DETERMINING TRIP AND TRAVEL MODE

FROM GPS AND ACCELEROMETER DATA

Aaron W Burgess

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Master of Science

in the Department of Geography,

Indiana University

May 2016

Accepted by the Graduate Faculty, Indiana University, in partial

fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

Jeff Wilson, Ph.D., Chair

Sarah Wiehe, M.D., MPH

Gabriel Filippelli, Ph.D.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Jeff Wilson, Sarah Wiehe and Gabe Filippelli for their tireless efforts in keeping me on track and teaching me to say "no." I particularly cite Jeff and Sarah for keeping me involved in their various projects. My skill set would be nowhere without that experience.

Special thanks are due to the Department of Geography, both faculty and students, for their support. As a group, they are always ready to listen, help or participate. Joyce Haibe, our department admin, has practically been my personal manager whose assistance will never be forgotten.

Finally, I thank my wife, Michelle Burgess, for putting up with all the neuroticism, obsessive behavior and other fun personality traits that come with being married to me. She is the only one that knows how to keep my good qualities on top and my not-so-good qualities at bay. Her qualities keep me calm, her love keeps me going.

Aaron W Burgess

DETERMINING TRIP AND TRAVEL MODE

FROM GPS AND ACCELEROMETER DATA

The use of Global Positioning Systems (GPS) and/or accelerometers to identify trips and transportation modes such as walking, running, bicycling or motorized transportation has been an active goal in multiple disciplines such as Transportation Engineering, Computer Science, Informatics and Public Health. The purpose of this study was to review existing methods that determined trip and travel mode from raw Global Positioning System (GPS) and accelerometer data, and test a select group of these methods. The study had three specific aims: (1) Create a systematic review of existing literature that explored various methods for determining trip and travel mode from GPS and/or accelerometer data, (2) Collect a convenience sample of subjects who were assigned a GPS and accelerometer unit to wear while performing and logging travel bouts consisting of walking, running, bicycling and driving, (3) Replicate selected method designs extracted from the systematic review (aim 1) and use subject data (aim 2) to compare the methods. The results were be used to examine which methods are effective for various modes of travel.

Jeffrey S. Wilson, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Global Positioning System (GPS) uses a constellation of satellites through which the position of a ground receiver can be attained. GPS has been utilized in many studies to track where, when and how subjects use their time (Mavoa et al. 2011). The ground receiver can be a handheld GPS unit in the form of a small, pocket-sized device or a smartphone that includes a GPS chipset. Some handheld GPS units can offer greater accuracy than a smartphone but can take 30 seconds or more to acquire the necessary satellite signals (known as a cold-start) while a smartphone is able to utilize wireless networks and cellular towers, albeit with less accuracy than a stand-alone GPS unit, avoiding the impact created by cold-starts (Zandbergen and Barbeau 2011).

Accelerometers measure intensity of movement and have been used in health studies to estimate physical activity (Cooper et al. 2010, Shay et al. 2009, Pober et al. 2006). Many studies utilize both GPS and accelerometer devices for various purposes (Oliver et al. 2010, Troped et al. 2008). The most basic structure of studies involving GPS and/or accelerometers is movement beginning with whether a subject is stationary (static) or whether that subject is in motion. Most studies classify movement over a specific distance and/or time to when movement ends as a trip (Chung and Shalaby 2005, Bohte and Maat 2009). A trip has no standard and exact definition varies by study. Gong, et al defines a trip as periods of movement in between two periods of “dwell time” that meet a set time threshold (Gong et al. 2012). Bohte, et al states that defining a trip is dependent on the travel mode itself (Bohte et al. 2009).

The use of GPS and accelerometers to identify trips and transportation modes such as walking, running, bicycling or motorized transportation has been a goal in multiple disciplines such as transportation engineering, computer science, informatics and public health (Stopher et al. 2008, Reddy et al. 2008, Cho, Rodriguez and Evenson 2011). Determining travel mode is important to these disciplines for varying reasons. Transportation engineers seek to use trip and travel mode data in order to create multi-modal planning models (Quddus, Noland and Ochieng 2006) . Public health researchers focus on trip and travel mode to understand physical activity practices and trip purposes (e.g. destinations) (Troped et al. 2010, Duncan, Badland and Mummery 2009). The fields of Computer Science and Informatics have sought methods to determine travel mode and trip for use with portable devices such as smart phones to personalize user experience by offering location-based services that relate to user's spatial context (Abe et al. 2009).

Traditional methods to determine transportation mode in study subjects used travel diaries in which subjects would record when a trip began, ended and the transportation mode used in order to validate GPS and accelerometer data. Travel diaries have been used in both a paper format and web-based digital reporting (Stopher, Fitzgerald and Xu 2007). Travel diaries are subject to many shortcomings such as disagreement between reported mode and actual mode (Stopher et al. 2007, Wolf et al. 2001b). This disagreement often occurs due to reporting fatigue or omission on the part of the study participants (Wolf, Guensler and Bachman 2001a).

Methods have been developed to estimate trips and travel mode using GPS or a combination of GPS and accelerometers without travel diaries (referred to as "passive") (Stopher et al. 2008, Gonzalez et al. 2009, Byon, Abdulhai and Shalaby 2009, Wu et al. 2011). The ability to process GPS and/or accelerometer date into trips identified by travel mode is a consistent goal across disciplines. The varying goals between disciplines, however, have led to limited interaction between disciplines regarding travel mode determination.

This study conducted a systematic review of existing trip and travel mode determination methods in order to identify methods that could be replicated. A convenience sample of X college students was recruited to wear a GPS and accelerometer unit while performing four different kinds of travel activities: walking, running, bicycling and driving. Data collected from the subjects was processed through existing methods found in the systematic review and the results were reported in order to compare methods.

BACKGROUND

Systematic Review

Search Strategy

A systematic review was conducted to identify existing methods to passively detect both trips and travel mode using GPS only, accelerometer only or a combination of both. The following eight databases were searched: EBSCO, IEEE Xplore, PubMed, ACM Digital Library, ProQuest Central, OVID, Web of Science and Google Scholar. All databases were searched from January 1st, 2000 to September 6th, 2012. Search terms used included: (travel OR transit OR activity OR transportation OR trip*) AND (mode OR survey OR detection) AND (GPS or Global Positioning System or accelerometer.) Searches were limited to published journal articles, dissertations and theses. Bibliographies of relevant articles were reviewed to extract appropriate articles not found by the initial search.

Inclusion Criteria

To meet inclusion criteria, studies needed to utilize a GPS unit (stand-alone or as a component of a cellular phone) and/or an accelerometer unit to infer trips and mode by trip for study subjects. The specific purpose for determining trips and travel mode by trip varied by study and was not used as an exclusionary criterion. Studies could assess trip and mode for multiple modes or for a single mode. Included studies used subject logs to validate the study's travel mode assessment.

