CONSTAnT - A Conceptual Data Model for Semantic Trajectories of Moving Objects

Vania Bogorny[1], Chiara Renso[2], Artur Aquino[1], Fernando de Lucca Siqueira[3], Luis Otavio Alvares[1]

[1]Department of Informatics and Statistics

Universidade Federal de Santa Catarina – Florianopolis – Brazil

{vania, alvares}@inf.ufsc.br, arturbr@gmail.com

[2]KDD Lab, Pisa, Italy

chiara.renso@isti.cnr.it

[3] Department of Automations and Systems

Universidade Federal de Santa Catarina – Florianopolis – Brazil

fernandols@das.ufsc.br

Keywords: semantic trajectories, conceptual model, trajectory behavior, movement data

Abstract

Several works have been proposed in the last few years for *raw* trajectory data analysis, and some attempts have been made for defining trajectories from a more semantic point of view. *Semantic* trajectory data analysis has received significant attention recently, but the formal definition of semantic trajectory, the set of aspects that should be considered to semantically enrich trajectories and a conceptual data model integrating these aspects from a broad sense is still missing. This paper presents a semantic trajectory conceptual data model named CONSTAnT, which defines the most important aspects of semantic trajectories. We believe that this model will be the basic stone for the design of semantic trajectory databases, where several aspects that make a trajectory "semantic" are taken into account. The proposed model includes the concepts of semantic subtrajectory, semantic points, geographical places, events, goals, environment and behaviour to create a general concept of semantic trajectory. The proposed model is the result of several years of work of the authors trying to add more semantics to raw trajectory data for real applications. Two case studies show the flexibility of the model for different application domains.

1. Introduction and Motivation

The price reduction of mobile devices as GPS, cell phones and sensor networks, significantly increased the generation of spatio-temporal data, called trajectories of moving objects. These devices generate raw data with the identifier of the object, and its position at a time instant. These data are called *raw trajectories*. The analysis that can be performed on these kinds of data are numerous, as witnessed by the interest of several research communities all around the world in the last 10 years. Trajectory research has lead to the creation of basic methods to reconstruct raw trajectories (Marketos et al. (2008)), new and complex datatypes, operations and languages for moving object databases Güting e Schneider (2005), Pelekis et al. (2008) and the creation of enormous amounts of novel data mining algorithms for trajectories, such as Lee et al. (2008), Giannotti et al. (2007), Cao, Mamoulis and Cheung (2007), Li et al. (2010), Alvares et al (2011), Nanni et al. (2008).

However, only a few works in the literature have focused on what are called *semantic trajectories,* that is a hot research topic at the moment. The idea of semantic trajectories is to add more meaning to the raw trajectories in terms of application and contextual knowledge with the aim of getting more meaningful results from the analysis of such trajectories. The first work that addressed semantic trajectories, looking at trajectories from a more conceptual level, was introduced in Spaccapietra et al. (2008), in the context of the European Project GEOPKDD (http://www.geopkdd.eu/) – Geographic Privacy Aware Knowledge Discovery and Delivery. This model is called *stops* and *moves*, and has as the main idea to show that a semantic trajectory is not a raw trajectory as collected by tracking devices, but a sequence of *important places,* where the moving object stays for a period, called stops. Stops are then connected by the moves where the object changes its position.

The trajectory conceptual data models presented so far are based on stops and moves. Although representing a first step towards defining a semantic trajectory, they do not represent several crucial information related to the semantics of the trajectories. For instance: which aspects should be considered in the semantic enrichment? How do they relate each other? Which parts of the raw trajectory do they enrich? Examples of these aspects are the goal of the trajectory, the trajectory semantic partition, the environment information related to both the moving object and the trajectory points and trajectory behaviors. We believe that there is a need to give a definition of what a semantic trajectory is in a broad sense, where the semantic aspects of a trajectory are defined and related to the raw trajectory. Therefore, this paper presents a novel conceptual data model for semantic trajectories, called CONSTAnT (CONceptual model of Semantic TrAjecTories) that aims at integrating and systematizing in a proper way all the aspects that relate to the concept of semantic trajectory. To the best of our knowledge this is the first approach that gives this general and comprehensive definition of semantic trajectory, including all the aspects that may contribute to the semantic enrichment of raw trajectories.

The model presented in this paper has its basis on several works on trajectory data analysis, modelling and mining. It is based on the experience and the knowledge of three of the authors that together with the group of Spaccapietra et al. (2008), since 2007 were the pioneers of *semantic trajectory research*. The model CONSTAnT is defined in the context of the current European Project called SEEK (http://www.seek-project.eu/) – Semantic Enrichment of Trajectory Knowledge Discovery, that has its focus on *semantic trajectory modeling, representation, analysis and mining*. We believe that a model that gives a wide and realistic definition of semantic trajectory should be the first step towards the definition of a semantic enriched trajectory knowledge discovery process.

The proposed model is more general then the previous proposals and supports, as a particular case, the original idea of stops and moves. We evaluate the proposed model with two case studies of different application domains, on human mobility and animal monitoring. The reminder of the paper is organized as follows: Section 2 presents the basic concepts and related works. Section 3 presents the model CONSTAnT. Section 4 presents two case studies for different application domains namely tourism mobility and animal behavior monitoring, while Section 5 concludes the paper and suggests directions of future research.

2. Basic Concepts and Related Works

In order to understand the proposed conceptual model for semantic trajectories we define some basic concepts. In general, existing works deal with raw trajectories, where a trajectory is a list of points located in space and time. So we start defining a point.

Definition 1. Point**.** A *point p* is a tuple (x,y,t), where x and y are the spatial coordinates that represent a place and t is the timestamp in which the point was collected.

A raw trajectory is an ordered list of points, as presented in Definition 2.

Definition 2. Trajectory. A *trajectory T* is an ordered list of points $\langle p_1, p_2, p_3, ..., p_n \rangle$, where $p_j = (x_j, y_j, t_j)$ and $t_1 < t_2 < t_3 < ... < t_n$.

A trajectory has several characteristics that we call *physic characteristics,* which can be either associated to the trajectory data or extracted from the points, as acceleration, speed, direction change, length, and so on.

A subtrajectory can have several definitions, but in general, in the literature a subtrajectory is defined as a segment of the raw trajectory. Therefore a subtrajectory is itself a raw trajectory. Definition 3 presents the concept of subtrajectory.

Definition 3. Subtrajectory. A *subtrajectory s* of *T* is a list of points $<p_k, p_{k+1}, . . . , p_l>$, where $p_k \subset T$ and $k \geq 1$ and $l \leq n$.

The first semantic model for trajectories was introduced in Spaccapietra et al. (2008), and is shown in Figure1. The travelingOT is the moving object, that has a trajectory, a begin, a set of stops and an end (BES), and moves. The stop is in a place where the object remains for a certain amount of time, while a move is the movement between two stops. In this model, the semantic trajectory is represented by a sequence of stops and moves and the main semantic information related to the trajectories is the places where the moving object either stops or moves. The model is limited since the stops have some constraints, as “the object must stay at a place for a certain amount of time for such a place to be considered important”, i.e., a stop.

