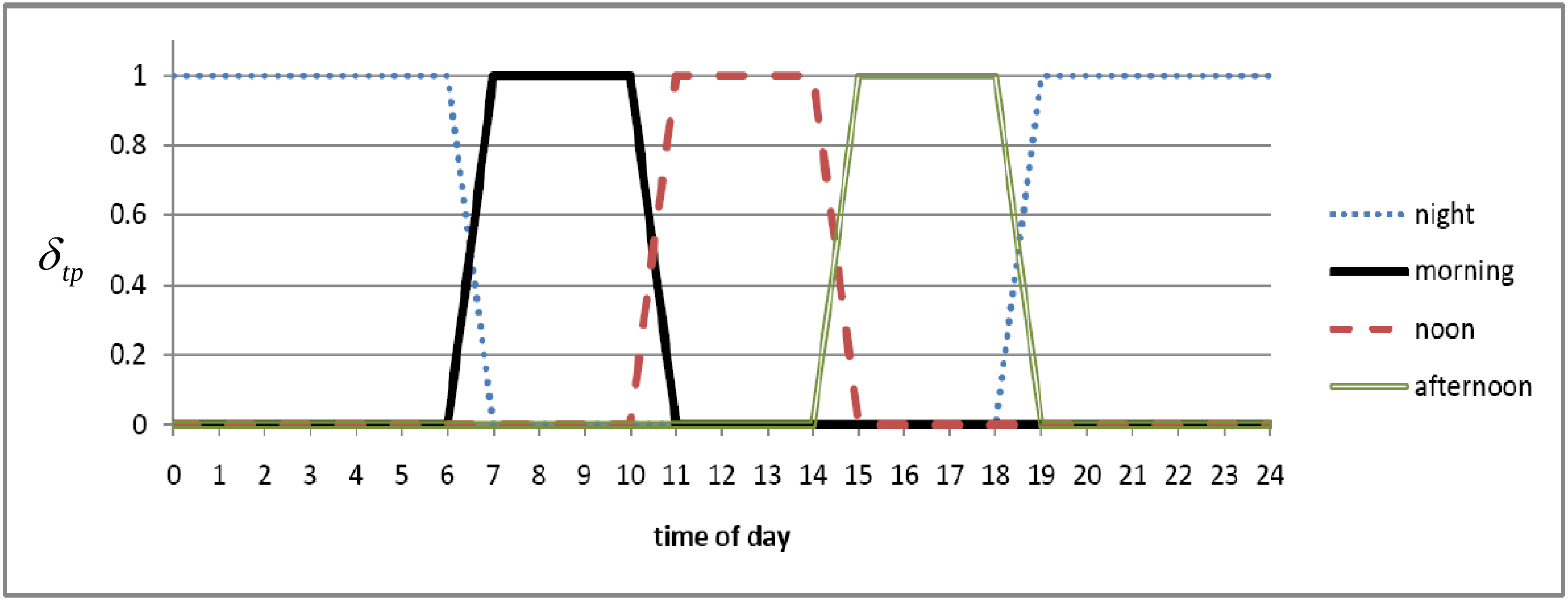
$$P_n(a \mid i, t) = \frac{\exp(U_{na}(z_i, z_n, \delta_t))}{\sum_{a'} \exp(U_{na'}(z_i, z_n, \delta_t))}$$

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

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Time discretization



$$\delta_t = (\delta_{tp}) \qquad p \in \{\text{night, morning, noon, afternoon}\}$$

Prior model estimation results

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

estimated using Biogeme (Bierlaire, 2003)

