Knowl Inf Syst (2013) 37:331–362
DOI 10.1007/s10115-012-0511-z

REGULAR PAPER

How you move reveals who you are: understanding human behavior by analyzing trajectory data

Chiara Renso · Miriam Baglioni · Jose António F. de Macedo · Roberto Trasarti · Monica Wachowicz

Received: 31 March 2011 / Revised: 4 March 2012 / Accepted: 24 March 2012 /
Published online: 19 June 2012

Abstract The widespread use of mobile devices is producing a huge amount of trajectory data, making the discovery of movement patterns possible, which are crucial for understanding human behavior. Significant advances have been made with regard to knowledge discovery, but the process now needs to be extended bearing in mind the emerging field of behavior informatics. This paper describes the formalization of a semantic-enriched KDD process for supporting meaningful pattern interpretations of human behavior. Our approach is based on the integration of inductive reasoning (movement pattern discovery) and deductive reasoning (human behavior inference). We describe the implemented Athena system, which supports such a process, along with the experimental results on two different application domains related to traffic and recreation management.

Keywords Behavior inference · Trajectory data mining · GPS data · Ontologies · Pattern classification

1 Introduction

Which is the movement behavior expressed by the commuters inside a city? How can we identify an exploring behavior of visitors inside a park? How is it possible to identify tourist behavior among the individuals moving in a city? These are only a few examples of

C. Renso (✉) · R. Trasarti
KDDLab, ISTI-CNR, Pisa, Italy
e-mail: chiara.renso@isti.cnr.it

M. Baglioni
IIT-CNR, Pisa, Italy

J. A. F. de Macedo
Federal University of Ceara', Fortaleza, Brazil

M. Wachowicz
University of New Brunswick, Fredericton, Canada

analytical questions that may arise when there is the need of understanding the mobility behavior of people moving in a geographical context. The widespread use of personal mobile devices enables the collection of huge quantity of positioning data representing people's movements. However, the knowledge gap between the raw trajectories data, representing the raw geographic coordinates as detected by the mobile device and the understanding of human behavior is still huge: there is the need for the analysis of raw trajectories and the extraction of *behavioral patterns*, which can be interpreted to get new knowledge on the mobility behavior.

Inferring behavioral patterns from electronic traces is a topic that has been faced by behavior informatics (BI). The growing interest of this research topic is witnessed by the increasing number of papers in the literature [11,12], where technologies that can support an accurate understanding of human behaviors are developed. The present paper gives a perspective where mobility data are collected and analyzed to extract mobility behavioral patterns, thus giving an interpretation of the patterns mined from data.

Improving the users' interpretation of discovered patterns has been mainly focused in the literature on the semantic annotation for the Web [41,48]. Syntactic patterns approaches have been used to classify patterns based on a set of extracted features. In contrast, approaches to semantic patterns have assumed that a pattern structure is quantifiable and extractable, so that structural similarity of patterns can be assessed [15]. However, syntactic patterns only reports the structural representation of patterns; meanwhile, the latter formulates hierarchical descriptions of complex patterns built up from simpler primitive elements and ontological commitments. In literature, several approaches have been proposed for closing this knowledge gap, such as Internet usage [17,20] and business data analysis [13]. We believe that there is the need of a new approach for dealing with the semantic complexity of movement patterns (i.e., the complexity of existing hierarchical structures of a domain knowledge that are essential for pattern interpretation). In other words, any movement pattern that emerges from trajectory data needs to be classified in light of an ontology in order to make its intrinsic meaning explicit to a user [18].

In this paper, we propose a *Semantic-Enriched Knowledge Discovery Process* for handling the syntactic and semantic complexity of movement patterns in order to support meaningful interpretations of human behavior. The classical knowledge discovery (KDD) process introduced by Fayyad [19] is mainly characterized by inductive reasoning, which begins with gathering data (i.e., facts) that are specific and limited in scope. Then, it proceeds to a generalized conclusion with a certain degree of uncertainty, depending on the accumulated evidences. By gathering data, seeking patterns, and building hypothesis, this process allows us to explain what has been observed, having the ultimate goal of improving human domain knowledge. However, the generalized conclusions are not absolutely certain, even after taking into account any premises on human behavior.

