Using Map Service API for Driving Cycle Detection for Wearable GPS Data

Preprint

Lei Zhu and Jeffrey Gonder
National Renewable Energy Laboratory

To be presented at Transportation Research Board (TRB) 97th Annual Meeting
Washington, DC
January 7-11, 2018

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Conference Paper
NREL/CP-5400-70474
December 2017

Contract No. DE-AC36-08GO28308

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

USING MAP SERVICE API FOR DRIVING CYCLE DETECTION FOR WEARABLE GPS DATA

Lei Zhu, PhD, Corresponding Author

National Renewable Energy Laboratory (NREL)

15013 Denver West Parkway, Golden, CO 80401

Tel: 303-275-3194; Email: lei.zhu@nrel.gov

Jeffrey D. Gonder

National Renewable Energy Laboratory (NREL)

15013 Denver West Parkway, Golden, CO 80401

Tel: 303-275-4462; Email: Jeff.Gonder@nrel.gov

Paper Number: 18-00236
Extended Abstract re-submitted for presentation at the 97th Annual Meeting of the Transportation Research Board
Submission Date: 11/15/2017

ACKNOWLEDGMENTS

The research was developed under the Transportation Secure Data Center (TSDC) project at the National Renewable Energy Laboratory (NREL). Application of the methodology to data in the TSDC was supported by joint funding for the TSDC project provided by the U.S. Department of Transportation and the U.S. Department of Energy. NREL is managed for the U.S. Department of Energy by the Alliance for Sustainable Energy under Contract No. DE-AC36-08GO28308. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

INTRODUCTION

Wearable global positioning system (GPS) data, such as from smartphones or dedicated wearable data acquisition devices, reflect all travel and activities by an individual who wears or pockets the GPS device during the data collection period (*1*). It encompasses an enormous amount of car mode driving cycles (*2, 3*). Typically, wearable GPS data require two post-processing steps for automatic estimation of travel mode: 1) detecting (stationary) activities and (single mode) trips between activity locations (*1, 4*), and 2) identifying unique travel mode trips.

The mode identification step recognizes the modes of each single-mode stage based on the GPS trajectory characteristics (*1, 5-7*) and other information through pattern recognition methods such as machine learning methods (*8-10*), probability methods (*11, 12*), and criterion-based methods (*13, 14*). It is straightforward to distinguish non-motorized modes and motorized modes according to GPS speed profiles because the speed of the non-motorized mode (walk, bike) trip is relatively lower. Within the motorized mode trips (passenger car, bus, etc.), the speed profiles (driving cycles) of the passenger vehicle ("car-mode") GPS trajectories (*15, 16*) are the focus of this research.

It is non-trivial to accurately extract car-mode trips or driving cycles from wearable GPS data flagged as motorized travel. The biggest challenge is to distinguish bus-mode and car-mode trips. Research efforts that have leveraged supplemental reference data include geographic information system (GIS) information to support mode detection (*11, 17, 18*). However, dedicated road GIS reference data (*19-21*) and the associated traffic data (*22*) are not always available for all places, and the quality of the GIS data is not guaranteed.

Routing web services, such as Google Maps Directions application programming interface (API) (*23*), which offers quality GIS route information for any given location, are easily accessed. However, two key challenges must be overcome to use the web services. First, it is not easy to find the car-mode API route best matched to the actual route as drivers do not necessarily follow the top API-identified routes. Second, once the best-matched car-mode API route is found, the feature differences between the two routes must be chosen and measured. The route feature and prediction model selections must maximize car-mode detection accuracy.

To address the challenges of applying an API route for mode detection, a novel driving-cycle detection method using a map service APIs is proposed. The method offers promising mode prediction results. The research makes contributions to both theory and practice. The major contributions include:

(1) The method can apply to any markets or cities without maintaining a costly GIS database. The method directly detects driving or car mode by considering both the actual route and the API route features.

