Continuous Trajectory Pattern Mining for Mobility Behaviour Change Detection

David Jonietz and Dominik Bucher

Abstract With the emergence of ubiquitous movement tracking technologies, developing systems which continuously monitor or even influence the mobility behaviour of individuals in order to increase its sustainability is now possible. Currently, however, most approaches do not move beyond merely describing the status quo of the observed mobility behaviour, and require an expert to assess possible behaviour changes of individual persons. Especially today, automated methods for this assessment are needed, which is why we propose a framework for detecting behavioural anomalies of individual users by continuously mining their movement trajectory data streams. For this, a workflow is presented which integrates data preprocessing, completeness assessment, feature extraction and pattern mining, and anomaly detection. In order to demonstrate its functionality and practical value, we apply our system to a real-world, large-scale trajectory dataset collected from 139 users over 3 months.

Keywords Mobility · Trajectory mining · Anomaly detection
Sustainability · Behavior change

1 Introduction

Human mobility is ubiquitous in modern societies and represents an integral part of our daily behavioural routines. At the same time, however, there are numerous undesirable effects, such as traffic jams or increased fossil fuel consumption (Taaffe 1996). With regards to Switzerland, for instance, roughly a half of the total CO_2 emissions are contributed by the transportation sector (including international aviation), with motorized individual mobility being responsible for around two thirds of these emissions (Bundesamt fuer Umwelt 2014). If no major changes occur in the

D. Jonietz (✉) · D. Bucher
Institute of Cartography and Geoinformation, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland
e-mail: jonietzd@ethz.ch

D. Bucher
e-mail: dobucher@ethz.ch

P. Kiefer et al. (eds.), *Progress in Location Based Services 2018*, Lecture Notes in Geoinformation and Cartography, https://doi.org/10.1007/978-3-319-71470-7_11

transport system, these numbers are widely expected to rise in the coming decades (Boulouchos et al. 2017).

Recently, the significance of emerging technologies which enable ubiquitous monitoring as well as real-time regulation and management of human mobility has been emphasized as potential game changing aspect for increasing the sustainability of travel behaviour (Boulouchos et al. 2017). Indeed, current developments in the field of location-acquisition technologies such as Global Navigation Satellite Systems (GNSS), Wireless Local Area Networks (WLAN), or Global System for Mobile Communications (GSM) allow to monitor and record human movement at an exceptional level of detail and at relatively low cost and effort (Feng and Zhu 2016). Due to the widespread use of modern smart phones, as well as a general trend towards digitalization in the transportation and mobility sector, Big Mobility Data are now widely available and ready to be utilized for gaining unprecedented insights into the fundamental mechanisms that guide human mobility (Brunauer and Rehrl 2016).

In fact, since the late 1990s, human movement trajectories, i.e. series of chronologically ordered x, y-coordinate pairs with time stamps (Andrienko et al. 2016), have increasingly been used for travel surveys (Shen and Stopher 2017). Apart from notable exceptions (e.g. Schlich and Axhausen 2003; Stopher et al. 2013), however, these studies have mainly applied a snapshot approach (e.g. Schüssler 2008; Kohla and Meschik 2013), with the center of interest being put on inter-personal variability (differences in the behaviour of different persons) rather than intra-personal variability (different behaviour of one person from day to day) (Schlich and Axhausen 2003). What has often been neglected, therefore, is analysing the dynamic dimension of mobility behaviour, i.e. behaviour changes such as trying out new travel alternatives, or forming new mobility habits.

Especially today, however, it would be worthwhile to be able to automatically detect and analyse such changes in mobility behaviour. On the one hand, in contrast to merely surveying mobility behaviour, there are now systems which move further by aiming to directly influence people's mobility behaviour towards more sustainable transport alternatives (cf. Banister 2008), e.g. by using mobile applications which continuously record the movements of users, stream the data to a server, and utilize them to provide their users with feedback or even suggest more sustainable travel options (Froehlich et al. 2009; Montini et al. 2015). To the best of our knowledge, currently none of these systems apply strategies for automatically detecting behaviour change, but instead require manual checking of the data for evaluating the effectiveness of the conducted persuasive measures. A fully automated system which continuously monitors movement behaviour based on a stream of trajectory data, and detects behavioural changes, however, could take over this tedious task and even trigger dynamic reactions to users based on their behavioural changes, e.g. encourage sustainable mobility behaviour adaptations and discourage in the opposite case. On the other hand, apart from application scenarios where behaviour change is actively induced, the development of methods for detecting such variations in movement data would also be useful for general transportation research and planning purposes. Thus, for instance, insights are still needed in terms of evaluating and predicting peoples reactions to today's novel mobility options, such as shared mobility,

mobility as a service, electric mobility and autonomous vehicles. Being confronted with these, one can expect numerous people to adapt their mobility behaviour, e.g. by testing novel alternatives and even forming new travel habits (Boulouchos et al. 2017). In order to accurately understand these behavioural changes, travel surveys are needed which involve tracking numerous participants over a long period of time. In addition, a set of suitable methods are necessary to analyse the collected data and be able to accurately understand these behavioural changes.

For developing such methods, however, a practical problem is posed by insufficient data quality. It is especially data incompleteness which represents a critical challenge for GNSS-based travel surveys, since it comprises missing records for parts of trips, one or more full trips, or even one or more full days of the recording period (Hecker et al. 2010). These gaps can have various causes, e.g. the cold start problem at the start of movement, bad signal reception, participants leaving the device switched off, or other technological problems (Shen and Stopher 2017). While shorter gaps can often be handled by means of map matching techniques (see Sect. 2.1), longer ones can heavily distort or bias the results of the following analyses. In the context of automated behaviour change detection, for instance, the occurrence of missing movement data could lead to misleading calculations, e.g. drastically lower values for CO_2 emissions produced during the respective week of recording. In this case, a system might erroneously interpret this drop in numbers as a behaviour change, whereas it is in fact merely the result of missing data. To avoid such misdetection of behaviour changes, methods need to be sensitive to recording gaps, i.e. distinguish them from cases where observed changes are actually due to changed mobility patterns.

Before this background, this study proposes a method for identifying and evaluating changes in human mobility behaviour by first detecting and quantifying spatio-temporal recording gaps in a stream of movement trajectory data, and then continuously mining it for anomalies with regards to various mobility features, i.e. a subset of variables which can be extracted from movement data, and describe selected aspects of mobility behaviour (e.g. average speed, travelled distances). Focussing on sustainable mobility as the application scenario, we simulate a real-time data stream using a real trajectory dataset collected from 139 users over 3 months in Switzerland.

