 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

Review

Travel mode detection based on GPS raw data collected by smartphones: a systematic review of the existing methodologies

Linlin Wu, Biao Yang* and Peng Jing

School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China;
E-mails: llwu@ujs.edu.cn (L.W.); jingpeng@ujs.edu.cn (P. J.)

* Author to whom correspondence should be addressed; E-Mail: 18796018897@163.com;
Tel.: +86-187-9601-8897

Abstract: Over the past couple of decades, Global positioning system (GPS) technology has been utilized to collect large-scale data from travel surveys. As the precise spatiotemporal characteristics of travel could be provided by GPS devices, the issues of traditional travel survey, such as misreporting and non-response, could be addressed. Considering the defects of dedicated GPS devices (e.g., need much money to buy devices, forget to take devices to collect data, limit the simple size because of the number of devices, etc.), and the phenomenon that the smartphone is becoming one of necessities of life, there is a great chance for the smartphone to replace dedicated GPS devices. Although, several general reviews have been done about smartphone-based GPS travel survey in the literature review section in some articles, a systematic review from smartphone-based GPS data collection to travel mode detection has none. The included studies were searched from six databases. The purpose of this review is to critically evaluate the current literature on the existing methodologies of travel mode detection based on GPS raw data collected by smartphones. Meanwhile, according to the systematic comparison among different methods from data-preprocessing to travel mode detection, this paper could carefully provide the Strengths and Weaknesses of existing methods. Furthermore, it is the crucial step to develop the methodologies and applications of GPS raw data collected by smartphones.

Keywords: travel mode detection; GPS raw data; smartphones

1. Introduction

Due to the worsening traffic congestion, transportation demand modeling and travel behavior research have played more and more significant roles in the formulation and evaluation of transportation demand management policies over the past two decades. In practice, travel surveys are widely used to collect necessary and required data which is the crucially infrastructural data for traffic demand analysis in transportation system planning [1, 2].

The traditional travel survey methods have gone through some stages. In the 1950s, the first travel survey approach, the face-to-face interview, was used in the field of urban transport planning, in which interviewers needed to visit participants' homes and ask questions about the household's travel information and the interviewers used paper and pencil to record the answers. In the 1960s, however, considering the weaknesses of the face-to-face interviews, such as the safety and cost issues, the method began to be replaced by the mail-out/mail-back survey [3]. But the low response rate was one of the problems of the mail survey. Furthermore, the data, the postal survey collects, needed to be transferred from paper to computers, which needed a great deal of manpower [4]. In the 1980s, for the purpose of surmounting the shortcomings of paper-and-pencil interviews (PAPI), computer-assisted surveys were introduced. Computer-assisted survey contains three main types: the computer-assisted telephone interview (CATI), the computer-assisted personal interview and the computer-assisted self-interview (CASI) [5]. However, all of these approaches had some disadvantages, such as misreporting [3], non-response [6]. Thus, in order to overcome the disadvantages of these approaches, the methods, collecting travel data automatically, must be considered.

For the sake of improving the accuracy and quantity of travel data and supplementing the traditional data elements which were collected on paper or electronic travel diaries, the Global Positioning System (GPS) technology, proving the accurate data, such as location, time, speed, heading and so on, was used in travel survey in the middle of 1990s [7, 8]. Over the past couple of decades, GPS-based surveys have been undertaken in lots of countries, such as the USA, the UK, Australia, Austria, Canada, China, Demark, France, Israel, the Netherlands, Japan, Sweden, Switzerland and so on[9-22]. Meanwhile, it has been widely recognized that GPS-based data collection methods can present obvious advantages over traditional travel methods. On the one hand, GPS-based data collection methods can impose fewer requirements on the respondents, provide greater spatial and temporal precision and are capable of reducing the labor and time costs [23]. On the other hand, removing the burden and fatigue from the survey respondents and allowing researchers to collect detailed travel data, are other important advantages of GPS-based data collection methods [24, 25]. Because of the very low level of burden and fatigue of respondents, the surveys' length can be extended from the traditional single day to multi-day travel information collection, which provides a chance to test the dynamics of multi-day travel patterns [1, 26]. Although time and positional characteristics of travel can be recorded accurately by GPS devices, the important attributes, such as travel mode, trip purpose and start and end of trip, cannot be extracted from the data collected by GPS devices. Therefore, data-processing procedures become useful and necessary, because the GPS raw data would be insufficient for travel modelling purpose without the results of the data-processing procedures [4, 26].

A number of methods for processing GPS data for application to a GPS-based travel survey have been studied. Among those studies, the majority concentrated on identification of travel modes. Lots of approaches have been applied in inferring travel modes based on GPS data collected by dedicated GPS devices, such as Rule-based Method [26, 27], Bayesian Model with Expectation Maximization [28], Fuzzy Logic Approach [20], Bayesian Belief Network Model [29], Multilayer Perceptron [30], Support Vector Machine [31], and so on. Travel surveys, however, based on dedicated GPS devices have the following disadvantages: (1) researchers need to spend large amounts of money buying dedicated GPS devices; (2) forgetting to take the GPS devices results in the data collection incomplete; (3) the number of dedicated GPS devices is a limitation of the sample size; (4) in GPS-based travel survey, dedicated GPS devices need to be distributed to and retrieved from participates [32].

In the past years, the smartphone is becoming one of the necessities of daily life and usually equipped with the GPS module, which provides a great chance to use smartphones to replace dedicated

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2016 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at Information 2016, 7, 67; doi:10.3390/info7040067

GPS device to collect travel data [33]. Some smartphone-based GPS travel surveys have been conducted in these studies [34-35]. Because of the increasing popularity of smartphones, the probability that people forget to carry their own smartphones is very low when they go out. Thus, utilizing smartphones to collect travel data would reduce expenditure of surveys and ensure data collection complete. In addition, the accelerometer sensor is also built into some smartphones, and using smartphones, equipped with GPS and accelerometer sensors, can record more data which could be used to recognize travel modes [29]. Of course, smartphone-based GPS surveys have some weaknesses: (1) the short battery lives of smartphones (compared with dedicated GPS devices); (2) the unstable signal acquisition in certain areas, such as urban canyons; (3) the high cost of transferring data from phones to data centers [4]. Due to the weaknesses of smartphone-based GPS surveys, it is of importance to choose a properly methodology to process these raw data and utilize the processed data.

The focus of this study is to assess the existing methodologies of detecting travel modes based on GPS data collected by smartphones. In this study, the purpose is to provide a systematic review of the existing methodologies of travel mode detection, compare the different data processing methods which have applied in existing researches and analyze the pros and cons of these methods. Meanwhile, it is the crucial step to develop the methodologies and applications of GPS data collected by smartphone.

The rest of this paper is organized as follows. The methods, searching eligible papers, are described in Section 2. Systematic review process is presented in Section 3. Section 4 describes the quality of reviewed studies. Limitations and strengths of this paper are proposed in Section 5. Finally, the discussion and conclusion are provided in Section 6.

2. Methods

2.1. Search Strategy and databases searched

In the light of the PRISMA (preferred reporting items for systematic reviews and meta-analyses) statement [36], six databases, including Web of Science (1997- December 2015), ScienceDirect (1997-December 2015), Academic Search Complete (1997-December 2015), Scopus (1997-December 2015), Cambridge Journals Online (1997-December 2015) and TRID (1997-December 2015), were searched using keywords contained in the title, abstract, mesh heading and eligible terms (the reason selecting 1997 as the start time of search is that the first time using GPS technology for travel survey is in 1997). There are three categories of search terms and at last one of each three categories of search terms must be used to combine: 1) smartphone, cell phone, and mobile phone; 2) GPS data, GPS trajectory, GPS raw data, and GPS track; 3) travel mode, transportation mode, movement patterns, travel mode detection, travel mode identification, travel mode recognition, infer travel mode, identify travel mode, detect travel mode, and detect transportation mode, infer transportation mode, deduce travel mode, classify travel mode, and identify transportation. Considering the specific structure of each database, the search must be adapted to match the database. It was an important step to examine the previous reviews. The references within identified articles, of course, were also reviewed for further studies.

2.2. Inclusion and exclusion criteria

In order to make sure that each study included in the review were eligible, studies had to: 1) be written in English and be published in a peer-reviewed English journal; 2) use smartphone as a tool to collect data; 3) at least the GPS data collected by smartphone; 4) at least relate to travel mode detection procedure; 5) at least one variable related to travel as a dependent variable.

2.3. Data extraction

Standardized data extraction table was extracted from included papers using matrix method. The information abstracted from each eligible article included study characteristics (e.g., study design, study area, study duration), data-preprocessing methods (e.g., data error recognition), trip/segment identification methods (e.g., feature selection reliability testing, parameter selection reliability testing), travel mode detection methods (e.g., feature selection reliability testing, comparison of experimental results). For the credibility of data extraction, authors drew a subsample of 8 papers and extracted the data independently. The authors approved of 80% of the extracted data indicating high inter-rater reliability.

2.4. Quality assessment

The quality of the included studies in the review was carefully assessed through a modified checklist combining the data collecting methodological quality scale and data-process methodological quality scale.

To capture influences affecting the quality of data collecting methods, the modified checklist includes five aspects: study design, the basis of sample size selection, survey duration, measures to overcome drawbacks and ground truth. The study design may include travel survey and experimental survey. As the quality of GPS data of experimental survey might be better than multi-day travel, which is likely to influence the accuracy of the data-processing methods [4]. Study design was added in the checklist as one of the criteria. Calculating the adequate sample size is of importance to determine the number of participants of travel survey [31]. Thus, the adequate sample size selection was chosen in the checklist as one of the criteria. The reason why survey duration is collected over the two weeks is to ensure the natural flow of the travel patterns of each participant [31]. The drawbacks of smartphone-based GPS travel surveys are likely to affect the accuracy of GPS raw data, which further influence the accuracy of the data-processing methods. Therefore, the measures, overcoming the drawbacks, are essential for ensuring the accuracy of GPS raw data. Ground truth is used to calculate the accuracy of travel mode detection, so ground truth is collected to be one of the criteria [4].

In order to catch influences affecting on the quality of data-processing methodological, the modified checklist includes three aspects: data-preprocessing methods, trip/segment identification methods and travel mode detection methods. It is very vital to recognize and clean the error data and increase the accuracy of data before they can be used in next part, so it is a necessary step to provide a statistical basis for the choice of independent variables in the trip/segment identification and travel mode detection [31]. The function of comparison of different experimental results is to show whether the proposed approach is the best.

All included studies were evaluated on 10 criteria listed in Table 1. The possible range of evaluation scores was 2 to 12.

Peer-reviewed version available at Information 2016, 7, 67; doi:10.3390/info7040067

Table1. Checklist for evaluating studies' quality.

Criteria	Description	Score
Assessing data collecting methodological quality		2-7
Study design	Travel survey	2
	Experimental survey	1
Adequate sample size selection	Included	1
	Not included	0
Survey duration	More than 2 weeks	2
	less than 2 weeks	1
Overcoming drawbacks of measures	Included(e.g., the short battery lives, the signal loss)	1
	Not included	0
Ground truth	Included (e.g., prompted recall survey)	1
	Not included	0
Assessing data-processing methodological quality		0-5
Data-preprocessing methods		0-1
Data error recognition	Included	1
	Not included	0
Trip/segment identification methods		0-2
Independent variables selection reliability testing	Testing	1
	Not testing	0
Parameter selection reliability testing	Testing	1
	Nor testing	0
Travel mode detection methods		0-2
Independent variables selection reliability testing	Testing	1
	Not testing	0
Comparison of experimental results	Included	1
	Not included	0

3. Systematic review process

The search and retrieval process were shown in Figure 1. The number of papers searched from each database mentioned above were 1584 (web of science), 1552 (ScienceDirect), 2107 (Scopus), 225 (Academic Search Complete), 107 (TRID), 79 (Cambridge Journals Online). After duplicates being removed, a total of 4137 different records were extracted from six databases, of which 265 were identified following the screening of titles and abstracts. There are three cases to exclude ineligible records: the first case is that the GPS raw data is not collected by smartphone; the second case is that the collected data is not used to detect travel modes; and the last case is that the full-text is not available. Thus, full texts of 12 publications were retrieved. Three reviews excluded and the reference lists of excluded reviews were reviewed and potential articles were identified. Finally, 12 published papers matching all the criteria were included in this review [1, 37-47], as shown in Table 2. It needs to give the reason to select the study of [38], because the segment identification method this study

doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

introduced is a significant data-preprocessing procedure for travel mode detection.

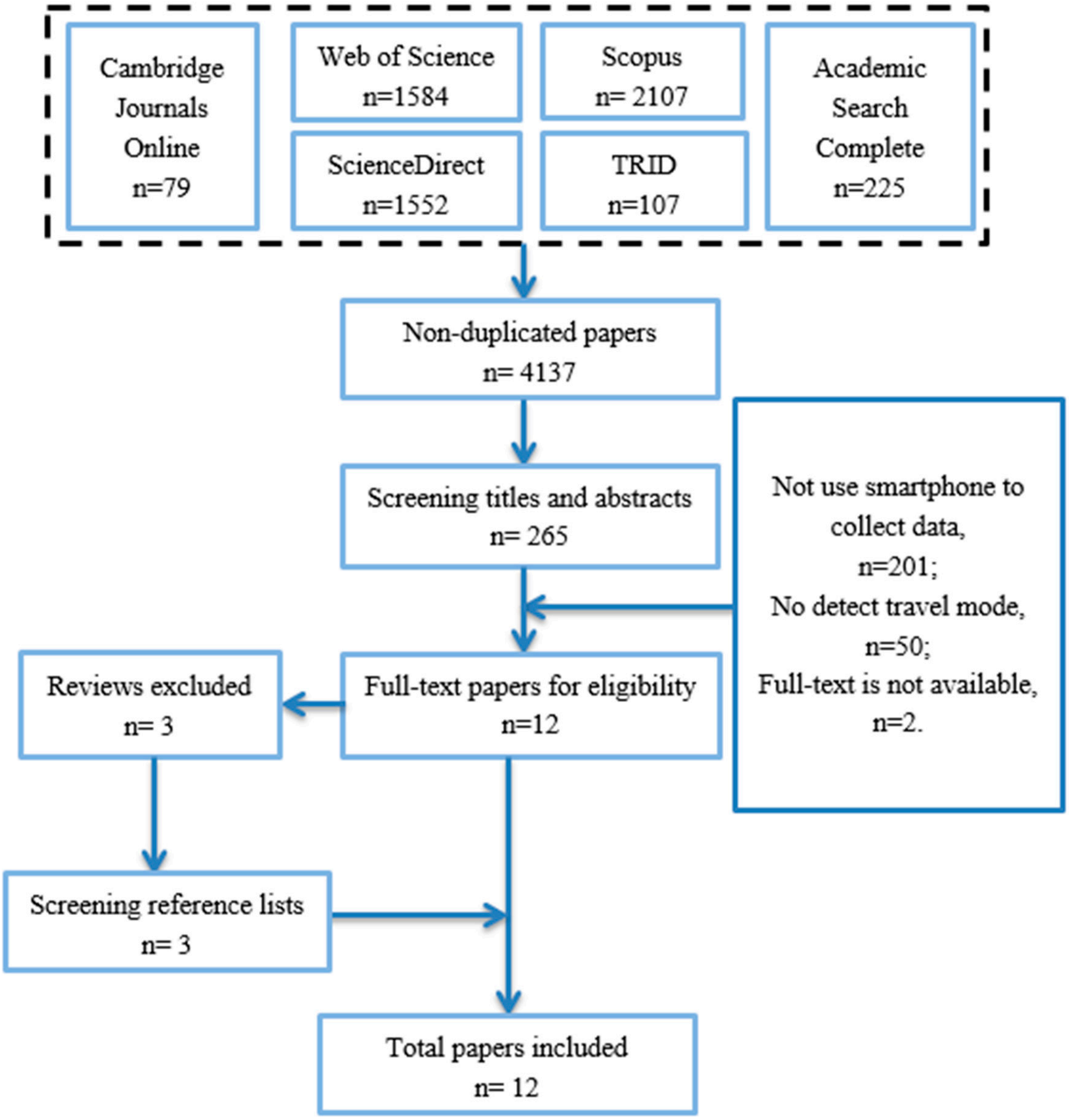

Figure 1. The flow chart of systematic review process.

doi:10.20944/preprints201607.0067.v1

Table 2. Summary of studies included in this systematic review.

Lead Author (Year)	Location	Journal	Sample Size	Collection Period	Device	Technical Details	Processing Involved [a]
Xiao, G. (2015)	Shanghai, China	Information	N/A	Mid-October 2013 to mid-July 2014	GPS-enabled smartphone	Random sampling; GPS-only survey;	TI,MD
Xiao, G. (2015)	Shanghai, China	Computers, Environment and Urban Systems	N/A	Mid-October 2013 to mid-July 2014	GPS-enabled smartphone	Random sampling; GPS-only survey;	TI,MD
Xiao, G. (2015)	Shanghai, China	Transportation Research Board 94th Annual Meeting	N/A	Mid-October 2013 to late-May 2014	GPS-enabled smartphone	Random sampling; GPS-only survey; Every participant is required to attend the survey at last five days.	TI
Lari, Z.A. (2015)	Tehran, Iran	Transportation Research Board 94th Annual Meeting	35 participants (25 males and 10 females)	2 weeks	Smartphone equipped with GPS and accelerometer sensors;	Random sampling: running the application from 6 a.m. to 9 p.m.	MD
Yang, F. (2015)	Chengdu, China	Transportation Research Record: Journal of the Transportation	20 persons	N/A	Mobile phone	Volunteers are required to collect data about special multimode trips.	TI,MD

		Research Board					
Nitsche, P. (2014)	Vienna, Austria	Transportation Research Part C: Emerging Technologies	15 volunteers	2 months	Android-based smartphone	Random sampling;	TI,MD
Stenneth, L. (2011)	USA	Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems	6 individuals (3 males and 3 females)	3 weeks	GPS-enabled mobile phone	Random sampling; GPS-only survey	MD
Zhang, L. (2011)	Hanover City, Germany	Remote Sensing and Spatial Information Sciences	137 sub-traces	N/A	Android-based smartphone	Random sampling; Tracer Android App has a travel-mode selection function.	SI,MD
Gonzalez, P.A. (2010)	USA	Intelligent Transport Systems	114 trips	N/A	GPS-enabled cell phone	Random sampling; GPS-only survey	MD
Reddy, S. (2010)	USA	ACM Transactions on Sensor Networks (TOSN)	16 individuals (8 males and 8 females)	75 minutes	Nokia N95 equipped with GPS and acceleration sensors;	Random sampling; GPS-only survey; fifteen minutes of data for each of the five transportation modes	MD

doi:10.20944/preprints201607.0[illegible]/v1

Reddy, S. (2008)	USA	12th IEEE International Symposium on.	6 individuals (3 males and 3 females)	12 hours	Mobile phone equipped with a GPS receiver and a 3-axis accelerometer	Individuals are required to traverse both typical street and highway roads.	MD
Zheng, Y. (2008)	Beijing, China	Proceedings of the 17th international conference on World Wide Web	45 persons	6 months	GPS phone and handheld GPS receivers	Random sampling; GPS-only survey;	SI, MD

[a] TI= Trip Identification, SI= Segment Identification, MD= Mode Detection

 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

3.1. GPS data-processing procedure

The common procedure of GPS data processing to detect travel modes consists of three significant parts: the first part is to transfer data collected from GPS-enabled smartphone to computer and create output files that could be used for next statistical analysis; the second part is to identify trips/segments; and the last part is to detect travel modes according to the previous processed data.

3.2. Data preprocess procedure

The manual mistakes, such as inaccurate time and underreporting trips, can be avoided by GPS raw data collected from smartphone. However, it is a pity that the GPS raw data may have the systematic errors. Hence, it is necessary to preprocess the raw GPS data before they can be utilized to the next steps. The typical data preprocess procedure can be divided into two parts: records' features used for error recognition and methods or steps of data transformation. And the summary of data preprocess procedure is shown in the Table 3. Detailed description of methods for data pre-processing in eligible papers is in the paragraphs below Table 3.

Table 3. Summary of data error recognition and pre-process in selected papers.

Year	Lead Author	GPS Devices	Records' Features Used for Error Recognition	Methods or Steps of Data Transforming
2015	Xiao, G.	GPS-enabled smartphone	Number of satellites, HDOP value, altitude value	Using three steps to implement data transforming
2015	Lari, Z.A.	Smartphone equipped with GPS and accelerometer sensors;	Maximum speed values of different modes	N/A
2014	Nitsche, P.	Android-based smartphone	N/A	Using the Kalman filter to preprocess the track data; and transforming the data of tri-axial accelerometer
2011	Stenneth, L.	GPS-enabled mobile phone	The GPS accuracy, the change in speed	N/A
2011	Zhang, L.	Android-based smartphone	The use of smoothing method	The values of speed and heading
2010	Reddy, S.	Nokia N95 equipped with GPS and acceleration sensors	The accuracy(vertical, horizontal, heading and speed), dilution of precision(time, vertical, horizontal), the changes in speed values of single and accelerometer sampling frequency considered	N/A

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2016 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

Xiao, G. *et al.* [38] utilized 3 rules to remove the incomplete or invalid data. The first rule is that the incomplete track points that may indicate wrong records were removed. The second one is that the record with less than four satellites (for 3-D use) or with the HDOP of four or more were eliminated. And the last one is that the track points with an altitude of more than 200 meters were deleted. They also used three steps to preprocess the cleaned track data. The first step is to convert the UTC time to local data and time. The second step is to extract track data for each person-day based on user ID and the local date. And the last step is to re-number all track points of one person-day to inference trip end conveniently.

Lari, Z.A. *et al.* [39] found that the speed, an important attribute calculated based on the three geographical parameters: altitude, longitude, and latitude, was a significant factor affecting the accuracy of recoding data, so they used the maximum speed values of different modes to clean the raw data.

Nitsche, P. *et al.* [41] utilized Kalman filter to preprocess the track data. For instance, they combined the raw GPS and cell location data with the predictions of a linear model assuming zero mean, Gaussian-distributed acceleration to calculate accurate and smooth tracks. They also transformed the data of tri-axial accelerometer into a rotation-invariant signal because the direction of the three-dimensional acceleration vector had a great impact on the accuracy of accelerometer data.

Stenneth, L. *et al.* [42] suppressed the invalid GPS points based on the accuracy level of the latitude and longitude coordinates and the change in speed.

Zhang, L. *et al.* [43] calculated the values of speed and heading from point positions and time stamps and utilized the smoothing method to reduce speed errors by averaging its neighborhood.

Reddy, S. *et al.* [45] recognized and discarded the invalid GPS points based on analyzing the accuracy of vertical, horizontal, heading and speed, dilution of precision of time, vertical, horizontal, and the changes in speed values of the GPS signal. Of course, they cleaned the accelerometer data as well. They pointed out that this type of data, too few samples received from the accelerometer, should be excluded.

However, data preprocessing procedure was not mentioned in these papers of [40, 44 and 46]. And the data preprocessing methods were just simply referred in the papers of [1, 37 and 47].

3.3. Trip/segment Identification

In this part, the concept of trip, being also defined as a segment, just mentions a one-mode trip. For all researchers, the first challenge of GPS data-processing procedure would be the trip identification (TI) or segment identification (SI). The data about each trip, travel mode detection and trip purpose imputation which the travel model needs, is based on the results of TI. A summary of attributes used for TI/SI in selected papers is shown in Table 4.Currently, the rule-based algorithms are used by most researchers to undertake the TI/SI procedure.

Table 4. Summary of records' parameters used for TI/SI in selected papers.

Year	Lead Author	Attributes	Accuracy
2015	Xiao, G.	Critical length, critical distance, dwell time	96.02%
2015	Yang, F.	Maximum lines	95%
2014	Nitsche, P.	Speed threshold, high amplitudes accelerometer signal,	N/A
2011	Zhang, L.	Small speed values, small change in position, large magnitude in heading change	N/A
2008	Zheng, Y.	Change point, uniform duration, and uniform length	N/A

Xiao, G. *et al.* [38] found that the specify values of parameters, used to infer trip ends with rule-based methods in most existing studies, might lead to the relatively bad performance of rule-based methods. Thus, in order to optimize the parameter values used to detect trip ends from GPS track data streams under two situations: single loss and normal recording, they focused on developing algorithm to inter trip ends. Under the single loss situation, they defined two different destinations: habitual destination and non-habitual destination. And two different parameters were defined for dwell time to judge whether there existed a trip end for a stop near habitual destinations or non-habitual destinations. Under the single available situation, due to the existence of two types of trip ends, they extracted two different parameters (critical length and critical distance) to identify trip ends. Each parameter had five values which could constitute 625 distinct parameter combinations. By calculating and comparing the results of different combinations, the best result could be obtained: 96.02% accuracy with a low error rate of 4.74%. Although the improved algorithm obtained the high accuracy of SI, the choice sets of four parameters actually lacked theoretical research to support.

Nitsche, P. *et al.* [41] used speed threshold and amplitudes of the accelerometer signal to avoid coverage of multiple travel modes. They, however, just considered the signal available situation, and missed the signal loss situation. Moreover, the value of speed threshold was not pointed out.

Zhang, L. *et al.* [43] derived individual travel-mode segments from GPS traces by identifying stops. As they mentioned, the very low speed and very small distance changes could be defined as the stops, and large magnitude in heading changes was vital to identify stops, so the three parameters were selected for the identification of stops. The thresholds of three parameters were chosen: the distance change for 5 continuous points is less than 5 meters; the speed value for 5 continuous points is less than 0.5m/s; and the change of heading of 5 continuous points is larger than 100 degrees.

Zheng, Y. *et al.* [47] utilized change point based on segmentation approach to divide each trip into segments. In order to verity the validity of this segmentation approach, they also selected two baseline methods, uniform duration based and uniform length based segmentation, to distinguish the trips.

Yang, F. *et al.* [40] adopted another method, the wavelet transform modulus maximum (WTMM) algorithm, to undertake the SI processing, because they found the thresholds of travel time and distance for SI could not be directly applied to another study. The Gaussian family [Gaus (n)] was selected because of the best performance in mode transfer time detection compared with Haar, the Daubechies family. As a result, the accuracy of SI is more than 95%. However, the GPS signal loss situation such as a long period of signal loss time, which could have an impact on the results of SI, didn't be considered. Meanwhile, seven types of special multimode travel couldn't represent all

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2016 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

combinations of travel modes, such as walk-bus-bicycle (public bicycle)-walk-subway-walk. Moreover, it might be unreasonable to see walk as the bridge of conversions of two different travel modes.

Lari, Z. *et al.* [39] used appropriate application to collect per GPS track that recorded an amount of vital information, such as the segments, time, date, instant speed, accuracy, bearing, altitude, latitude and longitude. Thus, the segment could be gained directed by the application, and the paper did not refer to the SI procedure. Unfortunately, the papers of [42, 44, 45, and 46] did not provide details of TI/SI procedure.

3.4. Travel mode detection

Travel mode detection is the third part of GPS data-processing procedure. Table 5 shows a summary of different approaches used for travel mode detection in the selected papers. Two categories of methods for travel mode detection applied in included papers consist of machine learning methods and hybrid methods. Detailed description of machine learning methods for MD in each selected papers is in the next seven paragraphs.

Gonzalez, P.A. *et al.* [44] chose neural networks (NNs) to identify travel modes. As they mentioned, NNs have two significant strengths. The first strength is that NNs can capture the information from data that is neglected by people and other analysis algorithms. The second strength is that NNs are able to summarize conclusions for new data, which don't completely match the training data. In their research, two datasets, namely all GPS points and critical GPS points, were used. And neural network input attributes varies depending on the dataset. They applied 10-fold cross-validation to train and test the neural network. From their results, the highest accuracy achieved for travel modes detection is as high as 91.23% for the only critical points using a learning rate of 0.1 and training time of 300 epochs. However, their research has two limitations. The first one is that the sample set of trips for the training is not enough. The second one is that the GPS data is manually segmented by the participators. It should be noted that the two limitations might affect the accuracy of mode detection.

Yang, F. *et al.* [40] selected the neural network (NN) algorithm to determine the travel mode of each trip segment. As a result, the accuracy of travel mode detection is more than 86%. Moreover, the accuracy of bus mode detection is higher than accuracy in any other studies. However, there is two limitations in this study: the first point, authors didn't consider the defect of traditional NN algorithm, for instance, traditional NN algorithm is likely to be trapped in local optimum when generally trained by back-propagation algorithms, which could affect the accuracy of travel mode detection; the second point, the results comparison lacks of rationality, due to the different studies use different quality of data, which has a great impact on the results.

Xiao, G. *et al.* [1] pointed out that traditional NNs were likely to be trapped in local optimum when they were generally trained by back-propagation algorithms. In order to solve this problem, they utilized particle swarm optimization (PSO) to search for a global optimum. In their analysis, they selected five features to infer travel modes according to exiting studies: the average speed, medium speed, average absolute acceleration, travel distance and 95th percentile speed. However, the result of distinguishing bus segments from car segments was not good when only the above-mentioned speed-related features were used. To address the issue, they developed a new feature named

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

"low-speed point rate" and made use of the two-sample Kolmogorov-Smirnov test between bus segment and car segment to ensure the value of "low-speed point rate". They divided raw data randomly into two separate subsets consisting of 25% and 75% of data used for testing and training, respectively. From their results, the accuracy of mode detection for training set is 95.81% and the mode-identification accuracy for test set is 94.44%. Because of the relatively lower accuracy of distinguishing bus and car segments, the potential features which could differentiate these two segments should be added to improve accuracy.

Another machine learning method currently adopted in travel mode detection is Bayesian networks. Xiao, G. *et al.* [37] utilized Bayesian networks to detect modes based on GPS data collected in a smartphone-based travel survey from mid-October 2013 to mid-July 2014. In authors' analysis, they used a K2 algorithm to establish the structure of Bayesian networks and estimated corresponding conditional probability tables with maximum likelihood methods. They extracted four features (the average speed, 95% percentile speed, the average absolute acceleration, and travel distance) to construct the Bayesian networks to identify the travel modes. In order to further improve the mode-identification performance of the Bayesian networks model, they added two targeted features, named low speed rate and average heading change, to the feature set. And according to the comparison of results of the original and updated Bayesian network, the updated Bayesian network is a better mode–identification performance. Although the improved Bayesian network achieved a better result, they also suggested that it still had much room for improvement, such as using GIS sources, and adding potential features which could markedly distinguish bike and e-bike segments.

Similar to the abovementioned machine learning methods, random forest was used by Lari, Z.A. *et al.* [39] to classify travel modes. In their research, several valuable attributes (e.g. speed, accuracy, delta bearing, delta speed, acceleration, and delta acceleration) that might affect the output were mentioned. In order to obtain the reliable and acceptable results, they strictly selected some parameters (e.g. the number of trees, and the number of attribute) to develop the random forest model. To assure the accuracy of the random forest model, they randomly regarded three tenths of the total sample as the test set and took seven tenths of the total sample as the train set. From their results, the accuracy of mode identification using their proposed approach is as high as 96.91%. And the two attributes, instant speed and accuracy of GPS track, were determined as the most influential attributes based on two significant indices: mean decrease accuracy and Gini index. Although, the accuracy of mode detection is very high, there are three main limitations of their research. First, the data error recognition procedure might be not considered comprehensively, because they just used the maximum speed value to clean the data set, but these conditions might not be considered by them, for example, the incomplete track points; the altitude of track points beyond the highest altitude of the area where a GPS travel survey was conducted. High accuracy data could possibly increase the accuracy of the random forest model. Second, they selected four different forests to examine, but they did not give a scientific basis for selection. Third, they used smartphones to collect GPS data which had a short battery life problem (compared with dedicated GPS devices). Aiming to this problem, they didn't give the right solution.

Stenneth, L. *et al.* [42] proposed a novel method to identify a user's travel mode according to the GPS data collected from his mobile device and external transportation network data. The novel features and the traditional features were selected to infer travel modes. The novel features consisted

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

of average bus location closeness, candidate bus location closeness, average rail line trajectory closeness, and bus stop closeness rate. And the traditional features were made up of average accuracy of GPS coordinates, average speed, average heading change, and average acceleration. They chose five classification models (Bayesian Net, Decision Tree, Random Forest, Naïve Bayesian, and Multilayer Perceptron) to detect modes and achieve a high accuracy level of 93.5%. The proposed approach, although can achieve high accuracy for detecting various travel modes, there are two major limitations of their research. First, the amount of training data is not enough. This limitation might influence the accuracy of mode detection. Second, this study doesn't contain the TI procedure which could also affect the results.

Zheng, Y. *et al.* [47] designed an automatic mode detection approach comprised of three aspects: a change point-based segmentation method, an inference model and a post-processing algorithm based on conditional probability. In the inference step, four inference models they selected including Decision Tree, Bayesian Net, Support Vector Machine and Conditional Random Field were leveraged in the experiments. Meanwhile, two criteria, the accuracy by length and the accuracy by Duration were chosen to evaluate the accuracy of transportation mode detection using the four mentioned inference model. From their results, change point based segmentation approach outperforms uniform duration based and uniform length based approaches. Furthermore, as compared to other inference models, Decision Tree achieves a higher degree of accuracy of travel mode detection over the change point based segmentation approach. However, raw data pre-processing procedure, such as recognizing and cleaning error data, is not mentioned, which would inference the accuracy of segment and further reduce the accuracy of identification.

Hybrid methods use two of the different methods together. Detailed description of hybrid methods for MD in each selected papers is in the next four paragraphs.

Nitsche, P. *et al.* [41] recruited 15 volunteers equipped with a smartphone with the developed logging application to collect data in the metropolitan area of Vienna, Austria over a period of 2 months. In order to cope with the problem that GPS device might be signal lost in shadowed areas (e.g. urban canyons and underground transportation systems), they collected accelerometer data and used these data to improve the precision of reconstruction trajectory. They utilized an ensemble of probabilistic classifiers combined with a Discrete Hidden Markov Model (DHMM) to detect eight travel modes. But the recognition accuracy of different travel modes varied considerably. For instance, the detection accuracy of train and subway are both 65%, while the detection accuracy of bicycle is 95%. Furthermore, they didn't provide more details for data error recognition and trip identification, and could not concluded that the method they proposed was better at detecting travel modes than other methods.

Zhang, L. *et al.* [43] present a novel multi-stage to deduce transport mode. In the first stage, they used four parameters (mean speed, maximum speed, and heading related changes) to identify three main travel modes (walk, bicycle, and motorized vehicles) on the basis of the identified segments. In the second stage, specific travel-mode of car, bus, tram, and train were classified based on Support Vector Machines (SVMs) from the motorized vehicles class. From their results, the classification of travel-mode they presented is qualitative and the accuracy of mode detection using the approach is as high as 93%. However, there are two limitations in their research. In the first place, they don't take incomplete GPS trajectory situation into account which could affect the accuracy of segment

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

identification, and further influence the accuracy of transportation mode detection. In the second place, the amount of training data and testing data is not enough. These issues could limit their approach to classify four specific travel modes and inference the results of the approach.

Reddy, S. *et al.* [45] presented a novel travel mode classification system consisting of a decision tree followed by a first-order discrete Hidden Markov Model. The dataset used to train and test the travel mode classification system was collected from an experiment by asking 16 volunteers to carry 6 phones placed on different positions to obtain 15 minutes of each of the five modes. From their results, the accuracy of mode detection is 93.6%. It should be noted that the dataset was just collected from an experiment, and the quality of data set was better. It is a pity that they didn't test the novel approach on a long-time travel survey in which the quality of data would be influenced easily because of the signal problems.

Reddy, S. *et al*. [46] put forward an innovative mode classification system employing a decision tree followed by a first-order discrete Hidden Markov Model. They conducted an experiment requiring six volunteers to obtain eight minutes of data for each mode. They used the novel system to deduct travel modes and achieve 98.8% accuracy overall. Nevertheless, the dataset collected from an experiment was not large enough to train and test the mode classification system.

doi:10.20944/preprints201607.0067.v1

Table 5. Summary of methods of travel mode detection utilized in selected papers.

Year	Lead Author	Methods/Steps	Input Variables	Travel Modes	Accuracy	Ground Truth
2010	Gonzalez, P. A.	Neural Networks	**For all GPS points case**: average speed, maximum speed, estimated horizontal accuracy uncertainty, percent Cell-ID fixes, standard deviation of distances between stop locations and average dwell time **For only critical points case**: average acceleration, maximum acceleration, average speed, maximum speed, ratio of the number of critical points over the total distance of the trip, ratio of the number of critical points over the total time of the trip, total distance, and average distance between critical points	Car, Bus, and Walk	88.6% (all GPS points) 91.23% (critical points-only dataset)	The travel modes were manually noted by research team.
2015	Yang, F.	Neural Network	Average speed, maximum speed, standard deviation of speed, and standard deviation of acceleration.	Walk, Bicycle, Bus, and Car	More than 86%	N/A
2015	Xiao, G.	Neural Networks and Particle Swarm Optimization(PSO-NNs)	Low-speed point rate, travel distance, average speed, average absolute acceleration, median speed, and 95% percentile speed	Walk, Bike, Bus, and Car	95.81%(training set) 94.44%(test set)	Prompted recall survey
2015	Xiao, G.	Bayesian Networks	travel distance, average speed, average absolute acceleration, 95% percentile speed, low speed rate, and average heading change	Walk, Bike, E-bike, Bus, and Car	94.74%(training set) 92.74%(test set)	Prompted recall survey

Peer-reviewed version available at *Information* **2016**, 7, 67; doi:10.3390/info7040067

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2016 doi:10.20944/preprints201607.0[illegible]/v1

2015	Lari, Z.A.	Random Forest	Speed, accuracy, delta bearing, delta speed, acceleration, and delta acceleration	Car, Bus, and Walking	Almost 96%	Users attach to each GPS file
2011	Stenneth, L.	Bayesian Net (BN), Decision Tree (DT), Random Forest (RF), Naïve Bayesian (NB), and Multilayer Perceptron (ML)	Average bus location closeness, candidate bus location closeness, average rail line trajectory closeness, bus stop closeness rate, average accuracy of GPS coordinates, average speed, average heading change, and average acceleration	Car, Bus, Aboveground Train, Walking, Bike, and Stationary	92.5%(BN), 92.2%(DT), 93.7%(RF), 91.6%(NB), 83.3%(ML)	Travel modes were labeled in sensor reports.
2008	Zheng, Y.	Decision Tree(DT), Bayesian Net(BN), Support Vector Machine (SVM) and Conditional Random Field (CRF)	Length, mean velocity, expectation of velocity, top three velocity and top three accelerations from each segment	Walk, Car, Bus and Bike	74%(DT), 70%(BN), 59%(SVM), 47%(CRF)	Prompted recall survey
2014	Nitsche, P.	An ensemble of probabilistic classifiers combined with a Discrete Hidden Markov Model (DHMM)	5^{th} , 50^{th} and 95^{th} percentile of speed, accelerations, decelerations, direction change, standard deviation of the high-frequency accelerometer magnitudes, and power Spectrum of the accelerometer signal for frequencies $i\omega/128$ Hz with $i = 1, \ldots, 64$ and the sampling frequency $\omega = 50$ Hz	Walk, Bicycle, Motorcycle, Car, Bus, Electric Tramway, Metro, Train, and Wait	Range from 65% (train, subway) to 95% (bicycle)	The current transport modes were annotated by the volunteers during travel.
2011	Zhang, L.	Two-stage approach, and Support Vector Machines (SVMs) used in second stage	**In the first stage**: mean speed, maximum speed, and heading related changes; **In the second stage**: mean and standard deviation of maximum speed, mean and standard deviation of average speed, mean and standard deviation of average	Walk, Bicycle, Car, Bus, Tram, and Train.	93%	User could pick and modify his travel modes correctly in the Tracer APP.

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

doi:10.20944/preprints201607.0[illegible]/v1

			acceleration, mean and standard deviation of travel time, mean and standard deviation of acceleration, and ratio of stop time in respect to travel time;			
2010	Reddy, S.	Decision Tree followed by a first-order discrete Hidden Markov Model (DT-DHMM)	GPS speed, accelerometer variance, accelerometer DFT components from 1-3 Hz calculated	Still, Walk, Run, Bike, and Motor	93%	Experiment (i.e. mode known)
2008	Reddy, S.	Decision Tree followed by a first-order discrete Hidden Markov Model (DT-DHMM)	Variance, energy, sum of FFT coefficients between 1-5 Hz from accelerometer and the speed	Still, Walk, Run, Bike, and Motor	98.8%	An external entity captured the ground truth labels

Peer-reviewed version available at Information 2016, 7, 67; doi:10.3390/info7040067

4. Quality of reviewed studies

The scores of the quality of eligible articles ranged from 2 to 12, as shown in Table 6.

Table 6. Distribution of quality characteristics across reviewed studies.

Criteria	Description	Score	N of studies	Percentage
Assessing data collecting methodological quality				
Study design	Travel survey	2	7	58.3%
	Experimental survey	1	5	41.7%
Adequate sample size selection	Included	1	0	0%
	Not included	0	12	100%
Survey duration	More than 2 weeks	2	6	50%
	Less than 2week	1	6	50%
Overcoming drawbacks of measures	Included(e.g., the short battery lives, the signal loss)	1	8	66.7%
	Not included	0	4	33.3%
Ground truth	Included (e.g., prompted recall survey)	1	11	91.7%
	Not included	0	1	8.3%
Assessing data-processing methodological quality				
Data-preprocess methods				
Data error recognition	Included	1	9	75%
	Not included	0	3	25%
Trip/segment identification methods				
Independent variables selection reliability testing	Testing	1	4	33.3%
	Not testing	0	8	66.7%
Parameter selection reliability testing	Testing	1	0	0%
	Nor testing	0	12	100%
Travel mode detection methods				
Independent variables selection reliability testing[1]	Testing	1	8	72.7%
	Not testing	0	3	27.3%
Comparison of experimental results[1]	Included	1	9	81.8%
	Not included	0	2	18.2%

Note: 1. the paper of [38] just descripts the approach of segment identification, so it doesn't contain the apart of travel mode detection.

We can see from Table 6, 7 studies (58.3%) applied the data collected from travel surveys. All studies didn't calculate the adequate sample size of travel survey. Most studies used ground truth to calculate the accuracy of travel mode detection or train classification models. Many studies (n=9, 75%) took appropriate measures to recognize the error data. 72.7% of studies tested independent variables selection reliability in travel mode detection methods. However, none of studies tested independent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 July 2016 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at Information 2016, 7, 67; doi:10.3390/info7040067

variables selection reliability in trip/segment identification procedure. And 9 studies (81.8%) compared results with other experimental results in order to highlight the superiority of the proposed method.

5. Limitations and strengths

The limitations in this review should be considered when explicating the current results. In the first place, the eligible articles must be published in English, which restrained relevant literature published in other languages from being selected. In the second place, the included studies focused only on the travel mode detection approaches, so the different approaches for trip purpose imputation could not be discussed. This study had two strengths. On the one hand, the included articles were rigorously screened based on the aforementioned well-defined inclusion/exclusion criteria in six databases. Second, the quality of included articles was evaluated in standardized way.

6. Discussion and conclusion

The aim of this systematic literature review was to summarize and critically appraise the travel mode detection methodologies on current literature. To our knowledge, this review may be the first one systematically searching eligible literatures and evaluating the existing methodologies of travel mode detection based on GPS data collected by smartphone.

Following strict evaluation process, this systematic review has provided a detailed discussion of methods of GPS data processing, such as GPS data pre-process procedure, TI/SI procedure, and travel mode detection based on GPS data collected by smartphone in the existing researches. This review has also carefully discussed both advantages and disadvantages for different methods used in included articles. Although new appropriate approaches or improved methods are utilized to obtain highly accurate results, there are several research gaps in the steps of GPS data processing.

The aforementioned drawbacks of smartphone-based GPS travel surveys that could influence the accuracy of data need to be taken into consideration when these surveys are conducted. It is a common sense that TI/SI procedure is undertaken before travel mode detection. Thus, the accuracy of TI/SI might highly affect the accuracy of travel mode detection. Moreover, signal noise and signal loss could reduce the accuracy of TI/SI, and further influence the accuracy of travel mode detection. Moreover, it is a necessary step to take the appropriate measures to avoid the negative impacts of drawbacks, such as reducing the accuracy of GPS raw data. With regard to travel mode detection, it can be found that it is ambiguous for deterministic approaches to infer similar modes, such as bike and e-bike, and bus and car. What is more, most of researchers only use segment or a single trip to deduce its mode. In addition, the participants recruited by researchers to take part in a travel survey have not been divided into different types based to their own social attributes, such as students, workers and so on, since different types of people have different travel features. For instance, the travel time, travel frequency and mode choice patterns of university students and workers might be not the same [48, 49].

At present, smartphone-based GPS travel survey, although, may still have some problems, it is widely recognized that the travel survey method can offer substantial advantages over traditional travel survey methods. Moreover, the travel survey method not only increases the accuracy of travel information but also provides a chance to explore the dynamics of multi-day travel patterns. With the

 doi:10.20944/preprints201607.0075.v1

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

development of methods, the accuracy of travel mode detection could be further improved, which could make the smartphone-based GPS travel survey better supplement traditional travel surveys.

Acknowledgments

This research has been supported by Six Peak of Talents Projects of Jiangsu Province (No. 2015-JY-025), by the Humanities and Social Sciences of Ministry of Education, China (No. 11YJA630152). The authors would like to express their appreciation to anyone who has provided suggestions and comments on this paper.

Author Contributions

Linlin Wu, Biao Yang and Peng Jing searched eligible papers from six databases together and read eligible papers carefully. Biao Yang wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Xiao, G.; Juan, Z.; Zhang, C. Travel mode detection based on GPS track data and Bayesian networks. *Computers Environment & Urban Systems* **2015**, 54, 14-22.
2. Yue, Y.; Lan, T.; Yeh, A.G.O.; Li, Q.Q. Zooming into individuals to understand the collective: A review of trajectory-based travel behavior studies. *Travel Behavior & Society* **2014**, 1, (2), 69-78.
3. Wolf, J. L. Using GPS data loggers to replace travel diaries in the collection of travel data, Dissertation, Georgia Institute of Technology, School of Civil and Environmental Engineering, 2000.
4. Shen, L.; Stopher, P.R. Review of GPS travel survey and GPS data-processing methods. *Transport Reviews* **2014**, 34, (3), 316-334.
5. Stopher, P.; Fitzgerald, C.; Zhang, J. Search for a global positioning system device to measure person travel. *Transportation Research Part C Emerging Technologies* **2008**, 16, (3), 350-369.
6. Zimowski, M.; Tourangeau, R.; Ghadialy, R.; Pedlow, S. NONRESPONSE IN HOUSEHOLD TRAVEL SURVEYS. *Travel Surveys* **1997**.
7. Gong, L.; Morikawa, T.; Yamamoto, T.; Sato, H. Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies ☆. *Procedia - Social and Behavioral Sciences* **2014**, 138, 557-565.
8. Stopher, P.; Clifford, E.; Zhang, J.; Fitzgerald, C. Deducing Mode and Purpose from GPS Data. *Institute of Transport & Logistics Studies Working Paper* **2008**, 06, 1-13.
9. Beijing Municipal Committee of Transportation. The fourth Beijing comprehensive transportation survey report (2012, July). Beijing Transport Research Centre, Beijing Municipal Committee of Transportation.
10. Bohte, W.; Maat, K. Deriving and validating trip purposes and travel modes for multi-day

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C Emerging Technologies* **2009**, 17, (3), 285-297.
11. Itsubo S.; Hato E. A Study of the Effectiveness of a Household Travel Survey Using GPS-equipped Cell Phones and a WEB Diary through a Comparative Study with a Paper based Travel Survey. In Proceedings of the Transportation Research Board 85th Annual Meeting, Washington, DC, USA, 22-26 January 2006.
12. Kelly, P.; Krenn, P.; Titze, S.; Stopher, P.; Foster, C. Quantifying the Difference Between Self-Reported and Global Positioning Systems-Measured Journey Durations: A Systematic Review. *Transport Reviews* **2013**, 33, (4), 443-459.
13. Kohla, B.; Meschik, M.; Zmud, J.; Lee-Gosselin, M. Comparing trip diaries with GPS tracking: An Austrian study. *Transport survey methods: Best practice for decision making* **2013**, 306-320.
14. Krygsman, S.; Nel, J. H. The use of global positioning devices in travel surveys – a developing country application. *Document Transformation Technologies* **2009**.
15. Marchal, P.; Roux, S.; Yuan, S.; Hubert, J.P.; Armoogum, J.; Madre, J.L.; Lee-Gosselin, M.E.H. A study of non-response in the GPS sub-sample of the French national travel survey 2007–08. In the 8th International Conference on Survey Methods in Transport, France, 25-31 May 2008.
16. Oliveira, M.; Vovsha, P.; Wolf, J.; Birotker, Y.; Givon, D.; Paasche, J. Global positioning system-assisted prompted recall household travel survey to support development of advanced travel model in Jerusalem, Israel. *Transportation Research Record: Journal of the Transportation Research Board* **2011**, (2246), 16-23.
17. Papinski, D.; Scott, D.M.; Doherty, S.T. Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. *Transportation research part F: traffic psychology and behavior* **2009**, 12, (4), 347-358.
18. Rasmussen, T.K.; Ingvardson, J.B.; Halldórsdóttir, K.; Nielsen, O.A. Using Wearable GPS Devices in Travel Surveys: A Case Study in the Greater Copenhagen Area. In Proceedings of the Annual Transport Conference at Aalborg University, Aalborg, Denmark, 26-27 August 2013.
19. Schnfelder, S.; Axhausen, K.W.; Antille, N.; Bierlaire, M. Exploring the potentials of automatically collected GPS data for travel behavior analysis --. *GI-Technologien für Verkehr und Logistik* **2002**, 66, (1), 455-456.
20. Schuessler, N.; Axhausen, K.W. Processing Raw Data from Global Positioning Systems without Additional Information. *Transportation Research Record Journal of the Transportation Research Board* **2009**, 2105, (2105), 28-36.
21. Stopher, P.R.; Moutou, C.J.; Liu, W., Sustainability of Voluntary Travel Behavior Change Initiatives: a 5-year Study. In Proceedings of Australasian Transport Research Forum, Brisbane, Australia, 2-4 October 2013.
22. Stopher, P.; Wargelin, L. Conducting a Household Travel Survey with GPS: Reports on a Pilot Study. In 12th World Conference on Transport Research, Lisbon, Portugal, 11-15 July 2010.
23. Vij, A.; Shankari, K. When is big data big enough? Implications of using GPS-based surveys for travel demand analysis. *Transportation Research Part C: Emerging Technologies* **2015**, 56, 446-462.

Peer-reviewed version available at *Information* **2016**, *7*, 67; doi:10.3390/info7040067

24. Wolf, J.; Hallmark, S.; Oliveira, M.; Guensler, R.; Sarasua, W. Accuracy issues with route choice data collection by using global positioning system. *Transp. Res. Rec. J. Transp. Res. Board* **1999**, *1660*, 66–74.
25. Zhou, J.; Golledge, R. Real-time tracking of activity scheduling/schedule execution within a unified data collection framework. *Transp. Res. Part A Policy Pract.* **2007**, *41*, 444–463.
26. Gong, H.; Chen, C.; Bialostozky, E.; Lawson, C.T. A GPS/GIS method for travel mode detection in new york city. *Comput. Environ. Urban Syst.* **2012**, *36*, 131–139.
27. Chen, C.; Gong, H.; Lawson, C.; Bialostozky, E. Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York city case study. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 830–840.
28. Patterson, D.J.; Liao, L.; Fox, D.; Kautz, H. Inferring high-level behavior from low-level sensors. In Proceedings of the Fifth International Conference on Ubiquitous Computing, Seattle, Washington, USA, 12-15 October 2003.
29. Feng, T.; Timmermans, H.J. Transportation mode recognition using GPS and accelerometer data. *Transportation Research Part C: Emerging Technologies* **2013**, 37, 118-130.
30. Rudloff, C.; Ray, M. Detecting Travel Modes and Profiling Commuter Habits Solely Based on GPS Data. In Proceedings of the Transportation Research Board 89th Annual Meeting, Washington, DC, USA, 10–14 January 2010.
31. Bolbol, A.; Cheng, T.; Tsapakis, I.; Haworth, J. Inferring hybrid transportation modes from sparse GPS data using a moving window svm classification. *Comput. Environ. Urban Syst.***2012**, *36*, 526–537.
32. Maruyama, T.; Mizokami, S.; Hato, E. A Smartphone-Based Travel Survey Trial Conducted in Kumamoto, Japan: An Examination of Voluntary Participants' Attributes. In Proceedings of the Transportation Research Board 93rd Annual Meeting, Washington, DC, USA, 12–16 January 2014.
33. Gilani H. Automatically determine route and mode of transport using a gps enabled phone. Graduate Theses and Dissertations, University of South Floride, State of Arizona, 2005.
34. Bierlaire, M.; Chen, J.; Newman, J. A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C Emerging Technologies* **2013**, 26, (1), 78-98.
35. Xiao, Y.; Low, D.; Bandara, T.; Pathak, P. In Transportation activity analysis using smartphones. In Proceedings of the Consumer Communications and Networking Conference (CCNC), Las Vegas, Nevada, USA, 14-17 January 2012.
36. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Group, T.P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *International Journal of Surgery* **2010**, 8, (5), 336-341.
37. Xiao, G.; Juan, Z.; Gao, J. Travel Mode Detection Based on Neural Networks and Particle Swarm Optimization. *Information* **2015**, 6, (3), 522-535.
38. Xiao, G.; Juan, Z.; Gao, J. Inferring Trip Ends from GPS Data Based on Smartphones in Shanghai. In Proceedings of the Transportation Research Board 94th Annual Meeting, Washington, DC, USA, 11–15 January 2015.
39. Lari, Z.A,; Golroo, A. Automated Transportation Mode Detection Using Smart Phone Applications via Machine Learning: Case Study Mega City of Tehran. In Proceedings of the

Transportation Research Board 94th Annual Meeting, Washington, DC, USA, 11–15 January 2015.

40. Yang, F.; Yao, Z.; Jin, P. J. GPS and Acceleration Data in Multimode Trip Data Recognition Based on Wavelet Transform Modulus Maximum Algorithm. *Transportation Research Record: Journal of the Transportation Research Board* **2015**, 2526, 90-98.
41. Nitsche, P.; Widhalm, P.; Breuss, S.; Brändle, N.; Maurer, P. Supporting large-scale travel surveys with smartphones——A practical approach. *Transportation Research Part C: Emerging Technologies* **2014**, 43, 212-221.
42. Stenneth, L.; Wolfson, O.; Yu, P.S.; Xu, B. In Transportation mode detection using mobile phones and GIS information. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, USA, 1-4 November 2011.
43. Zhang, L.; Dalyot, S.; Eggert, D.; Sester, M. Multi-stage approach to travel-mode segmentation and classification of GPS traces. In Proceedings of the ISPRS Guilin 2011 Workshop on International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Guilin, China, 20-21 October 2011.
44. Gonzalez, P.A.; Weinstein, J.S.; Barbeau, S.J.; Labrador, M.A.; Winters, P.L.; Georggi, N.L.; Perez, R. Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks. *IET Intell. Transp. Syst.* **2010**, *4*, 37–49.
45. Reddy, S.; Mun, M.; Burke, J.; Estrin, D.; Hansen, M.; Srivastava, M. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks* **2010**, 6, (2), 662-701.
46. Reddy, S.; Burke, J.; Estrin, D.; Hansen, M.; Srivastava, M. In Determining Transportation Mode on Mobile Phones. In Proceedings of the 12th IEEE International Symposium on Wearable Computers, Pittsburgh, PA, USA, 11-13 October 2008.
47. Zheng, Y.; Liu, L.; Wang, L.; Xie, X. Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web. In Proceedings of the International Conference on World Wide Web, Beijing, China, 21-25 April 2008.
48. Zhan, G.; Yan, X.; Zhu, S.; Wang, Y. Using hierarchical tree-based regression model to examine university student travel frequency and mode choice patterns in China. *Transport Policy* **2016**, 45, 55-65.
49. KHATTAK; Asad; Wang, X.; Son, S.; AGNELLO; Paul. Travel by University Students in Virginia: Is This Travel Different from Travel by the General Population? *Transportation Research Record* **2011**, 2255, (2255), 137-145.