DYNAMIC MOBILITY APPLICATIONS PROGRAM

SmarTrAC: A Smartphone Solution for Context-Aware Travel and Activity Capturing

Final report

Yingling Fan
Humphrey School of Public Affairs
University of Minnesota

Julian Wolfson
School of Public Health
University of Minnesota

Gediminas Adomavicius
Carlson School of Management
University of Minnesota

Kirti Vardhan Das
Humphrey School of Public Affairs
University of Minnesota

Yash Khandelwal
Department of Computer Science and Engineering
University of Minnesota

Jie Kang
Department of Computer Science and Engineering
University of Minnesota

February 2015

Research completed under contract with the Volpe Center at the U.S. Department of Transportation in support of the Intelligent Transportation Systems Joint Program Office

Acknowledgments

We wish to acknowledge those who made this research possible. The research was completed under contract with the Volpe Center at the U.S. Department of Transportation in support of the Intelligent Transportation Systems Joint Program Office. The research was developed on the basis of an earlier pilot study that was funded by the Center for Transportation Studies at the University of Minnesota.

CONTENTS

Keywords: smartphones; mobility data; travel survey; travel diary; experience sampling; accelerometer; GPS; mode detection; location-aware applications; destination identification; trip purpose.

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

The use of mobile phones in collecting travel behavior data has rapidly increased, especially after GPS tracking technology became widely available in commercial smartphones. Existing smartphone-based tools in the field have generally focused on capturing the “when”, “where”, and “how” of travel, i.e., using the smartphone’s automatic sensing functionality to detect travel mode and to collect position and route data. Although locations and modes of transportation derived from sensing data represent important travel behavior information, travel behavior has many other important dimensions—such as trip purpose, travel experience, and travel companionship (i.e., the “why”, “how”, and “who” of travel)—all of which are critical for understanding people’s travel choices. Some of these dimensions may be inferable from pure sensory data, but reliable inference will generally require long-term use data from a very large number of subjects. Other dimensions are simply inaccessible to passive sensing tools. In contrast, traditional travel diary methods and some first-generation smartphone-based travel survey tools enable the collection of multi-dimensional data through high-intensity sampling and qualitative survey techniques. However, these methods are often burdensome to study subjects and impractical for use in a diverse, mobile, and increasingly time-stressed population.

This project develops SmarTrAC, a user-friendly, open-source Android smartphone application which addresses the limitations of both the passive sensing and traditional travel survey tools. SmarTrAC has two major functionalities:

- SmarTrAC combines smartphone sensing with advanced statistical and machine learning techniques to automatically detect, identify, and summarize attributes of daily activity and travel episodes. Capitalizing on the smartphone’s computing capability, SmarTrAC will process the raw sensor data using rules derived from statistical and machine learning techniques, segment and partition time series into activity and travel episodes, and summarize attributes of the segmented activity and travel episodes based upon sensor data. Compared to traditional travel diary methods, SmarTrAC has a much lower respondent burden because it allows automatic data collection on activity and trip attributes such as activity/trip duration, travel mode, and activity type (trip purpose).
- SmarTrAC incorporates survey techniques to allow users to view and provide contextual information on the identified activity and travel episodes at their convenience on a daily basis. Capitalizing on the smartphone’s communication capability, SmarTrAC asks users to provide additional information on the automatically identified activity and travel episodes, thereby enabling the collection of more detailed contextual data on each activity and travel episode. SmarTrAC captures many more dimensions of travel behavior data than the existing passive sensing tools that focus on location and route tracking.

By bringing together automatic sensing, surveying, and statistical machine learning seamlessly, SmarTrAC yields travel data of a breadth and depth not available by using either smartphone sensory data or traditional travel diary methods alone, providing a simple, efficient, low-cost approach to collecting detailed, multimodal, and multi-dimensional travel data.

The research team has tested the functionalities of SmarTrAC through a series of laboratory tests and through two rounds of seven-day field tests in which a total of 17 real-world Android phone users participated. Test data indicates good performance of SmarTrAC. In particular, field test results suggest that SmarTrAC has a reasonable battery consumption rate, a moderate data storage/transmission requirement, a high accuracy in identifying activity vs. trip episodes and in classifying the travel modes of each trip episode, and a medium-high accuracy in classifying the types of activity episodes. More specifically:

- With SmarTrAC running continuously, 74% of the phones had a battery life longer than 6 hours, and about half of the phones (47%) had a battery life longer than 8 hours.
- SmarTrAC produces 50 megabytes of data per day, requiring 350 megabytes of data storage space for a seven-day participation. The associated weekly data transfer needs are roughly 150 megabytes after data compression.
- SmarTrAC has an overall accuracy of 90% in identifying activity vs. trip episodes and an overall accuracy of 96% in classifying motorized vs. non-motorized trip segments. The overall accuracy in classifying travel mode across all six mode options (car, bus, rail, wait, bike, and walk) is 86%.
- SmarTrAC has good overall accuracy in identifying activity type (trip purpose). For activities taking place at previously identified locations, activity type prediction accuracy is > 95%. When the activity is occurring at a new location not stored in the SmarTrAC database, the predicted activity matches the true activity type 25-40% of the time. From a user experience perspective, it may be acceptable if the correct activity type is among the top two or three most probable predicted activity types, and our results show that the correct activity type is among the top two most probable predicted activity types 70-80% of the time, and among the top three 80-95% of the time.

This final report describes SmarTrAC research work between June 15, 2013 and February 15, 2015. The remainder of this report is organized as follows:

Chapter 2 – Review of Related Methods, Applications, and Studies. This chapter includes reviews of (1) existing activity-travel data collection methods, (2) relevant smartphone applications, and (3) studies that have used sensing data to derive meaningful activity-travel information.

Chapter 3 – SmarTrAC Architecture and Core Components. This chapter describes key features of SmarTrAC, how these key features set SmarTrAC apart from existing activity-travel data collection methods, as well as SmarTrAC overall software architecture and core algorithms.

Chapter 4 – Laboratory Tests. This chapter describes a series of pre-designed feature inspection tests conducted by the research team to evaluate specific features of SmarTrAC. Implications of the test results are discussed.

Chapter 5 – Field Tests. This chapter describes two rounds of seven-day field tests conducted by 17 real-world Android smartphone users. Implications of the test results are discussed.

Chapter 6 – Conclusions and Future Directions. This chapter offers study conclusions with a lookout to future directions.

CHAPTER 2. REVIEW OF RELATED METHODS, APPLICATIONS, AND STUDIES

SmarTrAC differs substantially from the existing activity-travel data collection methods. In this chapter, we first review the methodological evolution in activity-travel data collection, offering insight into the increasing popularity of smartphone-based methods in the field. We further focus on recent developments in smartphone-based methods and detail the differences between SmarTrAC and other smartphone-based tools for activity-travel data collection. Finally, we review existing studies that have used sensing data to derive meaningful activity-travel information. These studies are highly relevant to the development of the core algorithms used in SmarTrAC for real-time data processing.

2.1 Methodological Evolution in Activity-Travel Data Collection

Figure 1 illustrates the methodological evolution in the field of activity-travel data collection. The earliest method to collect data on daily activities and trips is paper- or phone-based recall survey, which is prone to recall bias. Researchers later developed the diary survey approach with the intention to reduce recall bias. More recently, technological advances, such as mobile devices, Global Positioning Systems (GPS), and smartphones, have energized the use of GPS to record location changes and travel routes as well as the use of mobile devices and smartphones to collect self-reported data. More specifically, the existing smartphone technology on activity/travel data collection has enabled the collection of both sensor data and user input data (either real-time at the moment or at convenience) in one single device.

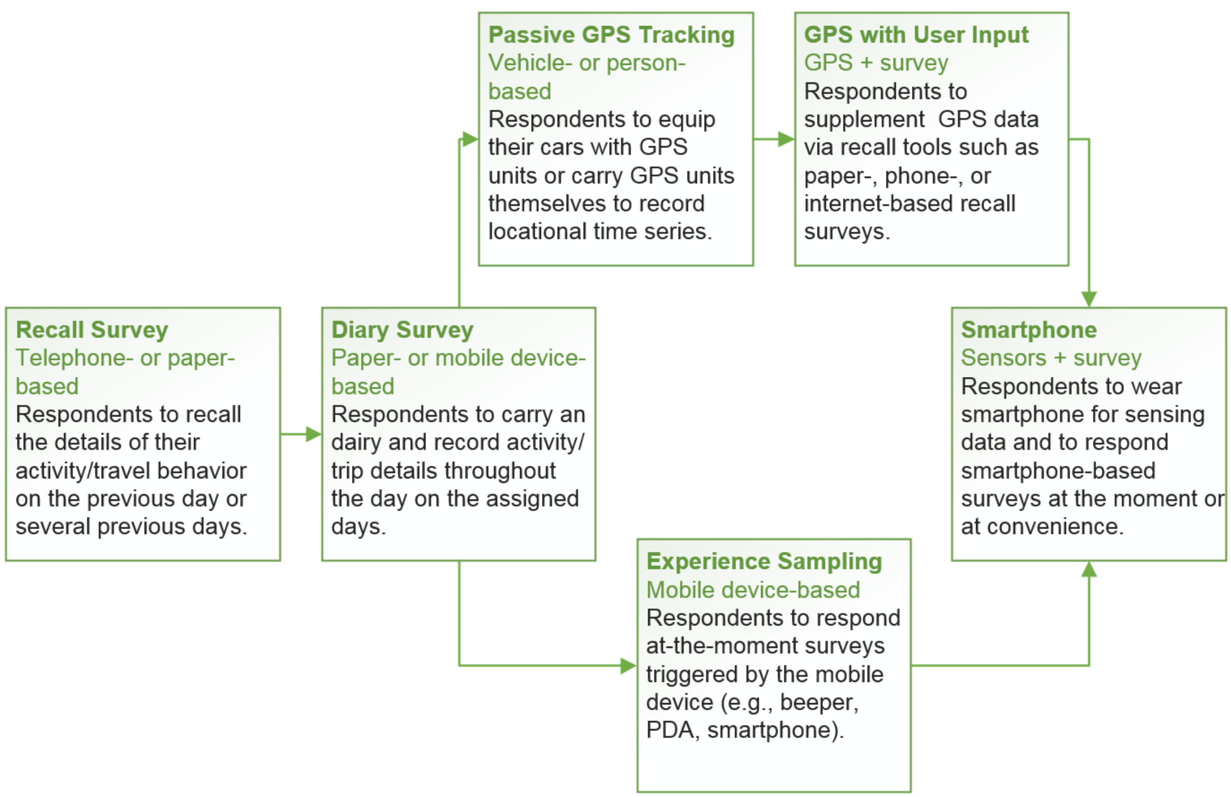


Figure 1. Methodological evolution in activity-travel data collection

Recall Survey: The recall survey approach, which is often phone- or paper-based, asks respondents to recall characteristics of their activity/travel behavior (e.g., frequency, duration, distance) on one or more days. Such activity/travel data are prone to recall bias and are typically low-resolution because respondents often do not memorize precise activity/travel details such as activity start/end time, trip arrival/departure time, and travel distances and routes (Bohte & Maat, 2009; Ettema et al., 1996).

Diary Survey: In the late 1990s, there was a shift from recall-based surveys to activity/travel diaries that require respondents to carry an activity/travel diary on assigned days and record activity/trip details throughout the day. While diaries do reduce recall bias when used correctly, many participants find that taking detailed notes on each activity/trip throughout the day is burdensome. Consequently, participants often postpone filling diaries, which leads to significant recall bias in the diary-based data collection efforts (Bohte & Maat, 2009; Schlich & Axhausen, 2003; Schönfelder et al., 2002).

Passive GPS Tracking: To reduce the burden placed on respondents in traditional recall/diary surveys, Global Positioning Systems (GPS) have gained popularity in recent activity/travel data collection efforts. GPS units have the advantage of providing accurate second-by-second data on location and velocity, which enables automatic collection of precise trip origin and destination data, trip route data, and trip start and end time data. Initially, GPS tracking of travel behavior was vehicle-based: cars were equipped with GPS units to record information when the vehicle was running. Such technology has shortcomings because it excludes non-motorized travel modes and does not provide reliable information on the real origins or destinations (origins and destinations tracked by vehicle-based GPS are often garages or parking lots) (Du & Aultman-Hall, 2007; Guensler et al., 2006). These shortcomings and improved technology (more compact GPS units) have led to the use of individual-based GPS units to collect travel behavior data across multiple modes. However, extracting meaningful travel behavior information from raw GPS data (which is in the format of locational time series) require intense, off-line data processing efforts that have to be carried out separately from GPS data collection (Schuessler & Axhausen, 2009). The processing efforts often include GPS data filtering, trip detection, and map matching. For example, speed and location change thresholds are often used to distinguish trip segments from non-trip (stationary) segments and to distinguish motorized trip segment from non-motorized trip segments. External land use and points of interest data are often used in conjunction with GPS data to predict activity location types and trip purposes (McGowen & McNally, 2007; Srinivasan et al., 2006).

GPS Tracking with User Input: No matter how sophisticated the classification rules or prediction models are in interpreting raw data collected from passive GPS tracking, results from classification and prediction inevitably contain errors and inaccuracies. Moreover, activity/travel behavior has many other important dimensions—such as experience and companionship—that may be impossible to infer from pure GPS data. To address this issue, researchers have begun supplementing GPS data with user input. This method primarily involves combining GPS data with recall or diary tools (these tools could be paper-, phone-, or web-based) (Gonget al., 2012; Lee-Gosselin et al., 2006; Li & Shalaby, 2008). The GPS data, combined with user input data, provide much richer and more accurate information on activities and trips throughout the day than data from using passive GPS tracking alone or using traditional survey/diary method alone.

Experience Sampling: Activity/experience sampling is another technique that has been developed to minimize recall bias and improve accuracy in activity and travel behavior data. This method focuses on getting real-time activity/travel information from respondents. Specifically, activity/experience sampling involves the use of a mobile signaling device —a pager, a PDA, or a cell phone—to prompt participants to respond to both open- and close-ended questions at several random time points throughout the day. The questions often include queries on what the respondent is currently doing, the physical and social contexts within which the current activity/travel occurs, and how people are actually experiencing the current activity/travel. Survey triggering of this kind can yield accurate information about activity/travel episodes at the moment in real time (Hektner et al., 2007). But experience sampling also comes with two major disadvantages: (1) Because activity/travel information is collected at random time points of the day, it does not provide continuous information on activity and travel behavior throughout the day without intensive post-estimation efforts; (2) it is often inconvenient and intrusive for study subjects as they are prompted to complete surveys several times a day (including times that are inconvenient for study subjects to respond).

Smartphone: The smartphone technology is more advantageous than the GPS with user input technology because smartphones come with a built-in GPS receiver and a convenient user interface to gather user input. Besides GPS, smartphones have other built in system sensors such as accelerometers, magnetometers, gyroscopes etc. It is easy for smartphones to seamlessly combine multiple data collection modules (GPS data, other sensory data, and user input data) into one single device (Nitsche et al., 2014; Wan & Lin, 2013). The activity/experience sampling technology can also be easily implemented in smartphones as smartphones are mobile and are capable of signaling the respondents and getting real-time input from the respondents. The smartphone-based methods have enabled the collection of both sensor data and user input data on daily activities and trips (either real-time at the moment or at convenience) in one single device.

2.2 Existing Smartphone-Based Tools in Comparison to SmarTrAC

Table 1 summarizes existing smartphone applications for activity-travel data collection. Despite the advantage of hosting sensor data collection and user input data collection in a single device, the existing smartphone applications have not been designed to allow much interaction between sensor data and user input data. As shown in Table 1, CycleTracks and CONNECT mainly use sensor data for displaying trip routes only. In the case of TRAC-IT and the Quantified Traveler, although they mine sensor data to extract additional trip characteristics such as travel mode and trip purpose, the mining is done remotely through servers and results from server-based data mining cannot be used in real time to facilitate user input. While Moves allows local data mining in real time to facilitate user input, its mining capability is limited to identifying non-motorized, active travel modes. In addition, Moves does not use user input data to optimize its data mining module. To give a specific example, if a user tags an unknown location using a place name from Foursquare API, Moves does not automatically identify that location the next time the user visits there but still show the location as an unknown location.

Table 1. Existing smartphone applications on activity-travel data collection

Name-Developer	Description	Comments
CycleTracks - San Francisco County Transportation Authority	CycleTracks allows active trip logging for users to register bicycle trips, initiate GPS tracking, and specify trip purpose of each bicycle trip. *Source:* *http://www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android*	GPS tracking with simple survey functions. Users initiate and discontinue tracking of each bicycle trip, and provide information on trip purpose. Sensor data are used to display trip routes.
CONNECT (previously MOVE) - Ghent University, Belgium	CONNECT allows active trip logging for users to register trips, initiate GPS and accelerometer tracking, and specify trip characteristics. It also allows automatic triggering of specific surveys based upon user input, e.g., surveys focused on bicycle trips, on shopping trips, or when passing by a certain location. *Source:* (Vlassenroot et al., 2014)	GPS and accelerometer tracking with advanced survey functions. Users initiate and discontinue tracking of each trip, and provide information on trip specifics that could be further used to trigger additional surveys. Sensor data are used to display trip routes.
TRAC-IT – University of South Florida	TRACT-IT logs GPS data continuously, transmits GPS data to a remote server for data processing, retrieves travel advisory feedback from the server for the user, and allows users to register trips and specify trip characteristics that cannot be extracted from GPS data. *Source:* (Winters et al., 2008)	GPS tracking with simple survey function, incorporated with server-based post-processing of GPS data. GPS tracking is continuous in the background and users provide limited trip specifics. GPS data are used to display trip route as well as mined at a remote server to provide travel advisory feedback.
Quantified Traveler – University of California Berkeley	The Quantified Traveler logs GPS and accelerometer data continuously, allows baseline survey of the user's travel habits, transmits the sensor data to a server in the cloud for post processing, and periodically requests the user to access travel advisory feedback through a website that displays post-processing results from the server. *Source:* (Jariyasunant et al., 2014)	GPS and accelerometer tracking with simple survey function, incorporated with cloud-based post-processing of sensor data. Sensor tracking is continuous in the background and users provide limited information on travel habits. Sensor data are mined in the cloud to provide travel advisory feedback.
Moves – ProtoGeo (Helsinki and London-based startup, recently purchased by Facebook)	Moves logs GPS and accelerometer data continuously, and automatically identifies activities, trips, as well as whether walking, cycling, and running modes are used in any of the trips. The app also allows users to tag activity location using three location types (home, work, or school) or place names available from Foursquare API. In addition, Moves has a pedometer function that counts the daily number of steps that users take and calculates daily calorie burn. *Source:* *https://www.moves-app.com/*	GPS and accelerometer tracking with simple survey function, incorporated with limited real-time processing of sensor data. Sensor tracking is continuous in the background and user provide limited information on travel mode and activity location. Sensor data are locally mined to provide instant information on travel mode, especially active travel modes such walking, cycling, and running.
UbiActive – University of Minnesota	UbiActive logs GPS and accelerometer data continuously and automatically identifies trips that are longer than 10 minutes. The app also allows users to add additional trip details (such as secondary activities, companionship information, and travel experience) immediately after each trip completion. The app also provides daily summary on the amount of physical activity associated with daily trips. *Source:* (Fan et al., 2012)	GPS and accelerometer tracking with simple survey functions, incorporated with limited real-time processing of sensor data. Sensor tracking is continuous in the background. Sensor data are locally mined to predict the beginning and completion of each trip that is at least 10 minutes long.

SmarTrAC: SmarTrAC is a technological innovation in the field of activity and travel behavior data collection. It allows sensor data and user input data to interact in an iterative manner:

- Sensor data are analyzed and processed locally on the phone and in real time to extract meaningful activity/travel information. The extracted information further reduces respondent burden and serves as an initial information basis to probe more detailed and more accurate information on each activity/travel episode from the user.
- The probed user input data in turn optimizes how sensor data is analyzed and processed so that there is increased accuracy over time in the information extracted from sensor data.

Such interaction between sensor data and user input data allowed by SmarTrAC forms feedback loops to perfect the sensor data processing procedure and the process of capturing user input in SmarTrAC over time.

2.3 Existing Studies on Deriving Activity-Travel Information from Sensing Data

Existing studies that derive activity-travel information from sensing data (especially location sensing data) involve significant processing due to the large amounts of data produced by GPS units/loggers or smartphones over time. Based on the current literature (Flamm & Kaufmann, 2007; Gong et al., 2012; Schuessler & Axhausen, 2009; Tsui & Shalaby, 2006; Wan & Lin, 2013), Figure 2 provides an overview of the steps employed by research studies when using GPS sensing data to identify activity-travel characteristics. In this section of the report, we will go over the various methods utilized by existing studies. These methods formed the foundation for developing SmarTrAC's real-time data processing algorithms.

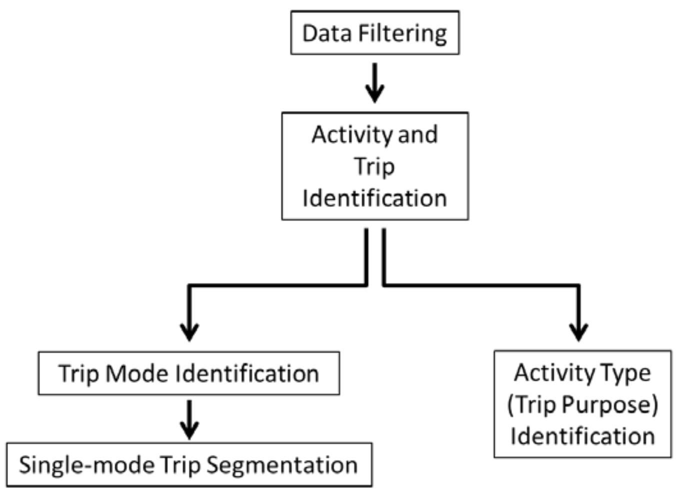


Figure 2. GPS travel data analysis summary

2.3.1 Data filtering

While GPS provides the most accurate passively collected travel data, it does have some limitations, such as redundant or poor quality data points that need to be dealt with in order to improve data quality and minimize analysis time. The most common issues encountered with GPS data collection include: (1) the 'urban canyon effect' caused by tall buildings that cause satellite signals to bounce and give inaccurate location information (Gong et al., 2012; Wu et al.,2011); (2) signal loss, which occurs when a GPS device is out of the range of satellites (e.g., in buildings, tunnels, etc.) and does not collect any readings (Biljecki et al., 2013); (3) warm starts, which occur when an out of range GPS device comes back into range and takes time (up to 2 mins) to recalibrate its location resulting in inaccurate mode and trip purpose prediction (Gong et al., 2012), especially in the case of short trips in dense urban environments (Flamm & Kaufmann, 2007). Other GPS data issues vary by the geographic and transportation network characteristics of the places the studies are conducted. For example, Gong et al. (2012) report issues in New York due to underground mode changes between subway and commuter rails as well as issues due to the overlap caused by subway lines running below roads, making the complexity of using GPS points for mode detection much higher.

Given these issues with GPS data it often requires filtering before it can be used for any analysis. Most of the studies use speed, distance, and time-based thresholds (individually or in combination) to filter bad data. For example, Wu et al. (2011) change the speeds of points with recorded speeds greater than 200km/h to zero. Wan & Lin (2013) remove points with speeds above 160km/h and points that deviate from the preceding point by at least 180m over a time of 5 seconds. In addition to using speed, Auld et al. (2009) look for invalid points by looking at 4 before and after points for each GPS point and identifying time gaps greater than 15 seconds. Tsui & Shalaby (2006) and Gong et al. (2012) add to time and speed-based thresholds by filtering data based on a number of satellites used to estimate GPS location and the value of the Horizontal Dilution of Precision (HDOP), which indicates the dispersion of satellites in view. Tsui & Shalaby (2006) remove points that are recorded by less than 3 satellites and have a HDOP higher than 5 while Gong et al. (2012) remove points based on at least 4 satellites being present and HDOP values higher than 4. Schuessler & Axhausen (2009) include altitude as a filter by removing points with an altitude less than 200m or greater than 4,200m. Table 2 summarizes the various data filters used in the reviewed studies.

Table 2. Data filters used in reviewed literature

Author, year	Data filter used					
	Speed	Heading	Distance	Time	Satellites	Altitude
Wan & Lin, 2013	X		X			
Schuessler & Axhausen, 2009	X					X
Wu, et al., 2011	X					
Zhang et al., 2011	X	X				
Auld et al., 2009	X			X		
Gong et al, 2012	X		X	X	X	
Wolf et al., 2004			X	X		
Flamm & Kaufmann, 2007			X			
Bohte & Maat, 2009	X		X	X		
Tsui & Shalaby, 2006	X	X			X	

In addition to filtering out poor quality location points, many studies address the issue of signal loss during data filtering. Wan & Lin (2013) adjust speed of points by: determining signal loss during an activity if the difference between two points is more than 10 minutes and distance is less than 200m, and signal loss across activities if difference between two points is more than 10 minutes and distance is more than 200m. Gong et al. (2012) define signal loss as the difference between two points is more than 120 seconds and distance is more than 250m. Other studies simply use a time threshold to determine signal loss. For example, Zhang et al. (2011) identify signal loss as whenever signal is lost for more than 20 seconds and Biljecki et al. (2013) identify it at 30s.

Finally, some studies also apply data smoothing techniques to further remove any noise and reduce errors. Schuessler & Axhausen (2009) use Gauss Kernel smoothing technique to smooth position (x and y coordinates of GPS points). Liao et al. (2007) use conditional random fields (CRF) cliques (measurement, consistency, and smoothness cliques) to associate GPS traces with street patches (Liao et al., 2007). Zhang et al. (2011) reduce speed errors by averaging its neighborhood by using a 5-second moving average smoothing for speed data. Nitsche et al. (2014) pre-process GPS tracks with a Kalman filter which computes accurate and smooth trajectories by combining the raw GPS and cell location data with predictions of a linear motion.

2.3.2 Activity and trip identification

Once the GPS-based time series data has been filtered and cleaned the next step of the analysis is to break the time series down into activity (dwelling activity) and trip segments. Typically, a minimum dwell time threshold (no or limited movement for a given time) is used in combination with a distance threshold (in some studies) to classify data streams into activities or trips connecting activities. Wolf et al. (2001) use a dwell time of 120 seconds to identify locations. Flamm & Kaufmann (2007) use a dwell time of 90 seconds and a maximum spatial divergence of 40m. Wolf et al. (2004) use time to identify dwelling activity. In particular, if dwelling time is greater than 5 minutes, they consider it a confident trip-end; however, if the dwelling time is less than 2 minutes, they consider it a suspicious trip-end, which is further evaluated and classified using path circuitry (the distance covered by vehicle over Euclidian distance between start and end points). Bohte & Maat (2009) assume that a dwelling is a location that the GPS indicates you have been at for over 3 minutes. Tsui & Shalaby (2006) identify activity locations by using different rules for signal loss and non-signal loss situations. When there is no signal loss, GPS records that have zero speeds for over 120s are identified as dwelling location. If the signal loss period is between 120 to 600 seconds and the Euclidian distance traveled is less than 50m it is considered an activity location. Auld et al. (2009) use a rule-based method where if a person is within a threshold distance (determined by block size) for at least a threshold amount of time (determined by travel speed) the average of the points is used as the activity location (Auld et al., 2009).

More complex methods for dwelling activity classification often involve change point analysis algorithms or spatial analysis algorithms (Flamm & Kaufmann, 2007). Wan & Lin (2013) identify dwelling activity using change point detection by flagging places where there is a speed switch of around 10 km/h or any points with *dt*[1] greater than 10 minutes and *dd*[2] less than 200m. They go further by trying to get more accurate location information by treating activities with signal loss and no signal loss. For actives without signal loss they use a modified kernel density method to identify activity locations, and for activities with signal loss they use the location of the last point before time gap as the location of the activity. Schuessler & Axhausen (2009)

[1] *dt*: time difference between the preceding point and the current point.
[2] *dd*: distance difference between the preceding point and the current point.

identify dwelling activities by flagging an activity if the speed is less than 0.01m/s for at least 120 seconds and flagging an activity based on GPS point density. Wu et al. (2011) use both a rule-based method and a random forest classification method (collection of many decision tree models) to identify dwelling activities. In the rule-based method, points within a minimum of one minute with speed lower than 3km/h are considered as a static cluster, and a line-detection process based upon "distance difference" of three sequential points is used to remove points that form linear alignments. The random forest classification of dwelling points is based upon speed, acceleration, distance difference, and distance ratio (distance between first and last sequential points in a series over the sum of the distance of all line segments formed by sequential points in the series).

Studies also use variants of K-means clustering algorithms that take a point and using a certain radius calculate the means of all the points within the radius and continue to do so till the mean of the points stop changing. Zhou et al. (2007) use a cluster-based approach developed by reviewing k-means, time, and density based clustering. In their model, a dwelling region must hold a minimum number of points within a certain distance threshold (Zhou et al., 2007). Gong et al. (2012) also use a clustering approach by identifying points within 50m of each other for more than 200 seconds.