Measurements

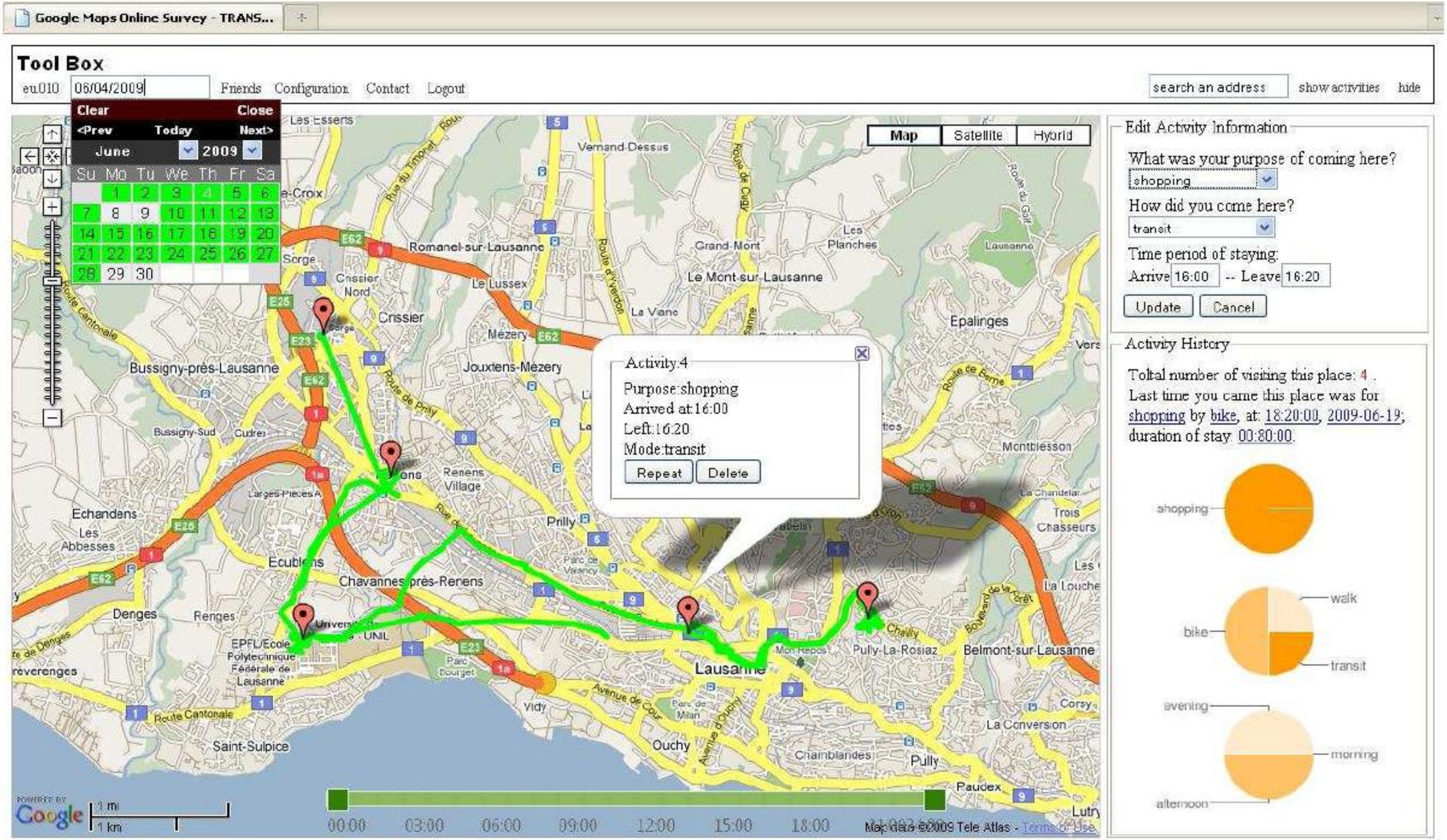
- Measurements from a smartphone (Nokia N95)
- Variables:
  - GPS location
  - Nearby networks (LAN,GPRS, cell id)
  - Nearby Bluetooth devices
  - Movement detection (accelerometer)
  - …