In contrast, the Semantic-Enriched Knowledge Discovery Process makes use of human domain knowledge (i.e., the users' a priori knowledge on human behavior) by developing a *mobility behavior ontology* where trajectory data and movement patterns are to be classified. We also propose the integration of inductive reasoning (pattern discovery) and deductive reasoning (behavior inference) that allows discovered movement patterns to be classified into human behavior. Essentially, this new process consists of a querying and mining process enhanced with reasoning tasks. The novelty of this approach is to allow the integration of mining and reasoning services within a comprehensive process, which provides the mechanism for the semantic annotation of trajectory data and the classification of movement patterns into behavior concepts.

The main research contribution in this paper is the formalization of a semantic-enriched KDD process by showing (i) how to make explicit the syntactic pattern structure of trajectory data; (ii) how to associate a semantic hierarchical structure of a mobility behavior ontology with discovered movement patterns; and finally, (iii) how to integrate inductive and deductive reasoning for discovering new knowledge on human behavior.

Toward the end, the first contribution of this paper is to introduce the Athena framework that incorporates an answer to these research challenges. In fact, the framework proposes a solution to the problem of understanding moving entities behavior based on the analysis of positioning tracks combined with encoded domain knowledge.

We show how this framework, integrating inductive and deductive steps, allows us to find answers to challenging analytical questions about mobility behavior properly analyzing the historical data collected from mobile devices and stored in a central repository.

A second, more practical, contribution of this present work is to introduce the Athena system as an implementation of the proposed framework and a synergic integration of trajectory mining with deductive inference performed by the ontology reasoning engine. Although the single components of the system are not innovative themselves (e.g., the mining algorithms and the ontology reasoning engines are "state of the art" tools), the design of a proper integration and interaction between these tools in a comprehensive running system is a new contribution in this area. We present key design features underlying Athena, emphasizing its combination of mining and deduction. We prove the usefulness of our approach by illustrating two experimental sessions run on GPS datasets expressing people movements in two different settings: movements of cars in an urban area in Italy and visitors of a park in the Netherlands.

The structure of the paper follows. Section 2 describes the related work. Section 3 gives some basic definitions used throughout the paper. Section 4 introduces the semantic-enriched KDD process. Section 5 describes the implemented Athena system, which supports such a process. Section 6 reports the experiments of Athena in two scenarios: traffic and recreation management. The performance results are presented in Sect. 7, and finally, in Sect. 8, we report the conclusions and future works.

2 Related work

Human behavior has been analyzed by a large number of scientists through psychological, social, geographical, and organizational perspectives [1,28]. Our research work was carried out from a multidisciplinary perspective and it is related to the emerging field of behavior informatics, which aims to develop methodologies, techniques, and practical tools for representing, modeling, analyzing, understanding, and utilizing human behavior [12].

Knowledge Discovery on movement data has been one of the most productive research communities, having generated substantial scientific output as seen by the vast amount of algorithms and methods developed in the last decade [23]. The majority of these methods have been focused on mining the geometric properties of a trajectory [34]. Recent research work has been carried out to address the challenges of generating new knowledge in understanding the human mobility behavior. The first steps toward this direction have been taken by enhancing the preprocessing step of a KDD process by using manually and semiautomated techniques for the semantic annotation of trajectory data [10,52]. Stops and moves have been proposed as important structures since any travel consists of stopping in an interesting place, and moving toward and away from this place [46,47,50]. The main difference in our approach compared to these works is that they consider only semantic enrichment of the trajectories as a stand-alone process not connected to the KDD process. In our approach, the construction of the semantic trajectory is just the beginning of the analysis, which includes both the inductive and the deductive steps.

Data mining ontologies have also been proposed in the past for ensuring quick and goal-oriented development [37]. They have been used to facilitate both forms of scientific discovery in providing a common framework for several systems and problem-solving methods. Their implementation has led to the so-called third-generation data mining and knowledge discovery services, which have served a number of different objectives in guiding the knowledge discovery process [4], building of workflows [31], and improving diagnostics [24]. Data mining ontologies have also been proposed as a framework to define the constraints associated with a specific class, which can be applied to filter discovered patterns [8]. In summary, data mining ontologies have been primarily applied at the core of the data mining step of a KDD process with the aim of driving pattern discovery or after the execution of a data mining step in order to support pattern filtering.

