Inferring Context of Mobile Data Crowdsensed in the Wild

Rachit Agarwal, Shaan Chopra, Vassilis Christophides, Nikolaos Georgantas, Valérie Issarny
Inria-Paris, France
Email: {rachit.agarwal, shaan.chopra, vassilis.christophides, nikolaos.georgantas, valerie.issarny}@inria.fr

Understanding the sensing context of raw data is crucial for assessing the quality of large crowdsourced spatio-temporal datasets. Accelerometer's precision can vary considerably depending on whether the phone is in-pocket or out-pocket, i.e., held in hand [1]. GPS accuracy can be very low in places like under-ground metro stations [2]. Further, jump-lengths are shorter and have higher frequency when a person is in-door. Hence, we focus on contexts such as in/out-pocket, under/over-ground, and in/out-door that can be essential for reliably inferring human mobility attributes and properties (e.g., location, jump-length, and mobility activity like walking or driving) from crowdsensed data. Our work is motivated by the fact that most of the publicly available crowdsensing datasets (e.g. PRIVA'MOV [3] and Beijing taxi dataset [4]) do not include data from specialized sensors such as light, barometer, etc. considered by state-of-the-art algorithms for detecting the above mentioned contexts. Therefore, we focus on mining context from the limited features available in the publicly available mobility related crowdsensing datasets. Moreover, as ground truth is typically not available in these datasets, we pay special attention to minimizing the training or tuning efforts of the introduced algorithms. Our algorithms are unsupervised binary classifiers with a small memory footprint and execution time. As the lack of certain features prohibits us to consider state-of-the-art algorithms as baselines, we compare the performance of our heuristic algorithms against Machine Learning (ML) models built by an AutoML tool [5] using the same set of features. Our experimental evaluation with a segment of the Ambiciti [6] dataset demonstrates that when compared to the best baseline ML model w.r.t. balanced accuracy (see Table I), our algorithm for in/out-pocket performs equally well, while for under/over-ground and in/out-door contexts, for a specific hyper-parameter, our corresponding algorithms are within 4.3% and 1%, respectively. Concerning memory, our algorithms require 0kB, 4kB, and 0kB, respectively, while they take 0.08sec, 0.17sec and 0.003sec, respectively, for execution. Our algorithms are lightweight enough to be integrated into smartphone applications. Context information mined onboard thus remains private and can be used to annotate users' personal trajectories and incentivize them to participate in crowd-measurement campaigns.

TABLE I
ACCURACY, PRECISION, RECALL AND F1 SCORE REPORTED BY DIFFERENT METHODS.

			Accuracy in % balanced	Precision		Recall		F1 score	
	Method	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
	Heuristics	-	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
	Heuristics	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	0.70	0.62	0.53	0.71
	Heuristics	313	65	0.62	0.68	0.62	0.68	0.62	0.68

† TPOT reported method

REFERENCES

[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 *First International Workshop on Sensing for App Phones (PhoneSense) at SenSys'10*, p. 21–25, 2010.

[2] K. v. Erum and J. Schöning, "SubwayAPPS: Using smartphone barometers for positioning in underground transportation environments," in *Progress in Location-Based Services 2016*, pp. 69–85, Springer International Publishing, oct 2016.

[3] S. Ben Mokhtar, A. Boutet, L. Bouzouina, P. Bonnel, O. Brette, L. Brunie, M. Cunche, S. D'Alu, V. Primault, P. Raveneau, H. Rivano, and R. Stanica, "PRIVA'MOV: Analysing Human Mobility through Multi-Sensor Datasets," in *NetMob: Book of Abstracts - Posters*, pp. 19–21, 2017.

[4] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: Driving Directions Based on Taxi Trajectories," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, (New York, NY, USA), pp. 99–108, ACM, 2010.

[5] 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.

[6] 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.