2.3.3 Trip mode identification

While there are studies that analyze a single mode of transportation, we focus on studies that have a broader research agenda and emphasize on differentiating between different modes of transportation as they are more relevant to our project. A model that has received significant attention in the recent years is the fuzzy logic model (mathematical models based on researcher/ user-based rules intended to permit degrees of imprecision in reasoning and knowledge). Tsui & Shalaby (2006) used fuzzy logic mode detection algorithms to test mode identification using only GPS, and a combination of GPS and GIS data. They used 58 person-days of GPS travel data for the analysis. Based on the characteristics of the GPS data they used variables within segments to identify modes. These variables included average speed, 95th percentile max speed, positive median acceleration, and data quality (number of valid GPS points in a segment / number of total GPS points in a segment). Finally, they formulated seventeen rules for inference (Tsui, 2005) and used the system to differentiate between walk, cycle, bus and automobile modes. While the total accuracy of mode prediction was close to 91% for both models, the addition of GIS data improved the bus prediction rate through route information (Tsui & Shalaby, 2006). Biljecki et al. (2013) use a fuzzy expert system to derive certainty factors for each mode (certainty factors are a quantification of confidence in the expert's conclusion based on the evidence). Indicators used included three speed variables, five average proximities for transportation mode infrastructure (e.g., bus lines, metro lines, etc.) and the location of trajectories with respect to water surfaces. They combined two datasets from Europe to get a 17 million point dataset. They were able to differentiate between 10 travel modes with an accuracy of almost 92%. Schuessler & Axhausen (2009) use a fuzzy engine for mode prediction based on three fuzzy variables, the median of speed distribution, and the 95th percentiles of the speed and acceleration distributions. After analyzing the available modes and characteristics of the GPS data, membership functions for the modes were devised. Then the likelihood of each mode was calculated based on all mode scores, and finally the reasonability of the derived mode chains was investigated by looking at the possibility of mode change. Their data included 4,882 participants who carried a GPS device for an average of 6.65 days. The results were compared to distributions from another dataset (the Swiss travel micro census) and deemed satisfactory but no validation of mode detection was conducted.

Other models used for mode prediction include cluster-based models, machine-learning methods (Random Forest Decision Trees), and sequence models (Conditional Random Fields and Discrete Hidden Markov Models). Wu et al. (2011) use both a rule-based model that developed rules based on time, speed, and spatial location and a random forest decision tree model to differentiate between walking and in-vehicle travel. They used a total of 152 person-day GPS data from two studies. The rule-based algorithm used time, speed, and spatial location from GPS points by first identifying static clusters, then identifying sequential points which represent periods of movement and finally refine the movement classification using speed to differentiate between walking and in-vehicle travel periods. The random forest decision tree was based on maximum speed in 4 minutes, maximum speed in 60 minutes, median speed in 30 minutes, maximum distance difference in 6 minutes and maximum distance difference in 30 minutes. They did not find much difference between the two models. However, the results for in-vehicle travel were predicted with more precision in both models when compared to walking.

In another study, Zhang et al. (2011) use a 2-stage mode identification system. For their analysis they used 125 GPS traces collected in Hannover City, Germany. In the first stage, speed, acceleration, heading, and second order polynomial coefficients are used to differentiate between walking, biking, and motorized vehicles with classification certainty of 94% and higher. In the second stage, they use a machine learning method, Support Vector Machines (SVM), based on eleven parameters (means and standard deviations of maximum speed, average speed, average acceleration, travel time, acceleration, and ratio of stop time with respect to travel time) to differentiate between motorized modes i.e. car, bus, tram, and train with classification certainty ranging from 78% for trams to 100% for trains and cars (Zhang et al., 2011). Zheng et al. (2008) used an approach that consisted of three parts: a change point-based segmentation method, an inference model (Conditional Random Fields), and a post-processing algorithm based on conditional probability. Their data consisted of GPS readings from 45 users across 15 cities over a period of 6 months. Their findings indicated that, of the four different models tested, Decision Trees outperformed Bayesian Net, Support Vector Machine (SVM), and Conditional Random Fields (CRF) in detecting modes and transitions between them (Zheng et al., 2008). Nitsche et al. (2014) used the ability of smartphones to get GPS and accelerometer data (extracting 77 features) and combined the information to differentiate between eight different transportation modes: walk, bike, motorcycle, car, bus, electric tramway, metro, and train. Their data consisted of 355 hours of probe travel data collected by 15 volunteers over 2 months. Using probabilistic classifiers and a Discrete Hidden Markov Model, they achieved classification results ranging from 65% (train, subway) to 95% (bicycle) (Nitsche et al., 2014).

Finally, some studies simply create their own rule-based mode detection algorithms by incorporating GIS data. Bohte & Maat (2009) used data from 1,104 respondents collected over one week and reported lower accuracies for mode detection (car – 75 percent; rail – 35 percent; bicycle – 72; and walk – 68 percent) when they used a model based simply on average and maximum speed (e.g., IF average trip_speed < 10 km/h AND max trip_speed < 14 km/h THEN set modality = 'foot'). They also used GIS data to separate rail from car mode. Gong et al. (2012) combined GPS data with GIS data to develop methods to predict travel modes. Their data consisted of GPS readings for 175 person travel days. The mode prediction model first identifies walk segments based on speed and time features, which are double-checked using similarity index where each consecutive pair of GPS points in walk segments are linked to the most similar link of street network. Next, subway and commuter rail are identified using distance features and trip segments (e.g., distance from each point of trip segment to nearest subway or commuter rail link < 60 m). Finally, car and bus segments are identified using speed, acceleration, and bus stop location data. Using the GIS-based mode prediction algorithm and data from two combined GPS and travel diary surveys, they were able to achieve a successful prediction rate of 82% for walk car, bus, subway, and commuter rail (Gong et al., 2012).

2.3.4 Single-mode trip segmentation

Biljecki et al. (2013) point out that most studies on mode identification focus on classification and omit the segmentation problem. These studies tend to presume a single mode of transportation for a trip that may lead to wrong classification, as many people use multiple modes during a single trip. Therefore, segmentation is essential in predicting travel modes for trips where a person uses more than one mode of transportation. Moreover, segmentation is also essential to deal with the issue of in-trip signal loss and noise which may be present even after data filtering (Biljecki et al., 2013).

A number of studies base their trip segmentation on the assumption that walking is required for every mode change. These studies look at the GPS points of a trip and try to identify walking segments (based on speed, time, distance etc.) within a trip which are then used to divide the trip into segments with potential mode changes. Schuessler & Axhausen (2009) use a walking speed threshold 2.78 m/s and a minimum duration of 60s to identify walk segments while all other modes have a minimum duration of 120s. Zhang et al. (2011) also segment trips by identifying stops and mode change points by using location non-changes (for 5 consecutive seconds, distance < 5meters), speed values (for 5 consecutive seconds, speed < 0.5 m/s), and heading changes (for 5 consecutive seconds, heading changes > 100 decimal degrees). Biljecki et al. (2013) use stops to segment trips based on identifying stops where consecutive points in an interval of 12 seconds do not have a speed higher than 2 km/h. Tsui & Shalaby (2006) segment trips using mode transfer points based on walking information such as end-of-walk, start-of walk, and end-of-gap. Gong et al. (2012) identify single-mode trip segments based on first and last points of signal loss gap or of a walk segment greater than 60 seconds. They identify a walking segment if: first point if speed $\geqslant$ 1.6 km/h and < 10 km/h; speed of each subsequent point $\leqslant$ 15 km/h; duration > 60 s; 85th percentile of speed of all points $\leqslant$ 10 km/h and; average speed of all points $\leqslant$ 6 km/h. Flamm & Kaufmann (2007) use a different approach that identifies significant changes in speed, by analyzing speeds 15 minutes earlier and 15 minutes later (for a given point) to use as segment breaks rather than specifically identifying walking (Gong et al., 2012).

One of the problems encountered while doing segmentation is signal loss, as lack of data may result in misclassification of trips, e.g., an undetected mode change between two segments. To deal with missing data due to signal loss, many studies develop and use methods for joining trip segments and consolidating activities. Wan & Lin (2013) combine segments if the ending point of a low speed segment and the starting point of the next low speed segment have a time span under 60 seconds. Schuessler & Axhausen (2009) merge activities if a new activity is found to start shortly after the last one has ended. For time gaps where average speed was higher than a walking threshold (2m/s) and neighboring segments were identified as non-walking, the segments are joined into one non-walking segment. Wu et al. (2011) combined moving points if a single unidentified point had two adjacent points (before and after) that were moving. Gong et al. (2012) deal with the issue of misidentified segments by looking at three consecutive identified segments. For example, if data suggest 3 segments: car-gap-car they assume that the gap is likely to be caused by a tunnel and join the three into one car segment. Zhang et al. (2011), combine signal loss segments based on their neighboring segments and rules that include, for example: no travel mode should last under 120 seconds, and stops between 2 segments should be more than 120 seconds.

2.3.5 Activity type (trip purpose) identification

In the existing literature the level of detail while looking at the activity type (more general, e.g., indoor/outdoor, home/neighborhood) or trip purpose (more specific, e.g., work, education, pick-up/drop-off, etc.) detection vary significantly and play a key role in determining the models used for prediction. We focus on studies that aim to achieve a greater degree of detail and try and identify specific activity locations (e.g., gas station, restaurant, etc.) and purposes (e.g., shopping, pickup/drop off, etc.) as they are more relevant to our analysis. These studies rely on a combination of rule-based models with land uses/points-of-interest at locations (Wolf et al., 2004; Wolf et al., 2001) or on a combination of rule-based models, land uses/points-of-interest, and personal information (home and work locations, vehicle ownership, etc.) (Bohte & Maat, 2009; Shen & Stopher, 2013; Stopher et al., 2008). Most of these studies use GIS to integrate and analyze land use and point-of-interest data which involves collection of GPS travel data and linking it to GIS land use information to identify trip purposes (Wolf et al., 2001).

Wolf et al. (2001) conducted a study in Atlanta, GA where the top three trip purposes were first identified for each land use description using the 1990 Atlanta regional household survey. The GPS data was then assigned a purpose by factoring in land use (and associated purpose), previous trips purpose, arrival time and activity duration. Using a data set of 156 trips, they were able to assign trips correctly for 79% of the trips. However, it is important to point out that about 26% of those identified trips were for land uses that were pooled in as one (i.e., all land use classes with potentially mixed uses or ambiguous uses were pooled into one group). Despite the challenges, the study highlighted the potential of GPS-derived travel data to replace conventional travel diaries for trip purpose identification. A similar study carried out using travel data from three Swedish cities found that post-processing of passively collected travel data through a Trip Identification and Analysis System (TIAS) successfully provided promising trip purpose information required for travel behavior analysis (Wolf et al., 2004). Their data set consisted of 49,667 vehicle days and 240,435 trips. They use a straightforward technique for identification purposes. For example, if a point cluster center is close to or identical to a driver's home location (less than 200m), the purpose is identified as home or return home. For other locations, they use points of interest (POI) data (e.g., restaurants, gas stations, etc.) along with local land use data. They use a buffer of 300m around point cluster centers; if only one POI is found in the buffer then it is stored, if more than one POI is identified, probabilities are calculated for each purpose type with higher heuristic weights given based on proximity. A polygon analysis for land use is conducted in a similar way. Next the structural and temporal nature of trips was compared to national survey data and, based on temporal and socio-demographic attributes, the most probable purpose was associated with the activity (based on heuristics). In the final step the purpose is selected based on the rules that took into account clustering, information from national survey data and probability predictions based on land use, POI, and temporal characteristics comparisons (Wolf et al., 2004).

In their study, Bohte & Maat (2009) use GIS map POI information to assign trip purposes based on trip-ends. Some of the rules used were: if a trip-end is within a radius of 50m from a known location, it is assumed the location is being visited, and if more than one POI lies within 50m, the closest POI was assumed to be the destination. Work and home information is collected from respondents beforehand and, if trip end is located less than 100m from either location, a purpose is assigned accordingly. For larger locations like train stations and shopping centers, if the trip endpoint is within a polygon that contained train stations or shopping centers, then trip purposes were assigned accordingly. The authors then compared the purpose shares in their data set and the Dutch Travel Survey and found that the trip purpose shares in the two were almost equal.

While this GPS-GIS approach has proved successful in identifying trip purposes to some extent, the authors of the studies do point out the importance of having accurate and well-documented GIS land use databases to carry out these studies. Wolf et al. (2001) reported facing numerous problems while working with GIS data from Atlanta, GA which required them to combine data from various sources and carry out significant data processing in order to get a workable dataset (Wolf et al., 2001).

2.3.6 Summary

We found that the selection of data filtration techniques depends greatly on the detail of the analysis to be conducted and the quality of GPS data being used. Speed, distance, and time were the most commonly used data filters. For activity/ trip identification the use of speed, time, and spatial density were amongst the most common methods used. For trip segmentation, the use of walking segments to identify mode transitions was most popular. While studies used different methods for mode detection, only Zheng et al. (2008) compared the detection capabilities of different models and concluded that decision tree models outperformed others. Trip purpose detection was found to rely heavily on the availability of accurate geographic information (land use or points of interest data), the quality of which varies from place to place. For future research endeavors, we have the ability to rely on mobile application programming interface (API) services for point-of-interest data. Examples of popular APIs include Foursquare[3] and Google[4].

[3] https://developer.foursquare.com/
[4] https://developers.google.com/places/

These services provide a ‘points of interest’ repositories which are constantly updated and provide location details that were not available in the past. In conclusion, the knowledge base generated from past research and the availability of new technologies has made it possible to significantly improve the mode and trip purpose detection models in SmarTrAC without increasing the burden on users.

CHAPTER 3. SMARTRAC ARCHITECTURE AND CORE COMPONENTS

This chapter first describes the software architecture of SmarTrAC. Following the description, we introduce three unique functions of SmarTrAC and how these functions enable the collection of activity and travel behavior data in more advanced ways than existing activity/travel data collection technology. Lastly, this chapter provides technical details on how the SmarTrAC Sensor Data Processor (SDP) and the SmarTrAC User Interface (UI) are implemented. The SmarTrAC SDP and UI are the two most important and innovative components of SmarTrAC.

3.1 Software Architecture

Figure 3 illustrates the high-level architecture of SmarTrAC. It decomposes SmarTrAC into logical software and data components. Major components include the Sensor Data Capturer (SDC), the SDP, the UI, and the Main Database.

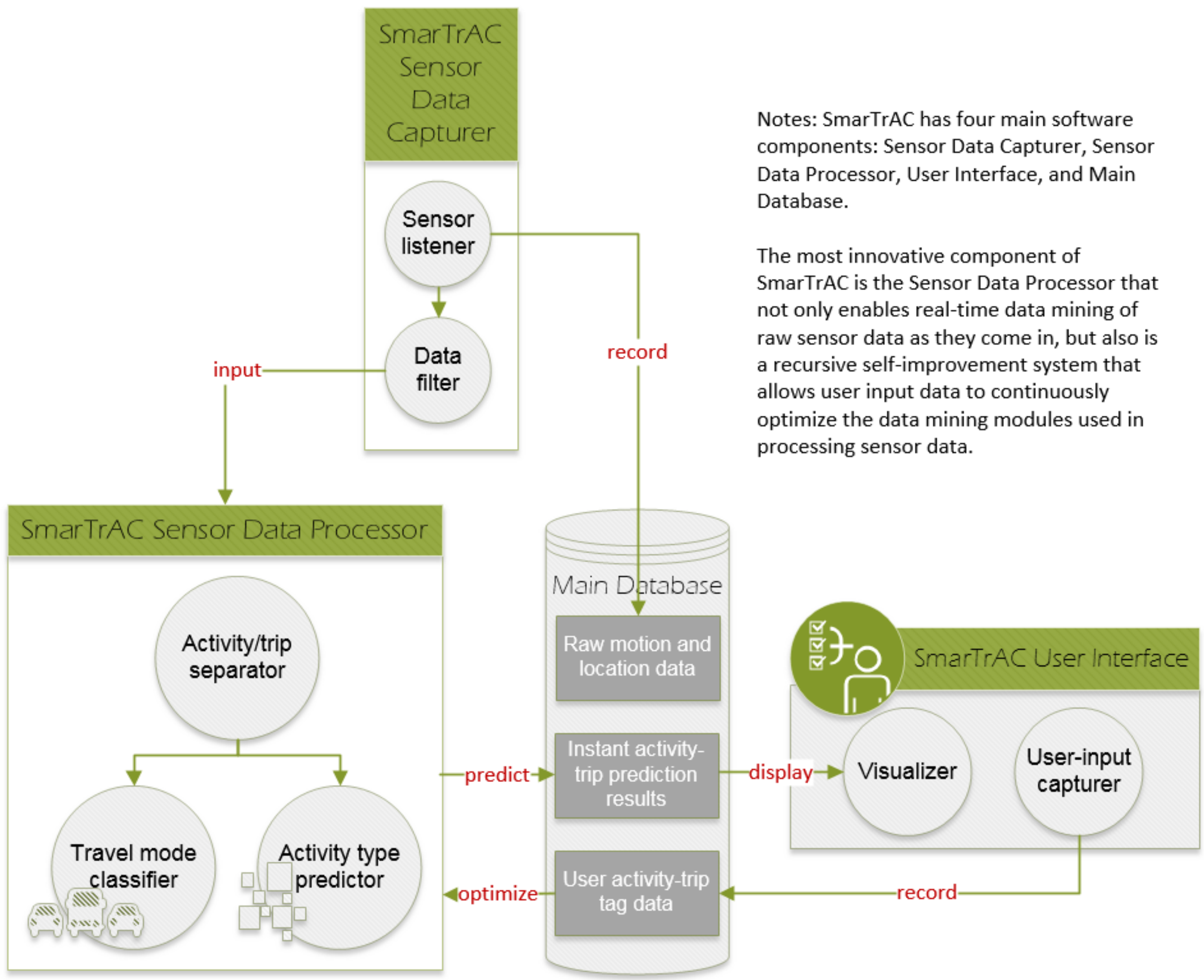


Figure 3. SmarTrAC high-level architecture

The SDC is responsible for recording and filtering raw sensor data from the smartphone's built-in sensors before the sensor data feed into the data processor. SDC consists of two real-time modules:

- *The sensor listener module* collects raw location and motion data from smartphone built-in sensors, as well as writes raw data at a 30-second interval to the Main Database. The collected location data include time-stamped latitude, longitude, speed, accuracy, and bearing. The collected motion data include time-stamped linear acceleration readings on x, y, z axis relative to the phone.

- *The data filter module* filters out poor quality location data based upon combined thresholds of accuracy, speed, and total acceleration (aggregate of raw linear accelerations along x, y, z axis mentioned above). More specifically, we remove locations with precision greater than 100 meters, or with a speed above 500 meters per second, or with a total acceleration of greater than 15 meters per square second.

The SDP is responsible for taking in the filtered location and motion data and for deriving meaningful activity and travel behavior information from the data. SDP is an important and innovative component of SmarTrAC—technical details of SDP are introduced in Section 3.4. SDP consists of three real-time prediction modules:

- *The activity/trip separator module* predicts whether the user is in the trip (travel) mode or activity (dwelling) mode at a current (or near-current, with a small time delay) time.

- *The travel mode classifier module* predicts the transportation mode at a current (or near-current, with a small time delay) time during trip episodes. The prediction outcome can be any of the following six categories: car, bus, rail, wait, bike, and walk.

- *The activity type predictor module* predicts the activity type of each activity (dwelling) episode after completion of the episode (this predictor can also be called trip purpose predictor as it identifies the trip purpose of each trip episode after completion of the activity episode for which the trip was conducted). The prediction outcome can be any of the following seven categories: home, work, education, shopping, eat out, social/recreation/community, and other personal businesses.

The UI is responsible for displaying the predicted results from SDP and for allowing the user to correct the predictions and add additional information. UI consists of two real-time modules:

- *The visualizer module* displays episode-level activity and trip information predicted by SDP, including travel mode and activity type predictions;

- *The user-input capturer module* allows the user to correct the predicted episode-level activity/trip information and add additional information on daily activities and trips.

The Main Database is responsible for storing and maintaining data. Besides raw location and motion data harvested from SDC, the Main Database maintains the following two sets of data:

- *Instant activity-trip prediction results* are obtained from SDP. Both the activity/trip separator and travel mode classifier are designed to identify dwelling vs. travel status and mode of transportation at the current time. The activity type predictor is designed to detect activity type right after the completion of a dwelling activity episode. These results are stored in the database and displayed on the UI.
- *User activity-trip tag data* are obtained via UI. User inputs on activity type and travel mode (corrections, augmentations, etc.) are stored in the database and are used to

optimize the SDP algorithms (as discussed in Section 3.4., the initial SDP algorithms are built upon general population data). Incorporating user tags on activity locations and trips in activity type and travel mode predictions will make the algorithms sensitive to individual users and improve the prediction results.

3.2 Uniqueness and Advantages

SmarTrAC is unique because it has three distinct features that set it apart from existing activity-travel data collection methods:

- *Real-time, high-accuracy prediction of travel mode and activity type:* SmarTrAC has a sophisticated data processor that is capable of extracting highly accurate episode-level activity and travel information from raw sensor data in real time or near real time.
- *On-the-fly (real-time) visualization and annotation:* SmarTrAC has an intuitive user interface that is capable of visualizing and annotating the activity and travel information extracted by its real-time data processor on the fly. Given that visualization and annotation are accessible immediately after the user enters SmarTrAC user interface, SmarTrAC provides to the user immediate read and write access to results from real-time data mining modules.
- *Recursive self-improvement:* SmarTrAC allows user input data to interact with and optimize its data mining modules so that they continuously adapt themselves to the tasks they need to perform. Specifically, user input data improve prediction models in the following two ways:
    - User input data can transform the initial "generic" prediction models for travel mode and activity type into "personalized" prediction models. In SmarTrAC, the initial prediction models are externally trained using labeled training data specifically collected by the researchers. Once SmarTrAC is installed by the user on the phone and used by the user to collect personal travel and activity data, the collected personal data (sensor data along with user tags) can be used to augment the external training data to update prediction models.
    - User input data are used to validate automatically identified activity locations based upon historical user input data on activity locations. Each time a user uses the annotation function of the SmarTrAC interface to confirm or correct the predicted activity type, the user input will be used to update a pre-existing table in the SmarTrAC Main Database that stores the relationships between locations and activity types. Because the pre-existing tables are designed as a rule-learning component in the activity type predictor module, any updates to the pre-existing table will improve the predictive accuracy of the activity type predictor module.

These three unique functions of SmarTrAC bring together smartphone-based sensing, data mining, and user communication seamlessly to enable the collection of activity and travel behavior data in more advanced ways than existing activity-travel data collection methods. Figure 4 shows that SmarTrAC improves on methods reviewed in Section 2.1 in four performance dimensions: respondent burden minimization, data comprehensiveness, data accuracy, and ease of distribution and management.

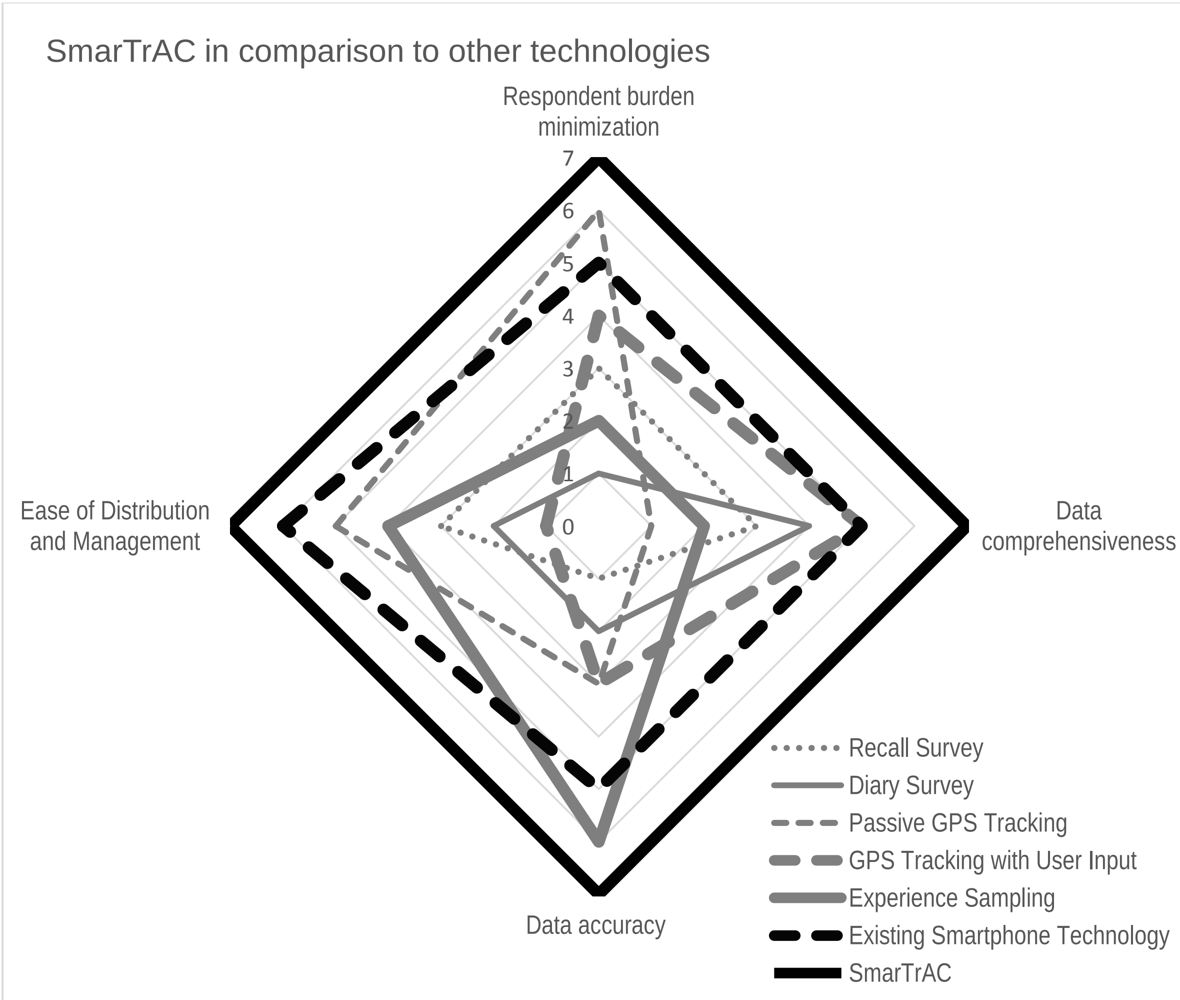


Figure 4. SmarTrAC's performance ratings compared with other activity-travel data collection methods

Specifically, why SmarTrAC is advantageous on each of the dimensions is detailed below.

- **Respondent burden minimization:** SmarTrAC does not require respondents to carry/use additional devices or instruments if the respondents already own a smartphone. SmarTrAC is also more advantageous than existing smartphone technology that simply uses smartphone-based sensing and surveying functions because SmarTrAC processes sensor data to extract meaningful activity/travel data to minimize the amount of input needed from the user.
- **Data comprehensiveness:** SmarTrAC is better than recall/diary surveys alone or passive GPS tracking alone because it is capable of capturing both time series data from built-in sensors (including GPS) and additional user input data. SmarTrAC is better than experience sampling even if experience sampling is embedded within smartphone technology because experience sampling gets user input at random timepoints of the day while SmarTrAC allows user input at any moment. SmarTrAC is better than the existing smartphone-based methods and the GPS with user input method in getting

comprehensive data because it uses information extracted from sensor data to form an information basis for the user to provide additional inputs while the existing smartphone-based methods and the GPS with user input method obtain sensor data and user input data separately even in the case that the data are obtained from a single device.

- **Data accuracy:** SmarTrAC provides advantages over other technologies because it allows sensor data to interact with user input data so that the two data sources can calibrate with each other. In this way, SmarTrAC minimizes recall bias or reporting errors in user input data because the user is able to access information derived from sensor data when providing user input.
- **Ease of distribution and management:** SmarTrAC is easier to distribute and manage than other technologies because it does not involve providing additional devices or instruments to respondents if they already own a smartphone. Users will only need to download the SmarTrAC app and install the app on their phone. Periodic updates (say, to add new features or tweak machine learning algorithms) can be provided to users quickly and easily. In addition, SmarTrAC collects activity and travel data in a way that reduces the need for post-processing. SmarTrAC generates both raw time series data and summarized episode-level data on activity and travel behavior. Both types of data can be opportunistically transmitted to any secured network storage space when connectivity is available, such as via cell phone or wireless networks. Transmitted data can be checked for quality and consistency and any necessary modifications or adjustments to the data collection process can be made quickly.

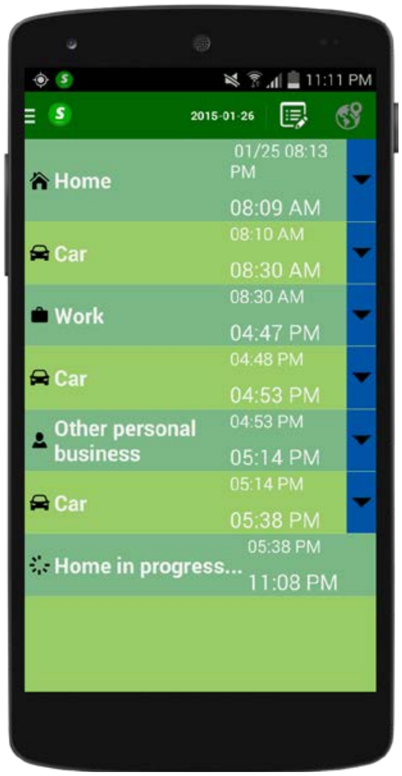
3.3 SmarTrAC UI

SmarTrAC UI visualizes activity and trip information predicted by SDP in real time and captures additional user-specified information about activities and trips. SmarTrAC UI provides a multitude of different views to the user to access and annotate information, including daily detail views, individual item views, and a daily summary view. Researchers also made instruction videos for users to familiarize with SmarTrAC UI. The videos are available online at http://smartrac.umn.edu/for-users.

3.3.1 Daily detail views

Daily detail views include the Calendar view to visualize the predicted or user-corrected sequence of activities and trips throughout the day, and the Map View to visualize the location traces of activities and trips in the space dimension (see Figure 5).