In case someone may want to represent information about the object or external attributes, in Spaccapietra et al. (2008) this information is added to the Trajectory entity or to the TravelingOT. However, there are some cases where semantic information needs to be added to a specific trajectory point, as for instance, the temperature of an object (e.g. an animal) at each collected point, and this cannot be expressed.

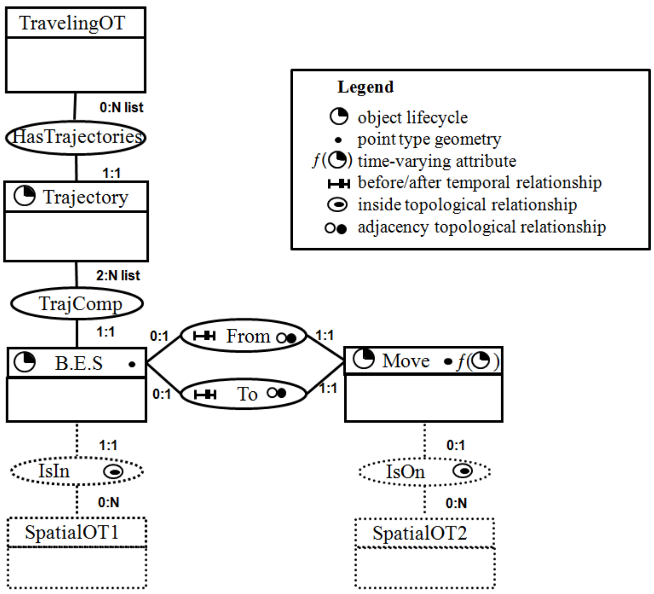

Figure 1 – The model of stops and moves [Spaccapietra et al 2008]

We believe that a conceptual data model for semantic trajectories should be more flexible, such that the user may add any kind of contextual information to the trajectories. The user should be free to choose how to split a trajectory and to choose the information to be added to different parts of a trajectory. Therefore, the model proposed in this paper, named CONSTAnT, considers that a semantic trajectory is build on different semantic subtrajectories, which is the novel concept that we introduce in this paper. The semantic subtrajectories can be generated based on the goals of the moving object, the transportation mean or the behavior, which are also new concepts introduced in CONSTANT. We organized all these aspects in a comprehensive model that relates several kinds of contextual information that may be used to semantically enrich trajectories. This makes the model more general, since it allows the trajectory splitting based on different semantic information and is much more general than the stops and moves approach, which fits in CONSTANT.

It is important to notice that the model presented in Figure 1 does also not support the explicit representation of the goal of a trajectory or of parts of a trajectory. For instance, the trajectory of a professor has the general goal of going to work, while the goal of his/her subtrajectories may be “teaching”, “reading a paper”, “attend a meeting in another department”. The model of stops and moves does not support these levels of representation. The same happens for behavior information.

Behaviors are normally extracted as patterns from the raw trajectories or can be extracted from semantic trajectories. Several works have been proposed using data mining methods to discover different types of trajectory behaviors. These works can be used to partially instantiate semantic trajectories, and are detailed in the following section.

Another conceptual model for trajectories, focussed on trajectory data mining and not on database modelling, is proposed in Bogorny, Heuser and Alvares (2010). That model associates several data mining tasks to the stops and moves of trajectories. However, since CONSTANT can represent stops and moves with the new concept of semantic subtrajectories, all data mining tasks presented in Bogorny et al. (2010) will directly fit into CONSTANT.

Hereafter we present some works that can be used to compute trajectory behaviors, to generate subtrajectories or to discover trajectory goals. Most of these works deal with raw trajectories, and were developed, independently, with specific objectives. Therefore, in this paper we consider these works to propose three types of semantic trajectory behaviours: *single, collective aware* and *collective non-aware.*

The first group of works tries to identify behaviors in groups of trajectories, where one trajectory is not aware of the other one, i.e., behavior patterns are extracted among objects that follow the same path by coincidence, like [Giannotti et al. 2007] [Hornsby and Cole 2007] [Cao, Mamoulis and Cheung 2005]. The objective is to find behaviors among objects with similar movement, but without common intentional behavior. In the work of Giannotti [Giannotti et al. 2007], for instance, the algorithm *T-Pattern* looks for sequences of regions, frequently visited by trajectories in a specified order and with similar transition times. Trajectory behaviors are generated as sequences of regions visited by a minimal number of trajectories.

A second group of works defines a set of behaviors between groups of trajectories that are aware of the movement of their neighbor trajectories. For instance, Laube in Laube, Imfeld and Weibel (2005) defines five types of trajectory collective behaviors: Convergence, Encounter, Recurrence, Flock and Leadership. In Cao, Mamoulis and Cheung (2006) Cao explores the collocation episodes in spatio-temporal data. The main objective is to find objects that move together for a certain amount of time and make another object move together too. Therefore, the concept of time window is used, where trajectories are divided in time slices, and the relationship between objects is identified through the distance between points in each time window. This method can be used for discovering behaviors of a puma hunting a dear. A more recent work of this group is the chasing behavior, defined in Siqueira and Bogorny (2011). This work proposes the Tra-Chase algorithm to find subtrajectories of objects that follow other objects for a certain amount of time.

The third group of works tries to identify common behaviors in single trajectories, trying to understand the behavior of the object by analyzing individual movements. Palma et al. (2008), for instance, proposes the method CB-SMOT which uses the idea of stops and moves of Spaccapietra et al. (2008) to find subtrajectories (stops) where the object moves slowly. Alvares in Alvares et al. (2011) addresses a new behavior of trajectories to find subtrajectories that avoid points/regions, like a thief avoiding a surveillance camera. It also evaluates the confidence of the pattern, to ensure that it was an intentional avoidance. Rocha in Rocha et al. (2010) proposes the algorithm DB-SMOT to find subtrajectories in single trajectories where the direction change characterizes the behavior, as for instance a vessel in a fishing region.

In the approaches proposed in Spinsanti et al. (2010] and in Baglioni et al . (2009) the term behavior is used as synonym of goal, and they both aim at annotating trajectories with goals inferred from a set of predefined goals. In Spinsanti et al (2010) a semantic trajectory is associated to a goal that is related to a stop. For example, a trajectory that has a high frequency of stops at touristic points of interest is associated to a "tourism" goal. However, this approach does not associate goals to subtrajectories. The work in Baglioni et al (2009) introduces a system called Athena that annotates trajectories and patterns with predefined goals (e.g. HomeWork, Tourism etc) using domain knowledge encoded in an ontology. However, similarly to the previous work, they do not consider subtrajectories.

In Lee et al. (2008), the method Tra-Class is proposed to classify trajectories in one class from a set of possible classes. However, this method does not classify subtrajectories.

The following section presents the proposed model and several definitions for semantic trajectories

3. CONSTANT: A Conceptual Data Model for Semantic Trajectories

In this section we present the CONSTANT conceptual data model for semantic trajectories. This model is shown in Figure 2. We can divide the model in two main parts: the first part (the non-colored classes) contains information related to: the *traveler (object)*, de *device that generated the trajectory*, the *semantic trajectory*, the *semantic subtrajectories*, the *semantic points, the environment, the place* and *events*. This part of the model is detailed in section 3.1.