Articles were examined for specific details regarding methods for trip and mode detection as well as evaluation reporting. In order for a paper to be included, specific steps for implementing the mode/trip detection process. A convenience sample of three methods found in these papers were selected for future use.

Resulting articles were categorized based on the number of discipline(s) represented by the authors in each study. Studies were then characterized based on the technology used (e.g. GPS and/or accelerometer; stand-alone GPS or phone), the number of subjects.

Results

Literature Search Results

EBSCO, IEEE Xplore, PubMed, ACM Digital Library, ProQuest Central, OVID and Web of Science returned a total of 552 articles while Google Scholar returned a total of 13,459 articles. The 552 articles were reviewed by title and abstract to eliminate articles that did not address trip and/or travel mode inference. The first 200 articles returned by Google Scholar, based on how Google returns the results, were examined under the same criteria. A total of 49 articles from the 752 candidate articles were extracted after review either from the articles themselves or article bibliographies. Out of the 49 articles there were 17 that met inclusion criteria. Excluded articles included those did not present a method designed to infer transportation mode or the methods were not tested. Any article that presented a method but did not meet full criteria was included in this paper's bibliography as some were utilized in included studies (Schüssler and Axhausen 2008, Stopher et al. 2008).

The included articles shared few interdisciplinary relationships based on the self-reported discipline of the authors. Articles including public health and urban planning coauthors were the most frequent with 4 articles in total (see Figure 1.) Other noticeable differences were a focus in Computer Science/Informatics and some engineers on mobile and real-time solutions for transportation inference of GPS data. Articles authored by traffic engineers concentrated primarily on developing methods to replace travel diaries in transit studies and primarily focused on vehicular travel. Some computer science articles on trip and travel mode detection focused on relationships with health but did include co-authors from health or medical disciplines.

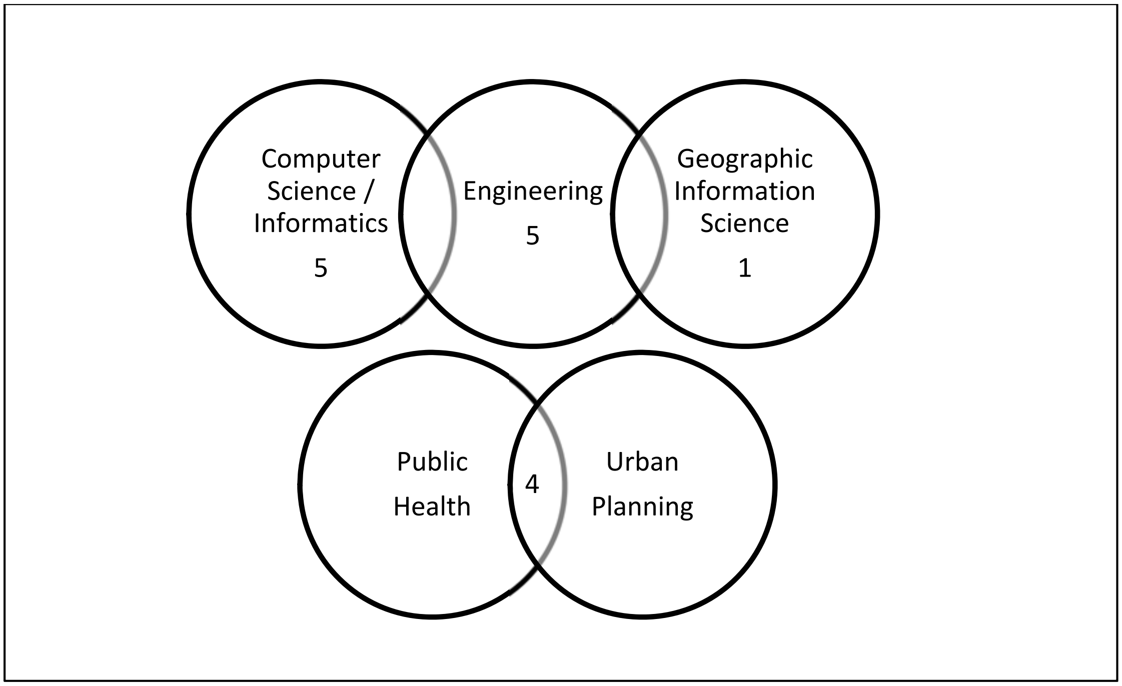

Figure 1 Venn Diagram of Disciplines from Included Articles

Characteristics

Seven studies used stand-alone GPS units, 2 used stand-alone GPS and accelerometer units and 8 used mobile phones. Travel mode classifications used included Walking (17), Car (13), Bicycle (8), Bus (7), Static (6), Rail/Train (4), Run (2), Subway (2), Inline Skating (1), Streetcar (1) and a general classification for urban transit (1). Most studies concentrated on 3 or more mode classifications while two studies focused on only walking trips (Cho et al. 2011, Rodriguez et al. 2012).

Lead Author	Year	Title	Author Discipline(s)	Device(s)	Focus (e.g. all modes, driving)
Anderson et al.	2007	Shakra: tracking and sharing daily activity levels with unaugmented mobile phones	Computer Science	Mobile	Car, Walk, Static
Bohte et al.	2009	Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands.	Transportation Engineering	GPS	Car, Bicycle, Walk, Train, Bus
Byon et al.	2009	Real-TimeReal-Time Transportation Mode Detection via Tracking Global Positioning System Mobile Devices	Civil Engineering	Mobile	Car, Walk, Streetcar, Bus
Chao et al.	2010	Identifying travel mode from GPS trajectories through fuzzy pattern recognition	Geographic Information Science	GPS	Walk, Bicycle, Bus, Rail
Chung et al.	2005	A Trip Reconstruction Tool for GPS-based Personal Travel Surveys	Transportation Engineering	GPS	Car, Bicycle, Walk, Transit
Cho et al.	2011	Identifying Walking Trips Using GPS Data	Public Health / Urban Planning	GPS	Walk
Gong et al.	2012	A GPS/GIS method for travel mode detection	GIS / Engineering	GPS	Walk, Subway, Rail, Car, Bus
Gonzalez et al.	2009	Automating mode detection for travel behaviour	Transportation / Computer Science	Mobile	Car, Walk, Bus
Reddy et al.	2008	Using Mobile Phones to Determinine Transportation Mode on Phone	Electrical Engineering	Mobile	Car, Walk, Run, Bicycle, Static
Rodriguez et al.	2012	Identifying Walking Trips From GPS and Accelerometer Data in Adolescent Females	Public Health / Urban Planning	GPS & Accelerometer	Walk
Sohn et al.	2006	Mobility Detection Using Everyday GSM Traces	Computer Science	Mobile	Car, Walk, Static
Stenneth et al.	2011	Transportation Mode Detection using Mobile Phones and GIS Information	Computer Science	Mobile	Car, Bicycle, Walk, Train, Bus, Static
Troped et al.	2008	Prediction of Activity Mode with Global Positioning System And Accelerometer Data	Public Health / Urban Planning	GPS & Accelerometer	Car, Walk, Run, Bicycle, In-line Skating
Wang, S. et al.	2010	Accelerometer based mode detection on phones	Computer Science	Mobile	Car, Walk, Bicycle, Bus, Subway, Static
Wu et al.	2011	Automated time activity classification based on global positioning system (GPS) tracking data	Public Health / Urban Planning	GPS	Car, Walk, Static
Zhang et al.	2011	Path2Go: Context-Aware Services for Mobile Real-Time MultiModal Traveler Information	Transportation Engineering	Mobile	Car, Bus, Rail, Walk, static
Zheng et al.	2010	Understanding transportation modes based on GPS data for web applications	Computer Science	GPS	Car, Bus, Walk, Bicycle

