Inferring Context of Mobile Data Crowdsensed in the Wild

Rachit Agarwal+, Shaan Chopra, Vassilis Christophides, Nikolaos Georgantas, Valerie Issarny

{rachit.agarwal, shaan.chopra, vassilis.christophides, nikolaos.georgantas, valerie.issarny}@inria.fr

MOTIVATION

- Identification of sensing context is essential to assess the quality of spatio-temporal sensed datasets
- Public crowd-sensed datasets have only limited features comparing to the many sensors available on a mobile device
- How to identify context in such feature limited datasets?

CONTEXT IDENTIFICATION

- Knowledge of context such as in/out-pocket, under/over-ground and in/out-door is key:
  - Accelerometer's precision varies with in/out-pocket [1]
  - GPS accuracy can be very low when underground [2]
  - Jump lengths are short when indoor
- Related work relies on rich features and algorithms (see Table 1)
- We propose simple and effective heuristics based unsupervised binary classifiers (see Figure 1) that can work with the available limited features

DATA COLLECTION

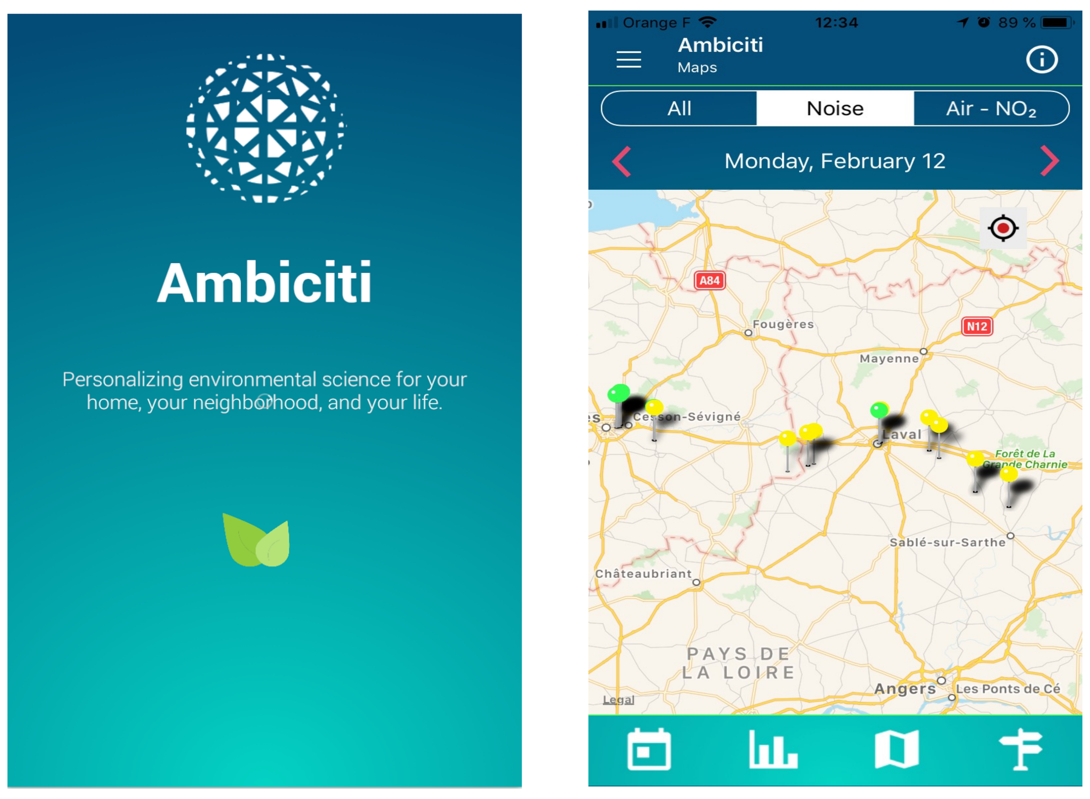


FIGURE 2: `Ambiciti` application monitors the environmental pollution a user is exposed to [3]

- Users provide ground-truth data for datapoints collected using Ambiciti application (see Figure 2)
- Use Paris metro data to identify underground points. All points within radius of τ^{uo} m of an underground station are tagged as underground

EVALUATION AND RESULTS

- Comparison with ML models built by TPOT (an AutoML tool) [4]
  - Training on 80% of the ground-truth dataset
- Identify Balanced Accuracy, Precision, Recall and F1 Score (see Table 2)
- Best balanced accuracy achieved when τ^{uo} = 313 m (see Figure 3)
- In/out pocket, under/over-ground, and in/out-door require 0kB, 4kB, and 0kB memory, respectively
- Our three heuristics based unsupervised binary classifier algorithms take 0.08sec, 0.17sec and 0.003sec, respectively, for execution

CONCLUSION/FUTURE WORK

- Our in/out-pocket achieves equivalent performance, our under/over-ground and in/out-door achieve balanced accuracy lower by 4.3% and 1%, respectively
- Our algorithms are very lightweight: Context can also be mined onboard and remain private to authorized applications