The second part of the model (dark gray classes) is more complex, and in general, advanced methods like data mining are needed to instantiate this part of the model, which includes: the *goals* of the semantic trajectories and semantic subtrajectories, the *transportation mean*, and the *behavior* of the semantic subtrajectories. This part is detailed in section 3.2

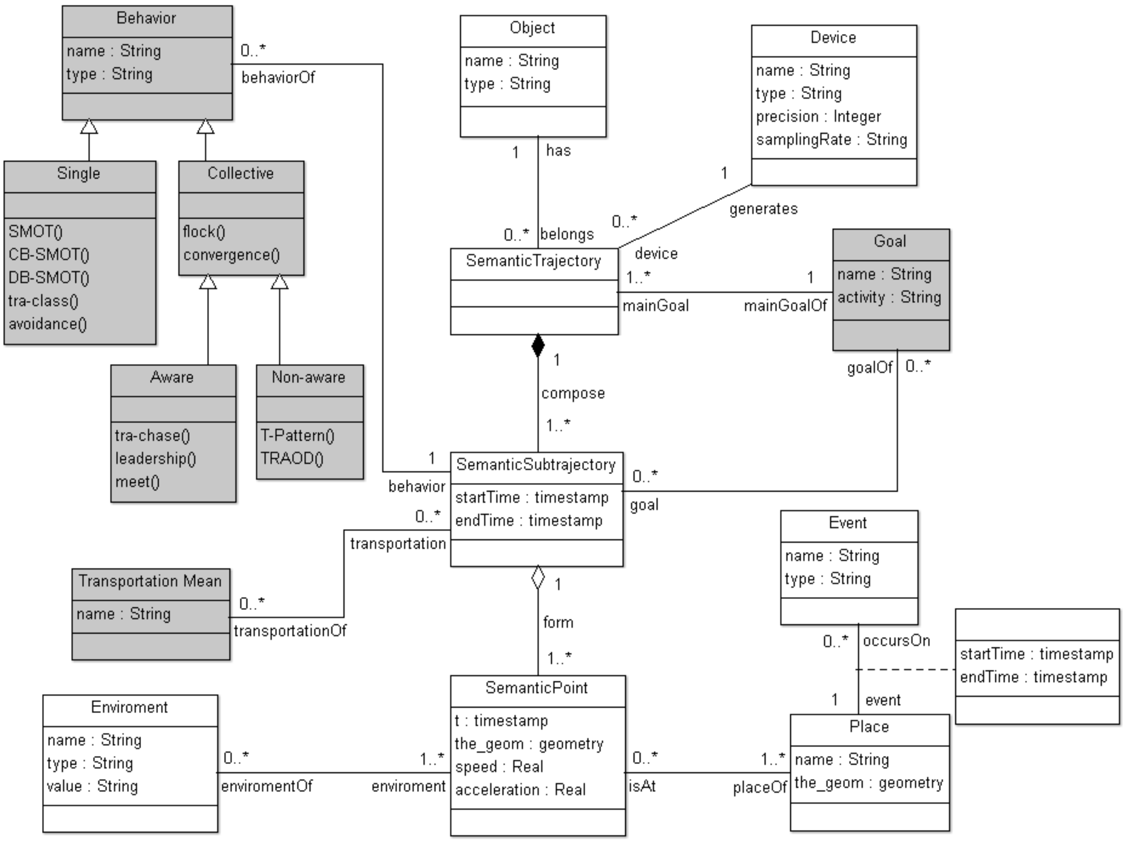

Figure 2 – CONSTANT: A conceptual data model for semantic trajectories

3.1 Semantics Related to Trajectory Points

In the semantic trajectory conceptual model, the object (traveler) represents the moving object which carries the mobile device that generates the trajectory. The object can be an animal, a person, a car, a plane, a robot, etc. We define the object according to definition 4.

Definition 4. Object. The *object O* is the moving object that carries the mobile device and has a name, a type and an identifier oid.

The device is the apparatus that collects the trajectory as a sequence of points, and can be a GPS, a cell phone, a sensor network, or any other device that captures the position of an object at a time. The main attributes of a device include, but are not limited to:

- Name: an identification of the device
- Type : its type (e.g. sensor, GPS)
- Precision: the geographical precision of the collected points
- SamplingRate: describes how the points are collected (e.g. each second, each 10 seconds, each 10 meters).

Each object has zero or more trajectories. These trajectories will be *semantic trajectories* only if they are associated to a *goal* and have at least one *semantic subtrajectory*. When this information is missing, the trajectory is simply a raw trajectory. Each semantic trajectory is build from at least one *semantic subtrajectory*. In our model, a *semantic trajectory* is defined according to definition 5.

Definition 5. Semantic Trajectory. A *semantic trajectory* $\mathcal{T}$ is a tuple (tid, oid, $\mathcal{S}$, g, d) , where *tid* is the trajectory identifier, *oid* is the object identifier, $\mathcal{S}$ is a list of *semantic subtrajectories*, g is the general goal of the trajectory expressing the reason why the object moves, and d is the device that generated the trajectory.

A *semantic subtrajectory* can be instantiated according to the application domain. For instance, it can incorporate the model of stops and moves, where stops and moves represent different semantic subtrajectories. A semantic subtrajectory can also be created based on different semantic information, like *goal* or *transportation mean*, as will be explained later in section 3.2.

A subtrajectory is a *semantic subtrajectory* if it has the semantic information of goal, behavior or transportation mean related to it and has a sequence of semantic points. If this information is missing, it is a raw subtrajectory. We define a semantic subtrajectory according to definition 6.

Definition 6. Semantic Subtrajectory**.** A *semantic subtrajectory* $s \subset \mathcal{T}$ is a tuple (tid, sid, $\mathcal{P}$, $\mathcal{G}$, $\mathcal{M}$, $\mathcal{B}$, startTime, endTime), where *tid* is the semantic trajectory identifier, *sid* is the semantic subtrajectory id, $\mathcal{P}$ is a list of consecutive semantic points, $\mathcal{G}$ is the set of goals of the subtrajectory, $\mathcal{M}$ is the set of transportation mean, and $\mathcal{B}$ is the set of behaviors. It has a start and end time, and at least one of $\mathcal{G}$, $\mathcal{M}$, or $\mathcal{B}$ must be non-empty.

The goal, transportation mean and behavior will be detailed in section 3.2.

A *semantic subtrajectory* has a set of points located in space and time. The duration of a semantic subtrajectory can be extracted from the time attribute of the first point and the last point of the subtrajectory. The attributes speed and acceleration of a point are calculated in relation to the previous point and can be given directly by the device or calculated by a method.

For a point to be a *semantic point* p it must be related to at least one place and can have some environment information. *For the place w*e consider that the instance stored in the database will be at the lowest level, such that any higher level analysis can be build by integrating the trajectory point with geographical information. For instance, let us consider that a subtrajectory point is located at a specific street address (e.g. Beira Mar Avenue, 800). Having this information stored, one may do higher level analysis, as for instance, this point at Beira Mar Avenue 800 is at a *Pub*, in the city centre *District*, in the *city* of Florianopolis. Such information can be inferred by using ontologies, a knowledge base or simply a geographic database storing the places that are relevant to the application.