(2) To apply a web service API, the proposed trajectory segmentation algorithm finds a "best-matched" car-mode API route corresponding to the actual route.

(3) A logistic regression-based travel mode detection model is built by the selected route features, which provides an accurate prediction of probability and is flexible for applications with various accuracy requirements.

METHODOLOGY

The proposed car-mode detection method detects car-mode trips from the broader set of motorized mode trips, including car and bus. The major modules of this approach include 1) cleaning and smoothing the trip trajectory data (*7, 24*), 2) a trajectory segmentation algorithm, and 3) a logistic regression model.

A trajectory segmentation algorithm guarantees a best-matched API car-mode route will be found for the actual route. The flow chart of the algorithm is illustrated in Figure 1. It is a recursive procedure, which keeps separating trajectory and scoring path similarity. The algorithm ends when the trajectory segmentation scheme does not change through consecutive iterations. Each trip is divided into segments by the algorithm, and each segment satisfies the longest common subsequence-based similarity score criterion (*24*)—that for an API call using the same origin and destination, a topologically similar API sub-path exists. All API sub-paths constitute an API route corresponding to the entire trajectory (i.e., the actual route). After that, the actual route and the API route features (i.e., similarity score, distance, and speed) with ground truth travel-mode data are used to develop a logistic regression model (*25*) to estimate the trip mode.

