Trajectory data mining: A review of methods and applications

Jean Damascène Mazimpaka and Sabine Timpf

Department of Geography, University of Augsburg, Germany

Abstract: The increasing use of location-aware devices has led to an increasing availability of trajectory data. As a result, researchers devoted their efforts to developing analysis methods including different data mining methods for trajectories. However, the research in this direction has so far produced mostly isolated studies and we still lack an integrated view of problems in applications of trajectory mining that were solved, the methods used to solve them, and applications using the obtained solutions. In this paper, we first discuss generic methods of trajectory mining and the relationships between them. Then, we discuss and classify application problems that were solved using trajectory data and relate them to the generic mining methods that were used and real world applications based on them. We classify trajectory-mining application problems under major problem groups based on how they are related. This classification of problems can guide researchers in identifying new application problems. The relationships between the methods together with the association between the application problems and mining methods can help researchers in identifying gaps between methods and inspire them to develop new methods. This paper can also guide analysts in choosing a suitable method for a specific problem. The main contribution of this paper is to provide an integrated view relating applications of mining trajectory data and the methods used.

Keywords: trajectories, mining methods, mining application problems, applications

1 Introduction

The development in the field of information and communication technology especially in mobile sensing and wireless communication is flooding us with data containing time-varying geographic positions. Though this kind of data is also associated with challenges such as exhausting our storage capacity and our data transmission bandwidth, researchers

have shown that these data sets constitute a precious resource. Their analysis can lead to solutions for important research problems in different fields such as urban planning, transportation, behavioral ecology, sports scene analysis, surveillance, and security [53].

Researchers in movement analysis have made important contributions by developing methods and tools to solve specific application problems. As researchers continue to address diverse application problems [73] we urgently need an integrated view of the application problems solved using trajectory data and the mining methods available for addressing each class of application problem. Such an integrated view would allow other researchers to easily identify the issues already solved, the methodological gaps, and to draw inspiration for new method development from the relations between problems and methods. The objective of this paper is to contribute to the provision of an integrated view of trajectory mining application problems and methods. The paper also presents a survey of applications that exploit the solutions to these application problems. Available analysis methods can be classified as computational, statistical, visual, or a combination of these. This paper focuses on computational methods of trajectory mining.

The remainder of this paper is organized as follows: after this introduction, Section 2 presents a brief overview of trajectory data and trajectory data mining. Section 3 discusses mining methods for trajectory data while Section 4 discusses application problems addressed using these methods. In addition, Section 4 relates the mining application problems to the methods presented in the paper, which are used to solve them. Section 5 reviews applications that exploit the solutions to application problems discussed in Section 4. Section 6 briefly presents open issues that need to be addressed for future development of trajectory data mining. In Section 7, we conclude and list directions for future work.

2 Trajectory data and trajectory data mining

In this section, we briefly explain the background concepts that form the basis of the research surveyed. We define the concept of trajectory data and locate trajectory data mining in the broader domain of data mining. For reasons of space, we do not discuss these concepts in detail but interested readers can follow the corresponding references indicated in this section.

2.1 Trajectory data

Based on the technology by which they are recorded, mobility data is available in different forms. Spinsanti et al. [122] differentiated GPS (global positioning system), GSM (global system for mobile communications), and geo-social network based trajectory data. Pelekis and Theodoris [112] added two other forms; RFID (radio frequency identification) based and Wi-Fi based data. GPS based data is composed of temporally ordered sequences of geographic coordinates recorded by a GPS-enabled device carried by the moving object. GSM based data is composed of temporally ordered sequences of identifiers of the cells in which the moving object passes. Geo-social network based data is content found on Internet social media and to which geographic coordinates have been attached. RFID based data contain a sequence of identifiers of RFID readers through which the moving object passed, while Wi-Fi based data contain a sequence of identifiers of access points that communicated with the moving object. Though the properties of these forms of data, e.g., their accuracy, show

that they are very different [112, 122], they have been used to address similar or related application problems using similar or related mining methods.

A trajectory can be generally formally represented as: $T=\langle p_1...p_n \rangle$ where $p_k=(id_k, loc_k, t_k, A_k)$ is the kth position, id_k is the position identifier, loc_k is the spatial location of the position, t_k is the time at which the position was recorded, and A_k is a possibly empty list of additional descriptive data (e.g., direction, occupancy status, etc.). The spatial location of the position may be represented in different ways depending on the recording technology (e.g., $loc_k= (x, y)$ for GPS-based data, and $loc_k = cell\ ID$ for GSM-based data).

2.2 Trajectory data mining

Data mining is an important step of a process, commonly known as *knowledge discovery* [37, 95] that extracts useful information from huge datasets. Data mining methods and applications have been widely surveyed in the general data mining domain. For instance, a survey of data mining methods for classical relational and transactional data can be found in [37] and [57]. Beyond the classical relational and transactional data, some researchers [100–102] provided an overview of mining tasks on geographic data. The tasks they discuss are generally analogous to the tasks in the general data mining domain but the mining methods are adapted to handle the peculiarity of geographic data such as the spatial dependency and the combination of spatial and non-spatial components. Though the tasks presented in these surveys have analogous tasks in the movement data domain, the review does not consider the temporal dimension that is inherent in movement data.

Andrienko et al. [5] proposed a conceptual framework for classifying movement analysis techniques in general. In their framework, they consider a movement analysis process to consist of one or more tasks. The types of tasks are distinguished based on the type of information sought, which is defined in terms of the elements involved, i.e., objects, space or time, and the specific target which is either characteristic or relation. They also distinguish tasks based on the analysis level, which is either elementary or synoptic. Andrienko et al. [5] also proposed a taxonomy of methods for analyzing movement data. This taxonomy focuses on the methods in the domain of visualization and interaction. The contribution of our paper extends the work of Andrienko et al. [5] into the domain of computational analysis methods.

From the applications side, Castro et al. [22] surveyed the work on mining traces of taxis. They classified the surveyed work into three categories: social dynamics, traffic dynamics, and operational dynamics. Though their survey provided an extensive literature review, it considered only the work done on taxi traces. We therefore need a survey that considers applications of trajectory data mining without limit to a specific type of trajectories. Our work differs from the survey in [22] by considering traces in general and by discussing specific application problems that are of major interest in different application fields.

We survey the work done from the application side according to the framework shown in Figure 1. As the framework shows, there are different types of moving objects that can be tracked. Tracking these objects generates a set of trajectories and some questions of which the solutions can be used in some application fields. We call these questions application problems because their solutions are not application in themselves but knowledge used in application fields. The application problems are solved using specific trajectory mining methods. For instance, if moving people are tracked we get their trajectories and some
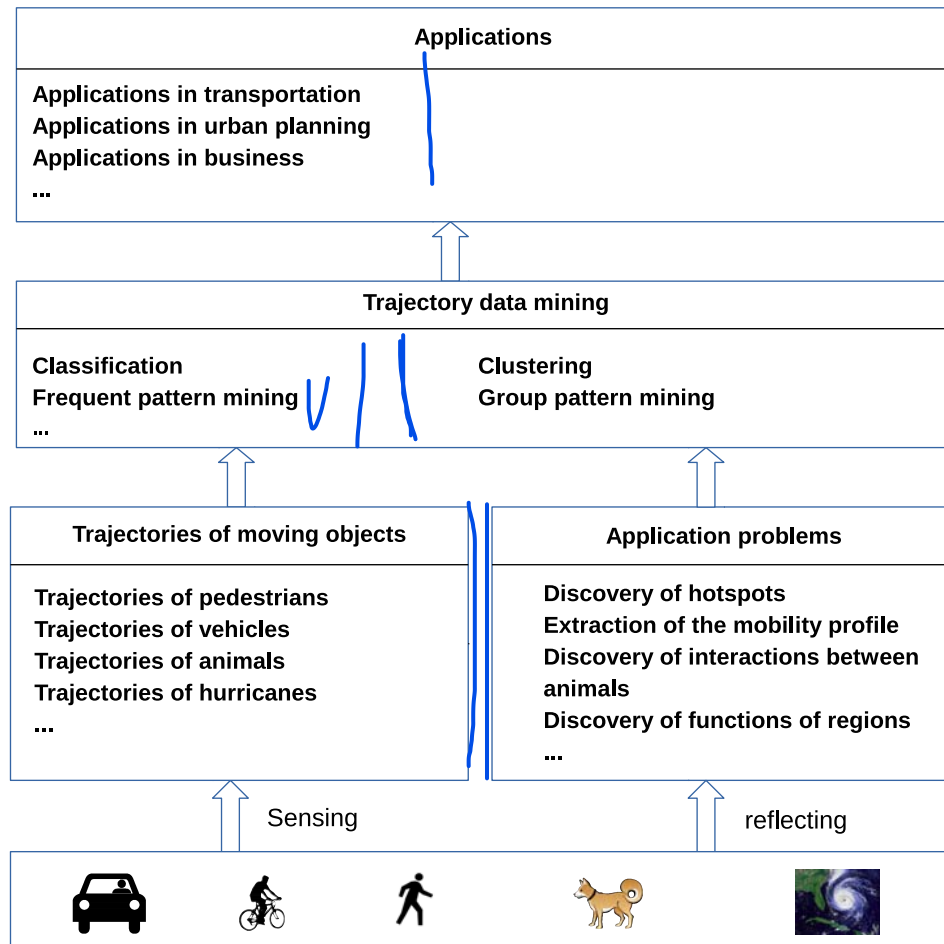
Figure 1: General framework of application-driven trajectory data mining.

questions come to our mind regarding, for instance, the characteristics of these people or the geographical space where they move. By mining these trajectories we may discover places where many people stop frequently and stay for a considerable long duration. These places are characterized as hotspots and the knowledge about hotspots can be used, for instance, in business for advertisement, in public security for crowd monitoring, and in transportation for taxi pick-up point recommendation. Therefore, hotspots discovery is an application problem of which the solution can have various applications.

The core task at the mining step is the application of a specific mining method. This is preceded by a trajectory data pre-processing step, which prepares the data for mining. This step includes tasks such as data cleaning, trajectory compression, map matching, and trajectory segmentation. Data cleaning aims at filtering noise from the data while trajectory compression aims at reducing the amount of trajectory sample points to reduce the volume of data to be processed. Map matching aims at adjusting trajectory sample points to the roads in case the movement follows a road network while trajectory segmentation

divides a trajectory into meaningful sub-trajectories required for subsequent operations. A discussion of trajectory pre-processing tasks can be found in [150].

3 Trajectory mining methods

Similar to the general domain of data mining, trajectory data mining aims at discovering interesting knowledge (e.g., patterns) from the data and it has two primary goals: prediction and description. Prediction consists in using some variables in the data to determine unknown or future values of other variables of interest, while description focuses on finding human-interpretable structures describing the data. We do not try to discuss all existing mining methods; instead we review generic ones and identify the relationships between them. By generic mining methods we mean mining methods applied on a wide range of trajectory data in a wide range of applications. We divide the methods discussed into two categories: primary methods and secondary methods. Given a trajectory dataset, primary methods aim at categorizing the trajectories based on their properties. On the other hand, secondary mining methods are composite methods in the sense that they apply a sequence of primary mining methods or classical statistical methods or a combination of the two.

The discussion of the methods focuses on the objective of the method, its general working principle and the data granularity level at which it works (whole trajectories versus sub-trajectories, single trajectory versus multiple trajectories, and trajectories of single moving object versus trajectories of multiple objects). The discussion ends with an attempt to identify relationships between different methods.

3.1 Primary mining methods

Primary trajectory mining methods are used to categorize the trajectories based on their properties. The application of a primary mining method followed by the interpretation of its results may be the sole intended analysis or the primary mining method may be used to prepare the data for an extended analysis possibly with other mining methods. Primary trajectory mining methods include trajectory clustering and trajectory classification.

3.1.1 Clustering

Trajectory data clustering aims at describing a trajectory dataset by grouping the trajectories into a finite set of categories, also called clusters, based on their movement characteristics. The trajectories in the same cluster exhibit movement characteristics that are similar and different from those of trajectories in other clusters.

There have been efforts to develop trajectory-specific clustering approaches mainly by adapting statistical and probabilistic models to account for the characteristics of trajectories. For example, the mixture models based clustering approach proposed by Gaffney and Smyth [42] groups together trajectories that are likely to be generated by a common representative trajectory by adding Gaussian noise. In the same line, Alon et al. [2] model trajectories as sequences of transitions between positions and use a hidden Markov model (HMM) that best fits the trajectories to model a cluster.