Calendar View



Notes:

- *Calendar View visualizes the sequence of activities and trips during the day. Each calendar item displays activity type in the case of an activity episode, travel mode in the case of a trip episode, and start and end times of each activity/trip episode.*
- *Map View visualizes the location traces of activities and trips in space. Activity episodes are shown as activity icons. Nearby activities are clustered to show the total number of activity episodes at the location. Travel episodes are shown as routes connecting activity episodes.*

Map View

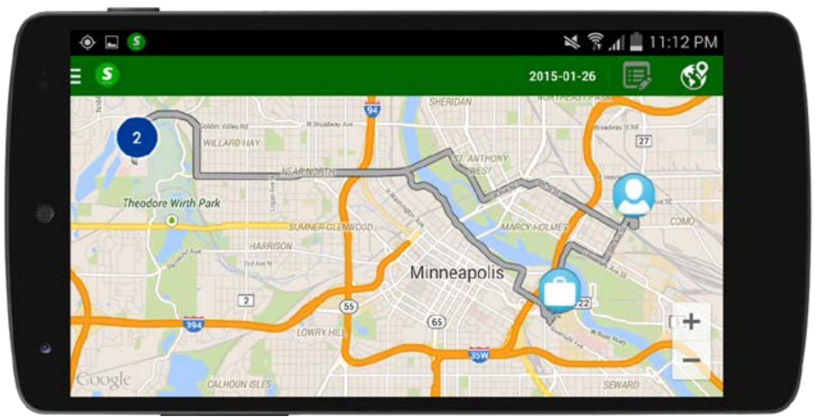


Figure 5. Daily detail views for displaying predicted daily activities and trips

3.3.2 Individual item views

The individual item views include visualization of predicted or user-corrected characteristics for each individual activity/trip episode, e.g., duration, activity type, travel mode, distance, location, route, and average speed (see Figure 6). SmarTrAC handles the visualization of travel mode prediction and activity type prediction differently. This is largely because SmarTrAC has a higher predictive accuracy in classifying travel mode than classifying activity type. The predicted travel modes for each trip are visualized in both the daily detail views and the individual item views. The predicted activity type for each activity is only visualized in the individual item Map View, as shown in Figure 6.

Information on individual items in the Calendar and Map Views

The drop down arrow for each calendar item expands to provide summary statistics for each calendar item. For example, for each trip episode, trip duration, trip distance, and average trip speed are shown. This individual component view also allow user to edit and/or add details to each calendar item through the Edit and Add Details buttons.

When the user clicks a calendar item, the user is taken to the map view of this individual calendar item. As shown below, an individual trip will be highlighted in green and information on this individual trip (e.g., start/end time and travel mode in the case of a trip item) is visualized.

Unlike the travel mode prediction module that automatically showing prediction results on the Calendar View, the activity type prediction module's prediction results are only accessible through the individual activity item's Map View. As shown below, this activity item is predicted to have a 56% probability of being other personal business, a 17% probability of being education.

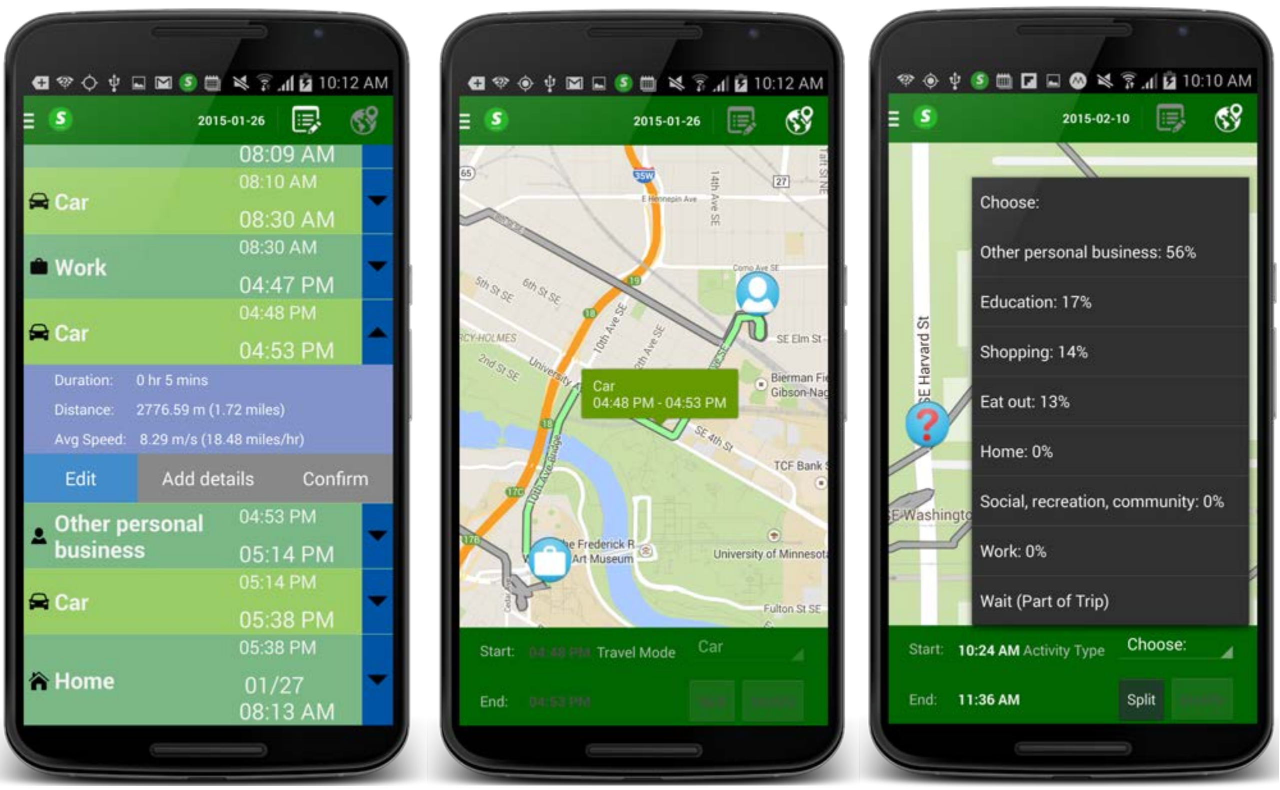


Figure 6. Individual item views for displaying an individual activity/trip episode

3.3.3 Daily summary view

The daily summary view offers summary statistics of the user's activities and trips. In the current version, the daily summary view is simplistic as shown in Figure 7. Future improvements to the daily summary view will include adding graphic presentations as well as adding week-to-date and month-to-date activity-travel statistics.

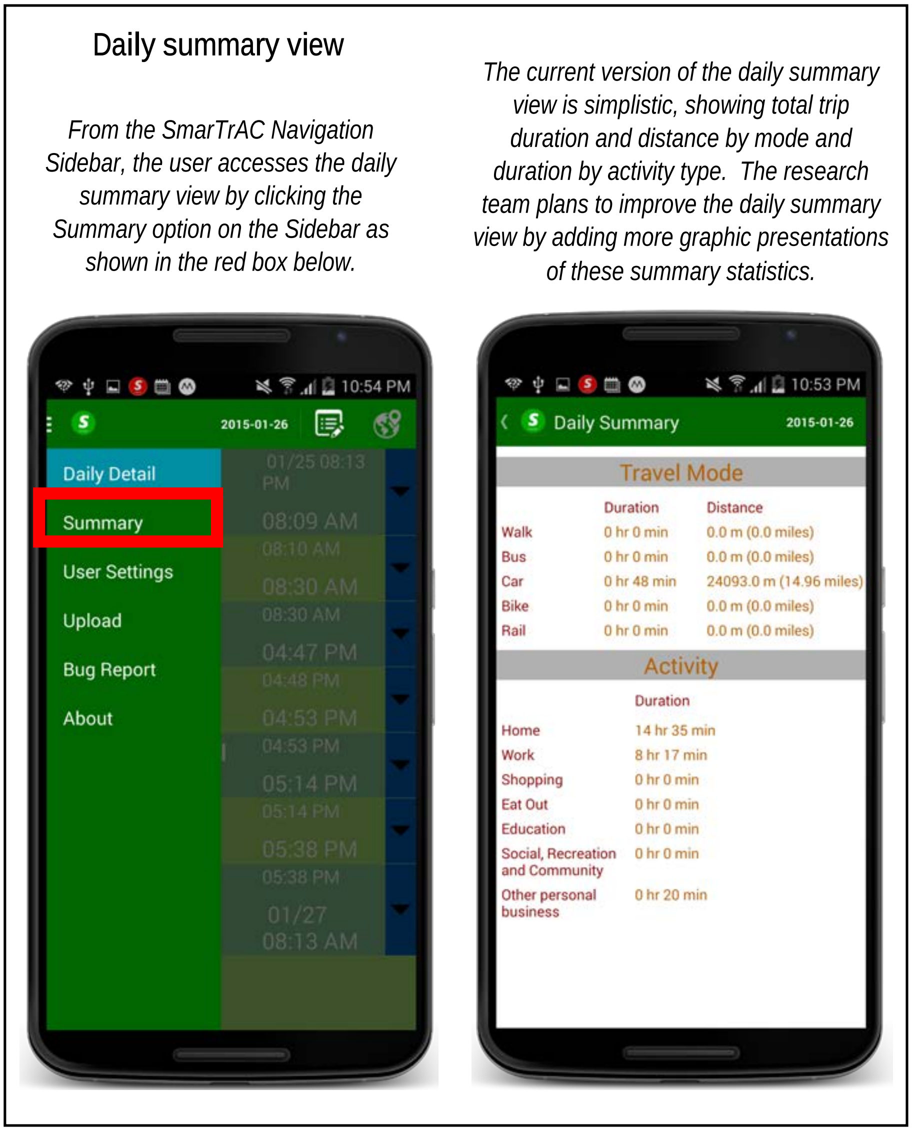


Figure 7. Daily summary view for displaying daily activity/trip summary statistics

3.3.4 Data transmission and user communication

Besides providing a multitude of different views to the user to access and annotate information, SmarTrAC UI is further enhanced by providing interfaces for the user to submit bug reports and transmit data collected by SmarTrAC to a secured cloud server at the Amazon Web Services (aws.amazon.com). Figure 8 illustrates how SmarTrAC allows easy data transmission and user communication.

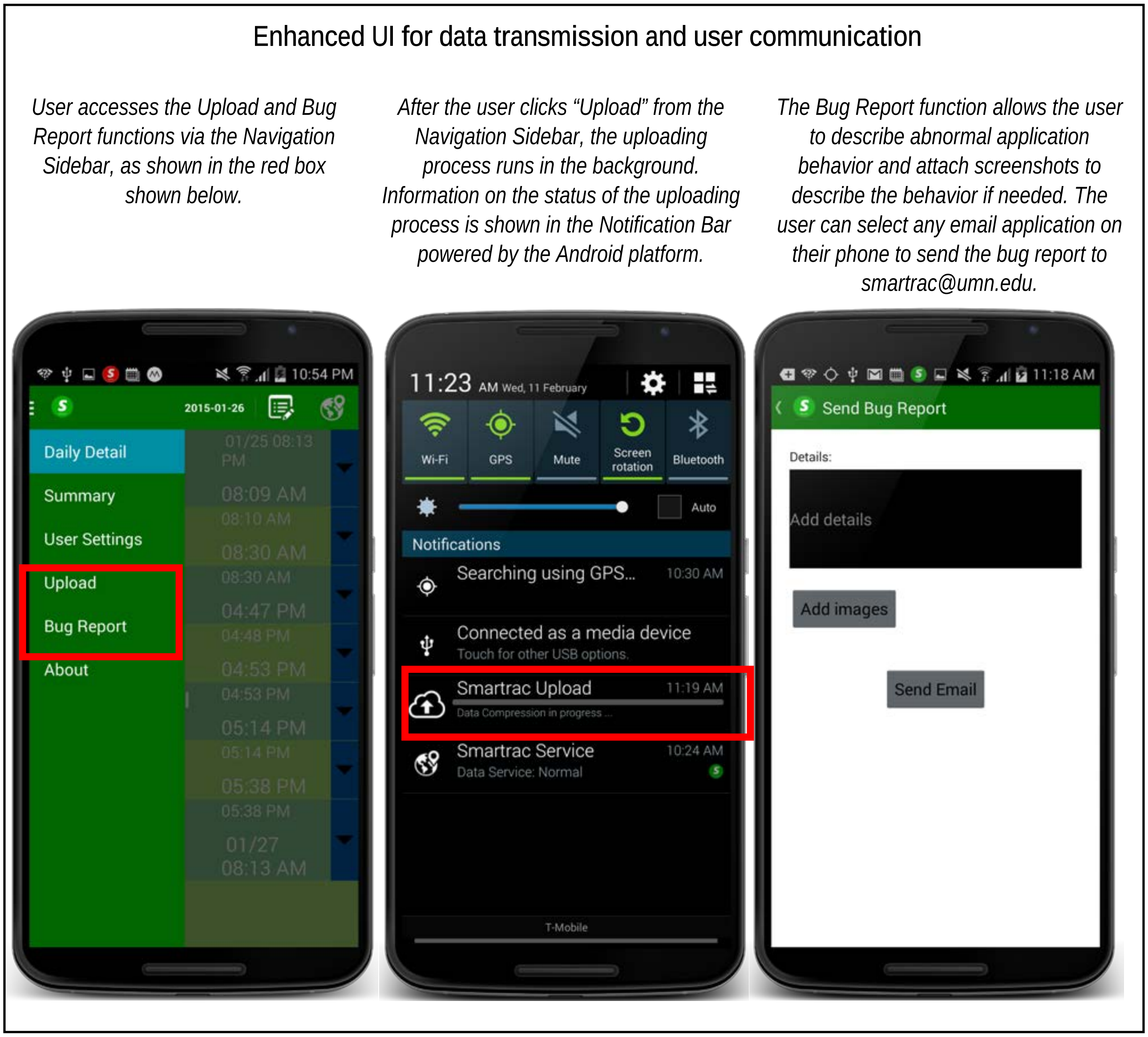


Figure 8. Enhanced UI for data transmission and user communication

3.4 SmarTrAC SDP

The most innovative feature of SmarTrAC is its sophisticated sensor data processor that not only enables real-time data mining of raw sensor data as they come in, but also is a recursive self-improvement system that allows user input data to continuously optimize the data mining modules used in processing sensor data. Below we introduce the three key SDP components: activity/trip separator, travel mode classifier, and activity type predictor.

3.4.1 Activity and trip separator

The academic literature on travel behavior generally divides time use during waking hours into two types: *trips* and *activities*. *Activities* take place in a particular location which we refer to as a *dwelling region. Trips* involve travel between activities. The activity/trip separator is responsible for identifying whether the phone is in the trip (travel) mode or activity (dwelling) mode at a current (or near-current, with a small time delay) time. The separator has three sub-functions:

- Identify the approximate starting and ending points of a dwelling (activity) episode;
- Find the precise starting and ending points for the dwelling (activity) episode; and
- Calculate the main location of the dwelling (activity) episode.

(A). Identify the approximate starting and ending points of a dwelling (activity) episode

Let *t* be a point defined by a unique (time, location) pair. SmarTrAC determines whether or not *t* is a dwelling point by assessing the diameter of the set of locations recorded within 2.5 minutes of *t* (i.e., both 2.5 minutes before *t* and 2.5 minutes after *t*). *t* is determined to be a dwelling point if the distance between all pairs of points within 2.5 minutes before and after *t* are shorter than 200 meters. In other words, *t* is labeled as a dwelling point if the GPS locations recorded within 2.5 minutes of *t* fall within a circle with diameter < 200 meters. Though GPS locations are recorded every second, SmarTrAC updates dwelling status every 30 seconds and using a coarsened (once per 30 seconds) subset of GPS data. The use of such coarser GPS track data provides a very significant reduction of computing time (i.e., allowing to compute dwelling points in real time) while maintaining high accuracy.

The dwelling episode detection algorithm used in SmarTrAC is as follows:

i. As soon as SmarTrAC is activated, cumulate GPS data for 5 minutes and sample the data at a 30-second interval to create a queue of 11 time points (point *t* and 5 time points before and after);
ii. Measure direct linear (airline) distances between all pairs of the points in the queue;
iii. If all the distances are shorter than 200 meters, declare *t* as belonging to a dwelling (activity) region. Otherwise, declare *t* as belonging to a trip.
iv. Move *t* ahead 30 seconds. This is achieved by removing the first (i.e., earliest) point in the queue and adding a new time point to the queue that is 30 seconds after the last point;
v. Repeat steps ii-iv continuously (i.e., every 30 seconds), as the GPS data are being collected.

(B). Find the precise starting and ending points for the dwelling (activity) episode

The steps above help to identify the *approximate* starting and ending points of a dwelling (i.e., activity) episode. To improve the identification and move the identified points closer to the true starting and ending points, we seek to identify the dwelling/trip change points (i.e., time points at which user's trip period ends and dwelling period starts, or at which dwelling period ends and trip period begins) in a more precise way. We find that the identification of such change points provides an excellent opportunity to refine the calculation of activity/trip boundaries. In particular, once an approximate dwelling starting point is identified, we carry out the following steps. Similar logic and steps are used to determine the ending point of a dwelling episode:

i. If steps above identified *t* as the starting point of a dwelling episode (i.e., the time point

30 seconds before *t* was identified as trip/travel mode and the time point *t* was identified as dwelling mode), get the five 30-second interval points after *t* and the five 30-second interval points before *t* to form an initial queue of 11 time points (including time point *t*), as was done in the previous steps;

ii. Calculate the maximum distance between any pairs of the points in the queue;

iii. Take out the first point in the queue and recalculate the maximum distance between all remaining pairs of points in the queue;

iv. Compare the new maximum distance with the previous maximum distance: If the new maximum distance is shorter than the previous maximum distance (indicating that the removed data point was still significantly away from the initial dwelling point *t*) and the difference between the two distances is larger than 5 meters, identify the point taken out of the queue in step iii as in travel mode; otherwise, identify this point as the revised starting point of the dwelling episode (i.e., more precise starting point than the approximate starting point *t*);

v. Repeat steps iii-iv until (a) either a more precise starting point is found in step iv, or (b) time *t* is the next point to be removed from the queue (indicating that the approximate solution *t* itself is the best candidate for the more precise starting point).

(C). Calculate the main location of the dwelling (activity) episode

Once the more accurate starting and ending points of a dwelling episode are identified, the following step is used to identify the main location of the dwelling episode:

i. Calculate the mean longitude and latitude values of all GPS points (collected at 1-second interval) between the precise starting and ending points of a dwelling episode and use the mean longitude and latitude values to identify the main location of the dwelling episode.

Please also note that the activity and trip separator is fully parameterizable. The current set of parameters (e.g., dwelling diameter of 200 m, coarser GPS interval of 30 seconds for real-time dwelling calculation, 5-minute window for dwelling determination, etc.) has been determined using pilot studies. If, after more comprehensive experiments (or in different usage settings), different parameter values prove to be more advantageous, the activity and trip separator module can be straightforwardly updated with the new parameter values.

3.4.2 Travel mode classifier

The travel mode classifier is responsible for identifying the transportation mode out of six mode options (Walk, Bike, Car, Bus, Rail, or Wait) at a current (or near-current, with a small time delay) time during trip (i.e., non-dwelling) episodes. The classifier has four subcomponents:

- Real-time data segmentation and feature construction for time segments;
- Point-wise prediction of the transportation mode at a given time instant based on the calculated characteristics of time segments (of, say, 30 or 120 seconds) immediately prior to the given time instant;
- Correction to point-wise mode prediction which incorporates smoothing based on mode

predictions of adjacent time points;
- Validation of mode transfer points based upon historical user input on mode-transfer locations.

(A). Real-time data segmentation and feature construction

The real-time data segmentation is designed to calculate segment-level aggregate characteristics of GPS and accelerometer data (including bearing, speed, and acceleration information) over 30-second and 120-second segments at 30-second intervals which are used as inputs to the embedded machine learning algorithms (currently classification random forests) described in the next section. The specific segmentation steps are:

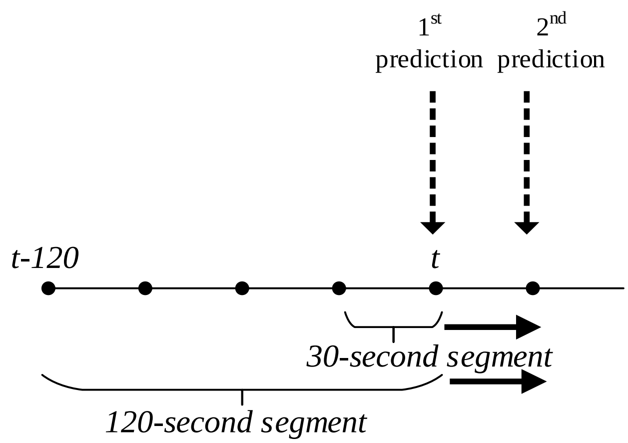


Figure 9. Segmentation and mode prediction

i. From the ending point of a dwelling episode—say time point *t* in Figure 9—which by definition is the starting point of a travel episode, cumulate GPS and accelerometer data for 120 seconds to construct the first 30- and 120-second segments;
ii. Calculate descriptive features describing patterns of bearing, speed, and acceleration in the 30- and 120-second segments;
iii. Construct the next 30- and 120-second segments with a sliding window of 30 seconds;
iv. Repeat steps ii-iii till the starting point of a dwelling episode (i.e., the ending point of this travel episode) is reached.

SmarTrAC calculates a large number of features from both speed and acceleration data. Table 3 summarizes the features calculated by SmarTrAC. Features may belong to the *time domain* or the *frequency domain*, and be *set-based* or *sequence-based* as detailed below.

Table 3. Features calculated by SmarTrAC to predict travel mode

<table>
<tr><th></th><th colspan="2">Speed data</th><th colspan="2">Acceleration data</th></tr>
<tr><th>Time domain</th><th>Set-based</th><th>Sequence-based</th><th>Set-based</th><th>Sequence-based</th></tr>
<tr><td>Mean</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Median</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Quantile</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>IQR</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Variance</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Coeff of Variation</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Minimum</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Maximum</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Kurtosis</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Skewness</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Autocorrelation</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Generalized Entropy</td><td>x</td><td>x</td><td>x</td><td>x</td></tr>
<tr><td>Bearing changes</td><td></td><td>x*</td><td></td><td></td></tr>
<tr><td>Frequency domain</td><td></td><td></td><td colspan="2"></td></tr>
<tr><td>FFT coeffs</td><td></td><td></td><td colspan="2">x</td></tr>
<tr><td>Sum of FFT coeffs</td><td></td><td></td><td colspan="2">x</td></tr>
<tr><td>Zero-crossing rate</td><td></td><td></td><td colspan="2">x</td></tr>
</table>

* = Uses sequential heading rather than speed data.

Time domain features are summary statistics which describe the distribution of the (speed or acceleration) measurements taken in a given time window. SmarTrAC computes the following time domain features for both speed and acceleration data unless otherwise noted:

- Mean: Arithmetic average of observations over a defined segment.
- Median: The sample median.
- Quantile: The sample 20th and 80th quantile.
- Inter Quartile Range: The difference between the 75th (Q3) and 25th quantile (Q1).
- Variance: The sample variance.
- Coefficient of Variation: The sample coefficient of variation.
- Minimum: Sample minimum.
- Maximum: Sample maximum.
- Kurtosis: Based on higher order moments, the kurtosis indicates the “sharpness” of peaks in the distribution of the observations.
- Skewness: Based on higher order moments, this feature describes the deviation from the symmetry of a probability distribution of a random variable around its defined mean. Skewness can be positive or negative depending on the nature of asymmetry.
- Autocorrelation: Measure of correlation between successive observations.
- Generalized Entropy: Quantifies the degree of disorder or variability in the observations.
- Bearing changes: Using data from the smartphone magnetometer, counts the number of second-to-second changes in bearing (e.g., N → NE) which exceed 15°.

Frequency domain features are calculated by viewing the set of measurements as a time series which can be described as a superposition of wave functions. The features are mostly based on the Fast Fourier transform (FFT). SmarTrAC computes the following frequency domain features on acceleration data only:

- First 6 real and imaginary components of the FFT
- Sum of the FFT coefficients: These sums are calculated separately for real and imaginary

components
- Zero crossing rate (ZCR): Measures how frequently the time series changes signs (i.e., crosses zero).

Set-based features are calculated from the actual measurements in a given time window. *Sequence-based features* are calculated from the sequential differences of measurements in a given time window. Both set- and sequence-based features only apply to data as viewed in the time domain.

(B). Point-wise travel mode prediction

The prediction of current mode of transportation is accomplished by using the features calculated for the previous 30- and 120-second segments as inputs to a classification random forests which predicts the probabilities that the current travel mode is one of: Walk, Bike, Car, Bus, Rail, or Wait. The classification random forests technique is an ensemble learning approach where predictions are made based on the aggregate of tree structures built on training data and the output generated is the mode of the predicted classes made by individual trees (i.e., prediction is based on the majority vote from individual trees).

In SmarTrAC, the random forests are externally trained and developed offline with periodical updates. The initial classification techniques were developed by using labeled data specifically collected by the researchers for building and testing mode classification models, i.e., sensor data with known transportation modes (See Figure 11 for the data collection app that the researchers developed to collect labelled data for travel mode prediction). These data were partitioned into training and test sets, of which training data were used to build predictive models and test data are used to evaluate the accuracy of the predictive models derived from training data. Before settling on random forests, the researchers explored a wide range of data mining techniques as classification and regression trees, conditional inference trees, neural networks, support vector machines, and gradient boosting techniques. (Travel mode classifier is designed in a modular fashion, allowing to “plug in” different kinds of predictive models, i.e., it is not restricted to the use of random forests only.)

Ultimately, the initial “generic” random forests used in point-wise mode prediction are expected to evolve into “personalized” prediction models, as SmarTrAC collects user-specific data that allow improvements to be made to the initial algorithms. In order to restrict the computational burden placed on the user’s smartphone, personalized prediction models will be generated on a central server and delivered to user devices as part of routine software updates.

(C). Correction to point-wise mode prediction

We use a backward-looking smoothing technique to correct point-wise, real-time mode predictions based upon the fact that sudden and brief changes in mode of transportation are highly unlikely. For example, given a sequence of predictions at 30-second intervals such as

"Car, Car, Car, Car, Bus, Car, Car, Car, Car", it is very probable that the "Bus" prediction is incorrect and should be replaced by "Car". We implement a procedure which adjusts the classification model-derived prediction at a time point by replacing it with the most commonly predicted mode in segments immediately prior to and following *the given time point.* Predictions are smoothed using some smoothing function *F* over 8 adjacent segments which span four minutes of data. In other words:

$$\text{SmoothedPred}_t = F(\text{Pred}_{t-120}, \text{Pred}_{t-90}, \text{Pred}_{t-60}, \text{Pred}_{t-30}, \text{Pred}_t, \text{Pred}_{t+30}, \text{Pred}_{t+60}, \text{Pred}_{t+90}, \text{Pred}_{t+120})$$

The current version of travel mode classifier uses the simple majority vote as the smoothing function *F*; however, other smoothing functions can be used as well. The specific steps for smoothing based upon the simple majority vote are:

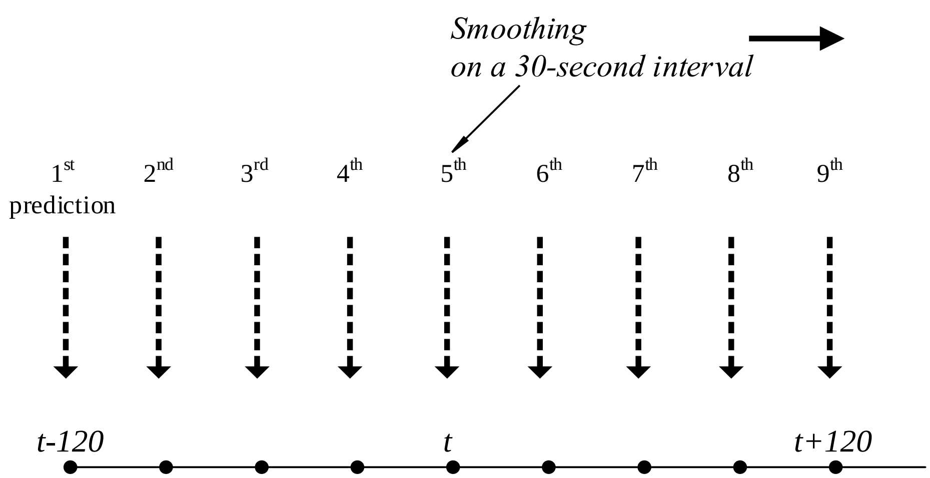


Figure 10. Smoothing of point-wise predictions

i. Compare the predicted mode at the time point *t* with the eight predicated modes at nearby points (defined as 4 points before and after as shown in Figure 10): If the predicted mode at *t* is not the most commonly predicted mode at the eight nearby points, replace the predicted mode at *t* with the most commonly predicted mode; otherwise, remain the original prediction;
ii. Move *t* ahead 30 seconds. This is achieved by remove the first (i.e., earliest) point in the nine-point queue (as shown in Figure 10) and add a new time point to the queue that is 30 seconds after the last point;
iii. Repeat step i-ii until the end of the travel episode.

Because it uses predicted modes from both before and after a given time point, the correction/smoothing procedure operates on predictions made two minutes before the current time. Hence, mode predictions in SmarTrAC can be viewed as “preliminary” for the first two minutes after they are made and “final” after that point, when the smoothing procedure has processed (and possibly corrected) the preliminary mode prediction. It is important to note that the initial mode predictions and smoothed mode predictions are stored separately; therefore, the all smoothing corrections can be observed and used as important information for further predictive model refinements.

Note that travel mode classifier (like other SmarTrAC modules) is fully parameterizable. The current set of parameters (e.g., using 30-second and 120-second feature segments, majority-based prediction smoothing over 4-minute horizon, random-forest-based classification, etc.) has been determined using pilot studies. If, after more comprehensive experiments (or in different usage settings), different parameter values prove to be more advantageous, the travel mode classifier module can be straightforwardly updated with the new parameter values.

