Transport Reviews, 2016
doi:

Routledge
Taylor & Francis Group

Transportation Mode Detection – an In-Depth Review of Applicability and Reliability

ADRIAN C. PRELIPCEAN[*,**§], GYŐZŐ GIDÓFALVI[*], YUSAK O. SUSILO[**]

ABSTRACT *The wide adoption of location-enabled devices, together with the acceptance of services that leverage (personal) data as payment, allows scientists to push through some of the previous barriers imposed by data insufficiency, ethics and privacy skepticism. The research problems whose study require hard-to-obtain data (e.g., transportation mode detection, service contextualization, etc.) have now become more accessible to scientists because of the availability of data collecting outlets. One such problem is the detection of a user's transportation mode. Different fields have approached the problem of transportation mode detection with different aims: Location Based Services is a field that focuses on understanding the transportation mode in real-time, Transportation Science is a field that focuses on measuring the daily travel patterns of individuals or groups of individuals, and Human Geography is a field that focuses on enriching a trajectory by adding domain-specific semantics. While different fields providing solutions to the same problem could be viewed as a positive outcome, it is difficult to compare these solutions because the reported performance indicators depend on the type of approach and its aim (e.g., the real-time availability of Location Based Services requires the performance to be computed on each classified location). The contributions of this paper are three fold. First, the paper reviews the critical aspects desired by each research field when providing solutions to the transportation mode detection problem. Second, it proposes three dimensions that separate three branches of science based on their main interest. Finally, it identifies important gaps in research and future directions, i.e. proposing: widely accepted error measures meaningful for all disciplines, methods robust to new datasets and a benchmark dataset for performance validation.*

Keywords Transportation Mode Detection; Transportation Segmentation; Location Based Services; Transportation Science; Human Geography.

1 Introduction

The in-parallel development of distinct research branches concerned with solving the same problem, makes it difficult to compare employed methodologies because

[§] Corresponding author: acpr@kth.se

of the definition of success that is specific to each research branch. Although finding the common ground between any proposed solutions is not an easy task, understanding the intrinsic differences between the fundamental aims of each research branch has unassailable value. This information can guide researchers to merging different approaches into a widely accepted solution.

Such a problem that has been studied by different research branches is detecting the transportation mode of a user. The different approaches are provided by the following research areas: 1) *Location Based Services* (LBS), which aims at detecting how a user is travelling as close to real-time as possible in order to provide useful information to the user (e.g., alerting the user when she should start walking towards the bus station in order to be on time for work - Prelipcean, Schmid, & Shirabe, 2015 -, or informing the user about the real time location of relevant buses - Reddy et al., 2010), 2) *Transportation Science* (TSc), which aims at generating reliable statistics on what transportation means the user employs when she performs her daily activities and chores for generating activity-travel diaries (Axhausen, Schönfelder, Wolf, Oliveira, & Samaga, 2003; Stopher, FitzGerald, & Zhang, 2008; Prelipcean, Gidófalvi, & Susilo, 2014, 2015), and 3) *Human Geography* (HG), which aims at enriching a trajectory with domain-specific semantics (e.g., detecting when a boat is fishing - Rocha, Oliveira, Alvares, Bogorny, & Times, 2010). The scope of the aforementioned research areas is considerably broader than transportation mode detection, and, as such, this paper does not attempt to provide definitions for each field, but only documents the solutions regarding transportation mode detection that originate in the mentioned fields.

While it is clear that different (partial) solutions have been studied and proposed by each of these fields, it is critical to understand the scope of each field. In LBS, the urgency of the response takes precedence over the correct segmentation of a trajectory since the focus is on providing answers to the question "Given a schema of transportation modes, how is a user travelling *now*?", and if an algorithm answers this question correctly 9 out of 10 times, it can report a precision of 90%. It is important to note that the question is answered every time a new location is received by a positioning device (e.g., GPS receiver). In TSc, the correct segmentation of a trajectory takes precedence over the urgency of response (which, in most cases, is ignored) since the focus is on providing answers to the question "Given a schema of transportation modes, how were users travelling during a defined period?", which usually implies that a segmentation algorithm can holistically process the trajectory and does not need to take into account an incoming stream of locations. In HG, the focus is on segmenting a trajectory into parts that can be semantically enriched and it is common to first segment the trajectories into segments where the object is stationary or moving. Since the focus is on semantically enriching a trajectory, there is no annotated dataset used to validate the stationary / moving segmentation and the ulterior semantic enrichment.

Since most of the research on transportation mode detection done by each of the fields continues the previous research performed in its own field, there is room left for improvement by analyzing whether solutions from different fields can complement one another, i.e., studying "fused" solutions. The aim of this paper is to enable the development of "fused" solutions by identifying the research gaps that can be filled by future research by proposing: 1) widely accepted error measures that are meaningful for LBS, TSc and HG, 2) methods that are robust to new datasets and are not prone to overfitting, and 3) a benchmark dataset that is useful for validating the performance of proposed methods.

While previous literature review papers also analyze transportation mode detection (Shen & Stopher, 2014; Gong, Morikawa, Yamamoto, & Sato, 2014), the papers only focus on transportation mode detection for automatic travel diary generation, i.e., TSc's aim, and do not take into account the variety of solutions available from other disciplines.

The remainder of the paper is organized as follows. Section 2 presents and discusses the different disciplines and their specifics. Section 3 summarizes the main differences and similarities between the main approaches. Section 4 presents the gaps in research and future research directions. Finally, Section 5 concludes.

2 Mode detection - main body of research

This section presents the differences between the current paper and other review papers, describes how the main body of reviewed literature was chosen, and overviews the body of literature. For a multi-dimensional summary of the body of literature, we direct the reader towards Table 6 in the Appendix.

2.1 Differences between the current review and other reviews

Two recent review papers have included transportation mode detection but only as part of a broad overview on generating travel diaries from GPS data (Shen & Stopher, 2014; Gong et al., 2014). The automatic generation of travel diaries is studied within the context of TSc because of two main factors: the utility of data collected / derived from such diaries (Prelipcean, Gidófalvi, & Susilo, 2015; Drchal, Certicky, & Jakob, 2015) and the decreasing response rate of respondents to classical travel diary collection methods (Ogle, Guensler, & Elango, 2005). In comparison, the current paper does not investigate TSc solutions only, but rather analyzes the problem of transportation mode detection in multiple research fields.

Shen and Stopher (2014) offer a good overview of what methods are used in TSc, but only centralize the precisions of each method as declared by its authors and do not analyze how the precision was computed or its implications. The authors also include important details such as what was the ground truth and which attributes were used in each method. In comparison, the current paper also investigates the implications of the declared precisions and how the precision is computed. Furthermore, the current paper investigates the relationship between the solutions offered by fields that include TSc but are not limited to TSc.

Table 1: An overview of the generic question interpretation, together with the interpretation's implications, for each studied discipline

Generic question	*Given a schema of transportation modes,* ***how does the user travel****?*		
Fields	LBS	TSc	HG
Question Interpretation	*Given a schema of transportation modes, how is a user travelling* ***now****?*	*Given a schema of transportation modes, how* ***were*** *users travelling* ***during*** *a defined* ***period****?*	*How can a trajectory be* ***segmented*** *into parts that can be* ***enriched*** *with domain specific* ***semantics****?*
Answer promptness	Real-time	Post-collection	Post-collection
Types of considered entities	**Raw** entities, i.e., GPS **points**, **windows** of accelerometer readings	**Aggregated** entities, i.e., same mode **triplegs**	**Aggregated** entities, i.e., **segments** of movement and stationarity
User benefits	Direct (tailored for the user)	Indirect (for groups of users)	Indirect (general knowledge)

Gong et al. (2014) summarize methodologies used for trip and transportation segmentation by analyzing the methods previously used in TSc and the used data sets. The current paper compares different research fields and also discusses the used datasets in terms of availability / promptness of response, thus offering more insights into how the data are useful and whether the benefits provided by using the data outweigh the latency and storage issues that accompany the data use.

Previous literature reviews have a focused approach to mode detection within TSc. The current paper explores the benefits and drawbacks of the major disciplines that provides solutions for mode detection, and presents possible future directions that could combine the knowledge gained from each discipline.

2.2 An overview of chosen literature

The chosen papers are either well established representatives of their fields (e.g., Zheng, Chen, Li, Xie, & Ma, 2010 for LBS, Wolf, 2000 for Transportation Science, and Alvares et al., 2007 for Human Geography), study the latest research trends or contain similarities with papers from the other disciplines. However, it is important to note that the grouping of papers in the three disciplines does not imply that the authors' main area of research is that of the proposed discipline, or that the papers were published in a journal that only caters for the proposed discipline, but rather that the applications of the paper fit the interpretation of the transportation mode detection problem of the proposed discipline. An overview of the studied

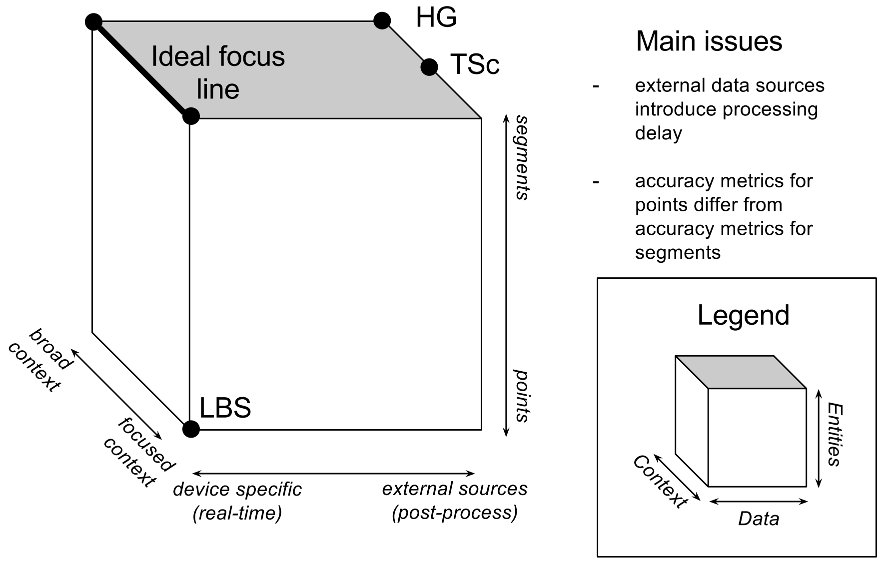

Figure 1: The differences between the approaches employed for transportation mode detection by the three studied disciplines. LBS focuses on understanding the transportation mode in real-time, TSc focuses on measuring the daily travel patterns of individuals or groups of individuals, and HG focuses on enriching a trajectory by adding domain-specific semantics.The dimensions along which the different disciplines differ substantially are visualized using a cube-display inspired by MacEachren, 1995.

disciplines, together with the main differences between them that constitute the discussions in the following sections, can be seen in Table 1 and in Figure 1.

The body of literature related to *LBS* contains those studies that focus on detecting the transportation mode of a user as close to real time as possible by relying mostly on data provided by GPS receivers (Stenneth, Wolfson, Yu, & Xu, 2011; Biljecki, Hugo, & van Oosterom, 2013), the accelerometer sensor (Hemminki, Nurmi, & Tarkoma, 2013; Yu, Lin, Yu, Chang, & Wang, 2014) or a combination of the two (Reddy et al., 2010; Manzoni, Maniloff, Kloeckl, & Ratti, 2010; Shah, Wan, Lu, & Nachman, 2014; Prelipcean et al., 2014).

The body of literature related to *TSc* contains studies that focus on automating the generation of travel diaries, which includes the task of segmenting a trajectory into same mode segments, i.e., triplegs (Axhausen et al., 2003; Wolf, 2000; Stopher et al., 2008). These studies employ similar three step approaches that segment trajectories into trips, split trips into triplegs, and classify triplegs.

The body of literature related to *HG* contains studies that deal with the broad task of enriching a trajectory with domain-specific semantics, which first identifies when an user is moving or stationary (Alvares et al., 2007; Palma, Bogorny, Kuijpers, & Alvares, 2008; Rocha et al., 2010). This task coincides with the TSc trip

Table 2: Overview of the main three LBS approaches. The main classification methods are Decision Trees (DT), Random Forests (RF), Support Vector Machine (SVM) and Discrete Hidden Markov Model (DHMM). The highest number of modes is obtained by the approaches that fuse GPS with accelerometer data. The references for the maximum number of considered modes, as well as the maximum precision correspond to the ones highlighted in Appendix A.

Sensors used	Acc.	GPS	GPS + Acc.
Accuracy entity	Sliding Window	GPS point	GPS point
Classif. method	DT, SVM	RF	DT, RF, DHMM
Max # modes	7 modes	6 modes	9 modes
Max precision	92.5% (5 modes)	92.8% (6 modes)	94.4% (9 modes)
Auxiliary data	No	Yes	No
Task	Given a schema of transportation modes, how is a user travelling now?		

segmentation task, which precedes the tripleg detection task.

2.3 Location Based Services

The predominant transportation mode detection version for LBS is "on-demand" mode detection (also known as online processing), where an algorithm has to provide the *current* transportation mode of the user whenever asked. Different "on-demand" transportation mode detection methods are, as mentioned previously, focused on obtaining solutions as close to real-time as possible, with a low footprint on battery consumption.

As there are different types of sensors that can provide data suitable for transportation mode detection, researchers have studied the following sensors: GPS receiver (Stenneth et al., 2011), accelerometer reader (Hemminki et al., 2013; Yu et al., 2014) and the two sensors combined (Reddy et al., 2010; Manzoni et al., 2010; Shah et al., 2014; Prelipcean et al., 2014). The main LBS specific approaches are summarized in Table 2.

2.3.1 Accelerometer-only studies

The studies that use only accelerometer data are oriented towards battery savings and compute features such as variance, wavelet entropy, peak frequency on sliding or tumbling windows that are labeled by the users. The used measures and dimensions are similar to those studied and presented by Figo, Diniz, Ferreira, & Cardoso, 2010. Notably, this type of low-battery consumption methods are also provided by the major phone operating systems via different Application Programing Interfaces (APIs), such as Google, 2015, whose `ActivityRecognition` API can detect whether the device is on a user that is travelling by car, walking, running, bicycling or stationary, and Apple, 2015, whose `CMMotionActivityManager` API can detect if the user is walking, running, in a vehicle, or stationary. Unfortunately, there are no publicly available documents regarding the precision of the aforementioned APIs. Research that focused explicitly on accelerometer only transportation mode detection report a precision of 70% for six transportation modes using decision tree classifiers (Wang, Chen, & Ma, 2010), 80.1% precision and 82.1% recall for seven transportation modes using AdaBoost together with Decision Tree classifiers (Hemminki et al., 2013) and 92.5% precision for five transportation modes using a Support Vector Machine classifier (Yu et al., 2014). While the data size available for each research group varies (150 hours from 16 users for Hemminki et al., 2013, 12 hours from 7 users for Wang et al., 2010 and an afternoon from 4 users for Yu et al., 2014), the precision and recall values are computed on accelerometer samples, which are labeled periods of a predefined duration where one accelerometer reading can be found either in one period only (tumbling windows) or in multiple adjacent periods (sliding windows) – (8 seconds for Wang et al., 2010, 1.2 seconds for Hemminki et al., 2013, and 10 seconds for Yu et al., 2014) associated to one activity. The validation of the proposed methods is done by allocating a percentage of the collected data set for testing purposes and training on the remaining set (Wang et al., 2010; Yu et al., 2014), or by using a leave one user out cross-validation (Hemminki et al., 2013).

The advantage of using an accelerometer only approach to transportation mode detection lies in the promptness of response and battery efficiency of the method. However, the main disadvantage is that it can not be reused as-it-is for other purposes where the spatial position of the users is needed. While, if used continuously during a time frame, it offers a good description of the temporal component of the user's activities, to offer the same level of description detail for the spatial component, it should make use of positioning technologies.

2.3.2 GPS-only studies

There are relatively few studies that try to estimate the transportation mode relying solely on GPS-derived features in real time, mostly because these features are usually not enough to distinguish between a great number of classes. To avoid this pitfall, research augments the mode detection with proximity to objects in relevant geographical datasets (e.g., transportation stations or transportation network), which is problematic from the point of view of scalability and usability, since the datasets are stored on a server. In general, these methods are inefficient in terms of battery consumption since there is no duty-cycling performed to leverage how much battery is consumed by the GPS receiver due to the lack of accelerometer use. This is detrimental because the battery consumption of a unsuccessful location request inside a building (where the user / phone is stationary, i.e., a large fraction of the time - Klepeis et al. 1996) is significantly higher than the cost of a location request outdoors (Prelipcean et al., 2014).

Stenneth et al., 2011 use data provided by the GPS receiver to distinguish between six transportation modes (walk, bus, driving, train, stationarity, bike). The GPS locations are sampled every 15 seconds, but the window used in the classification is formed of 30 seconds, which is needed to compute the used feature set. The authors use a transportation network GIS dataset together with the expected real time locations of buses (according to simulations based on the travel history of buses) obtained as a live feed (it is not specified whether the feed is a Generalized Transit Feed). The features used in the classification were: (1) average accuracy of GPS readings, (2) average speed, (3) average heading change, (4) average acceleration, (5) average bus closeness combined with candidate bus closeness, (6) rail line trajectory closeness, and (7) bus stop closeness rate. Since all these are not lightweight operations on average sized datasets, the processing cannot be performed by clients and the GPS points have to be sent to a central server, which in turns answers with a probable mode (when available). The authors find that a Random Forest (an ensemble learning method that aggregates the classification of multiple Decision Trees based on a voting scheme, where each tree gives a classification for a data point based on the set of data features revealed to the tree - Breiman, 2001) has the best performance metrics with a 92.8% precision and 92.9% recall. While the necessity of a central server that performs these operations is crippling in terms of availability for a large number of requests and the responses are not immediate due to the transportation feed (the bus positions are updated every 20-30 seconds), the authors obtain a good precision and recall even though the considered modes exhibit similar characteristics (e.g., speed for buses and cars).

Another study that relies exclusively on GPS data and GIS datasets is presented in Das, Ronald, & Winter, 2014, where the authors propose a method to detect transfer points. The methodology relies on clustering nearby points using a Density-Based Spatial Clustering of Application with Noise (DBSCAN) with an adaptive search radius and then classifying the clusters into transfer points based on the proximity to existing POIs (bus, train or tram stops) and the dwell time of the cluster (i.e., the time period between the earliest clustered point and the latest clustered point). The experiments were run in Melbourne on 29 valid trajectories collected by 4 users, where the trajectories had the following constraints: a minimum distance of 5 km and a minimum duration of 30 minutes. The authors compare the proposed approach with the traditional walking-based one in two experiments and report that the proposed one outperforms the traditional one, but given the dataset size it is not a definite result since more information is needed for the validation. Similarly to the previous GIS dataset dependent approaches, the applicability of this approach is dictated by the availability of the dataset. Also, the results of the test are inconclusive due to the modest dataset used for testing.

In summary, the studies that rely only on GPS-provided sensor data make use of the not-widely available GIS datasets and adds an extra layer of complexity on the server side. The use of GIS datasets is a downside from the usability aspect due to the well-known non-standard way of storing different attributes throughout different dataset sources (e.g., using transportation lines from OSM and a proprietary source), where overlapping entries cannot be easily filtered out.

2.3.3 GPS fused with accelerometer studies

In general, fusing GPS readings together with accelerometer derived features proved to be of great help in terms of precision gain (Reddy et al., 2010). The main advantages are the duty cycling of the GPS sensor based on the accelerometer readings to minimize battery consumption, and the availability of these sensors' readings directly on the client, which contrasts the GIS-dataset dependent approaches, where several extra steps have to be taken for a prediction (Prelipcean et al., 2014).

These methods collect all the accelerometer readings in between every two consecutive GPS locations (except Shah et al., 2014) and then compute features similar to those proposed by Figo et al., 2010, such as: time domain features (mean, standard deviation, mean crossing rate) and frequency domain features (dominant frequency, sub-band energy, discrete Fourier transform energy coefficients), which are then used as features for the applied machine learning methods (Reddy et al., 2010; Prelipcean et al., 2014). The features that are derived from the GPS readings are usually average heading change, average speed, average accuracy of a location, etc. These studies differentiate either between five different travel modes (still, walk, run, bike, and motor) with a 91.3% precision and a 91.3% recall (Reddy et al., 2010), between seven different travel modes (car, train, walk, subway, bus, bike, and ferry) with a 90.8% precision and 90.9% recall (Prelipcean et al., 2014), or between three different motorized travel modes (car, bus, and train) with a precision of 90% (Shah et al., 2014). The sampling method used for machine learning training is usually the 10-fold cross validation method (Reddy et al., 2010; Prelipcean et al., 2014; Manzoni et al., 2010), and Reddy et al., 2010 also discusses the implications of using a leave-one-user out cross validation technique to show that the trained classifiers can be used for making inferences for new users. The authors use different types of classifiers such as a Decision Tree in the initial stage followed by a post processing Discrete Hidden Markov Model based on a transition probability matrix (Reddy et al., 2010), a Random Forest classifier (Prelipcean et al., 2014), a Decision Tree (Manzoni et al., 2010), or a Markov Model Smoother followed by a Decision Tree (Shah et al., 2014). The sampling rates for the collected data vary: 1 location and 32 accelerometer readings per second (Reddy et al., 2010), 1 location every 50 meters and 5 accelerometer readings per second (Prelipcean et al., 2014), 1 location and 25 accelerometer readings per second (Manzoni et al., 2010), and 1 location every 5 seconds and 100 accelerometer readings per second (Shah et al., 2014). The authors do not explain the reasons behind the chosen collection sampling rates - the reader is directed towards Prelipcean et al., 2014 for a deeper discussion on different data collection sampling strategies. The size of the datasets and the number of users that collected data also vary: 16 users collecting data for 15 minutes per transportation mode

Table 3: Overview of the three main TSc approaches. The main employed classification methods are Subjective Rule Determination, Decision Trees (DT) and Random Forests (RF). The highest number of modes is obtained by the approaches that rely on fuzzy logic. The references for the maximum number of considered modes, as well as the maximum precision correspond to the ones highlighted in Appendix A.

Classif. type	Fuzzy Logic	Rule-based Heuristics	ML
Classif. methods	Subjective Rule Determination	Subjective Rule Determination	DT, RF
Max. # modes	10 modes	6 modes	6 modes
Max. precision	91.6% (10 modes)	95% (5 modes)	85.6% (6 modes)
Auxiliary data	Yes	Yes	Yes
Task	Given a schema of transportation modes, how were users travelling during a defined period?		

(Reddy et al., 2010), 11 users collecting data for 10 days (Prelipcean et al., 2014), or 15 users collecting 129 trips (Shah et al., 2014). With the exception of Shah et al., 2014, who use a generalized transit feed together with a street network for inferring the mode, all the methods rely solely on data readily available on the device. While the intention of the aforementioned approaches is real time travel mode detection, the algorithms are trained and tested after the data are collected and labeled, and are not re-tested on new users in real time, and therefore the generality of the obtained models is not investigated.

The LBS research reports the algorithms' performance via point-based precision and recall (i.e., per measured entity), which is sufficient if no further information needs to be derived from the measurements, such as the time spent by a user while traveling via a certain mode, the traveled distance via a mode, or the number of mode changes per trip. These methods put a higher emphasis on providing prompt and correct inferences for each location than on accurate aggregate statistics on same-mode travel segments.

2.4 Transportation Science

TSc has continuously pursued the study of how people travel and what impact do modifications on the infrastructure have on people. One of the most common methods used to gather this type of information is via travel diaries, where users declare how and why they are traveling to their destinations during a certain time period (usually of one day). The classical approaches relied on users filling up paper diaries, having phone interviews, or declaring how they traveled via web forms, which have two main downsides: the participants are under-reporting their trips (Bricka & Bhat, 2006; Wolf, Oliveira, & Thompson, 2003), and the response rate is decreasing (Ogle et al., 2005; Zimowski, Tourangeau, Ghadialy, & Pedlow, 1997). As a solution to these problems, transportation scientists tried to automate parts of these diaries.

Historically, among the first authors to carry out the automation of activity-travel diaries were Wolf, 2000, Schönfelder, Axhausen, Antille, & Bierlaire, 2002 and Axhausen et al., 2003, which collected data using GPS receivers mounted on vehicles. One of the first to move away from in-vehicle GPS devices to hand-held GPS receivers (accompanied by a PDA, all of which weighted approximately 2 kg) was Draijer, Kalfs, & Perdok, 2000 who studied the burden and bias imposed by these type of devices. These papers are mostly mentioned for historic reasons, since they constitute the launching points of the science that eventually ended up studying transportation mode detection as part of its efforts, which is of interest for the present paper.

One of the studies that proposes the methodology most widely used in TSc for segmenting trips into triplegs is Chung & Shalaby, 2005, which (besides noise filtering) contains two main pre-processing tasks prior to mode detection: trip detection, which is based on a heuristic rule that uses a threshold of 120 seconds dwell time between consecutive locations to detect trips, and tripleg identification, which is based on Mode Changing Points (MCPs). Unique to this study is the initial definition of MCPs (later branded by Tsui & Shalaby, 2006 as Mode Transfer Points – MTPs) and the set of rules used to define these points. A point is a MTP if: 1) the speed difference from the previous point's speed is more than 10 km/h and the time difference from the previous point is more than 5 seconds, or 2) the distance to its previous point is more than 150 meters and the speed difference from the previous point's speed is more than 10 km/h. As a continuation of this study, Tsui & Shalaby, 2006 formally define a Mode Transfer Point (MTP) as the point where travelers change from one transportation mode to another, and declare three types of MTPs: a start-of-walk point (SOW - the traveler changes from a mode to walking), an end-of-walk point (EOW - the traveler changes from walking to a new mode) and an end of gap point (EOG - the first point that occurs after a large time gap without GPS data). The detection of MTPs has become an integral part of most research performed in TSc as the most common pre-processing step of the transportation mode inference (Stopher et al., 2008; Schüssler & Axhausen, 2009; Schüssler, Montini, & Dobler, 2011). Another widely used approach for segmentation is the one proposed by Zheng et al., 2010, which relies on the same assumption as Tsui & Shalaby, 2006, namely that a traveler walks when changing transportation modes, and detects walking segments by using a loose upper bound for velocity and acceleration (2.5 m/s for velocity, and 1.5 m/s^2 for acceleration).

When analyzing the travel modes, there is usually a baseline for detected modes such as walking, bicycling and motorized (Chung & Shalaby, 2005; Tsui & Shalaby, 2006; Stopher et al., 2008; Schüssler & Axhausen, 2009; Schüssler et al., 2011; Montini, Rieser-Schüssler, & Axhausen, 2014; Bohte & Maat, 2009; Biljecki et al., 2013; Zheng et al., 2010; Rasmussen, Ingvardson, Halldórsdóttir, & Nielsen, 2013, 2015; Prelipcean et al., 2014; Prelipcean, Gidófalvi, & Susilo, 2016), but some research goes a step further and classifies the motorized into car, train, and public transportation (Stopher et al., 2008; Schüssler & Axhausen, 2009; Schüssler et al., 2011; Bohte & Maat, 2009; Biljecki et al., 2013; Rasmussen et al., 2015; Prelipcean et al., 2014, 2016). Of these papers, Biljecki et al., 2013 provides the most detailed classification scheme that includes 10 modes: walking, bicycling, car, ferry boat, sail boat, train, subway, bus, tram and flight.

In the body of research on mode detection in TSc, there seems to be a preference for inferring travel mode by using mostly subjective methods such as Fuzzy Logic (Tsui & Shalaby, 2006; Schüssler & Axhausen, 2009; Schüssler et al., 2011; Rasmussen et al., 2015) or any variation of fuzzy logic approaches such as methods based on Membership Functions (Biljecki et al., 2013), and Rule Based Heuristics (Chung & Shalaby, 2005; Bohte & Maat, 2009). Some methods also rely on objective methods such as Decision Trees (Zheng et al., 2010) or Random Forests (Montini et al., 2014), but these approaches constitute the minority. The most widely used features in mode detection are derived from segments, e.g., duration of tripleg, average speed of tripleg, median speed, 95th percentile speed, or proximity to road network segments or transportation stations. Most research on mode detection in TSc relies on external GIS datasets. The reported precisions vary: 91.7% precision for 4 travel modes (Chung & Shalaby, 2005), 94% precision for 7 travel modes (Tsui & Shalaby, 2006), 95% precision for 5 modes (Stopher et al., 2008), 83% precision for 5 modes (Schüssler et al., 2011), 85.8% precision for 7 modes, 70% precision for 6 modes (Bohte & Maat, 2009), 91.6% precision for 10 modes (Biljecki et al., 2013), 84.6% precision for 5 modes (Rasmussen et al., 2015), and 75.3% precision and 73.2% recall for 9 modes (Prelipcean et al., 2016).

Since scientists focused on different methods for automatic trajectory segmentation and mode detection, they reported the performance of the proposed algorithms using traditional metrics that do not cover all dimensions of error that are associated with dealing with continuous intervals and with the error propagation due to multi-step approaches (Prelipcean et al., 2016).

2.5 Human Geography

As Parent et al. (2013) noted, trajectory segmentation depends on the application of the services for which it is performed. While the output is a trajectory split into domain-specific semantic segments, one of the first steps in this process is splitting a trajectory into segments when the object is stationary, i.e., stops, and segments where the object is moving, i.e., moves. This type of trajectory segmentation is known as stops and moves. The main HG approaches are presented in Table 4.

The "stops and moves" segmentation methodology was initially proposed by Alvares et al. (2007), where the authors use a heuristic approach and check if a trajectory intersects a polygon associated with a POI (where the polygon is

Table 4: Overview of the two main HG approaches. These approaches rely on different heuristics and the precision is based on human interpretation.

Approach		Heuristics	Interactive Exploration
Thresholds	Dwell time	Yes	Yes
	Speed	Yes	Not specified
	POI proximity	Yes	Yes
	Direction change	Yes	Not specified
Task	How can a trajectory be segmented into parts that can be enriched with domain specific semantics?		

built as a buffer around a POI) for a minimal threshold duration. Whenever that happens, the subtrajectory for the duration of the intersection is labeled as a stop and the subtrajectory leading to it and the one following it are labeled as moves. The authors only show examples on how the algorithm is useful, but the algorithm is not tested on a real-world data set. New methods that rely on the "stops and moves" approach have been developed and tested (Palma et al., 2008; Rocha et al., 2010), but the methods either report only on identified stops and not on missed stops or falsely identified stops (Palma et al., 2008), or are tested on moving objects datasets that have a distinguishable aspect, which is known apriori (e.g., the fishing boat dataset studied by Rocha et al. (2010), where the authors rely on a minimal direction change threshold to detect stops, which is difficult to use on other types of moving objects, such as land vehicles). Krumm and Horvitz (2006) use a similar approach for segmenting trajectories collected by different vehicles into trips. They relied on a temporal threshold accompanied by a speed threshold, and identified 7355 trips; however, no ground truth data was provided for any validation, and the approach is tailored to work for GPS data collected by cars. Even though the "stops and moves" is rationally sound, the lack of reports on the performance of this approach makes it difficult to compare with other approaches that segment trajectories. Furthermore, the approach only differentiates between stationary and non-stationary periods, which is insufficient for more complex tasks (e.g., segmenting a trajectory into same transportation-mode segments). Finally, the approaches that rely on an existing POI dataset are restricted to places where such datasets are available.

Other approaches involve proposing exploratory analysis frameworks (Andrienko, Andrienko, & Wrobel, 2007; Yan, Chakraborty, Parent, Spaccapietra, & Aberer, 2011), which can then be used to identify temporal and / or spatial threshold values for segmenting trajectories into stops and moves. However, this approach can be thought of complementary to the "stop and moves" one, thus inheriting its shortcomings. Furthermore, due to its exploratory aspect, the approach relies on human interaction and assumes that the analyst can find optimal parameter values given her implicitly assumed correct notion of ground truth, which diminishes the ability to automate the task.

Xie, Deng, and Zhou (2009) define a trajectory semantic join and introduce two measures to identify parts of a trajectory semantically enriched by a set of activities associated to a POI, namely influence and influence direction. The first measure is used to generate segments as sequences of points that share the closest POI, and the second measure is used to choose the most probable activity from the activity set associated to its closest POI by using a POI activity mapping set. The POI activity mapping set contains the minimum and maximum elapsed time for an activity that happens at a POI. The authors simulate trajectories that contain five modes (driving, walking, short stop, stop and long stop) and perform the semantic join, but the purpose of the simulation is to measure the algorithm's efficiency rather than assessing the correctness of the semantic join.

The HG type of segmentation is useful for identifying stationary and movement parts of a trajectory, and could be used for detecting transportation modes if one of the following hypotheses holds: 1) any two consecutive movement periods with distinct modes are separated by a stop period, in which case the task is reduced to classifying a movement period into its travel mode, or 2) a movement period can be semantically enriched by using apriori knowledge specific to transportation detection derived from proximity to relevant geographic datasets. It is important to note that, in the absence of ground truth, the classification depends on the subjective interpretation of the expert that finely tunes the parameters of the classification to her expectation. Due to the similarity between the strategies used by TSc and HG for segmenting a trajectory, this paper considers any HG study that specifically focuses on travel mode detection as a TSc study. However, contrasting LBS and TSc, where the data are collected are labeled according to a predefined scheme, the HG approach only depends on the available auxiliary datasets and the knowledge of the interpretation expert. Finally, the HG approach is limited by the expert's prowess, as well as by other factors that can affect her decision making process, such as: fatigue or incomprehensibility induced by massive data volumes.

3 The main differences between disciplines

Since each of the three disciplines have an own interpretation of the generic transportation mode detection question (see Table 1), it is important to understand the main dimensions along which they differ the most. To facilitate the understanding of what follows, the discussion is accompanied by Figure 1, which is inspired by the "cartographic visualization cube" proposed by MacEachren, 1995.

3.1 Data used in the mode detection classification

One of the dimensions among which the disciplines vary considerably is the "data types and data processing" (the length of the cube in Figure 1) dimension. One end of the dimension contains *only* data that facilitates close to real-time transportation mode detection, which is directly linked to LBS. The opposite end of the dimension contains external data that are used in post-processing stages for transportation mode detection (e.g., road networks, POIs, etc.), which is directly linked to TSc and HG.

Since the priority of the LBS discipline is to respond to the requests in real-time (or as close to real-time as possible), most approaches rely on the data that are readily available on the devices, i.e., supplied by sensors, e.g. GPS (Stenneth et al., 2011), accelerometer (Hemminki et al., 2013; Yu et al., 2014) or both (Prelipcean et al., 2014; Shah et al., 2014). While some approaches have an extra penalty for the promptness of the classification because they rely on transit feeds, which are managed by 3rd parties, (Stenneth et al., 2011; Shah et al., 2014), most LBS approaches classify either every GPS location into its transportation mode (Prelipcean et al., 2014; Reddy et al., 2010), the penalty being the sampling frequency, or every window of accelerometer readings (Hemminki et al., 2013; Yu et al., 2014), the penalty being the size of the window.

Contrasting LBS, neither TSc or HG attempt to classify each sensor reading, but perform a bulk classification on a whole trajectory. Since there are no promptness constraints, TSc and HG approaches rely on external data sources such as road network datasets (Chung & Shalaby, 2005; Tsui & Shalaby, 2006; Stopher et al., 2008; Biljecki et al., 2013; Rasmussen et al., 2013, 2015) or POI buildings datasets (Alvares et al., 2007; Palma et al., 2008).

A direct consequence of the difference in focus between disciplines is the types of devices that are used to collect data and / or perform the mode detection. While there is a prevalence for using smartphones in LBS (Manzoni et al., 2010; Reddy et al., 2010; Wang et al., 2010; Hemminki et al., 2013; Montini et al., 2014; Prelipcean et al., 2014; Shah et al., 2014; Yu et al., 2014), TSc and HG mostly use dedicated devices such as the Geostats Geologger (Chung & Shalaby, 2005; Tsui & Shalaby, 2006) or other types of dedicated devices (Alvares et al., 2007; Palma et al., 2008; Stopher et al., 2008; Bohte & Maat, 2009; Schönfelder et al., 2002; Rocha et al., 2010; Zheng et al., 2010; Biljecki et al., 2013; Rasmussen et al., 2013, 2015) with a few exceptions that use smartphones (Schüssler et al., 2011; Montini et al., 2014). This is an important distinction because, with a few exceptions (Stenneth et al., 2011; Shah et al., 2014), the studies that collect data using smartphones have a *thick-client architecture*, in which the client performs most inference operations and the server acts only as a storage backup, which contrasts those that use dedicated devices, which use a *thin-client architecture*, in which the server performs inference operations in addition to the data management. These discussions are summarized in the first row of Table 5.

3.2 Types of entities that are studied by each discipline

Another dimension among which these fields vary is the measured entity, i.e., the type of entity used for assessing the performance of algorithms (height of the

Table 5: The differences induced by the three proposed dimensions. Most LBS studies perform the transportation classification on smartphones (SP) in real time, while TSc and HG collect data with dedicated devices (DD) and perform the transportation classification on a central server.

Dimension	Dimension implications	LBS	TSc	HG
Data	Answer	Real-time	Post	Post
	Device	SP	DD	DD
	Client	Thick	Thin	Thin
Entity	Type	Location	Tripleg	Segment
	P&R calculation	Direct	Multi-step	Subjective
	P&R disadvantages	Non holistic	Triplegs given	Manual
Context	Classif.	Robust	Subjective	Subjective
	User benefit	Short-term	Long-term	Long-term
	Classif. validation	Yes	Partial	No

cube in Figure 1). LBS measures precision and recall for every labeled raw entity, which is a GPS location (Stenneth et al., 2011), a window of accelerometer readings (Hemminki et al., 2013; Yu et al., 2014), or a GPS location fused with accelerometer derived features and summaries (Prelipcean et al., 2014). HG allows for subjective interpretation of the semantics attached to a trajectory, which is usually independent from ground truth data. TSc measures precision and recall (the recall is seldom disseminated in TSc) on segments that are obtained via different heuristic rules such as MTP (Tsui & Shalaby, 2006) or walk-segment detection (Zheng et al., 2010). As a consequence, the entities used for measuring the performance of each method vary between disciplines, the performance measures themselves are incompatible (e.g., a 90% precision in LBS does not translate to a 90% precision in TSc).

In LBS, the errors are easy to compute and understand, i.e., given the entities (GPS locations or accelerometer windows) that need to be classified, the performance measures are precision (percentage of the inferences that were correct, per mode) and recall (percentage of the total population that was inferred, per mode). This resonates with the promptness penalty that is present in LBS, where the inference is made in real-time. However, since the precision and recall are computed per entity, it does not directly capture information at a segment level such as over- or under-segmentation or segment matching, which are critical in TSc. The variation of the precision and recall values reported by the authors in LBS is more likely to be due to the classification scheme (i.e., tax complexity), the data features and the size of the learning corpus rather than due to the type of machine learning used (Banko & Brill, 2001 showed the size of the learning corpus of a machine learning algorithm to be more important than the type of machine learning algorithm).

In TSc, the errors are difficult to understand due to the vagueness induced by the multi-step approach that generates segments: segment trajectories into trips, segment trips into triplegs, and infer the travel mode of a segment. One of the few papers that provides a thorough definition of precision, recall and how the values are computed is written by Zheng et al., 2010. The authors define the following measures for the performance transportation mode detection: accuracy by segment computed as the number of segments correctly predicted divided by the total number of segments, and accuracy by distance computed as the distance covered by the correctly inferred segments divided by the total traveled distance. Similarly, the authors define precision and recall values for MTP, and emphasize that even though the recall has higher priority in this case, a balance between precision and recall should be kept. Even though this approach is thorough in terms of accuracy assessment, its output is discontinuous, the statistics are presented only for the inferred segments matched to ground truth segments, and all errors are treated as equal. However, the metrics proposed by Zheng et al., 2010 are not well suited for travel diaries automation since they have been developed for Geolife (a Location-Based Social Networking Service - Zheng, Xie, & Ma, 2009) while testing the performance of different transportation mode detection solutions (Zheng, Liu, Wang, & Xie, 2008). Most TSc research papers report on precision for the given ground truth triplegs, as opposed to the inferred segments that are obtained via detecting MTPs, which suggests that they are only valid for a 100% accurate segmentation. For example, Schüssler et al., 2011 report a tripleg segmentation precision of 68%, and then continue on with performing and reporting on mode detection on the 100% accurate segments, as opposed to the 68% accurate ones. Similarly, Biljecki et al., 2013 report that "the segments are not segmented exactly at the same transitions points (so one should accept small differences here)."

The issue of error propagation in multi-step approaches, and how the precision and recall values reported in LBS translate to triplegs are discussed at length in Prelipcean et al., 2016, where a generic framework for measuring the transport mode segmentation of trajectories is introduced. The main issue of disregarding multi-step error propagation lies in the uncertainty of predicting how well any of the proposed approaches can perform in new studies.

3.3 Breadth of the context in which the mode detection is performed

The dimension among which all disciplines vary the most is the breadth of context (depth of the cube in Figure 1), which directly relates to the subject of each study. LBS has a very focused context, which is inferring the transportation mode of a user (usually a stand-alone operation). TSc has a broader application than LBS, although it is also focused on a specific task, which is capturing how people travel, which contains the task of detecting how users travel en route to their destinations. HG has a very broad application that mostly segments the trajectories into stop and move segments as a prerequisite to domain specific semantics enrichment (e.g., understanding the effect of changes in the city's infrastructure - Noland, 2003 -, modeling the movements of fishing boats - Rocha et al., 2010 -, understanding why and how people migrate in between regions - Hägerstrand, 1962 -, or understanding decisions at an individual level - Hägerstrand, 1970).

The interpretation of each discipline of the mode detection question caters to different application scenarios, which in turn propose solutions that have different characteristics. The difference between disciplines is mainly governed by the time frame of the case study and the number of required participants. Mode detection is a task that relies on having a ground truth dataset that is usually obtained from users annotating their traces. In LBS, the interest is on obtaining data over long periods of time, in which case it is difficult to recruit users. Contrarily, most TSc efforts that are geared towards the automation of travel diaries put more emphasis on gathering data from a great number of users for a predefined period of time of several days (similar to traditional activity-travel diaries - Prelipcean, Gidófalvi, & Susilo, 2015) or more than one week (for research that also studies daily purpose variability -Axhausen et al., 2003). Even though these constraints might not be causal, there is a drastic difference between the methodology behind TSc mode detection and LBS mode detection.

LBS employs mostly statistically robust machine learning methods (Xu & Mannor, 2012) that perform well on different datasets: 1) supervised learning methods such as Decision Trees, Random Forests and Support Vector Machines, 2) boosted classifiers such as AdaBoost, or 3) multi-layered approaches such as Decision Trees followed by either a Discrete Hidden Markov Model, Rule-based Heuristics, or Support Vector Machines. TSc, on the contrary, uses highly empirical ad-hoc methods such as: 1) Rule-based Heuristics, and 2) Fuzzy Logic. The use of Fuzzy Logic has been both challenged (Elkan et al., 1994) and praised (Zadeh, 2008), and while there is no general consensus in the scientific community whether Fuzzy Logic should be used or not, there are problems with using Fuzzy Logic for mode detection. First, the empirical approach for generating rules for mode detection relies on the expert's understanding of the available dimensions (usually limited to a subset of the available dimensions). Second, empirical rule generation is not guaranteed to take into account inter-dimension correlation. Finally, since the traditional methods uses human judgment to define the rules, any class addition to the classification scheme is costly. Similarly, using machine learning methods has drawbacks, most notable the dependance on large amounts of labeled data needed to train classifiers (Banko & Brill, 2001; Zhu, Vondrick, Ramanan, & Fowlkes, 2012), the difficulty of dealing with imbalanced datasets (Van Hulse, Khoshgoftaar, & Napolitano, 2007), and the difficulty of obtaining and interpreting the rules learned by the algorithms (Andrews, Diederich, & Tickle, 1995; Krause, Perer, & Bertini, 2016).

Furthermore, the distinction between the classification methods affects how the models are validated, i.e., how well a model that is trained on a subset of the data (resampled dataset) performs on the remaining unrevealed dataset. Most common methods are either k-fold validation – the original sample is randomly partitioned into k sized subsamples, each of the sub-samples is used once as validation data – or hold-out validation – the original sample is randomly partitioned into two specified size samples, one of which is used for training the classifier, the other for validation. In most TSc approaches that rely on Fuzzy Logic or Rule-based Heuristic for mode detection, model validation is not mentioned since the dataset is not resampled and is revealed to the experts that build the rules (Chung & Shalaby, 2005; Tsui & Shalaby, 2006; Stopher et al., 2008; Bohte & Maat, 2009; Schüssler & Axhausen, 2009; Schüssler et al., 2011; Biljecki et al., 2013; Rasmussen et al., 2013, 2015), which raises the issue of overfitting (Jin, Von Seelen, & Sendhoff, 1999; Hawkins, 2004), i.e., the model is susceptible to not performing well on different datasets. Contrary to this, the approaches that employ classical machine algorithms are using k-fold validation (Stenneth et al., 2011; Prelipcean et al., 2014; Yu et al., 2014), hold-out validation (Zheng et al., 2008, 2010) or variations of cross-validation such as leave one user out (Hemminki et al., 2013) or grouped cross-validation (Montini et al., 2014).

The type of sampling used also indicates what is required from the classifier, where LBS approaches intend to propose methods that can provide real-time information to the user and can be used over long periods of time, and TSc approaches intend to classify the data collected over the case study period of time, without giving any output to a user. This raises also the issue of user benefit, where LBS provides fresh and relevant information in real time that can be linked into a service (e.g., informing users to avoid a congestion), and TSc provides information that affects the users in the long time (e.g., the improvement of public transportation infrastructure due to the study of how groups of individuals move). These discussions are summarized in the third row of Table 5.

4 Gaps in research and areas for future work

With regard to Figure 1, an "ideal focus line" can be defined as the convergence for the LBS, TSc and HG approaches to provide an answer for the generic question of mode detection (see Table 1). Considering the previous section, LBS performs mode detection in real time at a point level, and TSc and HG perform mode detection on bulk-trajectories at a segment level. An ideal solution would perform mode detection in real-time, and the precision and recall values would be high when computed for both point and segment entities. Because it is difficult to define the breadth of context of an ideal scenario, the ideal line is not constrained by it. Moving towards the ideal line poises several challenges.

The most difficult challenge is modeling both point and segment entities to allow for the real-time broadcast of the detected mode for LBS, which can also be used to obtain same-mode segments suitable for TSc and HG. While this might not seem problematic given the approximately 90% precision reported by most papers, an in-depth study (Prelipcean et al., 2016) reveals that using methods that classify locations into modes to form segments is accompanied by a precision drop of approximately 40%. Furthermore, the error is computed as the percentage of points correctly classified or as the percentage of segments correctly classified without taking into account the spatial and temporal factors such as the length / duration of a segment, or the distance / duration between two consecutive points, i.e., all errors are treated as equal (Prelipcean et al., 2016). To gain a deeper insight on how to overcome this challenge, scientists should report the performance of the proposed methods in a frame that is common for both point and segment entities, such as that proposed by Prelipcean et al., 2016, which could help researchers identify the most promising approaches for mode detection.

The next challenge is related to data acquisition and it is two-fold: 1) identify how to collect data from multiple users without (substantial) extra costs, and 2) having a "benchmark" dataset, which is very diverse, accompanied by "benchmark" performance measures that all scientists can use to test their methods. First, data collection can be achieved by using already developed seamless data collection systems (such as the MEILI Mobility Collector - Prelipcean et al., 2014), but reaching a large number of users is difficult without offering useful information back to the users, e.g., reminding users to hurry up to catch a bus, or informing users to change their most often used path to a place because of traffic congestion (which is reminiscent of calm technology – Weiser & Brown, 1996, 1997). Second, having a "benchmark" dataset such as the Geolife dataset (Zheng et al., 2008, 2009, 2010) collected from a great number of users travelling with different modes is difficult because of the magnitude of the collection, the willingness of the users to share their data, and the legal efforts behind making such a dataset available. Similarly, the "benchmark" performance measures have to be robust and accepted by the scientific community.

While it is not obvious how to develop solutions to overcome these challenges, future research in mode detection could focus on any of the aforementioned proposals to bring science closer to solving the issue of mode detection and to the acceptance of such a solution.

5 Conclusion

This paper presents an in-depth analysis of the research that surrounds transportation mode detection with an emphasis on how three of the major disciplines, i.e., LBS, TSc, and HG, approach this task. The paper is structured in three main parts that: provide an overview of the main body of literature specific to each discipline, identify how solutions are different in between disciplines, and identify the gaps in research that are good candidates for future work.

The first part of the paper identifies, for each discipline, both the seminal work, and the most recent research approaches that were built on that work. This part mentions the choice of body of literature, and presents an overview of the research, which is a methodology comparison in terms of: employed methods, classification features, and number of identified modes. The second part of the paper identifies three dimensions among which the studied disciplines vary greatly due to their different interpretation of the generic transportation mode question, i.e., the “data”, “context”, and “entities” dimensions. Each of these dimensions is thoroughly analyzed while identifying strong and weak points for the approaches proposed by each of the disciplines. The last part of the paper tries to establish the “common ground” in between disciplines and propose a direction of research that could constitute a meeting point in between disciplines, and identifies the most difficult challenges while doing so.

While a substantial amount of research has been invested in transportation mode detection, there is no acceptable solution, or a promising approach. Due to the fact that research in one field continues the research in its own field, and no inter-disciplinary solutions have been attempted and documented, there is a substantial amount of research that still has to be invested in defining:

1) widely accepted error measures that are meaningful for both entities (points and segments),

2) classification and model validation methods that are robust to new datasets and not prone to overfitting, and

3) a “benchmark” dataset to validate the performance of the proposed methods.

These steps would allow the studied disciplines to reach a “common ground” by avoiding the sole validation of new methods on newly collected and non-shared datasets, and hopefully achieving inter-disciplinary method convergence.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., & Vaisman, A. (2007). A Model for Enriching Trajectories with Semantic Geographical Information. In *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems* (pp. 22:1–22:8).

Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-based systems*, *8*(6), 373–389.

Andrienko, G., Andrienko, N., & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *SIGKDD Exploration Newsletter*, *9*(2), 38–46.

Apple. (2015). *Motion activity.* Retrieved 2015-08-06, from `https://developer.apple.com/library/ios/documentation/CoreMotion/Reference/CMMotionActivityManager_class/`

Axhausen, K. W., Schönfelder, S., Wolf, J., Oliveira, M., & Samaga, U. (2003). 80 weeks of GPS-traces: Approaches to enriching the trip information. *83rd Transportation Research Board meeting*.

Banko, M., & Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. In *Proceedings of the 39th annual meeting on association for computational linguistics* (pp. 26–33). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from `http://dx.doi.org/10.3115/1073012.1073017` doi: 10.3115/1073012.1073017

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

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

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Bricka, S., & Bhat, R., Chandra. (2006). Comparative Analysis of Global Positioning SystemBased and Travel SurveyBased Data. *Transportation Research Record: Journal of the Transportation Research Board*(1972), 9–20.

Chung, E.-H., & Shalaby, A. (2005). A Trip Reconstruction Tool for GPS-based Personal Travel Surveys. *Transportation Planning and Technology*, *28*(5), 381-401.

Das, R. D., Ronald, N., & Winter, S. (2014). Clustering based transfer detection with fuzzy activity recognition from smart-phone GPS trajectories. *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*(17), 3138–3143.

Draijer, G., Kalfs, N., & Perdok, J. (2000). Global positioning system as data collection method for travel research. *Transportation Research Record: Journal of the Transportation Research Board*(1719), 147–153.

Drchal, J., Certicky, M., & Jakob, M. (2015). Fully agent-based simulation model of multimodal mobility in European cities. In *Proceedings of the 4th International Conference on Models and Technologies for Intelligent Transportation Systems* (pp. 229–236).

Elkan, C., Berenji, H., Chandrasekaran, B., De Silva, C., Attikiouzel, Y., Dubois, D., Prade, H., Smets, P., Freksa, C., Klir, G., & Yuan, B. (1994). The paradoxical success of fuzzy logic. *IEEE expert*, *9*(4), 3–49.

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. a. M. (2010). Preprocessing Techniques for Context Recognition from Accelerometer Data. *Personal and Ubiquitous Computing*, *14*(7), 645-662.

Gong, L., Morikawa, T., Yamamoto, T., & Sato, H. (2014). Deriving Personal Trip

Data from GPS Data: A Literature Review on the Existing Methodologies. *Procedia - Social and Behavioral Sciences*, *138*, 557 – 565.

Google. (2015). *Activities.* Retrieved 2015-08-06, from `https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognition`

Hägerstrand, T. (1962). Geographic measurements of migration. Swedish data. *Human displacements: Measurements and methodological aspects*, 61–85.

Hägerstrand, T. (1970). What about people in regional science? *Papers in regional science*, *24*(1), 7–24.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, *44*(1), 1–12.

Hemminki, S., Nurmi, P., & Tarkoma, S. (2013). Accelerometer-based transportation mode detection on smartphones. *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*(13), 13:1–13:14.

Jin, Y., Von Seelen, W., & Sendhoff, B. (1999). On generating fc 3 fuzzy rule systems from data using evolution strategies. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *29*(6), 829–845.

Krause, J., Perer, A., & Bertini, E. (2016). Using visual analytics to interpret predictive machine learning models. *arXiv preprint arXiv:1606.05685*.

Krumm, J., & Horvitz, E. (2006). Predestination: Inferring Destinations from Partial Trajectories. In *Proceedings of the 8th International Conference on Ubiquitous Computing* (pp. 243–260). Springer-Verlag.

MacEachren, A. M. (1995). *How maps work : representation, visualization, and design.* New York, London: Guilford Press.

Manzoni, V., Maniloff, D., Kloeckl, K., & Ratti, C. (2010). *Transportation mode identification and real-time CO2 emission estimation using smartphones* (Tech. Rep.). Massachusetts Institute of Technology.

Montini, L., Rieser-Schüssler, N., & Axhausen, K. W. (2014). Personalisation in multi-day GPS and accelerometer data processing. In *Proceedings of the 14th Swiss Transport Research Conference (STRC).*

Noland, R. B. (2003). Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. *Accident Analysis & Prevention*, *35*(4), 599–611.

Ogle, J., Guensler, R., & Elango, V. (2005). Georgia's commute Atlanta value pricing program: Recruitment methods and travel diary response rates. *Journal of the Transportation Research Board*, *1931*(1), 28–37.

Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A Clustering-based Approach for Discovering Interesting Places in Trajectories. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 863–868).

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., Theodoridis, Y., & Yan, Z. (2013). Semantic Trajectories Modeling and Analysis. *ACM Computing Surveys*, *45*(4), 42:1 – 42:32.

Prelipcean, A. C., Gidófalvi, G., & Susilo, Y. O. (2014). Mobility Collector. *Journal of Location Based Services*, *4*(8), 229–255.

Prelipcean, A. C., Gidófalvi, G., & Susilo, Y. O. (2015). Comparative framework for activity-travel diary collection systems. In *Proceedings of the 4th International Conference on Models and Technologies for Intelligent Transportation Systems* (pp. 251–258).

Prelipcean, A. C., Gidófalvi, G., & Susilo, Y. O. (2016). Measures of transport mode segmentation of trajectories. *International Journal of Geographical Information Science*, *to appear*.

Prelipcean, A. C., Schmid, F., & Shirabe, T. (2015). A Space Time Alarm. In *Progress in Location-Based Services 2014* (pp. 187–198). Springer International Publishing.

Rasmussen, T. K., Ingvardson, J. B., Halldórsdóttir, K., & Nielsen, O. A. (2013). Using wearable GPS devices in travel surveys: A case study in the Greater Copenhagen Area. In *Proceedings from the Annual Transport Conference at Aalborg University.*

Rasmussen, T. K., Ingvardson, J. B., Halldórsdóttir, K., & Nielsen, O. A. (2015). Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area. *Computers, Environment and Urban Systems*.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2010). Using Mobile Phones to Determine Transportation Modes. *ACM Transanctions on Sensor Networks*, *6*(2), 13:1–13:27.

Rocha, J., Oliveira, G., Alvares, L., Bogorny, V., & Times, V. (2010). DB-SMoT: A direction-based spatio-temporal clustering method. In *Intelligent Systems, 2010 5th IEEE International Conference* (pp. 114–119).

Schönfelder, S., Axhausen, K. W., Antille, N., & Bierlaire, M. (2002). Exploring the potentials of automatically collected GPS data for travel behaviour analysis - A Swedish data source. *GITechnologien für Verkehr und Logistik*(13), 155–179.

Schüssler, N., & Axhausen, K. W. (2009). Processing Raw Data from Global Positioning Systems Without Additional Information. *Journal of the Transportation Research Board*(2105), 28–36.

Schüssler, N., Montini, L., & Dobler, C. (2011). Improving post-processing routines for GPS observations using prompted-recall data. *9th International Conference on Survey Methods in Transport*.

Shah, R. C., Wan, C.-y., Lu, H., & Nachman, L. (2014). Classifying the Mode of Transportation on Mobile Phones using GIS Information. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM.

Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, *34*(3), 316–334.

Stenneth, L., Wolfson, O., Yu, P. S., & Xu, B. (2011). Transportation mode detection using mobile phones and GIS information. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 54–63.

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

Tsui, S. Y. A., & Shalaby, A. S. (2006). Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems. *Transportation Research Record: Journal of the Transportation Research Board*(1972), 38–45.

Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2007). Experimental perspectives on learning from imbalanced data. In *Proceedings of the 24th international conference on machine learning* (pp. 935–942).

Wang, S., Chen, C., & Ma, J. (2010, April). Accelerometer Based Transportation Mode Recognition on Mobile Phones. *Wearable Computing Systems (APWCS)*, 44–46.

Weiser, M., & Brown, J. S. (1996). Designing calm technology. *PowerGrid Journal*, *1*(1), 75–85.

Weiser, M., & Brown, J. S. (1997). The coming age of calm technology. In *Beyond calculation* (pp. 75–85). Springer.

Wolf, J. (2000). *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.

Wolf, J., Oliveira, M., & Thompson, M. (2003). The impact of trip underreporting on VMT and travel time estimates: preliminary findings from the California statewide household travel survey GPS study. *Transportation Research Record*, *1854*, 189–198.

Xie, K., Deng, K., & Zhou, X. (2009). From Trajectories to Activities: A Spatiotemporal Join Approach. In *Proceedings of the 2009 International Workshop on Location Based Social Networks* (pp. 25–32). ACM.

Xu, H., & Mannor, S. (2012). Robustness and generalization. *Machine learning*, *86*(3), 391–423.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., & Aberer, K. (2011). SeMiTri: A Framework for Semantic Annotation of Heterogeneous Trajectories. In *Proceedings of the 14th International Conference on Extending Database Technology* (pp. 259–270).

Yu, M.-C., Lin, C.-J., Yu, T., Chang, E. Y., & Wang, S. (2014). Big data small footprint: The design of a low-power classifier for detecting transportation modes. *The Proceedings of the VLDB Endowment*, *7*, 1429-1440.

Zadeh, L. A. (2008). Is there a need for fuzzy logic? *Information sciences*, *178*(13), 2751–2779.

Zheng, Y., Chen, Y., Li, Q., Xie, X., & Ma, W.-Y. (2010). Understanding Transportation Modes Based on GPS Data for Web Applications. *ACM Transac-*

tions on the Web, *4*(1), 1:1–1:36.

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008). Learning Transportation Mode from Raw Gps Data for Geographic Applications on the Web. *Proceedings of the 17th International Conference on World Wide Web*, 247–256.

Zheng, Y., Xie, X., & Ma, W.-Y. (2009). GeoLife 2.0: A Location-Based Social Networking Service. In *Mdm 2009.* IEEE Computer Society.

Zhu, X., Vondrick, C., Ramanan, D., & Fowlkes, C. (2012). Do we need more training data or better models for object detection?. In (Vol. 3, p. 5).

Zimowski, M., Tourangeau, R., Ghadialy, R., & Pedlow, S. (1997). Nonresponse in household travel surveys. *Report to the Federal Highway Administration*.

A An overview of the existing research in terms of modes identified, methods used, performance measures, collected data, processing method (pre- or post-processing), devices used, sampling rate, auxiliary data, and geographical regions

Table 6: Overview for the considered body of research in chronological order. The values in bold represent studies that have the highest precision in their field (+) or highest number of classified modes in their field (*)

Author	Dom.	Modes[1]	Method.[2]	Perform.[3]	Col. data[4]	Other Proc.[5]	Dev.[6]	Sampl.	Data[7]	Region
Chung & Shalaby, 2005	TSc	W, Bk, C, Bu	RBH	P_S=92.7%	1 user 60 segments reproduced	A: NF A: MTP A:MM	GL	–	RN	GTA, CA
Tsui & Shalaby, 2006	TSc	W,Bk, C, Bu Sw, SC, OR	RBH FL	P_S=94%	9 users 237 segments 58 travel days	A: NF A: MTP A:MM	GL	–	RN	–
Alvares et al., 2007	HG	St, Mo	RBH	–	–	–	–	–	POI	–
Palma et al., 2008	HG	St, Mo	STC	–	– 487 segments	–	DD	–	POI	Amsterdam, NL
Stopher et al., 2008 (+)	TSc	W,Bk, C Bu, Tr	RBH PM	P_S=95%	–	A:SD A:MM	DD	GPS: 1Hz	RN	–
Bohte & Maat, 2009	TSc	W,Bk, C Tr, PT, O	RBH	P_S=70%	1104 users 33,686 segments 1 week	A:NF A:SD	DD	GPS: 0.17Hz	TS	NL

Continued on next page

Table 6 – *Continued from previous page*

Author	Dom.	Modes	Method.[2]	Perform.[3]	Col. data	Other Proc.	Dev.	Sampling	Data	Region
Schüssler & Axhausen, 2009	TSc	W,Bk, C UPT, Tr	FL	–	4,882 users –	A:NF A:SD A:MTP	DD	–	–	CH
Manzoni et al., 2010	LBS	St, W,Bk, C Bu, Tr,Sw, Mo	DT	P_W^{512}=82.14%	4 users – 10 days	A:FW	SP	GPS: 1 Hz Acc: 25Hz	–	–
Reddy et al., 2010	LBS	St, W, Bk, Run, C	DT DHMM	P_P=93.7% R_P=93.8%	16 users – 1 day	A:NF A:FW	SP	GPS: 1Hz Acc: 32Hz	–	–
Rocha et al., 2010 (+,*)	HG	St, Mo	DTC	P_S=90%	2 users – 22 days	–	DD	GPS: 30 mins	–	BR
Wang et al., 2010	LBS	W,Bk,St, C, Bu, Sw	DT-J48	P_W^{256}=70.7% –	7 users 5544 acc. win. 12 hours	A:FW	SP	Acc: 35Hz	–	–
Zheng et al., 2010	CSc	W, Bk, C, Bu	DT	P_S=75.6% R_S=73.8%	65 users – 10 months	P: GBP	DD	GPS: *var* Hz	–	–

Continued on next page

Table 6 – *Continued from previous page*

Author	Dom.	Modes	Method.[2]	Perform.[3]	Col. data	Other Proc.	Dev.	Sampling	Data	Region
Schüssler et al., 2011	TSc	W,Bk, C UPT, Tr	FL	P_S=83%	15 users 322 segments 1 week	A:NF A:SD A:MTP	SP	GPS: ?Hz Acc: 10Hz	TS	–
Stenneth et al., 2011	LBS	St, W,Bk, C,Bu,Tr	RF	P_W^2=92.8% R_W^2=92.9%	6 users – 21 days	–	SP	GPS: 0.07Hz	RN TF	Chicago, USA
Biljecki et al., 2013 (*)	TSc	W, Bk, C, Fb, Sb, Tr, Sw, Bu, Tm, Fl	CRF	P_S=91.6%	16m loc.	A:SD A:MTP	DD	GPS: *var* Hz	RN	NL
Hemminki et al., 2013	LBS	W,St, C, Bu Sw, Tr, Tm	HMM AdaBoost	P_W^{var}=84,9% R_W^{var} = 85.3%	16 users – 150 hours	A: FW	SP	Acc: 100Hz	–	Helsinki, FI
Montini et al., 2014	TSc	W,Bk, C Tm, Tr, O	RF	P_S=85.8%	56 users 6990 segments 1 week	A: SD	SP	GPS: 1 Hz Acc: ??Hz	TS	Zurich, CH
Prelipcean et al., 2014	LBS	W,Bk,C, Fb, Bu,Tr,Sw	RF	P_P=90.8% R_P=90.9%	11 users 14k loc. 10 days	A:FW	SP	GPS: 50 m Acc: 5Hz	–	Stockholm, SE

Continued on next page

Table 6 – *Continued from previous page*

Author	Dom.	Modes	Method.[2]	Perform.[3]	Col. data	Other Proc.	Dev.	Sampling	Data	Region
Shah et al., 2014	LBS	C, Bu, Tr	DT RBH	P_W^{512}=91.53% –	15 users 129 trips 50 hours	A:MS	SP	GPS: 0.2Hz Acc: 100Hz	–	San Francisco, US
Yu et al., 2014	LBS	W,St, Bk, Run, C	DT-J48 AdaBoost SVM	P_W^{512}=91.53% –	224 users 100GB 8,311 hours	A:FW	SP	Acc: 30Hz	–	–
Rasmussen et al., 2015	TSc	W, Bk, C Bu, Tr	FL	P_S=84.6%	183 users 521 segments 3-5 days	A: NF A: MTP A: MM	DD	GPS: 1 Hz	RN	Copenhagen, DK
Prelipcean et al., 2016 (+,*)	LBS & TSc	W, St, Bk Mp, C, Bu Tr, Tm, Sw	RF	P_S = 80.1% R_S = 82.4% P_P = 94.4% R_P = 94.5%	26 users 1307 segments 7 days	A: FW A: MTP	SP	GPS: 50 m Acc: 5Hz	-	Stockholm, SE

[1] The following abbreviations were used for modes: W-walk, Bk-Bike, C-Car, Fb-ferry boat, Bu-bus, Tr-train, Sw-subway, SC-street car, OR-off road, UPT-urban public transport, Tm-tram, O-other, PT- public transportation, Sb - sail boat, Fl-flight, Mp - moped;
[2] The following abbreviations were used for classification methods: RBH - rule based heuristics; FL- fuzzy logic; PM - probability matrix; STC - spatio-temporal clustering; DTC - direction and time based clustering;
[3] P and R denote precision and recall as computed for points (P_P and R_P), segments (P_S and R_S) or sliding windows (P_W and R_W)
[5] The considered procedures are either performed a-priori (A) or a-posteriori (P) with regards to the classification. The following procedures were considered: FW- form windows, MS - Markov smoother, MM - map matching, MTP - mode transfer points, NF - noise filter, SD - segmentation
[7] RN- road network; TS - public transport stops;