In case there is a need to store all the geographical places where a point is located, this can still be represented as defined in the model of stops and moves, following the OGC standards, by replacing *Place* by the set of geographical feature types that contain the instance of place (e.g. streets, pubs, hotels, restaurants, cities, districts). The place in the proposed model has the same meaning as the *spatial object* *type* in the model of stops and moves Spaccapietra et al. (2008), and represents a spatial feature type in the OpenGIS Consortium abstract specification model OGC (2011). With this representation it is possible to infer any spatio-temporal relationship of the point with any place, since the point has the time information. For example, if at 10AM the point is inside a hotel, that is inside a district, and that is inside a city, the point has a spatial relationship with its neighbours.

For a point to be semantic it must have associated at least one place or some environment information, otherwise it is considered a raw point. The Semantic point is defined in Definition 7.

Definition 7. Semantic Point. A *semantic point* $\mathcal{p} \subset \mathcal{s}$ is a tuple (pid,x,y,t,$\mathcal{V}$,$\mathcal{L}$, sid) where x and y are the geographic coordinates, collected at time t, $\mathcal{V}$ is the set of environment information related to the point, $\mathcal{L}$ is the set of places where the point is located and *sid* is the semantic subtrajectory identifier to which the point belongs.

In the model, *x* and *y* are represented by the attribute *the_geom,* according to OGC standards. The place where the point is located can be either related or not to an *Event.* An event occurs at a place during a time period (start and end time). Examples of events are, for instance, a musical show on May 13th, 2012 (from 8PM to 11PM), at Main Square in the city of Madrid, and a football match at Bernabéu Stadium on May 14th, 2012 (from 4PM to 6PM). With this representation it is possible to infer when a semantic subtrajectory point is at a place with an event, and the timestamp of the point is in the event period. The Event is defined in Definition 8.

Definition 8. Event**.** An *event* e is a tuple (ℓ, name, type, startTime, endTime), where ℓ is a place, *name* is the name of the event, *type* is the kind of event, *startTime* is the starting time and *endTime* is the ending time of the event.

The environment describes the physical context where the point is collected. It may represent information as temperature, wind speed, humidity, weather, atmospheric pressure. Depending on the domain, other measures can be used as, for instance, oxygen and nitrogen concentrations or pollution degree. To represent different environment information we consider the attributes *name, type,* and *value* for the environment, which can represent different information that is relevant to the application. The environment is defined in Definition 9.

Definition 9. Environment. *Environment* $\mathcal{V}$ is a triple (name, type, value) where *name* is the name of an attribute, *type* expresses if the attribute refers to external or internal information of the object, and value is the value of the attribute.

The environment information associated to each point can be either externally collected or collect by the device. For instance, the temperature of a point can be an interpolated value in a region or collected by the device. For instance, for an animal application, one may want to represent the external air temperature and the temperature of the animal at a specific point (or all points) of the trajectory to measure the health status of the animal. If this is the case, two instances can be created for environment with the attribute *name* being temperature and the *type* being internal (related to the object) and external for referring the outside world.

We related the environment information to the point since it represents the environment where the object moves, and this information may change during the movement. Therefore, having represented the most specific granularity level for environment allows the analyst to do more detailed analysis. In general, any additional information of the moving object that can change during its trajectory should be represented as Environment, which is related to each point.

3.2 Behavior, Goal and Transportation Mean

Every *semantic subtrajectory* may have *Transportation Means*, *Goals* and *Behaviors*. The transportation mean represents the mean that allows the object to move, for example by car, by bike, by plane, on foot, by boat. The transportation mean is related to the subtrajectoy and not to the object itself, since the transportation mean of an object can change during its movement. As a common example, consider the trajectory of a person going to work: the person can have three transportation means: start *on foot*, take *a bus*, *on foot*, take *a train,* and finish the trajectory *on foot.*

The transportation means of the trajectory of an animal that carries a GPS device or an Rfid, could be intended in a broader sense, depending on whether the animal is moving independently – therefore transportation means can be the way it moves such as flying, walking, running etc – or has been transported by a car, truck, boat, etc.

3.2.1 Goal

A *semantic trajectory* has a general goal (e.g. go to work) since objects always move for a reason. The goal is the objective of the movement and can be relative to the entire trajectory or part of the trajectory (the semantic subtrajectory). The goal represents the reason for the movement, or, in other words, it represents *why* the object moves.

Depending on the application, examples of goals can be *jogging*, *go to the gym*, *work*, *eat, having fun*. Goals can be represented as a taxonomy, from the more specific to more general. For instance: *go shopping* can be the more specific goal while *leisure* the more general, or *parking the car* could be the more specific goal while *going to work* the more general.

Goals can be related to activities, therefore *activity* is an attribute of goal. For example, if the goal is *leisure,* the activity could be *eating at a restaurant*. However, sometimes activity and goal may be used as synonyms, as for example when the goal is itself an activity (e.g. jogging). A definition for goal is presented in definition 10.

Definition 10. Goal. A *Goal* $\mathcal{G}$ is a tuple (id,name,activity) where *id* is the goal identifier, *name* is the objective that the moving object wants to achieve during his/her movement and *activity* is what the object is going to do.

A goal could have a start time and end time. This may coincide with the subtrajectory start and end time, but not necessarily, since it depends on the subtrajectory segmentation. Such attributes of the goal are up to the application.

(a)

(b)

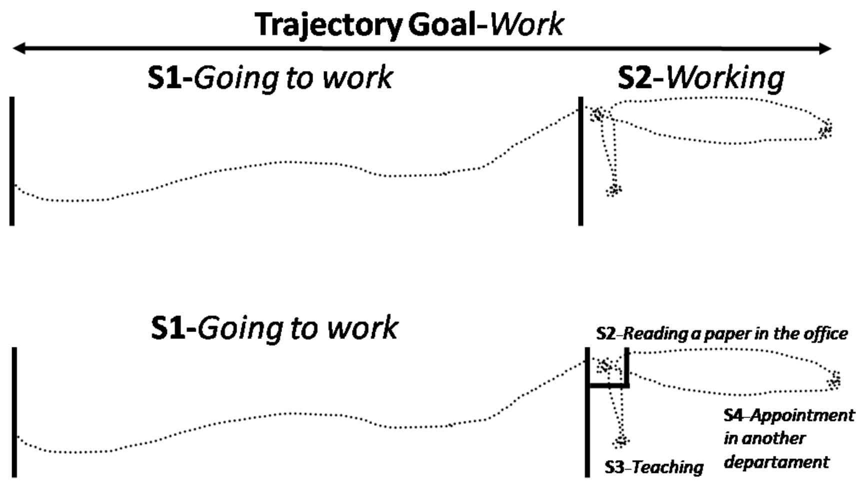

Figure 3 – Creating semantic subtrajectories by goal

According to the application domain, the user may want to create a semantic trajectory based on different semantic subtrajectories. An example is to generate subtrajectories based on the goal: let us consider the trajectory of a professor going to work, as shown in the example of Figure 3 (a). The general goal of the trajectory is to *work* since the person is moving with the objective of working. Two semantic subtrajectories S1 and S2 could be generated based on the general goal of the trajectory, as, for instance, *going to work* and *working.* Depending on the application, someone may want to consider more detailed subtrajectories based on the activities that the moving object performs. An example is shown in Figure 3 (b), where the same semantic trajectory of the professor is created based on the activity, generating 4 different subtrajectories: S1 – going to work, S2 - reading a paper in the office, S3 - teaching, and S4 - appointment in another department. In this example, every semantic subtrajectory has only one goal, because the subtrajectory is created based on the goal.

