Cambridge University Press
978-1-107-02171-6 - Mobility Data: Modeling, Management, and Understanding
Edited by Chiara Renso, Stefano Spaccapietra and Esteban Zimányi
Excerpt
More information

PART I

MOBILITY DATA MODELING AND REPRESENTATION

1

Trajectories and Their Representations

Stefano Spaccapietra, Christine Parent, and Laura Spinsanti

1.1 Introduction

For a long time, applications have been using data about the positions of the moving objects they are interested in. For example, city planning applications, in particular in the transportation and traffic management domains, have been observing and monitoring traffic flows to capture their characteristics, namely their importance and localization, with the aim to build better models for traffic regulation and to identify solutions for future development of the existing road network. Sociologists have also been examining the movement of cars equipped with GPS, focusing on individual cars rather than traffic flows, to understand the habits of their drivers. In the logistics domain, applications have been monitoring the localization of the parcels during their transportation from their source locations to their destinations. These applications use the data both to be able to locate a parcel at any time and to optimize the performance of the transportation and distribution strategy. Similar concerns rule the management of data tracking airline passengers and their luggage. Ecologists have been observing animals and, whenever possible, tracking them via transmitters and satellites, mainly to understand animals' individual and group behaviors. Nowadays many enterprises are looking to extract information about their potential consumers out of the tracks left by their smartphones, electronic tablets, or access to social networks such as Flickr and Foursquare that record the geographic position of their users.

Traditionally, data about movement have been captured using static facilities, for example, sensors producing traffic flow measures or detecting an animal's presence. Data acquisition facilities changed drastically with the availability of embedded positioning devices (e.g., GPS). Traffic data, for example, can now be captured as the sequences of positioning signals transmitted by the cars' GPS all along their itineraries.

These sequences may be very long, far longer than the ideal unit of processing of the application. Often the processing unit is some segment of the movement of the object instead of the whole movement itself. For instance, for animals' study the segments may correspond to the daylight time; for employees of an enterprise the segments are defined by working hours, for example, 8 A.M.–6 P.M.; for hikers in a natural park segments may be defined as going from one camp site to another camp site. These segments of movement are nowadays called *trajectories*. They are the unit of interest in applications' processing of movement data. They are the focus of this chapter.

While movement is inherently continuous, it cannot be captured as such in computers where stored data is by definition discrete. The movement track that stores movement data consists of a discrete sequence of records (transmitted by the acquisition device or input by humans) containing the position in space and time of the moving object. Movement tracks are application independent; their precise format and content depend on the device. Movement tracks are analyzed and transformed to produce application-dependent representations of trajectories. Because applications can require very different representations of trajectories (with differences in their structure as well as differences in their content) we define in this chapter three main kinds of trajectory representations that we identified as particularly significant and useful: continuous, discrete, and segmented.

Yet trajectories are not the only way to represent movement. Other representations have been designed to suit applications that need some global view of movement, resulting from the aggregation of the data about movement of individual moving objects. For example, movement can be represented as a *field of vectors* within a given space perceived as a continuous field. The vectors aggregate data from the individual tracks to represent, for a given instant, some characteristics – usually speed and direction – of the movements at every position in space. Similarly, applications willing to globally analyze the flow of objects moving among a discrete set of points (e.g., popular places within a city) will aggregate individual movement tracks into edges between nodes of a *flow network* as described in Chapter 15 on network systems. Various representations of aggregated movements in a continuous field are presented in Chapter 8. In this chapter we deal with trajectories only.

Furthermore, movement data is inherently uncertain, because of impreciseness of the data sensing and data transmission devices, or because of human inaccuracy and data entry errors if a position is manually acquired. This chapter does not address this issue, but Chapter 5 discusses uncertainty issues and approaches in detail.

Application users rarely reason about locations expressed as geographical coordinates: "I am at the Eiffel Tower" is easier to understand than "I am at 48°51′29″ North, 2°17′40″ East." To enable easier and richer use of movement data, recent research has been investigating ways to reformulate and enrich movement data to make them better correspond to application requirements and scenarios. This is done by adding to the movement data contextual data that describe where the object moved (e.g., the roads it followed, the places where it stopped), when (e.g., during which time period, during which event), how (e.g., using which transportation means), what for (e.g., which activity it performed when it stopped). Enriched movement tracks are nowadays referred to as *semantic trajectories*. Chapters 6 and 7 in this book discuss how to build and use semantic trajectories.

