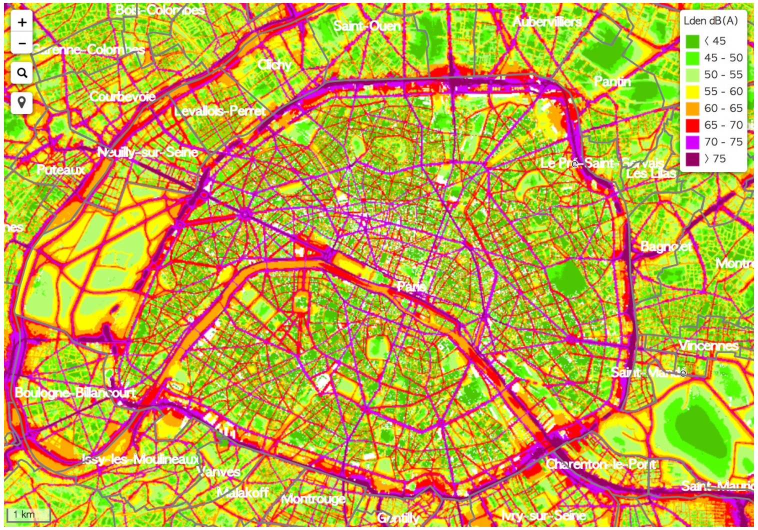
IEEE MDM’ 19

13th June

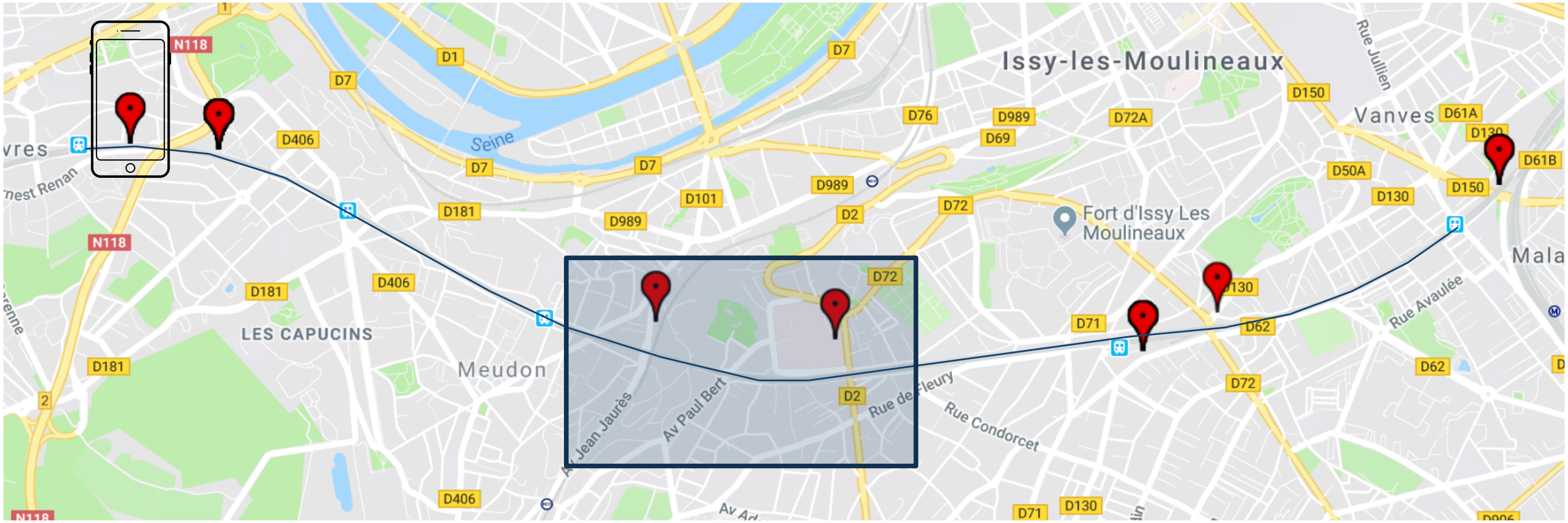
Detecting Mobile Crowd-sensing Context in the Wild

Rachit Agarwal,
Shaan Chopra,
Vassilis Christophides,
Nikolaos Georgantas,
Valérie Issarny