3.4.3 Activity type (trip purpose) predictor

The activity type predictor is responsible for identifying the type of each activity (dwelling) episode out of seven options (including home, work, education, shopping, eat out, social/recreational/community, and other personal business) after completion of the episode. It has two subcomponents:

- Tagging based upon historical user input about activity locations.
- Prediction of the probability of each activity option occurring at previously unvisited locations.

(A). Tagging based upon historical user input on activity locations

When first run, SmarTrAC asks users to pinpoint their home location and primary workplace (if any) on Google Maps. SmarTrAC stores the home/work locations along with other user-verified and previously visited activity locations in the main SmarTrAC database in a format as shown in Table 4.

Table 4. Historical activity locations

ID	Latitude	Longitude	Activity Type	Total number of visits at the location
1				
2				
3				

The specific steps for tagging are:

i. As soon as the activity and trip separator identifies a complete dwelling (activity) episode and the main location of the episode, calculate distance from the identified activity location to historical activity locations;
ii. If the identified activity location is within 50 meters of a historical activity location, automatically tag the location with the most frequent activity type associated with the historical activity location; various SmarTrAC user interface elements allow end user to confirm/validate the automatically tagged activity types;
iii. If the identified activity location is outside 50 meters of any historical activity locations, use the other sub-component of the predictor (see the section below) to

predict the likelihood (probability) of each activity option occurring at the location. SmarTrAC will sort all available activity options based upon the predicted likelihood (probability) as candidates for user-driven tagging via SmarTrAC user interface features.

(B). Prediction of the probability of each activity option occurring at previously unvisited locations

Similar to prediction of travel modes, prediction of the probability of each activity option occurring at a unvisited location is based upon classification random forests that are trained and developed offline and embedded in the SmarTrAC application, with periodic updates. The initial random forests are developed by using a pre-existing activity dataset that includes information on activity destination and type. The dataset is derived from the 2010 Travel Behavior Inventory (TBI) survey data collected by the Twin Cities Metropolitan Council. For each activity episode derived from the 2010 TBI data, the following information is available:

- Activity type (seven options, including home, work, education, shopping, eat out, social/recreational, and other personal business,)
- Type of the previous activity (same seven options as above)
- Day of the week (Day of week trip was taken; 1-Sunday and 7-Saturday)
- Holiday (Was the trip taken on an official holiday? Yes/No)
- First trip (Was this the first trip of the day?)
- Trip number (How many trips were taken before this trip?)
- Primary mode (Main mode of transportation used during trip)
- Latitude/Longitude (Location of trip destination)
- Arrival time (Arrival time at trip destination)
- Airline distance ("As the crow flies" distance between trip origin and destination)
- Activity duration (How long did the activity at the trip destination last?)
- Worker (Was the individual taking the trip employed?)
- Current student (Was the individual taking the trip a current student?)

We augmented the TBI data by adding neighborhood information from Google Places to describe nearby businesses of each activity location. This augmented activity dataset was partitioned into training and test sets to build and evaluate models for predicting activity type. The models used activity type (a categorical variable with seven options) as the outcome variable and all other information in the dataset such as information on nearby businesses and activity characteristics listed above as explanatory variables. Techniques used in deriving the initial predictive models include a wide range of data mining techniques available based on large research literature such as classification and regression trees, conditional inference trees, random forests, neural networks, support vector machines, and gradient boosting techniques.

The initial "generic" random forests could be updated periodically as SmarTrAC generates more personal data on activity episodes. Ultimately, the random forests used in activity type prediction are expected to evolve into "personalized" prediction models as SmarTrAC keeps generating user-specific data that allow improvements to be made to the initial rules. Similar to the travel mode classifier, personalized prediction models used in the activity type predictor will be generated on a central server and delivered to user devices as part of routine software updates for the purpose of restricting the computational burden placed on the user's smartphone.

CHAPTER 4. LABORATORY TESTS

Between September 2013 and November 2014, the researchers conducted a series of tests. These tests were not necessarily carried out in strictly controlled laboratory environments, but rather were carried out by the researchers to evaluate specific features of SmarTrAC. A total of eight unlocked, android phones were used in these tests, including two phones of the following four phone models: HTC One, Samsung Galaxy S4, Google Nexus, and Sony Xperia Z.

The four phone models are selected largely because they are highly ranked on popularity polls at online forums such as “PCWorld”,”GSMArena” “Brighthand”, “Lifehackers” and “the Verge”. As shown in Table 5, the four phone models all use the Qualcomm Snapdragon processor chip. The chip has a “dual-core” location system that is capable of accessing both the GPS and GLONASS satellite networks, which leads to similar GPS performance across the four models. The four phone models use different built-in accelerometers (Table 6), which produce acceleration data of different quality.

The laboratory tests can be categorized into three stages.

- The early stage from September 2013 to April 2014 that focused on extracting high-quality motion and location data from smartphone sensors;
- The second stage from May to June 2014 that focused on developing and testing data mining algorithms for travel mode classification and activity type prediction; and
- The final stage from July to November 2014 that focused on testing user experiences including issues of battery consumption, data storage and transmission, and UI enhancement.

Table 5. General hardware specifics across test phones

Phone model	Battery Life		Battery Type	Cores	Storage		Processor Speed	System Chip	Android OS	Source
	Talk time	Standby time	mAh		ROM	RAM				
HTC One	26.20 hours- At&T .6.70 hrs-T Mol	19.2 days AT&T 17 days - T Mob	2300	Quad	32 GB	2048 MB RAM	1700 MHz, Krait 300	Qualcomm Snapdragon 600 APQ8064T	Android (4.2.2, 4.1.2) Sense 5.0 UI	http://www.phonearena.com/phones/compare/HTC-One/phones/7678
Samsung Galaxy S4	17.00 hours - 3G	13.3 days (320 hours) -4G	2600	Quad	16 GB	2048 MB RAM	1900 MHz, Krait 300	Qualcomm Snapdragon 600 APQ8064T	4.2.2	http://www.phonearena.com/phones/Samsung-Galaxy-S4_id7597
Google Nexus	15.30 hours	16.2 days (390 hours)	2100	Quad	8 GB	2 GB	1500 MHz, Krait	Qualcomm Snapdragon S4 Pro APQ8064	Android (4.3, 4.2.2, 4.2.1, 4.2)	http://www.phonearena.com/phones/compare/Google-Nexus-4/phones/7531
Sony Xperia Z	14.00 hours - 3G	22.1 days (530 hours) - 3G	2330	Quad	16 GB	2048 MB RAM	1500 MHz, Krait	Qualcomm Snapdragon S4 Pro APQ8064	Android 4.2.2, 4.1.2	http://www.phonearena.com/phones/Sony-Xperia-Z_id7539

Table 6. Specifics of the built-in accelerometer across test phones

Smartphone	Company	Model No.	Axial	Range	Normal mode - Current consumption	Digital Resolution	Zero g-offset (over lifetime)	Sampling /Output Data rate	FIFO	Datasheet
HTC One	Bosch	BMA 250	3 Axis	±2g, ±4g, ±8g and ±16g	139µA	10 Bit	80 mg	8 – 1000 Hz	NO	http://ae-bst.resource.bosch.com/media/products/dokumente/bma250/BST-BMA250-DS002-05.pdf
Samsung Galaxy S4	STMicroelectronics	LSM330	6 Axis (Gyro + Acc)	±2g, ±4g, ±6g, ±8g, ±16g	250µA	8 Bit	±60 mg	3.125 - 1600 Hz	Yes[a]	http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/DM00059856.pdf
Google Nexus	Invensense	MPU-6050	6 Axis (Gyro + Acc)	±2g, ±4g, ±8g and ±16g	500µA	16 Bit	±60 mg	4-1000 Hz	Yes[b]	http://www.invensense.com/mems/gyro/documents/PS-MPU-6000A.pdf
Sony Xperia Z	Bosch	BMA 250	3 Axis	±2g, ±4g, ±8g and ±16g	139µA	10 Bit	80 mg	8 – 1000 Hz	NO	http://ae-bst.resource.bosch.com/media/products/dokumente/bma250/BST-BMA250-DS002-05.pdf

Note: [a]"LSM330 embeds a 32 slots of 16 bit data FIFO buffer for each of the three output channels, yaw, pitch and roll . This allows a consistent power saving for the system, since the host processor does not need to continuously poll data from the sensor, but it can wake up only when needed and burst the significant data out from the FIFO."
[b]1024 Byte- reduces power consumption by allowing host processor to read the data in bursts and then go into a low-power mode as the MPU collects more data.

4.1 Early Testing on Capturing Sensor Data

4.1.1 Data and approach

A separate app “SmarTrAC Data Collection” was developed to test sensory performance in capturing location (GPS) and motion (acceleration) data across phone models. The app, shown in Figure 11, allows for multiple options of sampling rates (i.e., listening frequency) and writing rates (i.e., file input/output frequency) when collecting location and motion data. The app also allows tagging of travel modes, which helps to make data collected in this early testing stage useful for developing data mining algorithms for identifying travel mode in the advanced testing stage.

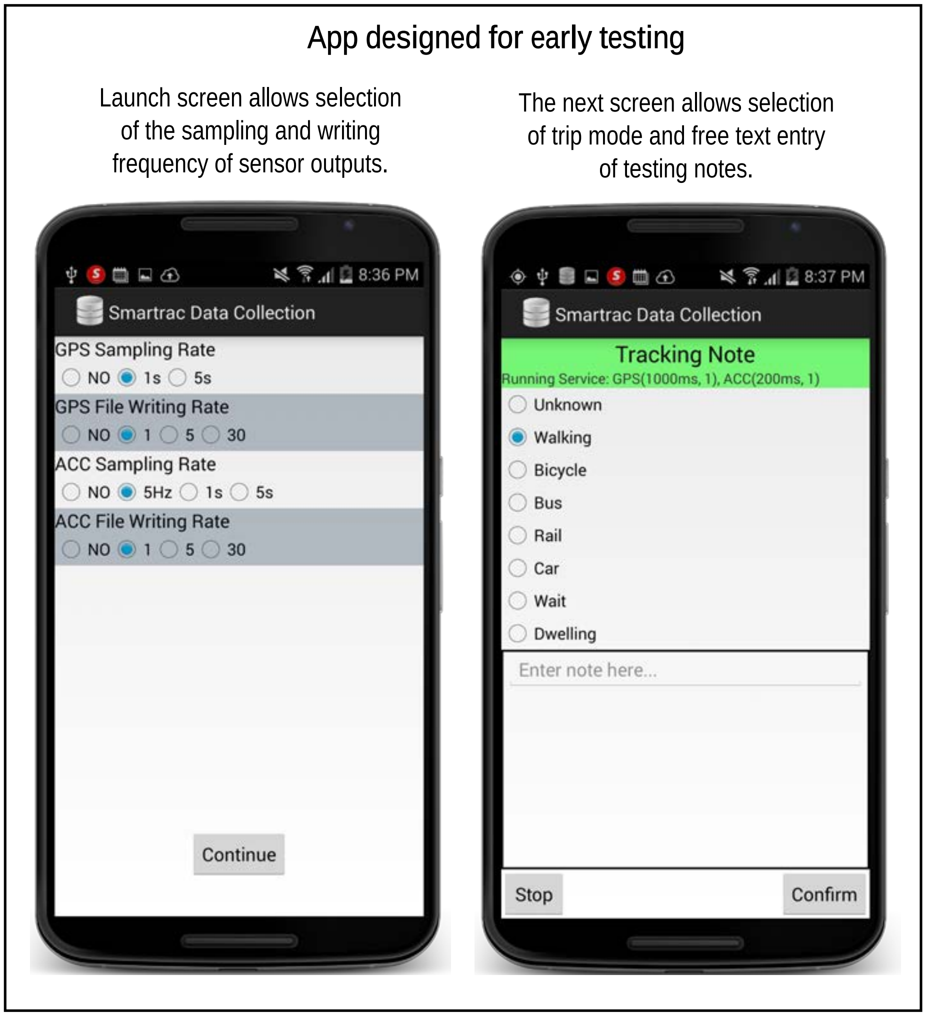


Figure 11. Screenshots showing the app desgined for early testing and collecting labelled data for travel mode prediction

This “SmarTrAC Data Collection” app was installed on all test phones and researchers often carry multiple phones during the same trips to generate comparable sensor data across test phones. To give an example, on Jan 8, 2014, a researcher conducted a 72-minute multi-modal trip (see Figure 12) and tested sensory performance across three different phones, including HTC

One, Samsung Galaxy S4, and Google Nexus. Other testing trips used a similar approach. Results from these testing trips are discussed in the section below.

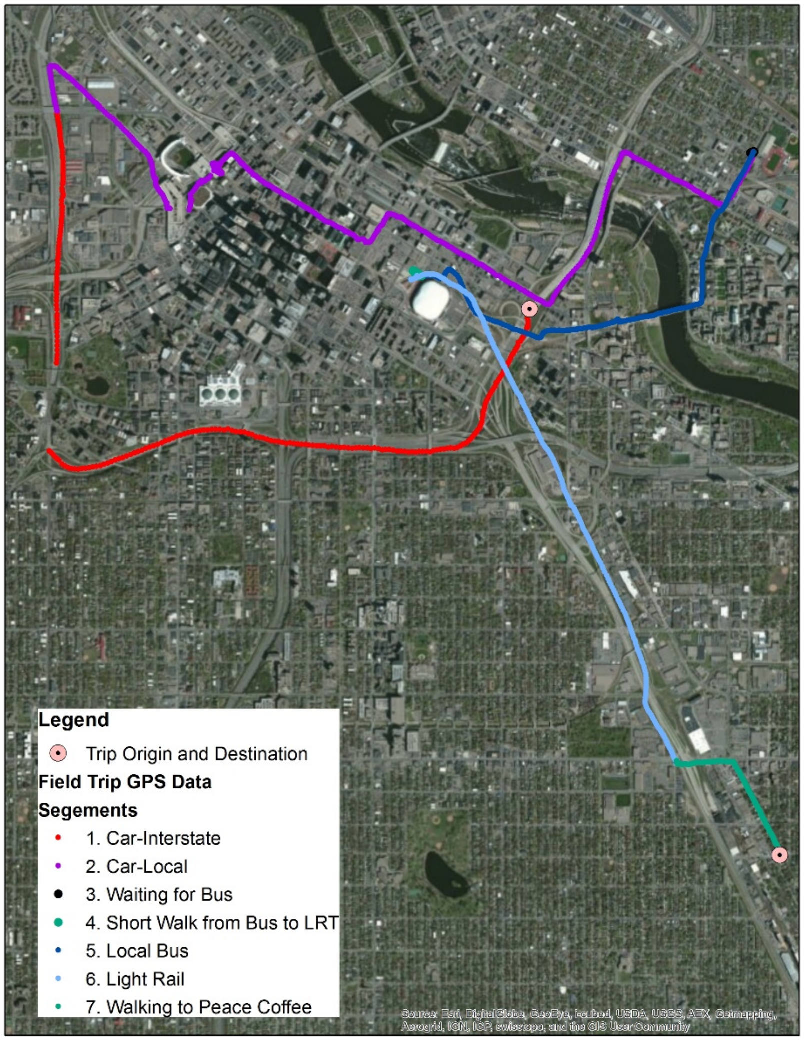


Figure 12. A example of early tests that focusing on sensory performance in capturing location and motion data

4.1.2 Results

The multiple testing trips show similar results that HTC One appears to use a different acceleration scale than the other phones. Table 7 summarizes the results from the testing trip illustrated in Figure 12. As shown in Table 7, for this same trip, the mean acceleration by travel mode ranges between 0.6 and 5.4 meters per square second in Google Nexus, between 0.5 and 3.7 meters per square second in Samsung Galaxy, and between 0.2 and 1.7 meters per square second in HTC One. This might be due to hardware differences. SamSung and Google respectively use accelerometers from STMicroelectronics (French-Italian) and Invensense (US-based). HTC uses accelerometer from Bosch (German-based).

Table 7. Acceleration statistics by trip segments and by phone model

Google Nexus						
	Duration	**Distance (miles)**	**Quantile**		**Acceleration - Mean**	**Acceleration – Std Dev**
			5 %	**95 %**	**Per minute**	
Car – Interstate	4 mins	3.5	0.296	1.779	0.912	0.479
Car – Local	19 mins	4	0.089	1.830	0.639	0.598
Waiting at bus stop	13 min	-	0.097	2.083	0.759	0.771
Bus (3A)	12 min	2	0.146	2.157	0.837	0.751
Short Walk	1 min	0.02	2.279	9.478	5.434	2.303
Light Rail	8 min	2.148	0.075	3.022	0.842	1.202
Walk	15 min	0.6	0.957	7.651	3.872	2.113
Total	**72 min**	**12.268**				
Samsung Galaxy S4						
	Duration	**Distance (miles)**	**Quantile**		**Acceleration - Mean**	**Acceleration – Std Dev**
			5 %	**95 %**	**Per minute**	
Car – Interstate	4 mins	3.5	0.294	1.715	0.881	0.460
Car – Local	19 mins	4	1.488	0.132	0.587	0.468
Waiting at bus stop	13 min	-	0.121	2.317	0.772	0.842
Bus (3A)	12 min	2	0.149	2.080	0.808	0.701
Short Walk	1 min	0.02	0.105	8.496	3.454	2.960
Light Rail	8 min	2.148	0.267	3.418	1.242	1.185
Walk	15 min	0.6	1.092	6.906	3.712	1.883
Total	**72 min**	**12.268**				
HTC One						
	Duration	**Distance (miles)**	**Quantile**		**Acceleration - Mean**	**Acceleration – Std Dev**
			5 %	**95 %**	**Per minute**	
Car – Interstate	4 mins	3.5	0.080	0.754	0.347	0.224
Car – Local	19 mins	4	0.026	0.609	0.208	0.210
Waiting at bus stop	13 min	-	0.058	1.415	0.474	0.483
Bus (3A)	12 min	2	0.044	1.166	0.389	0.404
Short Walk	1 min	0.02	0.323	3.767	1.711	1.108
Light Rail	8 min	2.148	0.060	1.738	0.557	0.651
Walk	15 min	0.6	0.422	3.938	1.821	1.098
Total	**72 min**	**12.268**				

Nonetheless, HTC still produces acceleration patterns similar to Google and Samsung phones. Figure 13 shows the mean of the total acceleration per second by phone model. Although the absolute values of mean acceleration in HTC One is generally much smaller than those of Google and Samsung phones (note the different range on Y axis for the HTC One phone in Figure 13), relatively the acceleration pattern over time appears to be similar across the phones. These results indicate the importance of using a wide range of summary statistics when describing acceleration patterns. If travel mode prediction is purely dependent upon statistics of non-normalized raw data, such travel mode prediction is likely to perform poorly on specific phone models. The acceleration statistics calculated by SmarTrAC include both set-based and sequence based measures in time and frequency domains (see Table 3 in Section 3.4.2), which helps to address this issue.

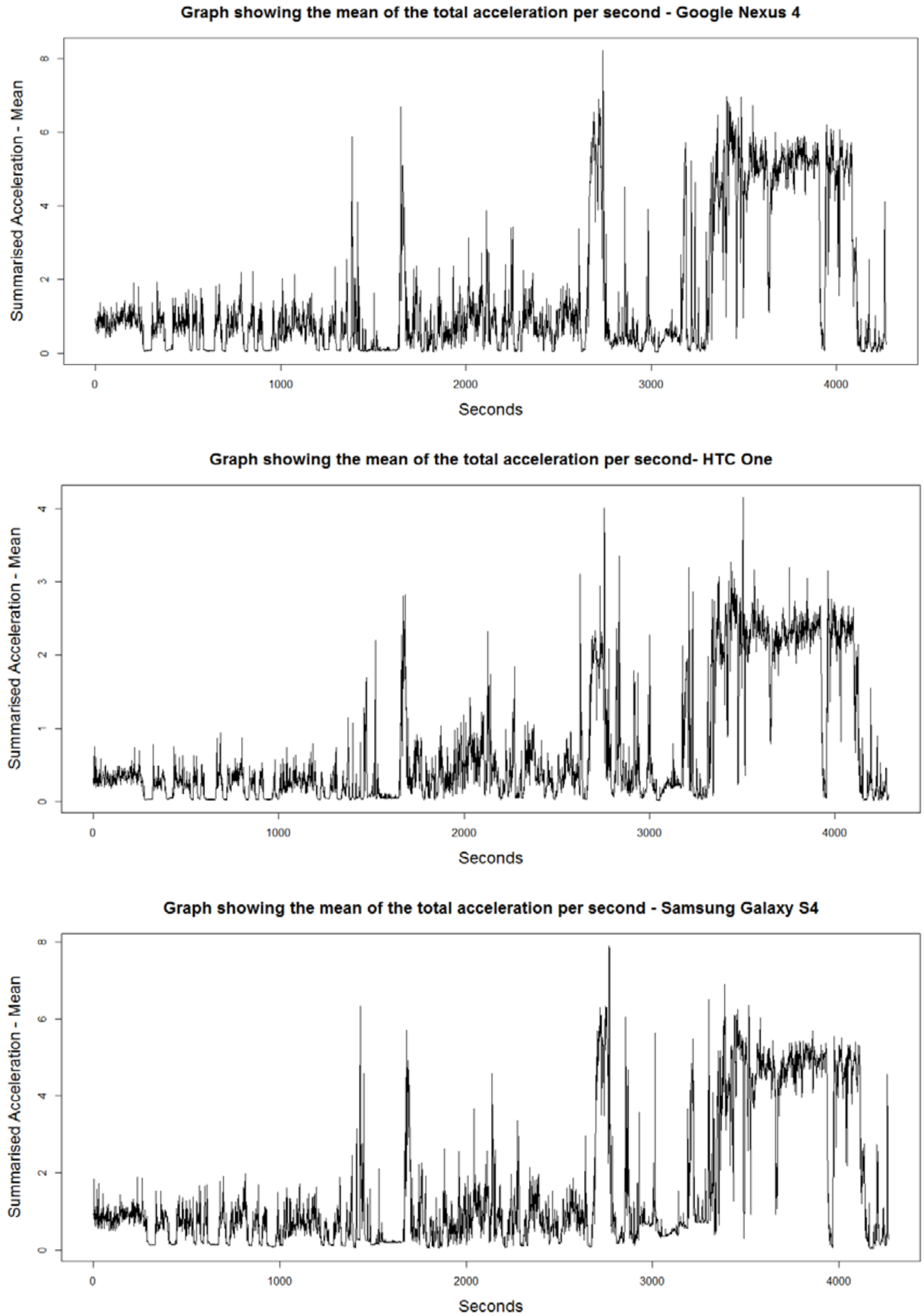


Figure 13. The mean of the total acceleration per second by phone model

Our early testing also shows that location data can have highly varying levels of accuracy. This makes plotting trip trajectories on Google Maps (and, more importantly, inferring travel mode from location data) challenging, as the line "jumps" to distant locations (see Figure 14). SmarTrAC implements some basic heuristics to improve the plotting of trajectories by filters out poor quality location data. The location filtering is based upon combined thresholds of accuracy, speed, and total acceleration. More specifically, we remove locations with an accuracy of greater than 100 meters, or with a speed above 500 meters per second, or with a total acceleration greater than 15 meters per square second. As shown in Figure 14, the post-filtering travel path is much cleaner and more accurately represents the actual path taken.

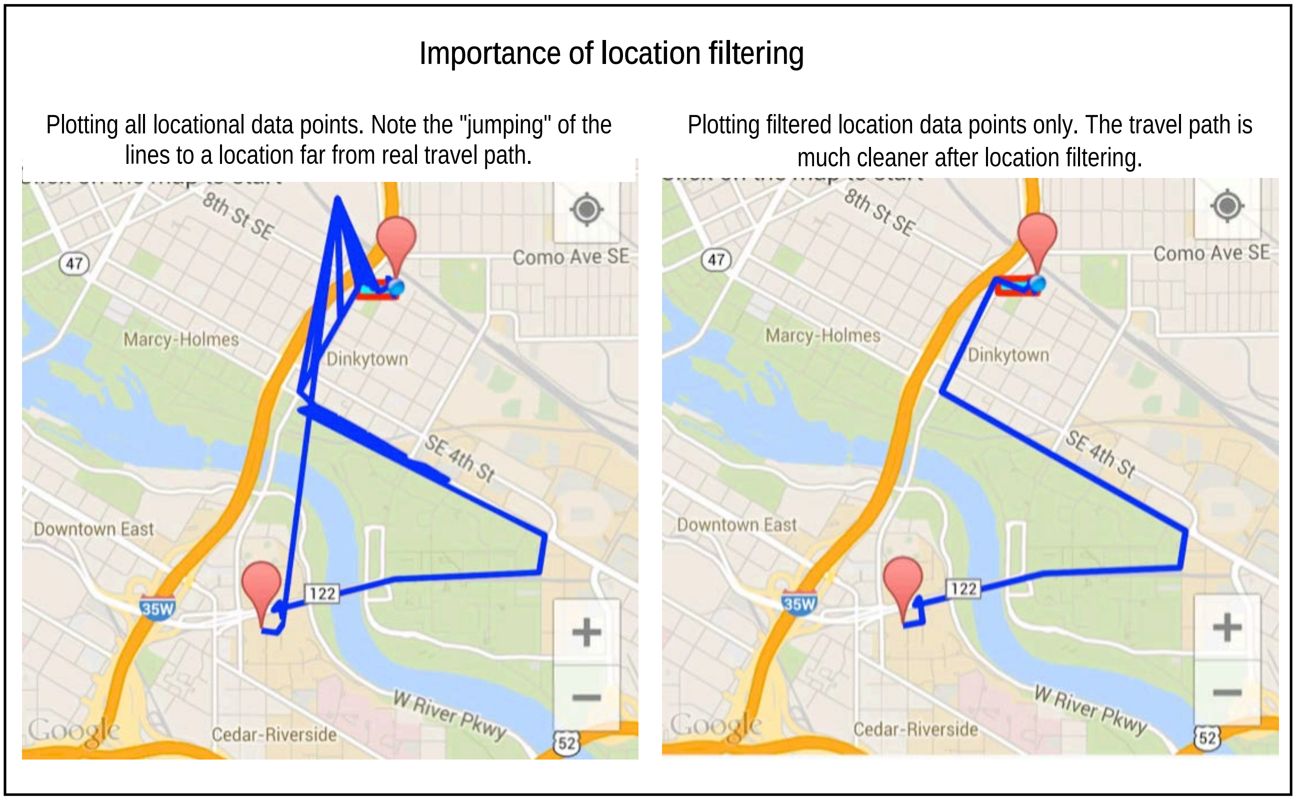


Figure 14. Eliminating poor quality data points from GPS tracks

4.2 Development and Testing of Sensor Data Processing Algorithms

4.2.1 Data

(A). Data for development and testing of travel mode classifier

The data collection application (Figure 11) was set to collect location data every second and motion data every 1/5 second. The application allows users to tag travel mode in real time. Three researchers carried the phones with them over a period of two weeks from April 21st to May 3rd 2014. During these two weeks, they were able to use the data collection app to capture 64 trips (a total of 697 minutes of travel) with each trip tagged with the actual modes used during the trip. As shown in Figure 15, the data include a total of six modes: walk (28 minutes), wait (54 minutes), bus (117 minutes), bicycle (161 minutes), car (147 minutes), and rail (189 minutes).

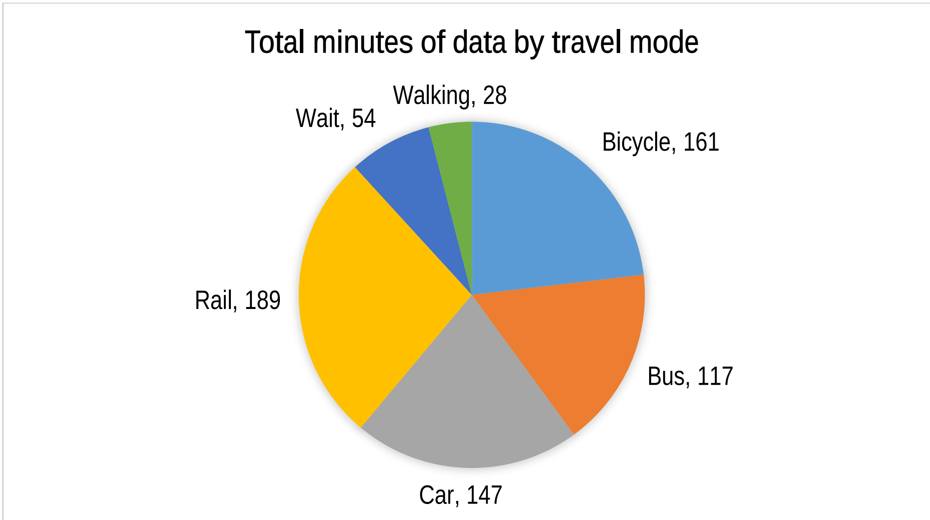


Figure 15 Amount of test data by travel mode in minutes

As mentioned above, the raw data were recorded on the phone(s) at 1-second (GPS) and 0.2-second (accelerometer) intervals. Our prediction algorithms do not operate on these raw data, but rather on features calculated from data aggregated over different time horizons (i.e., segments). We compute 35 features over 30-second and 120-second segments. The features calculated include both time and frequency domain features describing acceleration and speed data, as shown in Table 3 in Section 3.4.2.

(B). Data for development and testing of activity type predictor

The data for testing activity type predictor are derived from the 2010 Travel Behavior Inventory (TBI) survey data collected by the Twin Cities Metropolitan Council. In the 2010 TBI, approximately 14,000 households (including approximately 30,000 individuals) were asked to record trips they made during a single day, resulting in a total of 110,761 separate trips. The outcome of interest for the analysis was trip purpose. Survey participants recorded the purpose of each of their trips, choosing from the seven options as shown in Table 8.