This initial chapter introduces the reader to a global understanding of the trajectory domain. It spans from raw data to data transformation and enrichment, to end up with the analysis tasks needed to fulfill application requirements. The chapter covers both the static representation of the domain (what a trajectory is and how it can be represented) and its behavioral aspects (how to understand and characterize mobility in terms of why things move, what they do while moving, which are meaningful movement sequences, etc.). Given the diversity of application requirements, several representations of trajectories are considered. Basic concepts and terminology are defined, explained, and documented via examples.

1.2 Trajectory: Definition and Application Scenario

Mobility is a recent domain where people use diverse terminologies and concepts, without much consensus on choices and definitions. To limit misunderstanding and confusion, this section defines a set of concepts and vocabulary that together form a consistent framework for discussing trajectories and their analysis as understood in this book.

At the source of our movement data processing concerns there is a moving object, that is, an object that can over time change its position in space (its spatial coordinates). In this book, we don't address deformation issues raised when considering moving objects, such as hurricanes and oil spills, that span over a changing area or volume. We focus instead on moving objects represented as points. Keeping movement data about a moving object consists in keeping the history of its successive positions, that is, creating a record that holds, for this object, all past, present, and sometimes future positions and the associated instants. We will not discuss future positioning at this point, and call this record the movement track of the object. The sequence can be unbounded. The time intervals between successive positions may have the same duration or different durations.

Definition 1.1. The *movement track* of a moving object is the temporally ordered sequence of spatio-temporal position records captured by a positioning device during the whole lifespan of the object. Each record (instant, point, features) contains the instant of the capture, the 2D or 3D point of the object, and possibly other features captured by the device (e.g., the instantaneous speed, acceleration, direction, and rotation). There are no two records with the same instant value. □

Before going into a detailed analysis of what trajectories are and how they can be tailored into useful information for the targeted applications, we informally sketch an example application scenario that uses trajectories to describe the movement of tourists visiting Paris.

1.2.1 Tourists Application Scenario

Tourism represents an important source of revenue for many countries, regions, and cities. Its promotion has become a critical business. The efficiency of promotion activities can be boosted by the acquisition of knowledge about the habits of tourists, their preferences, and the local features that are likely to attract them in large numbers. Part of this knowledge can nowadays be extracted from the analysis of on-site movements of tourists, collected via their smartphones equipped with GPS and connected to social networks.

From a promoter's viewpoint, a tourist destination is a geographical area that offers tourists the opportunity of visiting a variety of places (e.g., museums, parks, monuments, and attractions) while using many services (e.g., restaurants, accommodations, shops, and travel agencies). All these tourist places and services are collectively referred to as points of interest (POIs), chosen from a tourist perspective. A tourist day consists in moving from one POI to another one, and so on, while stopping for some time in each one of the visited POIs for eating, resting, shopping, visiting, sleeping, attending a show, or meeting other people, as shown in Figure 1.1.

The oriented line in Figure 1.1 shows the spatial route of the trajectory made by a tourist during one day while visiting Paris. Very often, applications use only this spatial representation of movement on a background map. It is very intuitive, yet it provides very little temporal information. Time is only implicitly conveyed by the fact that the sequence of points forming the line is a temporally ordered sequence. In other words, going further down the line (from its beginning to its end) corresponds to moving later in time. In Figure 1.2, part of this trajectory is shown with a volume (x, y, t) visualization. The trajectory is represented by the thick line in the upper part of the figure, and its projection on the (x, y) plane shows its spatial route as a line lying on the map. As time never stops and always flows on, no two points can have the same time value, and the thick 3D line always moves further on the time axis. When a moving object stops, its position in the (x, y) plan does not change. In the (x, y, t)

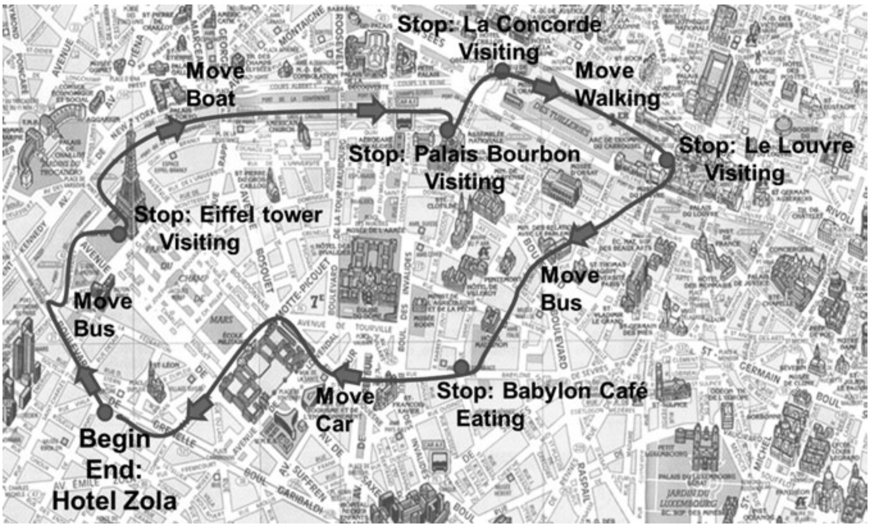

Figure 1.1 A daily trip of a tourist in Paris, visiting several tourist attractions.

visualization an object stopping results in a vertical segment whose length corresponds to the duration of the stop. Figure 1.2 shows three vertical segments that represent stops at Place de la Concorde, Le Louvre museum, and the Babylon café.

