

Semantic Trajectories: A Survey from Modeling to Application

**Basma H. Albanna, Ibrahim F. Moawad, Sherin M. Moussa
and Mahmoud A. Sakr**

Abstract Trajectory data analysis has recently become an active research area. This is due to the large availability of mobile tracking sensors, such as GPS-enabled smart phones. However, those GPS trackers only provide raw trajectories (x, y, t), ignoring information about the activity, transportation mode, etc. This information can contribute in producing significant knowledge about movements, which transforms raw trajectories into semantic trajectories. Therefore, research lately has focused on semantic trajectories; their representation, construction, and applications. This paper investigates the current studies on semantic trajectories so far. We propose a new classification schema for the research efforts in semantic trajectory construction and applications. The proposed classification schema includes three main classes: semantic trajectory modeling, computation, and applications. Besides, we discuss the current research gaps found in this research area.

Keywords Semantic trajectories · Activity recognition · Data modeling · Data segmentation · Semantic applications · Sensor data

1 Introduction

Around 80 % of all the available data have either an explicit or an implicit geographical reference [1]. Explicit references are the actual geometries e.g., city boundaries, lakes, whereas implicit references are textual references to geographical

B.H. Albanna (✉) · I.F. Moawad · S.M. Moussa · M.A. Sakr
Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt
e-mail: basma.hassan@cis.asu.edu.eg

I.F. Moawad
e-mail: ibrahim_moawad@cis.asu.edu.eg

S.M. Moussa
e-mail: sherinmoussa@cis.asu.edu.eg

M.A. Sakr
e-mail: mahmoudsagr@cis.asu.edu.eg

objects e.g., street names, city names, etc. There are objects that change their spatial reference with time, or so-called spatiotemporal objects. With the advancement of the current GPS technologies, large-scale capture of motion of those moving spatiotemporal objects became attainable. Typical examples of moving objects include cars and persons equipped with a GPS device, or animals wearing a transmitter whose signals are captured by satellites [2]. Understanding why and how people and animals move, which places they visit and for which purposes, what are their activities, and which resources they use, is of great importance for decision making in a variety of applications. Case in point, applications like road traffic monitoring, mobile health and animal data ecology, call for methods enabling rich and expressive representation of moving objects.

There have been works providing efficient mobile data management and mining techniques, but they focus on raw trajectories (i.e., a sequence of spatiotemporal observations (x, y, t) using geodetic coordinates). Thus, they ignore the background contextual information (e.g., transportation means and geographical objects) that can contribute in creating significant semantic knowledge about movements. Semantics refer to the contextual information available about the moving object, apart from its mere position data. Semantic is contained both in the geometric properties of the spatiotemporal stream (e.g., when the user stops/moves) as well as in the geographic space on which the object moves (e.g., shops, roads). An example of semantically enriched trajectory could be the following:

(Begin, home, 9 am) → (move, road, 9–10 am, on-bus) → (stop, office, 10 am–5 pm, work) → (move, road, 5–5:30 pm, on-metro) → (stop, market, 5:30–6 pm, shopping) → (move, road, 6–6:20 pm, walking) → (End, home, 6:20 pm)

Semantic trajectory is a growing trend that has recently emerged in geographic information science and spatiotemporal knowledge discovery. It is mainly concerned with understanding the motion of the moving object with respect to the application of interest. Adding semantics enhances the analysis of data and facilitates the discovery of semantically implicit patterns and behaviors. The community created within the FP6 GeoPKDD [3] has initiated most of the research on semantic trajectories with a special focus on privacy and security issues. Following the GeoPKDD, MODAP [4] and SEEK [5] continued the exploitation of knowledge about moving object data.

In this paper, we investigate the existing literature on semantic trajectories and propose a new classification schema for the research efforts done in semantic trajectory construction and applications till now. The proposed classification schema includes three main classes: semantic trajectory modeling, semantic trajectory computation, and semantic trajectory applications. Several similar survey efforts were presented in [6–8], but their main focus was defining the basic concepts and issues about mobility data and surveying techniques for semantic trajectory construction, annotation and knowledge extraction through mining. Our survey extends their work by covering the existing data models supporting semantic trajectory construction besides investigating the activity recognition means and modes (online and offline) for capturing spatiotemporal data. Furthermore, we present an in-depth

survey of trajectory segmentation criteria and demonstrate several applications of semantic trajectories rather than the data mining. Last not least, a major contribution of this paper is the classification schema developed, which maps the existing works in the semantic trajectory research area, discussing each area separately, and identifying the challenges and the potential opportunities within them.

The rest of the paper is organized as follows: Sect. 2 presents the proposed classification schema for the semantic trajectory research work whereas Sect. 3, 4, and 5 surveys the research efforts for semantic trajectory modeling, computation, and applications respectively. Section 6 analyzes the main gaps found in the current research works. Finally, Sect. 7 concludes our conducted study.

2 Classification Schema of Studies on Semantic Trajectories

We present a comprehensive study and analysis for the current research on semantic trajectories. There are three main areas of work that exist in the relevant literature: modeling semantic trajectories, their computation, and application. The modeling area studies which part of the trajectory data will be stored, how it will be accessed, and what kind of semantics will be annotated to it. The computational area discusses the extraction of raw data, its cleaning, compression, segmentation, and annotation.