TABLE 1: Features and Algorithms used by different related studies

	Features used	*Different Algorithms used*
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

```
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 point near underground station then
        underground = True
    end
end
```

(a)

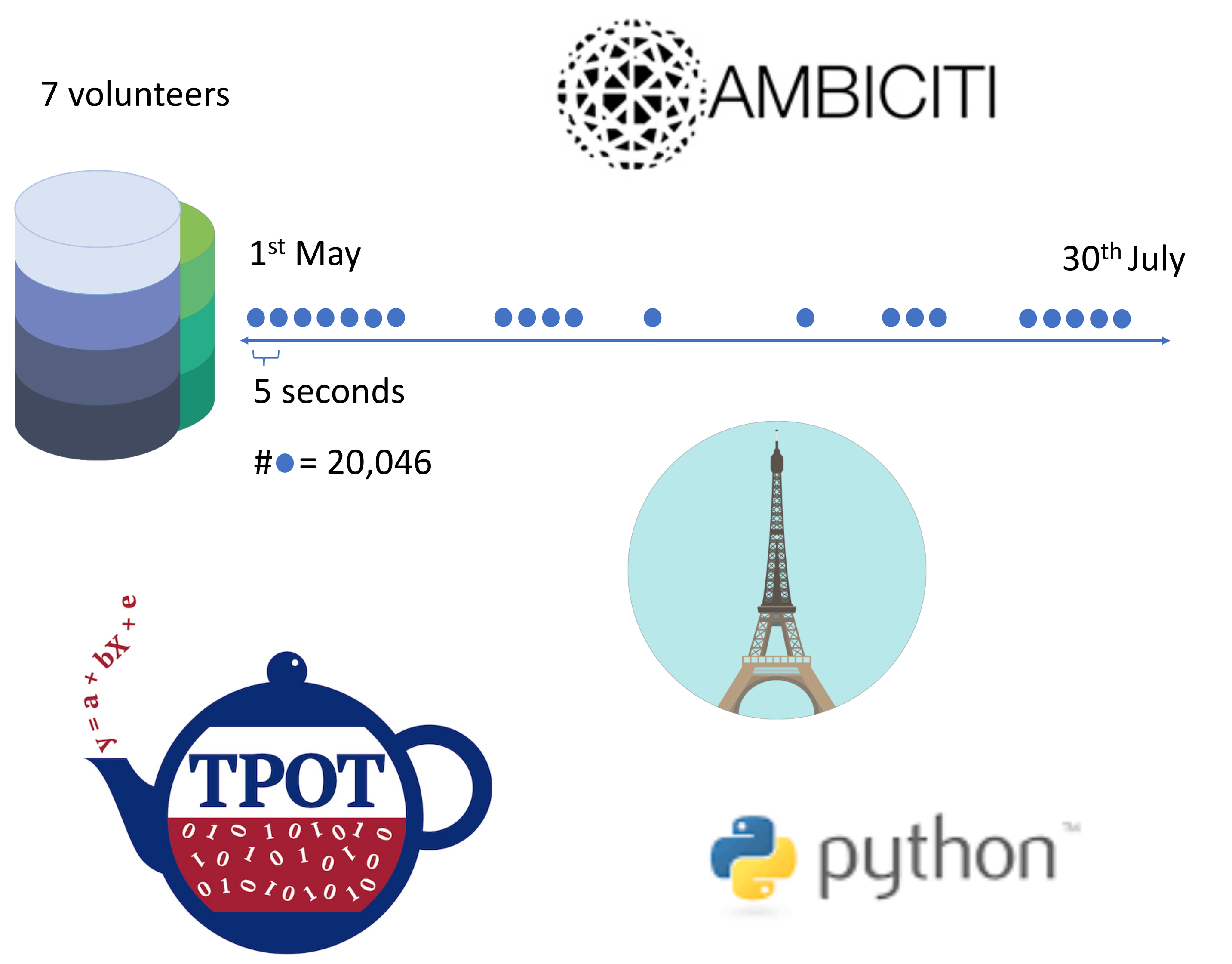
```
indoor = False
if underground == True then
    indoor = True
else
    if activity is still or stationary then
        indoor = True
    else
        if connected via WiFi then
            indoor = True
        end
    end
end
```

(b)

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

(c)

FIGURE 1: (a) Under/Over-ground assignment, (b) In/Out-door assignment, and (c) In/Out-Pocket assignment