Depending on the application, the user may want to generate semantic trajectories considering subtrajectories for each transportation mean used in the trajectory. For instance, let us consider the example in Figure 4, where the professor goes by bus to the metro station, from the metro station he goes to the university, inside the university he moves on foot, and inside the university he goes by bus to a meeting at another department, where he moves on foot. In this case, the semantic trajectory of the professor would have six *semantic subtrajectories*, based on transportation mean. In this example, some subtrajectories may have more than one goal, which is the case of subtrajectory S3, where the professor is moving on foot inside the university and the goals are *teaching* and *reading a paper*.

Figure 4 – Creating semantic subtrajectories by transportation mean

As another example, let us consider the trajectory of a person *going to a shopping center.* The main goal is to relax at the shopping center, and every subtrajectory of the movement in the shopping has a different goal: eat something, watch a movie and buy a t-shirt. In this example, the goal of the subtrajectory is also the activity that is an attribute of goal.

In summary, a semantic trajectory has a general goal, but a semantic subtrajectory may have one or more goals. A semantic subtrajectory may have some behaviors, as detailed in the following section.

3.2.2 Behavior

Behavior representation is an important feature in a semantic trajectory. A trajectory behavior has a set of characteristics that identify a peculiar bearing of a moving object or a set of moving objects. In other words, the behavior can show how and why a trajectory is moving, adding semantics to the movement. In general, the behavior of a trajectory or a group of trajectories cannot be understood or inferred by simply looking at the data or by applying a simple query. The behavior of a trajectory or a group of trajectories, in general, has to be computed with intelligent methods, like data mining algorithms.

In the presented model, each semantic subtrajectory can have none to several behaviors, which describe its movement. We define subtrajectory behavior based on several trajectory data mining works that have been proposed so far to discover patterns from both raw trajectories and semantic trajectories. Therefore, we classify a general behavior as *Single* and *Collective. Single* behavior refers to individual trajectories analysis, i.e., the behavior is related to the subtrajectories of the same object. *Collective* behavior is a pattern that repeats among a group of trajectories, that can be either of the same object or not.

In the proposed model we consider that a trajectory can have many different behaviors in its lifetime, and therefore the behavior is related to a semantic subtrajectory, and not the whole trajectory. In the proposed model, a subtrajectory is related to a *Behavior* that has a name and a type. The name attribute represents the behavior that someone may look for, as for instance, chasing, flock, encounter, avoidance. The type attribute expresses if the behavior is collective or single.

Single behavior analysis is useful when the user is interested in analyising individual trajectories, based on the places where the moving object has been, like the stops. Examples of methods that can be used to compute single behaviors are SMOT (Alvares et al. 2007), CB-SMOT (Palma et al. 2008), DB-SMOT (Rocha et al. 2010) and Tra-class (Lee et al. 2008), detailed in Section 2. The first three methods can be used for both computing *semantic behaviours* or generating semantic subtrajectories. With the method CB-SMOT, for instance, that finds subtrajectories based on low speed individual subtrajectories, one can identify low speed traffic areas in a city. With the method DB-SMOT, that discovers subtrajectories where the direction of the movement varies more than in the rest of the trajectory, one can identify fishing areas of individual boats. Another single behavior can be, for instance, the avoidance Alvares et al. (2011), that occurs when a single trajectory avoids a static object, as a security camera.

In this paper we introduce a new classification for collective behaviour, considering different semantics: ***aware*** *behavior and* ***non-aware*** *behavior*. The main difference between these two behaviors is that in the aware behavior, at least one trajectory is aware of other trajectories, and changes its movement based on the movement of the other trajectory. In other words, the behavior of a trajectory is affected by other trajectories, and the collective movement *is not a coincidence*. For example, the *chasing behavior* (Siqueira and Bogorny, 2011) is an *aware* behavior because a trajectory of an object (called stalker) follows another trajectory (called target), and needs to change its movement to keep following the target to characterize the chasing behavior. The leadership movement (Laube et al., 2005) is also considered an aware behaviour, because a group of trajectories follows a leader trajectory.

The non-aware behavior occurs when a group of trajectories has similar behavior *by coincidence* or *without the intention to stay with the other trajectories*. The movement of an object does not intentionally affect the behavior of the other objects. An example of this behavior is a group of trajectories moving in a highway. The trajectories move together restricted by the road network, without the intention to being together. An example of collective non-aware behavior is the work of Giannotti et al., (2007) which proposed the T-pattern algorithm. This method extracts sequences of regions visited in a specified order and with similar transition times, but the behavior of each individual trajectory is not related to the neighbours. Most of the data mining algorithms proposed so far for discovering patterns look for non-aware collective behaviours.

Some behaviors can be either aware or non-aware, depending on the context. For instance, a flock behavior (Laube et al., (2005)) could be an aware movement when a team of football players travels to a match or birds flying together. These objects move together by intention and are aware of each other. On the other hand, a group of people in a train, or a group of cars in a highway move as a non-aware flock.

A semantic subtrajectory can present both single and collective behaviors at the same time. For instance, a trajectory can chase another trajectory while avoiding surveillance cameras. The behaviors of semantic trajectories, in general, differently from the location or the events, are calculated by intelligent methods and, in general, are not defined by the user. The user should choose the types of behavior that are important for his/her context and application, and choose the methods that compute these behaviors.

Although this model has been developed to be as general as possible to represent the semantic trajectories aspects, it has to be intended as a core model that can be further extended and developed for more specific application cases. For example, the goal can be further developed into a taxonomy of goals and activities, or the behavior can be further extended towards a more structured behavior taxonomy.

4. Experimenting CONSTANT in two case studies

In this section we present two case studies on two different applications, based on real trajectories. The first case study is based on a tourist application, while the second one is based on a bird migration application.

4.1 CONSTANT instantiation for a Tourism Application

In this example we instantiate the CONSTANT model in a very common application for mobility: Tourism. Usually, this kind of application needs to study the movement of people visiting a location – that can be a city (e.g. Rome, or Paris), a specific monument (e.g. Stonehenge) or natural attractions like parks. Several actors related to tourism management may benefit from analyzing tourists mobility: from local municipalities to attraction managers, tourism offices, accommodation and transportation organizations. The motivations for this kind of analysis can be manifold, ranging from improving the user experience in the visiting places, to better transportation planning and urban sustainability. There is a need to perform a deeper analysis that goes beyond a simple counting of the hotel presence or sold museum tickets: we need to understand how these tourists use the city, its transportation network, its attractions and the services. For this reason, the need for semantic enriched mobility data that integrates the pure mobility aspect with the contextual knowledge is significantly growing in this area.