Table 8 Frequency and percentage distribution by trip purpose

Trip Purpose		Frequency	Percent
1	Home	39,225	35%
2	Work	17,033	15%
3	Education	5,514	5%
4	Shopping	11,031	10%
5	Eat out	5,865	5%
6	Social /recreation/community	11,854	11%
7	Other personal business	20,239	18%

Section 3.4.3 (B) provides a list of the TBI variables used in predicting activity type (trip purpose) and evaluating the prediction methods. As mentioned in Section 3.4.3 (B), we augment

the TBI data by adding neighborhood information from Google Places to further describe each activity location (i.e., trip destination).

Specifically, we used the longitude and latitude of each activity location to query the Google Places Application Programming Interface (API) and obtain a list of points of interest within 100 meters of the activity location. Note that Google Places API only returns a maximum of 20 results per query, with results ordered by popularity. Points of interest are labeled according to their type. As shown in Table 9, there are approximately 100 Google Places tags organized in a loose hierarchical structure, from the rather generic ('establishment') to the very specific ('beauty salon'). A location may have multiple tags associated with it, or none. For each activity episode in the TBI, we recorded the number of tags of each type assigned to the points of interest within 100 meters of the activity location. So, for example, a location whose nearby points of interest include 7 restaurants, 2 gas stations and a grocery store would have the numbers 7, 2, and 1 recorded for the ‘restaurant’, ‘gas_station’ and ‘grocery_store’ tags. Since at most 20 nearby locations are returned by the Google Places API, the maximum count for each tag is 20. Some activity locations (typically in residential areas) had no nearby points of interest in the Google database, and this was also recorded.

For activity prediction purposes, we summarized the tags associated with the nearby locations to a trip destination by the percent of nearby locations having that tag. For example, if 20 nearby locations are returned of which 4 are tagged 'restaurant', the 'restaurant' predictor value would take on the value 4/20 = 0.2. We further included a variable counting the number of nearby locations tagged 'establishment' as well as a variable indicating when no nearby locations returned (typically, these were locations in residential neighborhoods).

Table 9 Place categories defined by Google

accounting	embassy	moving_company
airport	establishment	museum
amusement_park	finance	night_club
aquarium	fire_station	painter
art_gallery	florist	park
atm	food	parking
bakery	funeral_home	pet_store
bank	furniture_store	pharmacy
bar	gas_station	physiotherapist
beauty_salon	general_contractor	place_of_worship
bicycle_store	grocery_or_supermarket	plumber
book_store	gym	police
bowling_alley	hair_care	post_office
bus_station	hardware_store	real_estate_agency
cafe	health	restaurant
campground	hindu_temple	roofing_contractor
car_dealer	home_goods_store	rv_park
car_rental	hospital	school
car_repair	insurance_agency	shoe_store
car_wash	jewelry_store	shopping_mall
casino	laundry	spa
cemetery	lawyer	stadium
church	library	storage
city_hall	liquor_store	store
clothing_store	local_government_office	subway_station
convenience_store	locksmith	synagogue
courthouse	lodging	taxi_stand
dentist	meal_delivery	train_station
department_store	meal_takeaway	travel_agency
doctor	mosque	university
electrician	movie_rental	veterinary_care
electronics_store	movie_theater	zoo

Note that before we settled on the Google API, we also tested location tag data from the Foursquare API. The following table gives a brief summary of how Foursquare data compares to Google data. Our tests show that Google Places API gives the most accurate places results. As a result, we decided to use Google Places API exclusively in SmarTrAC.

Table 10. Foursquare API data in comparison to Google API data

Foursquare API	Google API
Well defined hierarchies	Location-types are ambiguous
Effective crowdsourcing	Minimum crowdsourcing
Dependence on socio-demographic characteristics	No such dependence emerges
Coverage limited to areas ‘popular’ among certain demographic	Broad coverage
Dependent only on self-data collection	Collaboration with multiple data sources

4.2.2 Approach

(A). Machine learning techniques for predicting travel mode and activity type

We applied the four decision-tree-based classification techniques in both travel mode classification and activity type (trip purpose) prediction. All these techniques are based on the recursive partitioning approach that works by splitting the training data into segments using available independent variables. The purpose is to obtain segments that are *homogeneous,* which would allow to correctly classify the data. Brief descriptions of the four techniques are shown below; names in `monospaced font` refer to packages in the R statistical software program:

- Classification and regression trees (CART) (Breiman et al., 1984) as implemented in `rpart`. This method learns decision trees based on the reduction in error or impurity index that each level of the split generates. When no further reduction in impurity or error is obtained, the tree ceases the splitting process and the predictions are so obtained;
- Conditional inference trees (Hothorn et al., 2006) as implemented in the `party` package in R. This method predicts decision trees based on p-values that are calculated from a permutation test framework. P-values drive the determination of variables to split on as well as the procedure of splitting, removing the need for the pruning process that is required in the CART method. This inherent structure serves to remove the problem of over-fitting and biased selection towards variables with greater number of splits that the traditional CART method showcases.
- Random forests (Breiman, 2001) as implemented in `randomForest`. This is an ensemble learning approach where predictions are made based on the aggregate of tree structures built on training data and the output generated is the mode of the predicted classes made by individual trees. Visual representation of the predictions from this method is virtually impossible because, unlike the CART and the conditional inference methods that generate an individual tree structure, the random forests method produces a large number of tree structures and uses all the trees to generate prediction results. Nonetheless, the random forests method typically produces a more reliable and accurate prediction than the traditional CART method and the conditional inference trees method since the inherent variation present among the various predicted tree structures is accounted for.
- Gradient boosting (Friedman, 2001) as implemented in `gbm`. This is another ensemble learning approach where trees learn on the loss function of residuals available from a previous tree built and with high weight given to those datapoints inefficiently classified in the previous model. This is constantly repeated with the aim to improve the accuracy.

(B). Generating training and test datasets

We followed the standard data mining procedure to partition data into training and test sets. Specifically, training data were used to build predictive models, and test data were used to evaluate the accuracy of the predictive models derived from training data. Note that the data used in data mining are data specifically collected for building and testing predictive models, i.e., data with known transportation modes and activity types/trip purposes. These data are not the raw sensor data that SmarTrAC collects, but data we collected to build the models for automatically tagging the data that SmarTrAC collects.

For mode classification, features covering both 30-second and 120-second segments are computed every 30 seconds. The 30-second features are therefore computed on non-overlapping segments while the 120-second features are calculated on segments with a 90-second overlap. In this case, a random training-test split for mode classification data could create non-independent training and test sets (i.e., some overlap between training and test sets). To create non-overlapping training and test sets for the 120-second segment data, we used the following partitioning procedure to split segments into training and test sets:

- Group the data into blocks defined by (true) mode of transportation.
- Split each block into two sub-blocks of contiguous segments, and randomly assign one sub-block to the training set and the other to the test set.
- Coarsen the test set sub-block by including only every third segment (note that we do not coarsen the training set).

The final training set for mode classification using the 120-second segments included features computed from 507 segments and the final test set for mode classification included features computed from 173 segments.

For activity type (trip purpose) data, the trips in the 2010 Twin Cities TBI survey were split into a training set consisting of 70% of the trips, and a test set consisting of the remaining 30%.

(C). Evaluating predictive accuracy of classification techniques

Predictions were summarized using the confusion matrix as outputted by the *caret* package in R, which gives the overall predictive accuracy along with the sensitivity, specificity, and positive and negative predictive value. For travel mode, both unsmoothed and smoothed predictions were summarized. We also generated custom visualizations which were used to compare the predicted probabilities (from the original models, i.e., unsmoothed) with true class labels and assess where models had the greatest room for improvement.

4.2.3 Results

(A). Initial comparison of travel mode classification techniques

Table 11 presents information on the overall classification accuracy by classification technique and smoothing options. As shown in Table 11, smoothing prediction results improve the predictive accuracy significantly, regardless of the classification techniques used in prediction. By incorporating the prediction smoothing techniques, SmarTrAC could have an accuracy ranging from 78.6% to 89.8% in predicting instant travel mode.

Out of the four classification techniques, random forests generated the most accurate predictions, followed by gradient boosting, conditional inference trees, and classification and regression trees. Based on these findings, for our application we chose to employ random forests for travel mode classification.

Table 11 Overall classification accuracy by classification technique and smoothing options

	Unsmoothed	Smoothed
Classification and regression trees	69.9%	78.6%
Conditional inference trees	74.0%	83.8%
Gradient boosting	81.5%	88.5%
Random forests	86.3%	89.8%

(B). Initial comparison of activity classification techniques

Tests similar to those for travel mode detection were performed to compare the suitability of a variety of machine learning techniques for the problem of predicting activity type from the Met Council TBI survey data. The random forest classification technique once again performed best, with an overall prediction accuracy of 56% on the test set. The level of accuracy achieved is substantially lower than for mode prediction, but this is unsurprising given the nature of the source data (survey responses which are subject to uncertainty and misclassification). Furthermore, from a user experience perspective, it may not be necessary to have the most likely (predicted) activity match the true activity; it may be acceptable if the correct activity type is among, e.g., the top 3 most likely predicted activity types. Our results show that the correct activity type is among the top 3 most probable predicted activity types 92% of the time.

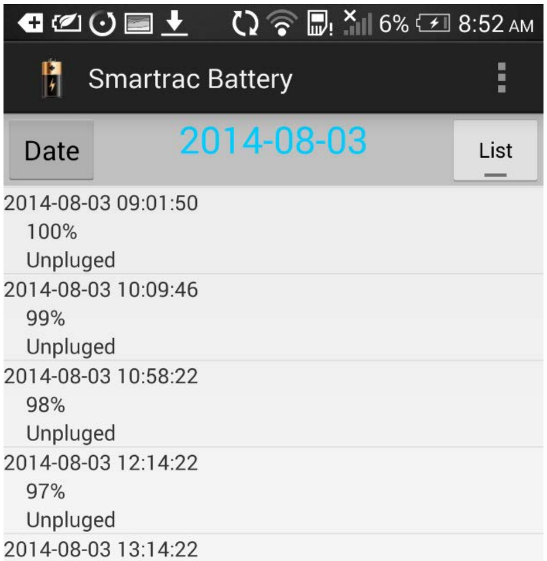
4.3 Final Testing on User Experiences

Laboratory tests at the final stage focused on SmarTrAC's battery consumption rate, data storage requirements, and the ease of data transmission.

4.3.1 Data and approach

A separate app (Figure 16) was developed to test for battery consumption associated with SmarTrAC operations. The app records battery life percentage over time. A SmarTrAC resarcher carried six phones over a 24-hour period (started at 9 am on August 3, 2014) during all waking hours using a multi-pocket belt (Figure 16). The six phones included two HTC phones, two Sony phones, one Google Nexus phone, and one Samsung phone.

Top-half screenshot of the app designed to record battery consumption.



A multi-pocket belt was used by the researchers to carry phones together during tests.

Figure 16. The app desgined and the instrument used for measuring SmarTrAC battery conspumtion

To test for data storage requirement, SmarTrAC Main Database was exported and decomposed into spreadsheet files.

4.3.2 Results

(A). Results on battery consumption

During the 24-hour testing period, the SmarTrAC Data Collection app was on between 9 am and 10:22 pm on August 3, 2014. Each phone had a unique pair of GPS/accelerometer sampling rates at the time of data collection. Three of the phones (HTC1, Samsung, and HTC2) had a zero accelemoter sampling rate, yet varied GPS sampling rate. And the other three phones (Nexus, Sony 2, and Sony 1) had a accelermeter sampling rate of 5 Hertz and varied GPS sampling rate. Figure 17 illustrates the test results, which show that battery life percentage drops sharply with any GPS sampling rates higher than zero. Accelerometer listening consumes much less battery than GPS listening.

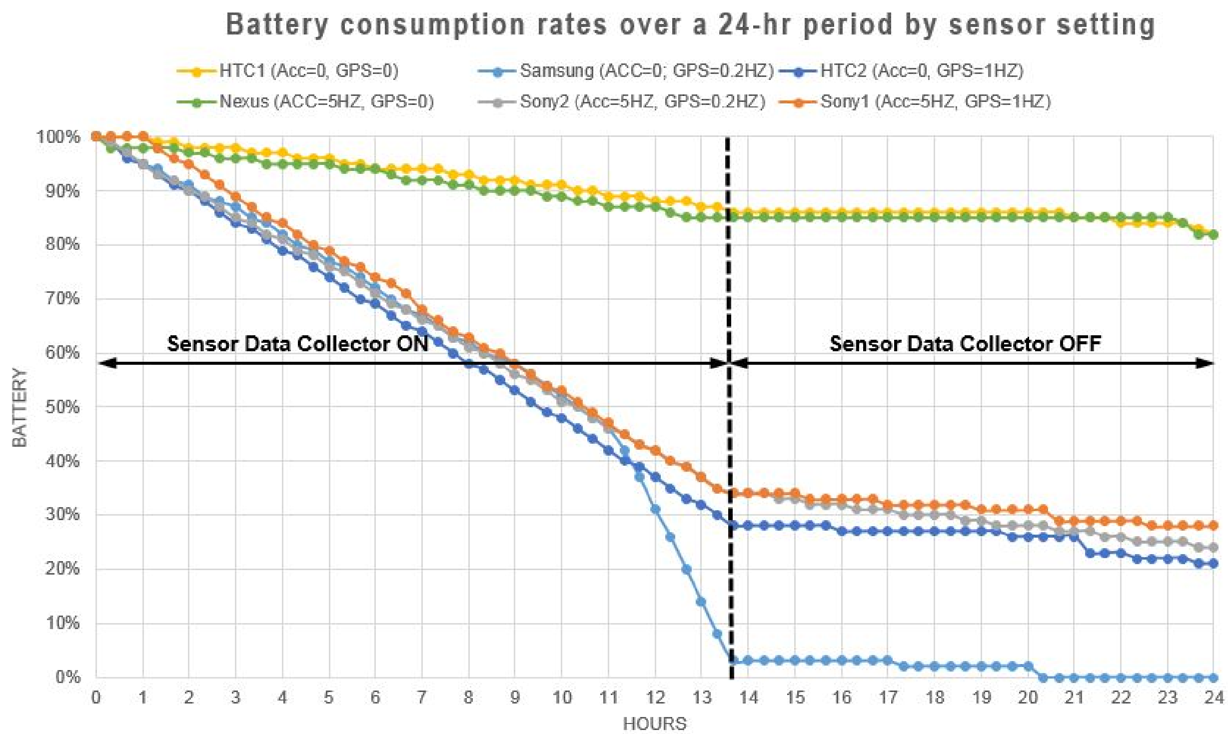


Figure 17. Test results on battery consumption and sensor sampling rates

Table 12 provides more quantified results on the tests. The phones (HTC1 and HTC2) consumed minimal power (1%-1.1% per hour) when GPS sampling was turned off. As long as GPS was turned on, the phones (Samsung, Nexus, Sony1, and Sony2) consumed 4.9%-7% power per hour.

Table 12. Battery consumption rate by GPS/accelerometer sampling rate (% change per hour)

	HTC1 Acc=0 GPS=0	**Samsung Acc=0 GPS=0.2Hz**	**HTC2 Acc=0 GPS=1Hz**	**Nexus Acc=5Hz GPS=0**	**Sony2 Acc=5Hz GPS=0.2Hz**	**Sony1 Acc=5Hz GPS=1Hz**
SmarTrAC Data Collection ON (9am-10:22pm)	1%	7%	5.2%	1.1%	4.9%	4.9%
SmarTrAC Data Collection OFF (10:22pm-8:59am)	0.4%	0.5%	0.8%	0.3%	1%	0.7%

The tests show that phone battery life is dramatically shortened by GPS data collection and is only modestly shorten by accelerometer data collection. Once GPS and accelerometer listening is turned on, sampling rates and file writing rates play a minimal role in influencing battery consumption. Based upon the testing results, the researchers developed a mitigation strategy to reduce battery consumption of SmarTrAC: having accelerometer run continuously in the background and using the accelerometer readings to stop GPS sampling during periods of detected inactivity (i.e., non-movement).

(B). Results on data storage and transmission requirements

Data storages tests show reasonable storage requirements of SmarTrAC. SmarTrAC produces 49.65 Mbytes of data over a 24-hour period (Table 13). The two largest tables (table_motions and table_locations) in SmarTrAC are the tables containing raw sensing data. The table_motions table contains accelerometer readings 5 times per second (5Hz). The table_locations table contains per-second GPS readings. Together, these two files occupy about 97.43% of the total data storage required by SmarTrAC.

Table 13. Size of SmarTrAC data over a 24-hour period

Data tables	Size (Mbytes)	Percentage (%)
Raw sensor data		
table_motions	43.58	87.76
table_locations	4.80	9.66
Instant prediction/processing results		
table_intermediate_locations	1.15	2.31
table_dwellings	0.08	0.16
table_modes	0.04	0.09
Summary data for facilitating user interactions	0.01	0.02
Total	49.65	100.00

Given that SmarTrAC produces 50 megabyte of data per day, SmarTrAC requires 350 megabytes of data storage space for a seven-day participation. To avoid situations that SmarTrAC consumes significant storage space on the user's phone (e.g., in cases that users keep SmarTrAC running for more than seven days), the research team implemented an automatic data cleaning mechanism. Specifically, SmarTrAC is designed to automatically delete raw senor data (the table_motions and table_locations files as shown in Table 13) that are more than 7 days old. SmarTrAC also has an automatic reminder mechanism which reminds the user to upload their data to the cloud sever weekly. The associated weekly data transfer needs are roughly 150 megabytes after data compression.

CHAPTER 5. FIELD TESTS

Two 7-day field tests were conducted to test the effectiveness of SmarTrAC in terms of data collection, processing, and ease of use. This chapter includes 5 sections: (1) study design; (2) study implementation; (3) tracking and addressing field test issues; (4) data analysis; (5) analysis of web-based exit survey data.

5.1 Study Design

The primary objective of the 7-day field tests was to test SmarTrAC for its feasibility and reliability in terms of tracking, collecting, and processing participant trip and activity data. In addition, we wanted to test the app's impact on the smartphones and its ease of use for participants. For both field tests, participants were asked to install the application on their phone and carry it with them at all times during the 7-day field test. In addition, the participants were assigned 3 daily in-application tasks. These included: (1) reviewing data collected at the end of each day and editing, adding details, and confirming the collected information; (2) uploading the confirmed information using SmarTrAC's built-in upload function; (3) reporting any bugs/unusual behavior (if encountered) through the application's bug report function. To ensure compliance with daily tasks, the project team sent out daily reminders to the participants at the end of each day to complete them. For both field tests, all data uploaded by the users was stored online using Amazon Web Services' S3 (scalable storage in the cloud) tool.

To help participants complete these tasks and to make using the application easier, the participants were provided with access to 9 online video tutorial modules (available on the SmarTrAC website). The video modules provided instructions on the following topics: getting started, application navigation, editing trips via calendar view, editing trips via map view, editing activities via calendar view, editing activities via map view, clearing changes, uploading data, and reporting issues.

To keep track of the issues participants faced during the field tests and to record their experience using SmarTrAC, a daily field test log (see Appendix B) was used where participants were asked to give details regarding their smartphone performance, unusual application behavior, and crash reports at the end of each day. Once the participants completed the 7-day field test, they were asked to fill out an exit survey (see Appendix F) based on their experience and uninstall SmarTrAC from their phones. Information collected through the daily field test log and exit survey was used to make improvements to and fix issues with the application. Improvements made to SmarTrAC and findings for the exit survey are discussed in greater detail later in this chapter.

5.2 Study Implementation

Participants were recruited using convenience sampling and comprised of members from the project team's own social network and no compensation was provided for participation. Once potential participants who were willing to participate in the study were identified, they were sent an introduction email with a brief description of the project and asked to fill out general information survey (see Appendix A) through which information regarding their phone (make, android firmware, battery life and IMEI number for data upload tracking) was collected. Finally, participants were selected for the field tests if they had android smartphones with the Android version of 4.0.4 or later.

Once the eligible participants were selected, they were sent all the information required to participate in the field test via email. This information included: (1) the start and end dates for the field test; (2) their daily tasks for the field test; (3) an information sheet (see Appendix C) with details about data confidentiality, the applications burden on their phone (e.g., battery, memory, data, etc.) and their overall responsibilities during the field test; (4) a Dos and Don’ts sheet (see Appendix D) to ensure that the application was used correctly; (5) a link to the SmarTrAC training videos; (6) the SmarTrAC installation file and installation instructions (see Appendix E).

The first 7-day field test was held between Thursday, January 8 and Wednesday, January 14, 2015. A total of 13 participants started the field test; however, only 9 participants completed it. Of the 4 that did not complete the field test 2 were due to phone issues that were unrelated to SmarTrAC, one was unable to install the application due to firmware issues, while the other had issues with the phone’s GPS. The other two participants dropped out of the study due to personal reasons.

The second 7-day field test was held between Thursday, January 22 and Thursday, January 29, 2015. Twelve participants started the field test and 8 participants completed it. Of the 4 that did not complete the test, 2 were unable to do so due to faulty GPS recorders on their phone, 1 due to general phone malfunction, and 1 dropped out of the study due to personal reasons.

In total, 17 participants completed the two 7-day field tests. Of the 17 participants that completed the field tests, 9 (53%) were between the ages of 21 - 30 years, 7 (41%) were between ages of 31 - 40 years, and 1 (6%) was between 41 - 50 years. Nine participants (53%) reported being full-time students, 9 (53%) reported being employed full-time, and 2 (12%) being employed part-time (percentages add up to more than 100 as it was possible to select more than one option). Eleven unique phone models were used in the field tests, these included, HTC One, HTC One M7, Google Nexus 5, Google Nexus 4, Sony Experia, Motorola Moto G, Motorola Moto X, Samsung Exhibit, Samsung Galaxy S4, Samsung Galaxy S3, and Samsung galaxy S5. The median battery life reported at the beginning of the survey without SmarTraAC running was 10-12 hours.

Upon further investigation of the data uploaded by the participants after the field tests were completed, it was discovered that only 15 of the 17 databases created were usable. Of the 2 databases that were excluded, one was due to a faulty phone accelerometer which led to SmarTrAC not capturing travel data and the second was because the participant failed to adequately follow field test instructions during testing.

5.3 Tracking and Addressing Field Test Issues

To a keep a track of application behavior and issues encountered by participants the study team used 3 tools, the built-in crash report generator, the built-in bug/unusual behavior reporting option, and the daily field test log. Through the field test log, participants were able to provide the study team with additional detail on crashes and bugs and also provide crash information in case the built-in crash report function did not automatically generate a report. The field test log also provided information on battery consumption that allowed us to compare battery use before and after the use of SmarTrAC.

5.3.1 Battery consumption

Of the 17 participants that completed the field tests, only 15 filled out the daily field test log at least once to enable a comparison of battery consumption with and without SmarTrAC. Battery consumption rates recorded in the general information collected before the field test was compared with battery consumption rates reported in the daily field test log during both field tests. Before the use of SmarTrAC, all participants' phones (100%) had a battery life longer than 6 hours and 94% of the phones had a battery life longer than 8 hours. With SmarTrAC running continuously, 74% of the phones had a battery life longer than 6 hours, and about half of the phones (47%) had a battery life longer than 8 hours. Overall, the field tests indicate that due to the need for accurate GPS data and the associated sampling frequency of GPS data, SmarTrAC does reduce battery life of phones in most cases. The battery consumption rate of SmarTrAC varies from phone to phone. Besides the phone brand and model, additional factors such as non-SmarTrAC-related phone usage may explain the variation in the battery consumption rate. For example, between two identical Sony Experia phones were used for the field tests, one did not show any change in battery life, the other saw a decrease of up to 2 hours in battery life.

5.3.2 Crash reports and bugs reported

Based on the crash reports, user-generated bug reports and the daily field test log the study team tracked and identified issues that needed to be addressed during the field test. During the first field test, 8 crash reports and 12 bug reports were generated. During the second field test, 5 crash reports and 1 bug report were generated. A summary of the issues identified through the reports and the daily field test log is given in Table 14. Almost all the SmarTrAC crashes and associated issues reported by users occurred in the editing module of the application (issues 1-6 in Table 14). After identifying the issues the application design team re-wrote the entire editing module of the application and tested it to ensure that all the issues had been resolved. During the first field test, a number of participants reported bugs and other issues with the upload function's user interface (issues 7-9 in Table 14). The upload interface was redesigned to take into account all reported issues and the new interface was tested in the second field test where no bugs or issues were reported. On some phones, the accelerometer recording capabilities were affected by the phone going into sleep-mode (when not used or moved for some time). This issue (issue 10 in Table 14) was fixed by implementing a wake-lock function that did not allow the phone's processor to go to sleep while SmarTrAC was running. Finally, a number of participants reported that crash reports were not being generated automatically by SmarTrAC, especially in cases where the application crashes occurred while it was running in the background (issue 11 in Table 14). To fix this concern the application design team modified the crash report module to ensure crash reports were generated each time. In addition, the team also implemented a function that saves all crash reports in case the user is unable to send them out instantly and enables the user to send them out at a more convenient time.

Table 14. Field test issues, issue descriptions and solutions

ISSUE	DESCRIPTION	SOLUTION
1. Lost trip segments	Editing a trip caused other exiting trips to disappear from calendar view in some cases.	Fixed by rewriting the edit module to eliminate all editing related issues.
2. Cluster rendering issue	In some cases, SmarTrAC crashed during map navigation when there were multiple activity locations on map.	Fixed by updating the google map utility library.
3. Google map cluster	Some users reported that SmarTrAC crashed when they zoomed in and out frequently in map view.	Fixed by updating the google map utility library.
4. Trip editing crash	In some cases it was found that when users selected the "wait as a part of a trip" option the application crashed.	Fixed by rewriting the edit module to eliminate all editing related issues.
5. Activity selection	SmarTrAC crashed in some cases when users attempted to save changes without selecting a predicted activity type.	Fixed by automatically saving activity as "unknown" if no user selection was made.
6. Nearest places null pointer	In some cases, SmarTrAC crashed when nearest places details were not received from the Google API.	Fixed by exiting prediction (not calibrating) when nearest places details were not received.
7. Upload progress bar	A number of users reported issue with the upload data progress bar not providing real-time upload status.	Fixed by redesigning the upload progress bar to provide real-time upload status.
8. Upload time	For all users, data upload was taking more than a few minutes.	Fixed by compressing data by approximately 60 % before upload to reduce upload time.
9. Upload notification	Some users reported that the upload confirmation message did not appear or only flashed briefly when data upload was complete making them doubt if the data had been uploaded.	Fixed by redesigning the upload function to include a data upload confirmation that stays on the screen till the user clicks "OK"
10. Sleep and accelerometer	On some phones it was found that the accelerometer stopped working when phone went into sleep mode thereby limiting data collected.	Fixed by applying a wake-lock to keep phone processor awake whenever SmarTrAC is running
11. Crash report not generated	In many cases it was found that SmarTrAC did not generate crash reports automatically, particularly when the application was running in the back ground.	Fixed by modifying the crash report module and saving all crash reports that the user is unable to send instantly so they can be sent at a more convenient time.

5.4 Analysis of Field Test Data Generated by SmarTrAC

Data from the field test were summarized and analyzed to assess the success of the field test procedure and to estimate the predictive accuracy of the machine learning algorithms employed by SmarTrAC. Individual phone databases were downloaded from Amazon S3 cloud storage, and extracted, combined, and analyzed using the open-source statistical software package R.

5.4.1 Field test data summary

The field test provided data on a total of 649 activities and 216 trips (742 individual trip segments) from 15 individuals. The grid of plots in Figure 18 shows the trip and activity segments for each of the individuals who provided field test data.

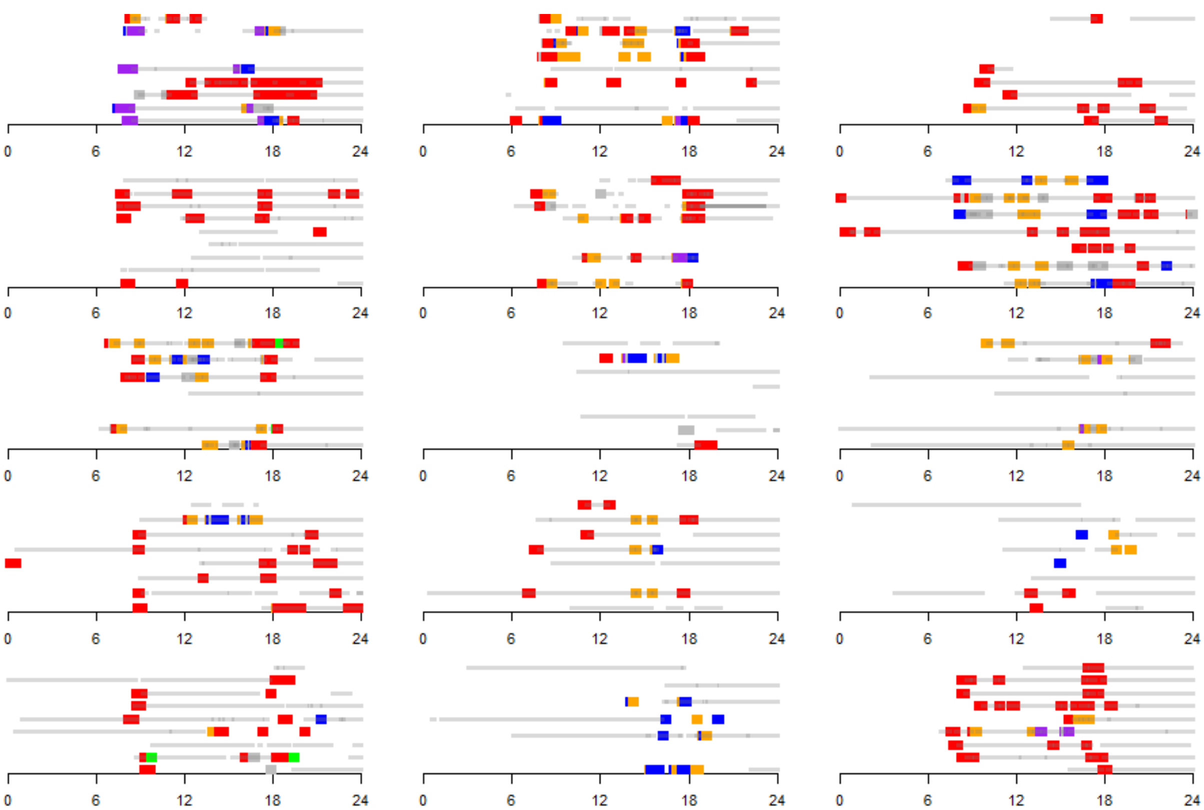


Figure 18. Daily trips and activities for field test participants

Note: Each separate plot shows the trips and activities from a single participant; trips and activities from the same day are connected horizontally. Colored boxes correspond to trips (Red = Car, Blue = Bus, Orange = Walking, Green = Bike, Purple = Rail, Gray = Wait). Thinner lines represent activities.

