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
NOTICE

The submitted manuscript has been offered by an employee of the Alliance for Sustainable Energy, LLC (Alliance), a contractor of the US Government under Contract No. DE-AC36-08GO28308. Accordingly, the US Government and Alliance retain a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

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

Available electronically at SciTech Connect http://www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI http://www.osti.gov
Phone:  865.576.8401
Fax:  865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS http://www.ntis.gov
Phone:  800.553.6847 or 703.605.6000
Fax:   703.605.6900
Email: orders@ntis.gov

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.
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 $\text{dist}_{API}$ 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**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Prop. method</th>
<th>Fuzzy logic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>Car</td>
<td>90.35%</td>
<td>94.93%</td>
</tr>
<tr>
<td>Non-car</td>
<td>86.73%</td>
<td>76.56%</td>
</tr>
<tr>
<td>Total</td>
<td>89.39%</td>
<td>-</td>
</tr>
</tbody>
</table>

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.
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

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