While the application area proposes different uses of the semantic trajectory data in a variety of applications. Figure 1 presents the proposed classification schema of research studies on semantic trajectories.

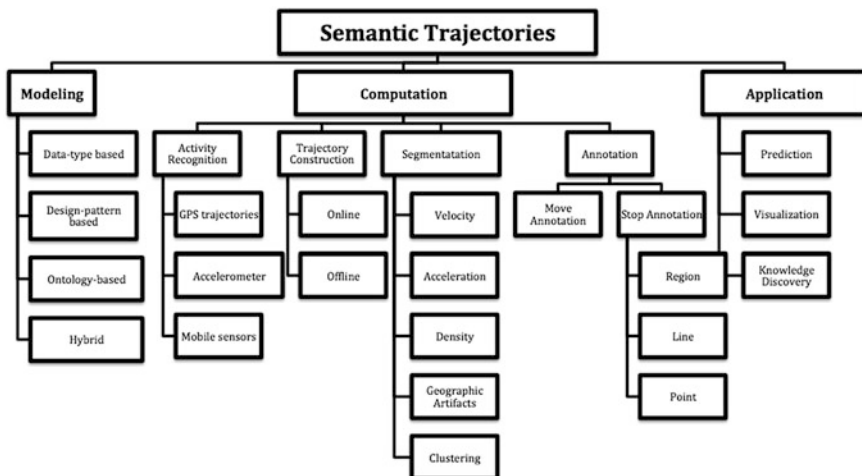


Fig. 1 A new classification schema of studies on semantic trajectories

3 Semantic Trajectories Modeling

The semantic trajectory data modeling is the main task of the semantic trajectory construction. It is the process of defining and analyzing data requirements to support the application of trajectories. There are three main levels of data models that evolve as we progress from the initial requirements to the actual database. The conceptual model maps the initial requirements as technology independent specifications. Following it is the logical data model that defines the document structures that will be used in the database. And finally, the logical data model is transformed into the physical data model, organizing data physically in the database for storage and access. In this paper we classified the semantic trajectory models, regardless of their level of abstraction, into four classes: (1) data type-based, (2) design pattern-based, (3) ontology-based and (4) hybrid data models.

3.1 Data Type-Based Modeling

The research presented in [9] introduced an algebraic model that represents a spatiotemporal trajectory (STT) as an abstract data type (ADT), encapsulating dynamic and semantic features. The ADT was designed in a way that if it got integrated in any database management system, it acquires the same status as built-in data structures. It is also supported with operations covering its spatiotemporal and semantic properties. The STT data type requires different data types varying from integer, boolean, string, enumeration, and constants to represent time, location and activities of spatiotemporal trajectories. A value of type STT is a pair (\mathbf{A}, \mathbf{D}) of temporally ordered sets, where \mathbf{a} , an element in \mathbf{A} is defined as $\mathbf{a} = (l, t_s, t_e, \text{purpose})$ where $l \in \text{Point}$ represents the location of the moving object, t_s and $t_e \in \text{Time}$ and $\text{purpose} \in \text{Enum}$ is the activity description. While \mathbf{d} , an element in \mathbf{D} , is a trip defined as $\mathbf{d} = (l_s, l_e, t_s, t_e, \text{mode}, \text{path})$ where l_s and $l_e \in \text{Point}$, t_s and $t_e \in \text{Time}$, $\text{mode} \in \text{Enum}$ which is the movement mean and the attribute path represents the geometric semantic of the path taken. Along with the data structure proposed, they also introduced a manipulation language composed of operations on the STT data type to formulate semantic operations (e.g., `Activity_Before_Activity`), spatial operations (e.g., `STT_EndsBy_Point`), temporal operations (e.g., `Time_Begins_STT`) and set-based operations (e.g., `Union`, `Intersect`.) A major drawback in this work is that the way the STT data type was designed made it application-dependent, as it represents the concept of space-time trajectories by a series of connected trips and activities. Yet, it provided useful data manipulation operations.

A conceptual model supporting the various requirements of the applications of semantic trajectories was still needed; a model that covers the characterization of trajectories with attributes, semantic and topological constrains and links to application objects. To fill this gap, the authors in [10] introduced *dedicated data types*.

In this research, they brought the minimal information common to all trajectories like the begin, end, moves, stops, as well as their sample points and interpolation functions, and encapsulated them in a generic data type. Whereas the application-specific information that cannot be encapsulated in the generic data type was modeled explicitly using dedicated data types. Those data types contain attributes representing the travelling object or its trajectories and have relationships linking them to the application objects.

To summarize, the ADT modeling approach is best used when the movement track is represented as a set of trips and activities. Whereas the dedicated data type modeling approach is preferable when dealing with trajectories having minimal stops, and where moves are on network-constrained paths that need basic semantics.

3.2 Design Pattern-Based Modeling

Data type modeling approaches alone are not sufficient to support the semantic trajectories application requirements. This is due to the inefficiency of using a generic data type for all application domains. In [11], the authors introduced an extensible model (i.e. trajectory design pattern) relying on the Model Analysis and Decision Support (MADS) model [12] to minimize the effort. MADS supports spatial and temporal objects and relationships (i.e. objects and relationships that have a geometry attribute describing their spatial extent and have a lifecycle attribute describing their temporal extent (the lifespan), and their activity status; active-suspended-disabled).