Motivation



Lden dB(A)
< 45
45 - 50
50 - 55
55 - 60
60 - 65
65 - 70
70 - 75
> 75
Saint-Ouen
Aubervilliers
Clichy
Pantin
Courbevoie
Levallois-Perret
Neuilly-sur-Seine
Puteaux
Boulogne-Billancourt
Issy-les-Moulineaux
Vanves
Malakoff
Montrouge
Gentilly
Vincennes
Charenton-le-Pont
Bagnolet
1 km



Issy-les-Moulineaux
Vanves
Fort d'Issy Les Moulineaux
Meudon
LES CAPUCINS
Seine
N118
D1
D7
D406
D76
D989
D69
D72A
D150
D61A
D61B
D50A
D130
D101
D181
D2
D72
D71
D62
Av Jean Jaurès
Av Paul Bert
Rue de Fleury
Rue Condorcet
Rue Avaulée
Rue Jullien

Motivation

- Identification of context is essential to assess quality of crowd-sensed spatio-temporal datasets collected in the wild and infer human mobility attributes and properties
  - Accelerometer precision can vary when devices are in/out-pocket
  - GPS accuracy can vary in underground places
  - Jump-lengths are shorter and have higher frequency when indoor

Related work

- Most of context identification algorithms rely on rich sensed features
- However, publicly available crowd-sensed datasets lack data from many sensors available on a device

	Features used	*SOTA Algorithms use*
In/out-Pocket	Light Intensity, Proximity distance, Noise level, Acceleration	Conditional checks, temporal smoothening, GMM, SVM, Variance, FFT
Under/over-ground	Pressure	Moving average
In/out-door	Light intensity, Magnetic strength, WiFi RSSI, Proximity distance, RSSI level, Time, Mobility Activity, Acceleration, Altitude, S/N Ratio, Direction, # turns when moving	HMM, CIMAP, semi-Markov CRF, conditional checks, KNN, modified GPS info. detection, SVM, sliding window

Challenge

- How contexts such as in/out-pocket, under/over-ground, in/out-door can be identified in feature limited datasets that are collected in the wild?

Our Proposal: Context Identification

- Based on available features in a dataset
- Rule based
- Unsupervised binary classifiers
- Heuristics
- Simple and effective

```
underground = False
if altitude >= 0 then
     underground = False
else if altitude < 0 then
     underground = True
else
     # also valid in case when altitude is
     # not given;
     if distance(point, underground station)< τ^uo then
          underground = True
     end
end
```

Evaluation: Dataset

- Sparse dataset with features: GPS information, Activity Mode, Network connection type, Proximity, Collection mode

1^{st} May

30^{th} July

5 seconds

$\# \bullet = 20{,}046$

- Ground truth also collected
- 4838 in-pocket, 3310 underground, and 5672 indoor points.
- Use Paris metro[3] dataset to identify underground points

[2] http://www.ambiciti.io

[3] https://github.com/ragarwa2/metroStations/tree/master/IledeFrance

Evaluation: Comparison

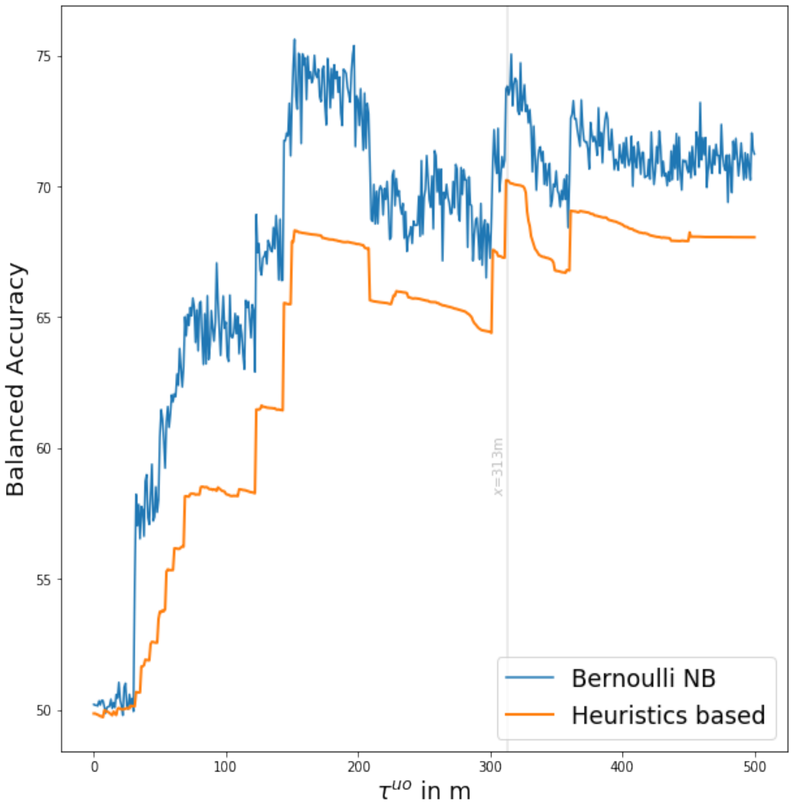
- Trained on 80% dataset
- Identify Balanced Accuracy, Precision, Recall and F1 score