This paper is structured as follows: First, in Sect. 2 background information is provided starting with a brief review of available methods for surveying human mobility behaviour on the basis of movement trajectory datasets. Then, the focus is shifted to the potential of similar techniques for inducing and analysing changes in mobility behaviour. In the following Sect. 3, our concept is presented and discussed with regards to data preprocessing, completeness assessment, feature extraction and pattern mining, and finally anomaly detection. In Sect. 4, the framework is applied to a test dataset, before the results are discussed and the paper is concluded in Sect. 5.

2 Related Work

In the context of this study, relevant prior work applies one of two distinct perspectives on mobility behaviour and movement data analysis, and is briefly reviewed in this section:

1. Assessing the **present state of mobility behaviour**, i.e. *where*, *when* and *how* a person travels. This is normally achieved by means of GNSS-assisted travel surveys.
2. Aiming to **change existing mobility behaviour** in order to increase its sustainability, e.g. by means of mobile applications which provide both tracking and user feedback functionalities.

2.1 Movement Trajectories for Surveying Human Mobility Behaviour

Before the rise of position tracking technologies, the traditional ways of gaining insights about the mobility behaviour of people were face-to-face interviews, mail-out/mail-back or telephone surveys. Since the late 1990s, however, GNSS-assisted travel surveys emerged as a novel method, and gradually replaced these approaches due to numerous advantages, such as a relatively high accuracy in recording time and position, low cost (especially with modern smartphones), and less problems with regards to trip-misreporting by respondents (Shen and Stopher 2017). Nowadays, exemplary approaches are manifold, and have spread from pilot studies undertaken in the USA (Wagner 1997) to a range of other countries, including Switzerland (Shen and Stopher 2017).

After recording the movements of test persons, the data require extensive processing in order to extract relevant mobility features, in particular places that have been visited for a certain purpose and the travelled routes between these places. With regards to the former category, stay points are typically detected based on various clustering techniques (e.g. Palma et al. 2008), or the movement speed (e.g. Li et al. 2008). With regards to the travelled routes, via map matching, the exact path taken through a road network can be inferred from the tracking points, e.g. by simple point-to-curve snapping (e.g. White et al. 2000) or advanced techniques such as evolutionary algorithms (Quddus and Washington 2015). Apart from the routes, numerous studies have proposed approaches to infer the used traffic mode, for instance based on identifying walking transitions between mode changes (Zheng et al. 2010), analysing a range of movement descriptors (Sester et al. 2012), or the underlying transportation network (Stenneth et al. 2011).

In order to describe a person's mobility behaviour based on trajectory data, these (and other) mobility features need to be further analysed to extract patterns, i.e. observable regularities in movement behaviour such as habits or long-lasting preferences and restrictions. Thus, one can calculate general statistics over certain time intervals, such as the average duration and length of trips, the modal split, or the usual times of travel (Axhausen and Frick 2005), but also more use-case specific aspects such as frequently visited places other than home or the work location (Siła-Nowicka et al. 2015) or the location of regularly performed activities like eating, shopping or physical exercise (e.g. Zheng et al. 2010; Furletti et al. 2013). When being properly interpreted, mobility features and their regular patterns can serve as indicators for higher-level attributes, such as the sustainability of mobility behaviour. In this context, for instance, (Nicolas et al. 2003) formulated a set of potential sustainability indicators which can be extracted from travel survey data. Among others which refer to the aggregate city level, those which could be extracted from trajectory data include the daily number of trips, the structure of trip purposes (e.g. commuting versus leisure), the daily average time budget spent for travelling, the modal split (especially the share of slow mobility, i.e. walking and cycling), the average distance travelled daily, and the average movement speed. Other relevant indicators which have been formulated in the literature include the amount of CO_2 emissions and the degree to which trips are intermodally integrated, i.e. use different traffic modes in combination (World Business Council 2015).

Naturally, the validity of the results computed for mobility features depend to a large degree on the quality of the input trajectory data, in particular the completeness of the recorded movement. Missing trips or even full day gaps will lead to erroneous, in some cases even heavily biased, results (Hecker et al. 2010), however, are a regularly occurring issue in travel surveys (Shen and Stopher 2017). Although this issue is frequently discussed in the literature (e.g. Shen and Stopher 2017; Wolf et al. 2003), only few studies propose solutions, such as evaluating the intrinsic trajectory data quality based on the spatial and temporal resolution (Prelipcean et al. 2015), a statistical approach to detect dependencies between mobility behaviour, socio-demography and missing data (Hecker et al. 2010), or imputation, the process of inferring the missing trips based on observed data using statistical relationships (Polak and Han 1997). Another popular option to improve and ensure the completeness and correctness of the movement data in travel surveys are prompted recall (PR) methods, in which during the tracking phase, respondents are regularly asked to manually validate and complete their recorded movements, for instance at the end of each day (e.g. Bucher et al. 2016).

In traditional travel surveys, the focus is usually put on analysing the status quo of mobility behaviour, since, as (Schlich and Axhausen 2003) argue, there is a general assumption that travel behaviour mainly consists of highly habitual routines, and remains relatively static over time. Thus, in most cases, mobility features are calculated once on the basis of the entire available data in order to assess the present state of transportation system usage (e.g. Schüssler 2008; Kohla and Meschik 2013) rather than analysing its temporal dynamics. Additionally, this snapshot approach is often caused by practical limitations with regards to the available movement data, with durations of the tracking period rarely exceeding two weeks (Shen and Stopher 2017). There are, however, also examples of longitudinal analyses of travel behaviour (e.g. Hanson and Huff 1988; Schlich and Axhausen 2003; Stopher et al. 2013; Gonzalez et al. 2008; Song et al. 2010). These studies were mostly concerned with detecting day-to-day variations, stability measures, and statistical properties of mobility behaviour from movement data of various kinds, such as those obtained with GSM or GPS, or traditional travel survey methods. While GSM data typically covers long durations and large numbers of users, transport surveys and GPS recordings stem from much less persons over the course of merely a few weeks. Gonzalez et al. (2008), for instance, developed an aggregated model of human mobility based on extensive mobile phone data, and found strong inter-personal regularities, but did not distinguish between individual users or temporal changes. Schlich and Axhausen (2003) report on different mobility indicators, and how they can be used to compute similarity measures between mobility behaviour on two different days.