The trajectory design pattern aims at the explicit representation of trajectories and their components (stops, moves, begin and end) as object types in the database schema and linking those components with application objects. This model requires from the designer to add the semantic information specific to the application. The model provides the designer with a predefined sub-schema that supports the basic data structures for data modeling. Therefore, the trajectory design patterns act as a half-baked schema containing the basic objects and components of the trajectory, and show the relationships between those objects and the application objects. It is considered as half-baked, since it needs from the designer to adjust it and connect it to the rest of the application components.

3.3 Ontology-Based Modeling

Ontology is the conceptualization of a specific domain showing relationships between concepts in the form of a hierarchy. Spatial ontologies became a major research issue for most semantic-aware GIS (Geographic Information Systems) studies.

In [13], a case study was presented on the use of an ontological-based approach for modeling seal semantic trajectories. The modeling approach is based on two main components: domain ontology and time ontology. Those ontologies are a transformation of the semantic seal trajectory after developing it from the World Wide Web consortium. The ontologies represent basic domain and time concepts for the application and show the relationships between them. Along with the ontologies, rules were defined. Some are declarative (ex: travelling is an activity), and others are imperative requiring implementation using Oracle database supporting semantic technologies (ex: travelling is when maximum depth length is larger than 3 m). After that, a semantic integration between the domain and time ontologies is done using queries to understand temporal relationships.

While in [6, 14], the authors presented an ontological approach for modeling semantic trajectories, which integrated domain ontologies with spatial ontologies. It is similar to the approach mentioned earlier integrating domain ontologies with time ontologies. However, it integrates domain ontologies with spatial ontologies to answer queries based on spatial instead of temporal relationships (ex: the activity happened at which area instead of answering a query asking what activities happened during a specific time interval).

A good example of a model based on multiple ontologies is represented in [8], where the authors analyzed modeling requirements for trajectory modeling and proposed a multi-layered trajectory model. First, the raw movement data is transformed into a cleaner version called raw trajectories. These raw trajectories are then transformed into structured trajectories to get a more informative view, where segments correspond to more meaningful steps. Finally, those trajectories experience ontology mapping to add semantics. In this approach, they used three ontologies: (1) Geometric ontology, where the trajectory is perceived as the evolution of geometric location of a moving object during a given time interval, usually captured by mobile devices, (2) geographical ontology, turning the geometric polylines into something with more semantics, and (3) application domain ontology linking application domain knowledge. Figure 2 is an abstract representation we developed to illustrate the model's framework.

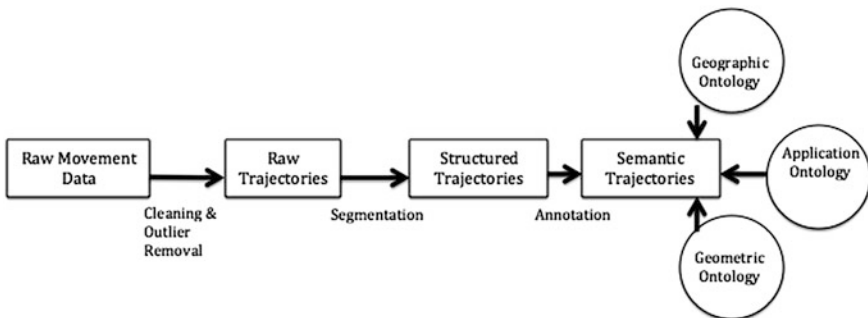


Fig. 2 Modeling using multiple ontologies

3.4 Hybrid Modeling

The proposed hybrid model in [3] encapsulated both the geometry and semantics of mobility data, supporting several levels of abstraction. It contained three models to represent the different levels of abstraction of spatiotemporal trajectories: (1) Raw data model, where raw GPS trajectories are cleaned from uncertainties and outliers to be represented as a stream of spatiotemporal tuples, (2) conceptual model, which abstracts tuples with a certain correlation (like velocity, acceleration, angle of movement, density, time interval, etc.) to become a series of non-overlapping episodes, (3) semantic model, where structured trajectories from the conceptual model are enriched with knowledge from third party sources. This research also introduced a computational platform for the progressive construction and evolution of those three models. An important contribution of this approach was that it offered a consistent framework that aimed at covering the requirements of a variety of applications, starting from those that are only interested in the raw data, to those looking for high-level of application semantic enrichments.

To summarize, choosing the right modeling approach for semantic trajectories depends on several factors. Among them is the application used, the availability of the domain's ontology, the level of trajectory abstraction required and the extent of intervention required by the database designers. *Data type-based* models are generic models that fit into a wide range of applications. They can be made persistent by extending a database model, and can be queried by extending SQL (Structured Query Language). *Design pattern-based* models are even more generic than the data type-based models, as they don't restrict to a specific data type. Instead, a dedicated type relevant to the application in hand can be added to the generic data types but will need the help of a database designer.