In our approach, we do not intend to embed an ontology in a data mining step, but, instead, we are proposing to embed a mobility behavior ontology in a new deductive reasoning step of the proposed semantic-enriched KDD process. The aim is to steer human behavior inference from the discovered movement patterns. The conceptualization of such a semantic-enriched KDD process was first introduced in [6] and a prototype implementation was described in [7]. The main contribution of this paper compared to these ones is the concise formalization of this semantic-enriched KDD process and its robust system implementation as a proof-of-concept.

3 Preliminaries

In this section, we give some basic concepts and preliminary definitions that will be used to formalize the proposed semantic-enriched KDD process. We first introduce the concepts of trajectory and semantic trajectories based on stops and moves as the basic conceptualization step for representing mobility data. Then, we describe the mobility behavior ontology whose aim is the conceptualization of the ground concepts of the domain. The third component is mobility data mining; therefore, the methods used in the semantic-enriched KDD process are introduced in Sect. 3.3.

3.1 Trajectories

A trajectory is the footprint of different positions of a moving object. Therefore, it is typically represented as a sequence of sample points, describing the spatial and temporal positions detected by a tracking device [27]. We distinguish three types of trajectories: *measured*, *syntactic*, and *semantic*, defined as follows:

Definition 1 (*Measured Trajectory*) A *measured (or raw) trajectory* T of an object O is represented as: $T_O =< \mathrm{p}_1 \cdots \mathrm{p}_\mathrm{n} >$, where $p_k = (x_k, y_k, t_k)$, and n are the number of sample points recorded during the movement of the object O. A *measured trajectory* is built from a sample of points recorded by mobile tracking devices, such as GPS, GMS, or WI-FI sensors.

The definitions of stops and moves we used in this paper have been adapted from [9] and [48] as follows:

Definition 2 *(Stops and Moves)* Given a *measured trajectory* $T = \langle \mathrm{p}_1 \cdots \mathrm{p}_\mathrm{n} \rangle$, a spatial threshold $\mathrm{th}_{\mathrm{spatial}}$ and a temporal threshold $\mathrm{th}_{\mathrm{temporal}}$, a *stop* is defined as a maximal subsequence S of T where

$$\begin{aligned} \mathrm{S} &= \{\mathrm{p_m} \cdots \mathrm{p_k}\} |\ 1 < \mathrm{m} \leq \mathrm{k} \leq \mathrm{n} \wedge \mathrm{m} \leq \mathrm{i} \leq \mathrm{k}\ \mathrm{Dist}\left(\mathrm{p_m},\ \mathrm{p_i}\right) \\ &\leq \mathrm{th}_{\mathrm{spatial}} \wedge \mathrm{Dur}\left(\mathrm{p_m},\ \mathrm{p_k}\right) \geq \mathrm{th}_{\mathrm{temporal}} \end{aligned}$$

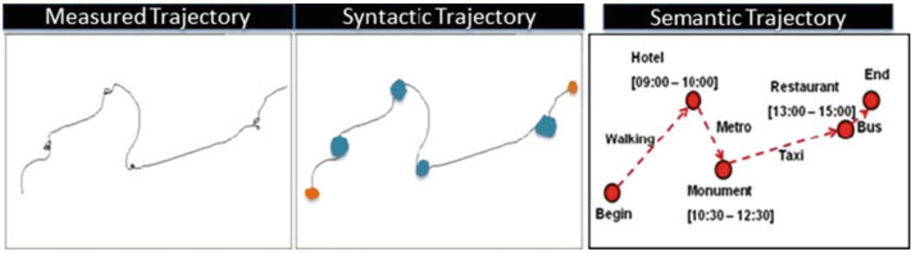


Fig. 1 A measured trajectory and its corresponding syntactic and semantic trajectory

Moreover, a *move* is defined as the maximal subsequence in between two consecutive stops:

$$\mathrm{M} = \{\mathrm{p_a} \cdots \mathrm{p_b}\} | \ \mathrm{S_k} = \{\mathrm{p_m} \cdots \mathrm{p_k}\} \wedge \mathrm{S_{k+1}} = \{\mathrm{p_{m'}} \cdots \mathrm{p_{k'}}\} \wedge \mathrm{a} = \mathrm{k}+1 \wedge \mathrm{b} = \mathrm{m'} - 1$$

Where *Dist* is the Euclidean distance function defined between the spatial coordinates of the points, and *Dur* is the difference in the temporal coordinates of the points.

