An Ontology-based Approach to Represent Trajectory Characteristics

Tales Paiva Nogueira*, Reinaldo Bezerra Braga†, Hervé Martin*
*Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France
†Federal Institute of Education, Science and Technology (IFCE)
Aracati, Brazil
paivan@imag.fr, reinaldo.braga@ifce.edu.br, herve.martin@imag.fr

Abstract—The behavior of moving objects has been a relevant source of information to intelligent mobile systems. However, most of existing works on trajectory representation deal only with basic characteristics of trajectories, such as space and time, while these attributes may be not enough to provide the required information to intelligent systems. We observe that the analysis of other characteristics (e.g. speed and acceleration) of mobile objects enriches the trajectory description as well as open opportunities to novel applications. However, the dynamic nature of these characteristics brings several challenges related to the preprocessing and analysis of raw data. In this paper, we show how these additional characteristics may be integrated in trajectory modeling. We address the problem of representing trajectories with qualitative descriptions of movement modeled as an ontology. We validate our approach with real data from a sport tracking application.

I. INTRODUCTION

The increasing adoption of semantic web technologies has changed the way that information is made available on the web. As we can see, the web of documents is going toward the web of data, where machines are able to understand and reason about the connections among different datasets and therefore enable the development of smart applications.

In parallel, the growing of geolocation technologies embedded in smart phones, smart watches, tablets, and even glasses has facilitated the development of intelligent mobile systems for the acquisition of data. The convergence of these technologies with the Web of Data allows the easy acquisition of information about user trajectories through the use of mobile devices. The management, modeling, and analysis of such data provide many challenges related to its integration with existing systems. Therefore, the multidimensional and multifaceted aspects of geolocated data should be taken into account due to their rich semantics. In consequence, a vast amount of works can be found in the literature proposing new approaches to bridge the gap between trajectory representation and Semantic Web technologies, mainly regarding the representation with ontologies [1]–[8].

Once all these aforementioned aspects are considered, we are able to identify mobility patterns of mobile objects. The identification of mobility patterns has been an constant topic of interest in the GIScience area in several domains like tourism, road traffic, crisis management, marketing, among others [9]. Several works have dealt with trajectory analysis proposing new ways of comparing, segmenting and clustering moving object’s paths. Nevertheless most of them only handle the geometric aspect of trajectories and just a few deal with dynamic parameters like speed and acceleration explicitly [10], [11].

GPS-recorded locations are influenced by problems of noise and error that makes difficult to derive some trajectory characteristics with a great degree of accuracy. If the speed of a moving object is calculated taking into consideration only longitude, latitude, and time data, it is likely that the output signal show a high variability due to the intrinsic error of commercial GPS systems. Therefore, there is a need to simplify this calculated data. One possible approach is to have a qualitative view on this data. This simplification becomes a useful and optimized information into existing data models. Accordingly, the development of new methods, models and algorithms to analyze movement characteristics is still a great challenge.

In this work, we analyze the dynamic characteristics of trajectories (e.g. speed, acceleration) in order to show how these characteristics may be integrated in trajectory modeling. Besides that, we show how qualitative representation can be extracted from raw trajectories in order to enable the building of new queries, which answer many application domain needs. Finally, we validate our proposal with real GPS traces collected by a sport tracking mobile application. The results show that our approach achieved a reasonable representation of qualitative aspects of any characteristic from trajectories that can be represented as a time series.

The remaining of this paper is organized as follows: in section II, we present the related works and highlight the differences between them and our approach. In section III we introduce the theoretical background for the main aspects considered in this paper. The main contribution of this paper, the QualiTraj ontology, is presented in section IV. In section V we address our attention to explain how we go from raw data to symbolic representation. In section VI we show the main results obtained by our validation process. Finally, in section VII we conclude explaining the main contributions and future works related to this work.
II. RELATED WORK