On the other hand, *ontology-based* models are application specific, as the ontology needs to reflect the application domain. They can represent richer semantic, and involve any kind of semantic annotations (e.g., multimedia object). In contrast to data type models, ontological models are naturally extensible because ontologies are designed to extend. Whereas the *hybrid model* is the only model that supports applications requiring several levels of abstraction, i.e., performs operations throughout the process of semantic trajectory evolution, going through the raw and structured trajectories. It also fits a wide range of applications, enabling semantic enrichment from several third party sources.

4 Computation

Semantic trajectory computation is the process of extracting and constructing spatiotemporal instances from large-scale GPS feeds, followed by semantic enrichment to comply with a predefined data. We overviewed the various stages of

semantic trajectory computation, going through the activity recognition to the segmentation and annotation. Besides investigating the different modes of computation (i.e., online and offline).

4.1 Activity Recognition

This section illustrates the extraction component of the semantic trajectory computation by showcasing the activity recognition studies conducted from GPS, accelerometers, and mobile sensing device data.

4.1.1 Activity Recognition from GPS Trajectories

Many studies focus on activity recognition using GPS-based trajectory data, where the movement history of the individual is extracted in conjunction with the semantics of the location (typically from geographic or application data repositories).

To identify the important locations from GPS trajectories, studies like [15, 16] proposed methods of joining the GPS trajectory data with predefined points of interest (POI), having specific time constraints for inferring activities. For example, given a set of trajectories, a set of POIs, and an activity mapping set show possible activities that might take place and their corresponding durations; find the sequence of activities that might be performed during those set of trajectories. The rationale behind it is that if a user stays at a POI for long enough time, then some activity might take place. So, it answers questions like which POI's did the user stay in? And what activities were performed in it?

When there are no predefined POI's, a clustering method can be used as suggested by [11, 17] to automatically discover hotspots in the trajectory data. In [17], the authors discovered stops or interesting places using speed-based methods, where the distance between points is calculated along the trajectory instead of the traditional Euclidean distance. They considered the notion of minimal time instead of minimal number of points for a region to be considered dense. The *minimum time* duration indicates the minimum time necessary to generate a cluster. It is calculated by subtracting the timestamp of the first point in the cluster from the last point's timestamp in the same cluster. While in [11], the POIs were detected using the DJ-Cluster algorithm, where for each point, a neighborhood is calculated. The neighborhood consists of points within distance Eps , under the condition that there are at least $MinPts$ of them. If no such neighborhood is found, the point is labeled noise. The DJ-Cluster algorithm has several important technical advantages: it allows clusters of arbitrary shape; ignores outliers, noise, and unusual points; has more easily chosen parameters; and has deterministic results.

The previous activity recognition studies are about the location part of trajectory data, stating, "What they move for". Another very interesting study in the literature was the recognition of the transportation modes to understand "how they move".

For example, researchers in [18] designed a methodology for detecting the transportation mode using a set of variables like acceleration, velocity, median speed, etc. Following this direction, the authors in [19] provided a more solid approach for identifying the transportation mode through a three-step framework to recognize means of transportation; first, by the segmentation of change points, second is the mode detection through a predefined decision tree and the third stage is to apply graph-based post-processing to refine the results.

4.1.2 Activity Recognition from Accelerometer Data

“A tri-axial accelerometer is a sensor that can collect a real valued estimate of acceleration along three axes, i.e., x, y and z” [6]. It has been largely used in activity recognition specifically in activities like running, walking, climbing steps, gym instruments, etc.

The most cited study in this regard was conducted in [2], where the authors were the first to use multiple accelerometer sensors worn in different parts of the body to detect common activities. The problem with this approach was that it required certain laboratory conditions, i.e. not easily applicable in normal circumstances. Further research has been developed in [20], making this approach more user friendly and enhancing its mobility by only using one accelerometer.

4.1.3 Activity Recognition from Mobile Phone Sensors

Mobile phone sensors activity recognition is done through the use of wireless devices like smart phones to understand what people do, where they go, and how they interact with each other. Combining accelerometer data with mobile phone audio data through a microphone to better detect the activity is an example. Several studies have been conducted in this field [1, 21, 22], where they used smart phone embedded sensors and data records, like GPS, GSM cell tower, call and SMS logs, Wi-Fi, Bluetooth, accelerometer, and audio features for mining people’s activities.

4.2 Trajectory Construction

The semantic trajectory construction is the process of integrating the spatiotemporal movement characteristics with useful information regarding objects movement patterns and social activities. There are two modes for semantic trajectory construction: (a) Offline mode, where all trajectory construction processes are done offline, and (b) online mode, where parts of the trajectory construction processes are done in real time.

4.2.1 Online Mode

In the current literature, timely trajectory computation to serve real-time queries for today's trajectory applications (ex: traffic monitoring) is not sufficient. To fill this gap, the authors in [23] proposed SeTraStream, a platform for online semantic trajectory construction. The main contributions of SeTraStream can be summarized as follows:

1. Online trajectory preprocessing: trajectory preprocessing was redesigned to include online cleaning using the kernel smoothing method, and online compression using the synchronized Euclidian distance and correlation coefficient.
2. Online trajectory Construction, where they designed techniques for episode identification during the online trajectory segmentation. Some of the above offline works can adapt to an online context. Yet, none of them support the exploitation of the profound semantics that exist in the computed trajectories in real-time.
3. Platform implementation and evaluation: an online framework that enables semantic trajectory construction over streaming movement data tackling real time streaming environments.

