NICR
NATIONAL INSTITUTE FOR CONGESTION REDUCTION

Evaluating the Accuracy of Smartphone-Based Travel Behavior Data

Sean J. Barbeau (PI-USF) [1], Shawn Turner (PI-TTI) [3], Ipek N. Sener (Co-PI TTI) [3], Michael Maness (Co-PI USF) [2], Wilson Lozano[1], Trang Luong[2]

[1]Center for Urban Transportation Research and [2]Civil and Environmental Engineering at the University of South Florida

[3]Texas A&M Transportation Institute

For more information, contact: Sean Barbeau
Email: barbeau@usf.edu

POLICY BRIEF

BACKGROUND AND OBJECTIVES

Multimodal transportation such as transit, bike, walk, ride-hailing (e.g., Uber, Lyft), carshare, and bikeshare are vital to supporting livable communities. To build safe and effective multimodal infrastructure, Departments of Transportation (DOTs), Metropolitan Planning Organizations (MPOs), and transit agencies need quality data about how the public is currently traveling via these modes. However, current data collection techniques for multimodal travel behavior have significant limitations, including limited duration (e.g., due to expense and/or lack of ongoing incentive to participants), user fatigue, opaqueness in data pre-processing and aggregation, and usually focus on single modes without knowledge of how other modes connect.

In this project, the research team evaluated the accuracy and precision of activity transition time and location data gathered using the open-source OneBusAway (OBA) mobile transit application. Activity transitions are generated from the Android Activity Transition Application Programming Interface (AATAPI). It is important to emphasize that the OBA application itself is not detecting activity transitions or classifying the activity type—this functionality is performed by the mobile device hardware and software. Therefore, the data collected within OBA should be identical to that collected by any other app using the AATAPI. By examining this data, we can therefore have a more transparent view into the raw data that other apps may be using to produce travel behavior information.

METHODOLOGY

The accuracy and precision were determined by comparing collected OBA data with ground truth information manually recorded by the research team. Trip trajectories with established start and stop locations (identified using aerial imagery in Google Earth) were planned before data collection so that research team members could avoid large uncertainties in location latitude and longitudes. Research team members carried several mobile devices in cities in Texas (6 simultaneous devices) and Florida (5 simultaneous devices) on 286 trips along these trajectories to collect mode transition data via the OBA app, resulting in 1568 activity observations across all devices. When stopping at each location, the research team member manually noted the time (e.g., by snapping a picture with their phone camera that includes a timestamp) and entered their ground truth trip information into a spreadsheet. Vehicle, bicycle, and walk trips were captured in College Station while data collection in Tampa included these modes plus transit and scooters.

The research team created several open-source software tools (CUTR @ USF 2021a and CUTR @ USF 2021b) to extract and match the collected activity transition data with the recorded ground truth data. These automatically matched records were then manually reviewed by the researchers to correct any match errors. The resulting data was then used to calculate the difference between the ground truth trip start and end times and the activity transition start and end times (temporal error), as well as the distance between the ground truth start and end locations and the activity transition start and end locations (spatial error).

 DOI: doi.org/10.5038/CUTR-NICR-Y1-3-1

RESEARCH FINDINGS

The accuracy and precision analysis indicated several findings. The median values for distance and time error for all modes across all devices for detected trips was 47 meters and 59 seconds, although distance error was larger at the origin than at the destination (a median value of 72 meters vs. 30 meters). The primary reason for the greater distance error at the origin is that the user is in motion during the time delay that is typically observed between when a participant starts a new activity and when the device acquires a location. Based on these results, unsurprisingly WALKING had the best performance—a median error of 32 meters in Tampa and 39 meters in College Station. However, ON_BICYCLE in College Station had a larger median distance error than IN_VEHICLE (66 meters vs. 60 meters, respectively). In Tampa ON_BICYCLE had a smaller median error than IN_VEHICLE, although perhaps not as high as expected (50 meters vs. 72 meters). TRANSIT had the highest median distance error of all modes at 110 meters; this may be due to the AATAPI having a hard time recognizing transitions to and from the activity, as the AATAPI doesn't natively support detection of transit (or scooter) trips.

Overall, around 19% of the observable activities in College Station and around 39% of observable activities in Tampa were missed (i.e., a false negative), leading to an overall false negative rate of 29%. Trips taken in Tampa that were intentionally planned in urban canyon areas (i.e., downtown) and consisted of less dwell time between trips (i.e., STILL activities) seemed to result in higher false negative rates for all devices than trips taken in non-urban canyon areas and with longer duration STILL activities in College Station. In other words, all devices seemed to struggle more in Tampa than College Station to track these trips. More false positive activity transitions occurred during trips in Tampa—49% of transit trips, 31% of e-scooter trips, and 10% of bicycle trips. One device model, the Samsung Galaxy S9, was a large contributor to these false negatives and was an outlier compared to other devices, especially for simpler trips with longer dwell times such as those taken in College Station. College Station false negative rate drops from 19% to 11% if the Galaxy S9 data is removed. Additionally, the Samsung devices were noisier and less accurate when detecting activity transitions when compared to the Google Pixel devices. In automated processing of the data the outliers will need to be addressed. However, across all device models the median values of time and distance error for observed activities was surprisingly relatively similar. Median values for distance error are between 25 meters and 75 meters, and median values for time errors across all devices are between 31 seconds and 82 seconds.

The accuracy of mode classification for trips was very high, in particular for 1:1 matches between the ground truth and collected data trips. For these trips the mode accuracy was 100% for IN_VEHICLE, 100% for transit, 99% and 100% for WALKING in Tampa and College Station respectively, 100% and 94% for ON_BICYCLE in Tampa and College Station respectively, and 100% for ON_SCOOTER.

POLICY AND PRACTICE RECOMMENDATIONS

The research team conducted semi-structured expert interviews with stakeholders (including researchers, practitioners, and policy makers) who have used other sources of GNSS data as part of their studies. Stakeholders had varying expectations and requirements for spatial and temporal accuracy depending on the exact application of the data (e.g., trip counts for origin-destination analysis, traffic counts for link-level analysis, path detection for route choice analysis, traffic counts for intersection-level analysis, mode detection for multimodal user analysis).

The data from the AATAPI seems best suited for trip counts for origin-destination analysis and mode detection for multimodal user analysis and meets the time and distance accuracy and precision requirements stated by the stakeholders for these categories. The interviews indicated that while the stakeholders see value in activity transition data such as that collected by OneBusAway, further data cleaning and mode inference (e.g., for public transit modes) is necessary for the data to be readily usable by stakeholders. Generally, however, all stakeholders welcomed the addition of more data, especially to supplement existing datasets.