Assessment of GPS and Self-reported Data in the Collection of Travel Activities

Xiaomeng Xu, Pedram Fard
Jeffrey M. Casello, & Andrea Mikkila
School of Planning, University of Waterloo

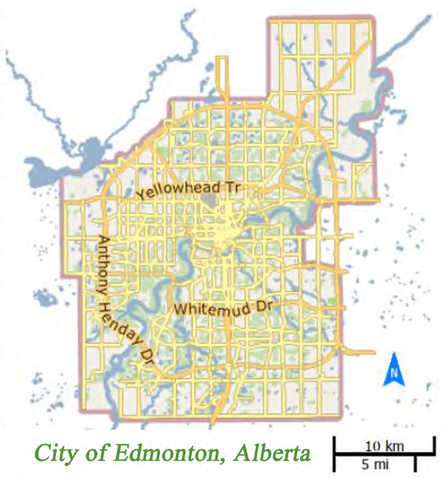
WPTI
WATERLOOPUBLICTRANSPORTATIONINITIATIVE

Introduction

We were asked to work with a consulting firm in the City of Edmonton, Alberta, to develop a passive data collection method to help inform the development of a regional travel forecasting model. This research reports on the methods and outcomes of that work.

This study answers three major research questions:

1. Is GPS data an effective way to validate the self-reported trip activities when building a travel forecasting model? a) Can we build an automated algorithm to detect activities? b) What's the relationship between the GPS detected trips, self-reported trips, and the true number of trips? c) Under what condition does the GPS work well?
2. What is the rate of underreporting of trips?
3. What GPS attributes can be helpful in identifying activity stops?

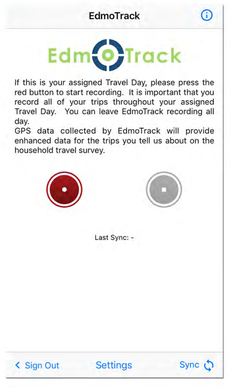


Data

Gathered data using a Waterloo developed smartphone App called: EdmoTrack.

51 Users' GPS data (Over 156,000 points) collected by smartphone	
GPS_Survey_ID	Anonymous participant identifier
Point_ID	Individual point identifier
TimeStamp	Date and time for each GPS point collected
GPSLat	Latitude value for each GPS point
GPSLog	Longitude value for each GPS point
Bearing	Bearing in degrees East of true North
GPS_Accuracy	Distance (m) by which the true location differs from reported location with 90% confidence.

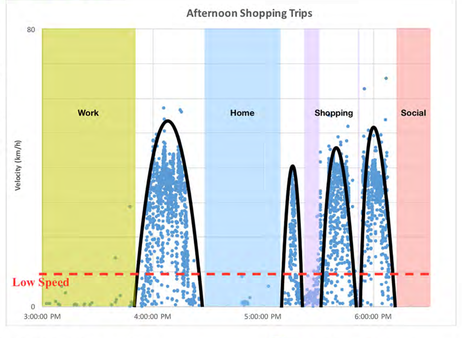
51 Users' self-reported data collected by survey on 1 day	
GPS_Survey_ID	Anonymous participant identifier
SegOrigin	Trip origin address
SegDestination	Trip destination address
SegDepartHrMin	Trip depart time
SegArriveHrMin	Trip arrive time
SegMode_desc	Trip travel mode



Methods

Activity Detection

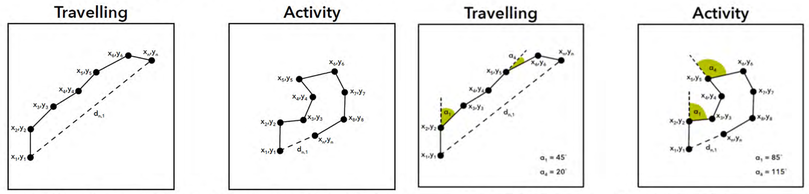
1. Low Speed: We consider a series of points below the speed threshold as a stop which may potentially be an activity.



2.Circuitous Trajectories: The (in)directness of travel can be measured by a) the ratio of cumulative travel distance to Euclidean distance between start point and end point; b) the cumulative deviation in bearing.

a. $\frac{\sum_{n=2}^{n} Distance(x_{n-1},y_{n-1};\, x_n,y_n)}{EuclideanDistance\ (x_1,y_1;\, x_n,y_n)}$ Close to 1 represents traveling

b. $\sum_{n=2}^{n} angle(x_{n-1},y_{n-1};\ x_n,y_n)$ Close to 0 represents traveling