- 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

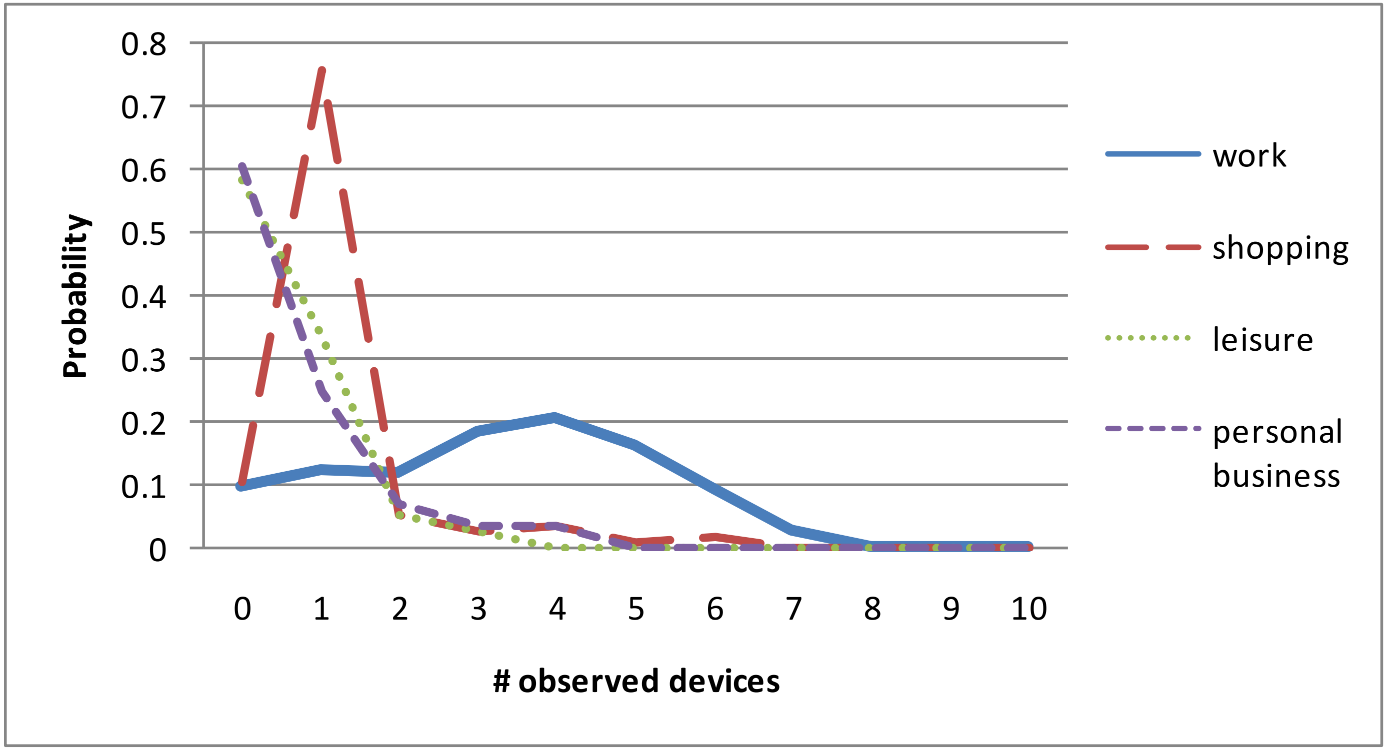
Survey



Google Maps Online Survey - TRANS...
Tool Box
eu010
06/04/2009
Friends
Configuration
Contact
Logout
search an address
show activities
hide
Clear
Close
<Prev
Today
Next>
June
2009
Su Mo Tu We Th Fr Sa
Map
Satellite
Hybrid
Edit Activity Information
What was your purpose of coming here?
shopping
How did you come here?
transit
Time period of staying:
Arrive 16:00 -- Leave 16:20
Update
Cancel
Activity History
Toltal number of visiting this place: 4 .
Last time you came this place was for shopping by bike, at: 18:20:00, 2009-06-19; duration of stay: 00:80:00.
shopping
walk
bike
transit
evening
morning
afternoon
Activity:4
Purpose:shopping
Arrived at:16:00
Left:16:20
Mode:transit
Repeat
Delete
Map data ©2009 Tele Atlas - Terms of Use