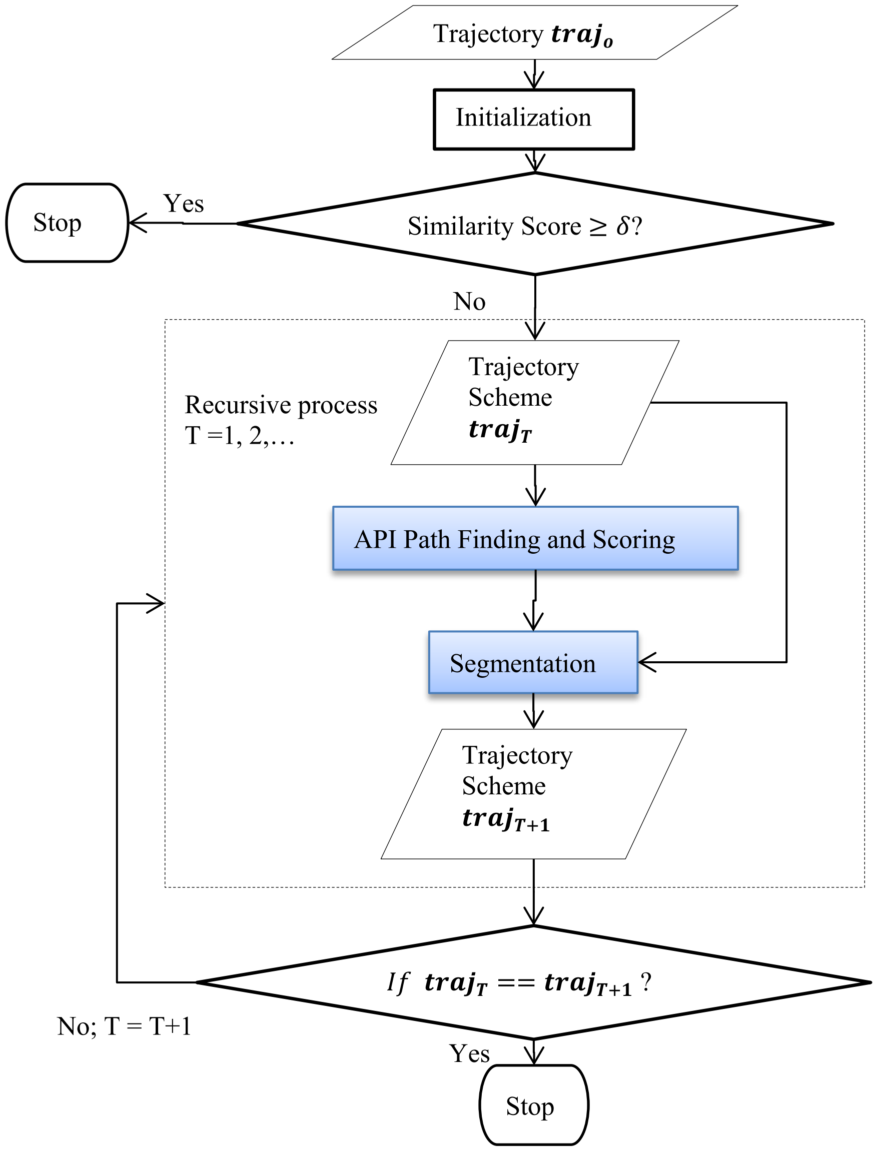

Figure 1 Flow chart of trajectory segmentation algorithm.

The Caltrans wearable GPS data used in this study were collected during 2010–2012. The data are accessible from the National Renewable Energy Laboratory's Transportation Secure Data Center (*26, 27*). In the data set, the travel mode to each GPS data point is given and considered as the ground truth.

Google API inputs such as origin and destination locations are directly obtained from the actual route start and end locations. The trajectory segmentation algorithm helps to find the "best-matched" API route from all API returned routes for each actual route. The Google API provides the "route duration in traffic" when a future departure time is assigned. In addition, the API route distance and duration in traffic are extracted, and the route polylines are decoded as a link endpoint coordinate sequence and are fed into the longest common subsequence-based similarity score calculation procedure to obtain the similarity score (denoted as $score$) of the API route. The API route distance $dist_{API}$ and duration are directly procured from the matched

API route. Thus, the API route average speed, spd_{API}, can be computed as the route distance divided by the route duration. The actual route distance, $dist_{actual,}$ and average speed, $spd_{actual,}$ are directly obtained from the actual route trajectory. Thus, the maximum distance ratio of the actual and API route, $ratio_{dist}$, (distance ratio) and the maximum speed ratio of the actual and API route, $ratio_{spd}$, (speed ratio) are calculated as well.

Therefore, the five input variables of the logistic regression model are $score$, $ratio_{dist}$, $dist_{actual}$, $ratio_{spd}$, spd_{actual}. The logistic regression classifier provides the prediction probabilities of two dependent variables, car mode or non-car mode for an actual route.

FINDINGS

The precision accuracy and recall accuracy measure the model accuracy. *Precision accuracy* is the number of correctly detected car (or non-car) trip segments divided by the total number of estimated car (or non-car) mode trip segments. *Recall accuracy* is the number of correctly detected car (or non-car) trip segments divided by the total number of ground truth car (or non-car) mode segments. Table 1 illustrates that the overall accuracy of the mode estimation is about 89%. For car-mode detection, the precision accuracy is about 90%, and the recall accuracy rate is roughly 95%.

A prediction performance comparison between the raw GPS data-based fuzzy logic model (*12*) and the proposed method on the same data set is also illustrated. The comparison shows the proposed method significantly outperforms the raw GPS data-based fuzzy logic method.

Table 1 Comparison of Mode Detection Accuracy Performance

	Proposed method		Fuzzy logic	
Mode	Precision	Recall	Precision	Recall
Car	90.35%	94.93%	71.75%	85.81%
Non-car	86.73%	76.56%	40.00%	21.88%
Total	89.39%	-	66.51%	-

Since the logistic regression model provides a car-mode detection probability for each trip, it gives an opportunity to further refine car-mode detection and to boost the accuracy by analyzing the car-mode detection probability values. The cumulative precision accuracy at threshold p is defined as the ratio of the number of correctly estimated car-mode trips within the total number of car-mode estimated trips under the condition that car-mode probability values for all estimated trips are greater than p. The cumulative precision accuracy against the probability threshold of car mode is plotted in Figure 2. At the high-probability portion (left-hand side), the curve fluctuates dramatically because of the small total number of car-mode estimates with high probabilities, and the cumulative precision accuracy value is sensitive to the total number of those trips. As the probability threshold value of car mode is reduced, the cumulative precision accuracy decreases.

