MOBILITYNET: TOWARDS A PUBLIC DATASET FOR MULTI-MODAL MOBILITY RESEARCH

Kalyanaraman Shankari
UC Berkeley
shankari
@eecs.berkeley.edu

Jonathan Fuerst
NEC Labs Europe
jonathan.fuerst
@neclab.eu

Mauricio Fadel Argerich
NEC Labs Europe
mauricio.fadel
@neclab.eu

Eleftherios Avramidis
German Research Center for AI (DFKI)
eleftherios.avramidis@dfki.de

Jesse Zhang
UC Berkeley
jessezhang@berkeley.edu

ABSTRACT

Influencing transportation demand can significantly reduce CO_2 emissions. Individual user mobility models are key to influencing demand at personal and structural levels. Constructing such models is a challenging task that depends on a number of interdependent steps. Progress on this task is hamstrung by the lack of high quality public datasets. We introduce *MobilityNet*: the first step towards a common ground for multi-modal mobility research. MobilityNet solves the holistic evaluation, privacy preservation and fine grained ground truth problems through the use of artificial trips, control phones, and repeated travel. It currently includes 1080 hours of data from both Android and iOS, representing 16 different travel contexts and 4 different sensing configurations.

1 INTRODUCTION

The transportation sector accounts for about a quarter of energy-related CO_2 emissions and decarbonization is challenging since it depends on changes at the individual consumer level (Rolnick et al., 2019). An accurate individual mobility profile, constructed with minimal user input, can help influence both individual demand through incentivization (Zhang et al., 2019) and infrastructure changes that overcome barriers to mode shift (Dill & McNeil, 2012).

There has been much work, both in academia and industry, on collecting and analyzing fine-grained individual location traces, relying on user smartphones as sensing devices. However, the procedure to *evaluate* the performance of these systems (e.g., accuracy/power trade-off) and the associated machine learning algorithms (e.g., transport mode classification) has largely been an afterthought. There are few public datasets shared (see Table 1) and due to privacy reasons, only two of them include location information. This is a problem, as reproducible evaluations on common benchmark datasets (e.g. ImageNet; Deng et al., 2009) are critical to improve accuracy and generate robust solutions on which policy makers can base decisions.

In this paper, we propose *MobilityNet* as a first step towards a common ground for multi-modal mobility research. MobilityNet is comprised of: (i) a public, privacy non-sensitive, multi-modal mobility dataset, (ii) a data collection and evaluation procedure to capture such datasets and (iii) a system that can be used to expand this dataset in the future by collecting data in other regions. Our aim is to build a community of machine learning experts, transportation specialists and smartphone sensing platform providers, who jointly improve data collection methods and the performance of common transportation specific classification problems. We strongly believe that this will also ease the way for AI experts from other domains to work on novel solutions that can contribute in the fight against climate change.

2 A PROCEDURE FOR EVALUATING HUMAN MOBILITY SYSTEMS

Human Mobility Systems (HMSes) need a procedure for rigorous evaluation that allows users of the data to understand their limitations and their accuracy in various settings (e.g., what is the resulting transport mode detection performance

Table 1: Summary of published mobility datasets, collected from android phones or dedicated GPS devices

Name	Year	Description
Opportunity Activity Recognition Challenge (Chavarriaga et al., 2013)	2011	12 users, 6 runs each, 72 wearable, object and ambient sensors, indoor setting, no GPS
Microsoft Geolife Dataset (Zheng et al., 2008)	2012	182 users, 3 years, GPS data from dedicated devices
US-TransportationMode Set (Carpineti et al., 2018)	2018	13 users, 31 hours, multiple smartphone sensors, no GPS.
Sussex-Huawei Transportation-Locomotion (SHL) Recognition Challenge (Wang et al., 2018)	2018	Single participant and single phone, 4 months, accelerometer, gyroscope, magnetometer, linear acceleration, gravity, orientation (quaternions), ambient pressure, fused location (GPS/WiFi/cell)

under various sampling regimes?). Before establishing such a method, we need to understand the evaluation requirements, and the challenges associated with meeting each requirement:

• **Holistic evaluation - power vs. overall accuracy**: There is a clear power/accuracy trade-off for smartphone sensing. Naïve high accuracy sensing, even for low power sensors, quickly drains the battery (Srinivasan & Phan, 2012), but techniques to lower battery drain also lower the accuracy. So it is critical that the evaluation considers accuracy in the context of the power consumed.