Table 1 Included Studies

Table 2 Accuracy of Mode Estimation

Lead Author	Sample Size*	Classification Method(s)	Walking**	Biking**	Driving**	Bus**	Train/Rail**	Static**	Other**
Anderson et al.	9	Artificial Neural Network; Hidden Markov Model	87%	--	73%	--	--	83%	--
Bohte et al.	1104	Rule Based	68%	72%	75%	0%	34%	--	--
Byon et al.	60 hours	Neural Network	98%	--	Highway = 60%; Arterial = 87%	82.0%	--	--	Streetcar = 84%
Chao et al.	32	Rule Based	96.4%	88.9%	--	57.1%	93.2%	--	--
Chung et al.	1	Rule Based	100%	100%	87.5%	--	--	--	Transit = 90%
Cho et al.	45	Rule Based	86%	--	--				--
Gong et al.	49	Rule Based	91.8%	--	84%	46.2%	25%	--	Subway = 58.8%
Gonzalez et al.	114 Trips	Multilayer Perceptron	100%	--	92.11%	81.6%	--	--	--
Reddy et al.	6	C4.5 Decision Trees; K-Means Clustering; Naïve Bayes; Nearest Neighbor; Support Vector Machines; continuous Hidden Markov Model	92.4%	90.6%	96.2%	--	--	97.8%	Run = 96.4%
Rodriguez et al.	51	Rule Based	88%	--	--	--	--	--	--
Sohn et al.	3	5-fold cross validation	70.2%	--	84.3%	--	--	95.4%	--
Stenneth et al.	6	Naïve Bayes, Bayesian Network, Decision Trees, Random Forest, Multilayer Perceptron	92%	93%	89%	95%	93%	98.5%	--
Troped et al.	10	Discriminant Function Analysis	96%	96.2%	61.1%	--	--	--	Run = 87.9%; Inline Skating = 91.7%
Wang, S. et al.	7	Decision Tree; k Nearest Neighbor; Support Vector Machine	88.78%	70.63%	53.62%	45.85%	--	63.31%	Subway = 40.64%
Wu et al.	47	Rule Based; Random Forest	74.7%***	--	88.6%***	--	--	42.3%	--
Zhang et al.	109 Trips	Rule Based Bayesian	--	--	--	--	--	--	LIMITED DETAIL: 92% Accuracy overall
Zheng et al.	65	Decision Tree	82.5%	75.4%	83.7%	61.5%	--	--	--

* Number of Subjects unless Otherwise Stated
** Best reported results in the event of multiple methods
*** Rule-based Classification

Sample size ranged from 1 participant performing several types of predetermined trips in a day (Chung and Shalaby 2005) to over a thousand participants performing their normal commuting routine for a week but with most using 50 or less (Bohte and Maat 2009). Almost all subject cohorts were convenience samples with most studies recruiting from university students or employees (Gi-Hyoug et al, 2011) while a few were recruited or extracted from government surveys (Bohte and Maat 2009) Age range for participants was seldom reported in detail although one study focused on adolescent females (Rodriguez et al. 2012). Subjects aged 65 or older were underrepresented and, for those studies that reported age ranges of participants, the youngest age was X (Troped et al. 2008, Bohte and Maat 2009).

Most methods require some type of geographic information system for spatially processing GPS data points be it for map matching algorithms or manual assessment (Gong et al. 2012, Stenneth et al. 2011). Utilized methods employed a rule-based system (Bohte and Maat 2009), some form of machine learning (such as a Naïve Bayes) (Gonzalez et al. 2009) or a combination of both (Zhang et al. 2011). Systems such as Path2Go (Zhang et al. 2011), Shakra (Anderson et al. 2007) and Trac-IT (Gonzalez et al. 2009) attempted to determine transportation mode in real-time on a smart-phone. All other studies used methods to transportation mode by trip after data collection had occurred.

Reported results were usually presented via precision and recall tables (Reddy et al. 2010) while others reported results in a confusion matrix or other table (Chung and Shalaby 2005, Gong et al. 2012, Troped et al. 2008, Xu et al. 2010).

<u>Selected Methods Examined</u>

Bohte et al. (Bohte and Maat 2009) used the largest subject sample (n = 1104) from three municipalities in the Netherlands. This study was focused on improving information about commuter habits for policy decisions. Subjects were recruited from respondents to a previous internet survey. Subjects carried a GPS logger for one week. Collected data was placed into a RDBMS with spatial capabilities (PostgreSQL/PostGIS.) GPS points less than 10 meters apart from the previous GPS point were removed. Trip end was defined as when GPS points remained in a particular location for 3 minutes or more with a trip beginning defined as when the next GPS point was 10 meters or more away from the previous point. This method used GIS data to assign trip purposes (13 categories) and for map matching in order to facilitate better mode matching such as when GPS points occurred along a train line. The rules used followed a decision path: (1) Remove unreliable GPS points and divide points in trips, (2) Set category of trip, (3) Set modality of trip and (4) Merge and add train trips. Subjects would then review trips, purpose and mode through a web-based interface to validate the method. Rule-based assignment of trip purpose failed to match the correct subject-identified trip purpose for most categories (e.g. work, recreation.) Correct mode identification was most effective for car, bicycle and foot (75%, 72% and 68%) while ineffective for train or bus/tram/metro (34% and 0%.)