The flow is as follows: The server receives from the mobile object device a batch of GPS data with a predefined window size and stream complementary features, like acceleration, speed, displacement, etc. Consequently, cleaning, smoothing and compression techniques are applied. Finally, feature vectors are extracted, and a corresponding matrix is formed and the batch is buffered until segmentation takes place. During the segmentation, previously buffered batches are de-queued and matched with dissimilar batches (based on RV-coefficient) to form an episode. Having detected an episode, SeTraStream defines the triplet (semantic tagging) describing its start and end time bounded to a specific geometry. With this mode of computation, new challenges to the conventional methods came to existence. As in the offline mode algorithms, threshold tuning is common. While in the online context, parameter tuning is prohibitive. In Fig. 3, we represented the flow of the online trajectory construction.

4.2.2 Offline Mode

In this mode, the movement data in the form of large-scale GPS datasets is collected in advance. The processing undergoes several stages starting from data refinement, tuning, map matching and compression to trajectory identification, and eventually trajectory segmentation and annotation. The offline trajectory-computing framework used for a specific application should reflect its semantic trajectory modeling requirements. For example, the authors in [3] designed an offline trajectory computational framework matching the requirements of the hybrid spatiotemporal model they proposed. The framework is composed of three layers:

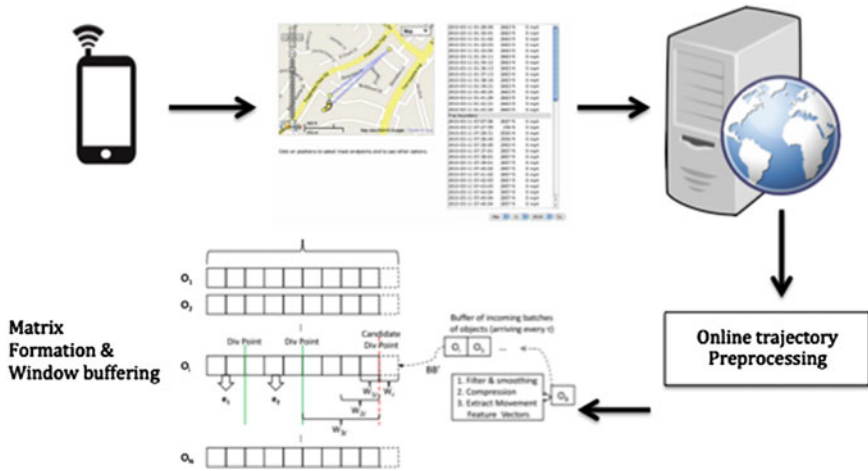


Fig. 3 Flow of online trajectory construction

1. The data preprocessing layer, where the outlier removal, kernel smoothing and compression stages occur. Several works have been conducted in this specific area as in [10, 24–27].
2. The trajectory identification layer, which is responsible for dividing the processed GPS raw data into trajectories using different policies (ex: GPS gap, predefined time interval, predefined space extent) [3, 8 28].
3. The trajectory structure layer that works on the identified trajectories. It further divides them into episodes i.e., meaningful stop and moves ready for semantic tagging/annotation using geographic artifacts, speed, velocity and direction based methods [3, 15, 29, 30].

In [8], a similar computational model was used, adding the semantic enrichment stage to the trajectory structure layer. It was customized for the multi-layered model mentioned earlier by linking the spatiotemporal units with semantic knowledge from the geographic data and application domain data.

Research is still needed to substantially reduce the amount of raw data, while not missing valuable information. On the fly analysis techniques are also required for data processing. This is because it is unaffordable to store first then reduce afterwards with the data’s exponential inflation nature. Existing work also assumes well-recognized constraints on valid data or well-understood error models; but for many emerging big data domains, these do not exist.

4.3 Segmentation

The authors in [31] proposed the first data model looking at the trajectories from a conceptual point of view, where they divided the trajectory into a set of stops and

moves. From this starting point, different works have been proposed to instantiate the model of stops [30, 32]. A stop can be defined as “the important places where a trajectory has passed and stayed for a reasonable time duration” [6]. For this kind of segmentation, different approaches have been proposed as follows.

4.3.1 Velocity-Based

The velocity-based approach [6] focused on stops and moves, where it determines if a GPS point belongs to a stop or to a move episode by using a speed threshold. Hence, if the instant speed of p is lower than the threshold, it is a part of a stop; otherwise it belongs to a move.

4.3.2 Density-Based

Using only velocity for identifying stops is not enough for some scenarios. Therefore, the authors in [6] designed a density-based stop discovery approach. It considered not only the speed but also the maximum diameter that the moving object has traveled during a given time duration.

4.3.3 Geographic Artifacts

Trajectories and geographic data overlap in space. In [15], the authors integrated geographic data with sub-trajectories overlapping in geometry. This is done in a user-dependent way, where the user identifies which places are of interest to his specific application to disregard any geographic places out of the application’s interest. They devised the algorithm SMoT (Stop and Moves of Trajectories) that verifies for each point of the trajectory if it intersects the geometry of a candidate stop (i.e. a geographical place related to the application) and that the duration of the intersection is at least equal to a specific predefined threshold.