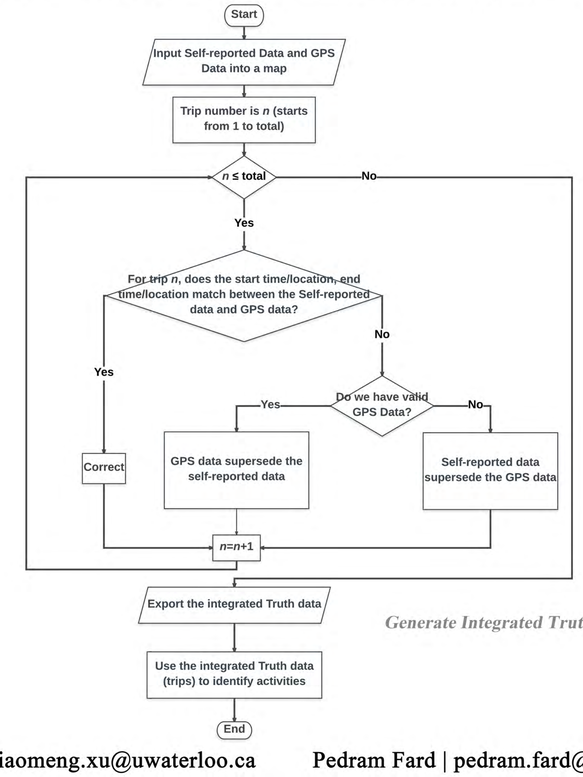


a: ratio of cumulative distance to Euclidean distance b: cumulative deviation in bearning

3. Distance from the roadway network: The approach is to determine if the cluster of points occur in a location that would suggest that travel is occurring at an intersection or at a transit stop.

Integrated Truth Data

If the survey response and the GPS data both indicate an activity, then this activity has occurred. When the automated GPS data indicates an activity but none is recorded on the survey, then the GPS data are investigated more closely, to either verify an unreported activity, or to confirm no activity took place. When the survey response indicates an activity, but the automated GPS script does not identify an activity, then further investigation is made. If the GPS data are believed to be valid, then this will be removed from the participant's travel record.



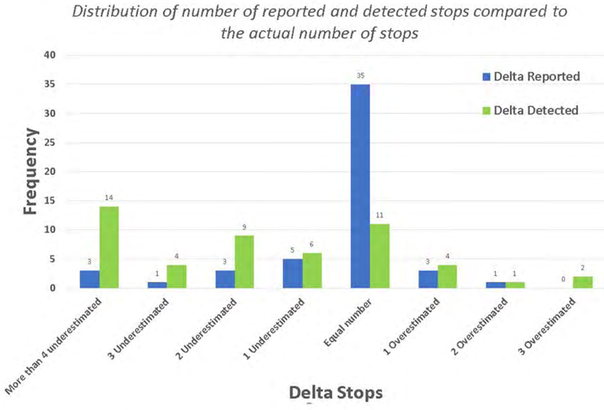
Generate Integrated Truth Data

Summary of Results

There are three types of errors identified among 51 users. 13.1% of the errors were caused by self-reported data, 37.7% from activity detection. 49.2% of the errors were caused by GPS data. However, 70% of GPS errors were caused by smartphone app or GPS signal.

Source of Errors	Number
Self-reported Errors	**Total 8**
User forgot to report trips	7
Misreported by users (i.e. nearby locations)	1
Activity Detection Errors	**Total 23**
Both methods identified a stop-location, but algorithms unable to detect (e.g. Gas station)	16
Local walking trips	7
GPS Errors	**Total 30**
Lost GPS signal or smartphone app crashed	11
Self-reported, but not collected by smartphone: app turned-off	10
Infrequent GPS reporting	4
GPS data duplicated	3
Extra stop recorded when user waited too long at transfer point	1
GPS emitting erroneous trips	1

Distribution of number of reported and detected stops compared to the actual number of stops



Percentage of Underestimated and Overestimated Data

Self-Reported Data		GPS Data	
Underreported Users	12 (23.53%)	Underestimated Users	33 (64.71%)
Overreported Users	4 (7.84%)	Overestimated Users	7 (13.73%)
Correct	35 (68.6%)	Correct	11 (21.57%)
Total	51 (100.00%)	Total	51 (100.00%)

Conclusion

The integrated truth data indicated that about 24% of respondents underreported their trips. Only 8% of respondents overreported their trips. The integrated data set also interprets that 65% of GPS data underestimate trips, 14% overestimate. In this test case, the GPS data were only effective in those instances where the users was extremely attentive to maintaining good records. The circumstances under which the Activity Detection algorithm works well include: 1. Attentive users and consistent GPS data; 2. No entry into tunnels; 3. Travel by modes other than LRT; 4. When travel does not include "drive-through".

There are four essential GPS attributes that can be helpful in identifying activities. 1. The velocity as a function of time; 2. The ratio of cumulative travel distance to Euclidean distance; 3. The cumulative deviation in bearing 4. The distance from roads.

UNIVERSITY OF WATERLOO
FACULTY OF ENVIRONMENT
School of Planning

Contacts

Xiaomeng Xu | xiaomeng.xu@uwaterloo.ca
Pedram Fard | pedram.fard@uwaterloo.ca
Professor Jeff Casello | jcasello@uwaterloo.ca
Andrea Mikkila | aemikkil@gmail.com