Chapter 1
Modelling mobility data and constructing large knowledge graphs to support analytics: The datAcron ontology

Georgios M. Santipantakis, George A. Vouros, Akrivi Vlachou and Christos Doulkeridis

Abstract This chapter presents modelling and representation techniques for mobility data, focusing on semantic representations that build around the central concept of semantic trajectory. Moving from mobility data to enriched representations of positional information, associated with contextual data but also with weather data, and furthermore with events that occur during the movement of an object is critical to support advanced mobility analytics. Motivated by these requirements, we propose a new ontology that satisfies these requirements to a larger extent than previous works on semantic representations of trajectories, at multiple, interlinked levels of detail. In addition, we show that this ontology supports data transformations that are required for advanced analytics, such as visual analytics, and we present meaningful use-case scenarios in Air Traffic Management and maritime domains.

This chapter presents modelling and representation techniques for mobility data, emphasizing on the representations of semantic trajectory. These representations enable the transition from mobility data to positional information associated with contextual data that is critical for advanced mobility analytics, including weather data, and events that occur during the movement of an object. We propose a new ontology that satisfies these requirements to a larger extent than previous works on semantic representations of trajectories, at multiple, interlinked levels of detail. We show that this ontology supports data transformations that are required for advanced analytics, such as visual analytics, demonstrating meaningful use-case scenarios in Air Traffic Management and in the maritime domains.

Georgios M. Santipantakis
University of Piraeus, e-mail: gsant@unipi.gr

George A. Vouros
University of Piraeus e-mail: georgev@unipi.gr

Akrivi Vlachou
University of Piraeus e-mail: akrivi.vlachou@gmail.com

Christos Doulkeridis
University of Piraeus e-mail: cdoulk@unipi.gr

Analysis of mobility data is often at the core of several tasks in critical domains w.r.t. economy and safety. This tasks often are based on the combination of surveillance data with descriptions of the moving objects, (e.g. geometric information, objects' physical and operational characteristics), and contextual information (e.g. areas and points of interest, weather information, traffic, etc.), originating from disparate and heterogeneous data sources.

Challenging problems include effective information provision for situation awareness, identification of recurrent patterns, decision making at different scales and levels of abstraction, as well as, the prediction of moving objects' behaviour under specific circumstances. These challenges are significant, given that their achievement aims to reduce factors of uncertainty regarding operations, enhance punctuality of activities, advance planning efficiency, and reduce operational costs in time critical domains, such as aviation and maritime. The complexity of these challenges increases significantly to the number of moving objects. Towards reducing this complexity, a shift of operations' paradigm from *location-based* to a *trajectory-based* has been proposed. Trajectories are turned into the main asset and placed in the core of decision making, assessment of situations and planning of operations tasks.

Towards addressing these challenges, we need to consider how we represent trajectories to satisfy the data needs and requirements of analysis tasks. The proposed approach is based on two principles:

- Trajectories should reveal objects' behaviour in explicit terms, at different levels of abstraction considering their geometric, contextual, and analysis-specific features. In doing so, analysis tasks can retrieve data about trajectories at any level of abstraction that is appropriate for their purposes, switching between abstraction levels, delving into the details of mobility phenomena and providing overviews in generic terms.
- Data transformations (or conversions) require trajectories to integrate spatial events into temporal sequences, while, on the other hand, these events need to be aggregated into spatial time series, associated to geographic contexts. Combining these abilities, allows identifying re-occurring patterns of behaviour at varying levels of abstraction, enhancing our understanding of mobility phenomena and thus, decision making. In this chapter, an "abstraction" is considered any possible combination of aggregation and generalization. When it is necessary, we specify explicitly which kind of abstraction is required.

This chapter describes an ontology for modelling semantic trajectories, integrating spatio-temporal information regarding mobility of objects at multiple, interlinked levels of abstraction, supporting appropriate data transformations, as needed by visual analysis tasks. Visual analytics impose specific requirements to support the combination of human and computational data processing through interactive visual interfaces, enabling analysis of spatio-temporal and mobility data, sophisticated data analysis, and informed decision making, at varying levels of abstraction.

Existing models and ontologies for the representation of semantic trajectories do not associate data and events at multiple levels of abstraction. They usually specify models for representing trajectories at different levels (from raw to semantic), where each level associates trajectories with a different kind of information. In models where some form of abstractions are supported, these are restricted to specific types and levels. Consequently, switching between levels of abstraction as needed by exploratory analysis tasks is limited, thus these representation models can hardly serve tasks for visual analytics.