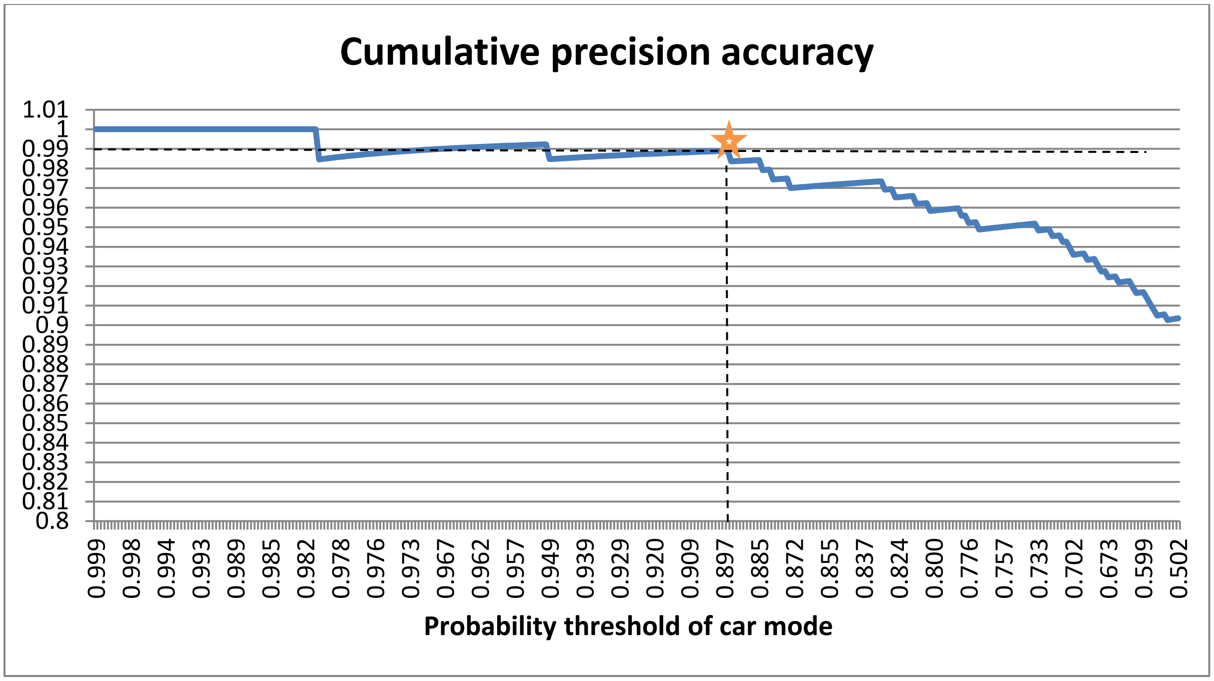

Figure 2 Car mode cumulative precision accuracy.

In Figure 2, when the probability is greater than 0.897, the precision accuracy reaches 99%. Based on that, the trips are categorized as a high-probability group (probability value > 0.897) and a low-probability group (0.5 < probability value ≤ 0.897). Table 2 illustrates the detail accuracy performance result for high and low probability groups. The high-probability group has a cumulative precision accuracy of 99%, while the cumulative precision accuracy of the low probability group is 78.46%. Most of the route attribute variables follow the car-mode estimation observations introduced previously.