4.3.4 Clustering-based

An extension to SMoT [15] was developed in [30] using the method CB-SMoT, which stands for Clustering Based—SMoT. It used a clustering technique in order to identify stops according to Spaccapietra’s stop definition. In [22], instead of comparing each and every point with the geometry of the geographic place, clusters of trajectories were identified beforehand according to their speeds and then they were mapped to geographic places to add semantics to those clusters.

4.4 Annotation Approach

This is the stage where trajectories are transformed into semantic trajectories in the computation stage. It is the task following the trajectory segmentation where meaningful information is assigned to specific intervals and sections of the moving object's movement track.

4.4.1 Annotating Moves

The annotation techniques mentioned above were mainly concerned with annotating the stops defined in [31] or annotating trajectory episodes introduced earlier in [33]. They defined an episode as “a discreet time period for which the user's spatiotemporal behavior was relatively homogeneous”. Very few research works [5, 23] had their focus mainly on annotating moves. Annotating moves is necessary because not every stop in the physical trajectory possesses (application dependent) interpretation. The semantic stops can happen without appearing in the data.

4.4.2 Stop Annotations

Stopping in a trip means that there is something of interest to do. So stop annotation is about mapping stops to places of interest, which can be geographical regions, roads in the form of lines, or POIs in the form of points.

- a. **Regions:** Annotating trajectories with regions of interest from geographical or application domain sources. It does so by computing topological correlations between trajectories and 3rd party data sources containing semantic places of regions [3, 6, 15].
- b. **Lines:** It is the annotation of trajectories with lines of interest like road networks. Given data sources of different forms of road networks, the purpose is to identify correct road segments, as well as, infer transportation modes such as walking, cycling, and public transportation like metro e.g., [19, 34].
- c. **Points:** It is the annotation of stop episodes of a trajectory with information about a suitable point of interest. Examples are shown in [4, 5, 32]. However, densely populated urban areas bring several candidate POIs for a stop. In addition, low GPS sampling rate (due to battery outage and GPS signal losses) makes the problem more intricate. Therefore, the authors in proposed the Hidden Markov Model (HMM)-based technique for semantic annotation of stops, which was able to overcome those problems. In the Hidden Markov Model, the state is not directly visible, but the output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Thus, the sequence of tokens generated by an HMM gives some information about the sequence of states. It can be presented as the simplest dynamic bayesian network.

5 Semantic Trajectory Applications

Adding meaning to the movement track of moving objects opened new perspectives for a large number of applications built on the semantic of movements of objects. This section classifies the state of the art applications into trajectory prediction, visualization and knowledge discovery applications.

5.1 Prediction

Many applications, such as location-based advertisement, navigational planning services and traffic management, have been developed for the location-based services market. Those applications require accurately predicting the next move of the moving object. The first to predict destinations from partial trajectories where the authors in [35] described a method called predestination that uses the history of a driver's destinations to predict how his trip will progress later on. Another example of prediction was a model developed in [36] where prediction was based on social spatial approximation, which utilizes current GPS coordinates of user friends to estimate GPS coordinate of the user. The authors in [37] proposed a novel approach named GTS-LP for mining and prediction of mobile user's movement behavior. They defined a new pattern, called the GTS-Pattern, to represent frequent moves, which based on it they proposed the location prediction strategy.

5.2 Visualization

An effective way for semantic trajectories analysis is to visualize the movement track. In [38], the main plot area used to visualize the trajectories was a 3D cube with three axes, the x-y geographical location and the time axis, where trajectory data and domain ontology were mapped into 3D cubes. Another research was conducted using Weka-STPM [39], with new pre-processing methods and a graphical GUI to visualize in a map the spatial entities and the generated stops and moves. Another example of a system enabling trajectory visualization is MoveMine [40], which provides a user friendly interface where users can select a data set and the corresponding raw data is plotted on the Google Map. Furthermore, a user can plot the results in Google Earth for 3-D visualization of the results.

5.3 Knowledge Discovery

There are approaches that exploit semantic trajectories for knowledge discovery, in particular movement patterns. Among them is [41], which proposed a novel

methodology for recognizing the behavior of moving objects within stops. This was done by further dividing a stop into sub-stops using velocity/direction based rules.

While in [34], the authors developed a pattern mining framework which detected moving patterns between two stops considering background geographical information, e.g., pattern of movement of tourists between touristic places. Several other works [15, 30] developed similar pattern and knowledge mining techniques for a pool of applications, ranging from identifying tourists' POIs to understanding moving object behaviors and trajectory goals.

Furthermore, a scalable reference framework for the semantic management of moving objects called SemanticMOVE was proposed in [42], which supports better mining, analysis and reasoning of semantic mobility data. It's a generic architecture with an infrastructure of distributed nature where each object collects, stores, processes and analyzes the semantics of its own data.