• **Privacy preserving**: The data collected by HMSes includes location traces, which are inherently privacy sensitive. Location traces allow re-identification from the raw data alone (Zang & Bolot, 2011; de Montjoye et al., 2013), *even after replacing personally identifiable information!*

• **Ground truthed**: In order to fully evaluate the data collected, we need ground truth for not just the mode, but also the trip start and end times, section start and end times and the travel trajectory. Labeling trips through prompted recall is a low effort technique to collect mode ground truth, but it is likely to be unreliable (Stopher et al., 2015, p.206-207). Similarly, for evaluating trajectories, spatiotemporal ground truth is almost impossible to obtain after the fact.

With MobilityNet, we propose the first public dataset that meets all these requirements. It utilizes three novel concepts in its construction: **(1) artificial trips**, which preserve privacy and provide spatial ground truth, **(2) control phones**, which provide temporal reference data and comparisons to a baseline, and **(3) repeated travel**, which controls for context sensitive variations in sensing. We implement a system that combines prior work on power evaluations (Shankari et al., 2018b) with an existing HMS platform E-MISSION (Shankari et al., 2018a). The system consists of 3 main parts:

• **Evaluation Specification**: The *spec* describes an evaluation that has been or will be performed. In addition to mode and trajectory ground truth, it includes the app configurations to be compared and the mapping from phones to evaluation *roles*. The spec automatically configures both the data collection app and the standard analysis modules.

• **Auto-configured Smartphone App**: We developed a custom user interface (UI), focused on evaluation. It allows evaluators to select the current spec from a public datastore and automatically download the potential comparisons to be evaluated, the role mappings and the timeline. When the data collectors perform the trips, they mark the transition ground truth in the UI and the app automatically displays the next step in the timeline (see Figure 1).

• **Public Data + Sample Access Modules**: Since there are no privacy constraints, our system uploads all collected data to a public server. The associated repository contains sample Binder notebooks (Bussonnier et al., 2018) that can download, visualize and evaluate the data associated with a particular spec.

3 MOBILITYNET

Our initial dataset contains 1080 hours of data from 3 artificial timelines. They cover 16 different travel contexts, including newly popular modes such as e-scooter and e-bike (see Table 2). For each timeline, we collect data with multiple phones and for different data collection regimes (e.g., sampling frequency). The detailed timeline specifications are included in the dataset.[1] Our data collection had three main goals:

(1) Dwell time: Instead of focusing only on trips, we evaluate a timeline that included significant dwell time to capture the impact of context sensitive behavior, such as Android's built-in duty cycling. We set our timeline trips as round trips with an intermediate dwell time $\sim 3\times$ the mean travel time to the location.

[1]Dataset and documentation: `https://mobility-net.org`

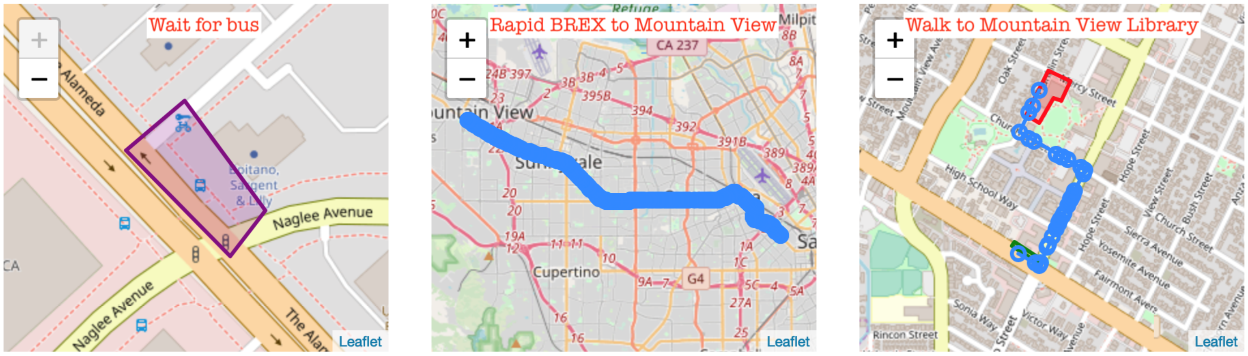


Figure 1: Shortened sample spec for a multi-modal trip, including transfers and waits for public transit.