2.2 Inducing Change in Human Mobility Behaviour

Apart from merely monitoring and analysing the status quo of mobility behaviour, other studies have built on similar analytical methods to actively influence users in order to make them travel in a more environmentally sustainable way. For this, mobile applications and a feedback loop were used, with examples including *UbiGreen* (Froehlich et al. 2009), *PEACOX* (Montini et al. 2015), or *GoEco!* (Bucher et al. 2016). In some cases, apart from merely summarizing the recorded mobility behaviour, the provided feedback also included the proposal of more sustainable travel alternatives. At present, however, most approaches suffer from either short study periods (Hamari et al. 2014), or from basing their feedback and suggestions for more sustainable mobility on a single snapshot, for example data which was recorded during a pre-study or a baseline-tracking phase. This shortcoming hinders the development of long-running applications that continuously monitor mobility behaviour and are thus able to provide feedback based on detected changes of current in comparison to past behavioural patterns.

Thus, a system would be worthwile with the ability to automatically detect changes in behaviour, which could then, based on established models of behavioural change processes, select actions to be taken to support (in case of increased sustainability) or prevent (in the opposite case) the observed behaviour change. A commonly used psychological conceptualization is the Transtheoretical Model (Prochaska and Velicer, 1997) which separates behaviour change into *precontemplation*, *contemplation*, *preparation*, *action* and *maintenance* phases. Upon detecting a change in mobility, one could for instance infer that a user started contemplating new behavior, and support a transition towards this behavior by supplying her with *information* (e.g. Tulusan et al. 2012; Taniguchi et al. 2003), *rewarding* further good choices (e.g. Ben-Elia and Ettema 2011), *dissuading* unsustainable behavior (e.g. Schade and Schlag 2003), or otherwise engage and motivate her to move to the *preparation* or *action* stage (Weiser et al., 2015). Alternatively, for users without changes in mobility (one could argue they are in a *precontemplation* or *maintenance* phase), a system might foster *self-experience of travel alternatives* (e.g. Abou-Zeid et al. 2012; Bamberg et al. 2003; Bamberg 2006) in order to make them try out new and more sustainable transport options.

Automatically exposing behaviour change is closely related to anomaly detection, the identification of deviations from a certain norm (Chandola et al. 2009). In contrast to filtering out noise, in this case the focus of interest is usually placed on the nature of the abnormalities themselves. In the transportation domain, researchers have been interested in detecting anomalies in large collective mobility datasets (cf. Souto and Liebig 2016; Yang and Liu 2011) for urban traffic applications and emergency management. Another line of research considers (geometrical) pattern matching on trajectory data (e.g. Florescu et al. 2012; Du Mouza et al. 2005), for example by building a higher-order Markov model of a user's transitions from one mobile phone cell to another (Sun et al. 2004). The authors encode the individual patterns in a *mobility trie*, which they in turn use to search for anomalies by computing distances between previous and new, potentially anomalous patterns. They explicitly note on the importance of dynamically updating "normal behaviour", and weighting recent patterns higher than ones which occurred longer ago. However, all these approaches are based on a relatively crude assessment of mobility, which either only considers transitions from one region to another, or aggregate data from many users to get a complete view of the current traffic situation. For detecting individual behaviour change over time, however, a method is needed which works with a continuous stream of non-aggregated movement data on an individual level, and tests multiple dimensions of mobility behaviour for anomalies, by comparing them to the user's past behaviour.

3 Method

In this section, we present a system for detecting mobility behaviour change based on a continuous stream of movement data from individual users. The proposed workflow is illustrated in Fig. 1. We assume that a user's raw movement trajectories, recorded via a smartphone application or a similar device, are constantly streamed to a server, and logged in a database. After a certain time period has passed (we propose one week), the data recorded in this interval are fed into a data processing engine, where they pass through four processing steps: first, the trajectories are preprocessed, i.e. filtered, segmented, annotated with the traffic mode, and matched to the road network. Then, the available data for this time period are tested for completeness in order to evaluate their sufficiency for the following analytical processes. If found insufficiently complete, the data are discarded, if rated appropriate, however, they are fed to the next module, which extracts a range of mobility features and mines for patterns. The results are stored in a database, and provide the input for an anomaly detection sub-process, which identifies behaviour change and triggers an appropriate reaction. As can be seen on the far right of Fig. 1, this may involve sending out notifications to the users or analysts, triggering a response (e.g. encouraging or discouraging the observed behaviour change), logging the occurrence of

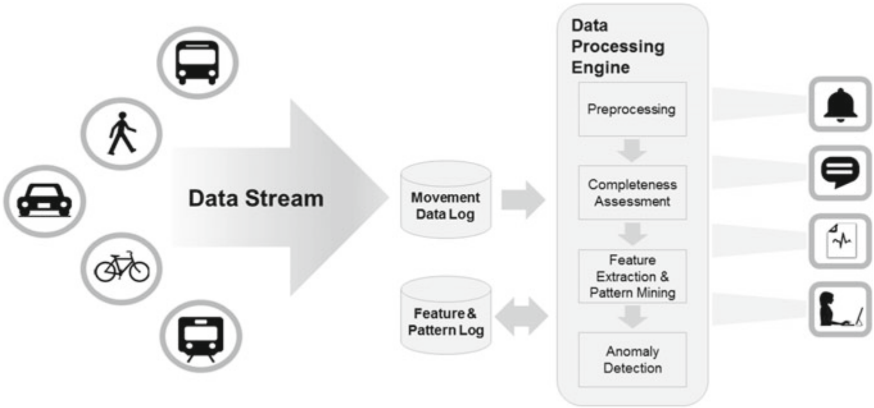

Fig. 1 Workflow

the anomaly, or providing information to an expert for decision support. The exact nature of these system reactions, however, is beyond the scope of this paper. Instead, since our focus is put on the data processing engine, its four sub-modules will be further described in this section.

3.1 Data Preprocessing

As it has been described, movement data are continuously streamed to a server, and logged in a database. In order to evaluate behavioural changes, however, it is necessary to define discrete time intervals (in the following: one week), which will serve as atomic units for later temporal analysis. Thus, after all available data for a full week have been stored in the database, they are fed into the data processing engine (cf. Fig. 1), and further analysed. In a first step, the data need to be preprocessed, which involves the sub-processes noise filtering, stay point detection, segmentation, mode detection, and map matching (Zheng 2015). Please note that whereas exemplary methods for these preprocessing steps are proposed in the following, they could also be replaced by other solutions which are better suited to the respective study aims or data characteristics.