Though trajectory-specific clustering approaches exist, state-of-the-art clustering algorithms for trajectories are extensions of traditional clustering algorithms through a proper definition of similarity (or distance) functions. A survey of traditional clustering algorithms

can be found in [57]. Depending on the goal of analysis some similarity (or distance) function such as similar route, similar destination, similar source, similar route and destination, or similar direction is used to determine which trajectories belong to the same cluster. A detailed discussion of how distance and similarity functions are applied to determine cluster membership can be found in Rokach [119].

Among the algorithms that have been extended, two well-known can be mentioned: DBSCAN (density based spatial clustering of applications with noise) [36] and OPTICS (ordering points to identify the clustering structure) [6]. As examples of such extensions, T-OPTICS [106] was developed as an extension of OPTICS by defining a spatio-temporal distance for comparing and clustering trajectories while ST-DBSCAN (spatial-temporal DB-SCAN) [14] uses two parameters for similarity measure to improve the identification of clusters and noise.

Several clustering algorithms have been developed and categorized in different ways (e.g., see in [40, 57, 58]) based on different features but a crisp boundary between the categories is difficult to draw. Most of the categorization approaches identify partitioning, hierarchical, and density-based algorithms. Partitioning algorithms group all objects into a pre-specified number of clusters. These algorithms start from a random partitioning which is then refined through iterations that may move an object from one cluster to another. K-means [94] is an example of such algorithms. Hierarchical algorithms organize the objects in a multilevel structure of clusters and sub-clusters. Using loose proximity requirements, the algorithms find clusters at a higher level while by tightening the proximity requirements they find sub-clusters. An example of such an algorithm is the BIRCH (balanced iterative reducing and clustering using hierarchies) [147]. Density-based algorithms partition the objects into clusters based on their density, i.e., starting from one object, the cluster grows as long as new objects exist in the neighborhood, and the cluster is considered valid if its total number of objects exceeds a threshold. DBSCAN is an example of algorithm in this category.

Trajectory clustering works on a set of trajectories taking each trajectory in turn and assigning it to a cluster. However, there are some cases in which the input to the clustering operation is not a set of trajectories but only one trajectory. In these cases, the objective is to cluster the points of the single trajectory with the aim of characterizing some positions on it. This case is exemplified in Palma et al. [109] where the points on the input trajectory are clustered to discover stops on it.

Trajectory clustering can be applied on either whole trajectories or sections of trajectories depending on the goal of analysis and the similarity function applied. For example, if the similarity function is defined as similar origin and similar destination, the trajectories can be clustered as a whole because the route followed is not important. An example of this case can be seen in [43] where the interest is on the overall directions of extra-tropical cyclone trajectories for obtaining clusters such as *south-to-north oriented* and *west-to-east oriented*. On the other hand, if the interest is on different locations traversed by the trajectories, for example the similarity is defined as having visited the same types of places; the clustering is applied on sections of trajectories. The TraClus clustering algorithm [79] follows this approach.

3.1.2 Classification

The objective of classification is to find a rule to assign objects into pre-defined classes. The knowledge about class assignment is available in the form of a set of predefined classes and a sample set of objects already labeled with the class they belong to. This sample is called a training set. Similar to classification in general, trajectory classification aims at determining the class label of trajectories from a predefined set of labels based on the features of the trajectories. For example, the trajectory classification may be aimed at labeling each trajectory from a large set with its transportation mode, given a small set of trajectories manually labeled with the transportation mode used.

Most trajectory classification algorithms follow a traditional two-step approach: first extracting a set of discriminative features and then using the extracted features to train an existing standard classification model. The first step involves identifying the properties of the trajectories that are likely best for identifying the class to which each trajectory belongs, and determining the values of the properties which are then represented in a specific form (e.g., a vector). For example, properties such as the average speed of the trajectory, average acceleration, trajectory duration, and length can be used. The discriminative power of the considered feature depends on the types of classes at hand. For example, we may consider that the acceleration has a much higher discriminative power than the trajectory length for the vehicle type classes. The second step selects a standard classification model and applies it to the extracted features.

A detailed discussion of the traditional classification algorithms can be found in [57]. Among these methods, the following have been commonly applied in trajectory classification. The decision tree algorithm has been used by Zheng et al. [152] to classify trajectories into different transportation modes. They first segment the trajectories and extract discriminative features such as average velocity of a segment, heading change rate, and velocity change rate, which are then fed into a decision tree based inference model for classification. Support vector machines (SVMs) were used by Bolbol et al. [15] to solve the same issue of transportation mode classification. Through a statistical evaluation, they first analyze the discriminative power of four features (speed, acceleration, distance, and rate of change in heading) on six transportation modes (bus, car, cycle, train, underground, and walk). After identifying the speed and acceleration as the best discriminative features, they applied the SVM algorithm on these features to study equal-sized instances of several segments of trajectories for classifying them.

In many of its applications, trajectory classification is performed after other analysis operations such as segmentation and clustering which prepare the features on which the classification will be based. For example, the TraClass framework [78] applies trajectory segmentation and clustering to extract region and sub-trajectory features that are then used for a support vector machine (SVM) based classification of the trajectories.

3.2 Secondary mining methods

While primary trajectory mining methods aim at categorizing the trajectories, secondary mining methods generally aim at analyzing the spatial, temporal, or spatio-temporal arrangement of the individual trajectories, within their categories or between categories. Several mining methods can be put in this category, but we discuss in this section a few generic ones: pattern mining, outlier detection, and prediction.

3.2.1 Pattern mining

Trajectory pattern mining aims at discovering and describing the movement patterns hidden in trajectories. It provides information about when and where the pattern occurs, and the entities involved in it. A large number of types of movement patterns have been reported in the literature. An integrated view of them can be found in the survey done by Dodge et al. [35]. Likewise, several methods have been developed for mining these patterns. The methods can be put into three categories: repetitive pattern mining, frequent pattern mining, and group pattern mining.

Repetitive pattern mining applies basically on trajectories of a single moving object while frequent pattern mining applies on trajectories of multiple moving objects but in a relative time; that is, the objects may not move at the same time and the only condition is that they visit approximately the same places in the same sequence. Similar to frequent pattern mining, group pattern mining applies to trajectories of multiple objects but the movement is considered in an absolute time; that is, the objects move together.

Repetitive pattern mining concerns regular movement patterns such as the movement of a commuter, which is repeated every day, or the movement of a migratory bird, which is repeated every season. A repetitive pattern is also called a periodic pattern because the object follows approximately the same route after an approximately constant time period. The discovery of periodic patterns is complicated by the approximate nature of the patterns in terms of space and time; i.e., the object does not visit exactly the same location at corresponding time instants of the period and the period does not exactly have the same value in different cycles.

A common approach in discovering periodic movement patterns is to apply the mining on sequences of locations. In an earlier study, Cao et al. [20] required the period to be specified as input and clustered the locations corresponding to the same offset of the period to discover frequent regions. They then iteratively combined discovered frequent regions to get complete periodic patterns. The work presented by Cao et al. [20] extended the approach presented in a previous study [97], but both studies required the period as a user input to the mining process. The *Periodica* algorithm [85] overcame this undesirable requirement to specify the period in advance. This algorithm first detects automatically the period using an approach combining Fourier transformation and autocorrelation. The algorithm finds reference spots, which are dense regions containing more trajectory points than others, and detects the different periods in each. The algorithm then discovers periodic patterns from movement sequences between reference spots with the same period using a hierarchical clustering for which the distance measure is based on a probabilistic model. Compared to most of previous related work, the *Periodica* algorithm presents two features that address the unknown and varying period issue commonly found on real life movement data: automatic period detection and the consideration of different periods during the periodic pattern discovery.

Frequent pattern mining is about extracting (parts of) routes that have been frequently followed by the moving objects in the trajectory dataset. Frequent trajectory patterns can be defined using spatial or spatiotemporal characteristics of the trajectories [68]. The definition based on spatial characteristics considers only the sequence of the locations visited. The *frequent spatiotemporal sequential patterns* [19] and the *generalized sequential patterns* (GSP) [108] are examples of this case. The definition based on spatiotemporal characteristics considers, in addition to the sequence of the locations visited, the transition time between the locations. The *T-Patterns* (or simply *trajectory patterns*) [49] are an example of this
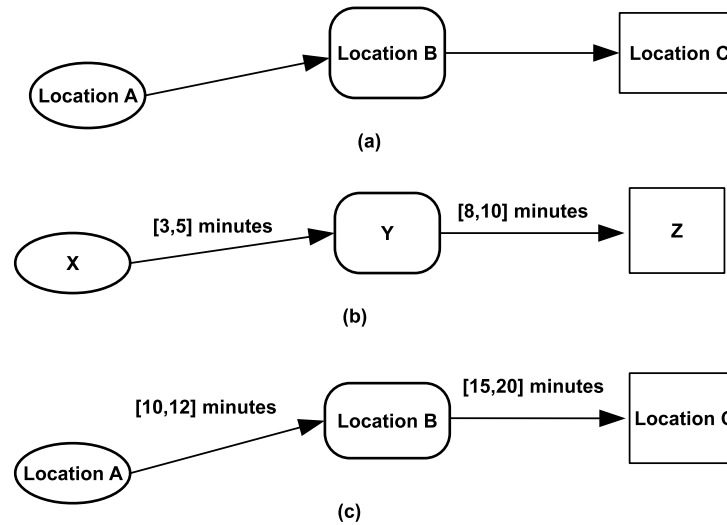
Figure 2: (a) A spatiotemporal sequential pattern, (b) a temporally annotated sequence (TAS), and (c) a T-pattern.

second case. Giannotti et al. [49] defined a T-pattern as "a set of individual trajectories that share the property of visiting the same sequence of places with similar travel times.". T-patterns evolved from the *temporally annotated sequences* or *TAS* introduced by Giannotti et al. [47]. In TAS, the elements of a sequence are generic events, e.g., purchases or logged web accesses, without any specific spatial information. Figure 2 shows three types of sequences that make different frequent patterns: for spatiotemporal sequential patterns (a) only the sequence of locations is considered while for T-patterns (c) the transition time between locations (e.g., from location A to location B it takes between 10 and 12 minutes) is also considered. T-patterns (c) combine the feature of location sequence from spatiotemporal sequential patterns (a) with the feature of transition time from TAS (b).

A common approach to mining frequent patterns consists in finding important regions from the trajectories and then applying sequence mining on a temporally annotated sequence of these regions [49, 65]. In this line, Giannotti et al. [49] developed two approaches for extracting T-patterns. The first approach involves a discretization of the space to identify the regions of interest. Then, the actual mining step applies the method introduced in Giannotti et al. [47] to mine sequences of regions of interest with temporal annotations. The second approach dynamically derives the regions of interest from trajectory segments and then translates the trajectories into a set of sequences of regions from which it iteratively builds T-patterns.

Group pattern mining aims at extracting movement patterns involving groups of objects that move together. A general condition is that the objects that form a group stay close in space for a considerable period of time. Several group patterns and their variants have been defined based on the general condition of spatiotemporal closeness, the internal structure of the group and the characteristics of group members. The mostly studied patterns are flocks [12, 130], convoy [63], and swarm [84]. Figure 3 shows examples of these three patterns.
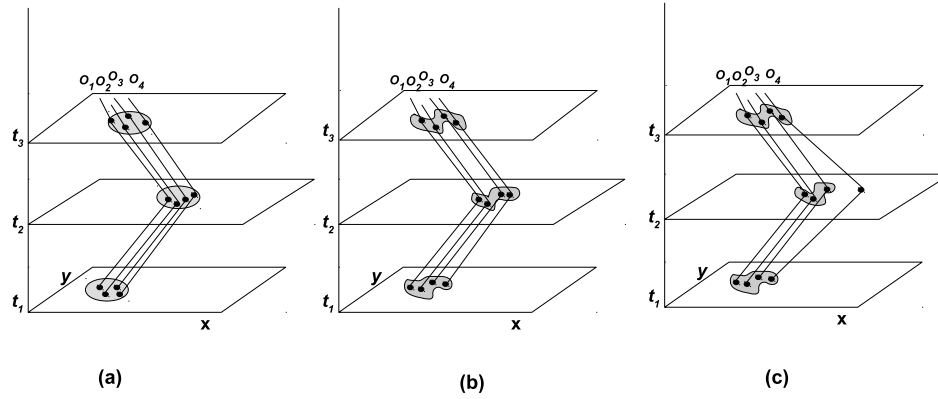
Figure 3: Examples of group pattern: (a) Flock, (b) Convoy, and (c) swarm.