Table 15 summarizes the durations and distances of trips and trip segments (by mode). We note that the field test provided ample data on travel episodes using Car, Bus, and Walking, but there were relatively few Bike and Rail episodes recorded. Table 16 summarizes the durations of activities for each of the 7 defined activity types (plus activities of Unknown type). As with trip segment modes, some activity types (Home and Work, each about 30% of activity episodes) were detected frequently but there was little data on others (e.g., Education, only 2 episodes during the field test period).

Table 15. Characteristics of trips and trip segments

	N (%)	Median (IQR) duration in minutes	Median (IQR) distance in miles
Complete trips	216 (100)	26.1 (33.5)	10.5 (16.5)
Trip segments			
Car	266 (35)	16.1 (14.7)	6.80 (13.90)
Bus	58 (7)	9.6 (10.8)	1.61 (2.80)
Rail	19 (2)	17.5 (20.3)	3.77 (6.63)
Bike	15 (2)	1.0 (1.5)	0.13 (0.13)
Walking	226 (30)	3.0 (4.6)	0.19 (0.35)
Wait	158 (21)	1.0 (3.0)	0.07 (0.16)
Total	742 (100)	5.6 (13.1)	0.72 (4.36)

Table 16. Characteristics of activities

Activity type	N (%)	Median (IQR) duration
Unknown	95 (14)	12 (37)
Home	213 (32)	360 (705)
Work	199 (30)	73 (189)
Education	2 (0)	68 (8)
Shopping	34 (5)	11 (16)
Eat out	44 (6)	39 (65)
Other personal business	39 (6)	13 (21)
Soc/rec/ent/comm	23 (3)	51 (88)
Total	649 (100)	60 (254)

Figure 19 and Figure 20 display the distribution of start times of trip segments (by mode) and activity episodes (by type). The shaded curves are estimated density plots (i.e., smoothed histograms), with peaks corresponding to times of day when trip segments or activities are more likely to start. Figure 19 displays expected patterns: car, bus, and rail segments peak in frequency around the traditional rush hour windows (7-9 am and 4-7 pm) while bike and walking segments are more evenly dispersed throughout the day. Figure 20 confirms that eating out typically occurs around noon and 6 pm, shopping activities are concentrated in the late afternoon and evening hours, while other personal business appears to be conducted during periods corresponding to the morning and afternoon commute. Due to the limited sample size available, these plots combine data from weekdays and weekends, but one would expect to see different patterns emerge if these periods were analyzed separately.

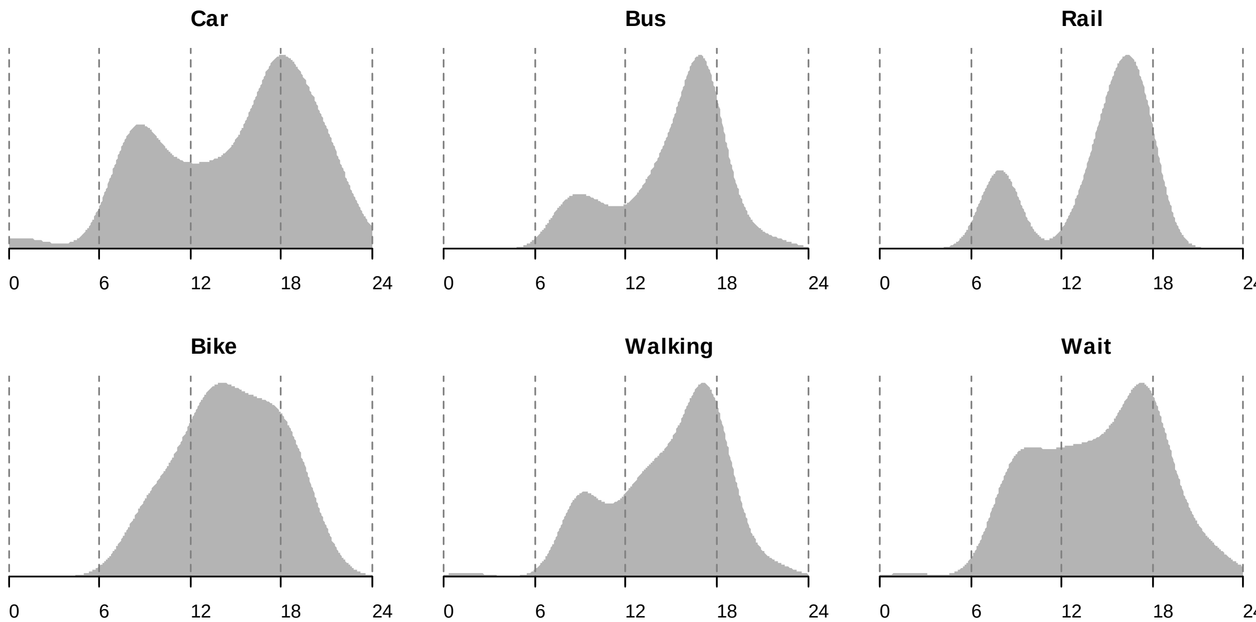


Figure 19. Distribution of trip segment start times, by mode

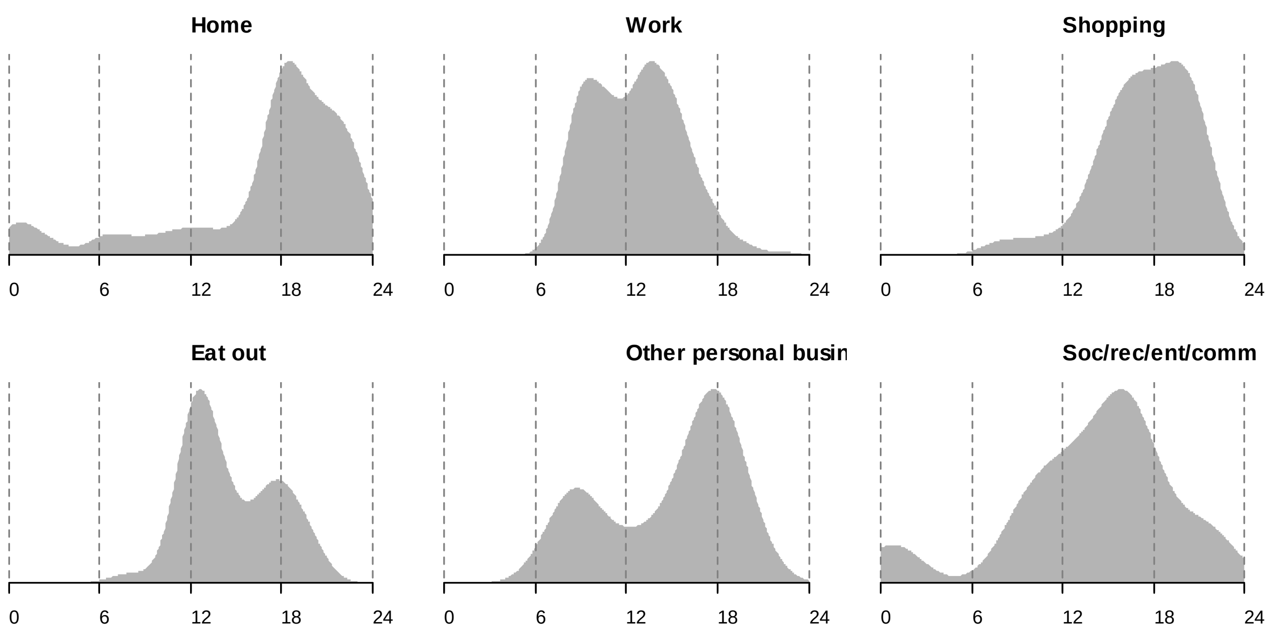


Figure 20. Distribution of start times, by activity type

5.4.2 Assessing the predictive performance of SmarTrAC algorithms

The data mining (machine learning) algorithms that form the basis for SmarTrAC's travel mode predictions were developed using a small set of pilot data collected in the summer of 2014. Activity predictions were made on the basis of models developed using data from the 2010 Travel Behavior Inventory survey conducted by the Twin Cities Metropolitan Council. The field test therefore provided both an important opportunity to evaluate the performance of these algorithms, and to collect data which can be used to improve them in the future. In the following sections, we summarize the success of SmarTrAC algorithms in carrying out three key tasks:

1) Separating activities and trips, i.e., identifying when an individual is engaging in an activity at or around a particular location, and when they are traveling between locations.
2) Predicting mode of travel during trips.
3) Predicting activity types following activity completion.

(A). Separating activities and trips

Figure 21 shows 100 randomly selected transition points when a trip ends and an activity begins; Figure 22 is similar, but shows points when an activity is ending and a trip beginning. The overlaid density plot represents the distribution of predicted transition points. As can be seen, SmarTrAC detects transition points with a high degree of accuracy; indeed, across all trip to activity transition points, SmarTrAC predicts a change within ± 30 seconds 88% of the time. For activity to trip transition points, the change is correctly predicted within ± 30 seconds 91% of the time.

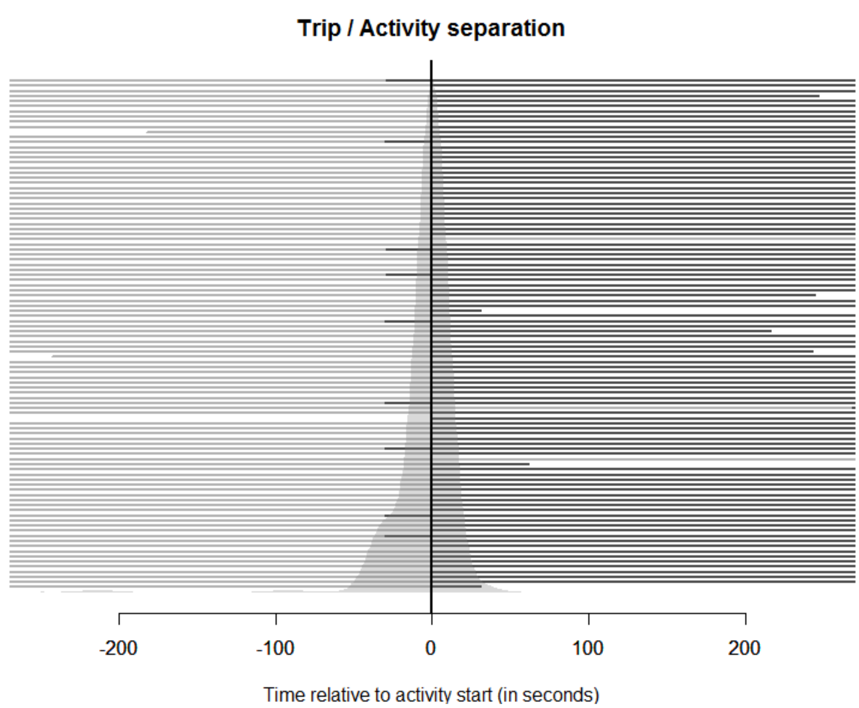


Figure 21. Trip / Activity separation

Notes: Gray and black horizontal lines are predicted trip and activity segments, respectively. The vertical zero line represents the true start time of the activity. Overlaid in gray is a smoothed density plot of the predicted activity start time.

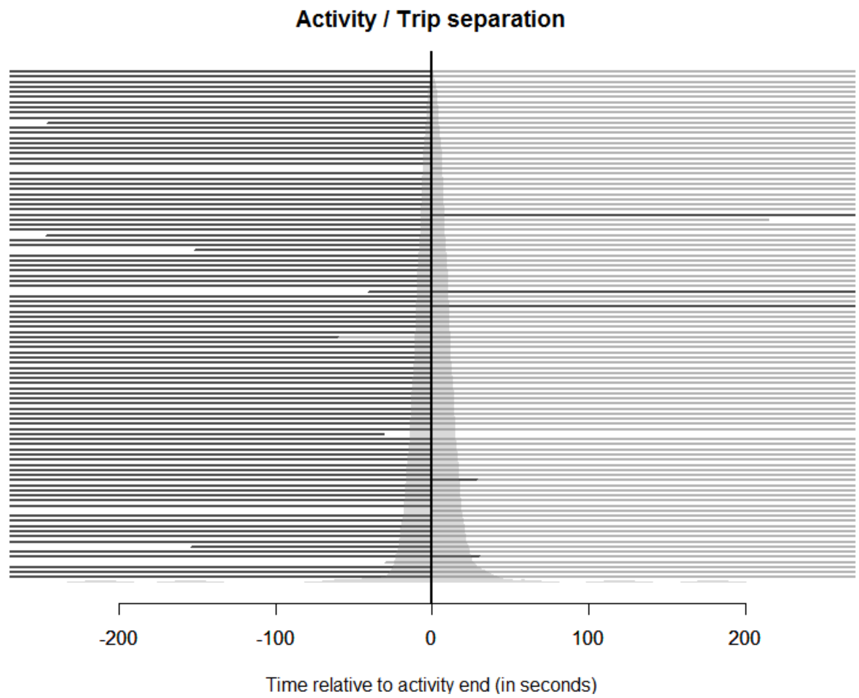


Figure 22. Activity / Trip separation

Notes: Gray and black horizontal lines are predicted trip and activity segments, respectively. The vertical zero line represents the true end time of the activity. Overlaid in gray is a smoothed density plot of the predicted activity end.

(B). Predicting modes of travel

Travel modes are predicted for each 30-second trip period during a trip, and these predictions are aggregated to predict the mode for an entire trip segment. Here, we summarize the predictive accuracy at both levels. Table 17 cross-tabulates the predicted and true (user-provided) for 30-second periods, and summarizes the sensitivity, specificity, positive and negative predictive values (PPV and NPV), and balanced accuracy of the SmarTrAC mode predictions. For each mode M, these performance metrics are calculated as follows:

$$Sensitivity = \frac{\#\ of\ periods\ predicted\ as\ mode\ M}{true\ \#\ of\ periods\ of\ mode\ M}$$

$$Specificity = \frac{\#\ of\ periods\ predicted\ as\ NOT\ mode\ M}{true\ \#\ of\ periods\ NOT\ of\ mode\ M}$$

$$PPV = \frac{\#\ of\ periods\ where\ predicted\ and\ true\ mode\ are\ both\ M}{\#\ of\ periods\ predicted\ as\ mode\ M}$$

$$NPV = \frac{\#\ of\ periods\ where\ predicted\ and\ true\ mode\ are\ both\ NOT\ mode\ M}{\#\ of\ periods\ predicted\ NOT\ of\ mode\ M}$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2}$$

Table 17. True and predicted modes, by 30-second trip period

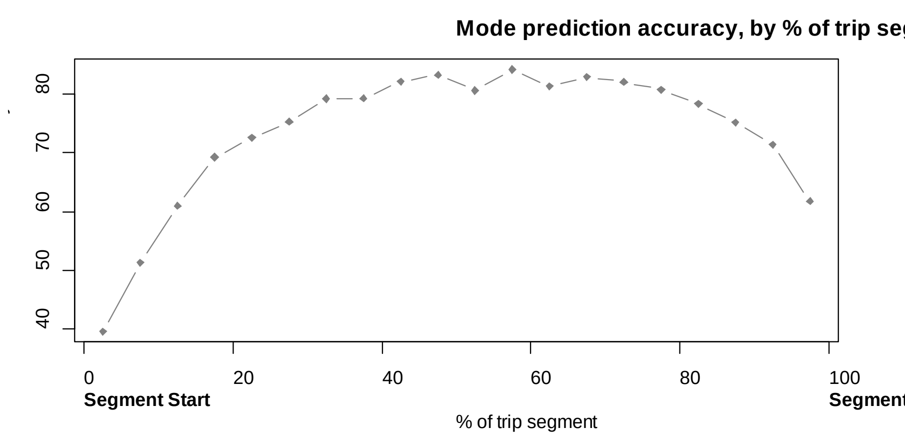
Predicted Mode	**True Mode**					
	Bike	*Bus*	*Car*	*Rail*	*Wait*	*Walking*
Bike	**46**	14	29	9	3	14
Bus	1	**1166**	1924	208	46	25
Car	0	312	**7004**	260	6	4
Rail	0	44	434	**240**	6	14
Wait	0	4	108	23	**798**	55
Walking	4	51	140	14	115	**1688**
Sensitivity	0.9	0.73	0.73	0.32	0.82	0.94
Specificity	1	0.83	0.89	0.96	0.99	0.98
PPV	0.4	0.35	0.92	0.33	0.81	0.84
NPV	1	0.96	0.64	0.96	0.99	0.99
Balanced Accuracy	0.95	0.78	0.81	0.64	0.9	0.96
Overall Accuracy	**0.74**					

Note: Bold numbers in the top part of the table indicate correct predictions.

Generally, SmarTrAC is quite successful in mode prediction. Mode detection sensitivities are above 70% for all but periods of mode Rail (which are frequently misclassified as Bus and Car), and the false positive rate (1 – specificity) is less than 20% across all modes, and less than 5% for all but Bus and Car. One weakness of the current algorithm is that it tends to overpredict the occurrence of Bike, Bus, and Rail periods, so that the positive predictive value (i.e., the probability that the true mode is M given that SmarTrAC predicts mode M) of these modes is less than 50%. The overall accuracy is 74%, indicating that approximately 3 out of every 4 30-second periods are predicted correctly. As we demonstrate below, by aggregating 30-second period predictions we are able to achieve even better accuracy for trip segments.

The estimate of overall accuracy for mode prediction at the 30-second period level ignores the fact that there is substantial variation in predictive accuracy as a trip progresses. As shown in Figure 23, accuracy is often relatively low over the first 10-20% of a trip segment, but increases to above 80% in the middle 60% of the segment before declining modestly in the last 20% of the trip segment.

Figure 23. Travel mode prediction accuracy for 30-second periods, by % of trip completed



The phenomenon in Figure 23 is mostly due to the fact that trips may consist of multiple segments using different modes of travel, and SmarTrAC faces the challenging task of identifying when the mode of travel changes during a trip. Improving predictions around these mode change points will be a point of emphasis in the future development of SmarTrAC.

Table 18 provides the same information as Table 17, but for predictions made at the trip segment level rather than for each 30-second period. As noted above, predictions are more accurate at this level than at the 30-second period level: Sensitivity is above 60%, specificity is above 94% (corresponding to a false positive rate below 6%), and the positive predictive value is greater than 50% across all modes. Bike, Bus, and Rail continue to be predicted somewhat more frequently than they actually occur (yielding smaller PPVs), but the overall accuracy rate of 86% shows that in nearly 9 out of every 10 segments, SmarTrAC accurately predicts the mode of transportation used during that segment.

Table 18. True and predicted modes, by trip segment

Predicted Mode	**True Mode**					
	Bike	*Bus*	*Car*	*Rail*	*Wait*	*Walking*
Bike	**15**	1	3	1	0	4
Bus	0	**55**	38	3	0	1
Car	0	1	**184**	2	0	0
Rail	0	0	7	**12**	0	0
Wait	0	0	18	0	**158**	9
Walking	0	1	16	1	0	**212**
Sensitivity	1	0.95	0.69	0.63	1	0.94
Specificity	0.99	0.94	0.99	0.99	0.95	0.97
PPV	0.63	0.57	0.98	0.63	0.85	0.92
NPV	1	1	0.85	0.99	1	0.97
Balanced Accuracy	0.99	0.94	0.84	0.81	0.98	0.95
Overall Accuracy	**0.86**					

Note: Bold numbers in the top part of the table indicate correct predictions.

(C). Predicting activity types

When the SmarTrAC app is first opened, users are asked to provide their Home and Work addresses, and subsequently SmarTrAC will predict Home and Work activities when the user is sufficiently close to these locations. Hence, the accuracy of predictions for these activity types is very high, exceeding 95%. Also, when a user identifies that a particular activity took place at a specific location, SmarTrAC saves that information to its internal database and future activities in that location are predicted to be of the same type (if a multiple activity types are associated with the same location, SmarTrAC will predict the activity which has been undertaken most frequently at that location). Since location-activity pairings are typically unique, activity type prediction is usually very accurate when a SmarTrAC user returns to a previously identified location.

The most challenging activity type prediction scenario, then, is when a prediction must be generated for an activity which took place at a new (to SmarTrAC) location. We employ a random forest model which incorporates information about the activity itself (start time, duration, day of week, etc.), the trips and activities preceding it, and points of interest in the vicinity of the location where the activity took place. The model was built using data from the Travel Behavior Inventory survey conducted in 2010 by the Metropolitan Council. There were 120 “de novo” activity predictions made during the SmarTrAC field test, and this is the set upon which we base our accuracy metrics.

SmarTrAC’s random forest model gives probabilities that an activity is one of five types (Eat out, Education, Shopping, Social/recreation/entertainment/community, or Other personal business; Home and Work are not predicted via the random forest as they are typically associated with previously visited locations). For many of the field test activity episodes, the most likely activity had a predicted probability less than 0.5. Hence, using the most likely activity as the prediction yielded relatively poor predictive accuracy, about 25-40%. However, the true activity type was among the *two* most probable activities approximately 70-80% of the time (versus 40% expected by random guessing) and among the *three* most probable activities 80-95% of the time (versus 60% expected by random guessing). This result suggests that SmarTrAC’s strategy of sorting predicted activity types by their predicted probability puts the true activity type in the top three the vast majority of the time.

5.5 Analysis of Web-Based Exit Survey Data

The SmarTrAC exit survey (see Appendix F) was sent out to all participants on the final day of each of the two field tests. The exit survey collected feedback on 3 facets of the field tests including: (1) user experience; (2) user responsibilities; (3) application performance. The data from the exit survey is intended to highlight the strengths and weaknesses of the SmarTrAC.

5.5.1 User experience

A majority of the participants reported being satisfied with the field tests. When asked to report their overall satisfaction with using the application 10 (59%) respondents reported being “satisfied”, 5 (29%) reported being “neither satisfied nor dissatisfied”, and only 2 (12%) reported being “dissatisfied”. See Figure 24.

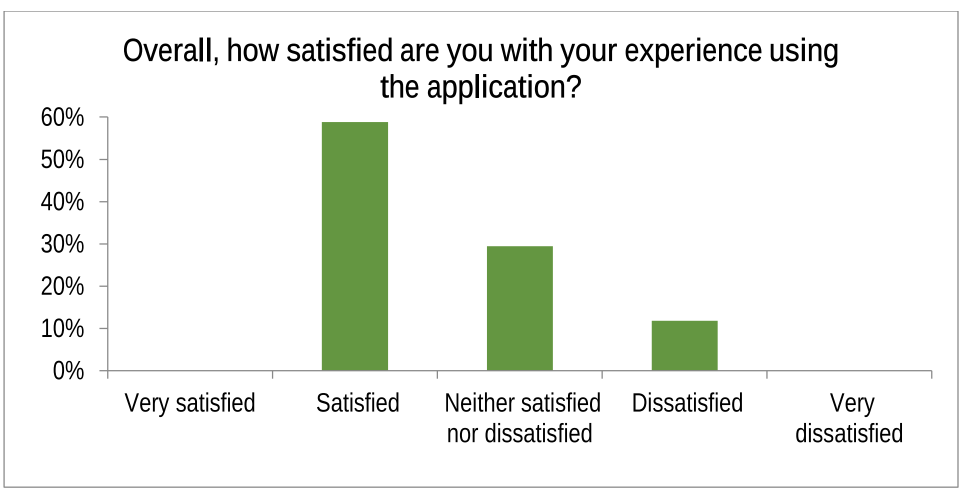


Figure 24. Overall satisfaction with the application

Participants were also asked to what extent the agreed or disagreed with statements about the effectiveness of 8 components of the field test. For non-application related components of the field tests, responses were very positive with between 14 (82%) to 16 (94%) reporting that they “strongly agree” or “agree” that communication with the research team was helpful, instruction on how to use SmarTrAC were helpful, and installing the application was easy. In terms of the overall navigation experience with SmarTrAC, 13 (76%) participants agreed that it was easy to navigate. While most participants agreed that study components functioned well or were neutral, the trip and activity detection questions received some negative feedback. When asked if all their activities were detected 6 (35%) respondents disagreed or strongly disagreed. For all trips being detected, 3 (18%) of the respondents disagreed or strongly disagreed. See Figure 25.

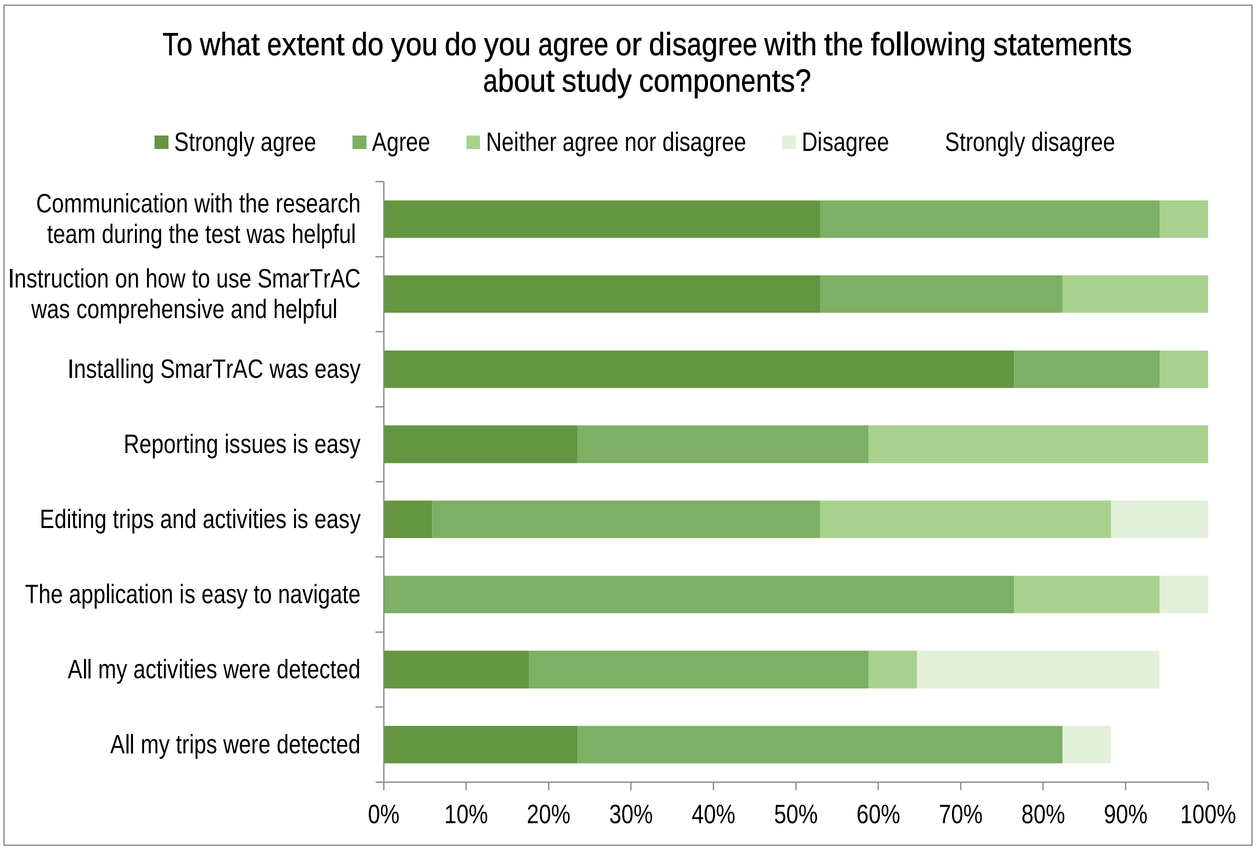


Figure 25. Evaluating study components

When asked whether using SmarTrAC had made them more aware of their travel and activity behavior, 14 (82%) of the respondents selected “agree” or “strongly agree”, while only 1 (6%) respondent selected “strongly disagree”. See Figure 26.

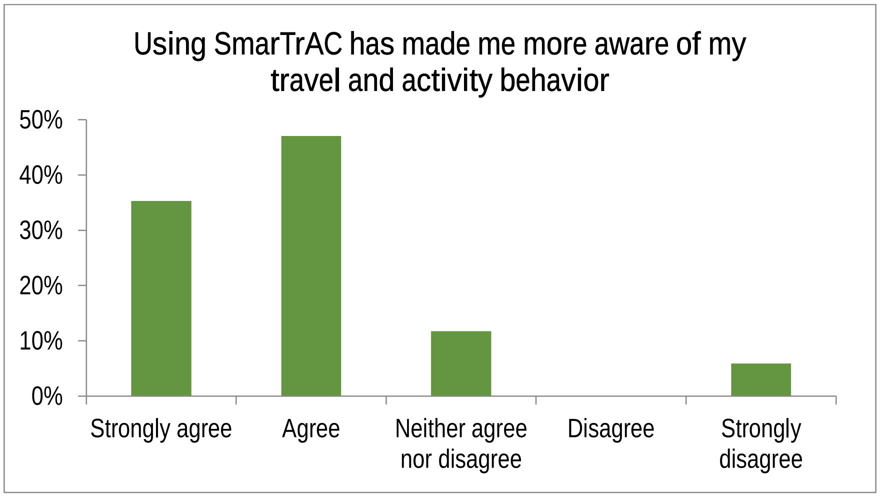


Figure 26. Awareness of travel and activity behavior

Overall, participants reported being comfortable with SmarTrAC tracking the details of their daily lives. Thirteen (76%) users indicated that that they were comfortable (“agree” or “strongly agree”) with SmarTrAC tracking their trips, while 4 (24%) indicated they were not (“disagree” or “strongly disagree”). For tracking activity locations, 12 (71%) indicated that they were comfortable, while 5 (29%) reported that they were not. For adding personal details to trips and activities (e.g., companionship, activity details, etc.), 10 (59%) indicated they were comfortable, while 3 (18%) said they were not. See Figure 27.

