2010 17th International Symposium on Temporal Representation and Reasoning

Dealing with uncertainty in trajectory databases

Bart Kuijpers
Hasselt University
bart.kuijpers@uhasselt.be

Research on spatial databases, which started in the 1980s and found its motivation in Geographic Information Systems, was extended in the 1990s to deal with spatio-temporal data [5, 11]. One particular line of research in the field of spatio-temporal data, started by Wolfson [12], focusses on *moving object data*. In moving object databases (MODs), several data models and query languages have been proposed to deal with moving objects whose position is recorded at, not always regular, moments in time. For an overview of models and techniques in the field of MODs, we refer to the recent textbook by Güting and Schneider [2]. The practical motivation for studying this type of data comes from the rapidly growing spatio-temporal, geo-referenced datasets that are collected via mobile phone and traffic control networks, sensors and other location-aware technologies such as GPS. Location-aware devices, carried by people, vehicles or animals, produce *trajectories*. A trajectory can be viewed as a continuous curve in the plane $\mathbf{R}^2$, that is parameterized by time. But since trajectories are collected by location-aware devices, it is more common to consider *trajectory samples* which are finite sequences of space-time points $\langle(x_0, y_0, t_0), (x_1, y_1, t_1), ..., (x_n, y_n, t_n)\rangle$, with $t_0 < t_1 < \cdots < t_n$ for the time values. We can regard a trajectory database as a finite collection of trajectory samples labelled by trajectory identifiers.

A popular way to reconstruct trajectories from trajectory samples is linear interpolation between consecutive sample points [2]. However, linear interpolation relies on the assumption that between sample points, a moving object moves at constant minimal speed, which is realistic when sample points are frequent and occur at regular time intervals. When this is not the case, cylinders around the linear interpolation were initially used to model the uncertainty of the moving object's location between sample points [2, 12]. More recently, an old idea coming from the work of Hägerstrand in the early 1970s in time-geography [3], has found its way into MOD research [4, 9]. Here, the assumption is that moving objects have some physically determined or law imposed speed bounds. Based on this, an uncertainty model has been proposed which constructs *space-time prisms* between two consecutive space-time points in a trajectory sample (some authors refer to prisms as *beads* [4, 9]). A space-time prism is the intersection of an upward and a downward pointing cone in the space-time space and all possible trajectories of the moving object between the two consecutive space-time points, given the speed bound, are located within the prism. More technically, the space-time prism between the consecutive sample points (x_i, y_i, t_i) and $(x_{i+1}, y_{i+1}, t_{i+1})$, given a speed bound $v_{\max}$, is described by the following polynomial constraints on the coordinates of space-time points (t, x, y):

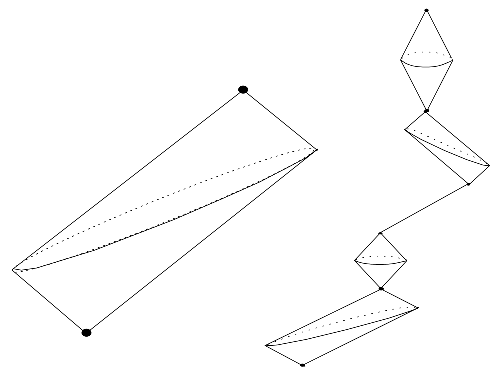

Figure 1. An example of a space-time prism (left) and a lifeline necklace (right).

$$\begin{cases} t_i \leq t \leq t_{i+1} \\ (x - x_i)^2 + (y - y_i)^2 \leq (t - t_i)^2 v_{\max}^2 \\ (x - x_{i+1})^2 + (y - y_{i+1})^2 \leq (t_{i+1} - t)^2 v_{\max}^2. \end{cases}$$

A chain of prisms connecting consecutive points in a trajectory sample is called a *lifeline necklace* and all possible trajectories of the moving object are loacted in this necklace [4]. Figure 1 illustrates the concepts of space-time prism and lifeline necklace.

Starting from the concept of space-time prism to model trajectory uncertainty, this talk will address a number of topics relevant for trajectory data managment: complete query languages that focus on speed and prisms; one particular

1530-1311/10 $26.00 © 2010 IEEE
DOI 10.1109/TIME.2010.27

query, the alibi query; space-time prisms for movement restricted to road networks; measurement uncertainty of sample points; and applications to map matching of trajectory data and trajectory data mining.