A flock is a group of at least *m* objects that travel together for at least *k* consecutive timestamps such that at any of these timestamps they are found within a disc of radius *r*. The pattern is defined by the three parameters *m*, *r*, and *k*.

The convoy pattern relaxes the disc shape constraint to represent a group of moving objects forming any shape. It is a group of at least *m* objects traveling together for at least *k* consecutive timestamps such that at each of these timestamps the group can be found using a density-based clustering with parameters *d* as the neighborhood distance and *m* as the minimum number of objects.

The swarm pattern is an extension of the convoy pattern, which further relaxes the constraint of consecutive timestamps. It is defined by the same parameters as the convoy except that in a swarm *k* is the minimum number of timestamps at which the group is found irrespective of whether the timestamps are consecutive or not. For example, with *m=4* and *k=2* the pattern in Figure 3 (c) is not a convoy because only three objects (O_1, O_2, and O_3) are found in the group for at least two consecutive timestamps. However, this pattern is a swarm because at two timestamps (t_1 and t_3) the group includes the required minimum of four objects and the swarm pattern does not require the timestamps to be consecutive.

The category of group patterns also includes patterns in which the members of the group have some interaction in which each member has a specific role. An example of such pattern is the *leadership* pattern [4] in which the group comprises a leader which moves ahead and followers for the time duration of the pattern. Another example is the *chasing* pattern [31] in which the object moving ahead aims at escaping the follower, which tries to reach it. Different examples of group patterns can be found in [76] under the general concept of relative motion (REMO).

Common approaches to group pattern mining involve clustering methods and checking the condition on parameters that define the pattern such as the minimum number of group members and the minimum duration of the pattern. For example, density-based clustering has been applied in [129] to mine flocks, in [62,63] to mine convoys, and in [86,139] to mine swarms. For swarms, the analysis starts by applying a clustering method at the level of a single timestamp, and then links spatial clusters belonging to different timestamps but sharing an appropriate fraction of objects regardless of their temporal distance.

3.2.2   Outlier detection

The objective in trajectory outlier detection is to find trajectories that do not comply with the general behavior of the trajectory dataset. While pattern mining focuses on patterns that are common in the trajectory dataset, outlier detection focuses on rare patterns (e.g., following a path different from the common path followed by most of the other trajectories).

Similar to trajectory classification, trajectory outlier detection can be performed by mining either whole trajectories or parts of trajectories. The work presented in [146] is an example of outlier detection applied on whole trajectories, while examples of outlier detection applied on sub-trajectories are found in [77, 88, 140]. For outlier detection applied on parts of trajectories, each trajectory is first segmented into sub-trajectories and then some distance function or clustering method is used to detect outlying sub-trajectories. Finally, the trajectories to which the detected outlying sub-trajectories belong are identified as outliers. This approach has been followed in the TRAOD algorithm [77].

Outliers often appear as a by-product of other mining methods especially clustering. Because of this, most outlier detection approaches in the literature adopt some clustering algorithm and identify as outliers those objects that are not found in any cluster. Hence, the main issue is finding the appropriate distance measure that will discriminate these deviating trajectories.

A basic approach for detecting trajectory outliers analyzes the neighborhood of each trajectory by counting the number of neighbors or using a density-based clustering method. In this approach, the trajectories that have too few neighbors are categorized as outliers. The iBAT [146] anomaly detection framework follows an extended variant of this approach. In iBAT, the study area is partitioned into a grid structure and the trajectories crossing the same source-destination cell-pair are grouped. Finally, the approach uses an isolation mechanism to identify outlying trajectories exploiting the property that they are *few and different* in their groups.

Another category of approaches to trajectory outlier detection follow the procedures used in classification methods. For instance, a set of predefined features can be extracted from the trajectories, and then a standard distance measure applied on the extracted vectors. This approach was followed in [140] where distance measures have been applied on four features (direction, speed, angle, and location) to discover trajectory outliers. Alternatively, trajectory outliers can be discovered by training a two-label classifier model in which one label corresponds to normal trajectories while the other corresponds to abnormal trajectories according to the features under consideration. This approach was followed in [82].

3.2.3   Prediction

Prediction using trajectory data mainly aims at guessing the future location of a moving object based on existing trajectories. It has been especially motivated by the fast growing development of location-based services, one of its major application areas [50]. The objective of most of the studies on trajectory based prediction is to predict a location (destination or next location), but there are some studies which aim at predicting the entire route generally based on a road network. Trajectory based location prediction applies to a set of trajectories.

The studies on location prediction are based on two approaches (Markov models and sequential rules or Trajectory patterns) and can be classified into three categories: i) those considering only the data of the concerned moving object [50, 61, 71], ii) those considering only the data of other moving objects [9], and iii) hybrid approaches; i.e., considering both the data of the concerned object and those of other objects [104, 105, 137]. Studies based on Markov models use probabilistic models for location prediction. Examples of work following this approach can be found in [7, 98]. Studies based on trajectory patterns mine frequent patterns and association rules by defining a trajectory as an ordered sequence of time-stamped locations, and then using sequence mining methods. Examples of work following this approach can be found in [104,136,137]. The WhereNext approach [104] extracts T-patterns and builds a decision tree in which root-to-node paths are used for prediction. Ying et al. [136] use a similar approach extracting trajectory patterns to capture the movement behaviors that are motivated by geographic, temporal and semantic intentions from trajectories of users, and then matching the current movement of a user to the extracted patterns.

Unlike location prediction, which results in a location, the route prediction must produce a sequence of route segments to a location. The common approaches to route prediction fall in the following three categories:

- *Trip observations*: The approach involves observing several trips from a driver. For a new trip, the approach considers the start point and possibly a short starting section of the new trip, attempts to find a good match with a previous trip, and then uses the matched trip as a route prediction [41].

- *Markov model*: The approach builds a probabilistic model from the long-term trip history of the driver for predicting short-term route. The model looks at the driver's just-driven path and predicts the next road segment [69].

- *Turn proportions*: The approach builds a route by predicting which ways a driver will turn at successive intersections. For instance, Krumm [70] combines the information on which proportion of drivers will choose each turn option at an intersection and the driving times between pairs of road segments to make the turn prediction. This approach is based on the assumptions that drivers tend to choose the road that will give them more destination options and that they take the most efficient route in terms of time to their destination.

3.3   Relationships between trajectory mining methods

In this section we present the relationships between the different mining methods we have discussed. Table 1 presents these relationships showing primary methods and secondary methods as headings of rows and columns respectively. The secondary mining method at the top of a column uses the primary mining method on the left of a row to perform the task found in their intersection cell. For instance, the outlier detection method may apply clustering to group similar trajectories or sub-trajectories, which may be a required step for detecting trajectory outliers. In addition, an example of this relationship may be found in the reference entry [77]. The empty table cell means that we have not seen in the literature where pattern mining uses classification.

		Secondary methods		
		Pattern mining	Outlier detection	Prediction
Primary methods	**Clustering**	• Grouping periodically related locations [20, 85] • Grouping close trajectories [62, 129] • Extraction of places of interest for frequent pattern mining [49] • Aggregating close locations for sequential pattern mining [45]	• Grouping similar trajectories or sub-trajectories [77]	• Grouping similar users [137] • Grouping similar trips of a user [3, 41] • Grouping user's stay points for building trajectory patterns [136] • Grouping visited locations for building periodic patterns [61]
	Classification		• Categorization into normal and abnormal trajectories [82]	• Categorizing on-going trajectory to one of defined trajectory clusters [3], or trajectory patterns [104]

Table 1: Relationships between trajectory mining methods.

4   Trajectory mining application problems

In this section, we discuss different application problems that have been or can be solved by mining trajectories. We categorize the application problems and relate them to generic trajectory mining methods discussed in previous sections. It is important to note that there is no one-to-one mapping between the methods and application problems. Different application problems may be addressed using the same method and one application problem may use more than one method; it depends on the specific tasks involved in addressing the application problem as shown in Table 2. The table relates application problems to the methods discussed in previous sections and provides references on where the application problem has been addressed using the methods mentioned.

4.1   Characterization of moving objects

The aim is to analyze the trajectories of moving objects for deriving some structure or information that describes the moving objects. Several instances of this application problem have been studied such as inferring activities of humans [117] or animals [120], determin-

Application problems	Application problem instances	Main tasks	Mining methods
Characterization of moving objects	Inferring activities, inferring transportation modes	Extraction of characteristic features and categorization	Classification [13, 15, 117, 120, 152]
	Human mobility profiling, detecting animal movement patterns	Extraction of sequences of places frequently visited	Frequent pattern mining [[26, 27] clustering [21]
Discovery of social relationship	Detection of social ties between two individuals, community discovery	Grouping based on stay locations of individuals	Clustering [44]
		Extracting temporally ordered sequences of stay locations	Frequent pattern mining [44]
	Discovery of interaction between animals	Finding moving animal groups, evaluating temporal change in group structure	Group pattern mining [54, 86]
Characterization of places or regions	Discovery of land use type or socio-economic function	Extraction of regions	Clustering [92, 111]
		Extraction of characteristic features and categorization	Classification [92, 111]
	Discovery of hotspots	Grouping based on spatial and/or temporal characteristics of the data	Clustering [89, 132]
	Detection of traffic jams	Finding moving car groups, speed-based filtering of groups	Flock pattern mining [107]
Discovery and characterization of the connectivity between places or regions	Connection discovery, connection strength estimation, mobility-based region delineation	Origin-Destination mapping, deriving link statistics, grouping	Clustering [87, 118]
		Separating normal links from abnormal links	Outlier detection [39]

Table 2: Trajectory mining application problems and corresponding methods.

Application problems	Application problem instances	Main tasks	Mining methods
Detection and recognition of social events	Discovery of social events	Grouping based on spatio-temporal properties	Clustering [148]
	Recognition of discovered social events	Extraction of characteristic features and categorization	Classification [18]
Trajectory-based prediction	Location prediction, route prediction	Building a probabilistic model of the mobility, comparing the current trajectory to the model	Prediction based on statistical methods (e.g., Markov Models) [7, 23, 69, 98]
		Extraction of ordered sequences of visited locations, comparing current trajectory with historical sequences	Frequent pattern mining (e.g., T-patterns, sequence analysis) [104, 136, 137]
	Prediction of traffic-related events (e.g., traffic jams)	Extraction of trajectory density and examining its variation over time on different paths	Frequent pattern mining (e.g., T-patterns) [48]
Trajectory-based recommendation	Place recommendation	Extracting sequences of locations visited, popularity-based ranking of locations, and selecting top popular ones	Frequent pattern mining (e.g., T-patterns) [81, 134, 156]
	Route recommendation, friend recommendation	Extracting sequences of locations visited, ranking them based on their similarity with those of the current user, selecting most similar ones or the owners of the most similar ones	
	Trip recommendation	Extracting sequences of locations visited, extracting travel preferences in terms of stay at locations and transition between them, selecting sequences most matching the current user's preferences	Frequent pattern mining (e.g., T-patterns) [16, 155]

Table 2 (continued): Trajectory mining application problems and corresponding methods.

ing the mobility profile such as human mobility routine [26, 27] and animal movement patterns [21, 30, 86], and inferring transportation modes [15, 152].

There has been a high interest in understanding the behavior of individuals by *inferring the activities* they performed or the *transportation mode* they used during their movements. Researchers have addressed this characterization problem instance commonly using classification methods. Another common application problem in the direction of understanding human behavior is *determining human mobility profiles*. A common approach to this problem applies frequent sequential pattern mining to extract the sequences of places that the user frequently visits. A similar problem of understanding the mobility pattern has been studied on animals. A common approach to *detecting animal movement pattern* uses various movement pattern analysis methods especially methods based on clustering. For example, Carneiro et al. [21] applied different clustering methods to study the migration habits of white storks. In particular, they detected flocks by using fuzzy c-means clustering technique.

4.2   Discovery of social relationships

An aim is frequently to analyze the trajectories of moving objects to discover the existence of some interaction between the objects and the properties of this interaction. Instances of this application problem include the detection of social ties between two individuals, community discovery, and the discovery of interaction between animals (e.g., the predator-prey interaction).

The success of online social media platforms in connecting people and encouraging them to share their mobility history has stimulated researchers to study social relationships from mobility data. Some researchers addressed the problem of discovering social relationships based on co-presence [28], others based on visits to same places [44, 131]. Co-presence, defined as being in the same place at the same time, has been studied by Crandall et al. [28]. Using a probabilistic model, they found that when two people exhibit multiple co-presences, it strongly indicates the existence of a social tie between them. Wang et al. [131] used a probabilistic model to study the similarity in mobility patterns of two individuals and its correlation with their social ties.