Here, we assume to have a dataset of trajectories of tourists moving in Rome (Italy) to visit the main city attractions. We assume the person holds a GPS device where the movement tracks are stored with a sampling rate of 10 seconds and a precision of 5 meters. In this application we present an example of a typical daily trajectory of a specific tourist (called John Smith). Figure 5 shows the route of John Smith on day April, 2, 2012. The starting point of the trajectory is the hotel, corresponding to the point labelled (H. Each labeled point in the figure corresponds to a long (e.g. at least one hour) stop. From point H, John proceeds walking to point B, then walking to point C and so on. During this daily trip, the movements are performed by different transportation means such as on foot, bus, and metro.

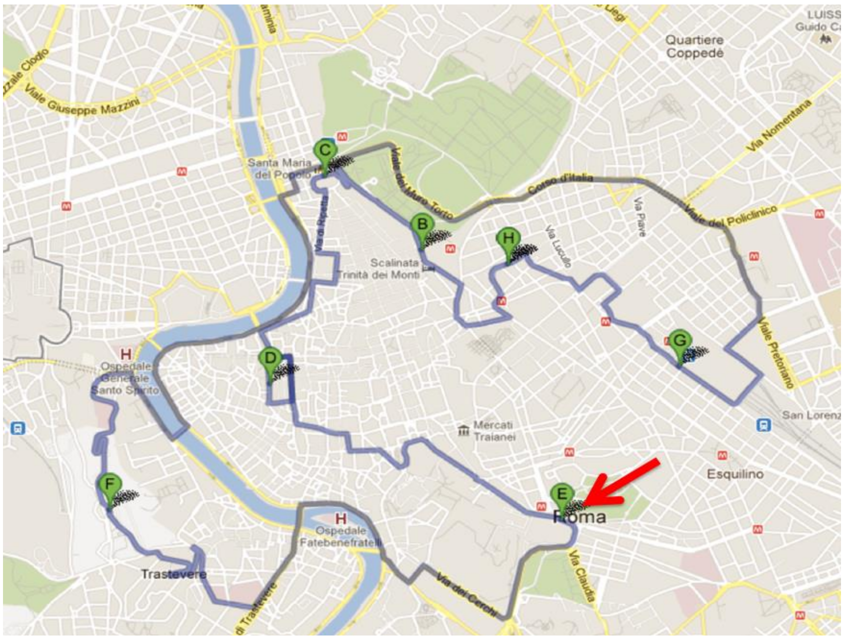

Figure 5. Image of a tourist trajectory in Rome

To illustrate the use of CONSTANT we instantiate the model on a specific visiting place which is labelled E, and corresponds to the Colosseum, a well known Roman Amphitheater. The instantiation of CONSTANT is shown in Figure 6.

The visit at the Colosseum corresponds to a semantic subtrajectory starting at 13:50 when John arrived at the monument and ending at 15:10. The semantic point represented in the model is a specific point inside this subtrajectory collected at 14:00 during the stop. Indeed, the velocity is low since it corresponds to the small movement inside the Colusseum. We can notice in the model that this point corresponds to a geographical *place* which is the Colosseum, and this place is linked to an *event* taking place there from 8:30 to 22:00 (Photographic exhibition). *Transportation mean* in this subtrajectory is *on foot*, while the *environment* – for the purpose of this application – corresponds to the weather in the city. In fact, the weather can affect the visit of a city since when it is raining tourists usually give preference to visit indoor places like museum, while in sunny days tourists may prefer to walk and visit outdoor attractions. During this day the weather was sunny and this may explain the several "walking" subtrajectories. The walking subtrajectories are several instances of the model, which in this example represents only one instance (so only one subtrajectory).

Weather conditions may be collected from meteo web sites for the different time of the day and different areas of the city, and can be either automatically added to the trajectory points or manually annotated by the user.

The *goal* of this semantic subtrajectory is “Attending an Event“ since John stopped at the Colosseum specifically to visit the photographic exhibition. We can notice how the goal of the entire trajectory is different from the goal of the subtrajectory, being at a different granularitrity, since the objective of the whole day is “to visit Rome”.

We also assume that several tourists that are participating in the same event at E (exhibition) go to point F “Gianicolo”, a well known touristic place in Rome. Therefore, this subtrajectory may belong to a T-Pattern behavior, representing a high frequency movement pattern from location E to location F. For this reason, the *behavior* of the selected subtrajectory is *collective non-aware,* since there is not a specific will from all tourists to “go together”. The method to compute this behavior is *T-Pattern()*, and is applied to the whole tourists dataset.

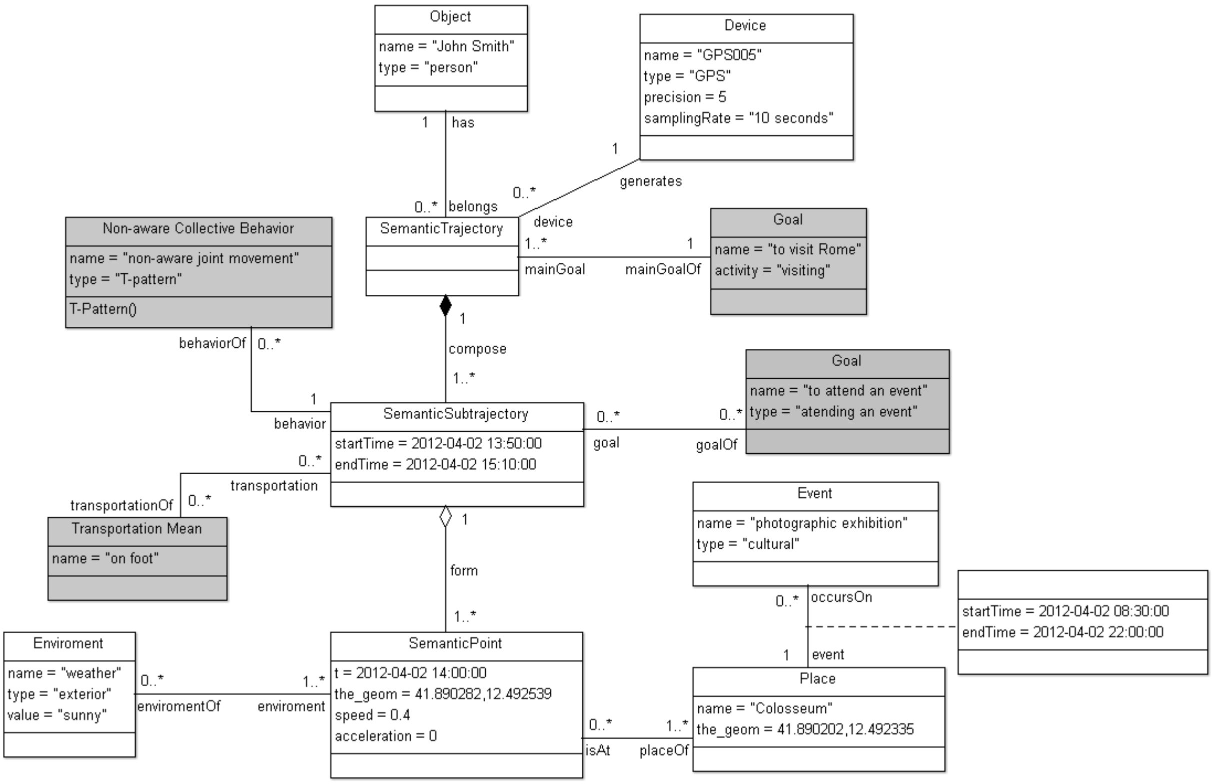

Figure 6. CONSTANT instantiation for John´s tourist trajectory.