6 Research Gaps

During our extensive study of semantic trajectories, several research gaps have been deduced throughout the previous studies and literature concerned with semantic trajectory construction and application. These gaps include:

- The data type-based models need to be less application-dependent and more generic to include the wide range of scientific domains, besides the advancement in manipulation languages for querying and knowledge discovery.
- In ontology-based modeling, research on applying more domain conditions on rules is becoming a necessity to reduce time and space storage inference complexity.
- In semantic trajectory extraction and activity recognition, more research is needed to address their use, and how they can be integrated with online computational platforms and geographical maps.
- There is a huge research opportunity in the area of trajectory segmentation using means rather than the episodes and stop and moves identification models.
- More research is required to focus on annotating moves, because a huge part of the semantics of moving objects lies in the movement activity rather than activities done at stops, besides adding to the logic behind the semantics at stops. Also, better stop analyses can be made via careful tuning (e.g., tuning stop identification and interpretation to make it work even for short stops).
- To the best of our knowledge, online mode algorithms for semantic trajectory construction are significantly missing. In current online mode research, the tagging needs to be customized according to different application contexts by modifying the feature vector (with features like segment distance, duration ... etc.), besides using the corresponding suitable tagging technique including decision trees, neural, and bayesian methods.
- We are living in the era of 'Big Data'. Spatiotemporal trajectories, whether captured through remote sensors or large-scale simulations, has always been

‘Big’. However, recent advances in instrumentation and computation made spatiotemporal data even bigger, putting several constraints on data analytics capabilities. Spatial computation needs to be transformed to meet the challenges posed by the big spatiotemporal trajectories.

- For semantic trajectory application, more innovative research is also expected through integrating traditional knowledge extraction techniques with visualization approaches, and with knowledge extracted from social network interactions.

7 Conclusion

In this paper, we discussed the main components of the semantic trajectory processing by analyzing the state of the art and past research contributions in this field. The relative novelty of the domain leaves many challenges, opportunities and extended studies open for future work, which we addressed most of them in our deduced research gaps. We were able to conclude the analysis and insights of our study as follows: (1) Starting with the trajectory extraction component, most of the literature focused on the conventional GPS tracking devices disregarding the wide penetration of smart phones that can be used for a broader range of applications in real time context, (2) from a data modeling perspective, several spatiotemporal models have been developed to include the semantic dimension. The hybrid models are the only variant that support different levels of data abstraction by representing trajectories in terms of both spatial and semantic mobility characteristics, (3) an essential component of semantic trajectory construction is the segmentation. The most common method of segmentation is the stop and move, which was the basis of many studies focusing on stop discovery techniques relying on speed, velocity, acceleration, direction, geographic artifacts and clustering algorithms, (4) research in semantic annotation of trajectories is either in annotating moves or in stop annotation, where stops are characterized as regions, lines, or points, (5) semantic trajectory applications fall in three main categories; knowledge discovery, visualization and prediction. There is a need to develop applications targeting large and deforming objects (e.g., oil spills, diseases ... etc.), network-constrained movements, relative movement, and collective movement for any kind of collections of objects, and finally (6) we have given an extensive survey of works done on the aspects of semantic trajectories. We have also highlighted research gaps in those areas to call for future work.

References

1. Yan Z, Chakraborty D, Misra A, Jeung H, Aberer K (2012) Semantic activity classification using locomotive signatures from mobile phones
2. Bao L, Intille S (2004) Pervasive computing, vol 3001. Springer, Berlin, pp 1–17
3. Yan Z, Parent C (2010) The semantic web: research and applications, vol 6088. Springer, Berlin, pp 60–75