A community is a group of entities that share some common interests. In the context of this paper, the common interests are shown by the existence of interaction between the entities, which is in turn detected from their mobility behavior. A common approach to community discovery applies clustering methods. For instance, considering that the stay-locations of individuals during their movement correspond to their interest, Gaito et al. [44] extracted the *geo-locations* as clusters of the stay-locations of individuals and then performed a density-based clustering of the overlapping geo-locations to discover the communities. They further detected interactions between communities using a sequential pattern analysis.

4.3   Characterization of places or regions

Characterization of places or regions aims to analyze the trajectories of moving objects that visit places and regions under study to derive some structure or information that describes these places and regions. This application problem has been studied under many instances

such as discovery of the semantics of regions in the form of socio-economic functions, the discovery of city hotspots, and the discovery of habitats of particular animals in ecology.

The *discovery of land use types or socio-economic functions* in a city is commonly performed by applying some clustering followed by a classification. Clustering extracts regions, while the classification assigns land use types or functions to the regions. For example, Liu et al. [92] derived urban land use types (commercial, industrial, residential, institutional, and recreational) and the dynamics of urban land use changes by mining taxi trajectory data. They extracted the temporal distributions of both pick-ups and drop-offs in the pixels of the study area and then used the k-means clustering method to classify all pixels in the study area into different classes of traffic source-sink areas. A classification tree method was then used to quantitatively examine the association of traffic source-sink areas with land uses. For the same problem, Pan et al. [111] applied a different clustering method, a modified version of DBSCAN algorithm (iterative DBSCAN), to extract regions based on the characteristics of the data. They then applied four different classification techniques to classify the regions into different social functions (land use types) based on the pick-up and drop-off dynamics. Yuan et al. [141] address the same problem with a different approach, a topic-based model applied on taxi trajectories and points of interest, and go further to estimate the functional intensity for each functional regional using a kernel density estimation model. Though this work takes a different approach, it still applies a k-means clustering to aggregate regions with the same function.

The *discovery of city hotspots*, which are regions in a city with high and regular intensity of activity, is commonly performed using clustering methods. Chang et al. [132] define hotspots as areas with a high number of taxi requests. They first filter trajectories to keep those with the same context (weather condition, day of the week, and hour of the day). Then, they apply a clustering on GPS points corresponding to passenger pick-up points. A hotness index is finally calculated based on the property of the clusters and the distance from the taxi to the cluster. Liu et al. [89] define hotspots based on the crowdedness of areas. Similar to other hotspots discovery studies, they use clustering approach. However, considering that taxis used in the study are a biased sample of vehicles and that taxis are highly mobile, they propose a new clustering approach: *mobility-based clustering*. This approach is based on a crowdedness function, which is computed taking into account the instant speed of the taxis in the area based on the consideration that low speed may be a good indicator of high crowdedness and vice versa.

The *detection of traffic jams* is another instance of region characterization. The common approach to addressing this problem involves discovering flock patterns, which are then filtered based on the speed to discover areas of slow flock patterns [107].

4.4   Characterization of the connectivity between places or regions

Discovering and characterizing the existence of a connection between places or regions can be based on the trajectories of objects moving between them. Instances of this problem include discovering connected regions, estimating the strength of the connection between regions, and delineating regions based on the connection between locations within them.

The problem is commonly addressed by deriving an origin-destination mapping in the form of origin-destination (OD) matrix, deriving some statistics and making groupings using clustering methods. For example, Liu et al. [87] analyzed taxi GPS traces and the check-in and check-out data produced by travel smart cards to discover the connections

and connection strength between different locations in a city on different days. To this end, they derived an OD matrix using clustering techniques. As an extension to the discovery of a connection between regions, the interest may be on the discovery of abnormal connections such as unexpected routes between two regions. This problem instance is commonly addressed through outlier detection methods, which identify and distinguish normal connections from abnormal ones [39]. Once the abnormal connection is discovered, a further analysis can be performed to identify the cause of abnormality and the location of this cause, which becomes a place characterization problem [110].

Other studies went further, from discovering the connection between locations to delineating regions based on human mobility. The common approach in such studies is to first build a network between locations based on human interaction, and then use some community discovery technique to partition the network into regions. The later step commonly involves clustering methods, which find location groupings with denser interaction between locations in the group than between groups. The work of Rinzivillo et al. [118] is an example addressing this problem instance by building a network representation of human movement from vehicle GPS tracks and then extracting regions using a hierarchical clustering algorithm.

4.5 Detection and recognition of social events

Detection and recognition of social events aims to analyze the trajectories of moving objects in order to detect the existence of a gathering. The analysis may go further to recognizing the type of event. The two common approaches used to address this problem are statistical methods combined with a classification, and clustering-based methods.

Statistical methods are used to discover a high concentration in a specific area and during a specific time interval of stationary objects that were initially moving [48]. Once the presence of events is detected, a further analysis can be performed to recognize their types. In this direction, Calabrese et al. [18] analyzed the origins of event attendees from cellphones mobility data to recognize the types of events. They used a classification on feature vectors representing the distribution of attendees' origins to determine the type of events.

Clustering-based approaches extend the standard clustering to discover groups of individuals that are not only dense but also of which the shape and location generally do not change and the members can enter and leave at any time provided there are some members who commit to the group for a certain time period. In this direction, Zheng et al. [148] apply a density-based clustering to discover snapshot clusters, which are then analyzed across a number of timestamps in terms of their locations and members to extract valid gathering patterns.

4.6 Trajectory-based prediction

This problem aims at analyzing the trajectories of moving objects to make some prediction, such as the next location of the moving object, its destination, the route it will take, and the occurrence of traffic related events such as traffic jams. The commonly used methods for trajectory-based prediction are based on Markov models [7,69,98] or movement pattern mining [104,136,137].

Giannotti et al. [48] predict traffic congestions by predicting future areas of dense traffic. The location of areas with a high density of trajectories is predicted using a tree structure of

T-Patterns, which are patterns capturing a sequence of visited locations and corresponding transition time. The support value at each node of the tree, which is the number of T-patterns having the path from the root to the current node as prefix, makes it possible to quantify the trajectories on different paths which is analyzed over time for prediction. Another approach applied to the prediction of traffic density is based on Markov model. This approach was adopted by Castro et al. [23], who built a prediction model using the probabilities of transition between road segments. In order to predict traffic congestion, they compare the predicted traffic density and the capacity of the road segments, which they determine based on the historical traffic density and speed readings.

4.7 Trajectory-based recommendation

Trajectory-based recommendations are enabled by the similarity between people inferred from their mobility histories. The similarity between people is commonly inferred by determining and comparing their mobility profiles. The assumption behind these recommendation systems is that people who have the same mobility behavior are likely to share interests and preferences, and hence are potential friends of each other or are likely to visit the same types of places and follow the same types of routes. For a given individual, once similar users are found they can be recommended to him or her as friends. Alternatively, the places that they visited and the routes they used can be recommended to the individual as places to visit and routes to follow, respectively.

The common approach used to address the trajectory-based recommendation problem is to apply a frequent pattern mining to extract and compare sequences of locations visited. In this direction, Li et al. [81] proposed a general architecture for a friend and place recommender system. The architecture comprises three processes: location history representation, user similarity exploration, and friend and location recommendation. For each user, they perform a hierarchical clustering of stay points to obtain his location history, which is then represented as a hierarchical graph. In the user similarity exploration, they retrieve a sequence of shared graph nodes between two users, and then retrieve a sequence containing such graph nodes from all users' graphs. The similar sequences represent a movement pattern that other researchers called T-patterns [49], based on which a similarity score is calculated for the pair of users. In the friend and place recommendation step, people can be ranked based on their similarity scores with respect to a given user. Then a set of people with relatively high scores can be recommended to the user as potential friends. A further analysis integrating other data such as points of interest (POIs) with the location histories of these friends can be used to recommend places to the user. The same approach was applied in [134] and [156]. Like Li et al. [81], Xiao et al. [134] focus on the first two processes of the proposed architecture but with an added consideration of the semantics of locations; that is, the location histories are represented as sequences of places (locations with semantic meaning such as restaurant, shopping mall, ...) instead of geographic coordinates. The work in [156] presents a complete recommender system built with this architecture.

Another instance of the trajectory-based recommendation problem is *trip recommendation*. This problem instance differs from the place recommendation instance in that trip recommendation produces not only a set of places to be visited but also the sequence in which they are to be visited. On the other hand, place recommendation provides one or more places but without proposing the order of visit. Trip recommendation differs from

the route recommendation in that trip recommendation does not specify the route to be followed from one place to the next in the sequence.

Similar to other trajectory-based recommendation problem instances, trip recommendation is addressed through frequent pattern mining methods with some customizations to handle user travel preferences in terms of stay at locations and transition between locations. For example, Brilhante et al. [16] model the mobility history of a user as a temporally ordered sequence of POIs he or she visited, each POI being annotated with the start and end time of visit. They then determine the travel preference of each user based on the functional categories of the POIs in his mobility history. Their system finds a personalized itinerary for a person by comparing his preference to the preferences of users taking into account the time budget the person has for the visit. Zheng and Xie [155] distinguished a generic trip recommendation from a personalized one and developed both. A generic trip recommendation proposes to an individual top interesting locations and travel sequences in a given region. On the other hand, a personalized trip recommendation provides an individual with locations and travel sequences matching his travel preferences.

5   Applications

The knowledge generated through trajectory data mining is useful in different application areas. The solution to one application problem can be used in different application fields. For instance, characterizing individuals in terms of their mobility profile can be used in urban planning for the management of recreational facilities [116], in business for planning business promotion activities, and in transportation to organize a car sharing system. Table 3 lists application examples, relating them to the application problems on which they are based, and giving corresponding references in the literature. This section discusses important requirements in respective application domains and how these requirements translate to application problems for trajectory data mining.

5.1   Transportation

In the transportation sector there is often a need to know the characteristics of regions or places (e.g., to identify major origins and destinations) and the characteristics of moving entities (e.g., when they move in a large number, and where they mainly stop). These two issues translate respectively to the application problems of characterizing regions (mainly hotspots discovery) and characterizing moving objects (mainly determining people's mobility profile). In some cases these problems extend to trajectory-based recommendation and trajectory-based prediction. In this category trajectory data mining aim, for example, at improving driving experience [142–144], the service of taxis and buses [46,80,83,113,158], and the planning and management of transportation systems [11,25,93,126].

5.2   Urban planning

Urban planners traditionally rely on labor-intensive surveys for addressing their concerns to know up-to-date characteristics of different urban regions and to determine the existence and characteristics of an interaction between regions. The first concern translates to region characterization, which is achieved by mining trajectories to identify regional functions or

Applications	Application problems[*]					
	A	**B**	**C**	**D**	**E**	**F**
Transportation						
Improving driving experience						[143, 144]
Improving the service of taxis and buses	[46, 80]		[83, 113]		[158]	
Improving the planning and management of transportation systems	[93, 126]		[11, 25]			
Urban planning						
Monitoring urban land use and urban boundary evolution			[92, 99]	[115]		
Detecting problems underlying the implementation of planning decisions				[24, 91, 154]		
Environment						
Controlling air pollution			[33, 153]			
Controlling noise pollution			[96, 114, 124]			
Energy						
Estimating energy consumption	[64]		[64, 121]			
Supporting site selection for charging infrastructure of electric vehicles	[103]		[103]			
Social applications						
Supporting the establishment of friendship and collaboration	[134, 135, 156]					[134, 135, 156]
Supporting on local navigation	[10, 138, 149, 157]					[10, 138, 149, 157]
Understanding individual and community lifestyle	[38, 145]	[145]				

*****A**: Characterization of moving objects, **B**: Discovery of social relationship, **C**: Characterization of places or regions, **D**: Discovery and characterization of the connectivity between places or regions, **E**: trajectory-based prediction, **F**: Trajectory-based recommendation

Table 3: Applications of trajectory data mining and corresponding application problems.