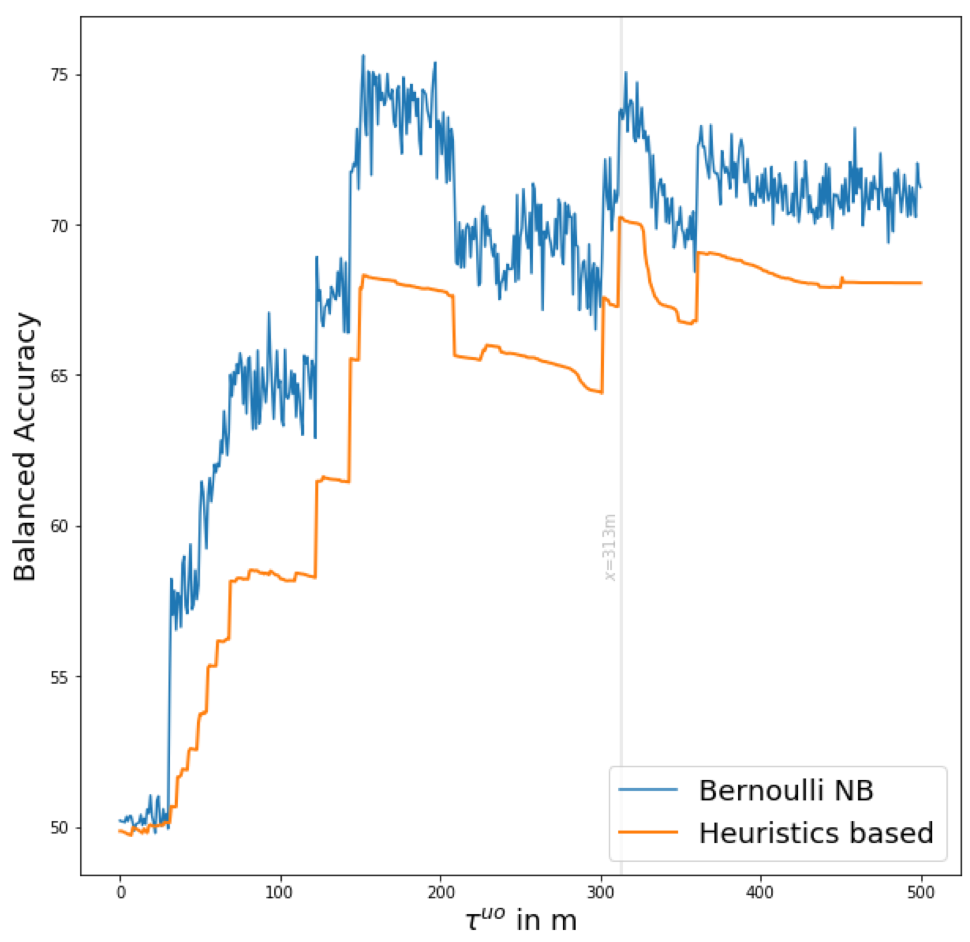


FIGURE 3: Effect of τ^{uo} on balanced accuracy

TABLE 2: Accuracy, Precision, Recall and F1 Score reported by different methods

	Method	τ^{uo} *in m*	*Balanced Accuracy (%)* *80-20 split*	Precision *in*	Precision *out*	*recall* *in*	*recall* *out*	*F1 score* *in*	*F1 score* *out*
In/out-pocket	Gaussian NB+,#	-	54	0.33	0.78	0.19	0.89	0.24	0.83
	Ours	-	54	0.19	0.89	0.19	0.89	0.19	0.89
Under/over-ground	Bernoulli NB+,*	313	74.5	0.33	0.97	0.81	0.68	0.46	0.79
	Ours	313	70.2	0.62	0.78	0.62	0.78	0.62	0.78
In/out-door	Bernoulli NB+,-	313	66	0.42	0.84	0.7	0.62	0.53	0.71
	Ours	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}, *: α = 1, binarize = 0.0, fit_prior = False, class_prior = None, -: α = 0.001, binarize = 0.0, class_prior = None, fit_prior = False

BIBLIOGRAPHY

[1] E. Miluzzo, M. Pap, N. D. Lane, H. Lu, and A. T. Campbell, "Pocket, Bag, Hand, etc.- Automatically Detecting Phone Context through Discovery," in Proceedings of First International Workshop on Sensing for App Phones (PhoneSense) at SenSys'10, pp. 21–25, 2010.

[2] K. V. Erum and J. Schoning, "SubwayAPPS: Using smartphone barometers for positioning in underground transportation environments," in Proceedings of Progress in Location-Based Services 2016, pp. 69–85, Springer International Publishing, October 2016.

[3] V. Issarny, V. Mallet, K. Nguyen, P. Raverdy, F. Rebhi, and R. Ventura, "Do's and Don'ts in Mobile Phone Sensing Middleware: Learning from a Large-Scale Experiment," in Proceedings of the 17th International Middleware Conference, (Trento), pp. 17:1–17:13, December 2016.

[4] 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, GECCO '16, (New York, NY, USA), pp. 485–492, ACM, 2016.

+ Corresponding Author

The authors would like to thank `Ambiciti` - `http://ambiciti.io` - for its support and granting us access to the specific data as a part of research collaboration.