In [12], Rehrl et. al. proposed a method for semantic processing of GPS traces where information is extracted from raw data. Based on the assumption that the basic parameters to express motion in space and time are velocity and course, they defined six motion patterns with associated rules. The patterns are the following: stand still characterizes the absence of motion and is assumed when the velocity is less than 1m/s; steady motion represents the periods when there is motion with steady velocity and is distinguished when velocity is greater than 1m/s and acceleration lies between $-0.3m/s^2$ and $0.3m/s^2$; positive acceleration happens when the velocity increases and is greater than 1m/s and acceleration is greater than $0.3m/s^2$; negative acceleration is similar but acceleration should be less than $-0.3m/s^2$; positive course change is identified when there is a course change rate above $0.4^\circ/s$, and negative course change is determined when this change is below $-0.4^\circ/s$. While this categorization has as objective to elevate the level of abstraction of motion data, the authors rely too heavily in thresholds to characterize speed, acceleration, and course changes. In an heterogeneous dataset, this approach does not seems adequate as these thresholds may vary depending on the mean of transportation. In our work, we preferred the usage of statistic measures whenever it was possible to avoid relying on thresholds that depends on the nature of the data being analyzed.

In [5], van Hage et. al. presented the Simple Event Model (SEM) and its application in the maritime domain. In their use case, events are automatically recognized from the Automatic Identification System (AIS) raw data and represented as SEM instances. From that, it was possible to characterize some types of ship behavior like slowing down, speeding up, and anchored. Three types of data were collected in the form of time series: location, speed, and course. Due to the large dimensions of the tracked ships and the fact that they do not accelerate nor change their courses quickly, their movement are very regular and much more easy to compress by a piece-wise linear algorithm. In our paper, instead of AIS data with speed information already included, we have at first just GPS raw data from which we have to calculate the speed profile. Moreover, the nature of motion data is very different: runners instead of ships. Runners may have a much more irregular speed, acceleration, and changes in course direction when compared to ships. Besides, we take a qualitative approach towards the characterization of movement.

III. THEORETICAL BACKGROUND

One of the most important features of spatio-temporal systems is the ability to trace the path that a moving object follows during some time. A trajectory can be defined as the user defined record of the evolution of the position (perceived as a point) of an object that is moving in space during a given time interval in order to achieve a given goal [13]. More formally, a trajectory $\tau$ is an ordered list of positions in space and time instants, such as:

$$\tau = \{(s_1, t_1), (s_2, t_2), ..., (s_n, t_n)\},$$  

(1)

where each position $s_i = (lat, lon)$ is a pair of latitude and longitude coordinates, and each time instant $t_i$ is represented by a timestamp.

A research topic that is constantly studied in the trajectory analysis domain is related to the representation of these spatio-temporal paths. While the representation of trajectories with ontologies has already been subject of many studies, the dynamic characteristics of trajectories are generally mentioned as simple attributes or even not mentioned. Most works about trajectory analysis limit themselves to the geometric representations of trajectories as a static curve [14].

The dynamic characteristics, which we refer in this paper, are treated with varying terminology in the literature. We use the taxonomy proposed by Dodge et. al. [14] who use the term movement parameters and classify these parameters in three groups: primitive parameters, primary derivatives, and secondary derivatives. Each group is further organized in three dimensions: spatial, temporal and spatio-temporal. In the next paragraphs, we define each of the terms and give a formal definition when necessary.

The only primitive spatial parameter is position, which is a point $s = (lat, lon)$. The primitive temporal parameters are time instance and interval, which represent a point in time and the temporal sampling rate, respectively.

The primary spatial derivatives are distance (see Equation 2), direction, and spatial extent (when the locations are defined as space compartments rather than as points [15]). For the distance derivative we have

$$\Delta s = s_i - s_{i-1},$$  

(2)

where $s_i$ is the current position and $s_{i-1}$ is the previous position.

The secondary spatial derivatives consist of spatial distribution (how a set of entities is spread over, e.g. evenly, aligned, concentrated), change of direction, i.e. the rate that a mobile object changes its direction over consecutive observations, and sinuosity which measures the ratio of the actual length and the distance represented by a straight line between the depart and arrival points.

