

Semantic Trajectories and Beyond

(Advanced Seminar)

Maria Luisa Damiani
University of Milan
Milan, Italy
Email: maria.damiani@unimi.it

Ralf Hartmut Güting
FernUniversität in Hagen
Hagen, Germany
Email: rhg@fernuni-hagen.de

I. OVERVIEW

This seminar overviews recent research on novel data models for the advanced representation of mobility data encompassing semantic and contextual information.

Mobility data generally takes the form of *geometric* trajectories, reporting the history of the objects' locations. Abstractly, geometric trajectories are time-varying functions ranging over space, or, put differently, geometries whose position and possibly shape, changes in time. This view is at the heart of the moving object data model developed in early 2000 [1], and implemented in major moving object databases such as Secondo and Hermes [2].

In modern applications, however, there is increasing evidence that describing the movement exclusively in geometric terms is no longer sufficient. For example, an important feature of the movement is the context in which such a movement takes place, such as the weather conditions during the travel, the places of interests being visited, the people in proximity and so forth. Moreover the movement can be described at different levels of abstraction, possibly resulting from the analysis of fine-grained mobility data, such as the activities performed by mobile individuals. Capturing this multiplicity of views is important especially in the light of the increasing availability of potentially *Big Data* collected from LBS and mobile sensing applications which paves the way to the development of challenging applications. All that motivates the growing concern for solutions advancing the classical moving object data model.

Semantic trajectories is a relatively recent paradigm aiming to provide applications with knowledge about the movement of entities. Basically a semantic trajectory is a conceptual artifact denoting a geometric trajectory augmented with annotations regarding the whole trajectory or parts of it. Probably because of its simplicity, the concept of semantic trajectory has attracted the interest of numerous researchers working in different areas, such as data analytics, conceptual modeling, Semantic Web, privacy. A comprehensive survey is reported in [3].

A question that so far has been only marginally addressed regards the data management dimension of semantic trajectories, in essence: how can we represent and query large amounts of semantic trajectories? The issue poses several challenges: it requires the definition of a rigorous data model, the specification of an effective query language enabling the retrieval of semantic trajectories or parts of them, and the development of efficient data access mechanisms. To deal with the question, different directions of research are emerging. For

example, a recent approach relies on the use of Semantic Web technologies to represent and query semantic trajectories. An orthogonal and more ambitious research direction focuses on the specification of a novel database model, formally defined, equipped with a dedicated query language and appropriate indexing mechanisms, and embedded into an existing moving object database. This data model is referred to as *symbolic trajectories* [4].

This seminar is structured in three main components: the first component overviews on-going research on semantic trajectories, focusing in particular on the three dimensions of semantics, data analytics and privacy. The second part overviews on-going research on symbolic trajectories. Scalable analysis on symbolic trajectories using MapReduce/Hadoop is also addressed. The third and final part discusses some research directions.

II. SEMANTIC TRAJECTORIES

Early work on semantic trajectories was conducted within the UE research project GeoPKDD (2006-2008) [5], and next continued in the MODAP project (2009-2013). The basic idea was triggered by the experimental analysis of a set of trajectories about a group of birds [6]. By using the standard functionalities of a GIS, we found that the sequences of points, just pairs of timestamped coordinates, associated with birds identifiers were actually representing the migration routes from Central Europe to Africa and vice versa. Such discovery, that was somewhat unexpected, inspired the proposal of a model for the high level representation of movement later on called *semantic trajectories* [7]. Since then, research has been developing along diverse streams. The driving research questions can be summarized as follows:

- Semantics: which abstractions allow capturing and structuring the application-dependent meaning of trajectories?
- Mobility knowledge discovery: which techniques allow the extraction of semantic-enriched trajectories from raw, geometric trajectories?
- Privacy: in case of human mobility data, how can privacy be safeguarded without sacrificing the semantic trajectory concept?

In the seminar, for each of these questions we overview some major lines of research. A brief synthesis is reported in the next.

- *Semantics*. The question poses two related issues: to devise suitable abstractions, capturing general, high level features of the movement; and to choose a notation to represent such abstractions.

As concerns the former aspect, the first conceptualization of trajectories is the stop-and-move mobility pattern [6]. The idea is basically to describe the movement as a sequence of transitions from one stop to another stop where the stop indicates the temporary suspension of the movement at the chosen level of abstraction. The level of abstraction depends on the application requirements and its temporal scale. For example, stops can describe the points of interest visited by tourists during their travels or the cities where users have lived in the last 20 years. The problem with this pattern-focused approach is that, in reality, there can exist multiple patterns each suitable for a different class of applications. That motivates the on-going research on a more general model not specifically focused on any pattern, but providing the constructs to express, ideally, any kind of mobility pattern [3].