Table 2 Cumulative Precision Accuracy Result for High and Low Probability Groups

	High probability (>0.897)	Low probability (≤0.897)
# of ground-truth car-mode trips	179	102
# of ground-truth non-car mode trips	2	28
# of total trips	181	130
Ratio of group trips	58.2%	41.8%
Precision accuracy	99%	78.46%
Avg. Score	0.813	0.751
Avg. distance ratio	1.039	1.091
Avg. actual distance (mile)	11.68	7.128
Avg. speed ratio	1.158	1.218
Avg. actual speed (mph)	37.173	22.127

CONCLUSION

By using a ubiquitous and easily accessible map service API, the proposed driving-mode detection method uses reliable API routing information to accurately detect driving travel modes and derive driving cycles. The features of the API route and the actual route are used for developing a logistic regression classifier to predict the trip mode with high probability. The proposed trajectory segmentation algorithm finding a matched car-mode API route for the actual route is the key to leveraging the map service API.

The numerical experiment results demonstrate that the proposed driving-cycle detection method is accurate and promising. The overall mode detection accuracy rate reaches about 89%. The correct detection rate of car-mode trips reaches about 95%, and the detection precision accuracy is about 90%. Those significantly outperform the estimation results of a fuzzy logic method. Furthermore, a cumulative precision accuracy curve method is proposed for various driving-mode detection applications to help determine the best probability threshold value.

In addition to drive-cycle detection, the proposed car mode trip or drive-cycle detection method can also be applied to other travel modes (bus, rail, etc.) to improve detection accuracy due to the flexibility of the map service API approach to provide route information on other modes.

REFERENCES

1. Hu, X., Y.-C. Chiu, Y.-L. Ma, and L. Zhu. Studying Driving Risk Factors Using Multi-Source Mobile Computing Data. *International Journal of Transportation Science and Technology*,. Vol. 4, No. 3, 2015, pp. 295–312. https://doi.org/10.1260/2046-0430.4.3.295

2. Hu, X., Y.-C. Chiu, and L. Zhu. Behavior Insights for an Incentive-Based Active Demand Management Platform. *International Journal of Transportation Science and Technology*, Vol. 4, No. 2, 2015, pp. 119–133. https://doi.org/10.1260/2046-0430.4.2.119

3. Holden, J., E.W. Wood, L. Zhu, and J.D. Gonder. Development of a Trip Energy Estimation Model Using Real-World Global Positioning System Driving Data: Preprint. 2017, National Renewable Energy Laboratory, Golden, CO (United States). https://www.nrel.gov/docs/fy17osti/69121.pdf

4. Schüssler, N. Accounting for Similarities Between Alternatives in Discrete Choice Models based on High-Resolution Observations of Transport Behaviour. 2010, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 19093, 2010.

5. Wang, M., X. Bao, L. Zhu, and Y.L. Bao. A Map-Matching Method Using Intersection-Based Parallelogram Criterion. *Advanced Materials Research*, Vols. 403–408, 2012, pp. 2746–2750.

6. Zhu, L., Y. Bao, S.-G. Wang, Q. Zhou, and X. Bao. Map-Matching Compatible with Junction Adjusting in Vehicle Navigation System. In *Recent Advances in Computer Science and Information Engineering*. Springer, 2012, pp. 451–457.

7. Zhu, L., J. Gonder, and L. Lin. Prediction of Individual Social-Demographic Role based on Travel Behavior Variability using Long-Term GPS Data. *Journal of Advanced Transportation*, Vol. 2017, 2017. https://doi.org/10.1155/2017/7290248

8. Moiseeva, A. and H. Timmermans. Imputing Relevant Information from Multi-Day GPS Tracers for Retail Planning and Management using Data Fusion and Context-Sensitive Learning. *Journal of Retailing and Consumer Services*, Vol. 17, No. 3, 2010, pp. 189–199. DOI: 10.1016/j.jretconser.2010.03.011

9. Zong, F., Y. Bai, X. Wang, Y. Yuan, and Y. He. Identifying Travel Mode with GPS Data Using Support Vector Machines and Genetic Algorithm. *Information*, Vol. 6, No. 2, 2015, pp. 212–227. DOI: 10.3390/info6020212

10. Zhang, L., S. Dalyot, D. Eggert, and M. Sester. Multi-Stage Approach to Travel-Mode Segmentation and Classification of GPS Traces. In *Proceedings of the ISPRS Guilin 2011 Workshop on International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Guilin, China. 2011.