Definition 3 *(Syntactic Trajectory)* A *syntactic trajectory St* is a feature vector (Tid, Oid, B.E.S, M) where Tid and Oid are positive integers identifying, respectively, the trajectory and its moving object, B.E.S is a set of Begin, End, or Stop (B.E.S.—the begin, end, or stop of a trajectory), and M is a set of moves. A B.E.S is characterized by a time interval representing the duration of the stop, and a spatio-temporal point representing the actual place where the B.E.S happens.

A semantic trajectory [45] represents the movement taken in a specific context, which can be geographical (the physical world where the entity moves) and/or socioeconomical (the domain knowledge which is specific in a human behavior). Context information is the main source of semantics in the semantic enrichment process. The syntactic trajectory annotated with context information is called *semantic trajectory.* A commonly used form of semantic trajectory is the *stop and move* representation:

Definition 5 *(Semantic Trajectory)* Given a syntactic trajectory *Ts* and an application context C, the semantic trajectory is obtained labeling the set of stops $\mathrm{S} = \{\mathrm{p_m} \cdots \mathrm{p_k}\}$ and moves $\mathrm{M} = \{\mathrm{p_a} \cdots \mathrm{p_b}\}$ of *Ts* by using two semantic enrichment functions ω(C,S) and ϕ(C,M).

In the examples presented in the paper, we focus only on stops and we have realized only the function ω(C,S) as the *spatial intersection of the stops* in *Ts* with a set of points of interests (e.g., stations, restaurants, and monuments) representing the spatial context in *C*. An example of the function ϕ(C,M) is a method assigning each move in *Ts* a mean of transportation (e.g., walking, metro, taxi, and cycling).

In Fig. 1, we illustrate a measured trajectory and the corresponding syntactic and semantic trajectories with the stops.

Notice that in this approach to trajectory processing, differently from [39], we do not deal explicitly with uncertainty and imprecision of the data. In fact, our assumption is that the data are already cleaned using some heuristics (e.g., map matching [35] or external information (e.g., the power of signal).

3.2 Mobility behavior ontology

We introduce the basic concepts of the mobility behavior ontology. The aim of this ontology is to represent the concepts and relations of the mobility domain. In fact, the term ontology is defined by Gruber in 1993 as "explicit specifications of conceptualizations" [25]. In other words, ontologies are a conceptual framework to formally model the semantic hierarchical structure of a system, that is, the relevant entities and relations that emerge from the observation of the world, and which are useful to our purposes [26]. Given a conceptualization, a *formal ontology* is a logical theory specifically designed to capture the description of the world corresponding to the conceptualization. The concepts of a domain are formalized by specifying them along with their *subconcepts* (*isa* relation, also called "taxonomic relations" or "hierarchical relations"). They express the "kind of" relationship between two concepts. Other generic relationships between concepts can also be modeled and they are called *properties*.

While the conceptualization specifies which are the concepts of our world and which are the relationships between them, the explicit specification regards the *instances* within the ontology that are the actual elements of the domain. Notice that one instance can belong to more concepts simultaneously. Ontologies may be classified into different types, depending on the way they are used. *Top-level ontologies* provide a broad view of the world suitable for many different target domains. On the contrary, *domain ontologies* model a specific domain, which represents part of the world. The primary purpose of *core ontologies* is to contain only those concepts, which are strictly necessary, such as the *basic categories* of domain knowledge which should be *coherent* with the domain in which they are inserted [49,51]. Finally, *application ontologies* are oriented to specific applications within a domain and are suitable for direct use in reasoning engines or software.

