Title: Semantic Trajectory Data Mining: a user driven approach

Trajectories left behind cars, humans, birds or other moving objects are a new kind of data which can be very useful in decision making process in several application domains. These data, however, are normally available as sample points, and therefore have very little or no semantics. Knowledge discovery from trajectory samples is very difficult from the user's point of view [1] [2]. Although several works have been developed so far for both querying and mining trajectories of moving objects, these works have focused on the geometric properties of trajectories and, therefore, deal with trajectory sample points. These approaches, however, suffer from some general problems that are essentially important for trajectory knowledge discovery: (i) focus on the mining step itself, basically considering the geometric properties of trajectory sample points, without taking into account the semantics of the data; (ii) do not cover the whole trajectory knowledge discovery process, which requires complex data preprocessing and post-processing tasks in order to generate meaningful patterns understandable by humans; (iii) do not consider the geography behind trajectories, which is the essential information to understand patterns in most application domains; and (iv) do not provide preprocessing/transformation mechanisms to manipulate the data at different granularities (e.g. morning/afternoon, rush hours, weekday/weekend), which may be of fundamental importance in the knowledge discovery process. Because such approaches do not consider data semantics, there is an emerging need for new data models, manipulation techniques, and tools to extract human understandable patterns from trajectories.

In this talk we address the problem of querying and mining trajectories from a semantic point of view. We introduce a new way for knowledge discovery from trajectories. We consider trajectories from a higher abstraction level, and have focus on the semantic properties of trajectories, therefore providing mechanisms to extract more meaningful and human understandable patterns. More specifically, we present through a data mining query language for trajectories a new methodology for extracting multiple-level patterns from moving object data. We introduce powerful data preprocessing operations which allow the user to transform trajectory sample points into trajectories at higher levels of abstraction, in order to extract semantic and comprehensive patterns. To accomplish this goal we adopt the model of stops and moves [3], where the user can specify the important parts of trajectories that are especially relevant for the application.

Stops are important places of the trajectory from the application point of view, where the moving object has stayed for a minimal amount of time. Moves are sub-trajectories between two consecutive stops. The aggregation of trajectory sample points into stops and moves allows more sophisticated analysis over trajectories, and without sacrificing the user with this task. This aggregation is performed a priori and only once, and therefore can be reused as many times as necessary in the discovery process. As a consequence, the output of the data mining task will be semantic patterns, which will facilitate their interpretation by the user in the post-processing phase.

We have recently developed two different methods for adding semantics to trajectories, i.e., for computing stops and moves: SMoT [1] and CB-SMoT [4]. SMoT is based on the intersection of trajectories with geographic places that are important for the application places of interest). For each relevant geographic object type a minimal amount of time is defined, such that a trajectory should continuously intersect this feature in order to be considered a stop. CB-SMoT is a spatio-temporal clustering method where the important places (stops) are the low speed parts of the trajectory that satisfy a minimal time duration threshold. Both methods cover a relevant set of applications. SMoT is interesting for applications where the speed may not be important, like tourism and urban planning. For traffic management, however, CB-SMoT, which is based on speed, would be more appropriate.

Stops and moves are extracted at the lowest granularity level (feature instance) by both methods. The stop duration is recorded when the moving object enters and leaves the stop. If a trajectory stops at the Eiffel tower, for instance, this will be the name of the stop. Starting from this low granularity level, the mining task can then transform the stop into a higher granularity, like touristic place.

The time dimension is generalized from a time stamp to different granularities, according to the user's interest. It is aggregated into a semantic discretized label. For example, an interval [07:00-09:00] can be labeled as "rushHour".

The contributions of this work can be summarized as follows:

(a) We extend Spatial SQL with operations for semantic trajectory knowledge discovery. The aim is to facilitate the process to the user, which may use spatial SQL without having to learn a new language;
(b) We provide a data preprocessing strategy to add semantic geographic information to trajectories, before applying data mining techniques. This strategy significantly simplifies and optimizes the data mining step. More specifically, we provide two different methods to automatically integrate trajectory samples and geographic information, where the user can define the important parts of trajectories. This integration is of fundamental importance for the discovery of patterns in real applications of different domains;
(c) We provide powerful data discretization functions for both space and time dimensions. This allows the user to extract patterns at several granularity levels;
(d) The output of the proposed language are semantic trajectory patterns, which are stored in the database for further querying. By generating semantic patterns we significantly reduce the complexity of pattern interpretation in post-processing, which is still a human-dependent task in the discovery process;
(e) The language is simple and can be implemented in several spatial or spatio-temporal database management systems, or in a toolkit with friendly and graphical user interfaces. We have implemented a first prototype of the language in a tool, in order to not restrict the user to one specific DBMS.

By adopting the model of stops and moves of trajectories we provide to the user aggregated information about space and time. Therefore, we facilitate not only trajectory mining, but also trajectory querying. Any information about trajectories can be directly obtained with queries over stops and moves. It reduces the query complexity in both formulation and computational time. The limitation of the approach is that it provides support for only three data mining tasks, but we believe that these tasks are the most important for extracting semantic patterns from trajectories.

The next steps include the evaluation of this work from the end user point of view. Indeed, we will extend the proposed approach to support semantic trajectory classification and clustering, as well as investigate the use of ontologies to infer new knowledge in the trajectory knowledge discovery process. We will also evaluate the possibility to implement the proposed approach as an extension of a spatio-temporal DBMS.

More details about this work may be found in [2].

[1] Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A.: A model for enriching trajectories with semantic geographical information. In: ACM-GIS, New York, NY, USA, ACM Press (2007) 162–169.

[2] Bogorny, V.; Kuijpers, B.; Alvares, L.O.: *A Spatio-temporal Data Mining Query Language for Moving Object Trajectories.* In: UFRGS, Brazil, Technical Report T-357, pages 22, 2008.

[3] Spaccapietra, S., Parent, C., Damiani, M.L., de Macedo, J.A., Porto, F., Vangenot, C.: A conceptual view on trajectories. Data and Knowledge Engineering 65(1) (2008) 126–146.

[4] Palma, A.T., Bogorny, V., Kuijpers, B., Alvares, L.O.: A clustering-based approach for discovering interesting places in trajectories. In: ACMSAC, New York, NY, USA, ACM Press (2008) 863–868