11. Tsui, S., and A. Shalaby. Enhanced System for Link and Mode Identification for Personal Travel Surveys based on Global Positioning Systems. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1972, 2006, pp. 38–45. https://doi.org/10.3141/1972-07

12. Schuessler, N. and K. Axhausen. Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2105, 2009, pp. 28–36. https://doi.org/10.3141/2105-04

13. Stopher, P.R., Q. Jiang, and C. FitzGerald. Processing GPS Data from Travel Surveys. *2nd International Colloquium on the Behavioural Foundations of Integrated Land-Use and Transportation Models: Frameworks, Models and Applications*, Toronto, 2005.

14. Chen, C., H. Gong, C. Lawson, and E. Bialostozky. Evaluating the Feasibility of a Passive Travel Survey Collection in a Complex Urban Environment: Lessons Learned from the New York City Case Study. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 10, 2010, pp. 830–840. https://doi.org/10.1016/j.tra.2010.08.004
15. Zhu, L., J. Holden, E. Wood, and J. Gonder. Green Routing Fuel Saving Opportunity Assessment: A Case Study using Large-Scale Real-World Travel Data. in *Intelligent Vehicles Symposium (IV), 2017 IEEE*. 2017. IEEE.
16. Chen, Y., L. Zhu, J. Gonder, S. Young, and K. Walkowicz. Data-Driven Fuel Consumption Estimation: A Multivariate Adaptive Regression Spline Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 83, 2017, pp. 134–145. https://doi.org/10.1016/j.trc.2017.08.003
17. Stenneth, L., O. Wolfson, P.S. Yu, and B. Xu. Transportation Mode Detection using Mobile Phones and GIS Information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2011. ACM.
18. Gong, H., C. Chen, E. Bialostozky, and C.T. Lawson, A GPS/GIS Method for Travel Mode Detection in New York City, *Computers, Environment and Urban Systems*, Vol. 36, No. 2, 2012, pp. 131–139. https://doi.org/10.1016/j.compenvurbsys.2011.05.003
19. Zhu, L., Y.-C. Chiu, and Y. Chen. Road Network Abstraction Approach for Traffic Analysis: Framework and Numerical Analysis. *IET Intelligent Transport Systems,* Vol. 11, No. 7, 2017, pp. 424–430. DOI: 10.1049/iet-its.2016.0287
20. Zhu, L., and Y.-C. Chiu. Transportation Routing Map Abstraction Approach: Algorithm and Numerical Analysis, *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2528, 2015, pp. 78–85. https://doi.org/10.3141/2528-09
21. Zhu, L. *Routing Map Topology Analysis and Application*. 2014, The University of Arizona.
22. Yuanlu, B.L.Z.L.B., Traffic Broadcasting Digital Map Service System with Multi-Correspondence. *Electronic Technology*, Vol. 9, 2009, pp. 017.
23. Svennerberg, G. *Beginning Google Maps API 3*. 2010: Apress.
24. Zhu, L., J.R. Holden, and J.D. Gonder. A Trajectory Segmentation Map-Matching Approach for Large-Scale, High-Resolution GPS Data. In *Transportation Research Board 96th Annual Meeting*. 2017.
25. Peng, C.-Y.J. and T.-S.H. So. Logistic Regression Analysis and Reporting: A Primer. *Understanding Statistics*, Vol. 1, No. 1, 2002, pp. 31–70. http://dx.doi.org/10.1207/S15328031US0101_04
26. Transportation Secure Data Center, National Renewable Energy Laboratory. 2005.
27. Gonder, J., E. Burton, and E. Murakami. Archiving Data from New Survey Technologies: Enabling Research with High-Precision Data While Preserving Participant Privacy. *Transportation Research Procedia,* Vol. 11, 2015 pp. 85–97. https://doi.org/10.1016/j.trpro.2015.12.008