Formal ontologies are described by languages, which are formal and machine readable. They often include reasoning facilities that support the automatic processing of that knowledge. Web ontology language (OWL)[1] is a well-known standard that originated from the semantic web and it is a W3C recommendation. An interesting feature of OWL is that it relies upon a family of languages known as description logics (DL) that provides a deductive inference system based on a formal well-founded semantics [5]. The basic components of DL are suitable to represent concepts (called*classes)*, properties *(roles),* and instances *(individuals)*. Furthermore, complex expressions, called *axioms*, can be used to define implicit new concepts. The reasoning tasks usually provided by the ontology inference engines are *consistency checking* (to check whether the ontology is consistent), *subsumption* (to find whether a concept is subsumed by another one), and *instance checking* (to check what classes a given instance belongs to). OWL comes with some reasoners[2,3,4] that perform sound and complete and terminating decision procedures. OWL has sublanguages of increasingly expressive power from *OWL Lite*, to the most used *OWL DL*, to the computationally expensive *OWL Full*. Other OWL languages have been proposed to find a good trade-off between expressive power and efficiency in reasoning tasks. OWLPRIME[5] is the language offered by the Oracle platform as a subset of OWL DL. It comes with some limitations in expressiveness. To overcome these OWLPRIME limitations, Oracle provides *rules* that allow the ontology engineer to complement the basic OWLPRIME reasoning with more sophisticated and application-dependent inference mechanisms. The rules can be added to a rule base that can be used conjointly with the ontology during the semantic query execution. OWLPRIME plus

[1] OWL: W3C Consortium. The web ontology language. http://www.w3.org/TR/owlfea\discretionary-ture.

[2] **Racer**: The Racer Reasoner: http://www.sts.tu-harburg.de/~r.f.moeller/racer/.

[3] **Pellet**: The Pellet Reasoner http://pellet.owldl.com/.

[4] **FACT** ++: The FACT++ Resoner: http://owl.man.ac.uk/factplusplus/.

[5] OWLPRIME: http://www.w3.org/2007/OWL/wiki/OracleOwlPrime.

the Oracle rules will be exploited for the implementation of the Athena system, as illustrated in Sect. 5.4.2.

In our approach, the mobility behavior ontology represents the domain knowledge where trajectory data and movement patterns are to be interpreted. It is composed of two conceptualization levels, the core and the application:

Core Ontology: This ontology component formalizes the concepts of human behavior, which are independent from a specific application domain. These concepts are trajectory, stop, move, time, and pattern. The core ontology is illustrated in Sect. 5.1.

Application Ontology: This ontology component describes the concepts of human behavior that are of interest within a particular application context. Consider an example in the context of tourism management. In this case, the application ontology explains the movement of tourists in urban cities and can be used to infer human behavior activities such as shopping and visiting. An example of application ontology is depicted in Fig. 4 in Sect. 5.

One of the main issues in using ontologies in large trajectory datasets is the efficiency, since the complexity of the reasoning task is exponential. For this reason, methods to reduce the size of the data not only in terms of number of instances, but also classes, may improve the running time of the reasoning tasks. One method for the instance checking improvement, broadly used in the literature, consists of using a database to store the ontology instances to get advantage of the database power in managing large sets of data. In this way, we combine the advantages of a powerful management system for large datasets, such as a DBMS, with the semantic richness of a structure capable of representing concepts and relationships between them. A mandatory requirement is the creation of a mechanism to link ontologies to the data stored in the relational database. We call this mechanism *ontology-relational data mapping* as stated by the following definition:

Definition 6 *(Ontology-RelationalData Mapping)* An *ontology-RelationalData mapping* is a triple (DO, DB, μ) where:

- DO is a *domain ontology*, which represents the domain knowledge.
- DB is a relational database;
- μ is a set of correspondence assertions, called a *mapping*, each one of the form $A \leftarrow q$, where A is a class or property of DO, q is a relational query over DB database. The relational query q always projects one object identifier, to the mapped concept or property.