Applications	Application problems[*]					
	A	**B**	**C**	**D**	**E**	**F**
Business						
Improving advertisement strategies	[67]		[67]			[67]
Supporting on choosing a business placement	[66]		[66]			[66]
Improving shopping and selling services	[52,72]		[52,72]			
Public security and safety						
Detecting misbehavior of individuals	[29]					
Monitoring big events and crowd behavior	[128]					
Monitoring phenomena potential to cause natural disasters	[17]					
Ecology						
Supporting the management of free-ranging cattle	[8,59]					
Understanding of habitat use and supporting effective animal habitat management	[56]	[30,133]				
Sports analytics						
Game evaluation to support coaching activities	[55,60,123]					
* **A**: Characterization of moving objects, **B**: Discovery of social relationship, **C**: Characterization of places or regions, **D**: Discovery and characterization of the connectivity between places or regions, **E**: trajectory-based prediction, **F**: Trajectory-based recommendation						

Table 3 (continued): Applications of trajectory data mining and corresponding application problems

land use (e.g., [92, 99]). The second concern translates to discovering and characterizing the connectivity between regions. This has been addressed using trajectory data mining for exploring urban boundary evolution [115] and detecting problems underlying a transportation network [24, 91, 154].

5.3 Environment

Quantifying the pollution level at different locations and integrating pollution data with other data is important for pollution control. This has been translated to the problem of characterizing places and regions by mining trajectories which are associated with air pollution [33, 151] or noise pollution [90, 96, 114, 124] measurement values at different locations. In this case, trajectory data mining is a method for increasing the number of sampled locations and integrating and analyzing the data [153].

5.4 Energy

The energy sector is often concerned with determining the energy consumption patterns of specific regions and individuals. These concerns translate to region and moving object characterization, which are addressed by mining trajectories of vehicles [121]. The approach has been followed in implementing eco-driving feedback systems [64, 127], and in site selection for electric vehicle charging infrastructure [103].

5.5 Social applications

Trajectory data mining can support discovering the behavior of individuals and comparing them for different types of recommendations. The issues addressed translate to the application problems of determining individuals' mobility profiles, discovering and characterizing their social relationships, and eventually trajectory-based recommendation. For example, the similarity determined from individuals' mobility profiles [81, 157] enable a recommendation of friends [134, 135, 156], or location and path [10, 138, 149]. Likewise, the activities of individuals inferred from their trajectories enable understanding individual and community lifestyle [38, 145].

5.6 Business

The business sector often needs to determine the visit potential of specific places and the people's mobility pattern in these places. Trajectory data mining addresses this need by solving the application problems of characterizing places or regions, characterizing moving objects, and trajectory-based recommendation. For example, using trajectory data mining best business location [66, 125], best advertisement site [67], and best layout in a store [52, 72] have been recommended.

5.7 Public security and safety

An important need of applications in this domain is to identify and monitor places and moving objects that have a high potential to cause or be targeted by threats to public security and safety. Since situations affecting public security and safety are often associated

with a mass movement, trajectory data mining can contribute to timely detecting or even predicting these threats. The common application problems that need to be solved are characterizing places and moving objects, which can extend to trajectory-based prediction. The approach has been followed for detecting misbehavior [29], monitoring big events and crowd behavior [128], and exploring the movement of hurricanes [34], which can further enable predicting their future occurrence [17].

5.8   Ecology

People managing or doing research on animals are mainly concerned with determining the behavior of individual animals, their interactions, and the habitat use. This work is viewed as solving the application problems of characterizing individual moving objects and discovering and characterizing the interactions between them. For instance, trajectory mining has enabled automated detection of activities of animals [8, 56, 59], and their interaction [133], as well as understanding their habitat use [30].

5.9   Sports analytics

The dynamic nature of sports with a lot of movement and events causes a challenge for a timely evaluation. The major concern is to extract and analyze the mobility patterns of a player and his interactions with other players. This can be viewed as the application problems of characterizing moving objects and discovering and characterizing their interactions. By mining the trajectories of players, their movement [55, 60] and their interactions [123] have been evaluated.

6   Open issues for trajectory data mining

Trajectory data mining has attracted a high research interest and given its applications it is likely to continue attracting this interest. However, trajectory data mining is mainly progressing in the breadth where most of current work concentrates on developing variants of existing methods and applying the method variants to the same application problems. In order to take trajectory data mining to the next higher level there are important open issues that need to be addressed. This section discusses some of the issues.

6.1   Coping with massive data

Most of the data mining methods applied on trajectories were developed in the era of data scarcity and/or in a different domain. Since trajectory data is inherently voluminous and they are increasingly becoming available, future work on trajectory data mining should be scalable to handle massive data. This calls for among others more efficient indexing and aggregation methods that achieve a high data reduction while minimizing information loss.

6.2   Data integration and handling of heterogeneity

Most of the existing studies on trajectory data mining are based on one or one type of dataset (e.g., taxi GPS-based tracks). However, as acknowledged in some papers [92, 99]

such studies face the problem of missing data in some regions or during some periods, which could be resolved by integrating other types of data (e.g., GSM-based trajectories, geo-Social media, etc.). Therefore, future work should look into integrating different types of data by addressing challenges due to their heterogeneity such as reconciling different data formats and different sampling rates.

6.3   Cross-scale trajectory data mining

A trajectory data mining method is commonly developed and demonstrated on one dataset with one specific scale and granularity. However, as demonstrated in [74,75] the scale and granularity at which the method is applied may affect the result. Therefore, future work on trajectory data mining methods should take into account the scale and granularity to evaluate their impact on the method parameters.

6.4   Mining methods specialists and application domain specialists

While researchers in trajectory data mining methods have been mainly studying the geometric trajectory with no consideration of application context, recent collaboration efforts (e.g., the MOVE project [32]) have demonstrated that best results with real life relevance of trajectory data mining can be achieved by a collaboration between method specialists and application domain specialists. Such collaboration facilitates the access to relevant data for methods specialists. It can also support methods specialists in setting suitable values for parameters used in mining methods, a task which is usually left to the method user without guidance.

6.5   Privacy handling

While trajectories of animals are easily made openly available, human tracking is associated with serious privacy issues making high quality human trajectory data hardly accessible to mining method researchers. Despite some work done to address these issues (e.g., anonymization [1,51]) they prevail. Therefore, future work on trajectory data mining should develop mining frameworks that are more convincing on balancing the extraction of useful knowledge and preserving the privacy of tracked individuals.

7   Conclusion and future work

The increasing use of location-aware devices has led to an increasing availability of trajectory data representing the movement of moving objects. The potential of these data in solving important research problems has raised researchers' interest in analysis methods for them. In particular, inspired by the progress in mining classical large datasets, researchers have developed different data mining methods for trajectories. However, the provision of an integrated view of solved problems and methods in trajectory mining, as well as the applications based on solutions to these problems has not been explored.

   This paper reviewed the literature to identify the application problems that have been solved by mining trajectory data and the generic mining methods used. It further surveyed the applications that are based on solutions to these application problems. By surveying

the objective in using each method and how it works, the paper identified and discussed relationships that exist between different methods. The paper established a grouping of application problems and related the problems to the mining methods used to address them.

The major contribution of this paper is in providing an integrated view of trajectory mining application problems, the methods used to address them, and the applications that exploit solutions to these problems. The classification of application problems can help researchers in identifying new problems where trajectory mining could be profitably applied. The relationships between the methods and the connection between the application problems and methods can help researchers in identifying method gaps and inspire them to develop new methods. The work presented in this paper can also guide analysts in choosing a suitable mining method for a specific application problem.

The work presented in this paper can be extended by a comparative study of algorithms implementing a specific trajectory mining method on specific application problems with a possibility to provide recommendations on the choice between the algorithms. One of the current trends in movement data analysis is to relate the movement to its embedding context [73]. In view of this trend, our future work will be on customizing algorithms implementing some of the presented methods to perform a context-aware mining of trajectory data for selected application problems.

Acknowledgments

References

[1] ABUL, O., BONCHI, F., AND NANNI, M. Never walk alone: Uncertainty for anonymity in moving objects databases. In *Proc. 24th International Conference on Data Engineering* (2008), IEEE, pp. 376–385. doi:10.1109/icde.2008.4497446.

[2] ALON, J., SCLAROFF, S., KOLLIOS, G., AND PAVLOVIC, V. Discovering clusters in motion time series data. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2003), IEEE, pp. 375–381. doi:10.1109/cvpr.2003.1211378.

[3] ANAGNOSTOPOULOS, T., ANAGNOSTOPOULOS, C., AND HADJIEFTHYMIADES, S. An online adaptive model for location prediction. *Autonomic Computing and Communications Systems 23* (2010), 64–78. doi:10.1007/978-3-642-11482-3_5.

[4] ANDERSSON, M., GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. Reporting leaders and followers among trajectories of moving point objects. *Geoinformatica 12*, 4 (2008), 497–528. doi:10.1007/s10707-007-0037-9.

[5] ANDRIENKO, G., ANDRIENKO, N., BAK, P., KEIM, D., KISILEVICH, S., AND WROBEL, S. A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages and Computing 22*, 3 (2011), 213–232. doi:10.1016/j.jvlc.2011.02.003.

[6] ANKERST, M., BREUNIG, M. M., KRIEGEL, H.-P., AND SANDER, J. OPTICS: Ordering points to identify the clustering structure. In *Proc. ACM SIGMOD International Conference on Management of Data* (1999), ACM Press, pp. 49–60. doi:10.1145/304182.304187.

[7] ASAHARA, A., MARUYAMA, K., SATO, A., AND SETO, K. Pedestrian-movement prediction based on mixed markov-chain model. In *Proc. 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2011), ACM Press, pp. 25–33. doi:10.1145/2093973.2093979.

[8] AUGUSTINE, D. J., AND DERNER, J. D. Assessing herbivore foraging behavior with GPS collars in a semiarid grassland. *Sensors 13*, 3 (2013), 3711–3723. doi:10.3390/s130303711.

[9] BACKSTROM, L., SUN, E., AND MARLOW, C. Find me if you can: Improving geographical prediction with social and spatial proximity. In *Proc. 19th International Conference on World Wide Web* (2010), ACM Press, pp. 61–70. doi:10.1145/1772690.1772698.

[10] BAO, J., ZHENG, Y., AND MOKBEL, M. F. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proc. 20th ACM SIGSPATIAL Conference on Advances in Geographic Information Systems* (2012), ACM press, pp. 199–208. doi:10.1145/2424321.2424348.

[11] BASTANI, F., HUANG, Y., XIE, X., AND POWELL, J. W. A greener transportation mode: Flexible routes discovery from GPS trajectory data. In *Proc. 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2011), ACM Press, pp. 405–408. doi:10.1145/2093973.2094034.

[12] BENKERT, M., GUDMUNDSSON, J., HÜBNER, F., AND WOLLE, T. Reporting flock patterns. *Computational Geometry 41*, 3 (2008), 111–125. doi:10.1016/j.comgeo.2007.10.003.

[13] BILJECKI, F., LEDOUX, H., AND OOSTEROM, P. V. Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science 27*, 2 (2013), 385–407. doi:10.1080/13658816.2012.692791.

[14] BIRANT, D., AND KUT, A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data and Knowledge Engineering 60*, 1 (2007), 208–221. doi:10.1016/j.datak.2006.01.013.

[15] BOLBOL, A., CHENG, T., TSAPAKIS, I., AND HAWORTH, J. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems 31*, 6 (2012), 526–537. doi:10.1016/j.compenvurbsys.2012.06.001.

[16] BRILHANTE, I., MACEDO, J. A., NARDINI, F. M., PEREGO, R., AND RENSO, C. Where shall we go today? planning touristic tours with TripBuilder. In *Proc. 22nd ACM International Conference on Information & Knowledge Management* (2013), ACM Press, pp. 757–762. doi:10.1145/2505515.2505643.

[17] Buchin, M., Dodge, S., and Speckmann, B. Similarity of trajectories taking into account geographic context. *Journal of Spatial Information Science 9* (2014), 101–124. doi:10.5311/josis.2014.9.179.

[18] Calabrese, F., Pereira, F. C., Lorenzo, G. D., Liu, L., and Ratti, C. The geography of taste: Analyzing cell-phone mobility and social events. In *Pervasive: Proc. 8th International Conference on Pervasive Computing* (2010), Springer-Verlag Berlin, pp. 22–37. doi:10.1007/978-3-642-12654-3_2.

[19] Cao, H., Mamoulis, N., and Cheung, D. W. Mining frequent spatio-temporal sequential patterns. In *ICDM: Proc. Fifth IEEE International Conference on Data Mining* (2005), IEEE Computer Society Press, pp. 82–89. doi:10.1109/icdm.2005.95.

[20] Cao, H., Mamoulis, N., and Cheung, D. W. Discovery of periodic patterns in spatiotemporal sequences. *IEEE Transactions on Knowledge and Data Engineering 19*, 4 (2007), 453–467. doi:10.1109/tkde.2007.1002.