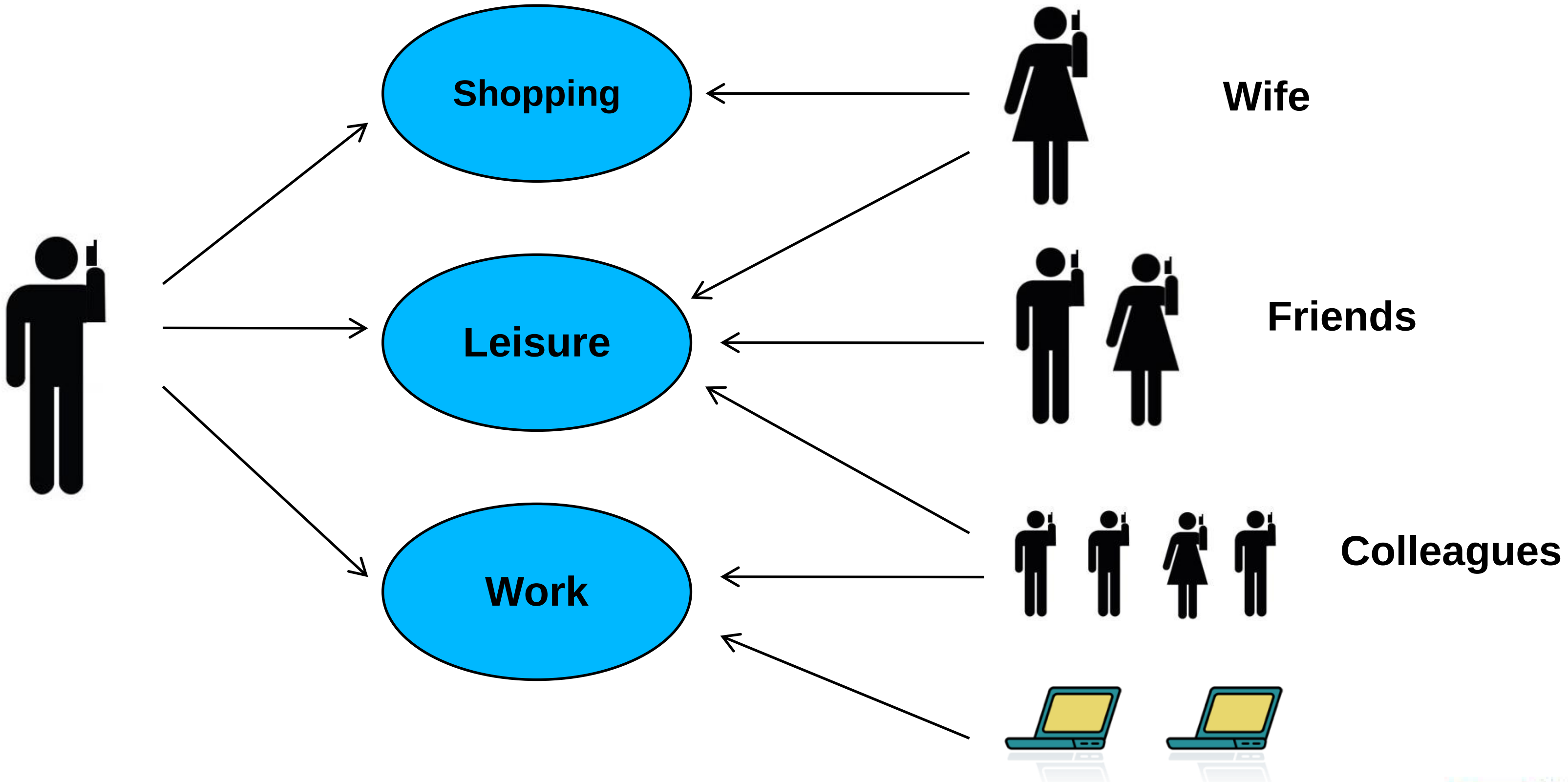
Measurements (Bluetooth devices)