[1] R. S. Olson, N. Bartley, R. J. Urbanowicz, and J. H. Moore, “Evaluation of a tree-based pipeline optimization tool for automating data science,” in Proceedings of the Genetic and Evolutionary Computation Conference 2016, ser. GECCO ’16. New York, NY, USA: ACM, 2016, pp. 485–492

Results

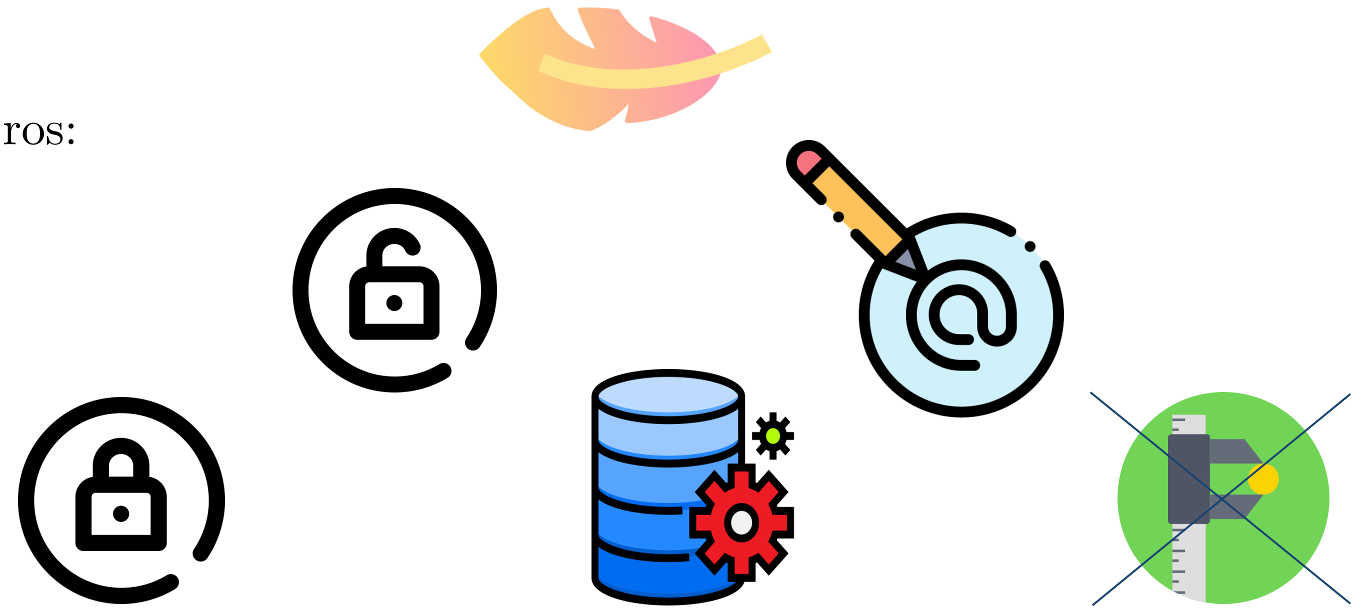
- Best balanced accuracy for underground context is achieved when $\tau^{uo} = 313$ m.
- In/out-pocket achieves equivalent performance, under/over-ground and in/out-door achieve balanced accuracy 4.3% and 1%, respectively, lower