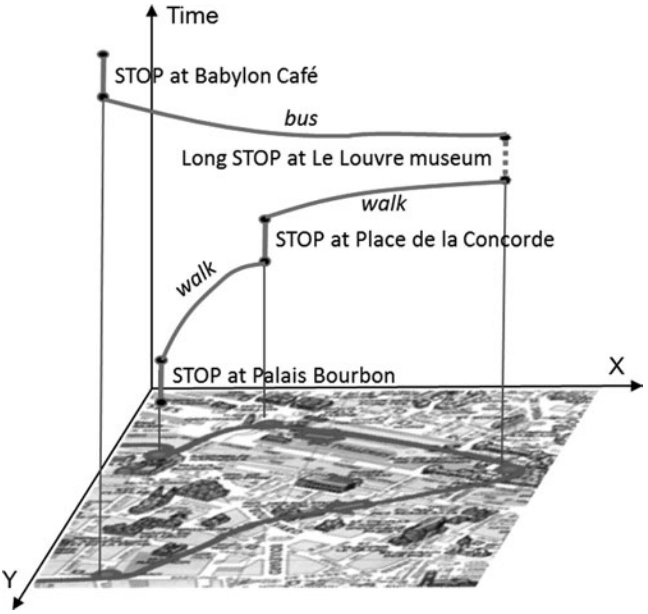

Figure 1.2 A volumetric representation of part of the tourist's daily trip of Figure 1.1.

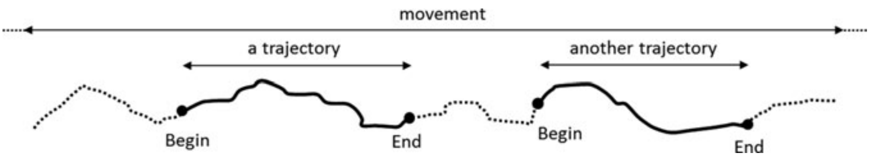

Figure 1.3 Two trajectories extracted from the movement of an object.

Collecting information on daily travels of tourists enables extracting knowledge on their favorite places, in which order places are visited, how much time tourists spend at each attraction, etc. This can be used to tune the facilities to better match tourist expectations and regulate the flow of tourists to avoid large waiting lines. It can also be used to build tourist profiles, propose personalized tours and services, and suggest to tourists on the move their next preferred destination. Similar kinds of moving persons' scenarios are used in many research papers to illustrate various kinds of analysis. We will use it throughout this chapter for illustrating the concepts.

1.2.2 Trajectory Definition

As stated in Section 1.1, while some applications keep and analyze whole movement tracks, many other applications are interested in specific segments of the movement. We call trajectories the segments of the object's movement that are of interest for a given application. Obviously the whole movement is a particular case of trajectory.

Definition 1.2. A *trajectory* is the part of the movement of an object that is delimited by a given time interval $[t_{\text{Begin}}, t_{\text{End}}]$. It is a continuous function from the time interval $[t_{\text{Begin}}, t_{\text{End}}]$ to Space. The spatio-temporal position of the object at t_{Begin} (resp. t_{End}) is called the Begin (resp. End) of the trajectory. □

Figure 1.3 shows (as a dotted line) a section of the movement of an object and, superimposed as continuous lines, two segments identified as relevant trajectories.

The criterion to identify trajectories within movement is application dependent. For instance, in the tourists scenario, to globally analyze the activities performed by a tourist during his/her stay in Paris, the whole track left by the tourist will generate a single trajectory (spatial criterion "inside Paris"). On the other hand, in order to analyze what tourists do in one day in Paris (whatever the length of their stay), or what they do on specific days (e.g., on Sundays, on December 25), each daily track of each tourist in Paris will generate a separate trajectory as in Figure 1.1.