Gong et al. (Gong et al. 2012) also used a rule-based method but focused on detecting travel mode in New York City. Like Bohte et al., this study is focused on improving detection of transportation mode for planning purposes. This study is unique in that methods were developed in a city with a population density of 10,000 per square kilometer. A total of 49 subject datasets were used from 63 subjects gathered through two separate surveys. Fourteen subject datasets were excluded for incomplete diaries, poor GPS data or erroneous diaries. Volunteers carried a handheld GPS logger for five consecutive weekdays and completed a travel diary for one of the five days. A rule based algorithm was developed using Visual Basic for Applications (VBA) requiring ArcGIS 9.3 or after with the Network Analyst Extension. Trip ends

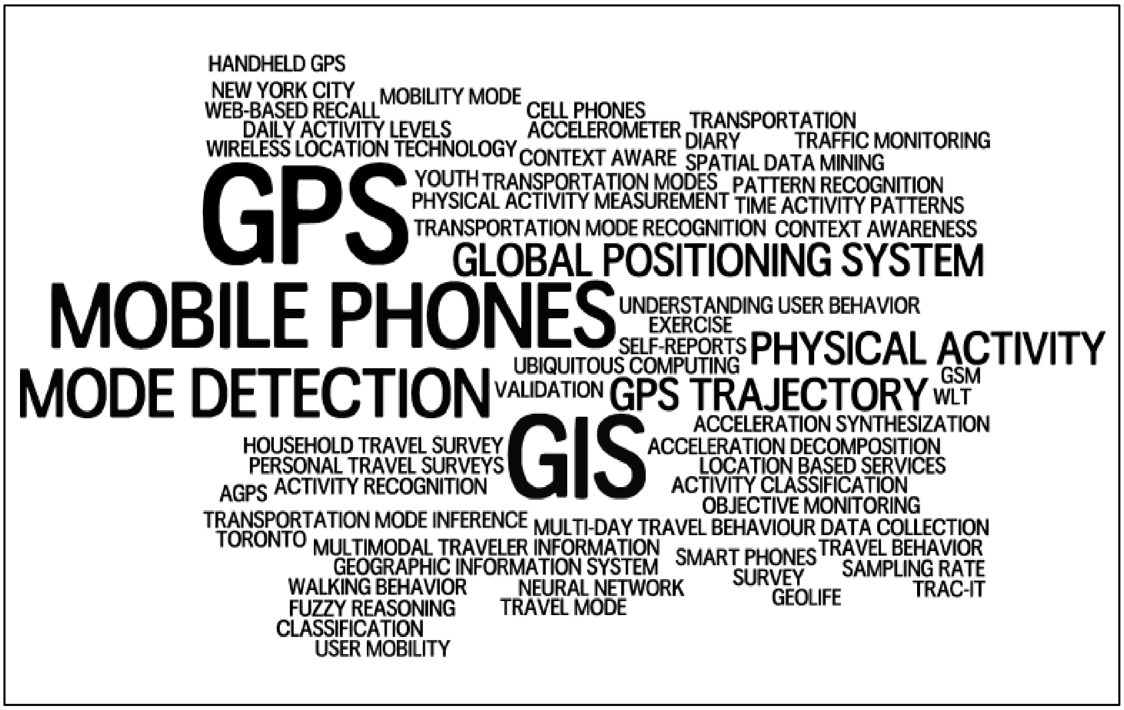

Figure 2 Keyword Cloud

were defined as the earliest point in a point cluster. Point clusters were defined as points that were within 50 meters of each other for greater than 200 seconds. A trip began with the next point that was not contained within the aforementioned point cluster. The algorithm steps were as follows: (1) Prepare GPS table, (2) Divide GPS data into trips, (3) Divide trips into trip segments and (4) Detect mode. The authors acknowledge the significant challenge that the urban canyon effect can hold in NYC. Urban canyon effect is created when tall buildings interfere with radio signals from GPS satellites. Walking trips were correctly identified 182 out of 197 times with 11 trips identified as a car trip and 4 as a bus trip. Thirty-seven out of 44 car trips were correctly identified with 2 trips incorrectly categorized as walking and 5 trips as bus. Subway and bus trips were not as effectively identified (65.6% and 62.5% correct) with the majority of incorrect identifications labeled as unknown or car trips.

Gonzalez et al. (Gonzalez et al. 2009) utilized the TRAC-IT program developed by the Center for Urban Transportation Research and the Department of Computer Science and

Engineering at the University of South Florida. TRAC-IT is installed on smart-phones and is designed to transfer GPS data to a server in real-time. This study did not report the number of subjects but instead reported the number of trips taken (n = 114) which were performed by an unknown number of research staff. The trip modes were known: 38 car, 38 bus and 38 walking trips. Trips were manually identified. This study used two datasets, one that included all collected GPS points and another that used the "critical point" algorithm. This algorithm retains only those GPS points that indicate a change in direction. Conceptually, this method retains the vertices associated with path creation during a trip. If a user was moving in a straight line then only the end vertices of that line were retained. The study used a type of neural network called a multilayer perceptron (a perceptron being a type of algorithm for supervised classification) to automatically assign travel mode to trip. Variables from the dataset of all GPS points were: (1) average speed, (2) maximum speed, (3) estimated horizontal accuracy uncertainty, (4) percent of points determined by cellular network, (5) standard deviation of distances between stop locations and (6) average dwell time. Classifiers for the "critical point" dataset used (1) average acceleration, (2) maximum acceleration, (3) average speed, (4) maximum speed, (5) ratio of critical points over total trip distance, (6) ratio of critical points over total trip time, (7) total distance and (8) average distance between critical points. The multi-layer perceptron was trained and tested in Weka, software designed for machine learning, using 10-fold cross-validation. (Hall et al. 2009) Using critical points proved the most effective as all GPS points can provide too much noise to the multi-layer perceptron. Additionally, this study showed that one can over-train a neural network which leads to model over-fit. Only accuracy was reported (see Table 1).

Wu et al. (Wu et al. 2011) developed both a rule-based approach and a machine learning approach (decision tree) simultaneously. 47 volunteers were recruited from two sources in the Los Angeles metropolitan area. The authors' interest was in developing methods to determine transportation mode to study people's exposure to air pollutants. Participants used handheld GPS loggers for varying amounts of time (all >= 2 days.) The rule based method used R and PostgreSQL to classify mode. First, static clusters were identified as all points in a minimum of one minute that had a speed < 3 km/h. A robust rule set was used to determine trip by one of the following rules: (1) Points with speed > 15 km/h, (2) a maximum of five sequential points bounded by two moving points identified by rule 1, (3) minimum of six continuous points with speed > 2.5 km/h, (4) points within 10 meters of a roadway but more than 25 meters of the center of any static cluster, (5) all other untagged points with speed > 10 km/h. Periods of movement were further refined then classified as indoor, outdoor static, outdoor walking or in-vehicle travel. The decision tree approach utilized R and Weka to classify transportation modes. Variables used were: (1) acceleration rate, (2) speed, (3) distance difference and (4) distance ratio. Overall, performance was similar in both models. Both models were efficient in correctly classifying vehicle trips and indoor points but were inconsistent in identifying outdoor static and walking trips.