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



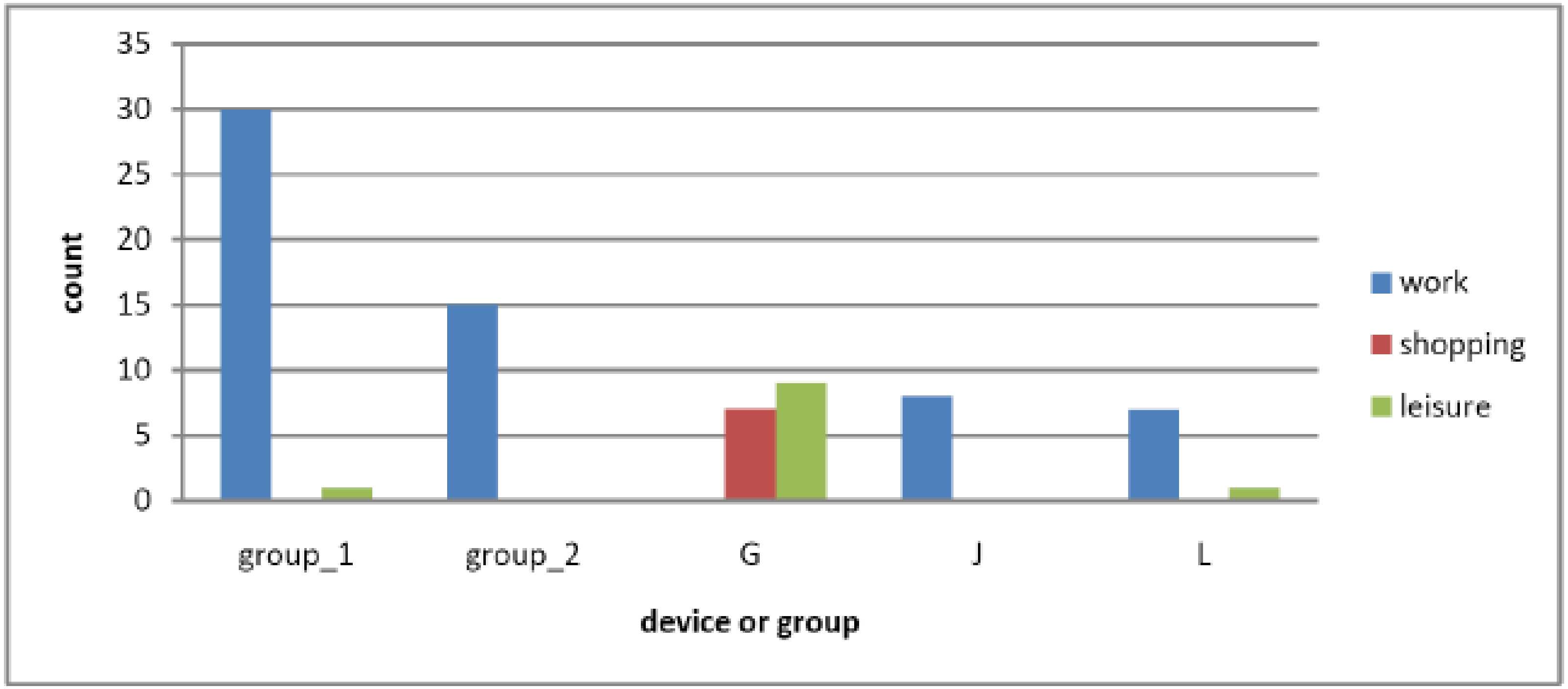
Measurements

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



Measurements

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



Measurements

- Definitions:

$$j \in \{\text{group\_1}, \text{group\_2}, \text{G}, \text{J}, \text{L}\}$$

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

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

$$P(Y|a,t) = \prod_j (P(y_j = 1|a,t) \cdot y_j + (1 - P(y_j = 1|a,t)) \cdot (1 - y_j))$$