The primary temporal derivatives are duration (see Equation 3) and travel time. For the duration derivative we have

$$\Delta t = t_i - t_{i-1},$$  

(3)

where $t_i$ is the current time instant and $t_{i-1}$ is the previous time instant.

Secondary temporal derivatives comprise temporal distribution, i.e. how frequent the observations happen, and change of duration, i.e. the rate at which the observations are made.

The primary spatio-temporal derivatives are velocity (see Equation 4) and its scalar value, speed $||v_i||$. The velocity $(v_i)$ is obtained from the results of Equations 2 and 3, such as:

$$v_i = \frac{\Delta s}{\Delta t}.$$  

(4)
Based on the velocity value, we are able to obtain the variation of velocity ($\Delta v$), such as:

$$\Delta v = v_i - v_{i-1},$$

where $v_i$ is the current velocity and $v_{i-1}$ is the previous velocity.

The secondary spatio-temporal derivatives are acceleration (see Equation 6) and approaching rate, which describes whether and how intensely a moving object approaches its destination [10]. To obtain the value of acceleration ($a$) we have

$$a = \frac{\Delta v}{\Delta t},$$

where ($a$) is computed with the values obtained of Equations 5 and 3.

The primitive parameters (position and time) have been adopted in most studies in GIS. In this work, we focus on the primary spatio-temporal derivative speed to illustrate all the steps to represent it qualitatively. It is important to keep in mind that although only the speed is used in the examples along this paper, our approach handles any other derivative that can be represented as a time series.

### IV. THE QUALITRAJ ONTOLOGY

Considering all theoretical background that has been presented, we propose an ontology to represent qualitative data of time-based characteristics of trajectories. Fig. 1 shows the QualiTraj ontology represented as a directed graph. The normal rectangles are nodes (entities) and the dashed rectangles represent nodes attributes that assume literal values. First, we have an entity to centralize one spatio-temporal path, Trajectory, which is composed by one or more Profiles.

A Profile represents a dynamic characteristic of a Trajectory and is the entry point to the qualitative representation of the evolution of this characteristic over time. The name of the node indicates which kind of movement characteristic is stored, e.g. “Speed”, “Acceleration”, “Direction”. It was also added an Global Attribute entity linked to a Profile to represent relevant information about the whole evolution of the characteristic (e.g. median, standard deviation).

Each Profile is composed by a sequence of Segments, which are qualitative representations of relevant changes of the characteristic over time. The kind of change is stored in the Qualitative Value (e.g. “Increase”, “Decrease”, “Steady”). The Coefficient serves to store the angle of the segment line related to the x axis (see Fig. 2). This coefficient is important to determine the slope of the line and infer approximate values along the segment.

We also introduce the concept of Key Points. These points are identified in space and time through Location and Timestamp attributes and may be used to represent important application-specific features of segments. For instance, due to the inherent loss of information, caused by qualitative transformation of raw data, a developer could store the highest and lowest quantitative value of a given characteristic for one or more segments. In this case, there would be two additional points for each Segment having the names Highest speed and Lowest speed and their respective values. These points would be represented by the optional relationship has_point. Although the Key Point entity gives a great flexibility to the user regarding which kind of quantitative information she wants to store, each segment must be linked to at least two points represented by the relationships start_point and end_point in order to specify the times at which the segment starts and ends.

### V. FROM RAW DATA TO SYMBOLIC REPRESENTATION

The data model specified by the QualiTraj ontology is the final step of a process of semantic enrichment of raw data. We describe in this section some tasks that should be performed in order to clean and condense the data without losing its meaning.

Any raw data transformation process depends on the format of both data input and required output. Different datasets may require different preprocessing strategies. In our case, we work with GPS data collected by smart phones through a sport tracking application. The data includes the starting date and time of trajectory, followed by a set of records containing each a geographic position (latitude and longitude) in signed degrees format, as well as the time offset, which is related to the starting time. Although the sampling rate is not constant among trajectories, it is nearly equal within each individual trajectory. Besides, due to existence of a spacial filter, locations are not recorded when the object does not move.