Discussion

Methods have been developed to determine trip and/or transportation mode but no "gold standard" exists (Wu et al. 2011). The 17 studies reviewed had some overlap but little agreement exists between methods. Gi-Hyoug, et al. states that a walking trip occurs between time gaps of three minutes or more, must last five minutes or more and have a speed between two and eight km/h whereas Gong, et al. assigns a walking trip between time gaps of two minutes or more, must last one minute or more and have a speed less than 15 km/h. Gonzalez et al. claims their trained neural network can determine walking trips 100% of the time and car trips 92.11% of the time while Byon, et al. claimed their trained neural network was 92% during "rush hour" (4 P.M. to 7 P.M.) but dropped to 66% if used during other times.

Most articles show homogeneity of disciplines among the authors. This has resulted in non-standard language about the same topic. Transportation mode, travel mode, activity mode, travel behavior and other phrases describe the same idea but in different fields and for different purposes (see Figure 2.) Developing a "gold standard" requires the development of standardized language used across disciplines.

No study examined focused on multiple age ranges and/or geographies. The absence of studies across age groups leaves an assumption that a 20 year old has the same physical activity signature as a 70 year old. Although possibly true in reference to a car, this is most likely incorrect with respect to walking, bicycling and running. A study conducted in New York City reported different complications and classifiers than a study conducted in a smaller metro or rural area resulting from the number of travel mode choices (e.g. subway, commuter train, bus, etc.) and the density of travel routes (Gong et al. 2012).

Rule-based and machine learning approaches demonstrate specific strengths and weaknesses. Rule-based models can be robust but resource heavy. A properly trained machine learning model can be fast and accurate in determination but an inappropriate training set and/or classifiers can lead to a model that struggles to distinguish between subtle differences in travel modes. Machine learning methods examined in this review did not determine trip, which puts the burden of manual classification on the investigator.

Using handheld GPS, accelerometers or smart phones presents different challenges. Handheld GPS can be more accurate and battery efficient than a smart phone if a proper signal is achieved. However, a handheld GPS cannot rely on a wireless signal or cellular network in the absence of a proper signal nor can it transmit data as easily. An accelerometer is an excellent tool for characterizing motion but is not optimal for mode detection in comparison to GPS. A smart phone, however, features both elements. It is likely that smart phones will eventually be as powerful as any handheld GPS unit. This makes smart phones, along with their accelerometer feature, a likely candidate for trip and mode detection in the future.

METHODS

Data Collection

Participants

A convenience sample of 12 adults was recruited to perform specific travel modes while wearing a GPS and accelerometer unit. Inclusion criteria for subjects required they be currently enrolled students at Indiana University-Purdue University at Indianapolis (IUPUI), own a car, own a bicycle and capable of performing moderate physical activity. Subjects were recruited from various classes during the fall 2012 semester with permission from the presiding professor or instructor. All subjects were assigned a de-identified ID number for data collection. Demographic data was not collected from participants. Subjects were able to opt out at any time.

Protocol

Subjects were assigned a GPS unit (GlobalSat BT-335 Receiver) and an accelerometer (GoSmart HJ-151) in order to test travel mode detection algorithms. The GPS unit was set to collect location information every second while the accelerometer collected data in 1-min epochs. Subjects were asked to perform four bouts of walking, running, bicycling and driving within a two-month window. A bout was defined as a continuous performance of a given mode for a minimum of 15 minutes and a minimum distance of half a mile. Subjects were instructed to turn on the GPS device at least one minute prior to trip bout in order to avoid a “cold start”, where the GPS requires a certain amount of time to acquire the appropriate number of satellite signals. Participants were asked to turn off the GPS device after completion of a bout to save memory space and battery power. The accelerometer was left on at all times.

Subjects maintained a paper travel log through which they reported trip start and stop times as well as travel mode used. Students received a $50 VISA gift card upon returning the devices and travel log. Both GPS and accelerometer data were uploaded to a secured server as comma delimited files (csv) identified by subject ID. Travel logs were manually entered onto the secure server, identified by subject ID and then the paper logs were promptly destroyed.

Subject Data Issues and Caveats

Data collected on four of the twelve participants was not used due to the device being accidentally turned on prematurely and collecting the maximum amount of storable waypoints before the subject had performed any activity.   Subjects would occasionally fail to turn off the device, which consequently ran out of memory before all activities were completed. As a result, data on 93 total trips were collected (Table 3.)

Activity	Count
Car	36
Walk	24
Bike	17
Run	16

Table 3 Frequency of subject activities

Collected accelerometer data was inaccessible when the researcher attempted to recover that data. As such, accelerometer data was unavailable for use in study efforts.

Mode Detection Method Implementation

Objective

The goal was to replicate a sample of existing trip and travel mode detection methods derived from the previously described systematic literature review. The immediate interest was to use collected subject data and compare the accuracy of multiple methods on this sample. The purpose of applying these methods was for potential future development of a programmatic interface that would allow other researchers to easily access and utilize these methods and facilitate comparison of results.

Method Selection

Methods were selected from the aforementioned systematic review to be coded into this program. Criteria for method inclusion were: (1) Method was not developed for a mobile phone, (2) method was rule based and (3) all method details and/or algorithms were fully articulated in the associated publication. Phone based methods were excluded due to complexity and support needed to develop mobile-based programs for study use. Machine learning methods were excluded due to the complexity of comparing methods developed using a training set of data from another study. Methods that utilized accelerometers were also excluded due to aforementioned issues with collected accelerometer data. A total of three methods were selected for inclusion (Table 4.) The representation of mode in the four methods is (3 trips) Walk, (2 trips) Car, (1 trip) Bus, (1 trip) Bicycle, (1 trip) Rail/Train, and (1 trip) Subway.

Implementation

All methods were implemented using PostgreSQL 9.2 (The PostgreSQL Global Development Group, version 9.2. Available at http://www.postgresql.org), PostGIS 2.0 (PostGIS Project, version 2.0. Available at http://postgis.net) and Python 3.3.4 (Python Software Foundation, version 3.3.4. Available at http://www.python.org.) SQLAlchemy 0.9.6

(Michael Bayer, mike(&)zzzcomputing.com, version 0.9.6. Available at http://www.sqlalchemy.org) was the only Python dependency.