In the beginning, the data are cleaned by removing noisy trackpoints based on a set of filter functions such as a spatial query with a certain study area, or plausibility checks with regards to speed constraints (Zheng 2015). Then, the stay points are detected in the remaining trackpoints, e.g. by means of a clustering technique (Palma et al. 2008). The next preprocessing step detects the traffic mode(s) used, e.g. by computing and analysing various movement descriptors such as the speed or acceleration (Sester et al. 2012). Finally, map matching needs to be performed for all

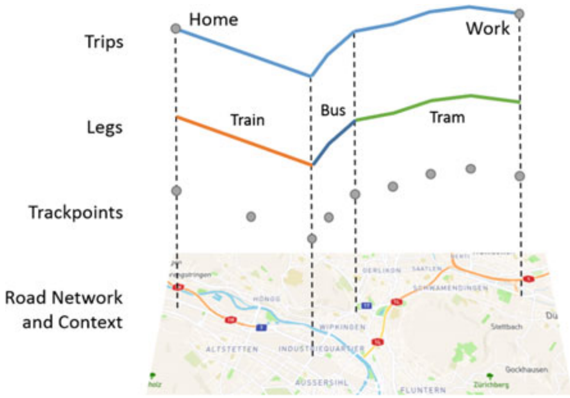

Fig. 2 The different layers of movement data aggregation used in this study. Note that in contrast to "home" and "work", the transition points between *train*, *bus* and *tram* are not considered activities. Basemap© Mapbox.com

points using one of the available techniques, e.g. evolutionary algorithms (Quddus and Washington 2015).

After basic preprocessing, it is necessary to structure the movement data into meaningful units. Inspired by prior approaches (Axhausen and Frick 2005), we propose to distinguish between the following elements: At the most fundamental level, trajectories (the complete trace of a users movement over a given time frame) are made up of **trackpoints**. In a first layer of aggregation, trackpoints are grouped into **trip legs** based on the used transport mode. Finally, a **trip** consists of one or more legs, and describes the journey from one 'activity' to another. A **stay point** simply denotes a location where someone spent longer than a certain time span, and can qualify as an **activity** if it represents an actual destination of travel (e.g. work, home or a shop), and not merely a location where a user spent time waiting for a bus or stuck in a traffic jam. Figure 2 shows an exemplary trip with its constituting elements.

3.2 Data Completeness Assessment

After preprocessing, the available data for the current week are tested for their completeness. As has been discussed in Sect. 2.1, missing trips or other gaps in recording can have negative effects on downstream analysis processes (e.g. Shen and Stopher 2017; Wolf et al. 2003). In our case, for instance, missing data, if not identified and filtered previously, might result in misdetections of behaviour changes due to drastically altered values for mobility features. Please note that in this step, we assume the norm to be continuous tracking over the whole study period, as it is often the case in related surveys (e.g. Montini et al. 2015; Bucher et al. 2016).
As a first step, we distinguish between different types of recording gaps:

- *Temporal* gaps: the duration with no recorded data between the last recorded time stamp of a trip leg or stay point and the first recorded time stamp of the sub-sequent trip leg or stay point. The spatial deviance between the position of the last track

point of the former, and the first track point of the latter tripleg or stay point is smaller than an expected GPS error (e.g. 250 m).
- *Spatio-temporal* gaps: gaps for which the spatial distance between the last track point of the former, and the first track point of the latter trip leg or stay point is larger than an expected GPS error.

This distinction is motivated by the fact that in the first case, chances are high that no mobility behaviour has been missed since the user might simply have remained stationary during the recording gap, whereas in the second case, the user's change in position proves that movement has certainly taken place but was not recorded.

Both types of gaps can be easily extracted from the database by calculating the time differences as well as spatial distances between the start and end points of subsequent pairs of trip legs and stay points. The data completeness for the current time interval can then be evaluated based on two index values:

$$gdur_i = \frac{\sum \Delta g_i}{\Delta t_i}$$

$$gdist_i = \frac{\sum dist(g_i)}{\sum dist(triplegs_i)}$$

where $gdur_i$ is the ratio of the summed durations Δg_i of all temporal and spatio-temporal gaps g_i and the total duration Δt_i of week i. In the second index, $gdist_i$ is the ratio of the summed distances $dist(g_i)$ of all spatio-temporal gaps g_i and the summed distances $dist(triplegs_i)$ of all trip legs $triplegs_i$ recorded within week i. In combination, these index values express the temporal extent of recording gaps, as well as the relative magnitude of missed mobility behaviour. For instance, in a week in which a user has travelled relatively less compared to others, recording gaps of similar temporal length can be rated as less critical, since less travelled distance, i.e. mobility behaviour, might be missing in the data.

3.3 Mobility Feature Extraction and Pattern Mining

After the available data has been confirmed to be of sufficient completeness, selected mobility features can be extracted. Of course, these will depend to a large degree on the study aims. As our focus is on sustainability, we compute durations, distances, speed, and produced CO_2 emissions for each trip leg to serve as basis for computing the indicators listed in Sect. 2.1. Next, in addition to segmenting the movement trajectories based on their semantics (e.g. trip legs by traffic mode, trips between activities), as described in Sect. 3.1, we also induce a temporal structure by grouping all movement on a daily basis. Of course, the pre-defined discrete time interval at which the data is processed (here: one week) provides a further temporal analytical unit.

Table 1 Units of analysis for deriving mobility features and patterns

Analysis unit	Delimiting factor	Description
Trip leg	Transport mode/vehicle	Mono-modal trip segment between two points without changing mode or vehicle
Trip	Purpose	Trip between two locations for a certain purpose; consists of one or more trip legs
Day	Time	All trips within 24 h; contains one or more complete or incomplete travels (incomplete: beyond temporal delimitation)
Week	Time	All trips within 7 consecutive days

Table 2 Mobility features

Descriptor	Day	Week
Total number of trips		x
Average number of triplegs per trip		x
Total distance travelled		x
Total distance travelled (per trip purpose)		x
Total distance travelled (per traffic mode)		x
Average distance travelled	x	x
Total duration spent travelling		x
Total duration spent travelling (per trip purpose)		x
Total duration spent travelling (per traffic mode)		x
Average duration spent travelling	x	x
Total CO_2 emissions		x
Average travel speed		x
Average travel speed (per traffic mode)		x
Frequently visited places		x

The resulting analytical units for computing mobility features are summarized in Table 1.