4. Yan Z, Chakraborty D, Parent C, Spaccapietra S, Aberer K (2011) SeMiTri. In: Proceedings of the 14th international conference on extending database technology—EDBT/ICDT '11, p 259
5. Yu F, Ip HHS (2008) Semantic content analysis and annotation of histological images. *Comput Biol Med* 38(6):635–649
6. Zhixian Y (2011) Semantic trajectories: computing and understanding mobility data In: Ph.D. dissertation, Swiss Federal Institute of Technology, Information and Communication Dept., Lausanne
7. Parent C, Pelekis N, Theodoridis Y, Yan Z, Spaccapietra S, Renso C, Andrienko G, Andrienko N, Bogorny V, Damiani ML, Gkoulalas-Divanis A, Macedo J (2013) Semantic trajectories modeling and analysis. *ACM Comput Surv* 45:1–32
8. Yan Z, Spaccapietra S (2009) Towards semantic trajectory data analysis: a conceptual and computational approach. *VLDB PhD Work* 15(2):165–190
9. Zheni D, Frihida A (2009) A semantic approach for the modeling of trajectories in space and time. In: *Advances in conceptual modelling*, pp 347–356
10. Marketos ARG, Frentzos E, Ntoutsis I, Pelekis N, Raffaetà A, Theodoridis Y (2008) Building real-world trajectory warehouses. In: *MobiDE*
11. Zhou C, Frankowski D, Ludford P, Shekhar S, Terveen L (2004) Discovering personal gazetteers. In: *Proceedings of the 12th annual ACM international workshop on geographic information systems—GIS '04*, p 266
12. Parent C, Spaccapietra S, Zimányi E (2006) *Conceptual modeling for traditional and spatio-temporal applications: the MADS approach*. Springer, New York, p 450
13. Wannous R, Malki J, Bouju A (2013) Time integration in semantic trajectories using an ontological modelling approach A case study with experiments, optimization and evaluation of an integration approach. Springer, Berlin
14. Wannous R, Malki J, Bouju A, Vincent C (2013) *Modeling approaches and algorithms for advanced computer applications*, vol 488. Springer International Publishing, Cham
15. Alvares LO, Bogorny V, Kuijpers B, de Macedo JAF, Moelans B, Vaisman A (2007) A model for enriching trajectories with semantic geographical information. In: *Proceedings of the 15th annual ACM international symposium on advances in geographic information systems—GIS '07*, p 1
16. Xie K, Deng K, Zhou X (2009) From trajectories to activities. In: *Proceedings of the 2009 international workshop on location based social networks—LBSN '09*, p 25
17. Zhao XL, Xu WX (2009) A clustering-based approach for discovering interesting places in a single trajectory. In: *2009 2nd international conference on intelligent computing technology and automation, ICICTA 2009*, vol 3, pp 429–432
18. Schuessler N, Axhausen KW (2009) Processing raw data from global positioning systems without additional information. *Transp Res Rec J Transp Res Board* 2105(1):28–36
19. Zheng Y, Chen Y, Li Q, Xie X, Ma W-Y (2010) Understanding transportation modes based on GPS data for web applications. *ACM Trans Web* 4(1):1–36
20. Ravi N, Dandekar N, Mysore P, Littman M (2009) *Distributed computing, artificial intelligence, bioinformatics, soft computing, and ambient assisted living*, vol 5518. Springer, Berlin
21. Kiukkonen N, Blom J, Dousse O (2010) Towards rich mobile phone datasets: Lausanne data collection campaign. *Proc ICPS*, Berlin
22. Choujaa D, Dulay N (2010) Predicting human behaviour from selected mobile phone data points. In: *Proceedings of the 12th ACM international conference on Ubiquitous computing—Ubicomp '10*, p 105
23. Yan Z, Giatrakos N, Katsikaros V (2011) *Advances in spatial and temporal databases*, vol 6849. Springer, Berlin, pp 367–385
24. Schüssler N, Axhausen KW (2009) Processing GPS raw data without information. *Transp Res Rec J Transp Res Board* 8:28–36
25. Douglas D, Peucker T (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Can Cartogr* 10(2):112–122

26. Meratnia N, de By RA (2004) Spatiotemporal compression techniques for moving point objects. In: EDBT
27. Potamias M, Patrumpas K, Sellis T (2006) Sampling trajectory streams with spatiotemporal criteria. In: 18th international conference on scientific and statistical database management (SSDBM'06), pp 275–284
28. Yan Z, Chakraborty D, Parent C, Spaccapietra S, Aberer K (2013) Semantic trajectories. *ACM Trans Intell Syst Technol* 4(3):1
29. Yan Z, Spremic L, Chakraborty D, Parent C, Spaccapietra S, Aberer K (2010) Automatic construction and multi-level visualization of semantic trajectories. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems—GIS '10, vol 12, p 524
30. Xiu-Li Z, Wei-Xiang X (2009) A clustering-based approach for discovering interesting places in a single trajectory. In: 2009 second international conference on intelligent computation technology and automation, pp 429–432
31. Spaccapietra S, Parent C, Damiani ML, de Macedo JA, Porto F, Vangenot C (2008) A conceptual view on trajectories. *Data Knowl Eng* 65(1):126–146
32. Vald F, Damiani ML, Güting RH, (2010) The semantic web: research and applications, vol 6088, Part 2. Springer, Berlin, pp 450–453
33. Mountain D, Raper J (2001) Modelling human spatio-temporal behaviour: a challenge for location-based services. In: Proceedings of the sixth international conference on geocomputation
34. Alvares LO, Bogorny V, de Macedo JAF, Moelans B, Spaccapietra S (2007) Dynamic modeling of trajectory patterns using data mining and reverse engineering. In: Tutorials, posters, panels and industrial contributions at the 26th international conference on conceptual modeling, vol 83, pp 149–154
35. Krumm J, Horvitz E (2006) Predestination: inferring destinations from partial trajectories. In: Proceedings of the 8th international conference on Ubiquitous computing (UbiComp '06), pp 243–260
36. Backstrom L, Sun E, Marlow C (2010) Find me if you can: improving geographical prediction with social and spatial proximity. In: Proceedings of the 19th international conference on world wide web, pp 61–70
37. Ying JJ-C, Lee W-C, Tseng VS (2013) Mining geographic-temporal—semantic patterns in trajectories for location prediction. *ACM Trans Intell Syst Technol* 5(1):1–33
38. Bakshev S, de Macêdo J, Spisanti L (2010) Semantic visualization of trajectories. In: ICEIS, pp 1–10
39. Alvares L, Palma A, Oliveira G, Bogorny V (2010) Weka-STPM: from trajectory samples to semantic trajectories. In: Proceedings of the Workshop Open Source Code, vol 1
40. Li Z, Ji M, Lee J, Tang L, Yu Y, Han J, Kays R (2010) MoveMine. In: Proceedings of the 2010 international conference on management of data—SIGMOD '10, p 1203
41. Moreno B, Times V, Renso C, Bogorny V (2010) Looking inside the stops of trajectories of moving objects. In: Geoinfo
42. Ilari S, Stojanovic D, Ray C (2015) Semantic management of moving objects: a vision towards smart mobility. *Expert Syst Appl* 42(3):1418–1435