Lead Author	Year	Title	Modes
Cho et al.	2011	Identifying Walking Trips Using GPS Data	Walk
Gong et al.	2012	A GPS/GIS method for travel mode detection	Bus, Car, Rail, Subway, Walk
Bohte et al.	2009	Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands.	Bicycle, Bus, Car, Train, Walk

Table 4 Selected Methods for Review

Results

Bohte et al. Results

The Bohte et al. method (Bohte and Maat 2009) used four linear steps/rules with multiple subroutines for each step/rule. These steps were (1) Remove unreliable waypoints and divide points into trips, (2) Set the category of the trip (e.g. purpose), (3) set the mode of the trip and (4) merge and add train trips. Step 2 was not implemented because trip/travel purpose was inconsequential to both this study and had no bearing on mode determination outcome. Step 4 was ignored due to no participant using a train (or any public transit.)

The Bohte et al. method determined too few trips dividing subject data into 71 trips instead of the actual 93. The method's determination of time between trip start and trip end often did not line with the trip start and end times as logged by the subjects. This method claimed too many car/auto trips at 63 trips rather than the actual 36 trips. The method also determined trip and mode for series of waypoints not declared by the subjects in their logs. This raised questions about not declaring static trips (periods of non-movement or limited movement.) The Bohte et al. method failed to identify any walking or running trips from the subject logs.

Bohte Trip Travel Mode						
	Actual Car	Actual Walk	Actual Bike	Actual Run	Actual Uncategorized	Test Totals
Test Car	37	8	1	2	15	63
Test Walk	0	0	0	0	2	2
Test Bike	3	0	1	1	1	6
Test Run	0	0	0	0	0	0
Test Uncategorized	0	0	0	0	0	0
Actual Totals	40	8	2	3	18	71

Table 5 Bohte Confusion Matrix

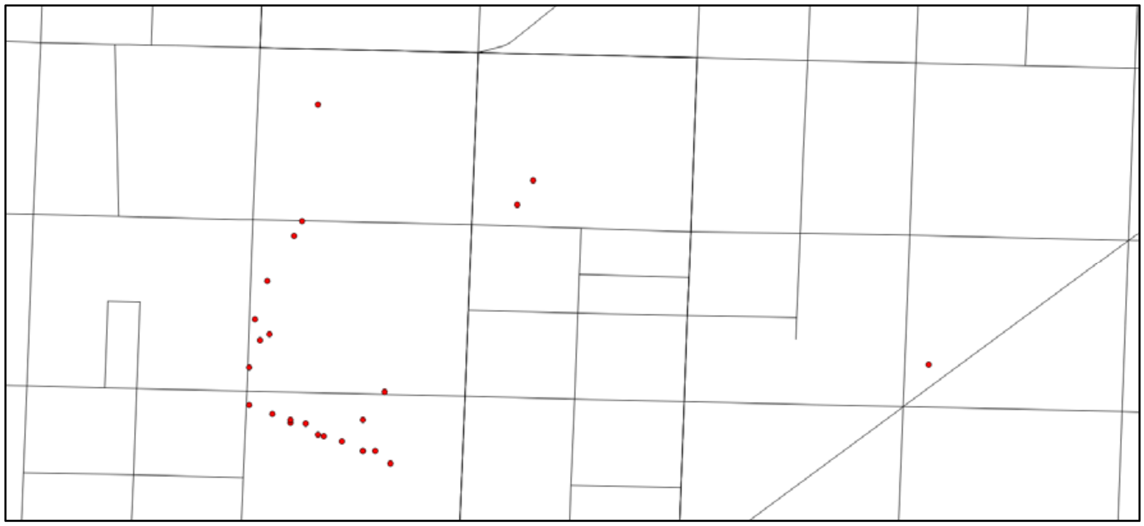

Figure 2 Subject Walk Trip - Categorized as Car by Bohte

Gong et al. Results

The method developed by Gong et al. (Gong et al. 2012) presented the highest level of false negatives and declared every single trip as a car/auto trip. There were four steps: (1) Prepare GPS data, (2) Divide GPS data into trips, (3) Divide Trips into trip segments and (4) Detect Mode. All four steps were successfully implemented.

Gong Trip Travel Mode						
	Actual Car	Actual Walk	Actual Bike	Actual Run	Actual Uncategorized	Test Totals
Test Car	33	14	5	3	38	93
Test Walk	0	0	0	0	0	0
Test Bike	0	0	0	0	0	0
Test Run	0	0	0	0	0	0
Test Uncategorized	0	0	0	0	0	0
Actual Totals	33	14	5	3	38	93

Table 6 Gong Confusion Matrix

This method correctly identified the total number of trips as recorded in the subject logs and claimed 33 car trips compared to the 36 as determined by the logs. The determined trips did not match start and stops times in the subject logs. The method was derived from data collected in New York City with certain expectations that subjects would walk to and from public transit stops. No subject in this study used public transportation during data collection.

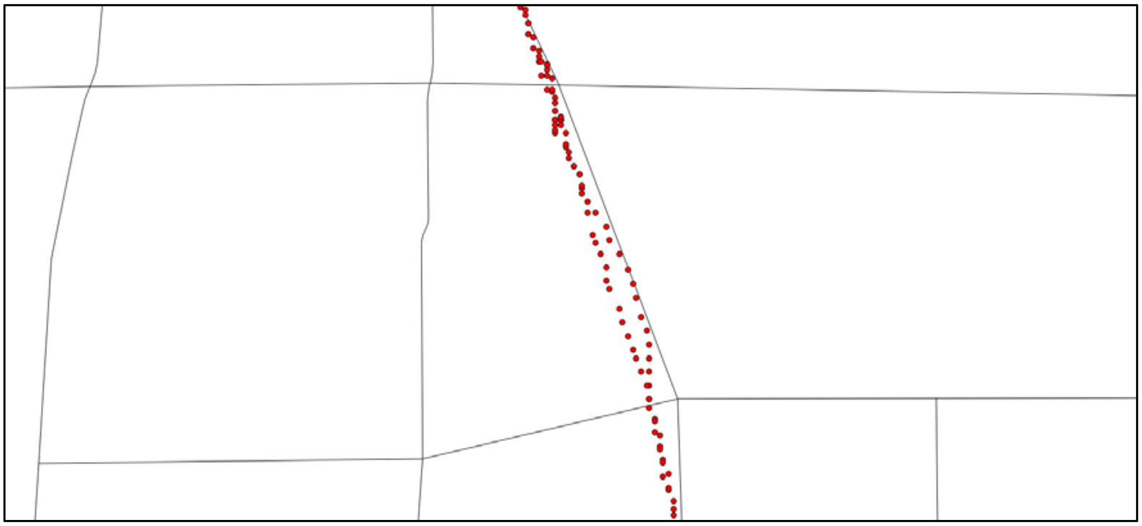

Figure 3 Subject Bike Trip - Classified as Car by Gong Method

Cho et al. Results

The method developed by Cho et al. (Cho et al. 2011) is focused solely on identifying walking trips. A decision tree was used to create this method with all steps repeated for this study. The decision tree for this method had rules for avoiding classifying indoor walking trips due to GPS being unreliable indoors.