[21] Carneiro, C., Alp, A., Macedo, J., and Spaccapietra, S. Advanced data mining method for discovering regions and trajectories of moving objects: "ciconia ciconia" scenario. In *Proc. 11th AGILE Conference* (2008), L. Bernard, A. Friis-Christensen, and H. Pundt, Eds., Springer Berlin Heidelberg, pp. 201–224. doi:10.1007/978-3-540-78946-8_11.

[22] Castro, P. S., Zhang, D., Chen, C., Li, S., and Pan, G. From taxi GPS traces to social and community dynamics: A survey. *ACM Computing Surveys 46*, 2 (2013), 1–34. doi:10.1145/2543581.2543584.

[23] Castro, P. S., Zhang, D., and Li, S. Urban traffic modelling and prediction using large scale taxi GPS traces. In *Pervasive Computing, LNCS 7319*, J. Kay, P. Lukowicz, H. Tokuda, P. Olivier, and A. Krüger, Eds. Springer, 2012, pp. 57–72. doi:10.1007/978-3-642-31205-2_4.

[24] Chawla, S., Zheng, Y., and Hu, J. Inferring the root cause in road traffic anomalies. In *ICDM: Proc. IEEE 12th International Conference on Data Mining* (2012), IEEE, pp. 141–150. doi:10.1109/icdm.2012.104.

[25] Chen, C., Zhang, D., Zhou, Z.-H., Li, N., Atmaca, T., and Li, S. B-planner: Night bus route planning using large-scale taxi GPS traces. In *PerCom: Proc. IEEE International Conference on Pervasive Computing and Communications* (2013), IEEE, pp. 225–233. doi:10.1109/percom.2013.6526736.

[26] Chen, X., Pang, J., and Xue, R. Constructing and comparing user mobility profiles for location-based services. In *SAC: Proc. 28th Annual ACM Symposium on Applied Computing* (2013), ACM Press, pp. 261–266. doi:10.1145/2480362.2480418.

[27] Chen, X., Pang, J., and Xue, R. Constructing and comparing user mobility profiles. *ACM Transactions on the Web (TWEB) 8*, 4 (2014), 1–25. doi:10.1145/2637483.

[28] Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., and Kleinberg, J. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences 107*, 52 (2010), 22436–22441. doi:10.1073/pnas.1006155107.

[29] DA SILVA, F. P., AND FILETO, R. A method to detect and classify inconsistencies of moving objects' stops with requested and reported tasks. *Journal of Information and Data Management 6*, 1 (2015), 71–80.

[30] DE GROEVE, J., DE WEGHE, N. V., RANC, N., NEUTENS, T., OMETTO, L., ROTA-STABELLI, O., AND CAGNACCI, F. Extracting spatio-temporal patterns in animal trajectories: an ecological application of sequence analysis methods. *Methods in Ecology and Evolution 7*, 3 (2015), 369–379. doi:10.1111/2041-210x.12453.

[31] DE LUCCA SIQUEIRA, F., AND BOGORNY, V. Discovering chasing behavior in moving object trajectories. *Transactions in GIS 15*, 5 (2011), 667–688. doi:10.1111/j.1467-9671.2011.01285.x.

[32] DEMŠAR, U., BUCHIN, K., CAGNACCI, F., SAFI, K., SPECKMANN, B., DE WEGHE, N. V., WEISKOPF, D., AND WEIBEL, R. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology 3*, 1 (2015), 1–24. doi:10.1186/s40462-015-0032-y.

[33] DEVARAKONDA, S., SEVUSU, P., LIU, H., LIU, R., IFTODE, L., AND NATH, B. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *UrbComp'13: Proc. 2nd ACM SIGKDD International Workshop on Urban Computing* (2013), ACM Press, pp. 15:1–15:8. doi:10.1145/2505821.2505834.

[34] DODGE, S., LAUBE, P., AND WEIBEL, R. Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science 26*, 9 (2012), 1563–1588. doi:10.1080/13658816.2011.630003.

[35] DODGE, S., WEIBEL, R., AND LAUTENSCHÜTZ, A.-K. Towards a taxonomy of movement patterns. *Journal of Information Visualization 7* (2008), 240–252. doi:10.1057/palgrave.ivs.9500182.

[36] ESTER, M., KRIEGEL, H.-P., SANDER, J., AND XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD'96: Proc. Second International Conference on Knowledge Discovery and Data Mining* (1996), pp. 226–231.

[37] FAYYAD, U., PIATETSKY-SHAPIRO, G., AND SMYTH, P. From data mining to knowledge discovery: an overview. In *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, Amith, P. Smyth, and R. Uthurusamy, Eds. AAAAI Press, 1996, pp. 1–34.

[38] FILHO, R. M., BORGES, G. R., ALMEIDA, J. M., AND PAPPA, G. L. Inferring user social class in online social networks. In *SNAKDD'14: Proc. 8th Workshop on Social Network Mining and Analysis* (2014), ACM Press. doi:10.1145/2659480.2659502.

[39] FONTES, V. C., DE ALENCAR, L. A., RENSO, C., AND BOGORNY, V. Discovering trajectory outliers between regions of interest. In *GeoInfo 2013: Proc. XIV Brazilian Symposium on Geoinformastics* (2013), pp. 49–60.

[40] FRALEY, C., AND RAFTERY, A. E. How many clusters? which clustering method? answers via model-based cluster analysis. *The Computer Journal 41*, 8 (1998), 578–588. doi:10.1093/comjnl/41.8.578.

[41] FROEHLICH, J., AND KRUMM, J. Route prediction from trip observations. In *Society of Automotive Engineers (SAE) World Congress* (2008), SAE International. doi:10.4271/2008-01-0201.

[42] GAFFNEY, S., AND SMYTH, P. Trajectory clustering with mixture of regression models. In *KDD' 99: Proc. Fifth International Conference on Knowledge Discovery and Data Mining* (1999), ACM Press, pp. 63–72. doi:10.1145/312129.312198.

[43] GAFFNEY, S. J., ROBERTSON, A. W., SMYTH, P., CAMARGO, S. J., AND GHIL, M. Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics 29*, 4 (2007), 423–440. doi:10.1007/s00382-007-0235-z.

[44] GAITO, S., ROSSI, G. P., AND ZIGNANI, M. From mobility data to social attitudes: A complex network approach. In *NEMO'11: Proc. International Workshop on Finding Patterns of Human Behaviors in Networks and Mobility Data* (2011), pp. 52–65.

[45] GAO, P., KUPFER, J. A., ZHU, X., AND GUO, D. Quantifying animal trajectories using spatial aggregation and sequence analysis – a case study of differentiating trajectories of multiple species. *Geographical Analysis 48*, 3 (2016), 275–291. doi:10.1111/gean.12098.

[46] GE, Y., XIONG, H., TUZHILIN, A., XIAO, K., GRUTESER, M., AND PAZZANI, M. An energy-efficient mobile recommender system. In *KDD'10: Proc. 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), ACM Press, pp. 899–908. doi:10.1145/1835804.1835918.

[47] GIANNOTTI, F., NANNI, M., AND PEDRESCHI, D. Efficient mining of temporally annotated sequences. In *Proc. 2006 SIAM International Conference on Data Mining* (2006), Society for Industrial & Applied Mathematics (SIAM), pp. 348–359. doi:10.1137/1.9781611972764.31.

[48] GIANNOTTI, F., NANNI, M., PEDRESCHI, D., PINELLI, F., RENSO, C., RINZIVILLO, S., AND TRASARTI, R. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB Journal 20*, 5 (2011), 695–719. doi:10.1007/s00778-011-0244-8.

[49] GIANNOTTI, F., NANNI, M., PINELLI, F., AND PEDRESCHI, D. Trajectory pattern mining. In *KDD'07: Proc. 13th International Conference on Knowledge Discovery and Data Mining* (2007), ACM Press, pp. 330–339. doi:10.1145/1281192.1281230.

[50] GIDÓFALVI, G., AND DONG, F. When and where next: Individual mobility prediction. In *MobiGIS'12: Proc. First ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems* (2012), ACM Press, pp. 57–64. doi:10.1145/2442810.2442821.

[51] GIDOFALVI, G., HUANG, X., AND PEDERSEN, T. B. Privacy-preserving data mining on moving object trajectories. In *2007 International Conference on Mobile Data Management* (2007), IEEE, pp. 60–68. doi:10.1109/mdm.2007.18.

[52] GIL, J., TOBARI, E., LEMLIJ, M., ROSE, A., AND PENN, A. R. The differentiating behaviour of shoppers: clustering of individual movement traces in a supermarket. In *Proc. 7th International Space Syntax Symposium* (2009), no. 036.

[53] GUDMUNDSSON, J., LAUBE, P., AND WOLLE, T. Computational movement analysis. In *Springer Handbook of Geographic Information*, W. Kresse and D. Danko, Eds. Springer, 2012, pp. 725–741. doi:10.1007/978-3-540-72680-7_22.

[54] GUDMUNDSSON, J., VAN KREVELD, M., AND SPECKMANN, B. Efficient detection of motion patterns in spatiotemporal data sets. In *GIS'04: Proc. 12th annual ACM international workshop on Geographic information systems* (2004), ACM, pp. 250–257. doi:10.1145/1032222.1032259.

[55] GUDMUNDSSON, J., AND WOLLE, T. Towards automated football analysis: Algorithms and data structures. In *Proc. 10th Australasian Conference on Mathematics and Computers in Sport* (2010), pp. 145–152.

[56] GUILFORD, T., MEADE, J., WILLIS, J., PHILLIPS, R. A., BOYLE, D., ROBERTS, S., COLLETT, M., FREEMAN, R., AND PERRINS, C. M. Migration and stopover in a small pelagic seabird, the manx shearwater puffinus puffinus: insights from machine learning. *Proceedings of the Royal Society B: Biological Sciences 276*, 1660 (2009), 1215–1223. doi:10.1098/rspb.2008.1577.

[57] HAN, J., KAMBER, M., AND PEI, J. *Data mining: concepts and techniques*, 3rd ed. Morgan Kaufmann, 2012. doi:10.5860/choice.49-3305.

[58] HAN, J., LEE, J.-G., AND KAMBER, M. An overview of clustering methods in geographic data analysis. In *Geographic Data Mining and Knowledge Discovery*, H. Miller and J. Han, Eds., 2nd ed. Taylor and Francis Group, 2009, pp. 149–187. doi:10.1201/9781420073980.ch7.

[59] HOMBURGER, H., SCHNEIDER, M. K., HILFIKER, S., AND LÜSCHER, A. Inferring behavioral states of grazing livestock from high-frequency position data alone. *PLoS ONE 9*, 12 (2014), e114522. doi:10.1371/journal.pone.0114522.

[60] HORTON, M., GUDMUNDSSON, J., CHAWLA, S., AND ESTEPHAN, J. Automated classification of passing in football. In *PAKDD 2015: Proc. 19th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining* (2015), Springer International Publishing, pp. 319–330. doi:10.1007/978-3-319-18032-8_25.

[61] JEUNG, H., LIU, Q., SHEN, H. T., AND ZHOU, X. A hybrid prediction model for moving objects. In *Proc. 24th IEEE International Conference on Data Engineering* (2008), IEEE, pp. 70–79. doi:10.1109/icde.2008.4497415.

[62] JEUNG, H., SHEN, H. T., AND ZHOU, X. Convoy queries in spatio-temporal databases. In *ICDE '08: Proc. 2008 IEEE 24th International Conference on Data Engineering* (2008), IEEE Computer Society, pp. 1457–1459. doi:10.1109/ICDE.2008.4497588.

[63] JEUNG, H., YIU, M. L., ZHOU, X., JENSEN, C. S., AND SHEN, H. T. Discovery of convoys in trajectory databases. *Proceedings of the VLDB Endowment 1*, 1 (2008), 1068–1080. doi:10.14778/1453856.1453971.

[64] JIRKA, S., REMKE, A., AND BRÖRING, A. envirocar – crowd sourced traffic and environment data for sustainable mobility. In *ENVIP'2013: Proc. Environmental Information Infrastructures and Platforms workshop* (2013).

[65] Kang, J.-Y., and Yong, H.-S.   Mining spatio-temporal patterns in trajectory data.   *Journal of Information Processing Systems 6*, 4 (2010), 521–536. doi:10.3745/JIPS.2010.6.4.521.

[66] Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., and Mascolo, C. Geo-spotting: Mining online location-based services for optimal retail store placement. In *KDD '13: Proc. 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (2013), ACM Press, pp. 793–801. doi:10.1145/2487575.2487616.