The second key issue is how to represent the semantic-enriched trajectories. One direction is to borrow from the database world the graphical notations used for the specification of conceptual models such as MADS or UML. This choice presents a number of shortcomings, not least the fact that the true purpose of the notation is not to create the conceptual schema of a moving object database, but rather to capture a conceptualization. A more recent approach is to rely on the standards developed by the Semantic Web [8]. An orthogonal direction, emphasizing the operational dimension, is represented by the symbolic trajectories data model that is introduced below.

- *Mobility knowledge discovery*. This line of research focuses on the study of techniques and methodologies for the extraction of rich content from geometric trajectories. Indeed this is a broad area of research initiated by different communities over the last years and extensively applied for the study of human mobility in urban settings [9], [10]. The analytical techniques explicitly inspired by the notion of semantic trajectories mainly focus on the extraction of mobility patterns for the automatic generation of semantically annotated trajectories. For example, early work by Alvares et al. [7] focuses on the extraction of stops and moves from trajectories of GPS points. A rich literature exists, either explicitly related to the notion of stop-and-move, or regarding the mining of both individual and collective behavior from geometric trajectories. A complementary direction of work focuses on the definition of methodologies driving the knowledge discovery process from the raw data to semantic trajectories such as in [11]. An additional research stream regards the development of platforms supporting analytical tasks over trajectories such as M-Atlas [12]. More recent work focuses instead on the mining of semantic trajectories [13].
- *Privacy*. Semantic trajectories and privacy inevitably clash when trajectories regard people because behavior

information can be explicitly represented and thus unfolded to potentially untrustworthy third parties. Recent research attempts to mitigate this conflict by forestalling the extraction of patterns conveying sensitive information such as visits to hospitals and religious places, and presence at specific locations revealing data about, for example, sex life. These patterns are called *sensitive stops*. Different classes of techniques for the protection of sensitive stops have been investigated [14], [3]. These techniques can be applied either before stops are recognized, therefore during the process of semantic trajectory construction, or after the semantic trajectories have been built. This line of research can be boiled down in the concept of *behavior privacy*.

III. SYMBOLIC TRAJECTORIES

The notion of semantic trajectories leads to a radical shift from the traditional, geometric view of trajectories to a new, application oriented vision of the movement. At the current stage, however, semantic trajectories is more a conceptual than an operational framework. Symbolic trajectories aim at filling this gap by providing a database model for the specification of text-annotated trajectories. The model is strongly focused on generality, efficiency and effectiveness of data representation and access as briefly described in the next.

A. The data model

Symbolic Trajectories is a simple generic data model able to capture different types of semantics [15]. A symbolic trajectory is, in its basic form, just a time dependent label, that is, a function from time into label values. Labels are just short character strings. Such a function can be represented as a sequence of pairs

$$\langle (i_1, l_1), \dots, (i_n, l_n) \rangle$$

where i_j is a time interval and l_j a label. Time intervals are disjoint (possibly adjacent) and the pairs in the sequence are ordered by time. For example, a simple symbolic trajectory would be:

$$\langle ([8 : 30 - 8 : 45], walk), ([8 : 45 - 9 : 13], train), \dots \rangle$$

The symbolic information can be computed from the movement itself or be obtained by relating the geometric trajectory to its environment, e.g. static geometries or other moving objects. Hence it is an abstraction that captures certain aspects of a precise geometric trajectory. Here are some examples of possible interesting aspects:

- Names of roads traversed by a vehicle, obtained by map matching.
- Cell identifiers of a cellular network
- Cardinal directions such as north, northwest, ...
- Speed profiles: slow, moderate, fast
- For animals: at home range, migrating, stopover, ...
- Indoor navigation: triples of (building, floor, room)
- Names of countries traversed on a long distance trip

- Activities of a tourist
- For animals: grazing, resting, at feeding station, ...

The core technical contribution is a novel language for pattern matching and rewriting on symbolic trajectories (a video of the system is available at <http://molle.fernuni-hagen.de/DfnA/SymbolicTrajectories.mp4>). The pattern language enables the extraction of subsequences from symbolic trajectories. Patterns are defined as regular expressions that can be matched by single units or sequences of units. For example, the query: *Which are the trajectories in which the individuals take more than 1 hour to move from home to work?* can be solved specifying the following pattern:

```
* (_ home) Z * (_ work) *
// getDuration(Z.time) > 3600
```

where:

- Z is a variable denoting a sequence of units, the symbol $*$ denotes a sequence of zero or more units,
- $* (_ home) Z * (_ work) *$ is the pattern,
- $getDuration(Z.time) > 3600$ is the condition that must be met by the matching sequences, in this case the duration in seconds of the transfer from home to work.