	In/out-pocket	**Under/over-ground**	**In/out-door**
Memory	0 kB	4 kB	0 kB
Time	0.08 sec	0.17 sec	0.003 sec



+/-

- Pros:



- Con:
  - Heuristics based on assumptions
  - Feature limited

Conclusion and Future work

- Conclusion
  - Heuristics based unsupervised binary classifiers that performs with satisfactory accuracy when compared with ML algorithm obtained by TPOT

- Future perspectives
  - Integrate our algorithms into mobile applications that target human mobility
  - Exploit additional sensor data such as, e.g., light, which can be used to improve the accuracy

Thank you

rachit.agarwal@inria.fr

@ragarwa2

http://rachit.gitlab.io

https://mimove.inria.fr

Icon made by Freepik from www.flaticon.com

Related study

DATA COLLECTED AND ALGORITHM USED BY RELATED STUDIES

Related study	Uses	Data collection: # of devices/users	Data collection: Duration	Data collection: Other details	Reported accuracy in %	Power, energy, Battery duration, CPU consumption
[14]	Conditional checks	no training needed			>98	6mW, CPU≈0
[20]	Variance FFT Angle	4	10m	-	>96	-
[6]	Temporal smoothening GMM, SVM	2	14m	‡	≈80	-
[7]	moving avg.	2	3d	20+19+	>55.6†	-
[12]	HMM	3	30d	19+	>82	>7.2hr
[19]	HMM, SVM, sliding window	-	1d	-	≈75	5.7mAh
[21]	SVM	19	169m	-	>86* >96	-
[17]	KNN	1	-	-	97.27	-
[18]	modified GPS info. detection	1	-	79+, 2595^p	85.6	143.1mW
[15]	semi-Markov CRF	3	env. dependent		96	>13.7hr
[13]	CIMAP	1	-	-	>98	4mW
[16]	conditional checks	no training needed			>92	-

d: days, m: minutes, s: seconds, *: baseline, +: trajectories, p: points in diverse environment, ‡: 116k samples (50% used in training), †: for different cities and cases different accuracies are reported.

In/out-pocket Context

- Assumptions
  - Features such as proximity and how the measurement is made are available

```
inPocket = False
if measurement is made “manually” then
    inPocket = False
else
    if proximity == True then
        inPocket = True
    end
end
```

Under/over-ground Context

- Assumption
  - Flat terrain at sea-level for simplification
  - Places within τ^{uo} meters of underground metro stations are underground
  - Features such as altitude is available

```
underground = False
if altitude >= 0 then
     underground = False
else if altitude < 0 then
     underground = True
else
     # also valid in case when altitude is
     # not given;
     if distance(point, underground station)<
τuo then
          underground = True
     end
end
```

In/out-door Context

- Assumption
  - Features such as connected via is available
  - If underground, the data point is deemed indoor
  - If stationary, we assume the data point is collected in indoor setting

```
indoor = False
if underground == True then
    indoor = True
else
    if activity is still or stationary then
        indoor = True
    else
        if connected via WiFi then
            indoor = True
        end
    end
end
```

Results

	Method		*Balanced Accuracy (%)*	Precision		*recall*		*F1 score*	
		τ^{uo} *in m*	*80-20 split*	*in*	*out*	*in*	*out*	*in*	*out*
In/out-Pocket	Gaussian NB$^{+,\#}$	-	54	0.33	0.78	0.19	0.89	0.24	0.83
	Our	-	54	0.19	0.89	0.19	0.89	0.19	0.89
Under/over-ground	Bernauli NB$^{+,*}$	313	74.5	0.33	0.97	0.81	0.68	0.46	0.79
	Our	313	70.2	0.62	0.78	0.62	0.78	0.62	0.78
In/out-door	Bernauli NB$^{+,-}$	313	66	0.42	0.84	0.7	0.62	0.53	0.71
	Our	313	65	0.62	0.68	0.62	0.68	0.62	0.68

+: Best algorithm identified by TPOT, #: priors = None, var_smoothing = 10^{-9}, *: $\boldsymbol{\alpha}$ = 1, binarize = 0.0, fit_prior = False, class_prior = None, -: $\boldsymbol{\alpha}$ = 0.001, binarize = 0.0, class_prior = None, fit_prior = False