For assessing the sustainability of the user's mobility behaviour within the week, we compute a set of indicators (Nicolas et al. 2003; World Business Council 2015) as listed in Table 2.

Whereas the first three indicators can be easily extracted from the preprocessed data, several others require a classification of the stay points and their related trips according to their purpose. Purpose and activity detection can either be achieved by computational methods, e.g. based on visited POI (e.g. Furletti et al. 2013), or by simply asking the users to annotate the data manually in the course of an accompanying PR survey. The total CO_2 emissions produced by travelling depend primarily on the modal split, and can for instance be calculated based on the Mobitool consumption and emission factors (Tuchschmid and Halder 2010), which provide the consumption and emissions of the full life-cycle of a mode of transport per single kilometre travelled in Switzerland.

Finally, although not being directly related to sustainability, the frequently visited places are nevertheless included in the list of mobility features. This is due to the fact that this attribute allows for drastic changes in the personal circumstances to be detected (e.g. moving to a different city). Thus, if other indicators such as the CO_2 emissions change, but the visited places remain unaltered, this could indicate that a user is testing new travel options (e.g. taking the bicycle to work) while her circumstances remain the same. For mining the frequently visited places in a way which allows them to be compared to the results obtained for previous weeks, we choose a clustering approach. Using the DBSCAN algorithm (Sander et al. 1998), we cluster all activities found during the week. Due to the fact that although a user might have visited the same place as in the week before, the recorded activities and their associated point geometries will not correspond spatially, we choose an alternative approach and compute a minimum bounding geometry of the points based on their cluster membership. In order to avoid creating multiple instances of the “same” place in the database, the resulting polygon is tested for overlaps with already existing places in the database. If an overlap is found, no new place instance is created, but rather the id of the overlapping place in the database is extracted and stored in a list of frequently visited places for the current week. If no overlap with already existing places is detected, a new place instance is created in the database, and a new id is assigned. Figure 3 shows an example of activities and the overlapping cluster geometries from different weeks for one user. Since they all overlap, only the first occurrence would be created as an instance and assigned an id. For all the other clusters, only the information that the place has been visited frequently enough to be detected as a cluster would be stored together with its id. After computation, the results for all indicators are stored in a database (see Fig. 1).

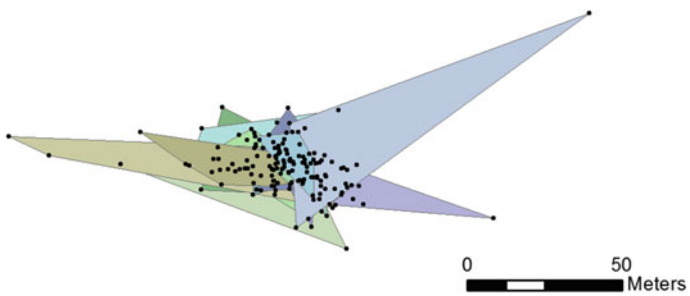

Fig. 3 The activities are shown on top of the overlapping minimum bounding polygons, as derived from the point clusters at different weeks

3.4 Anomaly Detection

At the present stage of the workflow, mobility features and patterns have been detected and stored for the current week. Now, it is possible to load similar data computed for the previous weeks from the database, and assess potential anomalies in mobility behaviour (see Fig. 1). Numerous algorithms available for anomaly detection simply classify individual data points (in our case, aggregations of all mobility features for the current week) as anomalous or normal, without allowing further insight into which feature exactly caused the data point to be classified as anomalous (cf. Chandola et al. 2009). This knowledge, however, is critical for our purposes since merely knowing that an anomaly occurred is not sufficient, but rather the results should allow deeper interpretation of the detected behaviour change. Thus, to decide which system action should be triggered as a reaction, it is critical to explicitly identify the mobility features which have changed, i.e. were detected as anomalous. For instance, an increase in bicycling distance could trigger encouraging feedback, whereas an increase in CO_2 production could lead to a discouraging response. There is work on explaining anomalies in more detail after their detection (e.g. Pevnỳ and Kopp 2014), which could therefore be used in combination with any anomaly detection algorithm. For our purpose, we found this unnecessary and rather detect anomalies for each feature individually.

For each mobility feature f_i (except the *frequently visited places*, which will be explained separately) we compute the mean μ_i and standard deviation σ_i of the n weeks preceding the week currently under investigation, where n is a tunable window size (set to 5 weeks in our tests). Comparing the values computed for the current week, it is now possible to assess if an existing deviation should be considered a normal fluctuation or an anomaly. This is controlled by another parameter λ, i.e. a feature f_i is considered anomalous if $|f_i - \mu_i| > \lambda \cdot \sigma_i$. Accordingly, if the feature re-centred around zero has a deviation larger than what can be expected given previous feature values, it is treated as anomalous. We found a value of $\lambda = 3$ to yield reasonable results.

To compute if a set of *frequently visited places* within a week should be considered anomalous, a similar approach is applied. We encode the presence of a certain place in a given week with a 1, and its absence with a 0. For every place, this results in a list of binary digits, e.g. the sequence $(0, 0, 1, 0, 1)$ encodes a place being visited in weeks 3 and 5, but not in any other week. Using this numerical representation, we can compute if the appearance of an individual place in any week should be considered anomalous or not by using a similar technique as above. However, as this results in every place being an additional mobility feature (which results in frequent cases with large number of anomalous features), we sum the number of anomalous places in every week, and perform another anomaly detection process on the resulting values. For example, a person frequently travelling for work purposes will constantly yield high numbers of anomalous places (i.e. first time visits at new places), a fact which is not particularly useful in terms of behaviour change detection. If, however, this number drops suddenly, and the visited places show a more regular pattern, it signals a behavioural change (which could be due to holidays, a job change, etc.). Summarizing anomalies in the frequently visited places as described allows us to handle them as a single mobility feature, and to report their anomalies for further interpretation by an automated system or an expert.

4 Case Study

We implemented the described method as a Python application (using a PostgreSQL database with the PostGIS extension for all spatial operations), and evaluated it on a large dataset collected over a period of three months, from approximately middle of December 2016–March 2017. 139 people used a smartphone tracking app, which passively recorded all their journeys, inferred a transport mode, and allowed them to change it in case the proposed one was wrong. The dataset consists of 52’370’797 trackpoints, which are divided into 125’759 trip legs and 71’099 trips.