The main contributions of the ontology proposed in this chapter are the following:

- We revisit fundamental data types for visual analysis tasks revolving around the notion of semantic trajectory, specifying conversions among these types of data. These types and conversions provide an in-principle framework for identifying trajectories' constituents, as well as, a comprehensive framework for validating ontological specifications towards the provision of appropriately transformed data, satisfying data requirements of visual analysis methods.
- We revisit the notion of "semantic trajectory", as a meaningful sequence of trajectory parts at any level of abstraction. By being meaningful, a semantic trajectory is associated with human-interpretable and machine-processable information, revealing objects behaviour in explicit terms. Dealing with multiple levels of abstraction, we support analysis of moving objects' behaviour at any scale and/or level of abstraction that is appropriate for analysis tasks.
- We demonstrate the ontology by means of enhanced SPARQL queries, using real-world data from the Air Traffic Management domain, and maritime domain.

The geometric, geographic and application specific information are some of the features necessary to the representation of semantic trajectories.

The geometric information enables queries like "Return objects which were located at x, y, z at time t" and may be specified at various levels of aggregation. It can reveal knowledge regarding the patterns of a moving object at different spatio-temporal scales. For example, computations regarding spatial/topological relations or patterns of movement are often easier when a trajectory is represented as a line, rather than a sequence of positions. Alternatively, a trajectory can be represented as a temporal sequence of lines representing sub-trajectories, each one of special interest on its own (e.g. each one crossing a specific region of interest, or corresponding to a specific phase of movement), or as a sequence of aggregated raw positions with high concentration in spatio-temporal regions or points of interest.

The geographic and application specific information associated to a trajectory at multiple levels of geometric abstraction, is also important. The usefulness of having multiple levels of geometric abstractions is that each one serves different purposes towards representing and analysing the behaviour of moving objects. This information enables queries like "Return objects that crossed the spatial region X during the time interval $[t_{begin}, t_{end}]$", "Return objects whose trajectories crossed spatial regions that properly include region X during the time interval $[t_{begin}, t_{end}]$", and "Return objects whose trajectories include an aggregation of positions close to a specific point of interest".

Different levels of geometric abstraction provide alternative constituents for structuring trajectories. A *structured trajectory* consists of a sequence of *trajectory parts* that can be either *raw positions* reported from any sensing devise, aggregations of raw positions referred as *nodes*, or *trajectory segments*. A *trajectory segment* is a trajectory itself, which may be part of a whole trajectory. Segments and nodes aggregate information that may instantiate a behaviour pattern. For example, a sequence of raw positions may instantiate a "turn" or a "stop" event. These aggregations can be represented by a single node or segment, associated to an event type (e.g. "turn" or "stop", respectively), and to the corresponding set of raw positions.

Segments of trajectories and nodes can be defined with different objectives depending on the application and target analysis, and are thus associated with application-specific information. A maximal sequence of raw data that comply with a given pattern defines an *episode*. In this work we consider *events* as a generalisation of episodes. *Events* represent specific or abstract happenings and are associated to trajectory parts, providing application specific information that is relevant to the trajectory. As a consequence, queries such as "Return objects whose trajectories contributed to congestion events in a specific spatio-temporal region", or "Return objects whose trajectories comprise a segment that is associated with a high-speed event" can be answered.

Geographical features allow turning the geometric information representing the spatial path into a geographical trace which is meaningful for humans and computational processing tasks. This requires associating trajectory parts to (types of) geographic regions: Shops/spots/buildings of different kinds, regions of special interest (e.g. touristic, commercial or industrial), etc. Generalising geographical features, we can draw semantic associations between trajectory parts, supporting further the abstraction of trajectories (e.g. any trajectory crossing many shops can be a "shopping trajectory", irrespectively the kind of shops crossed. Specific types of shopping trajectories may indicate specific types of shops crossed). In this chapter, we generalize geographical features to *contextual*. This comprises features of the moving objects, as well as features of moving objects' environment, considering that these features are associated to objects' movement. These may include weather attributes, space configuration features, as well as aggregated data about co-occurring trajectories – i.e., traffic. As a consequence, queries such as "Return trajectories that crossed any region with specific weather conditions [specified as conditions in weather attributes]" can be also answered.

A trajectory part may be associated with any event that co-occurs with it spatially and/or temporally: E.g. Bad weather conditions, or traffic regulations associated with a spatial region may co-occur with a trajectory crossing-it (thus, related spatially) during a time period (related temporally).

A *semantic trajectory* is a sequence of trajectory parts, associated with contextual information and related events. The association with such information, reveals objects' deliberative or accidental behaviour in explicit terms, thus contributes to understanding the rationale for that behaviour. The semantic trajectory can be also specified at different levels of abstraction, depending on the geometric features, contextual features, and events considered. Abstraction may happen by means of aggregation, generalization, or both. In doing so, we may retrieve semantically associated trajectories, based on the semantic features they aggregate and information to which they are associated. For instance we may request “trajectories crossing sensitive areas and associated to suspicious events”.