In the following section we present a case study on trajectories of migration birds, with different characteristics of the tourism application. This application is the same used to instantiate the model of stops and moves by Spaccapietra et al (2008).

4.2 Model Instantiation for a Bird Migration

This application case study is related to bird migration, since nowadays it is very common for biologists to study animal behavior with tracking techniques, which can be radio-station based (with low spatio-temporal granularity) or GPS-based (with higher spatio-temporal granularity). Here we assume that the tracked animals are birds equipped with a GPS collar that has a sensor that measures the temperature of the bird. Because of energy consumption, the satellite communication rate of the device is much lower than in the tourism application, where the GPS of tourist in a car can collect a point every second. In the birds application we assume that the sampling rate is more or less 30 minutes.

Many birds migrate long distances. Figure 7 shows an example of the migration route of the white stork bird, which is a long-distance migrant. White Storks fly south from their summer breeding grounds in Europe in August and September, heading for Africa, where they spend the winter. White storks diet varies according to the season, but in general they are highly opportunistic feeders who consume a wide variety of prey items including insects, frogs, toads, tadpoles, fish, rodents, snakes, lizards, earthworms between many others. They visually search for the prey while walking deliberately on the ground. When flying, white storks congregate in large flocks which may exceed a thousand individuals (Wikipedia, 2012).

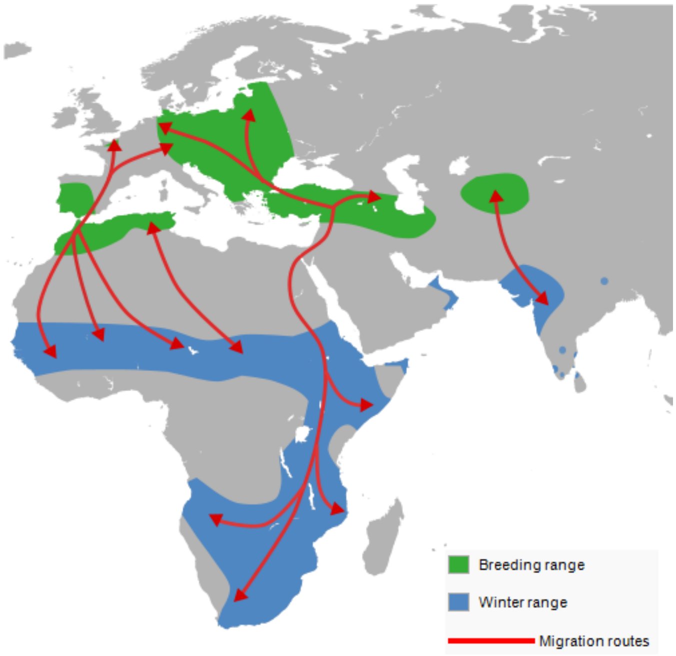

Figure 7 - White Stork bird migration map (Wikipedia 2012)

Considering the instantiation of our model, shown in Figure 8, the object is the white stork bird (we call bird1), that has a semantic trajectory generated by a GPS device with precision 20 and sampling rate around 30 minutes. The semantic trajectory has as the main goal "to migrate from Europe do Africa", where the activity is "migrating". We consider in this case study that semantic subtrajectories are generated by dividing a trajectory in specific goals: "flying", "feeding" and "breeding". The model shows one semantic subtrajectory, with the goal "flying", where the activity is flying as well. In a semantic subtrajectory with goal "breeding", for instance, the activity could be looking for food during the breeding for the baby. For the semantic subtrajectory "flying", represented in the model, we considered a time duration (starting at 12:20 and ending at 15:20). For this subtrajectory, the transportation mean is "wing", while for another semantic subtrajectory as "feeding", for instance, the transportation mean could be "leg", since the white stork walks looking for food. As by definition these birds fly in large groups, the most appropriate behavior when the semantic subtrajectory is "flying", is the "aware collective flock". For another semantic subtrajectory as "feeding" or "breeding", this behavior may be chasing (when the male follows the female or the female chases the baby). The behaviour depends on the analysis of the trajectories that the user is interested in.

The semantic subtrajectory has several points, in our example the specific semantic point has speed 20, and time 12:50 and is located in the place "Nigeria". The environment information considered in this application example is "temperature", and internal to the object, not external as in the tourist application.

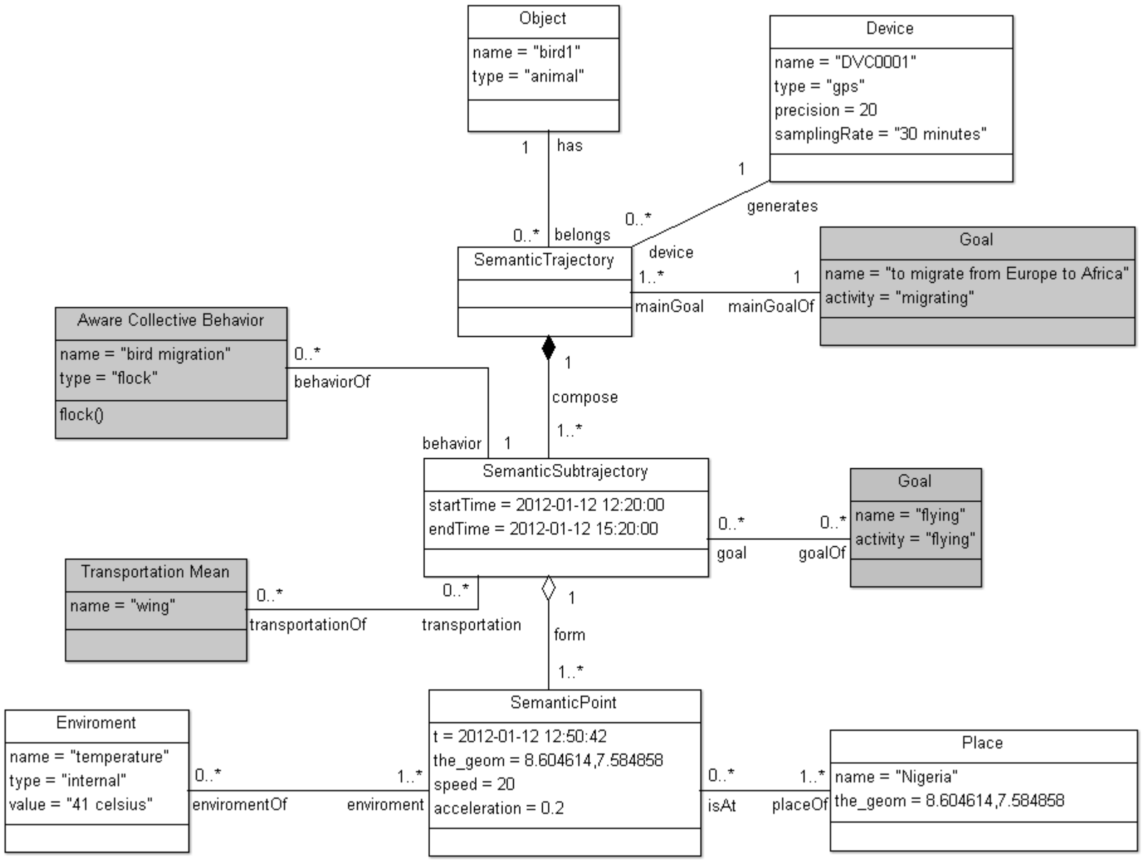

Figure 8 - CONSTANT instantiation for white stork bird1 trajectory

In this section we presented a case study with two different application domains to instantiate the proposed semantic trajectory conceptual model.