Using these data, we simulated a continuous data stream by feeding data for each week subsequently into the data processing engine. Below, the results for our mobility behaviour change detection process are provided for two exemplary users. Figures 4 and 5 show the detected anomalies for these users per week. The blue dots indicate the number of anomalies for each week, while the yellow dots show the number of anomalies with regards to frequently visited places. Please note that this does not correspond to the total number of places visited by a user, but only to those that were unexpectedly visited or skipped in the respective week. Not surprisingly, the place-related anomalies are relatively more frequent in the first weeks, which is due to the cold start problem, i.e., sparse data making it difficult to assess whether a frequently visited place should represent an anomaly. Weeks which are missing values were filtered out previously, due to insufficient data completeness. For this,

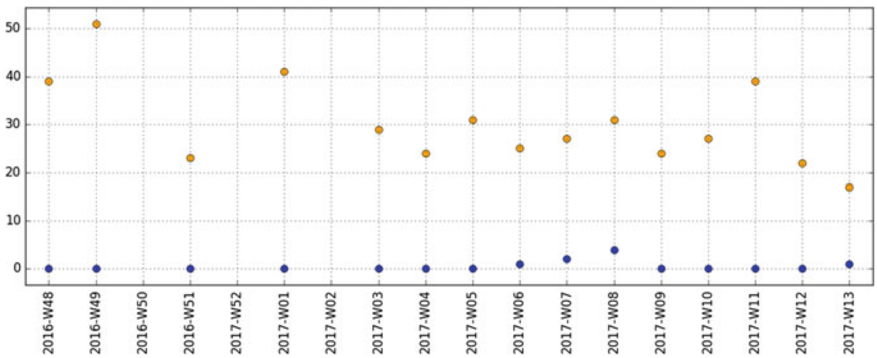

Fig. 4 All (blue) and only place-related (yellow) anomalies for user A of our test sample. In weeks 2016-50, 2016-52, and 2017-02, the data completeness was found insufficient to reliably assess mobility behaviour patterns

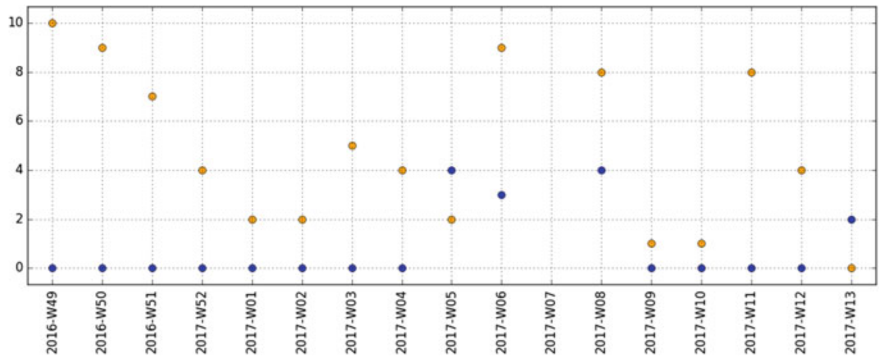

Fig. 5 All (blue) and only place-related (yellow) anomalies for user B of our test sample. In week 2017-07, the data completeness was found insufficient to reliably assess mobility behaviour patterns

we defined threshold values so that data for weeks were only further analysed if their $gdur_i \leq 0.25$ and $gdist_i \leq 0.25$.

The mobility behaviour of user A, whose anomalies are shown in Fig. 4, remains rather constant up until calender week 2017-06, where several anomalies are detected. Whereas in that week, only the average walking speed is noticeably higher compared to preceding weeks, in the following week 2017-07 we detect an increase in the distance ($\mu_d = 7.0$ km $\rightarrow f_d = 33.2$ km) and duration ($\mu_t = 19$ min $\rightarrow f_t = 1$ h 41 min) of travels made by bus. In week 2017-08, one can observe an additional increase in distance and duration of both walking (18.6 km → 58.4 km; 2 h 38 min → 9 h 43 min) and bicycling (1.9 km → 31.2 km; 5 min → 1 h 37 min). Due to the fact that in contrast to these anomalies, the frequently visited places still remain largely unchanged compared to the weeks before, we can conclude that this user indeed changed her mobility behaviour by increasingly using slow mobility (walking and bicycling) and public transport. An automated feedback system as described previously could now trigger reinforcing measures for this behaviour, e.g., by providing incentives, and thus assisting the user to transition to a phase where this new mobility behaviour is internalized and does not require further motivation.

The results for user B are shown in Fig. 5. Here, changes in mobility behaviour can be observed between weeks 2017-05 and 2017-08, which in this case, however, originate from increases in the totally travelled distance (e.g., 690 km → 1'836 km), the average speed (41.4 km/h → 97.1 km/h) the distance covered by car (307 km → 1'091 km), bike (1.1 km → 14.5 km) and walking (12.3 km → 34.1 km), as well as the related durations (plus the duration spent travelling by tram in week 2017-06). Based on the observation of such a general increase in mobility activities (not just one specific mode of transport), and set in combination with the occurrence of several place-related anomalies in weeks 2017-06 and 2017-08, one can interpret this pattern as an exceptional change of behaviour likely caused by altered personal circumstances, e.g., a holiday or business trip, rather than a gradual change of new habit formation. Indeed, when analysing the movement data for this user in more detail, we found several long distance car journeys with destinations outside of Switzerland during the respective weeks. Furthermore, in the user's home Kanton, the weeks 2017-07 and 2017-08 are usually winter holidays. This would also explain the observed data incompleteness in week 2017-07, since the smartphone tracking method deployed in this study relies on a mobile data connection, which is often unavailable when travelling abroad. In this case, an automated system reaction could be to rate the detected changes as likely temporary, and ignore them for the time being.

5 Discussion and Conclusion

In this study, we proposed a framework for continuously mining streams of movement trajectory data of users for detecting mobility behaviour changes. As it has been discussed, after data preprocessing, the completeness of the available movement recordings needs to be assessed in order to avoid misdetections of behavioural anomalies in the later steps of the analysis process. For this purpose, we presented a solution for quantifying recording gaps, hereby distinguishing between purely temporal and spatio-temporal gaps. Furthermore, we calculated a list of mobility features to serve as sustainability indicators, and proposed a method to compute and evaluate frequently visited places. Finally, the anomaly detection process was described which yields detailed results with regards to the exact mobility feature causing the anomaly occurrence. By applying the framework to a simulated stream based on a pre-recorded large-scale trajectory dataset, and evaluating the plausibility of the results obtained for two exemplary users, we could demonstrate its functionality and practical value.