Query languages for trajectory databases Since space-time prisms can be described by polynomial constraints, the constraint database formalism [6, 10] offers a good starting point to design query languages for trajectory databases. Here, we can use extensions of first-order logic over the real numbers as coordinate-based query languages. When we are interested in querying trajectory data, we mainly think of properties such as speed and in our setting also of queries that speak about uncertainty in term of prisms. By identifying the transformations of the ambient space that keep speed and prisms invariant, we can identify fragments of first-order logic over the reals that allow exactly the expression of those queries that talk about speed and prisms. We discuss a space-time point-based logic that captures exactly these first-order queries. It uses very simple predicates such as $\mathsf{Before}(p, q)$ and $\mathsf{inPrism}(r, p, q, v)$ that express that the space-time point p is temporally seen before q and that r is in the prism defined by p and q with speed constraint v. We also present a computationally complete extension of this first-order logic.

One specific query that is of particular interest to time-geographers is the *alibi query* [4]. It asks whether two moving objects could have met, given their trajectory samples and speed constraints. Although this query can be answered, in theory at least, by existing implementations of quantifier-elimination algorithms for the first-order theory of the reals, it turns out to be computationally too heavy and we propose an analytic solution to this problem.

Trajectories on road networks Thus far, we have only considered trajectories and space-time prisms for free movement in an isotropic two-dimensional space. But in practice, movement is often constrained, for instance, to a road network [7]. We view a *road network* as a graph embedding in $\mathbf{R}^2$, where road segments are straight line segments between vertices with an associated speed limit. In this context, we can also model, based on shortest path length, space-time prisms for road networks. We give algorithms to compute and visualize them. We also give an algorithm to solve the alibi query on road networks.

Measurement uncertainty of sample points Space-time prisms model the uncertainty of the location of a moving object between trajectory sample points. But there are other sources of uncertainty. Until now, sample points were considered to be exact. But in practice they are subject to measurement errors (GPS, for instance, has in the best case an accuracy of a few meters). We extend the space-time prism model on road networks to deal with this type of uncertainty.

Applications of space-time prisms When tracking the trajectory of a vehicle on a road network with a GPS device, the measured locations do not always fit the road network. They may be inside buildings or parks, for instance. *Map matching* techniques are used to obtain a corrected trajectory that is on the road network. Whereas many map matching algorithms use heading information that is derived from the GPS readings, space-time prisms can be used successfully for map matching without using this derived information. Another important application of prisms can be found in the very recent area of *trajectory data mining*, where prism-based distance measures can be used to cluster trajectories. The study of the relationship between uncertainty and privacy, which is a very relevant issue when mining trajectory data, has just started [1].

Acknowledgment I enjoyed discussing many topics presented in this talk with my former student Walied Othman. His thesis contains a wealth of the details [8].

References

[1] F. Giannotti and D. Pedreschi. *Mobility, Privacy, and Geography: a Knowledge Discovery Perspective*. Springer, 2008.

[2] R. Güting and M. Schneider. *Moving Object Databases*. Morgan Kaufmann, 2005.

[3] T. Hägerstrand. What about people in regional science? *Papers of the Regional Science Association*, 24:7–21, 1970.

[4] K. Hornsby and M. Egenhofer. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1–2):177–194, 2002.

[5] M. Koubarakis and T. Sellis et al. *Spatio-Temporal Databases: The CHOROCHRONOS Approach*. Springer, 2003.

[6] G. Kuper, L. Libkin, and J. Paredaens, editors. *Constraint Databases*. Springer, 2000.

[7] H. Miller. Modeling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems*, 5:287–301, 1991.

[8] W. Othman. *Uncertainty management in trajectory databases*. Phd thesis, Hasselt University (Belgium), 2009.

[9] D. Pfoser and C. Jensen. Capturing the uncertainty of moving-object representations. In *Advances in Spatial Databases (SSD'99)*, volume 1651 of *Lecture Notes in Computer Science*, pages 111–132. Springer, 1999.

[10] P. Revesz. *Introduction to Constraint Databases*. Springer, 2002.

[11] P. Rigaux, M. Scholl, and A. Voisard. *Introduction to Spatial Databases: Applications to GIS*. Morgan Kaufmann, 2000.

[12] O. Wolfson. Moving objects information management: The database challenge. In *Proceedings of the 5th Intl. Workshop NGITS*, pages 75–89. Springer, 2002.