Probability of observing device *j*

Probability of not observing device *j*

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 times activities type *a* are 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

- For a specific time of the day:

$$P(y_j = 1 \mid a, t) = \sum_p \delta_{tp} P(y_j = 1 \mid a, p)$$

Inference

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

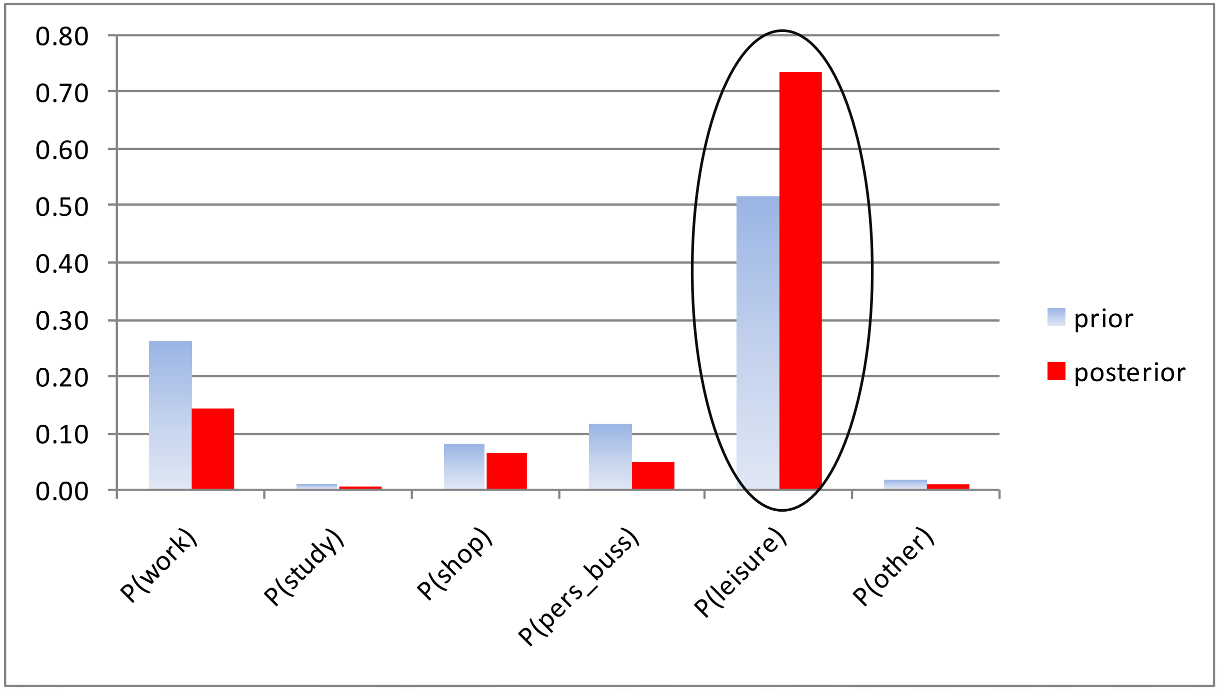
$$P(a|Y,i,t) = \frac{P(Y|a,t) \cdot P(a|i,t)}{P(Y|i,t)}$$

where:

$$P(Y|i,t) = \sum_{a'} P(Y|a',t) \cdot P(a'|i,t)$$

Case study

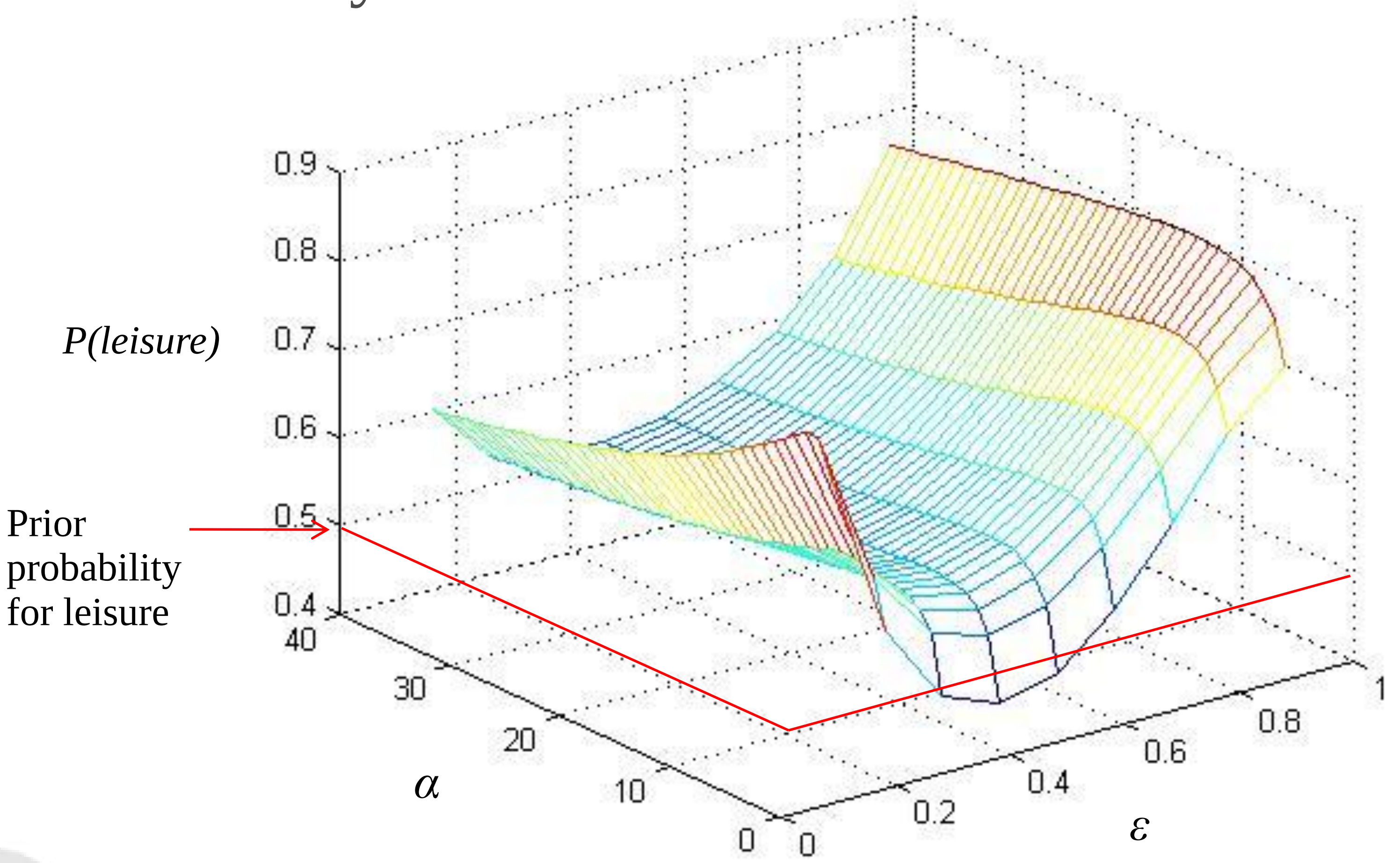
- A particular event
  - Leisure activity performed at work location during afternoon/night
  - Detection of devices:
    - Group_1 (frequent at work, also observed at leisure)
    - Device G (frequent at shopping and leisure, never observed at work)
    - Device J (observed only at work)