In our view, this work provides a first step towards the development of personalized, automated mobility support systems which provide adaptive intervention strategies for gradually changing people's mobility behaviour towards a higher sustainability. The proposed framework, however, is not restricted to this application domain, but could be applied for other purposes as well, e.g. for general monitoring of mobility behaviour and computing descriptive statistics, or for detecting anomalies in the movements of animals or even automated vehicles or drones. A practical advantage of our approach worth mentioning is the fact that whereas the derived mobility feature values are stored for every week (feature and pattern log in Fig. 1), the actual movement data (movement data log in Fig. 1) can be deleted immediately after processing. This not only reduces the resources necessary for data storage, but also addresses privacy concerns, since the most sensitive data are deleted regularly.

There are, however, still some limitations to our approach. Thus, although the most sensitive movement data can be deleted after analysis, there still remain concerns with regards to location privacy. With mobile devices constantly gaining in computation and storage capabilities, however, a potential solution could be to shift critical parts of the analytical process to the client, and simply transmit the computed index values to the server for anomaly detection. Moreover, the list of used sustainability indicators is not exhaustive, and more complex values, e.g. incorporating car occupancy, would increase the realism with which sustainability is quantified in our study. These restrictions, however, largely depend on the quality and level of detail of the available data. Furthermore, in the exemplary application of our system, we could clearly observe problems for the first iterations due to the cold start problem, which is a usual challenge for user profiling and sequence mining applications. The usefulness of our system would therefore be reduced to a certain degree in the first phase of application. In addition, it would certainly be worthwhile to include more detailed mobility features, e.g. the usual times of travel, distinguish between the weekend and working days, or incorporate contextual information (e.g. the weather) for better results. However, special care needs to be taken for correlating features (e.g., distance and duration), as they would be flagged as anomalous in the same weeks, thus leading to a wrong assessment of behaviour change. At the same time, it can be expected that an increase in the number of features could complicate their semantic interpretation. Decision support, e.g. in the form of automated feature classification could therefore be worthwhile. Finally, due to the fact that at the current stage of this study, we have no access to ground truth data with regards to the behavioural anomalies (e.g. in the form of user interviews), a systematic evaluation of the proposed method must be regarded as future work.

Apart from testing and evaluating the model with a subset of users who can provide additional information with regards to their mobility behaviour, it is planned to refine the list of mobility features and develop a prototype of an expert system capable of interpreting the detected behavioural changes. It would also be interesting to apply a semantic perspective to the interpretation of place-related anomalies, e.g. by incorporating POI from additional data sources to assess the type of places visited.

Acknowledgements This research was supported by the Swiss National Science Foundation (SNF) within NRP 71 "Managing energy consumption" and by the Commission for Technology and Innovation (CTI) within the Swiss Competence Center for Energy Research (SCCER) Mobility.

References

Abou-Zeid M, Witter R, Bierlaire M, Kaufmann V, Ben-Akiva M (2012) Happiness and travel mode switching: findings from a Swiss public transportation experiment. Transport Policy 19(1):93–104

Andrienko G, Andrienko N, Fuchs G (2016) Understanding movement data quality. J Loc Based Serv 10(1):31–46

Axhausen KW, Frick M (2005) Nutzungen—Strukturen—Verkehr

Bamberg S (2006) Is a residential relocation a good opportunity to change peoples travel behavior? results from a theory-driven intervention study. Env Behav 38(6):820–840

Bamberg S, Rölle D, Weber C (2003) Does habitual car use not lead to more resistance to change of travel mode? Transportation 30(1):97–108

Banister D (2008) The sustainable mobility paradigm. Transport policy 15(2):73–80

Ben-Elia E, Ettema D (2011) Changing commuters behavior using rewards: a study of rush-hour avoidance. Trans Res Part F Traffic Psychol Behav

Boulouchos K, Cellina F, Ciari F, Ciari F, Cox B, Georges G, Hirschberg S, Hoppe M, Jonietz D, Kannan R, Kovacz N, Küng L, Michl T, Raubal M, Rudel R, Schenler W (2017) Towards

an energy efficient and climate compatible future swiss transportation system. SCCER mobility working paper
Brunauer R, Rehrl K (2016) Big data in der mobilität–FCD modellregion salzburg. In: Big Data, pp 235–267. Springer
Bucher D, Cellina F, Mangili F, Raubal M, Rudel R, Rizzoli RE, Elabed O (2016) Exploiting fitness apps for sustainable mobility-challenges deploying the Goeco! app. ICT for sustainability (ICT4S)
Bundesamt fuer Umwelt (BAFU), Treibhausgasemissionen der Schweiz 1990–2014
Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: a survey. ACM Comput Surv (CSUR) 41(3):15
Du Mouza C, Rigaux P, Scholl M (2005) Efficient evaluation of parameterized pattern queries. In: Proceedings of the 14th ACM international conference on information and knowledge management, pp 728–735. ACM
Feng Z, Zhu Y (2016) A survey on trajectory data mining: techniques and applications. IEEE Access 4:2056–2067
Florescu S, Körner C, Mock M, May M (2012) Efficient mobility pattern stream matching on mobile devices. In: Proceedings of the ubiquitous data mining workshop (UDM 2012), pp 23–27
Froehlich J, Dillahunt T, Klasnja P, Mankoff J, Consolvo S, Harrison B, Landay JA (2009) Ubigreen: investigating a mobile tool for tracking and supporting green transportation habits. In: Proceedings of the sigchi conference on human factors in computing systems, pp 1043–1052. ACM
Furletti B, Cintia P, Renso C, Spinsanti L (2013) Inferring human activities from GPS tracks. In: Proceedings of the 2nd ACM SIGKDD international workshop on urban computing—13. Association for Computing Machinery (ACM)
Gonzalez MC, Hidalgo CA, Barabasi A-L (2008) Understanding individual human mobility patterns. Nature 453(7196):779–782
Hamari J, Koivisto J, Pakkanen T (2014) Do persuasive technologies persuade?-a review of empirical studies. In: International conference on persuasive technology, pp 118–136. Springer
Hanson S, Huff OJ (1988) Systematic variability in repetitious travel. Transportation 15(1):111–135
Hecker D, Stange H, Korner C, May M (2010) Sample bias due to missing data in mobility surveys. In: 2010 IEEE International conference on data mining workshops, Dec, pp 241–248
Kohla B, Meschik M (2013) Comparing trip diaries with gps tracking: Results of a comprehensive austrian study. In: Transport survey methods: best practice for decision making, pp 305–320. Emerald Group Publishing Limited
Li Q, Zheng Y, Xie X, Chen Y, Liu W, Ma W-Y (2008) Mining user similarity based on location history. In: Proceedings of the 16th ACM SIGSPATIAL international conference on advances in geographic information systems, p 34. ACM
Montini L, Prost S, Schrammel J, Rieser-Schüssler N, Axhausen KW (2015) Comparison of travel diaries generated from smartphone data and dedicated GPS devices. Trans Res Proc 11:227–241
Nicolas J-P, Pochet P, Poimboeuf H (2003) Towards sustainable mobility indicators: application to the lyons conurbation. Transport Policy 10(3):197–208
Palma AT, Bogorny V, Kuijpers B, Alvares LO (2008) A clustering-based approach for discovering interesting places in trajectories. In: Proceedings of the 2008 ACM symposium on Applied computing, pp 863–868. ACM
Pevnỳ T, Kopp M (2014) Explaining anomalies with sapling random forests. In: Information technologies—applications and theory workshops, posters, and tutorials (ITAT 2014)
Polak J, Han X (1997) Iterative imputation based methods for unit and item non-response in travel surveys. In: 8th meeting of the international association of travel behaviour research. Austin, Texas
Prelipcean AC, Gidofalvi G, Susilo YO (2015) Comparative framework for activity-travel diary collection systems. In: 2015 International conference on, models and technologies for intelligent transportation systems (MT-ITS), pp. 251–258. IEEE