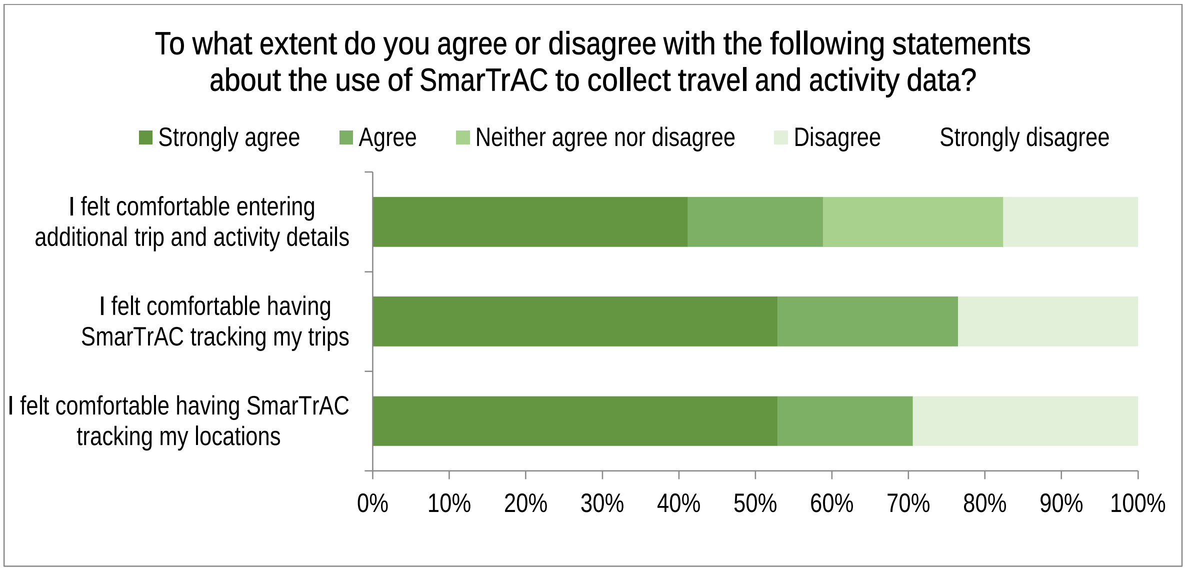


Figure 27. Feelings towards daily activity and trip tracking

5.5.2 User responsibility

To gauge convenience of use, participants were asked how burdensome various tasks associated with the field test were for them. The biggest issue identified by users was related to phone battery life. Thirteen (76%) of the respondents reported that keeping their phone charged during the field test was either “very burdensome” or “burdensome”. Other tasks that were reported as potential issues by a few users were editing trips and activities, which 4 (24%) respondents reported as “burdensome”; adding details to trips and activities, keeping WiFi activated, and filling out the daily field were all reported as “very burdensome” or “burdensome” by 3 (18%) of the participants. See Figure 28.

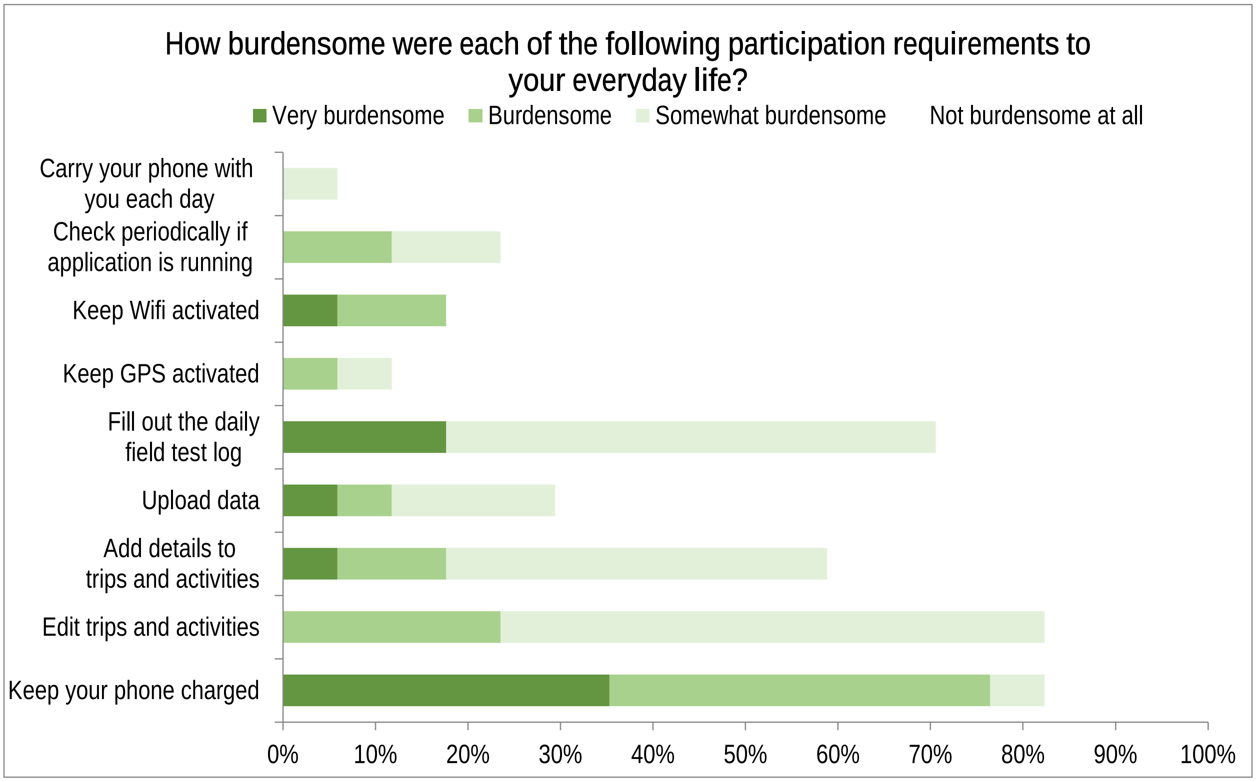


Figure 28. User burden

To ensure proper functioning of SmarTrAC, the users were provided with a “Dos and Don’ts” sheet (see Appendix D) and assigned daily tasks. To check compliance with these requirements, the participants were asked how often they were able to complete 6 tasks associated with the 7-day field test. Between 12 (70%) and 16 (94%) of participants reported completing all tasks on “4 to 6 days” during the tests. For editing and adding details to trips and activities, 5 (30%) participants reported completing the task on “2 to 3 days” or less. See Figure 29.

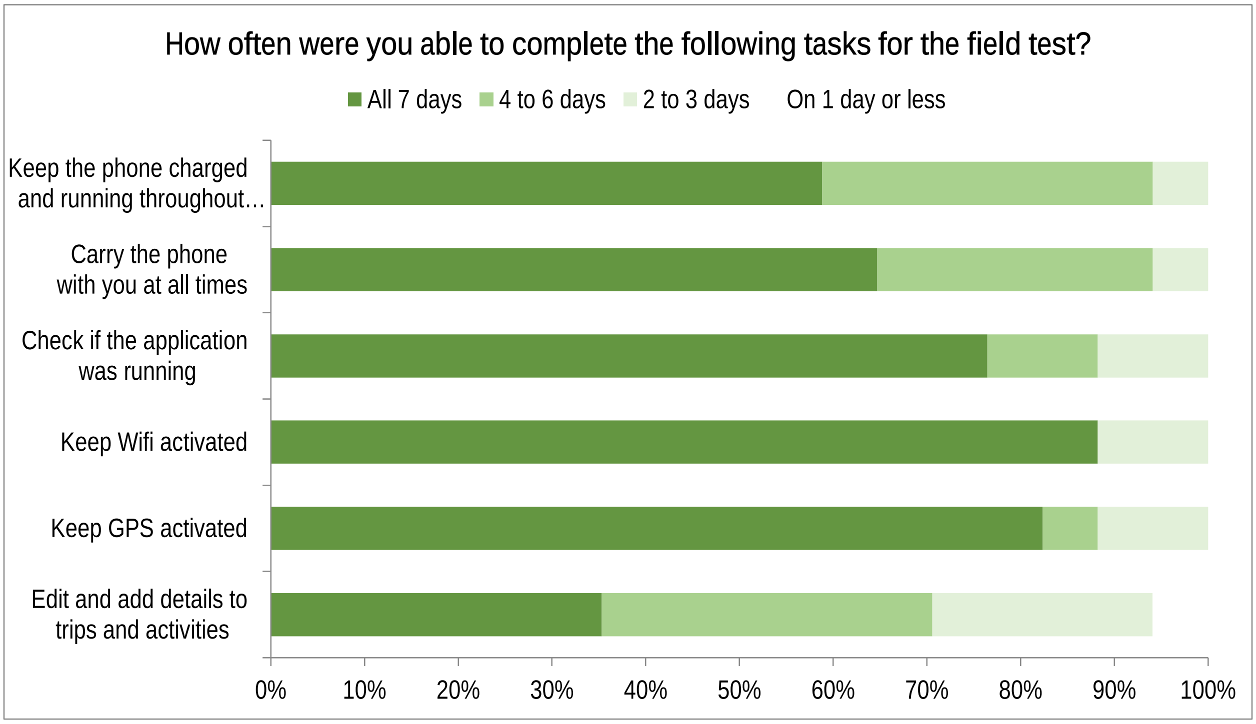


Figure 29. User compliance with field test tasks

5.5.3 Application performance

To gauge the accuracy of SmarTrAC in processing and displaying data, participants were asked to evaluate the application's performance in executing 6 functions. For two of these functions, start and end times of trips and activities being detected accurately and trip routes being accurately detected and mapped, 15 (88%) participants reported the application functioned accurately "always or almost always" or "most times". For "locations and activities were accurately detected and mapped" and "actual activity was among the top three predicted activity types", 13 (76%) and 10 (59%) participants reported the application functioned accurately "always or almost always" or "most times" respectively. Functions where users reported SmarTrAC did not function accurately were generation of crash reports and detection of travel modes, where 8 (47%) and 5 (30%) participants reported the application functioned accurately "a few times" or "almost never or never". See Figure 30.

The exit survey also included 2 open-ended questions. The first question asked participants what they enjoyed the most about the application. Of the participants that answered the question, 6 identified the accuracy of activity and trip detection, 6 identified being able to look at their whole day in calendar view, and 1 identified having the start and end times of activities as the best part about the application. The second question asked participants what features of SmarTrAC needed improvements. Of the participants that answered the question, 5 identified battery consumption, 3 identified the user interface in general, 2 identified the complexity of editing activities, and 2 identified application crashes while editing as areas that could be improved.

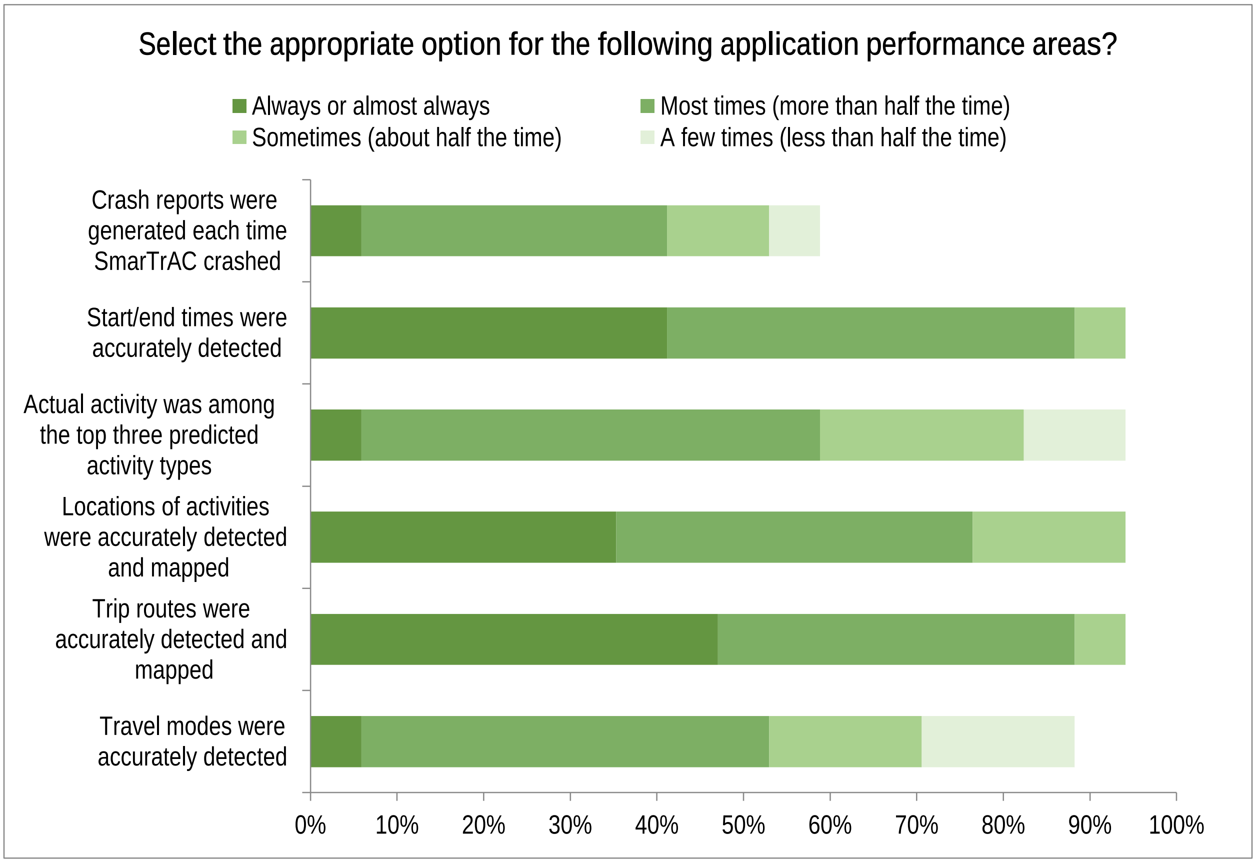


Figure 30. Evaluating application functions

CHAPTER 6. CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Project Conclusions

We have developed SmarTrAC, a robust Android-based smartphone application that collects detailed and multi-dimensional data on daily activity and travel behavior. *SmarTrAC* brings together automatic sensing, data mining, and surveying in a hybrid and seamless manner, generating more comprehensive and accurate travel behavior data with less respondent burden when compared to any other existing activity data collection method:

- SmarTrAC processes sensor data (including location sensing) to extract meaningful activity/travel data to minimize the amount of input needed from the user;
- SmarTrAC is capable of capturing both time-series data from built-in sensors and additional user input data, which generates more comprehensive data than recall/diary surveys alone or GPS/accelerometer tracking alone; and
- SmarTrAC allows sensor data to interact with user input data so that the two data sources can help calibrate and validate each other. Such interaction minimizes recall bias and reporting errors in user input data because information derived from sensor data serves as robustness check for user input.

SmarTrAC has passed both laboratory tests conducted by SmarTrAC researchers using testing smartphones and field tests among 17 real-world smartphone users. Tests confirmed that SmarTrAC has a reasonable battery consumption rate (with room for improvement), a moderate data storage/transmission requirement, a high accuracy in classifying episodes as activities vs. trip, a high accuracy in identifying travel modes for trips, and as a medium-high accuracy in classifying activity type for activities. Although SmarTrAC performed better in laboratory tests than field tests, results from the field test showed good real-world performance:

- SmarTrAC detects activity-to-trip and trip-to-activity transition points with a high degree of accuracy; indeed, across all trip to activity transition points, SmarTrAC detects the transition within ± 30 seconds 88% of the time. For activity to trip transition points, the change is correctly predicted within ± 30 seconds 91% of the time.
- SmarTrAC detects the travel mode of each single-mode trip segment with a high degree of accuracy. The overall accuracy rate of 86% shows that, in nearly 9 out of every 10 single-mode trip segments, SmarTrAC accurately predicts the mode of transportation used during that segment.
- SmarTrAC's predictive accuracy is low (about 25-40%) when one considers an activity to have been accurately predicted only if the most likely (predicted) activity type matches the true activity for each episode. However, the true activity type was among the two most probable activities approximately 70-80% of the time (versus 40% expected by random guessing) and among the three most probable activities 80-95% of the time (versus 60% expected by random guessing). This result suggests that SmarTrAC's

strategy of sorting predicted activity types by their predicted probability puts the true activity type in the top three the vast majority of the time. Also, it is important to keep in mind that this evaluation is done for the most challenging activity type prediction scenario – i.e., when a prediction must be made for an activity which took place at a new location (for this particular SmarTrAC user). However, after a user visits any location and specifies that a particular activity took place there, SmarTrAC is able to leverage this information for the user's subsequent visits to that location – activity type prediction accuracy in such cases is very high, exceeding 95%.

Further, the field tests show that SmarTrAC functioned well on eleven unique Android phone models having Android version of 4.0.4 or later. During the field tests, 74% of the phones had a battery life longer than 6 hours, and about half of the phones (47%) had a battery life longer than 8 hours. The quality of travel mode and activity type predictions delivered by SmarTrAC did not vary substantially across phone models, suggesting that app performance does not depend heavily on the characteristics of the handset on which it is deployed.

Users found SmarTrAC to be easy to use and had few issues entering and annotating data within the application. The most burdensome participation requirement reported by the field test participants was keeping the phone charged due to the fact that battery life was less than 10 hours on some phones. The majority of participants were satisfied with the overall experience with SmarTrAC, and felt comfortable to have SmarTrAC recording their daily activity and trip details. Despite the fact that SmarTrAC is designed to track and catalogue users' daily behaviors, which might be considered intrusive, no users expressed concern about sharing these data with the research team for research purposes.

To conclude, SmarTrAC is an innovative, user-friendly application that can effectively collect highly detailed, multi-dimensional activity-travel data with high accuracy and minimal respondent burden. SmarTrAC is functional across a range of phone models and is feasible for wide-scale deployment.

6.2 Future Directions

The SmarTrAC research team is committed to continuing work on this application. The team has received some internal funding from the Center for Transportation Studies at the University of Minnesota to explore the potential of working with U.S. metropolitan planning organizations to deploy SmarTrAC in their metropolitan-scale household travel behavior survey efforts.

Aside from this immediate improvement direction, SmarTrAC researchers will look for new funding opportunities to pursue the following improvements:

- Continue to explore and implement battery-saving techniques for reducing SmarTrAC's impact on phone battery life;
- Continue to explore and implement techniques for reducing the data storage requirements of SmarTrAC;

- Continue to improve the predictive accuracy of SmarTrAC, especially when it comes to travel mode predictions around mode change points as well as between motorized modes such as car vs. bus. Specifically, we will consider how individual user data might be used to further customize travel mode and activity type predictions based on previously detected trips and activities;
- Continue to explore possibilities of interfacing SmarTrAC with additional geographic data and knowledge bases (e.g., public transit bus route information);
- Enhance the data privacy and security features of SmarTrAC, including developing functionality for password protection of user data on the device, and data backup to a secure server;
- Enhance visualization and analytical tools that enable users to summarize their activity-travel characteristics using the activity and travel data SmarTrAC collects.
- Develop a web interface that allows users to explore, summarize, and analyze the activity and travel data they upload from SmarTrAC to the cloud; and
- Develop a version of SmarTrAC for Apple iOS devices.

References

Auld, J., Williams, C., Mohammadian, A., & Nelson, P. (2009). An automated GPS-based prompted recall survey with learning algorithms. *Transportation Letters, 1*(1), 59-79.

Biljecki, F., Ledoux, H., & van Oosterom, P. (2013). Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science, 27*(2), 385-407.

Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the netherlands. *Transportation Research Part C: Emerging Technologies, 17*(3), 285-297.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Boca Raton, FL: CRC Press. Retrieved from http://scholar.google.com/scholar?cluster=3017989415557051150&hl=en&oi=scholarr#0

Du, J., & Aultman-Hall, L. (2007). Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. *Transportation Research Part A: Policy and Practice, 41*(3), 220-232.

Ettema, D., Timmermans, H., & van Veghel, L. (1996). Effects of data collection methods in travel and activity research.

Fan, Y., Chen, Q., Liao, C., & Douma, F. (2012). *Smartphone-based travel experience sampling and behavior intervention among young adults*Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota. Retrieved from http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=2142

Flamm, M., & Kaufmann, V. (2007). Combining person based GPS tracking and prompted recall interviews for a comprehensive investigation of travel behaviour adaptation processes during life course transitions. Paper presented at the *11th World Conference on Transportation Research, Berkeley,*

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*, 1189-1232. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.9093

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in new york city. *Computers, Environment and Urban Systems, 36*(2), 131-139.

Guensler, R. L., Li, H., Ogle, J. H., Axhausen, K. W., & Schönfelder, S. (2006). Analysis of commute atlanta instrumented vehicle GPS data: Destination choice behavior and activity spaces. Paper presented at the *Transportation Research Board 85th Annual Meeting,* (06-1983)

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life* Sage.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics, 15*(3), 651-674. doi:10.1198/106186006X133933

Jariyasunant, J., Abou-Zeid, M., Carrel, A., Ekambaram, V., Gaker, D., Sengupta, R., et al. (2014). Quantified traveler: Travel feedback meets the cloud to change behavior. *Journal of Intelligent Transportation Systems,* (ahead-of-print), 1-16.

Lee-Gosselin, M. E., Doherty, S. T., & Papinski, D. (2006). Internet-based prompted recall diary with automated gps activity-trip detection: System design. Paper presented at the *Transportation Research Board 85th Annual Meeting,* (06-1934)

Li, Z. J., & Shalaby, A. S. (2008). Web-based GIS system for prompted recall of GPS-assisted personal travel surveys: System development and experimental study. Paper presented at the *Transportation Research Board 87th Annual Meeting,* (08-2868)

Liao, L., Fox, D., & Kautz, H. (2007). Extracting places and activities from gps traces using hierarchical conditional random fields. *The International Journal of Robotics Research, 26*(1), 119-134.

McGowen, P., & McNally, M. (2007). Evaluating the potential to predict activity types from GPS and GIS data. Paper presented at the *Transportation Research Board 86th Annual Meeting, Washington,*

Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., & Maurer, P. (2014). Supporting large-scale travel surveys with smartphones–A practical approach. *Transportation Research Part C: Emerging Technologies, 43,* 212-221.

Schlich, R., & Axhausen, K. W. (2003). Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation, 30*(1), 13-36.

Schönfelder, S., Axhausen, K. W., Antille, N., Bierlaire, M., Axhausen, K. W., Axhausen, K. W., et al. (2002). *Exploring the potentials of automatically collected GPS data for travel behaviour analysis: A swedish data source* Citeseer.

Schuessler, N., & Axhausen, K. W. (2009). Processing raw data from global positioning systems without additional information. *Transportation Research Record: Journal of the Transportation Research Board, 2105*(1), 28-36.

Shen, L., & Stopher, P. R. (2013). A process for trip purpose imputation from global positioning system data. *Transportation Research Part C: Emerging Technologies, 36,* 261-267.

Srinivasan, S., Ghosh, P., Sivakumar, A., Kapur, A., Bhat, C. R., & Bricka, S. (2006). *Conversion of Volunteer-Collected GPS Diary Data into Travel Time Performance Measures: Final Report,*

Stopher, P., FitzGerald, C., & Zhang, J. (2008). Search for a global positioning system device to measure person travel. *Transportation Research Part C: Emerging Technologies, 16*(3), 350-369.

Tsui, S. Y. A. (2005). *An enhanced system for link and mode identifications for GPS-based personal travel surveys* University of Toronto.

Tsui, S. Y. A., & Shalaby, A. S. (2006). 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, 1972*(1), 38-45.

Vlassenroot, S., Gillis, D., Bellens, R., & Gautama, S. (2014). The use of smartphone applications in the collection of travel behaviour data. *International Journal of Intelligent Transportation Systems Research, 13*(1), 17-27.

Wan, N., & Lin, G. (2013). Life-space characterization from cellular telephone collected GPS data. *Computers, Environment and Urban Systems, 39*, 63-70.

Winters, P. L., Barbeau, S. J., & Georggi, N. L. (2008). *Smart phone application to influence travel behavior (trac-it phase 3)* No. 549-35)National Center for Transit Research for Florida Department of Transportation.

Wolf, J., SchöUnfelder, S., Samaga, U., Oliveira, M., & Axhausen, K. (2004). Eighty weeks of global positioning system traces: Approaches to enriching trip information. *Transportation Research Record: Journal of the Transportation Research Board, 1870*(1), 46-54.

Wolf, J., Guensler, R., & Bachman, W. (2001). Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record: Journal of the Transportation Research Board, 1768*(1), 125-134.

Wu, J., Jiang, C., Houston, D., Baker, D., & Delfino, R. (2011). Automated time activity classification based on global positioning system (GPS) tracking data. *Environ Health, 10*(101), 22082316.

Zhang, L., Dalyot, S., Eggert, D., & Sester, M. (2011). Multi-stage approach to travel-mode segmentation and classification of gps traces. Paper presented at the *ISPRS Workshop on Geospatial Data Infrastructure: From Data Acquisition and Updating to Smarter Services,*

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). Learning transportation mode from raw gps data for geographic applications on the web. Paper presented at the *Proceedings of the 17th International Conference on World Wide Web,* pp. 247-256.

Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., & Terveen, L. (2007). Discovering personally meaningful places: An interactive clustering approach. *ACM Transactions on Information Systems (TOIS), 25*(3), 12.

APPENDIX A

GENERAL INFORMATION SURVEY

Page 1 Genereal information

SmarTrAC Field Test: Information Sheet

* Required

What is your name? *

Please enter the email address you use most often in the space below. *

Please indicate the make of your phone. *

e.g. Sony Experia, Samsung S3, Nexus 5 etc.

Please fill out your phone's 15 digit IMEI number *

To get the number simply dial *#06# on the phone and hit the call button. The number will pop up.

What is the android version of your phone?

Go to Settings and the About or About Device option to get the details (e.g. 4.4.2)

On a typical day how long does the phone battery last (from full to no charge)? *

- Less than 4 hours
- 4 to 6 hours
- 6 to 8 hours
- 8 to 10 hours
- 10 to 12 hours
- 12 to 14 hours
- More than 14 hours

Thank You!

Submit

Never submit passwords through Google Forms.

APPENDIX B

DAILY FIELD TEST LOG

Daily Log: Page 1 (part 1)

SmarTrAC Daily Field Test Log

Please answer the following questions about your experience with the application. You are required to fill this form once for each day of field testing.

* Required

Please select today's date *

mm/dd/yyyy

Please enter your name *

Please select your phone from the list below

If you phone is not listed above please enter your name followed by the phone's IMEI number in the box below

Battery Consumption

Please answer the following questions about your phone's battery life while using SmarTrAC

How long did it take for your phone to loose charge (from 100 – 0%)? *

Please select an option based on approximate observed battery consumption. For example, if the phone loses approximately 20% charge every hour it will last for 5 hours on full charge; if the battery loses approximately 50% charge in 4 hours it will last for 8 hours on full charge, etc.

- Under 4 hours
- 4-6 hours
- 6-8 hours
- 8-10 hours
- 10-12 hours
- 12-14 hours
- More than 14 hours

Daily Log: Page 1 (part 2)

Please provide details of any battery issues you had today while the SmarTrAC application was running on your phone

If you have no additional details to report leave the response box blank

Application crashing or bug issues

Please answer the following questions about any crashing issues, bugs or abnormal application behavior you may have encountered while using the SmarTrAC application

Did the application crash while you were using it today? *

Report all crashes regardless of whether you used the application's crash report function or not.