Cho Trip Travel Mode						
	Actual Car	Actual Walk	Actual Bike	Actual Run	Actual Uncategorized	Test Totals
Test Car	0	0	0	0	0	0
Test Walk	14	8	2	0	15	39
Test Bike	0	0	0	0	0	0
Test Run	0	0	0	0	0	0
Test Uncategorized	25	3	2	3	30	63
Actual Totals	39	11	4	3	45	102

Table 7 Cho Confusion Matrix

The method determined 103 trips rather than the correct 93 trips. Based on start and stop times, Cho et al. method claimed eight instances of a walking trip with eleven of the determined trips matching walking waypoints in the logs. However, the logs claimed 24 total walking trips.

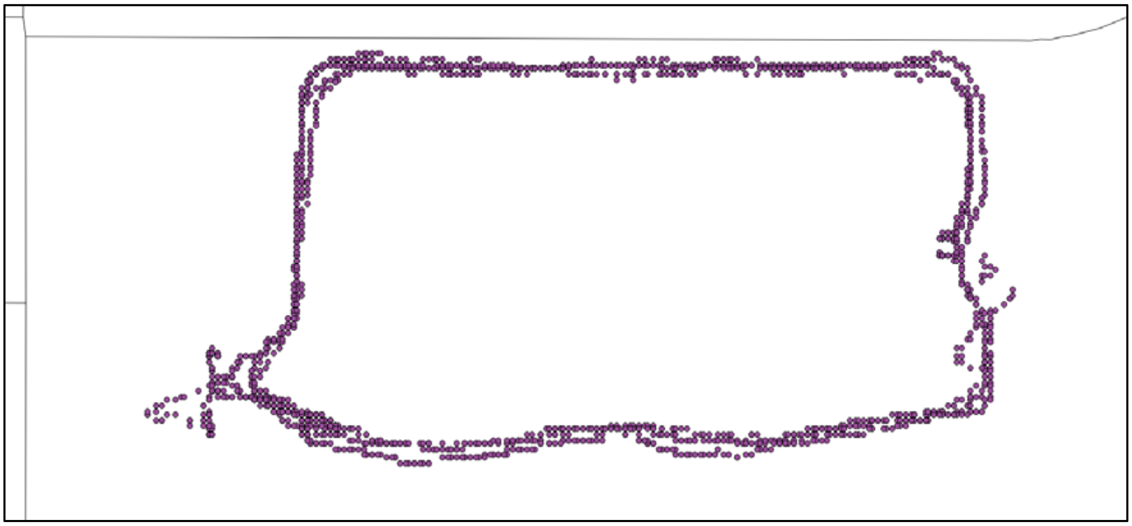

Figure 4 Subject Walking Trip - Correctly Classified by Cho

Conclusion

Determining trips and their travel mode is of significant value to multiple domains including transportation, informatics and health. Several existing methods were reviewed in this study with three being implemented on subject GPS data and corresponding logs. The number of methods along with the outcomes for the implemented methods demonstrates that no "gold standard" exists yet for determining trips and travel mode.

A few potential issues remain for developing a gold standard. A possible issue with inconsistent results is the geographic context of where the input data were collected. Trip patterns for various travel modes might have drastically different signatures between two cities, between an urban and rural environment or even between two cultures. Age of participants might be significant for both model development and model outcome. Existing methods are often created to serve a specific purpose as opposed to the development of a general purpose model or algorithm for determination of trip and travel mode.

A limitation in this specific study was the limited resources to test models designed around some form of machine learning. These models, in conjunction with modern smart phones, demonstrate the strongest potential in creating an "out-of-the-box" solution to determining trip and travel mode. Additionally, it would be optimal for researchers to make code or queries used to develop their model accessible by other researchers.

REFERENCES

Abe, M., Y. Morinishi, M. Atsuhiko, M. Aoki & H. Inagaki. 2009. A Life Log Collector Integrated with a Remote-Controller for Enabling User Centric Services. In *Consumer Electronics, 2009. ICCE '09. Digest of Technical Papers International Conference on*, 1-2, 10-14.

Anderson, I., J. Maitland, S. Sherwood, L. Barkhuus, M. Chalmers, M. Hall, B. Brown & H. Muller (2007) Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mob. Netw. Appl.,* 12**,** 185-199.

Auld, J., C. Williams, A. K. Mohammadian & P. Nelson (2008) An automated GPS-based prompted recall survery with learning algorithms. *Transportation Letters: The International Journal of Transportation Research,* 1**,** 59-79.

Bohte, W. & K. Maat (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,* 17**,** 285-297.

Byon, Y.-J., B. Abdulhai & A. Shalaby (2009) Real-Time Transportation Mode Detection via Tracking Global Positioning System Mobile Devices. *Journal of Intelligent Transportation Systems,* 13**,** 161-170.

Cho, G.-H., D. A. Rodriguez & K. R. Evenson (2011) Identifying Walking Trips Using GPS Data. *Medicine & Science in Sports & Exercise,* 43**,** 365-372.

Chung, E.-H. & A. Shalaby (2005) A Trip Reconstruction Tool for GPS-based Personal Travel Surveys. *Transportation Planning and Technology,* 28**,** 381-401.

Cooper, A. R., A. S. Page, B. W. Wheeler, P. Griew, L. Davis, M. Hillsdon & R. Jago (2010) Mapping the Walk to School Using Accelerometry Combined with a Global Positioning System. *American Journal of Preventive Medicine,* 38**,** 178-183.

Duncan, M. J., H. M. Badland & W. K. Mummery (2009) Applying GPS to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport,* 12**,** 549-556.

Gong, H., C. Chen, E. Bialostozky & C. T. Lawson (2012) A GPS/GIS method for travel mode detection in New York City. *Computers, environment and urban systems,* 36**,** 131-139.

Gonzalez, P. A., J. S. Weinstein, S. J. Barbeau, M. A. Labrador, P. L. Winters, N. L. Georggi & R. Perez (2009) Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks. *IET Intelligent Transport Systems,* 4**,** 37-49.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann & I. H. Witten (2009) The WEKA Data Mining Software: An Update. *SIGKDD Explorations,* 11.

Mavoa, S., M. Oliver, K. Witten & H. M. Badland (2011) Linking GPS and travel diary data using sequence alignment in a study of children's independent mobility. *International Journal of Health Geographics,* 10.

Oliver, M., H. Badland, S. Mavoa, M. J. Duncan & S. Duncan (2010) Combining GPS, GIS, and Accelerometry: Methodological Issues in the Assessment of Location and Intensity of Travel Behaviors. *Journal of Physical Activity & Health,* 7**,** 102-108.