As a concrete and simple example of a trajectory specified at multiple levels of abstraction, Figure 1 shows the representation of a trajectory crossing an airspace compartment: The trajectory is represented both as a geometry projected in two dimensions, and as a temporal sequence of trajectory segments, which are indicated in different colour, depending on whether each segment occurs within the compartment or not. This structure results through a topological link discovery process where the trajectory geometry is used as a first indication of the potential fact that the trajectory crosses the air compartment (filtering step). This is further verified by exploiting the raw trajectory positional data and identifying the trajectory segments that spatially occur within the compartment. Additional information to trajectory segments is provided by associated events that are not shown in the figure, to keep it simple. Hence, beyond the representation of the trajectory as a sequence of trajectory segments, at a second level of abstraction, the trajectory is represented as a temporal sequence of semantic nodes, each one signifying an important event occurring across the trajectory. For instance, trajectory nodes H, L, M, and K are associated to entry/exit events, representing the relation of raw positions with the airspace compartment. Trajectory segments and nodes are further associated to positional raw data.

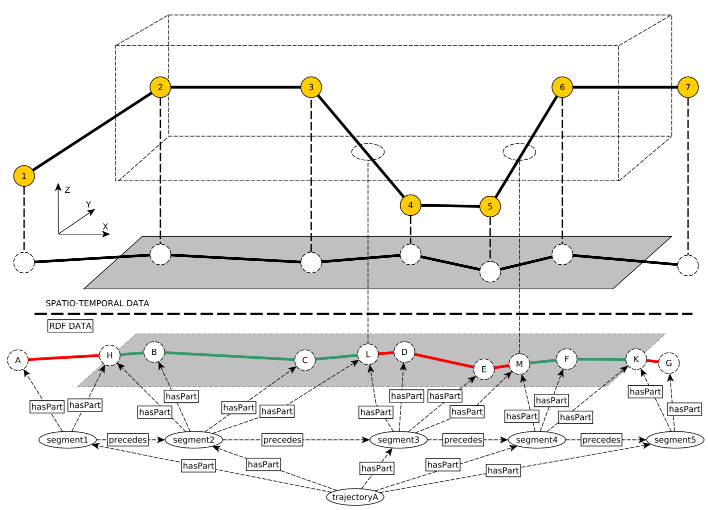

Fig. 1.1 Example of a multiple levels of abstraction of a trajectory.

Apparently, abstractions of a single trajectory should be interlinked so as any application to be able to get any relevant information that is necessary for its purposes. This also allows transitions between specialized / basic information and generalized / aggregated information, through querying and applying data transformations.

Individual trajectories provide information on the movement of individual objects. Aggregated traffic data are *spatial time series* describing how many moving objects were present in different spatial locations and/or how many objects moved from one location to another during different time intervals. The time series may also include aggregate characteristics of the movement, such as the average speed and travel time. Time series describing the presence of objects are associated with distinct locations, and time series describing aggregated moves (often called fluxes or flows) are associated with directed links between pairs of locations. In both cases *spatial time series* are represented as chronologically ordered sequences of values of time-variant thematic attributes associated with spatial locations or spatial entities (for example, regions of special interest).

Spatial events emerge at spatial locations and exist for a period of time. Spatial events are described by their spatial regions, existence times, and contextual features. Events may occur irrespectively of trajectories, but somehow be related to trajectories (e.g. weather events, regulations imposed in a spatio-temporal region), or may be derived from trajectories (e.g. a turn of a moving object, short distance between a pair of objects, or large number of moving objects in a spatio-temporal region).

Based on these types of spatio-temporal data, the fundamental types of queries can be seen as transformations combining three basic components: (a) space (*where*), (b) time (*when*), (c) object or event (*what*). These components can be used in three basic types of queries:

- Retrieve the trajectories/events in a region for a time period (*when*&*where*→*what*)
- Retrieve the region occupied by a trajectory/event or set of trajectories/events, at a given time instant or period (*when*&*what*→*where*)
- Retrieve the time periods that a non-empty set of trajectories/events appear in a specific location or area (i.e., *where*&*what*→*when*)

Exploiting these fundamental data types and queries, this chapter presents how to support the generic transformations depicted in Figure 2.

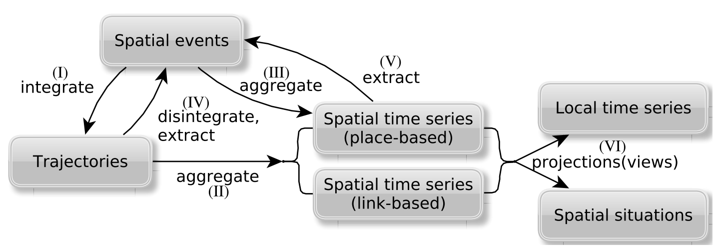

Fig. 1.2 Example of a multiple levels of abstraction of a trajectory.