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

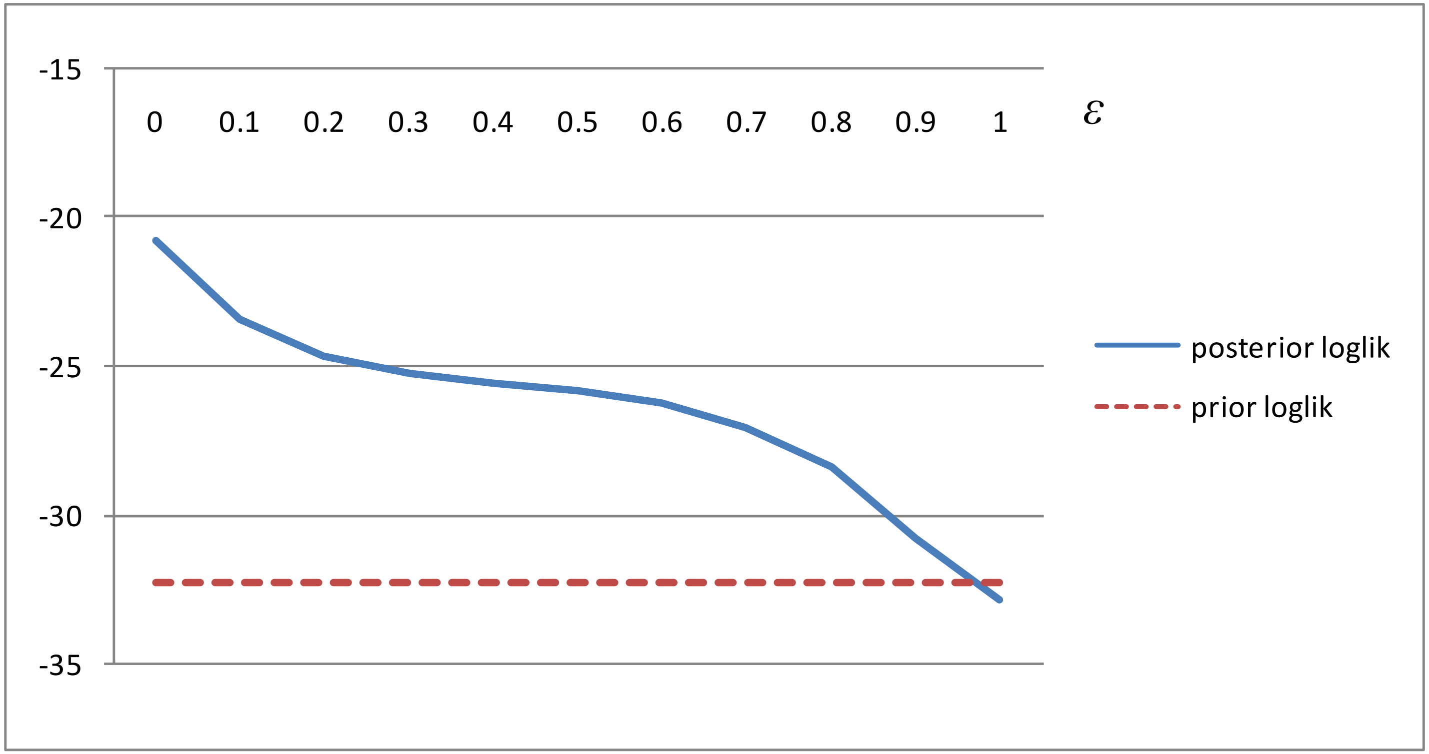
Case study

- 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

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

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

BACK

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE