

REVIEW

Open Access



Transport behavior-mining from smartphones: a review

Valentino Servizi^{*} , Francisco C. Pereira, Marie K. Anderson and Otto A. Nielsen

Abstract

Background: Although people and smartphones have become almost inseparable, especially during travel, smartphones still represent a small fraction of a complex multi-sensor platform enabling the passive collection of users' travel behavior. Smartphone-based travel survey data yields the richest perspective on the study of inter- and intrauser behavioral variations. Yet after over a decade of research and field experimentation on such surveys, and despite a consensus in transportation research as to their potential, smartphone-based travel surveys are seldom used on a large scale.

Purpose: This literature review pinpoints and examines the problems limiting prior research, and exposes drivers to select and rank machine-learning algorithms used for data processing in smartphone-based surveys.

Conclusion: Our findings show the main physical limitations from a device perspective; the methodological framework deployed for the automatic generation of travel-diaries, from the application perspective; and the relationship among user interaction, methods, and data, from the ground truth perspective.

Keywords: Smartphone-based travel surveys, Machine learning, User behavior, Transport, Map-matching, Mode detection, Activity inference, Data fusion

1 Introduction

To support the planning, design, and policy-making processes for improving transport systems [44], travel surveys capture essential aspects of user behaviors on which behavioral modeling relies [18]. For designing the representativeness of a user sample under study, the statistical approach in traditional travel surveys is prominent. The process involves person-to-person (P2P) interactions for data collection, a process overlapping with ground truth collection: Trained travel surveyors directly validate data with users and manually reconstruct users' travel-diaries for behavioral study.

In contrast, machine-learning plays a primary role in smartphone-based travel surveys (SBTS). The data collection process involves device-to-device interaction, with machine-learning algorithms automatically

reconstructing users' travel-diaries directly from data that might contain various sources of errors [50]. By submitting each travel-diary to the user for validation (i.e., to find out whether the user needs to change the travel-diary or not), the process can collect ground truth through a person-to-device (P2D) interaction between the user and an input/output interface, either via a website or smartphone [59].

Since the introduction of the first generation of smartphones equipped with assisted global positioning systems (AGPS) in the early 2000s, researchers have described smartphone-based travel surveys as a promising platform to measure user transport behavior. They can track the same user with an extended time horizon [91], collect data passively [125], detect previously unreported short trips, and avoid stereotypes of daily activity [104] (e.g., "I don't remember what I did, but here's what I usually do"). Given that SBTS would likely facilitate the discovery of inter- and intra-user behavior variations, the question

*Correspondence: valse@dtu.dk

Department of Technology, Management and Economics, Technical University of Denmark (DTU), Kgs. Lyngby, Copenhagen, Denmark

is why SBTS have not yet replaced traditional travel surveys [41].

For researchers and public authorities, standardized performance indexes based on standard datasets support optimal investment decision-making. This approach also applies to classification or regression methods underpinning the identification of user transport behavior variations. Nevertheless, standardization in this field is lacking. Instead, decision-making often relies on assumptions, such as (i) consistent performance indexes evaluation across studies; (ii) comparable performance indexes across studies, even when based on different datasets; (iii) adequate representativeness of the few public datasets available; (iv) exact ground truth. By definition, each necessary assumption represents a knowledge gap.

We ask and answer the following questions: What are the main machine-learning methods that are used in the field? What is the relationship between ground truth and machine-learning methods? What are the primary datasets studied? What characteristics do these datasets have, and what features can we extract from them, and how? What are the challenges for machine-learning in the field of SBTS? What are the main implications for transport science?

To tackle these questions, we proceed by snowballing first forward and then backward [114]. We cover deterministic and machine-learning methods based on different datasets collected from across the world. We examine how models and algorithms exploit various data sources such as AGPS, inertial navigation systems (INS), geographic information systems (GIS), and Internet-of-Things.

The paper analyzes technologies enabling SBTS data validation, such as data preparation and feature extraction, and focuses on machine-learning methods for mining user's behavior from smartphone data. These methods target why people travel, where along the transport network they travel, and which mode of transport they use. These technologies make an impact by reducing resources associated with running traditional travel surveys, while enhancing users' transport behavior data-resolution. Following this approach, we are able to review purpose imputation, map-matching, and mode detection methods.

Existing literature and reviews offer a clear picture of how algorithms and background technologies evolve to provide improved measures of users' travel behavior variations. For example, we list several specialized methods with impressive performance scores. We also find unilateral perspectives offering standardization pathways for both methods application and performance evaluation. In practice, limitations such as data representativeness, ground truth quality, and performance evaluation

procedures may often result in a biased perception of each method's potential.

Decisions based on wrong assumptions and biased perceptions represent a threat to the progress of this field. To bridge the gap, we provide the following contributions. We deliver a self-contained overview connecting the user transport behavior measures with the supporting smartphone-sensing-platform. We detail how available methods can be combined to extract behavioral information from various data streams. We show the convergence between research areas studying complementary aspects of transport behavior. We organize each reviewed work by task complexity, method requirements, and dataset representativeness. So we facilitate methods' assessment and comparison across specific use cases, mitigating the limitations of dry and incomparable performance scores. The paper reveals opportunities offered by device-to-device interactions for data validation instead of other interactions, and exposes gaps in deep learning strategic applications.

The first section below presents the dimensions describing transport behavior and the tools embodied in a smartphone device for data collection. The following section describes the methods used to identify transport behavior from data and an overview of the implications for transport science. The subsequent discussion presents a joint look on the results of the surveyed literature, which the conclusion summarizes from a big-data perspective. We include the Tables organizing the main features of the literature reviewed.

2 Measures and tools

To support the reader through the following analysis and discussion, we start by providing context and presenting concepts on which the paper rests, i.e., definitions, employment, and technological framework of SBTS.

2.1 Measures of transport behavior

The following terms are used to describe a user's journey (throughout a single day, for example; see Fig. 1) and represent the different variables, or measures, that SBTS is used to collect for studies on transport behavior.

2.1.1 Tour

Aggregation of trips, such that users' travels start and end at the same place, e.g. at home [28].

2.1.2 Trip

Travel entity identified with a set of attributes such as: start-location, start-time, purpose, transport mode, arrival time, arrival location [28].

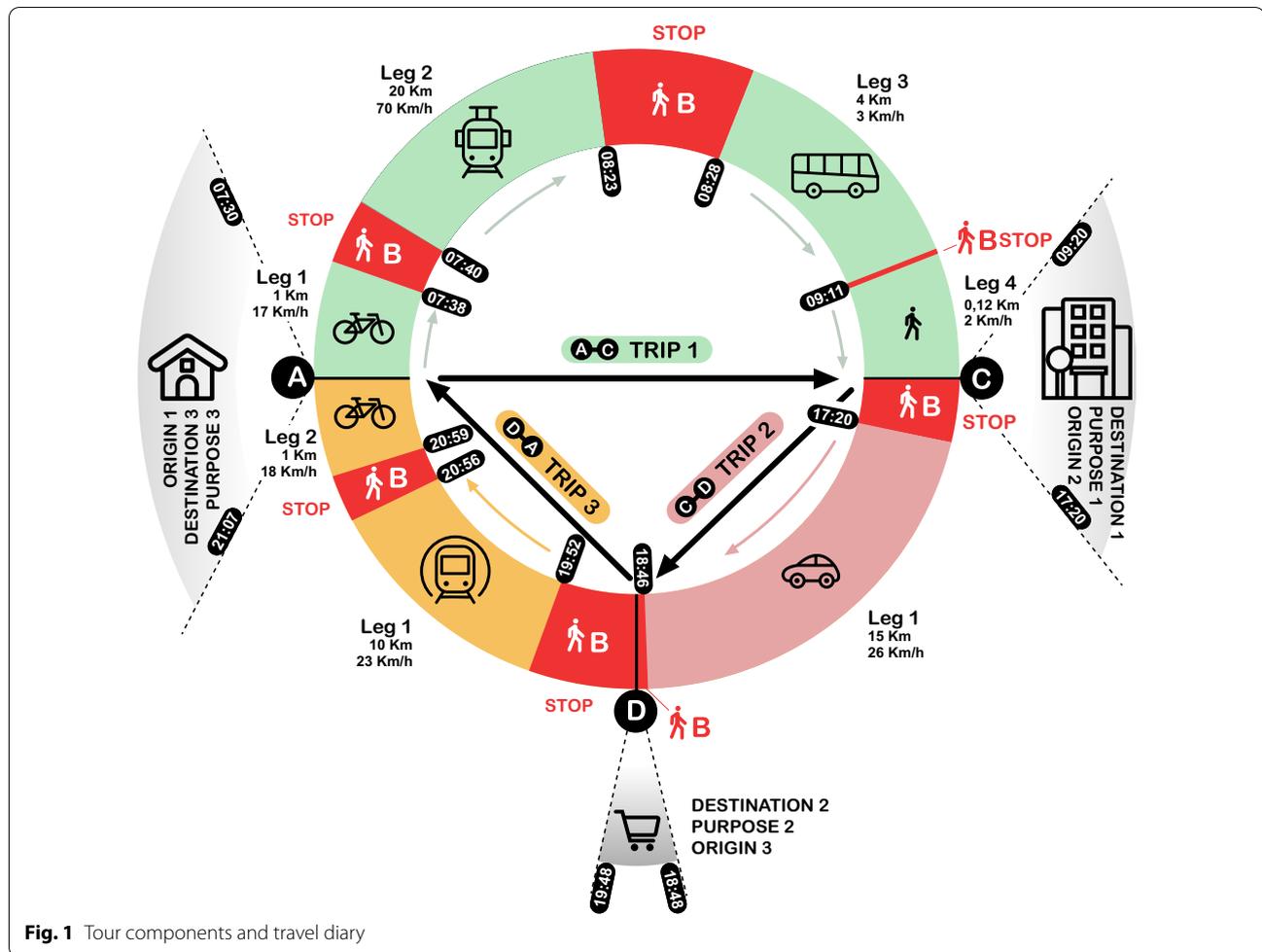


Fig. 1 Tour components and travel diary

2.1.3 Leg

Also identified as a “trip segment,” this is the unimodal segment between two stops. Each trip segment has a start-time and -location, end-time and -location, and stop-purpose at the end of the leg (see Fig. 1B) [28, 96].

2.1.4 Purpose

This represents what triggers the trip from origin to destination (see Fig. 1A, C, D), and identifies the “activity” performed at the end of a trip.

2.1.5 Stop

This can be reduced to two categories: stops at the end of legs (see Fig. 1B), and stops at the end of trips (see Fig. 1A, C, D).

2.1.6 Transport mode

This refers to a trip leg [120] and identifies, e.g., walking, cycling, car, train, bus, light rail (see Fig. 1).

2.1.7 Mode-chain-type

The literature provides no strict consensus on the definition of this term, and we define it as the list of transport modes one uses to get from the origin to the destination of a trip (see Fig. 1).

2.1.8 Travel-diary

This can focus on “one-day” (see Fig. 1) or on “multiple-days” and it describes the user trips through: (i) legs, where each leg has a unique transport mode; (ii) purpose; (iii) stops; and (iv) mode-chain-type. Generally, it is linked to a user, and his or her link-able personal information, such as: (i) age; (ii) occupation; (iii) education level; (iv) home address; and (v) work address. [28] presents a detailed list of further personal attributes.

2.1.9 Ground truth

This describes the true measurements of the target variables, for example the purpose of a trip, its

transport-mode-chain, and the route between origin and destination. In general, the literature refers to (i) travel-diary; (ii) prompted recall survey; (iii) user input in mobile phones [12]; (iv) experiments (e.g. mode known) [98]; (v) trips reported in-situ by the user participating in an experiment [33]; and (vi) “traffic counts” extracted from video recordings [53]. However, because ground truth is lacking in several studies [26], authors have introduced alternative methods to close this gap, the results of which serve as a benchmark [54]. In case of synthetic data, studies on map-matching refer to the random selection among a set of alternative shortest paths [71]; in case of real data, other studies refer to GPS receivers collecting two independent measures, where ground truth is the measure with a higher sampling rate [51]. When algorithms target public transportation, ground truth can be extracted as the combination of bus stops and intersections within the transport network [42]. In the best-case scenario, the information is reported by users. As ground truth always seems prone to errors, Prelipcean et al. [87] have introduced the concept of “*acceptable truth*,” which, while not truly absolute, may be considered sufficiently accurate relative to the application.

2.2 Pioneering smartphone-based travel surveys

Within the last 20 years, traditional travel survey methods have been subject to the pressure of disruptive technological evolution. The large penetration of smartphone devices equipped with low-cost sensors, the introduction of Web 2.0, and the emergence of other directly related phenomena, such as Big Data [5], could represent a tipping point for this research method [125]. There are several reasons to complement and/or substitute traditional travel surveys with smartphone-based technology, given the former’s shortcomings, as follows:

1. Statistic representativeness, improvable or decreasing in some population’s strata [80];
2. Trend of unreported short trips which the user tends to forget or does not want to mention [104];
3. Undetected behavior variations of the same user, due to the design of traditional travel surveys, which collects a cross-section sample of the population by focusing on one single day for each respondent [91];
4. Data collection cost per surveyed user [101].

The first large-scale SBTS deployments were the Future Mobility Sensing (FMS) in 2012, and the Sydney Travel and Health Survey in 2013. Most of the SBTS we know offer either web or app validation (seldom both), use machine learning, and are fully automated, as for example: (i) FMS/Mobile Market Monitor [126]; (ii)

TRAVELVU/Trivector [38]; (iii) RMOVE/RSG [23]; (iv) Itinerum [83, 84]; (v) MEILI [87]; (vi) Sydney Travel and Health Survey [45]; (vii) Dutch Mobile Mobility Panel [104]; and (viii) MTL Traject [123].

These SBTS no longer collect ground truth via person-to-person interaction. Instead, their interfaces provide users with options to validate travel-diaries accurately generated, and to correct errors of the inaccurate ones, collecting ground truth via person-to-device explicit interaction. Nonetheless, users seem unable to report inaccurate diaries that are too difficult for them to correct on their own [102]. Consequently, the risk of encountering incorrect data within ground truth seems unavoidable for survey data. Regardless of whether available ground truth is acceptable or inaccurate, it is important to assess each application on an individual basis in the context of field research.

Success depends also on users’ willingness to keep such an application installed on their smartphones. The main drivers determining the decision of a user to keep applications on his or her device are: (i) The information conveyed through the App; (ii) ease of use; (iii) perceived usefulness; (iv) perceived risks; and (v) general satisfaction of the user experience [10].

In (v) we mention a broad and very relevant field of research in which there is consensus about the negative impact of smartphone battery consumption on the user experience, which affects applications’ penetration and drop-out rates. Because of the impact on quality of data collection, we observe the same consensus on battery concerns in the field of SBTS [80]. Also, the need of high resolution data in SBTS clashes with the need for battery efficiency enforced by smartphone platform providers [7].

Due to the highly-accurate trajectories generated by smartphones (e.g., through AGPS) and used by SBTS researchers, users are concerned by the potential for privacy violation. These trajectories often expose very personal information of each surveyed user, thereby presenting new challenges [35] in terms of reconciling a need for high-resolution data and a need to ensure privacy for researchers and users, respectively [88, 95].

2.3 Smartphone capabilities

In Fig. 2 we present the abstraction of an SBTS platform. The main platform’s components are client and server. The client (see Fig. 2A) enables human interaction, e.g., for user travel diary validation (see Fig. 2A.1), and orchestrates sensors, user-generated data (e.g., location), and computer intelligence models. Processing data locally, the client prevents loss of information, and maximizes privacy (see Fig. 2A.3). A battery efficiency layer tunes and optimizes, e.g. data sampling or network

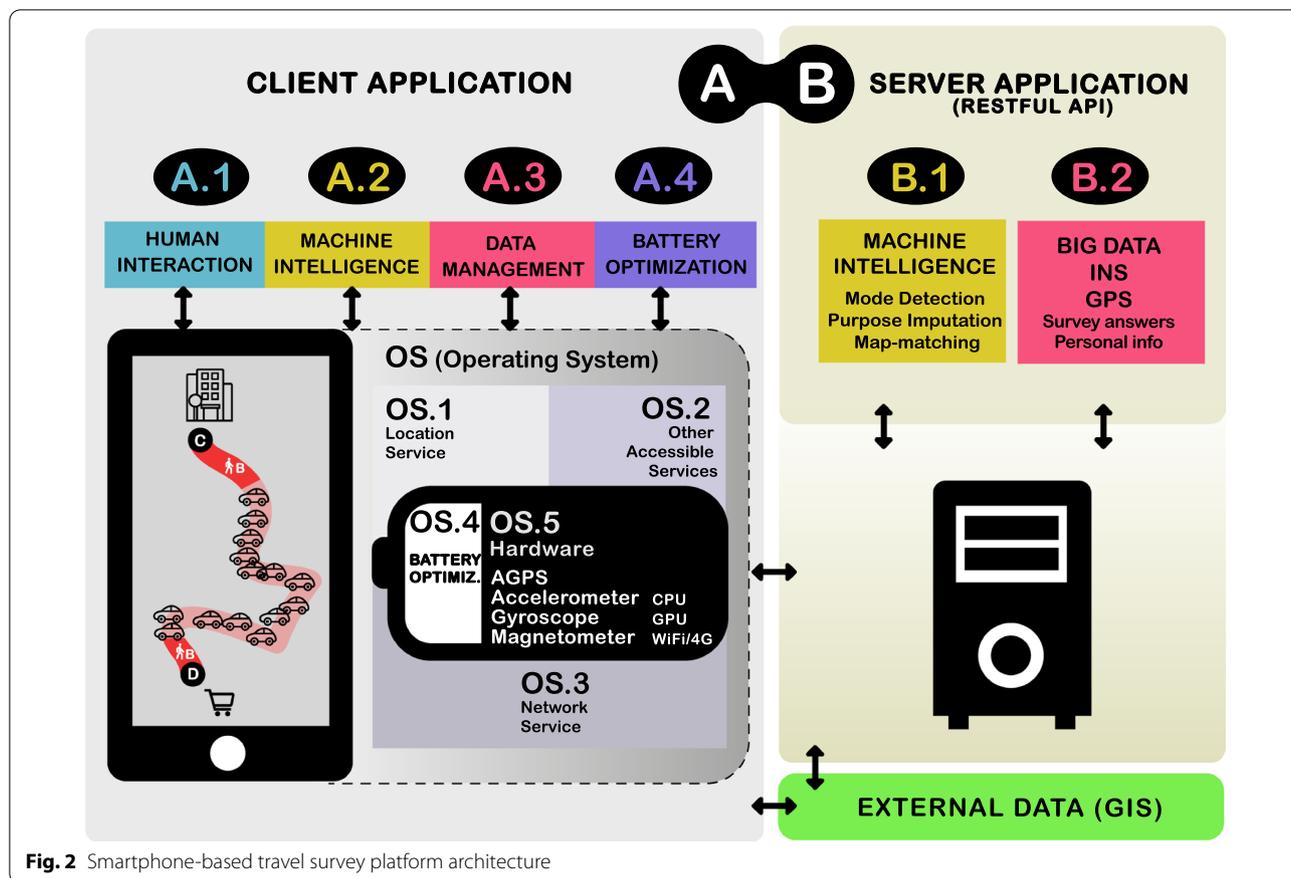


Fig. 2 Smartphone-based travel survey platform architecture

input/output operations among client, server, and external data sources (e.g., GIS).

The sensory system of the platform is the smartphone, represented by:

- Principal hardware components (see Fig. 2OS.5);
- Services exposed by the Operation System (OS, see Fig. 2OS.1–OS.3); and
- Operations beyond users and developers influence, such as those focusing on device battery life extension (see Fig. 2OS.4).

The following list of components is ranked by highest battery consumption to lowest [85, 112]:

1. Graphical processing unit (GPU) and screen, triggered when users interact actively with SBTS (e.g., validating travel-diaries).
2. Central processing unit (CPU), engaged also by computer intelligence models for online mode classification, for example, and for detecting conditions to switch off unnecessary sensors. While computation offloading to a server is possible, it implies transmitting data at its own energy cost.

3. AGPS. While GPS depends exclusively on satellites, in smartphones AGPS uses internet to look up the position of satellites and mitigate the cold-start problem. AGPS also uses cell-tower data. This feature is convenient when GPS signal is weak or disturbed, but it introduces challenges for position accuracy. To provide the location of a smartphone while reducing AGPS up-time, several effective strategies are available [82]. Finding the best trade-off between location accuracy, data resolution, and energy consumption is not trivial. Interestingly, we observe a convergence between approaches developed for the OS to improve the energetic efficiency of smartphones, and for datamining to fill data gaps resulting from missing or highly uncertain GPS observations. Both provide location coordinates, reducing GPS sensor need, and leveraging data from INS, GIS, and telecom networks. Nevertheless, some of the current smartphone operation systems do not allow direct access to telecom network data from independent applications [6].
4. Network. An efficient tuning should consider network selection (Cellular or WiFi), data transfer frequency, battery status, and size of the data-transfer.

5. Accelerometer, gyroscope, and magnetometer raw data is accessible on the main OS platforms. GPS up-time is often optimized by leveraging these sensors to detect whether a user starts or ends a trip [82]. In general, accelerometer and gyroscope readings from smartphones should be collected with a resolution compatible with the motion frequency of human bodies in daily routines, which is above 20 Hz [49]. The consumption of such high-frequency data streams within the device is not critical for the battery. However, in case of transfer for storage and data consumption offline, handling the number of sensors and the high frequency quickly become critical for the smartphone's battery and for the user's data plan.

Sensors up-time and data transfer to the back-end, as well as the Ground Truth collection on screen are very critical for smartphones battery life [69]. For example, given a fixed data sampling rate, AGPS battery consumption is relatively more sensitive to the up-time, while high frequency sensors consumption is relatively more sensitive to data transfer. If not properly handled within the SBTS, battery drain could occur twice as fast, limiting the battery life to few hours instead of the whole day. Consequently, the impact of service interruptions would result in increasing limitations on the data. Covering the entire day for certain users would no longer be possible, and such a negative user experience would even increase risk of drop-out [10].

2.4 Physical limitations for data validation

In addition to the aforementioned battery consumption issues, further critical implications of moving to this new technology are presented below.

2.4.1 Person-to-device validation

Design simplicity and intuitiveness should reduce any potential to distract the user while interacting with the survey application, as distractions could impact the quality of ground truth collected [80]. Furthermore, when the purpose of the interaction is directed to amend inaccurate travel-diaries, the impact that the design has on the quality of the ground truth collected from the respondents is even greater. A poor interaction between users and an SBTS interface could trigger a critical loop in which users validate wrong predictions instead of correcting them [3, 30].

2.4.2 Device-to-device validation

Arising from the convergence of Bluetooth and WiFi protocol in the Internet of Things context, and unlike the classic Bluetooth protocol, Bluetooth low-energy beacons communication is one-to-many (as traditional television

or radio), involves few bits of data to be broadcast frequently, and needs no pairing operations. These properties are suitable for proximity detection and interaction with smartphones, and for activity sensing [34, 56]. A pioneering device-to-device ground truth collection on bus trips [66] already experimented Bluetooth low-energy interaction with SBTS, as an independent and redundant measurement of users' bus trips. This system has the potential to release users' resources that could be exploited, for example, for filling in context-specific active surveys, and not for validating a travel diary. However, the authors highlight the challenge of finding a signal strength that allows for smartphones to detect beacons in conditions where signals may be attenuated or interfered with. A user's body or location, for example, may attenuate a signal, while interference with other beacons in range could result from passing by a bus stop or grouping with other buses.

3 Measuring transport behavior

The primary objective of SBTS consists of accurate ground truth collection from surveyed users. The correct reconstruction of travel-diaries, which encompasses both the transport mode and the purpose of any trip, allows for this goal to be achieved. Research on transport behavior also studies trajectories generated by the same sensors mentioned earlier. Therefore, it applies the same methods described in the following sections. In contrast with SBTS, however, research on transport behavior has the main objective of analyzing behavior, and not of collecting trip ground truth. This subtle difference may support the large community of researchers claiming that mode detection methods should be agnostic to personal and location context (see Tables 1, 2, 3). For example, the same method could generally serve different mode choice studies across the globe. In SBTS, this constraint does not seem to hold since travel-diaries also require predicting each trip's purpose, relying on both sensors and geospatial information (see Table 6). Successful hybrid approaches in this field further expose the shortcomings of such a purist approach. Data preparation is propaedeutic for learning the mode, purpose, and route of any trip. Simultaneously, cross-field convergence proves to be effective; for example, mode detection improves map-matching [26] and purpose imputation tasks [76, 120]. Inversely, map-matching GPS trajectories upfront improves the mode detection task [90]. When outputting a travel diary that allows ground truth collection on users' journeys, we do not find advantages from self-imposing restrictions on what data we should use or what method we should combine. Therefore, we find it beneficial to review purpose imputation and map-matching methods

Table 1 Classification task ranked by difficulty and score, for mode detection

References	No.	Classes	Score	Metric	Validation	Area
Zhou et al. [131]	6	Walk, Bike, Bus, Car, Rail, Plain	86.5%	Accuracy	Hold-out	Beijing
Bohte and Maat [21]	6	Car, Train, Bus-Tram-Metro, Foot, Bicycle, Other	70.00%	Accuracy	n.p.	Netherlands
Martin et al. [75]	5	Walk, Bike, Bus, Car, Rail	96.8%	Accuracy	Manifold-cross-validation	Minnesota
Jahangiri and Rakha [55]	5	Walk, Bike, Bus, Car, Run	95.1%	F-Score	Manifold-cross-validation, Out-of-bag-estimate	Tennessee
Semanjski et al. [96]	5	Walk, Bike, Bus, Car, Rail	94.00%	Accuracy	Manifold-cross-validation	Leuven
Zhou et al. [132]	5	Walk, Bike, Run, in-Vehicle, Stationary	93.8%	Accuracy	Hold-out	Georgia (USA)
Zhu et al. [134]	5	Walk, Bike, Bus, Car, Rail	93.45%	F1-Score	Manifold-cross-validation	Beijing
Xiao et al. [119]	5	Walk, Bike, el-Bike, Car, Bus	92.74%	Accuracy	Manifold-cross-validation	Shanghai
Rasmussen et al. [90]	5	Walk, Bike, Car, Bus, Rail	92.4%	Accuracy	n.p.	Copenhagen
Yazdizadeh et al. [122]	5	Walk, Bike, Public transit, Car, Car and Public transit	88.00%	F1-Score weighted average	Manifold-cross-validation	Montreal
Dabiri and Heaslip [31]	5	Walk, Bike, Bus, Car, Rail	84.8%	F-Score	Manifold-cross-validation	Beijing
Byon and Liang [22]	5	Auto, Bus, Streetcar, Bike, Walk	82.00%	F1-Score weighted average	Hold-out	Toronto
Thomas et al. [104]	5	Walk, Bike, Bus, Car, Rail	82.00%	Accuracy	n.p.	Netherlands, [43]
Dabiri et al. [32]	5	Walk, Bike, Bus, Drive, Train	76.4%	F1-Score weighted average	Manifold-cross-validation	Beijing
Jiang et al. [57]	4	Walk, Bike, Bus, Car	98.00%	Accuracy	Hold-out	Beijing
Assemi et al. [11]	4	Walk, Bike, Bus, Car	94.7%	Accuracy	Hold-out	New-Zealand
Yazdizadeh et al. [123]	4	Walk, Bike, Transit, Car	91.8%	Accuracy	Manifold-cross-validation	Montreal
Mäenpää et al. [72]	4	Walk, Bike, Bus, Car	90.7%	F1-Score	Manifold-cross-validation, Out-of-bag-estimate	Beijing. 1 week BUS trajectories, 1000 trajectories from Open Street Map (OSM)
Yazdizadeh et al. [124]	4	Walk, Bike, Transit, Car	83.4%	Accuracy	Manifold-cross-validation	Montreal

in this context. Tables 4, 5 and 6 present purpose imputation; Tables 7, 8 and 9 map-matching methods.

3.1 Smartphone data mining

Due to the disparity of progress drivers, we see a trend of increasing fragmentation, inconsistencies, availability, and volume of travel data. In response to this challenge, two main branches seem to arise as flip sides of the same coin [39, 58, 64, 98]. The first focuses towards data fusion, intended to compose and then mine high dimensional datasets collected from multiple sources, including GIS, INS, and GPS. The second targets the development of, for example, very sophisticated computer intelligence models, feature extraction methodologies, and optimal hyper parameters selection. These are constantly improving and therefore complementing traditional statistical methodologies, often substituting them for specific purposes [59].

Literature has shown that smartphone data is affected by several errors. For example, map-matching observations based on positions generated by a Nokia N95 would be much less reliable than those based on a dedicated GPS logger [19]. With current smartphones, however, the situation has improved substantially. For mode detection, neural network classifiers have shown higher performance on data collected from smartphones than from GPS devices [22]. Nevertheless, we should be aware that raw sensor measurements may vary between smartphones, as well as within the same model of smartphone [20]. Any measurement is affected by noise that is not necessarily random, since it may be correlated with: weather conditions; building density, materials, and height; crowdedness; physical placement of the smartphone (e.g. in the pocket is different than on a table); smartphone model; and software “bugs.” Therefore, achieving consistency of machine-learning methods across different smartphones requires a rigorous

Table 2 Dataset ranked by number of users, for mode detection

References	Person-day	Users	Ground truth	Observations	Time	Area	Smartphone App
Semanjski et al. [96]	24,900	8303	Validated-by-respondents	30,000 trips 3,960,243 GPS points 340,000 km	n.p.	Leuven	Routecoach
Yazdizadeh et al. [124]	88,630	6846	Validated-by-respondents (102,904 trips)	623,718 trips	2 months collection period	Montreal	MTL Traject App
Yazdizadeh et al. [123]	88,630	6846	Validated-by-respondents (P2D)	102,904 trips	2 months collection period	Montreal	MTL Traject App
Yazdizadeh et al. [122]	88,630	6846	Validated-by-respondents (P2D)	131,777 trips 33 mln GPS points	2 months collection period	Montreal	MTL Traject App
Bohte and Maat [21]	40,208	1104	Validated-by-respondents (P2D)	n.p.	7395 days	Netherlands	GPS logger and Web based validation
Thomas et al. [104]	n.p.	600	Validated-by-respondents	60,000 trips	3 batches per 1 month each	Netherlands, [43]	Move smarter
Xiao et al. [119]	1248	202	Validated-by-respondents	4685 Trip-legs	n.p.	Shanghai	Shangai City—Smartphone based travel survey
Dabiri et al. [32]	4000	189	Partially validated-by-respondents (69 respondents)	17,621 trajectories 1,292,951 km 50,176 h	3 years collection period	Beijing	Geolife [129]
Rasmussen et al. [90]	644	101	Validated-by-respondents (P2P)	6,419,441 GPS points 1783 h of travel	3–5 days per respondent	Copenhagen	GPS logger
Assemi et al. [11]	372	76	Validated-by-respondents	760,000 GPS observations, 530 h trajectories	2 months per respondent	New-Zealand	Advanced Travel Logging Application for Smartphones II (ATLAS II)
Mäenpää et al. [72]	4000	> 69	Validated-by-respondents	n.p.	n.p.	Beijing. 1 week BUS trajectories, 1000 trajectories from Open Street Map (OSM)	Geolife [129], Journeys API ^a , OpenStreetMap ^b
Dabiri and Heaslip [31]	4000	69	Validated-by-respondents	n.p.	3 years collection period	Beijing	Geolife [129]
Jiang et al. [57]	4000	69	Validated-by-respondents	n.p.	3 years collection period	Beijing	Geolife [129]
Zhou et al. [131]	4000	69	Validated-by-respondents	n.p.	3 years collection period	Beijing	Geolife [129]
Zhu et al. [134]	4000	69	Validated-by-respondents	n.p.	3 years collection period	Beijing	Geolife [129]
Zhou et al. [132]	n.p.	12	Validated-by-respondents	n.p.	6 days per respondent	Georgia (USA)	Self Developed App
Martin et al. [75]	n.p.	6	Validated-by-respondents	347,719 GPS points in 96.59 h (1 Hz) 1.7 mln points Acceleration in 98.62 h (5 Hz)	n.p.	Minnesota	Self Developed App
Byon and Liang [22]	n.p.	n.p.	n.p.	n.p.	50 h	Toronto	Self Developed App
Jahangiri and Rakha [55]	n.p.	n.p.	Validated-by-respondents	n.p.	n.p.	Tennessee	Self Developed App

^a Journeys API, retrieved from web 01/01/2019, http://wiki.itsfactory.fi/index.php/Journeys_API

^b Open-source Trajectories , retrieved from web 01/01/2019, <https://www.openstreetmap.org/traces>

process of data preparation, cleansing, and trajectory segmentation up front. We describe these processes in the next sections.

For each classifier, such as for mode detection and purpose imputation, the underlying features can be (i) location-agnostic versus location-specific; and (i) user-agnostic versus user-specific. For example, methods

Table 3 Methodology and features, for mode detection

References	Method	Main features	AGPS	INS	GIS
Assemi et al. [11]	Nested logit model, multinomial logistic regression, multiple discriminant analysis	Skewness of speed distribution, share of travel time with speed (m/s) $\in [2, 8]$, share of travel time with speed (m/s) $\in [8, 15]$, maximum speed, 95% percentile acceleration, maximum acceleration, acceleration variance, direct distance <i>origin</i> \rightarrow <i>destination</i> , travelled distance <i>origin</i> \rightarrow <i>destination</i>	Yes	No	No
Bohte and Maat [21]	Rule-based	Distance <i>GPS</i> \rightarrow Points-of-interest, Distance <i>GPS</i> \rightarrow <i>LandUse</i>	Yes	No	Yes
Byon and Liang [22]	Neural network	Speed, acceleration, magnetic field, satellites number	GPS	Accelerometer magnetometer	No
Dabiri and Heaslip [31]	Convolutional neural network, random forest, key nearest neighbor, support vector machines, multi layer perceptron	Speed, acceleration, jerk, bearing rate	Yes	No	No
Dabiri et al. [32]	Semi-Supervised Convolutional Autoencoder	GPS points: relative distance, speed	Yes	No	No
Jahangiri and Rakha [55]	Random forest, bagging model, support vector machines, key nearest neighbor, Max-dependency Min-redundancy	Acceleration spectral entropy, acceleration range, Max angular velocity, average absolute acceleration, average angular velocity	Yes	Accelerometer scope, rotation vector	No
Jiang et al. [57]	Recurrent neural network, Hampel filter	Speed, average speed, standard deviation speed	Yes	No	No
Mäenpää et al. [72]	Bayesian classifier, neural network, random forest, auto encoder	Maximum acceleration, maximum speed, minimum acceleration, minimum speed, average acceleration, average speed, acceleration variance, speed variance, speed skewness, speed kurtosis, acceleration skewness, acceleration kurtosis	Yes	No	No
Martin et al. [75]	Random forest, key nearest neighbor, principal component analysis, recursive feature elimination	Average change in acceleration ($\Delta T = 120$ s), 80% percentile speed ($\Delta T = 120$ s), variance change in acceleration ($\Delta T = 120$ s), maximum speed ($\Delta T = 120$ s), average speed ($\Delta T = 120$ s), average change in speed ($\Delta T = 120$ s)	Yes	Accelerometer	No
Rasmussen et al. [90]	Fuzzy logic	95% percentile acceleration, 95% percentile speed, median speed, network segment	GPS	No	Yes
Semanjski et al. [96]	Support vector machines	Distance from (DF) motorway, DF railway, DF bicycle lane, DF bus stop, DF railways station, DF car parking, DF bicycle parking, DF bus line	Yes	No	Yes
Thomas et al. [104]	Bayesian classifier	Personal trip history, speed, altitude, longitude, latitude, public transport time-table	Yes	Accelerometer	Yes
Xiao et al. [119]	Bayesian network	Average speed, 95% percentile speed, average absolute acceleration, travel distance, average heading change, Low-speed-rate (as the ratio of points with speed < threshold)	Yes	No	No
Yazdizadeh et al. [123]	Counvolutional neural network augmented with ensemble method, with random forest as meta learner	GPS points: relative distance, speed	Yes	No	No

Table 3 (continued)

References	Method	Main features	AGPS	INS	GIS
Yazdizadeh et al. [122]	Random forest	Measures between origin-destination: cumulative and direct distance (m), travel time (Min.), average and 85th percentile speed (km/h), maximum, minimum difference between Min. and Max. acceleration (km/h ²), minimum and maximum slope; Max time interval (min) and Max distance (m) between each consecutive pair of GPS point; time of day and time of week; age, gender, occupation; average value of residential buildings around each individual's home (in 250 m radius); direct distance between the origin and nearest public transit stop; direct distance between the destination and nearest public transit stop; average value of residential buildings around each individual's home (in 250 m radius)	Yes	No	Yes
Yazdizadeh et al. [124]	Semi-supervised Generative Adversarial Networks	GPS points: relative distance, speed	Yes	No	No
Zhou et al. [132]	Random forest with 3 layers	Speed, Acceleration – Gravity, fast fourier transform (frequency domain), energy of the signals, sum of spectral coefficients	Yes	Accelerometer	No
Zhou et al. [131]	Random forest	85% percentile speed, average speed, median speed, medium velocity rate, high velocity rate, low velocity rate, travel distance	Yes	No	Yes
Zhu et al. [134]	Auto encoder, deep neural network	Average speed, travel distance, average acceleration, head direction change, bus stop closeness, subway line closeness	Yes	No	Yes

Table 4 Classification task ranked by difficulty and score, for purpose imputation

References	No.	Classes	Score	Metric	Validation
Kim et al. [60]	15	Work, Study, Shopping, Social Visit, Recreation, Home, Business Meeting, Change mode/Transfer, Pick-up, Drop-off, Meal/Eating break, Personal Errand/Task, Medical/Dental, Entertainment, Sport/Exercise	98.68%	F1-Score	Out-of-bag-estimate
Feng and Timmermans [40]	10	Study, Social Visit, Recreation, Home, Service, Paid Work, Daily Shopping, Non-daily Shopping, Help parents/cildren, Voluntary work	96.8%	Accuracy	Out-of-bag-estimate
Montini et al. [76]	9	Work, Shop, Service, Recreation, Home, Pick-up, Drop-off, Business Meeting, Other	79.8%	Accuracy	Out-of-bag-estimate
Xiao et al. [120]	8	Work, Study, Shop, Social Visit, Home, Eeating Out, Pick-up, Drop-off	96.53%	Accuracy	Hold-out
Bohte and Maat [21]	7	Work, Study, Shop, Social Visit, Recreation, Home, Other	43%	Accuracy	n.p.
Yazdizadeh et al. [122]	6	Education, Health, Leisure, Shopping/Errands, Home, Work	72%	F1-Score weighted average	Manifold-cross-validation

relying on user- or location-agnostic features can be trained on any geographic area, and then either deployed on a different area to classify the activities of another population or reused to solve similar problems. The former depends on the generalization power of the model, while the latter is identified as transfer learning. Transfer learning is the discipline dedicated to using the knowledge gained by solving a problem in one domain (e.g. stop detection) to solve a different problem in another domain (e.g. mode and purpose classification). From our standpoint, these approaches could contribute in mitigating the cold-start problem [100], for example in the process of switching from a traditional to a smartphone-based travel survey.

The literature reviewed often works with location and user-agnostic features. In contrast, user- [60, 126] and location-specific [96] data seem to enable more accurate classifications. Although results presented in the relevant literature are hardly comparable across studies, within each relevant study we find evidence about the positive contribution of user- and location-specific data on the performance of the classifiers [104]. The cost is the volume of information to be handled, poor transferability and poor generalization power. From this angle, we challenge the conclusions of [62]: Transferability and generalization power may also be related to the supporting dataset, and not only to the machine-learning method.

3.2 Data cleansing

While performing data cleansing, data analysts should check whether basic features such as speed and acceleration are consistent with the context. The data cleansing purpose is to find and remove outliers, fill observation gaps, and possibly smooth the trajectories [1]. This crucial step should begin performing a sanity check on the

observations' timestamps. Common issues are multiple observations with the same timestamp, or discrepancies due to implicit time localization that keeps no trace, e.g., of periodical solar and legal time shifts. The first case can be mitigated using fine grained timestamps during data collection, such as milliseconds or microseconds; the second, using standard date representations such as the ISO 8601. Further, sensors trajectories are often stored inconsistently on database, e.g., due to smartphones temporary lack of internet connection. Therefore, to find "correct outliers", any basic feature—such as speed, space, and time variation between consecutive pairs of observations—should be computed after sorting these trajectories by timestamp. Once the basic features are available, to handle outliers there are different degrees of sophistication between rule-based, statistical, and model-based filters, such as threshold-, median-, and Kalman-filter. The measurements' sampling rate is a critical factor determining the filter choice. In general, the trade-off is between scalability and accuracy, with rule-based filters on the one hand, and more sophisticated tools like the Kalman-filters on the other. If the number of outliers is very high, such that removing these outliers we create unacceptable gaps in the trajectories, data analysts can resort to one of the several data imputation techniques available [108], such as an exponential weighted moving average.

To reduce the risk of noisy labels that could bias supervised classifiers already in the training phase, data cleansing should focus on labels too. Often labels come as a separate trajectory, which should have a common timeline with the sensors' observations. We are aware that during the validation users may overlook errors present on travel diaries. We cannot exclude human-computer interaction problems facilitating

Table 5 Dataset ranked by number of users, for purpose imputation

References	Person-day	Users	Ground truth	Observations	Time	Area	Smart-phone App
Yazdizadeh et al. [122]	88,629	6845	Validated-by-respondents (P2D)	131,777 trips, 33 mln GPS points	1 month collection period	Montreal	MTL Traject App
Bohte and Maat [21]	40,208	1104	Validated-by-respondents	n.p.	7395 days	Netherlands	GPS logger and Web based validation
Kim et al. [60]	7856	793	Validated-by-respondents (P2D)	22,170 days, 130 mln GPS points	5–14 days per respondent	Singapore	Future Mobility Survey
Feng and Timmermans [40]	n.p.	329	Validated-by-respondents (P2D)	10,545 activities	3 month per respondent	Netherlands (Rotterdam)	GPS logger and Web based validation
Xiao et al. [120]	2409	321	Validated-by-respondents (P2P)	7039 trips	7–12 days per respondent	Shanghai	Shangai City - Smart-phone Based Travel Survey
Montini et al. [76]	n.p.	156	Validated-by-respondents	6938 activities	7 days	Zurich	Self Developed App

Table 6 Methodlogy and features, for purpose imputation

References	Method	Main features	AGPS	INS	GIS
Bohte and Maat [21]	Rule-based	Distance <i>GPS</i> → Points-of-interest, Distance <i>GPS</i> → <i>LandUse</i>	GPS	No	Yes
Feng and Timmermans [40]	Random forest	Activity duration, activity start time, travel time to activity, distance <i>GPS</i> → Points-of-interest	GPS	No	Yes
Kim et al. [60]	Bagging decision tree, random forest	Activity probability, distance-based empirical probability, activity transition probability, activity duration	Yes	Accelerometer	Yes
Montini et al. [76]	Clustering, random forest	start time, end time, GPS points density, age, education, income, mobility ownership, activity duration, walk percentage	Yes	Accelerometer	Yes
Xiao et al. [120]	Multi layer perceptron, particle swarm optimisation, multinomial logit, support vector machines, Bayesian network	Age, gender, education, working hours, income, time of week, activity duration, time of day, transportation mode, distance <i>GPS</i> → Points-of-interest, distance <i>GPS</i> → <i>LandUse</i>	Yes	No	Yes
Yazdizadeh et al. [122]	Random forest	Features returned by Open Trip Planner ^a itinerary: GPS tracks average speed, time interval between the first and last GPS track of a trip, average distance between consecutive GPS point, attributes from, itinerary length, total transit time of each returned, total walking time of each itinerary, total waiting time of each itinerary, total travel time, number of transfers, walking distance, itinerary average speed attributes from GPS tracks, difference between GPS tracks length and itinerary length, overlapping percentage of itinerary and GPS tracks	Yes	No	Yes

^a Open Trip Planner (OTP) retrieved from web 01/01/2019, <https://github.com/opentripplanner/OpenTripPlanner>

human errors during the travel diary validation step. Human errors may also occur while extracting data from the database. Rather than outliers, in these cases we should be concerned of flipping-labels [92]. Given a set of labels that a travel survey collects, outlying-labels indicate one or more trajectories labeled with a

class not included in this set; flipping-labels indicate one or more trajectories belonging to one class and labeled with another class, both being present in the set. However, while the impact of both outlying- and flipping-labels on supervised classifiers is extensively studied for independent and identical distributed data [15, 16, 73, 74, 77]—for example on the popular

Table 7 Map-matching task ranked by difficulty and score

References	Mode	Category	Score	Metric	Validation
Chen and Bierlaire [26]	Walk, Bike, Car, Metro	Multimodal, global, shortest-path	[80%, 99%]	Path similarity indicator	n.p.
Torre et al. [106]	Bicycle	Match when possible, build when needed	n.p.	n.p.	n.p.
Quddus et al. [89]	Car	Unimodal, incremental, point-based	99.2%	$A = \frac{\#(\text{correctly matched GPS points})}{\#(\text{Total GPS points})}$	n.p.
Li et al. [68]	Car	Unimodal, incremental, point-based	99.8% (sub-urban), 97.8% (urban)	$A = \frac{\#(\text{correctly matched GPS points})}{\#(\text{Total GPS points})}$	n.p.
Wei et al. [115]	Car	Unimodal, incremental, shortest-path	98%	Accuracy	n.p.
Bierlaire et al. [19]	n.p.	Unimodal, global, shortest-path	[80%, 99%]	Path similarity indicator	n.p.
Wu et al. [116]	Taxi	Unimodal, incremental, point-based	93.58%	Prediction accuracy of next road by the road having the maximum probability	Hold-out
Hunter et al. [52]	Taxi	Unimodal, incremental, shortest-path, supervised, unsupervised	100% (1 s resolution), > 90% (30 s resolution)	Accuracy	Manifold-cross-validation
Li and Wu [67]	Taxi	Unimodal, incremental, point-based	87.18%	$A = \frac{\#(\text{correctly matched GPS points})}{\#(\text{Total GPS points})}$	Hold-out
Jagadeesh and Srikanthan [54]	Dataset 1: Taxi. Dataset 2: n.p.	Unimodal, global, shortest-path	91.3%	Average F-Score with: $Precision = \frac{Length_{correct}}{Length_{matched}}$, $Recall = \frac{Length_{correct}}{Length_{truth}}$, Input-to-output latency (Timelines)	Hold-out
Newson and Krumm [78]	Car	Unimodal, incremental, point-based	100% (1 s resolution), > 90% (30 s resolution)	$Accuracy = 1 - E_L$, where $E_L = \frac{(d_- + d_+)}{(d_0)}$, d_- = erroneous subtracted length, d_+ = erroneous added length, d_0 = length of correct route	Hold-out
Lou et al. [71]	n.p.	Unimodal, global, shortest-path	$A_N > 81\%$, $A_L > 87\%$	$A_N = \frac{\#(\text{correctly matched road segments})}{\#(\text{all road segments of the trajectory})}$ $A_L = \frac{(\sum \text{length of matched road segments})}{(\text{length of the trajectory})}$	Hold-out

handwritten digits dataset from the Modified National Institute of Standards and Technology database—we found no literature focusing on time series, as for example GPS.

3.3 Trajectory stop detection

The analysis of human trajectories can be reduced to two fundamental classes: **motion**, and **stop**. Tables 1 and 4 present how each class branches out. Tables 3 and 6 specify both features and methods enabling accurate classifications. Tables 2 and 5 present the dataset that enabled each study we reviewed. To perform any specialized inference on trip legs we need to identify homogeneous segments and relevant discontinuities from heterogeneous and complex mode-chain-types.

A GPS segment is considered a stop candidate if it lays within a topologically closed polygon for a certain

time [4, 128, 133]. The presence of GPS points nearby may be indicative of a stop—the absence of motion [107]. Rules to acquire a local density of points, for example, include a moving window linking 30 preceding and 30 succeeding points within a 15 m range [94]. Although compatible with the error amplitude of GPS devices declared in a survey by [37], this range seems too small compared to smartphone AGPS expected error [109]. Smartphones location output does not rely exclusively on GPS, but also on less accurate methods that fill GPS gaps. Zhao et al. [126], for example, extend the range to 45 m.

Based on the assumption that noise detected in transition points is temporary while the changes in speed are permanent, affinity propagation clustering methods can be effective in stop detection [133]. By building a network that links stationary events, identified as nodes within a critical space-time range, and clustering this network

Table 8 Dataset ranked by number of links and users, for Map-matching

References	Links	Users	Ground truth (GT)	Observations	Area	Device
Quddus et al. [89]	4605	n.p.	24-channel dual-frequency geodetic receiver	4 h trajectories, 1 s resolution	London, sub-urban areas	GPS logger, gyroscope, odometer
Li and Wu [67]	583	12,000	Hand match supported by Rule Based Algorithm	Training-set: 8678 GPS points (traces + synthetic from GIS), Test-set: 1334 GPS points (traces only), 10 s resolution	Beijing, urban areas	GPS logger
Wu et al. [116]	n.p.	442 + 13,650	No GT available. Hidden Markov Models map-matching results as benchmark with [78]	859,195 Traces, 3,709,666 Traces	Porto, Shangai	GPS logger
Lou et al. [71]	n.p.	189	Validated-by-respondents (69 users only)	Dataset 1: Synthetic generated from road network (error normally distributed 20 stdev, 0 mean). Dataset 2: 28 GPS Traces (Trips)	Beijing	Geolife, [130]
Chen and Bierlaire [26]	n.p.	180	No GT available. Unimodal map-matching result as benchmark	10 s resolution	Lausanne (CH) Urban and outskirts areas	Nokia EPFL Lausanne [61]
Hunter et al. [52]	n.p.	Dataset 1: 10. Dataset 2: 600	Dataset 1: 1 s resolution GPS considered as high accuracy GT. Dataset 2: no GT	Dataset 1: 700,000 GPS points, 1 s resolution. Dataset 2: 600,000 points, 1 min resolution	S. Francisco	Mobile Millennium system—GPS logger
Jagadeesh and Srikanthan [54]	n.p.	Dataset 1: 21,807 GPS points, 20 trips, 421 km. Dataset 2: 1000 trips, 13,139 km	Dataset 1: Manual Check on Map-matched GPS points from higher accuracy source (smartphone), leveraging on knowledge of taxi route. Dataset 2: User validation	Dataset 1: 21,807 GPS points, 20 trips (TAXI), 421 km, 1 s resolution. Dataset 2: 13,139 km, 1000 trips. Dataset 3: Synthetic Dataset adding noise to Dataset 1	Singapore	Dataset 1: Custom Smartphone App (Android), Dataset 2: Commercial Smartphone App
Newson and Krumm [78]	n.p.	1	Route planned before data collection and hand match	7531 GPS points, 80 km, 1 s resolution, degraded data simulation	Seattle	GPS logger
Bierlaire et al. [19]	n.p.	Dataset 1: 1 users. Dataset 2: 3 users.	Dataset 1: Known true path. Dataset 2: no ground truth. Dataset 3: high accuracy GPS device	Dataset 1: 10 points Dataset 2: 25 trips 1041 GPS points, 10 s resolution	Lausanne (CH), Urban and outskirts	Nokia EPFL Lausanne [61]
Li et al. [68]	n.p.	n.p.	Tightly-coupled carrier phase GPS receivers integrated with a high-grade inertial navigation system	3363 epochs (sub-urban), 2399 epochs (urban), resolution: 1 epoch/s	Nottingham rural sub-urban, Central London	GPS logger, digital elevation model
Torre et al. [106]	n.p.	n.p.	n.p.	128 GPS Traces, 185,000 GPS points, 360 km, 1088 min	Minneapolis (Twin Cities)	Cyclopath Android App
Wei et al. [115]	n.p.	n.p.	n.p.	14,436 GPS points (SIGSPATIAL Cup 2012 DS), 19,080 GPS points, 1 s resolution	Seattle Shanghai	

using two-level Infomap [93], a swift algorithm, available as python package [9], outputs a label for each stop event detected in a raw GPS trajectory.

Literature shows many developments in this direction, employing clustering techniques [46, 105, 117, 130], which can learn in an unsupervised fashion and find stops within GPS trajectories. In multiple-step approaches, personal- [46], and geographical-context [130] can augment trajectories' information and improve the classification of stop candidates. Density-based spatial clustering of applications with noise (DBSCAN) is at the base of most frameworks; some of these frameworks can even find stop candidates directly on raster image representations [111]. Many other effective probabilistic unsupervised methods are available, as for example kernel-based [48, 103]; generative [81, 118]; and discriminative [70], such as kernel-density algorithms, Hidden Markov Models, and conditional random fields.

Assuming that travelers walk to change mode, a rule-based algorithm can identify transition points by applying thresholds on speed, acceleration, range and time, as well as by checking GPS on-off status [90]. In fact, the most common rule-based stop detection techniques rely on range, time, speed or acceleration thresholds [98].

These rule-based algorithms can be further improved by statistical tests. For example, a Kolmogorov–Smirnov test on a random sample can be used to check for outliers [131], as the normal distribution is sometimes accepted as a suitable approximation for GPS. Assuming normal distribution of GPS error, though, GPS follows a bi-variate Raleigh distribution [19].

Rule-based algorithms are both effective and appropriate, and are independent of the subsequent classification task, as for example mode detection, or purpose imputation. However, thresholds inflexibility (for example, in handling GPS signal loss and signal noise) leads to poor performance in detecting short stops (such as alighting from a bus) and long permanence in the same position (such as sitting on the bus during and intermediate stop) [98].

3.4 Trajectory segmentation

Another approach specialized in “mode detection” is a GPS trajectory preparation through segmentation, which goes through four steps [31]. The first step splits the trajectory in fixed segments having the same size of the median number of points on all the available trips. The second step concatenates together consecutive segments with the same label. Let us note that the first two steps depend strictly on the availability of the ground truth, while the segment size depends on the data collection context. The third step discards segments with less than

10 GPS points. The fourth step smooths the trajectory through a Savitzky–Golay filter.

Segmentation methods can be distance-, time-, bearing- and window-based. While the last three are statistically equivalent, the first leads to varying sample sizes within each segment due to the different speeds in complex mode-chain-types. Discontinuities in the mode-chain-type, detected on these segments, represent stops [57].

The impact of stop-detection or trip segmentation on the quality of the travel diary generation process, and therefore on the quality of the ground truth collected from users that validate their trips, can be considerable [102]. Therefore, more advanced hybrid methods have been studied, as have multiple rules and machine-learning specializing in both trajectories and contexts. One hybrid method consists of the following six steps [126]: The first step is trajectory cleansing, based on the accuracy provided by the AGPS; the second step is rule-based detection of stop candidates, where stops are points within a 50-m range and a 1-min time window. The third step checks for stop candidates against users' frequent stop locations. The fourth step merges the resulting stops, with a rule-based algorithm configured with various range and time thresholds. The fifth step detects “still” mode, with a learned classifier based on acceleration. The sixth step removes, after mode detection, any orphan stop left.

3.5 Towards a standardized measurement of performance

All of the aforementioned methods are very critical for the classification steps downstream in the process, and they all lack of flexibility in adapting to different thresholds, which might depend on some users, context, or both. However, the choice of trip segmentation method determines the object to be classified in the next step of the process, which can be a single observation, such as a GPS point, or a set of observations, such as a GPS segment. Consequently, two methods presenting the same classification score might be very different, depending on whether these methods target points or segments. It is very unlikely that the same number of points and segments will identify two analogous trips in terms of space and time. Therefore, comparing the performance score between point- and segment-based methods is misleading. The scores presented in Tables 1, 4 and 7 are not comparable, nor harmonized. Since scores and respective results reflect the case of correct classifications related, e.g., to a stage, a trip, an excursion or the whole day, harmonization attempts should take these cases explicitly into account.

Prelipcean et al. [86] introduce penalty systems and metrics that look at where these methods lead to errors,

Table 9 Methodology and main features, for Map-matching

References	Method	Main features	AGPS	INS	GIS
Quddus et al. [89]	Fuzzy logic, extended Kalaman filter	Speed, heading error, perpendicular distance, horizontal dilution of precision	1.2-channel single frequency high sensitivity GPS receiver	Dead-reckoning	Yes
Li et al. [68]	Rule based, extended Kalaman filter, integrity check	Altitude, longitude, latitude, traffic flow directions, road curvature, grade separation, travel distance, heading	GPS	Dead-reckoning	Yes
Bierlaire et al. [19]	Probabilistic	Timestamp, longitude, latitude, speed, heading, horizontal error Std. Dev., network error Std. Dev.	Yes	No	Yes
Li and Wu [67]	Feed forward neural network	Longitude, latitude, timestamp, heading	GPS	No	Yes
Lou et al. [71]	Mixed method: topological, geometric, probabilistic	Distance GPS(t) → GPS(t + 1), distance GPS → network, shortest path between candidate points on network, average speed	Yes	No	Yes
Torre et al. [106]	Hidden Markov model, Viterbi	Distance GPS → node, maximum out-degree of the transportation graph	Yes	No	Cyclo-path map
Wei et al. [115]	Global Max-weight, hidden Markov model, Viterbi	Fréchet distance, shortest-path	GPS	No	Open Street Map
Chen and Bierlaire [26]	Probabilistic	Transport mode, distance, speed, acceleration	Yes	Accelerometer, Bluetooth Low Energy	Yes
Wu et al. [116]	Recurrent neural network, long short term memory	Longitude, latitude, timestamp, destination	GPS	No	Open Street Map
Newson and Krumm [78]	Hidden Markov model, Viterbi	Distance GPS(t) → GPS(t + 1), distance GPS(t) → network (only in range < 200 m)	Yes	No	Yes
Hunter et al. [52]	Undirected graph Bayesian network, Viterbi	Path length, distance point projection → GPS, number of signals, number of turns, average speed, Max/Min Num. lanes	GPS	No	560,000 links map
Jagadeesh and Srikanthan [54]	DS 1: hidden Markov model, Viterbi, conditional random fields (CRF), DS 2: multinomial logit model, k-shortest path with link-penalty approach	Path choice: free-flow travel time (s), number of traffic signals, average road class, number of class changes	AGPS, with WiFi and GPS off	No	Yes

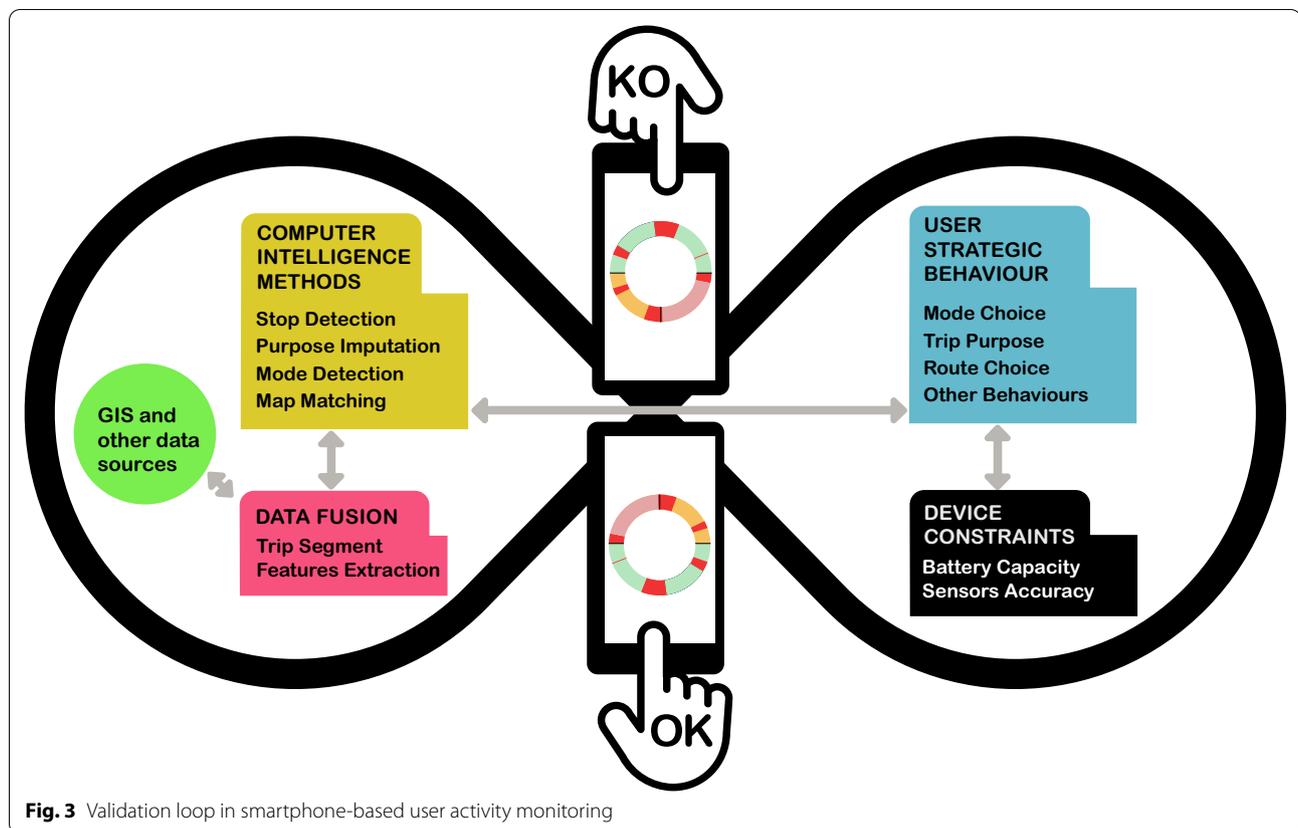


Fig. 3 Validation loop in smartphone-based user activity monitoring

and provide meaning to the comparison among different segmentation techniques. In particular, with respect to the ground truth, if precision and recall identify “hits” and “misses” of a classifier (the broadly used F1-Score is the harmonic mean of precision and recall) from such measurements, we do not understand how the error depends on over- or under-segmentation, e.g., of the trajectory that this method classified. Since errors in trajectory segmentation propagate to the classification of the trajectories, and classification performance depends on how the segmentation inference aligns with the ground truth, these penalties are proportional to time and space of segments misaligned with the ground truth. This is in opposition to previous studies where a count of the editing operations was proposed [2]. Interestingly, with this metric, point-based trajectory segmentation techniques seem to outperform segment-based techniques [86]. Since both segment- and point-based classifiers discard any segment below a certain threshold of (e.g.) GPS observations—which in the first case can be two magnitudes higher than in the second case—an intuitive explanation is that segment-based classifiers are incapable of classifying a larger fraction of a dataset.

3.6 Human activity recognition in mobility

To support the modeling of activity and travel choices at the heart, for example, of activity-based models [110], human activity recognition in mobility must include both stop, mode and purpose of any trip. The combination of feature extraction techniques and computer intelligence algorithms allows for a capturing of the correlation between features and the user’s strategic choices. As technology evolves, the inference of users’ strategic choices in the form of a travel-diary and user validation by means of such a diary (see Fig. 3), enable continuous improvement of the acceptable truth asymptotically approaching the theoretical ground truth. Computer intelligence algorithms are tightly coupled with the data necessary to allow and refine the inferences. Given an initial validated dataset, their performance can be measured only by comparing inferences with the ground truth (see Fig. 3). Errors propagate from trajectory segmentation, to trajectory classification, and then to the travel-diary generation [86]. Therefore, it is likely that errors propagate to the ground truth. From this standpoint, the output of this process might lead to systematically biased predictions. In SBTS, machine-learning is just a tool used to capture the information represented by data. The quality of models has a strong influence on the quality of the

ground truth we can collect through travel-diaries, and vice-versa.

There is consensus in the field about the lack of standardization for validating and comparing competing classifiers. There are several studies where, even though classifications are performed on the same dataset, differences in number and quality of classes predicted and in validation setup are enough to make F1-Score comparisons meaningless. For example, F1-Scores obtained as average on a 5-label transport mode classification task and a fivefold cross-validation [31], cannot be compared with F1-Scores from a 4-label transport modes classification task, computed on a random test-set only (hold-out method) [57].

We have identified three approaches that allow for a comparison to be made between different methods and datasets. The first is the same aforementioned penalization solution to ease the comparison between point- and segment-based classifiers [86]. The second approach could provide a standardized baseline by combining a public dataset and a cross-validation workflow [112]. The dataset includes the observations of 18 sensors on three users made over a period of 2812 h' worth of labeled data. Labels include the position of the phone as: in the hand, at the torso, at the hip, and in a bag. The workflow for cross-validation covers three tasks: user-independent, phone position-independent, and time-invariant. At the end of the three tasks, each one accomplished with manifold cross-validation, the paper suggests the standard deviation of F1-Scores computed across users, phone positions, and time periods as the benchmark of the predictive power of a model. This workflow cannot be applied in most of the datasets available, which are not as rich; for example, the widely used Geolife [129] provides GPS trajectories and transport mode labels only (see Table 1). The third approach leverages the Weka software [47], where several machine-learning algorithms are available off-the-shelf. Based on Weka software, Ectors et al. [36] compare a few rule-based and probabilistic machine-learning algorithms for purpose imputation on the same dataset.

However, we found no attempts at combining these three approaches, which are complementary to comparing different methods, but not self-sufficient. Another step should consider the feature extraction process. Indeed, this process is also subject to attempts of standardization. One candidate method is "minimum redundancy maximum relevance" [112] (MRMR, see Table 3). For classifiers relying on deep learning though, this feature extraction method is not effective, as the neural network extracts the features autonomously. In this case, the new challenge is finding

optimal hyper parameters for the neural network. Such hyper parameters may include, for example, architecture configuration, activation functions, batch size, regularization factor, and optimization step. Balaprakash et al. [14] propose an approach to selecting these hyper parameters automatically, moving towards standardized deep learning method optimization. Still, we did not find applications in this field; instead, optimal hyper parameters are still a craftsman product [31, 57, 121].

3.7 Implications for transport science

The choice of complementary sensors, such as the gyroscope, could mitigate the challenges that most of the algorithms encounter in discriminating between, for example, bike and walk or bike and bus in congested urban contexts. Similarly, the magnetometer could help distinguish between rails and cars, and the accelerometer between bike and e-bike. However, these high-frequency sensors require online rather than offline classifiers. Offline classifiers would suffer from the large footprint of the data, which would in turn have a negative impact on smartphone users' data plan and battery. This would ultimately lead users to dropout from travel surveys.

Several studies exhibit how useful GIS information can be on mode detection. However, when classifying the complement of the same trajectory, studies on purpose imputation expose the challenges associated with the proximity of heterogeneous points of interest, as various trips can start for different purposes and end in the same spatial range. In such a case generally helpful, personal patterns and a limited amount of personal information proved to support more accurate predictions (see Table 3 against Table 1, and Table 6 against Table 4).

Nevertheless, among the studies identified for map-matching, we find no examples of personal information use (see Table 9). Even in the assumption of unavailability of any personal information, map-matching and consequent route-choice records would amplify the impact of transport mode and trip purpose classification (see Table 7). Expressing a trajectory as a sequence of links and nodes on the transport network, instead of longitude and latitude, pinpoints specific micropatterns. Furthermore, it potentially reduces the confusion that users often face while validating their travel-diaries in the presence of GPS outliers.

For map-matching, we identify two problems. First, most of the methods specialize in cars and road network for cars, and few or none refer to emerging modes such as e-bikes and e-scooters (see Table 8). Second, in the literature, we did not find a good representation of adequate datasets and ground truth quality levels (see Table 9). In the first case, the assumption that GPS points should

belong to the road network does not hold. Map-matching for modes different from cars requires degrees of freedom to allow transit on, for example, sidewalks and bicycle lanes, often not mapped—few studies pinpoint this problem. In contrast, emerging shared modes such as e-bikes and e-scooters imply behaviors not strictly coherent with the mapped network. Furthermore, these emerging modes are introducing new public transport mode-chain-types with irregular patterns, alternating traditional public transport and emerging shared modes. The former offers reliable timetables, while the latter is volatile, as it depends on vehicle availability. Still, Sicotte et al. [99] show that looking at meaningful mode-chain-types also represent a tool to improve trip classification.

From the direct experience testing Mobile Market Monitor and TRAVELVU on a small user base, we realize that the sample of literature reviewed in this work does not express the differences between a raw trajectory, such as the one that SBTS use to generate travel-diaries, and a processed trajectory, such as the one that SBTS may output as ground truth. The first trajectory presents a level of noise that could even ease trip segmentation process and subsequent classification on uni-modal segments. The lack of noise of in the second trajectory, in contrast, might prevent accurate travel-diary generation. These obvious differences have an impact on the choice of method and performance of any transport-related analysis, such as for mode detection. For example, we expect better generalization of Bayesian temporal models or artificial neural network methods in the first case, and machine-learning techniques such as random forest or support vector machines in the second case.

Further, Tables 3, 6, and 9 clearly show that while artificial neural networks and temporal models do not require particular feature extraction methods, machine-learning approaches such as random forest or support vector machines must rely on time-series feature extraction. Hence, to find the best classification method, e.g. for transport mode, any attempt at ranking should be considered in light of whether the trajectories of interest embody any pre-processing, and possibly which one. A possible indicator is the proportion of point loss on the dataset after the application of simple filters, e.g. on point speed and time gaps between points.

For travel-diary generation in presence of multiple sensors and large datasets, artificial neural networks seem very promising. Artificial neural networks are flexible in learning with and without labels. They also act as powerful dimensionality-reduction, information-compression, and feature-extraction tools for simultaneous signal processing of multiple sensors monitoring the same event, and signaling at different and irregular frequencies. Let us consider, for example: (i) smartwatches

and other bio-metric devices complementary to smartphones [29]; (ii) ongoing software integration between cars and smartphones, which include navigation and INS sensors [8]; and (iii) development of edge-computing to augment the processing power of smartphones when consuming cloud services [113], where users' mobility patterns are studied to reduce service-latency in the information-technology-network.

A holistic approach could amplify the impact of studies sharing the scope of those identified in this review. Smartphones' onboard sensors represent only a fraction of the collectible signals, and the surveyed literature seem not fully aware the quickly-evolving context surrounding smartphone devices. To release new potential towards the disambiguation of transport patterns that in congested urban areas look exactly the same for the surveyed methods, while contrasting the curse of dimensionality [17], this field requires a new perspective. Compared to the advances in other fields, such as computer vision or social networks, transport science seems only at the beginning of the exploration of artificial neural networks.

4 Discussion

SBTS depends on a sophisticated multi-sided platform which is subject to often conflicting interests over the resources available, beginning with the battery. In current versions, the OS orchestrates the applications' use of sensors and battery, and some OS preclude direct access to AGPS. Therefore, developers have limited configuration possibilities. Furthermore, the data collected through these platforms is affected by large standard deviation, severe errors, and noise due to exogenous elements.

4.1 Sensors

When a smartphone outputs a location signal, whether the location comes from the onboard AGPS, from the triangulation with GSM antennas, the car GPS, or another external GPS connected to the smartphone, developers are not allowed to know. If not properly handled, this uncertainty may negatively affect datasets, method classification performance, user validation and finally ground truth.

Smartphone onboard sensors represent only a fraction of the bio-metric and ambient sensors that could be connected with these devices. Cornacchia et al. [29] present a survey of activity classification from wearable sensors. Differing effective frequencies of each sensor, e.g., 1–10 Hz for GPS, or > 20 Hz for accelerometer, require flexible frameworks as for joint features extraction, compression, and analysis. From this standpoint, artificial neural networks seem to have potential.

4.2 Data sources

From the perspective of smartphone-related trajectories, a better understanding of travel behavior requires the standardization of measures relevant for travel patterns, which should also rely on standard datasets. The options available are a good starting point, but still seem insufficient. For example, let us consider the following datasets. (i) Shankari et al. [97] deliver real GPS trajectories collected in the USA from real smartphones, in which ground truth, available on trip mode and not trip purpose, is generated synthetically to protect privacy exposure (users follow instructions provided by a custom App). (ii) Wang et al. [112] offer trajectories collected in the UK from multiple smartphone sensors at relevant frequencies, and from smartphones of the same model positioned on various part of the body, providing ground truth for trip mode only. (iii) Zheng and Fu [129] include GPS trajectories from China, with ground truth on trip mode for 69 users out of 189. (iv) Kubicka et al. [63] supply GPS trajectories collected in various parts of the world for map-matching, but not multi-modal. (v) Carpineti et al. [24] propose onboard high-frequency sensors with ground truth on transport mode, collected in Italy from multiple smartphones and users, but where GPS is unavailable. (vi) Chavarriaga et al. [25] provide data from over 72 wearable sensors, collected indoors with ground truth on performed activities, and no GPS. (vii) Laurila et al. [65] offer data collected in Switzerland over 18 months from 185 users of the Nokia N95 device with multiple sensors, including, for example, AGPS, accelerometer, Bluetooth, trip purpose labels, and no transport modes.

4.3 Methods

The collection of any acceptable ground truth depends on the reliability and accuracy of underlying measurement methods. The vast choice of alternatives requires a standardized way of comparing competing methods. Existing literature offers effective penalization systems for classic performance scores [87]. Invitations on standardized mode detection are available in form of feature extraction and cross-validation workflows [112]. However, these attempts do not seem sufficient to cover mode detection, purpose imputation, and map-matching at the same time across existing and emerging methodologies.

We identified excellent alternatives. Some perform best on low-resolution trajectories. Other classifiers are tight (e.g.) to the location where GPS trajectories are fused with data from GIS, users' personal information, or both. Among the best performers in terms of accuracy measurement, in general, we find: support vector machines, fuzzy logic, random forests, and probabilistic models (e.g., hidden Markov models). Classic rule-based

algorithms might not perform at the same accuracy level. However, they are still competitive when the application scenario is stable, and if execution speed and scalability are a priority over accuracy.

Methods based on artificial neural networks are rising quickly and are applicable across mode detection, purpose imputation, and map-matching, as probabilistic and Bayesian methods unlike other machine-learning techniques. For map-matching and purpose imputation, for example, we find applications combining GPS and GIS, while for stop and mode detection, we find applications with GPS only. Particular configurations of these methods, such as variational auto encoders and deep kalman filters, which represent the convergence with Bayesian methods, could offer a background facilitating methodological convergence that might also allow for a breakthrough in this mature field of research.

4.4 Ground truth

Whether a study targets, for instance, the whole day, week, month, season or year, modelers need a correct dataset ideally of a whole period. If this is not the case, the value of the whole dataset is limited. Since a "person to device" validation might introduce further errors; their magnitude and their impact on machine-learning methods performance should be investigated. We find no attempt of self-learning on multi-sensor datasets, which would raise expectations on a "device-to-device" ground truth evolution. We could achieve full automation of both travel-diary generation and validation by using independent measurements of the same event to substitute traditional labels with pseudo-labels. For example, instead of learning from labels, artificial neural networks could learn GPS patterns to reconstruct accelerometer patterns, and vice-versa. Meanwhile, where machine-learning algorithms do not provide correct travel-diaries to the user, "person to device" interaction could be enhanced by introducing the possibility for the user: (i) to trigger a specialized automatic evaluation of such segments; and (ii) to flag whether he or she was unable to correct the mistakes (see Fig. 3).

5 Conclusion

In transport science, the process of methodological perfection between paper-and-pencil personal interviews, and computer assisted personal interviews [13], towards computer assisted telephone interviews [79], and computer assisted web interviews [135] is still evolving towards SBTS [101, 127]. The leap between paper and computer determined a structural impact on the surveying costs, requiring software, IT-infrastructure, and personnel-training. According to [27], the shift to computer assisted web interviews requires to fall back to

telephone interviews in cases where the web interviews are incomplete.

From computer to smartphones, the impact seems negligible both on software and IT-infrastructure costs. In contrast, the impact on human resources seems to determine a significant reduction of personnel, and a shift towards highly specialized and more expensive skills of data scientists required to deploy a SBTS. Consequently, under a certain volume-threshold of, e.g., surveyed users in time, traditional surveys could be still competitive in terms of cost. However, to push transport science boundaries under the constraint of Big Data—which traditional travel surveys are unable to satisfy—SBTS bring a huge scalability potential and support higher resolution datasets, handling users during time horizons longer than just one day.

To expose SBTS potential, this paper selects and summarizes information on SBTS relevant for a qualitative comparison of the methods focusing on mode detection, purpose imputation, and map-matching. To ease such a comparison, since the standardization process in the field is still ongoing, we organized the literature into tables, which include information about classification objectives, datasets employed in the experiments, and validation approach of both data and experiments. Besides, by listing sensors, features, and dataset that each of the related works depends on, we identify the main methods underlying the process of ground truth generation.

Comparison based only on scores reflecting different variables, such as accuracy and F-Score, is misleading. As we find, scores depend on the underlying dataset, trajectory segmentation, classification method and experiment design. Evaluation of larger segment units leads to discarding significant portions of a dataset. The classification task is relatively more difficult with a larger number of classes. The accuracy bias is relatively lower when performing cross-validation, and when processing more representative datasets. For example, Tables 1 and 2 for mode detection, Tables 4 and 5 for purpose imputation, as well as Tables 7 and 8 for map-matching expose, from another perspective than Prelipcean et al. [86], that methods performance is beyond dry scores. When comparing methods, newcomers in this field would certainly benefit from considering task complexity, representativeness of the supporting dataset, and validation method. For example, task and method complexity, features collection and extraction cost (see Tables 3, 6, 9).

A converging thrust in the field seems represented by simultaneous methods focusing on, e.g., mode detection to improve map-matching or purpose imputation, and vice-versa. To support the disambiguation of travel patterns that are still challenging to detect in congested urban areas, for the future, emerging applications of

artificial neural networks seem to support further fruitful convergence. The study of smartphones onboard sensors in addition to other streams collectible through smartphones—from GIS, wearable sensors, or edge-computing—would benefit from the artificial neural networks flexible framework. This technology can be exploited on the one hand to learn from large and heterogeneous data streams, and on the other hand to compress and store such BIG bulk of information through relatively few trained parameters. To support the standardization of relevant measures for transport behavior, efforts should also be directed towards the solution of privacy concerns that represent an obstacle, in this field, for the generation of open-access datasets.

Abbreviations

AGPS: Assisted global positioning systems; CPU: Central processing unit; GIS: Geographic information systems; GPS: Global positioning systems; GPU: Graphical processing unit; INS: Inertial navigation systems; OS: Operation systems; P2D: Person-to-device; P2P: Person-to-person; SBTS: Smartphone-based travel surveys.

Acknowledgements

Not applicable.

Authors' contributions

VS: Conceptualization, methodology, data curation, investigation, formal analysis, writing—original draft. FCP: Conceptualization, validation, supervision. MKA: Validation, editing. OAN: Conceptualization, validation, supervision. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declaration

Competing interests

The authors declare that they have no competing interests.

Received: 27 November 2020 Accepted: 20 October 2021

Published online: 06 November 2021

References

1. Abbruzzo, A., Ferrante, M., & Cantis, S. D. (2021). A pre-processing and network analysis of GPS tracking data. *Spatial Economic Analysis*, 16(2), 217–240.
2. Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832–843.
3. Allström, A., Kristoffersson, I., & Susilo, Y. (2017). Smartphone based based travel diary collection: Experiences from a field trial in Stockholm. *Transportation Research Procedia*, 26, 32–38.
4. 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 *GIS: Proceedings of the ACM international symposium on advances in geographic information systems*.
5. Anderson, P., Hepworth, M., Kelly, B., & Metcalfe, R. (2007). What is Web 2.0? Ideas, technologies and implications for education by. *Technology*, 60(1), 64.
6. Apple. (2016). *Apple developers support resolution on network signal strength access*. Retrieved January 1, 2019, from web.

7. Apple. (2019). *Preventing unexpected shutdowns*. Retrieved January 1, 2020, from web.
8. Apple. (2021). *Car data integration on smartphones*. Retrieved March 17, 2021, from web.
9. Aslak, U. (2019). *Infostop, a Python package for detecting stop locations in mobility data*. Retrieved November 26, 2019, from web.
10. Assemi, B., Jafarzadeh, H., Mesbah, M., & Hickman, M. (2018). Participants' perceptions of smartphone travel surveys. *Transportation Research Part F: Traffic Psychology and Behaviour*, 54, 338–348.
11. Assemi, B., Safi, H., Mesbah, M., & Ferreira, L. (2016). Developing and validating a statistical model for travel mode identification on smartphones. *IEEE Transactions on Intelligent Transportation Systems*, 17(7), 1920–1931.
12. Auld, J., Williams, C., Mohammadian, A., & Nelson, P. (2009). An automated GPS-based prompted recall survey with learning algorithms. *Transportation Letters*, 1, 59–79.
13. Baker, R. P., Bradburn, N. M., & Johnson, R. A. (1995). Computer-assisted personal interviewing: An experimental evaluation of data quality and cost. *Journal of Official Statistics*, 11(4), 413–431.
14. Balaprakash, P., Salim, M., Uram, T. D., Vishwanath, V., & Wild, S. M. (2019). DeepHyper: Asynchronous hyperparameter search for deep neural networks. In *Proceedings—25th IEEE international conference on high performance computing, HiPC 2018* (pp. 42–51).
15. Barandela, R., & Gasca, E. (2000). Decontamination of training samples for supervised pattern recognition methods. In F. J. Ferri, J. M. Iñesta, A. Amin, & P. Pudil (Eds.), *Advances in pattern recognition* (pp. 621–630). Springer.
16. Beigman, E., & Klebanov, B. B. (2009). Learning with annotation noise. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP: Volume 1, ACL '09* (Vol. 1, pp. 280–287). Association for Computational Linguistics.
17. Bellman, R. (1957). *Dynamic programming*. Princeton University Press.
18. Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. The MIT Press.
19. Bierlaire, M., Chen, J., & Newman, J. (2013). A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, 26, 78–98.
20. Blum, J. R., Greencorn, D. G., & Cooperstock, J. R. (2013). Smartphone sensor reliability for augmented reality applications. In K. Zheng, M. Li, & H. Jiang (Eds.), *Mobile and ubiquitous systems: Computing, networking, and services* (pp. 127–138). Springer.
21. 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.
22. Byon, Y. J., & Liang, S. (2014). Real-time transportation mode detection using smartphones and artificial neural networks: Performance comparisons between smartphones and conventional global positioning system sensors. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 18, 264–272.
23. Calastri, C., Dit Sourd, R. C., & Hess, S. (2018). We want it all: Experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. *Transportation*, 47, 175–201.
24. Carpineti, C., Lomonaco, V., Bedogni, L., Felice, M. D., & Bononi, L. (2018). Custom dual transportation mode detection by smartphone devices exploiting sensor diversity. In *Proceedings of the 14th workshop on context and activity modeling and recognition (IEEE COMOREA 2018)*.
25. Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., del Millán, J. R., & Roggen, D. (2013). The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15), 2033–2042. Smart Approaches for Human Action Recognition.
26. Chen, J., & Bierlaire, M. (2015). Probabilistic multimodal map matching with rich smartphone data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 19(2), 134–148.
27. Christensen, L. (2013). The Role of Web Interviews as Part of a National Travel Survey. In J. Zmud, M. Lee-Gosselin, M. Munizaga, & J. A. Carrasco (Eds.), *Transport Survey Methods* (pp. 115–154). Emerald Group Publishing Limited. <https://doi.org/10.1108/9781781902882-006>.
28. Christiansen, H. (Author), & Warnecke, M.-L. (Author). (2018). The Danish National Travel Survey - declaration of variables TU 2006-17, version 1. Dataset, DTU Management.
29. Cornacchia, M., Ozcan, K., Zheng, Y., & Velipasalar, S. (2017). A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal*, 17(2), 7742959.
30. Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M., & Zegras, P. (2013). Future mobility survey. *Transportation Research Record: Journal of the Transportation Research Board*, 2354, 59–67.
31. Dabiri, S., & Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies*, 86(November 2017), 360–371.
32. Dabiri, S., Lu, C.-T., Heaslip, K., & Reddy, C. K. (2019). Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, 32, 1010–1023.
33. Das, R. D., & Winter, S. (2016). Automated urban travel interpretation: A bottom-up approach for trajectory segmentation. *Sensors (Switzerland)*, 16(11), 1962.
34. Davidson, P., & Piché, R. (2017). A survey of selected indoor positioning methods for smartphones. *IEEE Communications Surveys Tutorials*, 19(2), 1347–1370.
35. De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, 1–5.
36. Ectors, W., Reumers, S., Lee, W. D., Choi, K., Kochan, B., Janssens, D., Bellemans, T., & Wets, G. (2017). Developing an optimised activity type annotation method based on classification accuracy and entropy indices. *Transportmetrica A: Transport Science*, 13(8), 742–766.
37. Ehsani, R., Buchanan, S., & Salyani, M. (2009). GPS Accuracy for Tree Scouting and Other Horticultural Uses. EDIS, 2009(2). Retrieved from <https://journals.flvc.org/edis/article/view/117815>.
38. Ek, A., Alexandrou, C., Delisle Nyström, C., Direito, A., Eriksson, U., Hammar, U., Henriksson, P., Maddison, R., Trolle Lagerros, Y., & Löf, M. (2018). The Smart City Active Mobile Phone Intervention (SCAMPI) study to promote physical activity through active transportation in healthy adults: A study protocol for a randomised controlled trial. *BMC Public Health*, 18, 1–11.
39. Faouzi, N. E. E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges—A survey. *Information Fusion*, 12, 4–10.
40. Feng, T., & Timmermans, H. J. (2015). Detecting activity type from GPS traces using spatial and temporal information. *European Journal of Transport and Infrastructure Research*, 15(4), 662–674.
41. Gadziński, J. (2018). Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study. *Transportation Research Part C: Emerging Technologies*, 88(July 2017), 74–86.
42. Garg, N. (2018). Mining bus stops from raw GPS data of bus trajectories. In *10th International conference on communication systems & networks (COMSNETS), Bengaluru, India* (pp. 583–588). IEEE.
43. Geurs, K. T., Thomas, T., Bijlsma, M., & Douhou, S. (2015). Automatic trip and mode detection with move smarter: first results from the dutch mobile mobility panel. *Transport Res Proc*, <https://doi.org/10.1016/j.trpro.2015.12.022>.
44. 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.
45. Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C., & Crane, M. (2015). A web-based diary and companion smartphone app for travel/activity surveys. *Transportation Research Procedia*, 11, 297–310.
46. Guidotti, R., Trasarti, R., & Nanni, M. (2015). TOSCA: Two-steps clustering algorithm for personal locations detection. In *GIS: Proceedings of the ACM international symposium on advances in geographic information systems*.
47. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations Newsletter*, 11(1), 10–18.
48. Hariharan, R., & Toyama, K. (2004). Project lachesis: Parsing and modeling location histories. In M. J. Egenhofer, C. Freksa, & H. J. Miller (Eds.), *Geographic information science* (pp. 106–124). Springer.

49. Hoseini-Tabatabaei, S. A., Gluhak, A., & Tafazolli, R. (2013). A survey on smartphone-based systems for opportunistic user context recognition. *ACM Computing Surveys*, 45(3), 1–51.
50. Houston, D., Luong, T. T., & Boarnet, M. G. (2014). Tracking daily travel: Assessing discrepancies between GPS-derived and self-reported travel patterns. *Transportation Research Part C: Emerging Technologies*, 48, 97–108.
51. Huang, J., Qiao, S., Yu, H., Qie, J., & Liu, C. (2014). Parallel map matching on massive vehicle GPS data using MapReduce. In *Proceedings—2013 IEEE international conference on high performance computing and communications, HPCC 2013 and 2013 IEEE international conference on embedded and ubiquitous computing, EUC 2013* (pp. 1498–1503).
52. Hunter, T., Abbeel, P., & Bayen, A. (2014). The path inference filter: Model-based low-latency map matching of probe vehicle data. *IEEE Transactions on Intelligent Transportation Systems*, 15(2), 507–529.
53. Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40, 63–74.
54. Jagadeesh, G. R., & Srikanthan, T. (2017). Online map-matching of noisy and sparse location data with hidden Markov and route choice models. *IEEE Transactions on Intelligent Transportation Systems*, 18, 2423–2434.
55. Jahangiri, A., & Rakha, H. A. (2015). Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE Transactions on Intelligent Transportation Systems*, 16(5), 2406–2417.
56. Jeon, K. E., She, J., Soonsawad, P., & Ng, P. C. (2018). BLE beacons for internet of things applications: Survey, challenges, and opportunities. *IEEE Internet of Things Journal*, 5(2), 811–828.
57. Jiang, X., de Souza, E. N., Pesaranghader, A., Hu, B., Silver, D. L., & Matwin, S. (2017). TrajectoryNet: An embedded GPS trajectory representation for point-based classification using recurrent neural networks. Source code published on Github at <https://github.com/wuhaotju/TrajectoryNet>. Retrieved November 1, 2019, from web.
58. Kanarachos, S., Christopoulos, S. R. G., & Chronos, A. (2018). Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity. *Transportation Research Part C: Emerging Technologies*, 95(March), 867–882.
59. Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387–399.
60. Kim, Y., Pereira, F. C., Zegras, P. C., & Ben-akiva, M. (2018). Activity recognition for a smartphone and web-based human mobility sensing system. *IEEE Intelligent Systems*, 33(August), 5–23.
61. Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D., & Laurila, J. (2010). Towards rich mobile phone datasets: Lausanne data collection campaign. *Proc. ICPS, Berlin*, 68, 7.
62. Koushik, A. N., Manoj, M., & Nezamuddin, N. (2020). Machine learning applications in activity-travel behaviour research: A review. *Transport Reviews*, 40, 1–24.
63. Kubicka, M., Cela, A., Moulin, P., Mounier, H., & Niculescu, S. I. (2016). Dataset for testing and training map-matching methods [Data set]. 2015 IEEE Intelligent Vehicles Symposium (IV 2015), Seoul, South Korea. Zenodo. <https://doi.org/10.5281/zenodo.57731>.
64. Kubicka, M., Cela, A., Mounier, H., & Niculescu, S. I. (2018). Comparative study and application-oriented classification of vehicular map-matching methods. *IEEE Intelligent Transportation Systems Magazine*, 10(2), 150–166.
65. Laurila, J. K., Gatica-Perez, D., Aad, I., Blom, J., Borner, O., Do, T. M. T., Dousse, O., Eberle, J., & Miettinen, M. (2013). From big smartphone data to worldwide research: The mobile data challenge. *Pervasive and Mobile Computing*, 9(6), 752–771.
66. Li, C., Zegras, P. C., Zhao, F., Qin, Z., Shahid, A., Ben-Akiva, M., Pereira, F., & Zhao, J. (2017). Enabling bus transit service quality co-monitoring through smartphone-based platform. *Transportation Research Record: Journal of the Transportation Research Board*, 2649(1), 42–51.
67. Li, H., & Wu, G. (2014). *Map matching for taxi GPS data with extreme learning machine* (Vol. 8933). Springer.
68. Li, L., Quddus, M., & Zhao, L. (2013). High accuracy tightly-coupled integrity monitoring algorithm for map-matching. *Transportation Research Part C: Emerging Technologies*, 36, 13–26.
69. Li, X., Zhang, X., Chen, K., & Feng, S. (2014). Measurement and analysis of energy consumption on android smartphones. In *2014 4th IEEE International conference on information science and technology* (pp. 242–245).
70. Liao, L., Fox, D., & Kautz, H. (2007). Extracting places and activities from GPS traces using hierarchical conditional random fields. *International Journal of Robotics Research*, 26, 119–134.
71. Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., & Huang, Y. (2009). Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems—GIS '09*, (c) (p. 352).
72. Mäenpää, H., Lobov, A., & Martínez Lastra, J. L. (2017). Travel mode estimation for multi-modal journey planner. *Transportation Research Part C: Emerging Technologies*, 82, 273–289.
73. Teng, C. M. (2001, May). A Comparison of Noise Handling Techniques. In *Proceedings of the Fourteenth International Florida Artificial Intelligence Research Society Conference* (pp. 269–273).
74. Manwani, N., & Sastry, P. S. (2013). Noise tolerance under risk minimization. *IEEE Transactions on Cybernetics*, 43(3), 1146–1151.
75. Martin, B. D., Addona, V., Wolfson, J., Adomavicius, G., & Fan, Y. (2017). Methods for real-time prediction of the mode of travel using smartphone-based GPS and accelerometer data. *Sensors (Switzerland)*, 17(9), 2058.
76. Montini, L., Rieser-Schüssler, N., Horni, A., & Axhausen, K. (2014). Trip purpose identification from GPS tracks. *Transportation Research Record: Journal of the Transportation Research Board*, 2405, 16–23.
77. Nettleton, D. F., Orriols-Puig, A., & Fornells, A. (2010). A study of the effect of different types of noise on the precision of supervised learning techniques. *Artificial Intelligence Review*, 33(4), 275–306.
78. Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems—GIS '09* (pp. 336–343).
79. Nicholls, L. II., & Groves, R. M. (1986). The status of computer-assisted telephone interviewing: Part I—Introduction and impact on cost and timeliness of survey data. *Journal of Official Statistics*, 2(2), 93.
80. Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., & Maurer, P. (2014). Supporting large-scale travel surveys with smartphones—A practical approach. *Transportation Research Part C: Emerging Technologies*, 43, 212–221.
81. Nurmi, P., & Koolwaaij, J. (2006). Identifying meaningful locations. In *2006 3rd Annual international conference on mobile and ubiquitous systems: Networking and services, MobiQuitous*.
82. Oshin, T. O., Poslad, S., & Ma, A. (2012). Improving the energy-efficiency of GPS based location sensing smartphone applications. In *Proceedings of the 11th IEEE international conference on trust, security and privacy in computing and communications, TrustCom-2012—11th IEEE international conference on ubiquitous computing and communications, IUCC-2012* (pp. 1698–1705).
83. Patterson, Z., & Fitzsimmons, K. (2016). Datamobile: Smartphone travel survey experiment. *Transportation Research Record*, 2594, 35–53.
84. Patterson, Z., Fitzsimmons, K., Jackson, S., & Mukai, T. (2019). Itinerum: The open smartphone travel survey platform. *SoftwareX*, 10, 100230.
85. Perrucci, G. P., Fitzek, F. H. P., & Widmer, J. (2011). Survey on energy consumption entities on the smartphone platform. In *2011 IEEE 73rd Vehicular technology conference (VTC Spring)* (pp. 1–6).
86. Prelipcean, A. C., Gidofalvi, G., & Susilo, Y. O. (2016). Measures of transport mode segmentation of trajectories. *International Journal of Geographical Information Science*, 30(9), 1763–1784.
87. Prelipcean, A. C., Gidofalvi, G., & Susilo, Y. O. (2018). MEILL: A travel diary collection, annotation and automation system. *Computers, Environment and Urban Systems*, 70, 24–34.
88. Primault, V., Boutet, A., Mokhtar, S. B., & Brunie, L. (2019). The long road to computational location privacy: A survey. *IEEE Communications Surveys and Tutorials*, 21(3), 8482357, 2772–2793.
89. Quddus, M. A., Noland, R. B., & Ochieng, W. Y. (2006). A high accuracy fuzzy logic based map matching algorithm for road transport. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 10(3), 103–115.
90. 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*, 54, 301–313.
91. Renso, C., Baglioni, M., de Macedo, J. A. F., Trasarti, R., & Wachowicz, M. (2013). How you move reveals who you are: Understanding human behavior by analyzing trajectory data. *Knowledge and Information Systems*, 37(2), 331–362.
 92. Rolnick, D., Veit, A., Belongie, S., & Shavit, N. (2018). Deep learning is robust to massive label noise. Retrieved November 14, 2019, from the arXiv database.
 93. Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *European Physical Journal Special Topics*, 178(1), 13–23.
 94. Schuessler, N., & Axhausen, K. W. (2009). Processing raw data from global positioning systems without additional information. *Transportation Research Record*, 2105(1), 28–36.
 95. Seidl, D. E., Jankowski, P., & Tso, M. H. (2016). Privacy and spatial pattern preservation in masked GPS trajectory data. *International Journal of Geographical Information Science*, 30(4), 785–800.
 96. Semanjski, I., Gautama, S., Ahas, R., & Witlox, F. (2017). Spatial context mining approach for transport mode recognition from mobile sensed big data. *Computers, Environment and Urban Systems*, 66, 38–52.
 97. Shankari, K., Fürst, J., Fadel Argerich, M., Avramidis, E., & Zhang, J. (2020). MobilityNet: Towards a Public Dataset for Multi-modal Mobility Research. *ICLR 2020 Workshop on Tackling Climate Change with Machine Learning*. <https://www.climatechange.ai/papers/iclr2020/15.html>.
 98. Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 34(3), 316–334. <https://doi.org/10.1080/01441647.2014.903530>.
 99. Sicotte, G., Morency, C., & Farooq, B. (2017). Comparison between trip and trip chain models: Evidence from Montreal commuter train corridor (No. CIRRELT-2017-35). CIRRELT, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport = Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation.
 100. Silver, D. L., Yang, Q., & Li, L. (2013). Lifelong machine learning systems: Beyond learning algorithms. In *2013 AAAI spring symposium series*. Citeseer.
 101. Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, 41(5), 367–381.
 102. Stopher, P. R., Shen, L., Liu, W., & Ahmed, A. (2015). The challenge of obtaining ground truth for GPS processing. *Transportation Research Procedia*, 11, 206–217. Transport Survey Methods: Embracing Behavioural and Technological Changes Selected contributions from the 10th International Conference on Transport Survey Methods 16–21 November 2014, Leura, Australia.
 103. Thierry, B., Chaix, B., & Kestens, Y. (2013). Detecting activity locations from raw GPS data: A novel kernel-based algorithm. *International Journal of Health Geographics*, 12, 1–10.
 104. Thomas, T., Geurs, K. T., Koolwaaij, J., & Bijlsma, M. (2018). Automatic trip detection with the dutch mobile mobility panel: Towards reliable multiple-week trip registration for large samples. *Journal of Urban Technology*, 25, 1–19.
 105. Tietbohl, A., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the ACM symposium on applied computing*.
 106. Torre, F., Pitchford, D., Brown, P., & Terveen, L. (2012). Matching GPS traces to (possibly) incomplete map data. In *Proceedings of the 20th international conference on advances in geographic information systems—SIGSPATIAL '12* (p. 546).
 107. Van Dijk, J. (2018). Identifying activity-travel points from GPS-data with multiple moving windows. *Computers, Environment and Urban Systems*, 70(September 2017), 84–101.
 108. Velasco-Gallego, C., & Lazakis, I. (2020). Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study. *Ocean Engineering*, 218, 108261.
 109. von Watzdorf, S., & Michahelles, F. (2010). Accuracy of positioning data on smartphones. In *Proceedings of the 3rd international workshop on location and the web, LocWeb '10, New York*. Association for Computing Machinery.
 110. Vuk, G., Bowman, J. L., Daly, A., & Hess, S. (2016). Impact of family in-home quality time on person travel demand. *Transportation*, 43(4), 705–724.
 111. Wang, D., Zhang, J., Cao, W., Li, J., & Zheng, Y. (2018). When will you arrive? Estimating travel time based on deep neural networks. In *IJCAI*.
 112. Wang, L., Gjoeski, H., Ciliberto, M., Mekki, S., Valentin, S., & Roggen, D. (2019). Enabling reproducible research in sensor-based transportation mode recognition with the Sussex–Huawei dataset. *IEEE Access*, 7, 10870–10891.
 113. Wang, L., Jiao, L., Li, J., Gedeon, J., & Mühlhäuser, M. (2019). Moera: Mobility-agnostic online resource allocation for edge computing. *IEEE Transactions on Mobile Computing*, 18(8), 1843–1856.
 114. Wee, B. V., & Banister, D. (2016). How to write a literature review paper? *Transport Reviews*, 36(2), 278–288.
 115. Wei, H., Wang, Y., Forman, G., & Zhu, Y. (2013). Map matching: Comparison of approaches using sparse and noisy data. In *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems, SIGSPATIAL'13, New York* (pp. 444–447). Association for Computing Machinery.
 116. Wu, H., Chen, Z., Sun, W., Zheng, B., & Wang, W. (2017). Modeling trajectories with recurrent neural networks. In *IJCAI International joint conference on artificial intelligence* (pp. 3083–3090).
 117. Xiang, L., Gao, M., & Wu, T. (2016). Extracting stops from noisy trajectories: A sequence oriented clustering approach. *ISPRS International Journal of Geo-Information*, 5, 29.
 118. Xiao, G., Cheng, Q., & Zhang, C. (2019). Detecting travel modes from smartphone-based travel surveys with continuous hidden Markov models. *International Journal of Distributed Sensor Networks*, 15, 1550147719844156.
 119. Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Computers, Environment and Urban Systems*, 54, 14–22.
 120. Xiao, G., Juan, Z., & Zhang, C. (2016). Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. *Transportation Research Part C: Emerging Technologies*, 71, 447–463.
 121. Xiao, L., Li, Y., Han, G., Dai, H., & Poor, H. V. (2018). A secure mobile crowd-sensing game with deep reinforcement learning. *IEEE Transactions on Information Forensics and Security*, 13(1), 35–47.
 122. Yazdizadeh, A., Patterson, Z., & Farooq, B. (2019). An automated approach from GPS traces to complete trip information. *International Journal of Transportation Science and Technology*, 8, 82–100.
 123. Yazdizadeh, A., Patterson, Z., & Farooq, B. (2019). Ensemble convolutional neural networks for mode inference in smartphone travel survey. *IEEE Transactions on Intelligent Transportation Systems*, 21, 2232–2239.
 124. Kalatian, A., & Farooq, B. (2020). A semi-supervised deep residual network for mode detection in Wi-Fi signals. *Journal of Big Data Analytics in Transportation*, 2(2), 167–180.
 125. Yurur, O., Liu, C. H., Sheng, Z., Leung, V. C. M., Moreno, W., & Leung, K. K. (2016). Context-awareness for mobile sensing: A survey and future directions. *IEEE Communications Surveys and Tutorials*, 18(1), 68–93.
 126. Zhao, F., Ghorpade, A., Pereira, F. C., Zegras, C., & Ben-Akiva, M. (2015a). Stop detection in smartphone-based travel surveys. *Transportation Research Procedia*, 11(2010), 218–226.
 127. Zhao, F., Pereira, F. C., Ball, R., Kim, Y., Han, Y., Zegras, C., & Ben-Akiva, M. (2015b). Exploratory analysis of a smartphone-based travel survey in Singapore. *Transportation Research Record*, 2494(1), 45–56.
 128. Zheng, Y. (2015). Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 1–41.
 129. Zheng, Y., & Fu, H. (2011). Geolife GPS trajectory dataset—User guide. Technical Report November 31. Online. Retrieved July 19, 2008.
 130. Zheng, Y., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on world wide web—WWW '09*.
 131. Zhou, R., Li, M., Wang, H., Song, X., Xie, W., & Lu, Z. (2017). An enhanced transportation mode detection method based on GPS data. *Communications in Computer and Information Science*, 727, 605–620.
 132. Zhou, X., Yu, W., & Sullivan, W. C. (2016). Making pervasive sensing possible: Effective travel mode sensing based on smartphones. *Computers, Environment and Urban Systems*, 58, 52–59.

133. Zhu, Q., Zhu, M., Li, M., Fu, M., Huang, Z., Gan, Q., & Zhou, Z. (2016). Identifying transportation modes from raw GPS data. In *Communications in computer and information science*.
134. Zhu, X., Li, J., Liu, Z., Wang, S., & Yang, F. (2016). Learning transportation annotated mobility profiles from GPS data for context-aware mobile services. In *Proceedings—2016 IEEE international conference on services computing, SCC 2016* (pp. 475–482).
135. Zmud, J., Lee-Gosselin, M., Carrasco, J. A., & Munizaga, M. A. (2013). *Transport survey methods: Best practice for decision making*. Emerald Group Publishing.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