The identification of stops and moves from raw data is the first step of our trajectory semantic enrichment process. This
step is important because it allows the execution of interesting queries. Besides, not properly identifying moves and stops may affect future steps of the preprocessing phase. In our approach, candidate stops are identified based on the analysis of sampling rate and speed values in each time instant, taking into consideration the sampling rate standard score (z-value) and a speed threshold (\(\epsilon\)). The standard score is a statistic metric that indicates how far a sample is from the mean of a population. It is denoted by:

\[
z = \frac{x - \mu}{\sigma}, \tag{7}\]

where \(x\) is the individual sample value, \(\mu\) is the mean value of the population, and \(\sigma\) is the standard deviation.

For each point of a trajectory, we consider the speed \(v = \{v_1, v_2, v_3, \ldots, v_n\}\) and the time interval between the current and previous GPS fixes \(\Delta t = \{\Delta t_1, \Delta t_2, \Delta t_3, \ldots, \Delta t_n\}\). By calculation the standard score (as in Equation 7) of each element of \(\Delta t\), we discover the moments where the interval of time between fixes lasts longer than normal. As a normal interval of time, we consider the statistic mode of the collection \(\Delta t\), i.e., the interval length that happens the most during the entire trajectory.

However, only this metric is not enough to determine a stop because this gap in time may occur due to a period of lost of signal from the GPS satellites. When this kind of error happens, the speed is not significantly changed because the traveled distance is updated proportionally with the time. For instance, if an moving object travels at 2 meters per second and there is a lack of GPS signal for 4 seconds, the calculated distance would be 8 meters (considering constant speed). In the case of a stop, the traveled distance and speed would assume values near zero.

Therefore, we also verify if the speed has decreased to a level lower than a threshold \(\epsilon\). The whole verification process can be seen in Algorithm 1.

**Algorithm 1 Candidate stops detection algorithm**

1: \textbf{procedure} CANDIDATESTOPS\((v, \Delta t, \epsilon)\)  
2: \hspace{10mm} \(\mu \leftarrow \text{mean of } \Delta t\)  
3: \hspace{10mm} \(\sigma \leftarrow \text{standard deviation of } \Delta t\)  
4: \hspace{10mm} \(m \leftarrow \text{mode of } \Delta t\)  
5: \hspace{10mm} \text{for each speed and duration in } v \text{ and } \Delta t \text{ do}  
6: \hspace{20mm} z \leftarrow (\text{duration} - \mu)/\sigma \text{ } \triangleright \text{ according to Eq. 7}  
7: \hspace{20mm} \text{if } z > m \text{ and speed } < \epsilon \text{ then}  
8: \hspace{30mm} \text{Consider as stop}  
9: \hspace{10mm} \text{else}  
10: \hspace{20mm} \text{continue}  
11: \hspace{10mm} \text{end if}  
12: \hspace{10mm} \text{end for}  
13: \textbf{end procedure}

Basically, we assume that every trajectory has been recorded in an environment with good position accuracy. This is an important requirement because movement characteristic calculation has a direct relation with accuracy of signal and an adequate sampling rate. Thus, even though we do not cover this aspect in this paper, a quality evaluation of such trajectory attributes is encouraged in order to discard trajectories with highly disperse sampling rates, which would lead to inaccurate results.

Along any moving parameter time series it is usual to have some individual points that suddenly assume high or low values. These points are outliers caused by inconsistencies in the GPS signal quality. To deal with these cases, we have to filter the signal in order to have a better representation of the real evolution of the trajectory characteristic. Many filtering techniques can be used to smooth the signal, e.g., moving average and moving mean. We chose the Kalman filter as it is widely used in time series analysis and trajectory preprocessing [16].