Pober, D. M., J. Staudenmayer, C. Raphael & P. S. Freedson (2006) Development of Novel Techniques to Classify Physical Activity Mode Using Accelerometers. *Medicine & Science in Sports & Exercise,* 38**,** 1626-1634.

Quddus, M. A., R. B. Noland & W. Y. Ochieng (2006) A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. *Journal of Intelligent Transportation Systems,* 10**,** 103-115.

Reddy, S., J. Burke, D. Estrin, M. Hansen & M. Srivastava. 2008. Determining transportation mode on mobile phones. In *12th IEEE International Symposium on Wearable Computers*, 25-28. IEEE Computer Society.

Rodriguez, D. A., G.-H. Cho, J. P. Elder, T. L. Conway, K. R. Evenson, B. Ghosh-Dastidar, E. Shay, D. Cohen, S. Veblen-Mortenson, J. Pickrell & L. Lytle (2012) Identifying Walking Trips from GPS and Accelerometer Data in Adolescent Females. *Journal of Physical Activity & Health,* 9**,** 421-431.

Schüssler, N. & K. W. Axhausen. 2008. Processing GPS raw data without additional information. Eidgenössische Technische Hochschule, Institut für Verkehrsplanung und Transportsysteme.

Shay, E., D. A. Rodriguez, G. Cho, K. J. Clifton & K. R. Evenson (2009) Comparing objective measures of environmental supports for pedestrian travel in adults. *International Journal of Health Geographics,* 8.

Stenneth, L., O. Wolfson, P. S. Yu & B. Xu. 2011. Transportation Mode Detection using Mobile Phones and GIS Information. In *ACM SIGSPATIAL GIS '11*. Chicago, IL.

Stopher, P., E. Clifford, J. Zhang & C. FitzGerald (2008) Deducing mode and purpose from GPS data. *Institute of Transport and Logistics Studies Working Paper*.

Stopher, P., C. Fitzgerald & M. Xu (2007) Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation,* 34**,** 723-741.

Troped, P. J., M. S. Oliveira, C. E. Matthews, E. K. Cromley, S. J. Melly & B. A. Craig (2008) Prediction of Activity Mode with Global Positioning System and Accelerometer Data. *Medicine & Science in Sports & Exercise,* 40**,** 972-978.

Troped, P. J., J. S. Wilson, C. E. Matthews, E. K. Cromley & S. J. Melly (2010) The Built Environment and Location-Based Physical Activity. *American Journal of Preventive Medicine,* 38**,** 429-438.

Wolf, J., R. Guensler & W. Bachman. 2001a. Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data. In *Transportation Research Board 80th Annual Meeting*. Washington, D.C.

Wolf, J., M. Loechl, J. Myers & C. Arce. 2001b. Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys. In *International Conference on Transport Survey Quality and Innovation*. Kruger Park, South Africa.

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

Xu, C., M. Ji, W. Chen & Z. Zhang. 2010. Identifying Travel Mode from GPS Trajectories through Fuzzy Pattern Recognition. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*. Yantai, China.

Zandbergen, P. A. & S. J. Barbeau (2011) GPS Data from High-Sensitivity GPS-enabled Mobile Phones. *The Journal of Navigation,* 64**,** 381-399.

Zhang, L., S. D. Gupta, J.-Q. Li, K. Zhou & W.-b. Zhang. 2011. Path2Go: Context-Aware Services for Mobile Real-Time MultiModal Traveler Information. In *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 174-179. Washington, D.C.

CURRICULUM VITAE

Aaron W Burgess

Education

2016 M.S. Geographic Information Science, Indiana University earned at IUPUI

2011 B.A. Geography, Indiana University earned at IUPUI

Honors, Awards, Fellowships

New Geography Major Alumni Scholarship (2008)

Undergraduate Research Opportunities Program Scholar (2008)

IUPUI School of Liberal Arts Featured Student (Fall 2009)

Mary J. Crisler Scholarship for Research (2010)

Plater Medallion for Civic Engagement (2011)

Outstanding Graduating Senior in Geography (2011)

Center for Urban Health Fellow (2012 – 2013)

Research and Training Experience

2012-2013 Research Fellow, Center for Urban Health, IUPUI

2010-2013 Research Assistant to Sarah Wiehe, M.D., MPH, Children's Health Services Research, IU School of Medicine

2010-2012 Research Assistant to Jeff Wilson, Ph.D., Department of Geography

Professional Experience

2015-Present Software Data Engineer, Milliman, Indianapolis, IN

2015 Software Engineer, Blackboard, Indianapolis, IN

2013-2015 Senior Software Engineer, Maponics, White River Junction, VT

Presentations

Burgess A, Wilson JS, Hoch SC, Wiehe SE. Perl Girlz: Using GIS for a Health Ecological Model. *ESRI User Conference* (San Diego, CA), July 2012.

Wiehe SE, Wilson J, **Burgess AW**, Guterl S, Hoch SC, Fortenberry JD. Physical Activity and Weight Status Associated with Self−Reported Context, Not Census, Observed or Other Administrative Contextual Data. *Pediatric Academic Societies Meeting* (Boston, MA), April 2012.

Roth AM, Fortenberry JD, Van Der Pol B, Reece M, **Burgess A**, Wiehe SE. Situational Context & STD Risk among Street-based Sex Workers in Indianapolis. *Presented by Roth at the annual Midwestern regional meeting of the Society for the Scientific Study of Sexuality* (Bloomington, IN), May 2012.

Wiehe SE, Roth SM, **Burgess A**, Arno J, Fortenberry JD. Redefining the Social Geography of Community STI Risk: An Ecological Study of the Association, Mediators, and Moderators of Area-Level Prostitution Arrests. *CDC National STD Prevention Conference* (Minneapolis, MN), March 2012.

Wiehe SE, Kwan M, Hoch SC, Brooks B, **Burgess A**, Wilson JS, Fortenberry JD. Adolescent Sexual Intercourse and Neighborhood Social Disorder. *ISSTDR Conference* (Montreal, QC), July 2011.
Burgess A, Wilson JS, Hoch SC, Wiehe SE. Methods for Determining Linear Paths and Locational Position using GPS Data Points. *American Association of Geographers Annual Meeting* (Seattle, WA), April 2011.

Dugan T, Anand V, **Burgess A**, Brooks B, Wiehe, SE. An Open Source Solution for Tracking Adolescent Travel Patterns: The Pearl Grlz Application. *ICEPHI Symposium* (Indianapolis, IN), April 2011.

Burgess A, Wilson JS, Whitcomb HA, Troped, PJ. Methods for Detecting Use of Community Trails Using GPS Data. *Association of American Geographers West Lakes Annual Meeting*, (Macomb, IL), October 2010.