- ○ Yes
- ○ No

If you answered yes to the previous question please provide details about the application crash you encountered today

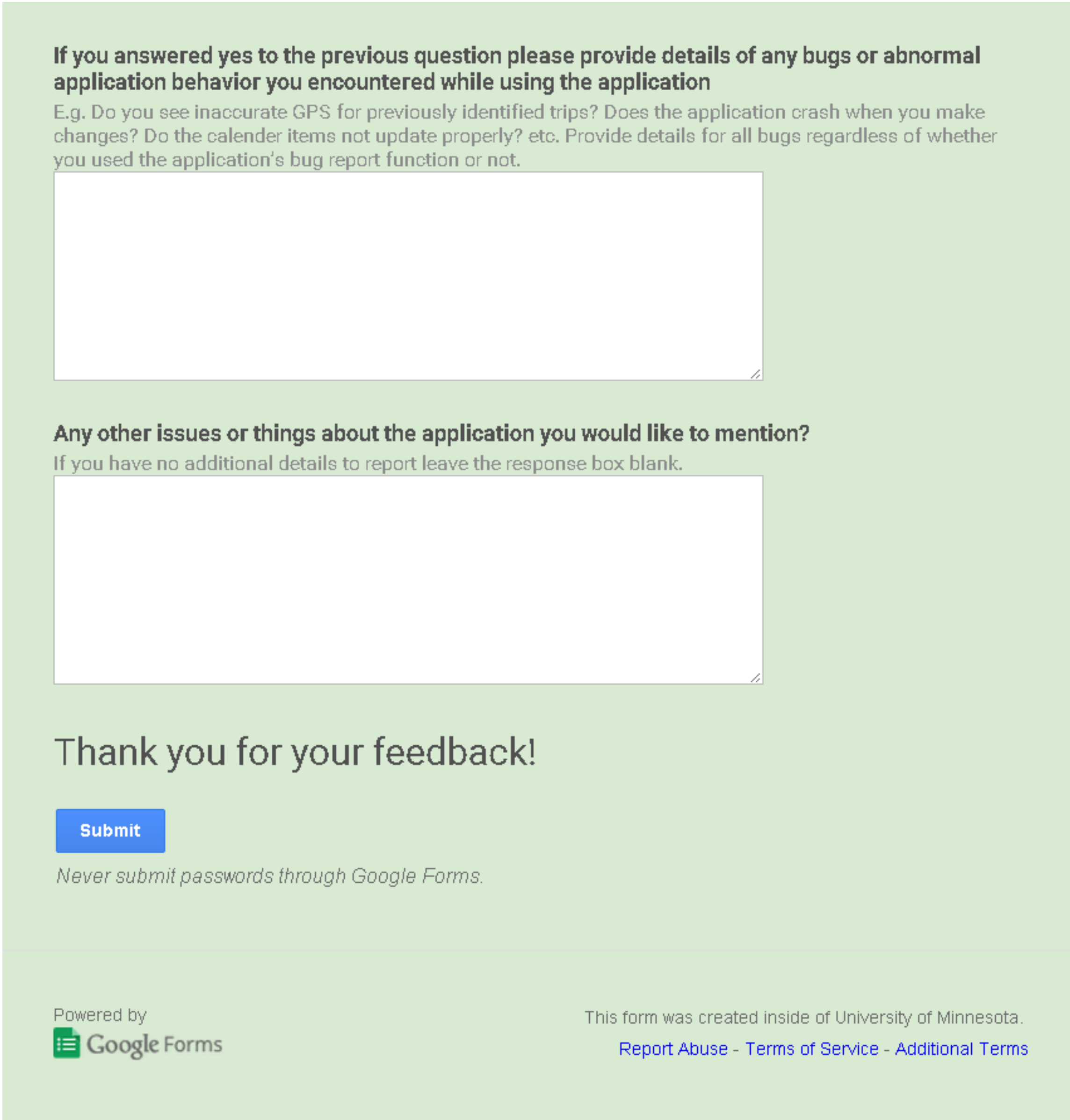
E.g. Did the crash report screen open automatically when the application crashed? Did it happen while you were editing trips, uploading data, clearing changes? etc. Provide details for all crashes regardless of whether you used the application's crash report function or not.

Did you encounter any bugs or abnormal application behavior while using SmarTrAC today? *

Report all bugs regardless of whether you used the application's bug report function or not.

- ○ Yes
- ○ No

Daily Log: Page 1 (part 3)



If you answered yes to the previous question please provide details of any bugs or abnormal application behavior you encountered while using the application

E.g. Do you see inaccurate GPS for previously identified trips? Does the application crash when you make changes? Do the calender items not update properly? etc. Provide details for all bugs regardless of whether you used the application's bug report function or not.

Any other issues or things about the application you would like to mention?

If you have no additional details to report leave the response box blank.

Thank you for your feedback!

Submit

Never submit passwords through Google Forms.

Powered by Google Forms

This form was created inside of University of Minnesota.

Report Abuse - Terms of Service - Additional Terms

APPENDIX C

SMARTRAC INFORMATION SHEET

Information Sheet for Research Participants

Thank you for your interest in this study that investigates how smartphones can be used to study travel activity and experiences. Please read this document and ask any questions you may have before participating in this research.

Project Introduction:

This study is collaborative effort between researchers at the Humphrey School of Public Affairs, the School of Public Health and the Carlson School of Management, University of Minnesota. In this study, we apply advanced smartphone technologies including computing, communication, and sensing to collect detailed information about people's daily travel behavior. Findings from this study are expected to shed light on the potential of smartphone applications for understanding and accurately capturing travel-related activity and experiences.

Participants' Responsibility:

If you agree to participate, please perform the following tasks:

1. **Install the SmarTrAC application on your phone**. The SmarTrAC application is designed by our researcher team.
2. **Do keep your phone with you - with our application running - as much as possible during all waking hours.**
3. **Do review and change/edit your travel and activity information whenever possible.** Review the daily travel data recorded by the application whenever possible and change/edit the recorded information as needed. We recommend doing this at the end of each day. Reviewing the recorded information at the end of each day helps ensure that you accurately recall your day's trips and activities.
4. **Upload data.** After reviewing data for each day please upload all data. We recommend doing this at the end of each day.
5. **Keep your phone charged.** The application does not record any information when it is turned off.
6. **Contact the study team with any issues:** If the SmarTrAC application is not functioning properly (i.e. has turned off or is not recording all trips/activities) and the issue is not resolved by simply turning the application OFF and ON, please inform the study team at your earliest convenience at smartrac@umn.edu
7. **Do not uninstall the SmarTrAC application from your phone.**

SmarTrAC's Impacts on Phone Performance:

We have tested the impact of our application on a phones battery life and memory requirement, for data storage.

- With the application running in the background continuously, the **battery life of a phone** is likely to be **8-10 hours** with normal voice/text/data usage. It is important to note that the impact on the battery life varies by phone and depends on your intensity of use.

- The application collects about 25 MB of raw sensor data and statistics per day which will be stored in the memory card of your phone. Therefore, for **7 days** of participation, **175Mb storage** space will be needed.

Confidentiality:

Information collected by our application includes location data (i.e., latitude and longitude) for all trips/activities conducted and trip features (i.e., time, speed and acceleration). To protect the information collected from you:

- All the collected information will be encrypted and stored on a server behind a firewall.
- No person other than members of the study group will have access to the collected information.
- Study results will not include any location information of test participants (e.g. home, work, activity etc.) and all participants will be represented by an alphanumeric identifier (rather than names) to protect their identity.
- In addition, if you feel the need to turn off the application for a period of time due to privacy concerns you can do so by going to the application's user settings and turning it off.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota.

Contacts and Questions:

The Principal Investigator for this study is Associate Professor Yingling Fan at the Humphrey School of Public Affairs. Assistant Professor Julian Wolfson at the School of Public Health is the Co-Principal Investigator. Other researchers conducting this study include: Professor Gediminas Adomavicius (co-investigator), Kirti V Das (research fellow and project manager), Yash Khandelwal (research assistant), and Jie Kang (research assistant). If you have any questions about the project please feel free to email us at smartrac@umn.edu or visit our website at http://smartrac.umn.edu/.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), you are encouraged to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

Please keep a copy of this information sheet for your record purpose.

APPENDIX D

DOS AND DON'TS SHEET

DO's and DON'T's (SmarTrAC)

DO:

- **Do charge your phone every night and when the battery is low.** Our application runs in the background continuously, which leads to short battery life (8-10 hours). Make sure your phone is fully charged every morning and has power during the day because the application does not collect any data when the phone is powered off.
- **Do keep your phone with you as much as possible.** This ensures the activity data collected by our application is valid and consistent for comparison between participants.
- **Do check the SmarTrAC icon occasionally.** No SmarTrAC icon or a red SmarTrAC icon indicates that our application may have stopped running. Try switching off the application and then turning it on again. If the green icon still does not appear contact us at your earliest convenience at smartrac@umn.edu.
- **Do contact us at your earliest convenience at smartrac@umn.edu if there is problem with our application.** For example, for a day, the application repeatedly fails to capture all or some of the trips/activities conducted.
- **Do review and change/edit your travel and activity information whenever possible.** We recommend doing this at the end of each day or earlier to ensure you accurately recall your trips and activities during the day.

DON'T:

- **Don't turn off your GPS or Wi-Fi.** Both the GPS and Wi-Fi help to detect activities and trips. If you wish to turn off the SmarTrAC application for privacy reasons, please go to "User Settings" in the SmarTrAC sidebar menu and turn the application OFF. Please, make sure you remember to turn the application back ON.
- **Don't uninstall the SmarTrAC application on your phone until the end of your participation in the study.**

For additional tips on application navigation, reviewing and changing/editing information, please watch our tutorial videos.

Videos can be viewed at: http://smartrac.umn.edu/for-users

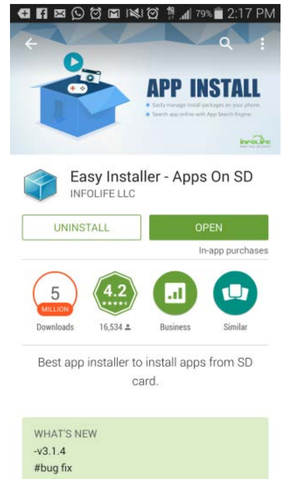
APPENDIX E

SMARTRAC INSTALLATION INSTRUCTIONS

Installing the SmarTrAC application

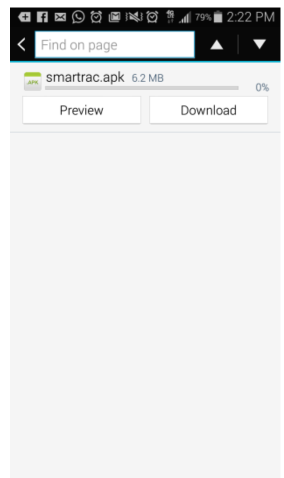
Step 1: Install Easy Installer

1. Go to the Google Play Store.
2. Search for "Easy Installer".
3. Install the application.



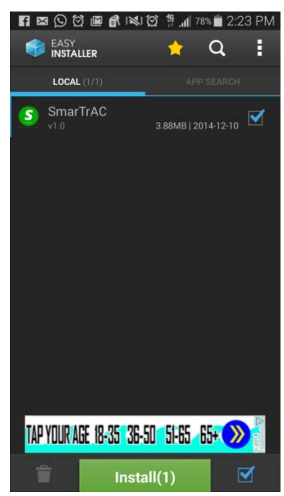
Step 2: Download the SmarTrAC installation file from your email (on your phone)

1. Go to the SmarTrAC application install file in the email we sent you.
2. Download the SmarTrAC.apk file.



Step 3: Installing the SmarTrAC alpplication

1. Open easy Installer on your phone
2. Select the SmarTrAC file and click on "Install"



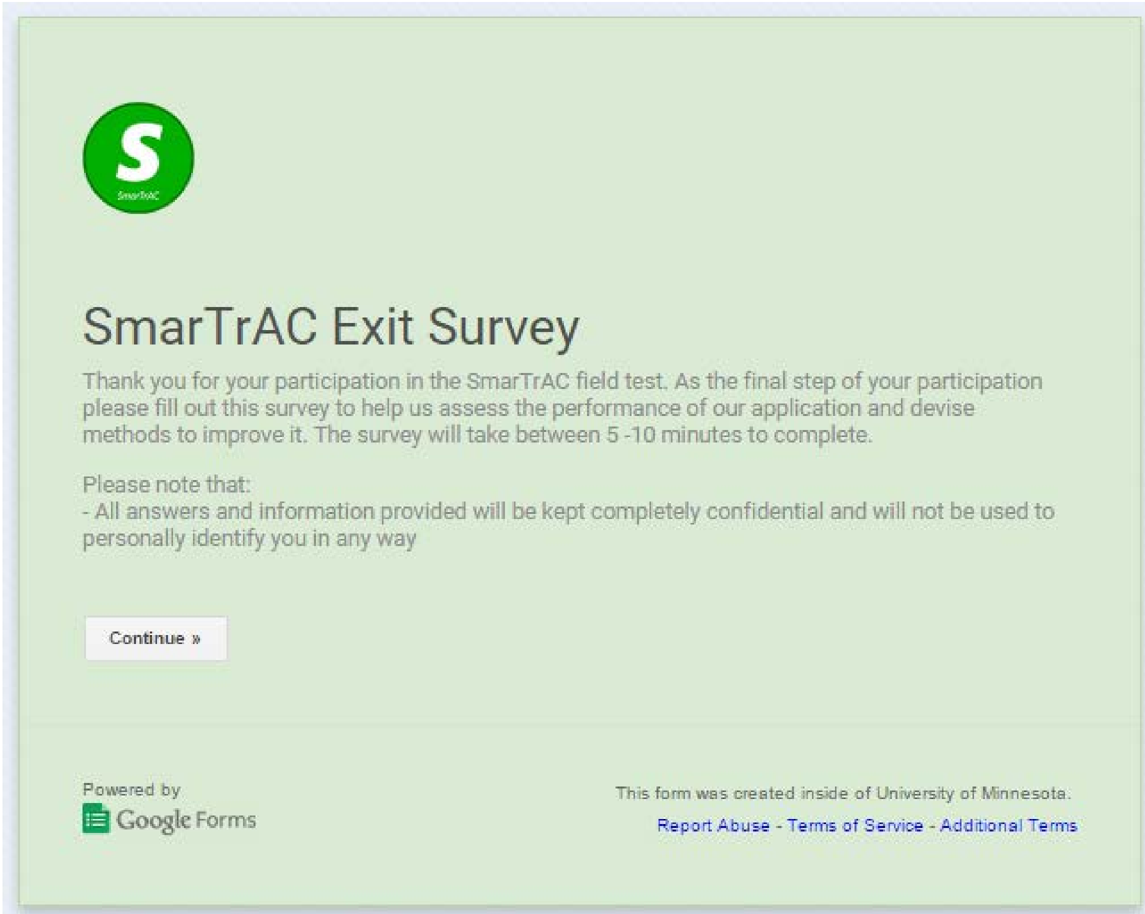
The application should now be set up on your phone!

If you have any questions about the installation process feel free to email us at smartrac@umn.edu or visit our website at http://smartrac.umn.edu/.

APPENDIX F

EXIT SURVEY

Introduction: Page 1



SmarTrAC Exit Survey

Thank you for your participation in the SmarTrAC field test. As the final step of your participation please fill out this survey to help us assess the performance of our application and devise methods to improve it. The survey will take between 5 -10 minutes to complete.

Please note that:
- All answers and information provided will be kept completely confidential and will not be used to personally identify you in any way

Continue »

Powered by Google Forms

This form was created inside of University of Minnesota.

Report Abuse - Terms of Service - Additional Terms

SmarTrAC User Experience: Page 2 (part 1)

SmarTrAC Exit Survey

* Required

Section 1: SmarTrAC user experience

Please answer the following questions about your experience using SmarTrAC

Overall, how satisfied are you with your experience using the application? *

- ○ Very satisfied
- ○ Satisfied
- ○ Neither satisfied nor dissatisfied
- ○ Dissatisfied
- ○ Very dissatisfied

To what extent do you do you agree or disagree with the following statements about study components? *

Answer scale (from left to right): Strongly agree - Strongly disagree. Select one answer for each row

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
All my trips were detected	○	○	○	○	○
All my activities were detected	○	○	○	○	○
The application is easy to navigate	○	○	○	○	○
Editing trips and activities is easy	○	○	○	○	○
Reporting issues is easy	○	○	○	○	○
Installing SmarTrAC was easy	○	○	○	○	○
Instruction on how to use SmarTrAC was comprehensive and helpful	○	○	○	○	○
Communication with the research team during the test was helpful	○	○	○	○	○

SmarTrAC User Experience: Page 2 (part 2)

To what extent do you agree or disagree with the following statements about the use of SmarTrAC to collect travel and activity data? *

Answer scale (from left to right): Strongly agree - Strongly disagree. Select one answer for each row

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
I felt comfortable having SmarTrAC tracking my locations	○	○	○	○	○
I felt comfortable having SmarTrAC tracking my trips	○	○	○	○	○
I felt comfortable entering additional trip and activity details	○	○	○	○	○

To what extent do you agree or disagree with the following statement? *

Answer scale (from left to right): Strongly agree - Strongly disagree.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Using SmarTrAC has made me more aware of my travel and activity behavior	○	○	○	○	○

If you responded "Strongly agree" or "Agree" to the previous question please provide details of how using the application has added to your awareness of your travel and activity behavior.

« Back   Continue »

Powered by
Google Forms

This form was created inside of University of Minnesota.
Report Abuse - Terms of Service - Additional Terms

SmarTrAC User Responsibilities: Page 3 (part 1)

SmarTrAC Exit Survey

* Required

Section 2: SmarTrAC user responsibilities

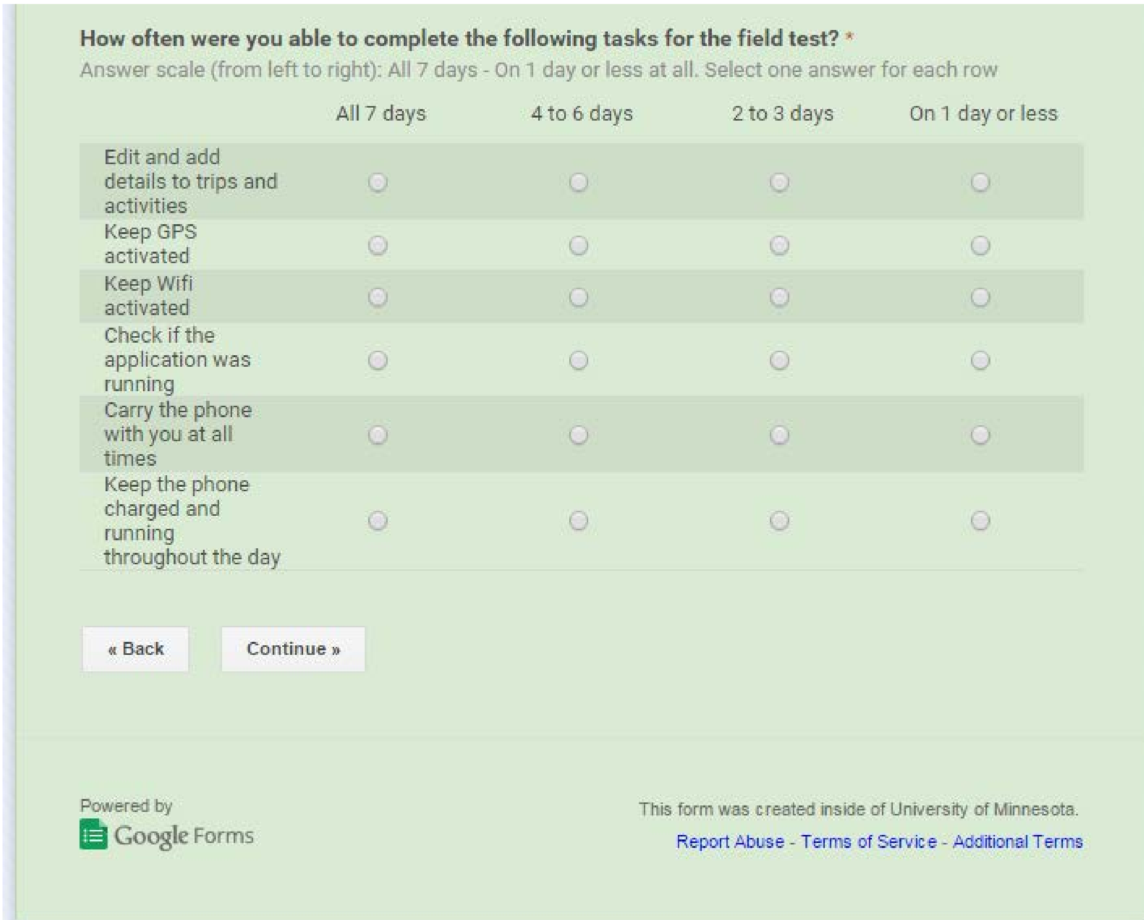
Please answer the following questions about tasks you were requested to complete during the SmarTrAC field test

How burdensome were each of the following participation requirements to your everyday life? *

Answer scale (from left to right): Very burdensome - Not burdensome at all. Select one answer for each row

	N/A (I did not fulfill the requirement)	Very burdensome	Burdensome	Somewhat burdensome	Not burden at all
Keep your phone charged	○	○	○	○	○
Edit trips and activities	○	○	○	○	○
Add details to trips and activities	○	○	○	○	○
Upload data	○	○	○	○	○
Fill out the daily field test log	○	○	○	○	○
Keep GPS activated	○	○	○	○	○
Keep Wifi activated	○	○	○	○	○
Check periodically if application is running	○	○	○	○	○
Carry your phone with you each day	○	○	○	○	○

SmarTrAC User Responsibilities: Page 3 (part 2)



How often were you able to complete the following tasks for the field test? *

Answer scale (from left to right): All 7 days - On 1 day or less at all. Select one answer for each row

	All 7 days	4 to 6 days	2 to 3 days	On 1 day or less
Edit and add details to trips and activities	○	○	○	○
Keep GPS activated	○	○	○	○
Keep Wifi activated	○	○	○	○
Check if the application was running	○	○	○	○
Carry the phone with you at all times	○	○	○	○
Keep the phone charged and running throughout the day	○	○	○	○

« Back   Continue »

Powered by Google Forms

This form was created inside of University of Minnesota.

Report Abuse - Terms of Service - Additional Terms

SmarTrAC Application Performance: Page 4 (part 1)

SmarTrAC Exit Survey

* Required

Section 3: SmarTrAC application performance

Please answer the following questions about the application's performance.

Select the appropriate option for the following application performance areas *

Answer scale (from left to right): Always or almost always - Almost never or never. Select one answer for each row

	Always or almost always	Most times (more than half the time)	Sometimes (about half the time)	A few times (less than half the time)	Almost never or never
Travel modes (e.g., car, bus, etc.) were accurately detected	○	○	○	○	○
Trip routes were accurately detected and mapped	○	○	○	○	○
Locations of activities were accurately detected and mapped	○	○	○	○	○
For new activity locations, actual activity was among the top three predicted activity types (e.g., Education: 9%; Shopping 20% etc.)	○	○	○	○	○
Start/end times were accurately detected	○	○	○	○	○
Crash reports were generated each time SmarTrAC crashed	○	○	○	○	○

SmarTrAC Application Performance: Page 4 (part 2)

If you answered "A few times" or "Almost never or never" to any of the items in the previous question please provide details or examples of the issues faced

What aspects or features of the SmarTrAC did you enjoy the most? *

What aspects or features of the SmarTrAC do you think need improvement? *

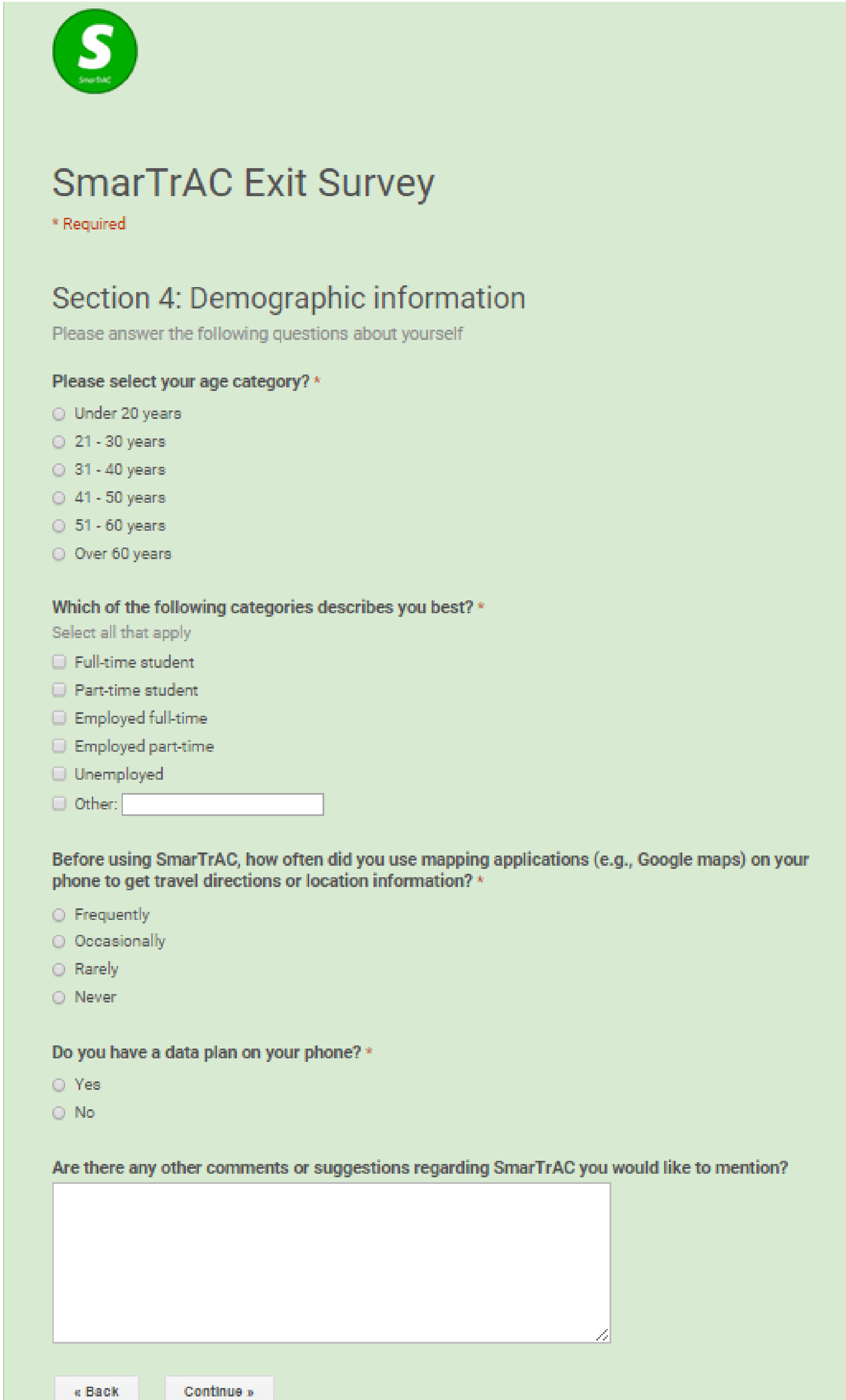
« Back   Continue »

Powered by Google Forms

This form was created inside of University of Minnesota.

Report Abuse - Terms of Service - Additional Terms

SmarTrAC Demographic Information: Page 5



SmarTrAC Exit Survey

* Required

Section 4: Demographic information

Please answer the following questions about yourself

Please select your age category? *

- ○ Under 20 years
- ○ 21 - 30 years
- ○ 31 - 40 years
- ○ 41 - 50 years
- ○ 51 - 60 years
- ○ Over 60 years

Which of the following categories describes you best? *

Select all that apply

- ☐ Full-time student
- ☐ Part-time student
- ☐ Employed full-time
- ☐ Employed part-time
- ☐ Unemployed
- ☐ Other:

Before using SmarTrAC, how often did you use mapping applications (e.g., Google maps) on your phone to get travel directions or location information? *

- ○ Frequently
- ○ Occasionally
- ○ Rarely
- ○ Never

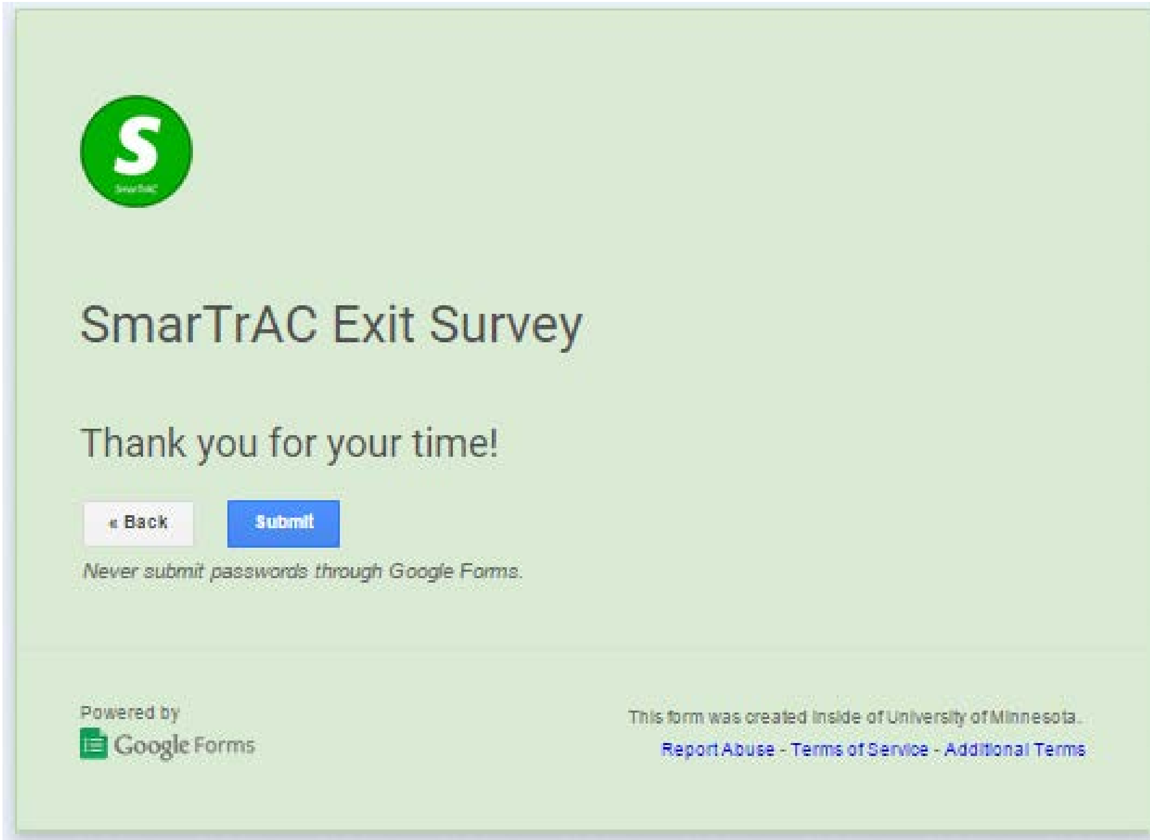
Do you have a data plan on your phone? *

- ○ Yes
- ○ No

Are there any other comments or suggestions regarding SmarTrAC you would like to mention?

« Back   Continue »

SmarTrAC Survey Completion Page: Page 6



SmarTrAC Exit Survey

Thank you for your time!

« Back

Submit

Never submit passwords through Google Forms.

Powered by

Google Forms

This form was created inside of University of Minnesota.

Report Abuse - Terms of Service - Additional Terms