The representation of all data points of a dynamic characteristic may be inefficient for most of the use cases due to the large amount of data to be stored. Thus, the application of data compression techniques is a important strategy to optimize storage and keep representative points. We use Piecewise Linear Segmentation (PLS), which approximates a time series in a set of straight lines, to achieve this generalized representation of data. There are different categories of algorithms to segment time series as well as approximating strategies to define the straight lines. We refer the reader to the work of Keogh, Chu, Hart et. al. [17] for a comprehensive comparison of the main strategies.

After passing through the PLS step, the data is ready to be represented using the QualiTraj ontology. For each straight line outputted by the PLS algorithm, we create a Segment that follows the structure of the ontology. If the speed evolution represented in Fig. 2 would be stored following the QualiTraj ontology, seven Segment instances would be created. The third one, for instance, would have could be called Segment 3 and have the Qualitative Value attribute set to “Increase”, the value in degrees of the \(\alpha\) angle would be the Coefficient, and two
VI. VALIDATION

In order to validate our approach, we performed some controlled experiments with a sport tracking application to assure that all the transformation processes of raw data to qualitative representation were correct. In this experiment, we chose to represent only one characteristic of a trajectory, its speed. Fig. 4(a) shows raw data of the speed evolution of a person who walked for about 10 minutes stopping three times along the path. As we can observe, the person was moving during most of the time, but there are some gaps in time that take longer than the average sampling rate with instants of very low speed. Although there were only three stops in the real trajectory, there are four intervals where the gap in time are noticeable.

As we discussed in section V, the first step consists in detecting stops and moves where Algorithm 1 was employed having as speed threshold ({$\epsilon$}) 1 m/s based on empirical observation of the dataset. The result can be seen in Fig. 4(b) where the time series is sliced based on stops’ starting and ending points. These points are identified by squares in the figure. As we can notice, near the end of the time series, there is a gap in time longer than the average. However, as the speed does not decrease to a level lower than the threshold, we do not consider this event as a stop.

The following steps consist in a simple application of the Kalman filter to smooth the time series resulting in the series of Fig. 4(c) and then a PLS algorithm was applied to the filtered data. In our implementation, the bottom-up with linear regression approach [17] yield the best results, as can be seen in Fig. 4(d).

Before creating instances of the QualiTraj entities, we have to define the lexical space, i.e. the set of possible values, for the Qualitative Value property. Thus, this entity may assume the values “Increase”, “Decrease”, “Steady”, and “Stop”. For the coefficient attribute, the degrees unit was used. Fig. 3 shows the first two segments represented with the ontology. The first segment consists in an increase of speed from 2.02 m/s to 2.40 m/s, which is represented by a line segment having an angle of 7.24 degrees. This angular information is important to retrieve other values in the same segment when necessary. Notice that the same Key Point, KP 2, has been reused by both segments, avoiding duplication of data thanks to the graph structure provided by the ontological modeling approach that we employed.

Key Points would be stored having 7 seconds as start time, and 10 seconds as end time.
VII. CONCLUSION

In this paper, we demonstrated how it is possible to enrich raw trajectory data with dynamic characteristics of movement and provide an infrastructure for representing this new knowledge through an ontology. With this work, we create a basis for a framework that enables the qualitative comparison of trajectories, taking into consideration the dynamics of motion. Such lack of qualitative dynamics-aware measure has already been noticed by Ranacher and Tzavella [18] where the authors review movement similarity metrics.

The representation of spatio-temporal data by means of ontologies is even more useful when the inference features of reasoners are explored. As a following activity of this work, we will investigate how reasoners can improve the analysis of patterns in trajectories’ dynamic movement characteristics. Queries that involve more than one moving object form an important group of queries about relative motion and should be studied in the future.

Another aspect that can be addressed in the future is the multiscale analysis of trajectory characteristics. While we achieve a compact representation of motion attributes, some applications may require different levels of granularity of the same data. Therefore, algorithms that allow different views of the qualitative data according to several multi-scale criteria are important tools to specialists.

ACKNOWLEDGMENT

The authors would like to thank the French Ministry of Higher Education and Research (Ministère de l’Enseignement Supérieur et de la Recherche de la France – MESR) for supporting this work.

REFERENCES