Table 2: Summary of the collected timelines

Description	Outgoing trip modes	Incoming trip modes	Overall
suburban round trip	car (suburban street)	bicycle	72h
downtown library	car (freeway)	escooter, bus rapid transit	216h
multi-modal trip across SF bay	suburb walk, commuter train, subway, city bus, university walk	ebike, express bus, downtown walk, light rail, commuter train with tunnels, suburb walk	792h
Total			1080h

(2) Broad range of modes: Since we create artificial trips, we structure them to maximize mode variety. To efficiently cover this space, we tried to ensure no mode was repeated. Even similar modes (car, commuter rail) were in different travel contexts (e.g. street versus freeway) for maximum variety. **(3) Multi-modal transfers**: Detecting multi-modal transfers is complex because there is not a clear signal similar to a trip end. We ensure that there are many transition examples when constructing our artificial trips.

The data is primarily from *virtual sensors* – closed source APIs built into the phone OS that generate location and motion results from raw sensor data. These include:

- **Fused location:** Virtual sensor from GPS/WiFi/cellular signals. It includes *timestamp (ts), latitude, longtitude* (always), and *accuracy*, *speed*, *heading* (sometimes).
- **Motion activity:** Virtual sensor from accelerometer/gyroscope/barometer signals. It includes *ts, confidence, type* (e.g. *walking, automotive*). It does not distinguish motorized modes.
- **Trip transition events:** Combination of virtual (e.g. *exited geofence, visit started/ended*) and custom platform duty cycling events (e.g. *stopped moving, tracking stopped, booted*). It includes *ts, current state, transition*. State and transition constants are defined in the platform.
- **Battery:** Voltage and current sensor. It includes *ts, battery status, battery level percent*.

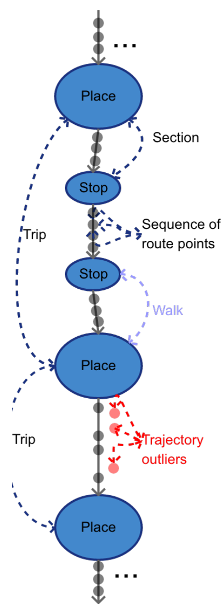


Figure 2

3.1 Challenge: From raw data to individual mobility diaries

A key challenge to deriving mobility insights for improving individual transportation behavior is converting the data into an individual mobility diary. Figure 2 depicts such a diary, where the raw time series data has been converted into trips and section trajectories with assigned transport modes. Multiple steps are necessary for its construction:

- **Trip segmentation:** Split into a linked sequence of *trips* and *places*. Since the phone OS automatically duty cycles sensing to low power, the input timeseries will have gaps, and the first few points of a trip will be lost to cold starts.
- **Section segmentation:** Converts trips into a linked sequence of *sections* and *stops*. Each section represents travel by one mode – multi-modal trips will consist of multiple sections while unimodal trips will consist of only one section.
- **Trajectory filtering:** Location data can frequently be very noisy and, particularly in underground sections, generate errors in the range of 25km. This step identifies these erroneous points so that they can be removed.

• **Mode inference:** Use inference algorithms to determine the mode for each section in the mobility diary. The accuracy of inference algorithms typically varies widely across modes; modes with similar speed characteristics (e.g., bicycling and buses on city streets) are hard to distinguish, especially at low accuracy/frequency sensing levels.

3.2 Metrics and baseline results

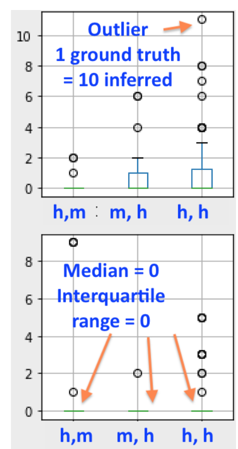
The analysis generates multiple outputs, so we need metrics along multiple dimensions: **(1) Trip and section segmentation**: As metrics, we define the differences in the (i) count and (ii) start and end timestamps for the inferred and ground truthed trips and sections. **(2) Trajectory outlier detection**: For *spatial* accuracy, the metric is the distribution of the perpendicular distance between the final route points and the ground truth trajectory. For *spatio-temporal* accuracy, since we don't have ground truth, we create a reference trajectory from the two accuracy control streams and the spatial ground truth to compute against. **(3) Section mode classification**: Since the ground truth sections are not guaranteed to match up to the inferred sections 1:1, we cannot directly use the F1 score. Instead, the metric is the percentage of the inferred value that matches the ground truth.

Our baseline results are evaluated directly against the data, with no additional post-processing algorithms: (i) the trips are segmented by the appropriate *trip transition events*, (ii) sections are segmented whenever the *motion activity* changes, (iii) the mode is set to the new *motion activity*, but (iv) trajectories are unchanged.