Symbolic trajectories are provided as abstract data types and integrated into the ADT model defined in [1]. This means that symbolic and geometric trajectories can be treated homogeneously within a well-founded framework. Moreover the pattern-based query language can be efficiently used with large data sets of symbolic trajectories. The scalability analysis is performed using a benchmark based on a data set of symbolic trajectories generated from synthetic geometric trajectories using MapReduce/Hadoop.

IV. CONCLUSIONS AND RESEARCH DIRECTIONS

Following the growing availability of semantic-enriched trajectories, the definition of unifying data modeling frameworks enabling the representation and access of different semantics across various application domains is definitely crucial. As we have seen, there are two emerging modeling paradigms, i.e. semantic trajectories and symbolic trajectories. In the former case, the emphasis is more on the ontological aspects, while in the latter on the effective and efficient data access. The two visions are however complementary. Another research area that can benefit from the definition of advanced trajectory data models is mobility data analytics. Analytical methods defined on generic, well-defined semantic-enriched trajectories (instead of narrow, ad hoc data models), can be of more general applicability and be deployed in different domains. Achieving in perspective some form of consensus on advanced trajectory data models can thus have a great impact.

On the other hand, there are numerous research issues that are still open. In particular the notion of symbolic trajectories opens up several challenges, such as the integration of the symbolic dimension with the geometric dimension to achieve multi-dimensional trajectories [16], the deployment of the concept in advanced applications, and the integration with

large scale data analytical techniques (e.g. trajectory data warehouses, data mining).

REFERENCES

- [1] R.H. Güting, M. Böhlen, M. Erwig, C. Jensen, N. Lorentzos, M. Schneider, and M. Vazirgiannis, "A foundation for representing and querying moving objects," *ACM Trans. Database Systems*, vol. 25, no. 1, pp. 1–42, 2000.
- [2] R.H. Güting, T. Behr, C. Düntgen, "Trajectory Databases," in *Mobility Data: Modelling Management and Understanding*. Cambridge University Press, 2013.
- [3] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M.L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, and Z. Yan, "Semantic trajectories modeling and analysis," *ACM Computing Surveys*, vol. 45, no. 4, pp. 1-42, Aug. 2013.
- [4] R.H. Güting, F. Valdes, M.L. Damiani, "Symbolic Trajectories," FernUniversität in Hagen, Informatik-Report 369 - 12/2013, Tech. Rep., 2013.
- [5] F. Giannotti and D. Pedreschi, Eds., *Mobility, Data Mining and Privacy*. Springer, 2008.
- [6] S. Spaccapietra, C. Parent, M.L. Damiani, J. de Macedo, F. Porto, and C. Vangenot, "A conceptual view on trajectories," *Data & Knowledge Engineering*, vol. 65, no. 1, pp. 126–146, 2008.
- [7] L. Alvares, V. Bogorny, B. Kuijpers, J. de Macedo, B. Moelans, and A. Vaisman, "A model for enriching trajectories with semantic geographical information," in *Proc. ACM GIS*, 2007.
- [8] R. Fileto, M. Krüger, N. Pelekis, Y. Theodoridis, and C. Renso, "Baquara: A holistic ontological framework for movement analysis using linked data," in *Proc. ER*, 2013.
- [9] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proc. ACM SIGKDD*, 2007.
- [10] Y. Zheng and X. E. Zhou, Eds., *Computing with Spatial Trajectories*. Springer, 2011.
- [11] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "Semitri: a framework for semantic annotation of heterogeneous trajectories," in *Proc. EDBT*, 2011.
- [12] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti, "Unveiling the complexity of human mobility by querying and mining massive trajectory data," *VLDB J.*, vol. 20, no. 5, pp. 695–719, 2011.
- [13] C. Zhang, J. Han, L. Shou, J. Lu, and T. L. Porta, "Splitter: Mining Fine Grained Sequential Patterns in Semantic Trajectories," *PVLDB*, vol. (7)9, pp. 769–780, 2014.
- [14] M.L. Damiani, C. Silvestri, and E. Bertino, "Fine-grained cloaking of sensitive positions in location-sharing applications," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 64–72, Oct. 2011.
- [15] F. Valdés, M.L. Damiani, and R.H. Güting, "Symbolic Trajectories in Secondo: Pattern Matching and Rewriting," in *Proc. DASFAA*, 2013.
- [16] M.L. Damiani, H. Issa, R.H. Güting, and F. Valdés, "Hybrid queries over symbolic and spatial trajectories: a usage scenario," in *Proc. MDM*, 2014.