In the real world, time, movements, and trajectories are continuous, but in the digital world, where applications are implemented, we can only store discrete

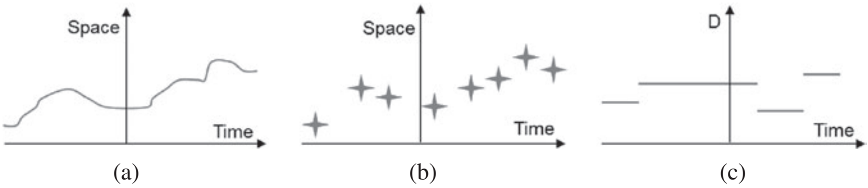

Figure 1.4 The three kinds of representations of movement: continuous, discrete, and stepwise.

implementations, such as the movement track. In order to satisfy applications that need a continuous view of trajectories, the discrete implementation may be enriched with interpolation functions that allow dynamically reconstructing a continuous representation of the discretized trajectory.

Definition 1.3. A *continuous representation of a trajectory* (or *continuous trajectory* in short) is a trajectory representation that describes in a continuous way the movement of the object for the time interval $[t_{Begin}, t_{End}]$ of the trajectory. It usually consists of a finite sequence of spatio-temporal positions, and the interpolation functions that enable the computation of the spatio-temporal position of the moving object for any instant in $[t_{Begin}, t_{End}]$. □

Whenever the movement track is too sparse for inferring the original continuous movement of the object, or the applications do not need the continuous movement, the finite sequence of spatio-temporal positions is used as a discrete representation of a trajectory. Currently this is the case, for example, of the applications that use the movement tracks generated by social networks (see Chapter 16).

Definition 1.4. A *discrete representation of a trajectory* (or *discrete trajectory* in short) is a trajectory representation that is made up of the finite list of spatio-temporal positions for the time interval $[t_{Begin}, t_{End}]$ of the trajectory, but not providing the continuity of the movement of the object. □

Figure 1.4a visualizes (as a line) a continuous representation of a trajectory. Figure 1.4b visualizes (as a set of points) a discrete representation. Figure 1.4c visualizes a stepwise (segmented) representation (see Section 1.3).

To complete the basic picture we briefly introduce two trajectory concepts, holes and semantics gaps, which address the understanding of missing points at the conceptual level. These concepts contribute to a more complete vision of trajectories. The reader has to be aware that they only play an important role in a limited number of application cases, which explains why researchers rarely take these concepts into account.

The term *missing point* denotes the existence, within a movement track, of an abnormal (longer than the sampling rate) temporal gap between two consecutive recorded positions: the information on the movement of the object is missing. If this is accidental (e.g., because of a device malfunction) we say there is a *hole* in the track. The typical case where this still happens is when a GPS is taken through a tunnel. The connection is cut as long as the GPS doesn't get out of the tunnel. Short-duration holes may sometimes be "filled," using, for example, linear interpolation algorithms that compute the missing positions. In this case the hole disappears.

If missing points are not due to some data acquisition accident (whatever the cause), it follows that their absence is due to a decision by the application designer to interrupt data acquisition during some specific periods. For example, a company running daily tourist tours in Paris may decide to track tourists' positions during its hours of operation (say from 8 A.M. to 6 P.M.) but not during lunchtime (say from 12:30 P.M. to 2 P.M.) when tourists on a tour are free to do whatever they want. Consequently, tourists' daily trajectory tracks will be filled with positions from 8 A.M. to 12:30 P.M. and from 2 P.M. to 6 P.M., and no positions during the lunchtime break. This lunchtime break is not an accidental hole in the trajectory; we call it a *semantic gap* (its semantic in this case is that of the lunch period).

A trajectory with semantic gaps is defined for a set of disjoint time intervals instead of a unique time interval. For the sake of simplicity, in the rest of the chapter we will deal only with trajectories defined on a single interval (i.e., without semantic gaps).

1.3 From Raw Trajectories to Semantic Trajectories

The two representations of trajectories defined above come directly from the movement track. It is why they are often called *raw trajectories*. They are well fitted if, for example, the aim of the application reduces to locating some moving objects (e.g., where was Mr. Smith on the evening of June 12, 2012?) or computing statistics on the spatio-temporal characteristics of the trajectory (e.g., which percentage of daily tourist trajectories show a global speed over 7 km/h?). On the other hand, many applications need more informative results, such as those that can be computed by combining raw data with the contextual data (e.g., geo-objects and events that show a spatial or temporal relationship with the trajectory data), and with the thematic data available for the moving object itself (e.g., age, gender). These applications can reach this goal by following one of two approaches:

1. The application dynamically accesses the contextual data during its computations.
2. The application first preprocesses the trajectory representations, enriching them with contextual data and appropriate restructurings, and after that it computes its results by using the enriched trajectories.