Prochaska JO, Velicer WF (1997) The transtheoretical model of health behavior change. Am J Health Promotion 12(1):38–48
Quddus M, Washington S (2015) Shortest path and vehicle trajectory aided map-matching for low frequency gps data. Trans Res Part C Em Technol 55:328–339
Sander J, Ester M, Kriegel H-P, Xu X (1998) Density-based clustering in spatial databases: the algorithm gdbscan and its applications. Data Mining Knowl Discovery 2(2):169–194
Schade J, Schlag B (2003) Acceptability of urban transport pricing strategies. Trans Res Part F Traffic Psychol Behav 6(1):45–61
Schlich R, Axhausen KW (2003) Habitual travel behaviour: evidence from a six-week travel diary. Transportation 30(1):13–36
Schüssler N (2008) Processing GPS raw data without additional information
Sester M, Feuerhake U, Kuntzsch C, Zhang L (2012) Revealing underlying structure and behaviour from movement data. KI—Künstliche Intelligenz 26(3):223–231
Shen L, Stopher PR (2017) Review of GPS travel survey and GPS data- processing methods. Trans Rev 1–19
Siła-Nowicka K, Vandrol J, Oshan T, Long JA, Demšar U, Fotheringham AS (2015) Analysis of human mobility patterns from GPS trajectories and contextual information. Int J Geograph Inf Sci 30(5):881–906
Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021
Souto G, Liebig T (2016) On event detection from spatial time series for urban traffic applications. In: Solving large scale learning tasks. Challenges and algorithms, pp 221–233. Springer
Stenneth L, Wolfson O, Yu PS, Xu B (2011) Transportation mode detection using mobile phones and GIS information. In: Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems—GIS 11. Association for Computing Machinery (ACM)
Stopher PR, Moutou CJ, Liu W (2013) Sustainability of voluntary travel behaviour change initiatives: a 5-year study
Sun B, Yu F, Wu K, Leung V (2004) Mobility-based anomaly detection in cellular mobile networks. In: Proceedings of the 3rd ACM workshop on Wireless security, pp 61–69. ACM
Taaffe EJ (1996) Geography of transportation. Morton O'Kelly, New Jersey, USA
Taniguchi A, Hara F, Takano S, Kagaya S, Fujii S (2003) Psychological and behavioral effects of travel feedback program for travel behavior modification. Trans Res Record J Trans Res Board 1839:182–190
Tuchschmid M, Halder M (2010) Mobitool-grundlagenbericht: Hintergrund, methodik & emissionsfaktoren. Tuchschmid und M, Halder im Auftrag von SBB, Swisscom, BKW und ÖBU
Tulusan J, Steggers H, Fleisch E, Staake T (2012) Supporting eco-driving with eco-feedback technologies: recommendations targeted at improving corporate car drivers' intrinsic motivation to drive more sustainable. In: 14th ACM international conference on ubiquitous computing (UbiComp), p 18. Ubicomp
Wagner DP (1997) Lexington area travel data collection test: GPS for personal travel surveys. Final report, office of highway policy information and office of technology applications. Federal Highway Administration, Battelle Transport Division, Columbus, pp 1–92
Weiser P, Bucher D, Cellina F, De Luca V (2015) A taxonomy of motivational affordances for meaningful gamified and persuasive technologies. In: Proceedings of the 3rd international conference on ICT for sustainability (ICT4S), ser. Adv Comput Sci Res 22, pp 271–280. Atlantis Press, Paris
White CE, Bernstein D, Kornhauser AL (2000) Some map matching algorithms for personal navigation assistants. Trans Res Part C Em Technol 8(1):91–108
Wolf J, Loechl M, Thompson M, Arce C (2003) Trip rate analysis in GPS-enhanced personal travel surveys. In: Transport survey quality and innovation. Emerald Group Publishing Limited, pp 483–498

World Business Council for Sustainable Development (WBCSD) (2015) Methodology and indicator calculation method for sustainable urban mobility. WBCSD, Geneva, Switzerland
Yang S, Liu W (2011) Anomaly detection on collective moving patterns: a hidden Markov model based solution. In: Internet of things (iThings/CPSCom), 2011 international conference on and 4th international conference on cyber, physical and social computing, pp 291–296. IEEE
Zheng VW, Cao B, Zheng Y, Xie X, Yang Q (2010) Collaborative filtering meets mobile recommendation: a user-centered approach. In: Proceedings of the twenty-fourth AAAI conference on artificial intelligence, ser. AAAI'10, pp 236–241. AAAI Press
Zheng Y (2015) Trajectory data mining. TIST 6(3):1–41
Zheng Y, Chen Y, Li Q, Xie X, Ma W-Y (2010) Understanding transportation modes based on GPS data for web applications. ACM Trans Web 4(1):1:1–1:36