[67] Körner, C., Hecker, D., May, M., and Wrobel, S. Visit potential: A common vocabulary for the analysis of entity-location interactions in mobility applications. In *Geospatial Thinking, Lecture Notes in Geoinformation and Cartography*, M. Painho, M. Y. Santos, and H. Pundt, Eds. Springer Berlin Heidelberg, 2010, pp. 79–95. doi:10.1007/978-3-642-12326-9_5.

[68] Körner, C., May, M., and Wrobel, S. Spatiotemporal modeling and analysis - introduction and overview. *KI - Künstliche Intelligenz 26*, 3 (2012), 215–221. doi:10.1007/s13218-012-0215-2.

[69] Krumm, J. A markov model for driver turn prediction. In *Society of Automotive Engineers (SAE) 2008 World Congress* (2008), SAE International. doi:10.4271/2008-01-0195.

[70] Krumm, J. Where will they turn: predicting turn proportions at intersections. *Personal and Ubiquitous Computing 14*, 7 (2009), 591–599. doi:10.1007/s00779-009-0248-1.

[71] Krumm, J., and Horvitz, E. Predestination: Inferring destinations from partial trajectories. In *UbiComp 2006: Ubiquitous Computing, LNCS 4206*, P. Dourish and A. Friday, Eds. Springer, 2006, pp. 243–260. doi:10.1007/11853565_15.

[72] Larson, J. S., Bradlow, E. T., and Fader, P. S. An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing 22*, 4 (2005), 395–414. doi:10.1016/j.ijresmar.2005.09.005.

[73] Laube, P. *Computational Movement Analysis*. Springer, Berlin-Heidelberg, 2014.

[74] Laube, P., Dennis, T., Forer, P., and Walker, M. Movement beyond the snapshot: Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems 31*, 5 (2007), 481–501. doi:10.1016/j.compenvurbsys.2007.08.002.

[75] Laube, P., and Purves, R. S. How fast is a cow? cross-scale analysis of movement data. *Transactions in GIS 15*, 3 (2011), 401–418. doi:10.1111/j.1467-9671.2011.01256.x.

[76] Laube, P., van Kreveld, M., and Imfeld, S. Finding remo – detecting relative motion patterns in geospatial lifelines. In *Developments in Spatial Data Handling*, P. F. Fisher, Ed. Springer, 2005, pp. 201–215. doi:10.1007/3-540-26772-7_16.

[77] Lee, J.-G., Han, J., and Li, X. Trajectory outlier detection: A partition-and-detect framework. In *ICDE 2008: Proc. International Conference on Data Mining* (2008), IEEE, pp. 140–149. doi:10.1109/icde.2008.4497422.

[78] Lee, J.-G., Han, J., Li, X., and Gonzalez, H. Traclass: Trajectory classification using hierarchical region based and trajectory-based clustering. *Proceedings of the VLDB Endowment 1*, 1 (2008), 1081–1094. doi:10.14778/1453856.1453972.

[79] LEE, J.-G., HAN, J., AND WHANG, K.-Y. Trajectory clustering: A partition-and-group framework. In *Proc. 2007 ACM SIGMOD International Conference on Management of Data* (2007), ACM Press, pp. 593–604. doi:10.1145/1247480.1247546.

[80] LI, B., ZHANG, D., SUN, L., CHEN, C., LI, S., QI, G., AND YANG, Q. Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)* (2011), IEEE, pp. 63–68. doi:10.1109/percomw.2011.5766967.

[81] LI, Q., ZHENG, Y., XIE, X., CHEN, Y., LIU, W., AND MA, W.-Y. Mining user similarity based on location history. In *GIS '08: Proc. 16th ACM SIGSPATIAL international conference on Advances in geographic information systems* (2008), ACM Press, pp. 1–10. doi:10.1145/1463434.1463477.

[82] LI, X., HAN, J., KIM, S., AND GONZALEZ, H. ROAM: Rule- and motif-based anomaly detection in massive moving object data sets. In *Proc. 2007 SIAM International Conference on Data Mining* (2007), Society for Industrial and Applied Mathematics (SIAM), pp. 273–284. doi:10.1137/1.9781611972771.25.

[83] LI, X., PAN, G., WU, Z., QI, G., LI, S., ZHANG, D., ZHANG, W., AND WANG, Z. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science 6*, 1 (2012), 111–121. doi:10.1007/s11704-011-1192-6.

[84] LI, Z., DING, B., HAN, J., AND KAYS, R. Swarm: mining relaxed temporal moving object clusters. *Proceedings of the VLDB Endowment 3*, 1-2 (2010), 723–734. doi:10.14778/1920841.1920934.

[85] LI, Z., DING, B., HAN, J., KAYS, R., AND NYE, P. Mining periodic behaviors for moving objects. In *KDD '10: Proc. 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), ACM Press, pp. 1099–1108. doi:10.1145/1835804.1835942.

[86] LI, Z., HAN, J., JI, M., TANG, L.-A., YU, Y., DING, B., LEE, J.-G., AND KAYS, R. Movemine: Mining moving object data for discovery of animal movement patterns. *ACM Transactions on Intelligent Systems and Technology 2*, 4 (2011), 1–32. doi:10.1145/1989734.1989741.

[87] LIU, L., BIDERMAN, A., AND RATTI, C. Urban mobility landscape: Real time monitoring of urban mobility patterns. In *CUPUM2009: Proc. 11th International Conference on Computers in Urban Planning and Urban Management* (2009), pp. 1–16.

[88] LIU, L., QIAO, S., ZHANG, Y., AND HU, J. An efficient outlying trajectories mining approach based on relative distance. *International Journal of Geographical Information Science 26*, 10 (2012), 1789–1810. doi:10.1080/13658816.2012.654792.

[89] LIU, S., LIU, Y., NI, L. M., FAN, J., AND LI, M. Towards mobility-based clustering. In *KDD '10: Proc. 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), ACM, pp. 919–928. doi:10.1145/1835804.1835920.

[90] LIU, T., ZHENG, Y., LIU, L., LIU, Y., AND ZHU, Y. Methods for sensing urban noises. Tech. Rep. MSR-TR-2014-66, Microsoft Research Asia, 2014.

[91] LIU, W., ZHENG, Y., CHAWLA, S., YUAN, J., AND XING, X. Discovering spatio-temporal causal interactions in traffic data streams. In *KDD '11: Proc. 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011), ACM, pp. 1010–1018. doi:10.1145/2020408.2020571.

[92] LIU, Y., WANG, F., XIAO, Y., AND GAO, S. Urban land uses and traffic 'source-sink areas': Evidence from gps-enabled taxi data in shanghai. *Landscape and Urban Planning 106*, 1 (2012), 73–87. doi:10.1016/j.landurbplan.2012.02.012.

[93] MA, S., ZHENG, Y., AND WOLFSON, O. In *ICDE 2013: Proc. 29th IEEE International Conference on Data Engineering* (2013). doi:10.1109/icde.2013.6544843, Pages = 410–421, Publisher = IEEE, Title = T-share: A large-scale dynamic taxi ridesharing service, Url = http://dx.doi.org/10.1109/icde.2013.6544843.

[94] MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In *Proc. 5th Berkley Symposium on Mathematical Statistics and Probability*, L. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.

[95] MAIMON, O., AND ROKACH, L. Introduction to knowledge discovery and data mining. In *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., 2nd ed. Springer, 2010, pp. 1–15. doi:10.1007/978-0-387-09823-4_1.

[96] MAISONNEUVE, N., STEVENS, M., NIESSEN, M. E., AND STEELS, L. Noisetube: Measuring and mapping noise pollution with mobile phones. In *Proc. 4th International ICSC Symposium on Information Technologies in Environmental Engineering* (2009), Springer Berlin Heidelberg, pp. 215–228. doi:10.1007/978-3-540-88351-7_16.

[97] MAMOULIS, N., CAO, H., KOLLIOS, G., HADJIELEFTHERIOU, M., TAO, Y., AND CHEUNG, D. W. Mining, indexing, and querying historical spatiotemporal data. In *KDD '04: Proc. 2004 ACM SIGKDD international conference on Knowledge discovery and data mining* (2004), ACM, pp. 236–245. doi:10.1145/1014052.1014080.

[98] MATHEW, W., RAPOSO, R., AND MARTINS, B. Predicting future locations with hidden markov models. In *UbiComp '12: Proc. 2012 ACM Conference on Ubiquitous Computing* (2012), ACM, pp. 911–918. doi:10.1145/2370216.2370421.

[99] MAZIMPAKA, J. D., AND TIMPF, S. Exploring the potential of combining taxi GPS and flickr data for discovering functional regions. In *AGILE 2015: Geographic Information Science as an Enabler of Smarter Cities and Communities*, F. Bacao, M. Y. Santos, and M. Painho, Eds. Springer International Publishing, 2015, pp. 3–18. doi:10.1007/978-3-319-16787-9_1.

[100] MENNIS, J., AND GUO, D. Spatial data mining and geographic knowledge discovery-an introduction. *Computers, Environment and Urban Systems 33*, 6 (2009), 403–408. doi:10.1016/j.compenvurbsys.2009.11.001.

[101] MILLER, H. J. Geographic data mining and knowledge discovery. In *Handbook of Geographic Information Science*, J. P. Wilson and A. S. Fotheringham, Eds. Blackwell Publishing, 2008, pp. 352–366. doi:10.1002/9780470690819.ch19.

[102] MILLER, H. J., AND HAN, J. Geographic data mining and knowledge discovery: An overview. In *Geographic Data Mining and Knowledge Discovery*, H. J. Miller and J. Han, Eds., 2nd ed. Taylor and Francis Group, FL, USA, 2009, pp. 1–26. doi:10.4324/9780203468029_chapter_1.

[103] MOMTAZPOUR, M., BUTLER, P., HOSSAIN, M. S., BOZCHALUI, M. C., RAMAKR-ISHNAN, N., AND SHARMA, R. Coordinated clustering algorithms to support charging infrastructure design for electric vehicles. In *UrbComp '12: Proc. 2nd ACM SIGKDD International Workshop on Urban Computing* (2012), ACM, pp. 126–133. doi:10.1145/2346496.2346517.

[104] MONREALE, A., PINELLI, F., TRASARTI, R., AND GIANNOTTI, F. WhereNext: A location predictor on trajectory pattern mining. In *KDD '09: Proc. 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), ACM, pp. 637–646. doi:10.1145/1557019.1557091.

[105] MORZY, M. Prediction of moving object location based on frequent trajectories. In *Computer and Information Sciences - ISCIS 2006, LNCS 4263*, A. Levi, E. Savas, H. Yenigün, S. Balcosoy, and Y. Saygin, Eds. Springer Berlin Heidelberg, 2006, pp. 583–592. doi:10.1007/11902140_62.

[106] NANNI, M., AND PEDRESCHI, D. Time-focused clustering of trajectories of moving objects. *Journal of Intelligent Information Systems 27*, 3 (2006), 267–289. doi:10.1007/s10844-006-9953-7.

[107] ONG, R., PINELLI, F., TRASARTI, R., NANNI, M., RENSO, C., RINZIVILLO, S., AND GIANNOTTI, F. Traffic jams detection using flock mining. In *Machine Learning and Knowledge Discovery in Databases, LNCS 6913*, D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, Eds. Springer Berlin Heidelberg, 2011, pp. 650–653. doi:10.1007/978-3-642-23808-6_49.

[108] ORELLANA, D., BREGT, A. K., LIGTENBERG, A., AND WACHOWICZ, M. Exploring visitor movement patterns in natural recreational areas. *Tourism Management 33*, 3 (2012), 672–682. doi:10.1016/j.tourman.2011.07.010.

[109] PALMA, A. T., BOGORNY, V., KUIJPERS, B., AND ALVARES, L. O. A clustering-based approach for discovering interesting places in trajectories. In *SAC '08: Proc. 2008 ACM symposium on Applied computing* (2008), ACM, pp. 863–868. doi:10.1145/1363686.1363886.

[110] PAN, B., ZHENG, Y., WILKIE, D., AND SHAHABI, C. Crowd sensing of traffic anomalies based on human mobility and social media. In *SIGSPATIAL '13: Proc. 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2013), ACM, pp. 344–353. doi:10.1145/2525314.2525343.

[111] PAN, G., QI, G., WU, Z., ZHANG, D., AND LI, S. Land-use classification using taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems 14*, 1 (2013), 113–123. doi:10.1109/tits.2012.2209201.