Our results (Table 3) vary by sensing configuration and phone OS – e.g. lowering the sensing quality can lower the power drain from 42% to 10% on android and from 10% to 2% on iOS. However, the median values in the table do not fully capture the data complexity. For example, the **median** Δ *section count* is consistently 0, so most ground truthed sections were matched 1:1 to inferred sections. However, a **boxplot** shows many outliers, indicating cases when one ground truth section (e.g. a fast bicycle ride) was broken up into multiple inferred sections (e.g. alternating bicycling and motorized). Using such a section for downstream analysis, e.g. at the *personal* level, to suggest an alternate transportation mode, or at the *structural* level, to determine the modes of transportation to work, is clearly incorrect. The challenge to the AI/ML community is to use post-processing algorithms (e.g. a recent, weakly-supervised approach for distinguishing between motorized modes; Fürst et al., 2020) to eliminate the outliers while supporting more classes of travel and using minimum battery drain.

Table 3: Median values for the raw phone data from the downtown library timeline under various accuracy and frequency sensing settings. The mode labels are the one detected on the phone (walk, bike, vehicle). While the median results are good, outliers, such as for section count (right) can reach up to 10 on both android (top) and iOS (bottom). Lowering them is the primary challenge.

		android			iOS		
metric	goal	h, h	m, h	h, m	h, h	m, h	h, m
battery drain (%)	low	42	30	10	10	2	10
trajectory error (m)	low	6	5	10	6	15	3
Δ trip count	low	0	0	0	0	0.5	1
Δ section count	low	0	0	0	0	0	0
Δ trip start (min)	low	4	5	5	5	5	4
Δ trip end (min)	low	5	30	5	2	1	0
Δ section start (min)	low	2	0	3	1	0.5	2
Δ section end (min)	low	0	1	0	3	5	2
Mode match ratio	high	1	1	0.99	0.9	0.8	0.9



4 Conclusion

We present MobilityNet, the first public, smartphone-based dataset for multi-modal mobility that: (i) includes data from multiple smartphone OSes, (ii) includes detailed ground truth, (iii) addresses power/accuracy tradeoffs, and (iv) preserves privacy thanks to artificial trips. We also define metrics for the segmentation, smoothing and mode inference required to model a mobility diary and implement them in jupyter notebooks, included in our dataset.

This dataset can form the basis of challenges such as WordNet/ImageNet. Such challenges can also motivate similar data collection in other regions, which can improve the scope and generalizability of future challenges. The standardization associated with such challenges can also enable *hybrid* competitions, in which the public dataset acts as a training set, while large-scale datasets which cannot be published since they contain real travel patterns (e.g. from the Transportation Secure Data Center; Holden et al., 2018), can act as test set. Such challenges allow direct comparisons between implementations and can help improve the state-of-the art in this important domain.

Acknowledgments

This work was supported in part by the CONIX Research Center, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. Additionally, support has been received through the research project "Open Source Lab for Sustainable Mobility", initiated and funded by the Volkswagen Sustainability Council.

References

Matthias Bussonnier, Jessica Forde, Jeremy Freeman, Brian Granger, Tim Head, Chris Holdgraf, Kyle Kelley, Gladys Nalvarte, Andrew Osheroff, M Pacer, Yuvi Panda, Fernando Perez, Benjamin Ragan-Kelley, and Carol Willing. Binder 2.0 - Reproducible, interactive, sharable environments for science at scale. In Fatih Akici, David Lippa, Dillon Niederhut, and M Pacer (eds.), *Proceedings of the 17th Python in Science Conference*, pp. 113–120, 2018. doi: 10.25080/Majora-4af1f417-011. URL `https://conference.scipy.org/proceedings/scipy2018/project{_}jupyter.html`.

Claudia Carpineti, Vincenzo Lomonaco, Luca Bedogni, Marco Di Felice, and Luciano Bononi. Custom Dual Transportation Mode Detection by Smartphone Devices Exploiting Sensor Diversity. In *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 367–372. IEEE, 2018. doi: 10.1109/PERCOMW.2018.8480119. URL `https://ieeexplore.ieee.org/abstract/document/8480119`.

Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. The Opportunity Challenge: A Benchmark Database for on-Body Sensor-Based Activity Recognition. *Pattern Recognition Letters*, 34(15):2033–2042, nov 2013. ISSN 01678655. doi: 10.1016/j.patrec.2012.12.014. URL `https://linkinghub.elsevier.com/retrieve/pii/S0167865512004205`.

Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific Reports*, 3, 2013. ISSN 2045-2322. doi: 10.1038/srep01376. URL `http://www.nature.com/doifinder/10.1038/srep01376`.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Miami, FL, 2009. IEEE. ISBN 978-1-4244-3992-8. doi: 10.1109/CVPR.2009.5206848. URL `http://ieeexplore.ieee.org/document/5206848/`.

Jennifer Dill and Nathan McNeil. Four types of cyclists? examination of typology for better understanding of bicycling behavior and potential. *Transportation Research Record: Journal of the Transportation Research Board*, 2387, 2012. ISSN 0361-1981. doi: 10.3141/2387-15. URL `https://journals.sagepub.com/doi/10.3141/2387-15`.

Jonathan Fürst, Mauricio Fadel Argerich, Kalyanaraman Shankari, Gürkan Solmaz, and Bin Cheng. Applying Weak Supervision to Mobile Sensor Data: Experiences with TransportMode Detection. In *AAAI-20 Workshop on Artificial Intelligence of Things*, New York, New York, USA, feb 2020.

Jacob Holden, Harrison Van Til, Eric Wood, Lei Zhu, Jeffrey Gonder, and Matthew Shirk. Trip Energy Estimation Methodology and Model Based on Real-World Driving Data for Green-Routing Applications. *Transportation Research Record*, 2672(24):41–48, 2018. doi: 10.1177/0361198118798286. URL `https://journals.sagepub.com/doi/10.1177/0361198118798286`.

David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D Sherwin, S Karthik Mukkavilli, Konrad P Kording, Carla Gomes, Andrew Y Ng, Demis Hassabis, John C Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling Climate Change with Machine Learning. *arXiv:1906.05433 [cs, stat]*, nov 2019. URL `http://arxiv.org/abs/1906.05433`.

Kalyanaraman Shankari, Mohamed Amine Bouzaghrane, Samuel M Maurer, Paul Waddell, David E Culler, and Randy H Katz. E-Mission: An Open-Source, Smartphone Platform for Collecting Human Travel Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(42):1–12, 2018a. ISSN 0361-1981, 2169-4052. doi: 10.1177/0361198118770167. URL http://journals.sagepub.com/doi/10.1177/0361198118770167.

Kalyanaraman Shankari, Jonathan Fürst, Yawen Wang, Philippe Bonnet, David E Culler, and Randy H Katz. Zephyr: Simple, Ready-to-use Software-based Power Evaluation for Background Sensing Smartphone Applications. Technical report, University of California at Berkeley, 2018b. URL http://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-168.html.

Vijay Srinivasan and Thomas Phan. An Accurate Two-Tier Classifier for Efficient Duty-Cycling of Smartphone Activity Recognition Systems. In *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, pp. 11, Toronto, Ontario, Canada, nov 2012. doi: 10.1145/2389148.2389159. URL http://dl.acm.org/citation.cfm?id=2389159.

Peter R Stopher, Li Shen, Wen Liu, and Asif Ahmed. The Challenge of Obtaining Ground Truth for GPS Processing. *Transportation Research Procedia*, 11:206–217, 2015. ISSN 23521465. doi: 10.1016/j.trpro.2015.12.018. URL https://linkinghub.elsevier.com/retrieve/pii/S2352146515003099.

Lin Wang, Hristijan Gjoreskia, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. Summary of the Sussex-Huawei locomotion-transportation recognition challenge. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 1521–1530, Singapore, Singapore, oct 2018. ACM. doi: 10.1145/3267305.3267519. URL https://dl.acm.org/doi/abs/10.1145/3267305.3267519.

Hui Zang and Jean Bolot. Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, pp. 145–156, Las Vegas, Nevada, USA, 2011. Association for Computing Machinery. doi: 10.1145/2030613.2030630. URL https://dl.acm.org/doi/10.1145/2030613.2030630.

Jesse Zhang, Jack Sullivan, Vasudev Venkatesh P B., Kyle Tse, Andy Yan, John Leyden, Kalyanaraman Shankari, and Randy Katz. TripAware: Emotional and Informational Approaches to Encourage Sustainable Transportation via Mobile Applications. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '19, pp. 183–186, New York, NY, USA, nov 2019. Association for Computing Machinery. ISBN 978-1-4503-7005-9. doi: 10.1145/3360322.3360871. URL https://dl.acm.org/doi/10.1145/3360322.3360871.

Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pp. 312–321, Seoul, South Korea, 2008. Association for Computing Machinery, Association for Computing Machinery. doi: 10.1145/1409635.1409677. URL https://dl.acm.org/doi/abs/10.1145/1409635.1409677.