5. Conclusions and Future Works

The need for analyzing the movement of entities (e.g. people, animals or objects) in a meaningful way is becoming more and more urgent, because of the flood of spatio-temporal data and emerging applications. Although several approaches have been developed as new analysis techniques for raw trajectories, we are witnessing a growing interest in finding more meaningful ways to represent movement. However, we believe that a comprehensive view of what a semantic trajectory is and which aspects make a raw trajectory a semantic one, is still missing. The proposed model CONSTANT goes in this direction of presenting a conceptual model for semantic trajectories, embedding and organizing several fragmented approaches proposed so far and for different purposes on trajectory data analysis. We demonstrate the usefulness of the model with two case studies of model instantiation in different applications: tourism and animal behavior.

We strongly believe that the proposed model CONSTAnT provides the basic concepts for semantic trajectories and is the baseline rock for future research on semantic trajectory databases, semantic trajectory data warehouses and semantic trajectory knowledge discovery.

As future work we intend to extend the goal entity, as for instance, developing a taxonomy of goals and activities. The behavior representation also needs further studies and representation, towards a more structured behavior taxonomy.

Acknowledgments

Authors if this paper would like to thank EU project FP7-PEOPLE SEEK (N. 295179

http://www.seek-project.eu) for partially supporting this work, CAPES, CNPQ

References

Alvares, L. O.; Bogorny, V.; Kuijpers, B.; Macedo, J. A. F.; Moelans, A.; Vaisman, A. "A Model for Enriching Trajectories with Semantic Geographical Information." *Proc. of the ACM 15th International Symposium on Advances in Geographic Information Systems (ACM-GIS'07).* Seattle, 2007. 162-169.

Alvares L O, Loy A, Renso C, and Bogorny V. 2011. An algorithm to identify avoidance behavior in moving object trajectories. J. Braz. Comp. Soc., 17(3):193-203, 2011.

Baglioni M, Macedo J A, Renso C, Trasarti R, and Wachowicz M. 2009. Towards semantic interpretation of movement behavior. In Monika Sester, Lars Bernard, and Volker Paelke, editors, AGILE Conf., Lecture Notes in Geoinformation and Cartography, pages 271-288. Springer, 2009.

Bogorny V, Heuser C A, and Alvares L O.2010. A conceptual data model for trajectory data mining. In Sara Irina Fabrikant, Tumasch Reichenbacher, Marc J. van Kreveld, and Christoph Schlieder, editors, GIScience, volume 6292 of Lecture Notes in Computer Science, pages 1-15. Springer, 2010.

Cao H, Mamoulis N, and Cheung D W. 2005. Mining frequent spatiotemporal sequential patterns. In ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining, pages 82-89, Washington, DC, USA, 2005. IEEE Computer Society.

Cao H, Mamoulis N, and Cheung D W. 2006. Discovery of collocation episodes in spatiotemporal data. In ICDM, pages 823-827. IEEE Computer Society, 2006.

Cao H, Mamoulis N, and Cheung D W. 2007. Discovery of periodic patterns in spatiotemporal sequences. IEEE Trans. Knowl. Data Eng., 19(4):453-467, 2007.

Giannotti F, Nanni M, Pinelli F, and Pedreschi D. 2007 Trajectory pattern mining. In Pavel Berkhin, Rich Caruana, and Xindong Wu, editors, KDD, pages 330-339. ACM Press, 2007.

Güting R H, and Schneider M.2005 Moving Objects Databases. Morgan Kaufmann, 2005.

Hornsby K S, and Cole S J. 2007 Modeling moving geospatial objects from an event-based perspective. Transactions in GIS, 11(4):555-573, 2007.

Laube P, Imfeld S, and Weibel R. 2005 Discovering relative motion patterns in groups of moving point objects. International Journal of Geographical Information Science, 19(6):639-668, 2005.

Lee J, Han J, Li X, and Gonzalez H. 2008 TraClass: trajectory classiFIcation using hierarchical region-based and trajectory-based clustering. PVLDB, 1(1):1081-1094, 2008.

Li Z, Ding B, Han J, Kays R, and Nye P. 2010 Mining periodic behaviors for moving objects. In Rao B, Krishnapuram B, Tomkins A, and Yang Q, editors, KDD, pages 1099-1108. ACM, 2010.

Marketos G, Frentzos E, Ntoutsi I, Pelekis N, Raffaeta A, and Theodoridis Y. 2008. Building real-world trajectory warehouses. In Delis A, Zadorozhny V, Kotidis Y, and Marrón P J, editors, MobiDE, pages 8-15. ACM, 2008.

Nanni M, Kuijpers B, Körner C, May M, and Pedreschi D. 2008. Spatiotemporal data mining. In Giannotti F and Pedreschi D, editors, Mobility, Data Mining and Privacy, pages 267-296. Springer, 2008.

OGC. OpenGIS Standards Abstract Specifications. 2012. Available at: http://www.opengeospatial.org/standards/as

Palma A T, Bogorny V,Kuijpers B, and Alvares L O. 2008. A clustering-based approach for discovering interesting places in trajectories. In Roger L. Wainwright and Hisham Haddad, editors, SAC, pages 863-868. ACM, 2008.

Pelekis N, Frentzos E, Giatrakos N, and Theodoridis Y. 2008. Hermes: aggregative lbs via a trajectory db engine. In Jason Tsong-Li Wang, editor, SIGMOD Conference, pages 1255-1258. ACM, 2008.

Rocha J A M R, Times V C, Oliveira G, Alvares L O, and Bogorny V. 2010. DB-SMOT: A direction-based spatio-temporal clustering method. In IEEE Conf. of Intelligent Systems, pages 114-119. IEEE, 2010.

Siqueira F de L, and Bogorny V. 2011. Discovering chasing behavior in moving object trajectories. T. GIS, 15(5):667-688, 2011.

Spaccapietra S, and Parent C. 2011. Adding meaning to your steps (keynote paper). In Manfred A. Jeusfeld, Lois M. L. Delcambre, and Tok Wang Ling, editors, ER, volume 6998 of Lecture Notes in Computer Science, pages 13-31. Springer, 2011.

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

Spinsanti, L., Celli, F., Renso, C. 2010. Where you go is who you are: understanding people's activities by places visited Proceedings of the Workshop on Behavior Monitoring and Interpretation BMI 2010.

Wikipedia 2012. http://en.wikipedia.org/wiki/White_Stork, accessed at May 9, 2012