[112] PELEKIS, N., AND THEODORIDIS, Y. *Mobility Data Management and Exploration*. Springer-Verlag New York, 2014. doi:10.1007/978-1-4939-0392-4.

[113] PHITHAKKITNUKOON, S., VELOSO, M., BENTO, C., BIDERMAN, A., AND RATTI, C. Taxi-aware map: Identifying and predicting vacant taxis in the city. In *Ambient Intelligence, LNCS 6439*, B. de Ruyter, R. Wichert, D. V. Keyson, P. Markopoulos, N. Streitz, M. Divitini, N. Georgantas, and A. M. Gomez, Eds. Springer, 2010, pp. 86–95. doi:10.1007/978-3-642-16917-5_9.

[114] RANA, R., CHOU, C. T., BULUSU, N., KANHERE, S., AND HU, W. Ear-phone: A context-aware noise mapping using smart phones. *Pervasive and Mobile Computing 17* (2015), 1–22. doi:10.1016/j.pmcj.2014.02.001.

[115] RATTI, C., SOBOLEVSKY, S., CALABRESE, F., ANDRIS, C., READES, J., MARTINO, M., CLAXTON, R., AND STROGATZ, S. H. Redrawing the map of great britain from a network of human interactions. *PLoS ONE 5*, 12 (2010), e14248. doi:10.1371/journal.pone.0014248.

[116] RENSO, C., BAGLIONI, M., DE MACEDO, J. A. F., TRASARTI, R., AND WACHOWICZ, M. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and Information Systems 37*, 2 (2012), 331–362. doi:10.1007/s10115-012-0511-z.

[117] REUMERS, S., LIU, F., JANSSENS, D., COOLS, M., AND WETS, G. Semantic annotation of global positioning system traces: Activity type inference. *Transportation Research Record: Journal of the Transportation Research Board 2383* (2013), 35–43. doi:10.3141/2383-05.

[118] RINZIVILLO, S., MAINARDI, S., PEZZONI, F., COSCIA, M., PEDRESCHI, D., AND GIANNOTTI, F. Discovering the geographical borders of human mobility. *KI - Künstliche Intelligenz 26*, 3 (2012), 253–260. doi:10.1007/s13218-012-0181-8.

[119] ROKACH, L. A survey of clustering algorithms. In *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., 2nd ed. Springer, 2010, pp. 269–298. doi:10.1007/978-0-387-09823-4_14.

[120] SHAMOUN-BARANES, J., BOM, R., VAN LOON, E. E., ENS, B. J., OOSTERBEEK, K., AND BOUTEN, W. From sensor data to animal behaviour: An oystercatcher example. *PLoS ONE 7*, 5 (2012), e37997. doi:10.1371/journal.pone.0037997.

[121] SHANG, J., ZHENG, Y., TONG, W., CHANG, E., AND YU, Y. Inferring gas consumption and pollution emission of vehicles throughout a city. In *KDD '14: Proc. 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), ACM, pp. 1027–1036. doi:10.1145/2623330.2623653.

[122] SPINSANTI, L., BERLINGERIO, M., AND PAPPALARDO, L. Mobility and geo-social networks. In *Mobility Data*, C. Renso, S. Spaccapietra, and E. Zimanyi, Eds. Cambridge University Press, 2013, pp. 315–333. doi:10.1017/cbo9781139128926.017.

[123] STEIN, M., HÄUSSLER, J., JÄCKLE, D., JANETZKO, H., SCHRECK, T., AND KEIM, D. Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information 4*, 4 (2015), 2159–2184. doi:10.3390/ijgi4042159.

[124] STEVENS, M., AND D'HONDT, E. Crowdsourcing of pollution data using smartphones. In *Proc. Workshop on Ubiquitous Crowdsourcing at UbiComp '10* (2010), ACM, pp. 1–3.

[125] TEKNOMO, K., AND GERILLA, P. G. Pedestrian static trajectory analysis of a hypermarket. In *Proc. 8th International Conference of Eastern Asia Society for Transportation Studies* (2009), pp. 220–220.

[126] TRASARTI, R., PINELLI, F., NANNI, M., AND GIANNOTTI, F. Individual mobility profiles: methods and application on vehicle sharing. In *Proc. Twentieth Italian Symposium on Advanced Database Systems* (2012), pp. 35–42.

[127] TULUSAN, J., STAAKE, T., AND FLEISCH, E. Providing eco-driving feedback to corporate car drivers: what impact does a smartphone application have on their fuel efficiency? In *UbiComp '12: Proc. 2012 ACM Conference on Ubiquitous Computing* (2012), ACM, pp. 212–215. doi:10.1145/2370216.2370250.

[128] VERSICHELE, M., NEUTENS, T., DELAFONTAINE, M., AND DE WEGHE, N. V. The use of bluetooth for analyzing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. *Applied Geography 32*, 2 (2012), 208–220. doi:10.1016/j.apgeog.2011.05.011.

[129] VIEIRA, M. R., BAKALOV, P., AND TSOTRAS, V. J. On-line discovery of flock patterns in spatio-temporal data. In *GIS '09: Proc. 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2009), ACM, pp. 286–295. doi:10.1145/1653771.1653812.

[130] WACHOWICZ, M., ONG, R., RENSO, C., AND NANNI, M. Finding moving flock patterns among pedestrians through collective coherence. *International Journal of Geographical Information Science 25*, 11 (2011), 1849–1864. doi:10.1080/13658816.2011.561209.

[131] WANG, D., PEDRESCHI, D., SONG, C., GIANNOTTI, F., AND BARABASI, A.-L. Human mobility, social ties, and link prediction. In *KDD '11: Proc. 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011), ACM, pp. 1100–1108. doi:10.1145/2020408.2020581.

[132] WEN CHANG, H., CHIN TAI, Y., AND JEN HSU, J. Y. Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining 5*, 1 (2010), 3–18. doi:10.1504/ijbidm.2010.030296.

[133] WU, F., LEI, T. K. H., LI, Z., AND HAN, J. Movemine 2.0: Mining object relationships from movement data. *Proceedings of the VLDB Endowment 7*, 13 (2014), 1613–1616. doi:10.14778/2733004.2733043.

[134] XIAO, X., ZHENG, Y., LUO, Q., AND XIE, X. Finding similar users using category-based location history. In *GIS '10: Proc. 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2010), ACM, pp. 442–445. doi:10.1145/1869790.1869857.

[135] XIAO, X., ZHENG, Y., LUO, Q., AND XIE, X. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing 5*, 1 (2012), 3–19. doi:10.1007/s12652-012-0117-z.

[136] YING, J. J.-C., LEE, W.-C., AND TSENG, V. S. Mining geographic-temporal-semantic patterns in trajectories for location prediction. *ACM Transactions on Intelligent Systems and Technology 5*, 1 (2013), 1–33. doi:10.1145/2542182.2542184.

[137] YING, J. J.-C., LEE, W.-C., WENG, T.-C., AND TSENG, V. S. Semantic trajectory mining for location prediction. In *GIS '11: Proc. 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2011), ACM, pp. 34–43. doi:10.1145/2093973.2093980.

[138] YOON, H., ZHENG, Y., XIE, X., AND WOO, W. Social itinerary recommendation from user-generated digital trails. *Personal and Ubiquitous Computing 16*, 5 (2011), 469–484. doi:10.1007/s00779-011-0419-8.

[139] YU, Y., WANG, Q., KUANG, J., AND HE, J. TGCR: An efficient algorithm for mining swarm in trajectory databases. In *Proc. 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services* (2011), IEEE, pp. 90–95. doi:10.1109/icsdm.2011.5969011.

[140] YUAN, G., XIA, S., ZHANG, L., AND JI, C. Trajectory outlier detection algorithm based on structural features. *Journal of Computational Information Systems 7*, 11 (2011), 4137–4144.

[141] YUAN, J., ZHENG, Y., AND XIE, X. Discovering regions of different functions in a city using human mobility and POIs. In *KDD '12: Proc. 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), ACM, pp. 186–194. doi:10.1145/2339530.2339561.

[142] YUAN, J., ZHENG, Y., XIE, X., AND SUN, G. Driving with knowledge from the physical world. In *KDD '11: Proc. 17th ACM SIGKDD international conference on Knowledge Discovery and Data mining* (2011), ACM, pp. 316–324. doi:10.1145/2020408.2020462.

[143] YUAN, J., ZHENG, Y., XIE, X., AND SUN, G. T-Drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Transactions on Knowledge and Data Engineering 25*, 1 (2013), 220–232. doi:10.1109/tkde.2011.200.

[144] YUAN, J., ZHENG, Y., ZHANG, C., XIE, W., XIE, X., SUN, G., AND HUANG, Y. T-drive: driving directions based on taxi trajectories. In *GIS '10: Proc. 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2010), ACM, pp. 99–108. doi:10.1145/1869790.1869807.

[145] YUAN, N. J., ZHANG, F., LIAN, D., ZHENG, K., YU, S., AND XIE, X. We know how you live: Exploring the spectrum of urban lifestyles. In *COSN '13: Proc. first ACM Conference on Online Social networks* (2013), ACM Press, pp. 3–14. doi:10.1145/2512938.2512945.

[146] ZHANG, D., LI, N., ZHOU, Z.-H., CHEN, C., SUN, L., AND LI, S. iBAT: detecting anomalous taxi trajectories from GPS traces. In *UbiComp '11:*

Proc. 13th international conference on Ubiquitous computing (2011), ACM, pp. 99–108. doi:10.1145/2030112.2030127.

[147] Zhang, T., Ramakrishnan, R., and Livny, M. BIRCH: an efficient data clustering method for very large databases. In *SIGMOD '96: Proc. 1996 ACM SIGMOD International Conference on Management of Data* (1996), ACM, pp. 103–114. doi:10.1145/235968.233324.

[148] Zheng, K., Zheng, Y., Yuan, N. J., and Shang, S. On discovery of gathering patterns from trajectories. In *ICDE '13: Proc. 29th IEEE International Conference on Data Engineering* (2013), IEEE, pp. 242–253. doi:10.1109/icde.2013.6544829.

[149] Zheng, V. W., Cao, B., Zheng, Y., Xie, X., and Yang, Q. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI '10: Proc. Twenty-Fourth AAAI Conference on Artificial Intelligence* (2010), AAAI press, pp. 236–241.

[150] Zheng, Y. Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology 6*, 3 (2015), 1–41. doi:10.1145/2743025.

[151] Zheng, Y., Capra, L., Wolfson, O., and Yang, H. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology 5*, 3 (2014), 1–55. doi:10.1145/2629592.

[152] Zheng, Y., Chen, Y., Li, Q., Xie, X., and Ma, W.-Y. Understanding transportation modes based on gps data for web applications. *ACM Transactions on the Web 4*, 1 (2010), 1–36. doi:10.1145/1658373.1658374.

[153] Zheng, Y., Liu, F., and Hsieh, H.-P. U-Air: when urban air quality inference meets big data. In *KDD '13: Proc. 19th ACM SIGKDD international conference on Knowledge Discovery and Data mining* (2013), ACM, pp. 1436–1444. doi:10.1145/2487575.2488188.

[154] Zheng, Y., Liu, Y., Yuan, J., and Xie, X. Urban computing with taxicabs. In *UbiComp '11: Proc. 13th international conference on Ubiquitous computing* (2011), ACM, pp. 89–98. doi:10.1145/2030112.2030126.

[155] Zheng, Y., and Xie, X. Learning travel recommendations from user-generated GPS traces. *ACM Transactions on Intelligent Systems and Technology 2*, 1 (2011), 1–29. doi:10.1145/1889681.1889683.

[156] Zheng, Y., Zhang, L., Ma, Z., Xie, X., and Ma, W.-Y. Recommending friends and locations based on individual location history. *ACM Transactions on the Web 5*, 1 (2011), 1–44. doi:10.1145/1921591.1921596.

[157] Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. Mining interesting locations and travel sequences from GPS trajectories. In *WWW '09: Proc. 18th international conference on World Wide Web* (2009), ACM, pp. 791–800. doi:10.1145/1526709.1526816.

[158] Zimmerman, J., Tomasic, A., Garrod, C., Yoo, D., Hiruncharoenvate, C., Aziz, R., Thiruvengadam, N. R., Huang, Y., and Steinfeld, A. Field trial of tiramisu: crowd-sourcing bus arrival times to spur co-design. In *CHI '11: Proc. 2011 annual conference on human factors in computing systems* (2011), ACM, pp. 1677–1686. doi:10.1145/1978942.1979187.