For example, let's consider a very simple database DB and the ontology DO. DB contains tables "Hotel" representing standard information such as hotel ID, hotel name, address, and a "TouristAttraction" table representing standard information such as the attraction ID, the location, and the type of attraction (e.g., monument, theater, and museum). The ontology DO expresses the concepts "Hotel" and "Monument." An ontology-relational data mapping between DO and DB maps "Hotel" concept to "Hotel" table and "Monument" concept to a selection query on "Tourist Attraction" table. It is therefore defined as

$$\text{“Hotel}'' \leftarrow \text{“SELECT * FROM Hotel}'';$$
$$\text{“Monument}'' \leftarrow \text{“SELECT * FROM TouristAttraction WHERE type} = \text{“monument}''.$$

The Ontology-RelationalData mapping is the key mechanism to link the inductive and deductive steps in the semantic-enriched KDD process. This is implemented as the *MapFile* component in the Athena architecture introduced in Sect. 5.

3.3 Trajectory data mining

Mobility patterns can be extracted from masses of trajectories (measured or semantic) using data mining techniques [34].

Among the most studied kind of patterns, there are clustering methods. They come from the common need in analyzing large quantities of raw trajectory data splitting the dataset into logically well distinct groups, such that the objects in each group share some property, which does not hold (or holds much less) for other objects. A basic approach to define a similarity between trajectories is to consider *similar* the couples of moving objects that follow approximately the same spatio-temporal trajectory, that is, at each time instant, they are approximately in the same place [42]. A particular case of clustering is *flock*, where the spatio-temporal coincidence of trajectories could happen only for a segment of trajectories. Other kinds of patterns are sequential and frequent patterns—possibly adapted for the case of trajectories—where sequences or sets of objects are found frequent in the dataset (e.g., specific areas visited during the same hours of the day). We integrate into the Athena framework the trajectory clustering algorithm T-Clustering, T-Pattern as a sequential pattern specialized for trajectories, flock, frequent, and sequential pattern. Following, we briefly describe them, leaving the reader with the appropriate references for further details.

T-Clustering. The aim of this algorithm is the extraction of groups of similar trajectories. A group is called T-Cluster and is defined as a set of labeled trajectories, which share the same membership. The trajectories of a T-Cluster are grouped on the basis of their similarity according to a specified similarity function, chosen from a repertoire of possible choices—as presented in [42]—such as route similarity, common start or common end.

T-Pattern. This algorithm extracts a set of patterns $\mathrm{tp} = (\mathrm{R,T,s})$ where $\mathrm{R} = \langle \mathrm{r}_0, \ldots, \mathrm{r}_\mathrm{k} \rangle$ is a sequence of regions, $\mathrm{T} = \langle \mathrm{t}_1, \ldots, \mathrm{t}_\mathrm{k} \rangle$ is a sequence of relative time intervals $\mathrm{t}_\mathrm{j} = [\mathrm{ts}_\mathrm{j},\ \mathrm{te}_\mathrm{j}]$ associated with each region and s is the support of tp, that is, the number of trajectories that are compatible with tp in space and time. Originally introduced in [21], a T-Pattern is a concise description of frequent behaviors, in terms of both space (i.e., the regions of visited space) and time (i.e., the movement duration).

Flock. The Flock algorithm, introduced in [52], extracts a set of flocks defined as $\mathrm{f} = (\mathrm{I,r,b})$, which represents a spatio-temporal coincidence of a group of moving points, where $\mathrm{I} = [\mathrm{t}_{\mathrm{min}},\ \mathrm{t}_{\mathrm{max}}]$ is the time interval of the coincidence, b is the base moving point, and r is the spatial buffer around b which is used to determine the coincidence.

Frequent pattern. This algorithm is a frequent item set data mining algorithm applied to semantic trajectories represented as sets of stops [2,9]. The algorithm extracts the most frequent stops according to a minimum support threshold. The implementation used in this paper is introduced in [2].

Sequential pattern. This algorithm is a standard data mining algorithm aimed at finding frequent sequences of items. Here, the algorithm is applied to semantic trajectories represented as sequences of stops. The set of frequent sequences are extracted according to a minimum support threshold, implementing the algorithm presented in [9].

4 The semantic-enriched KDD process

The KDD research field is one of the most productive areas in the last decade since it has provided a huge amount of algorithms and methods. However, the problem is that the knowledge produced by the KDD process is, generally, not actionable in the sense that it is not really applicable to the business domain. This is mainly due to the lack of semantics of the

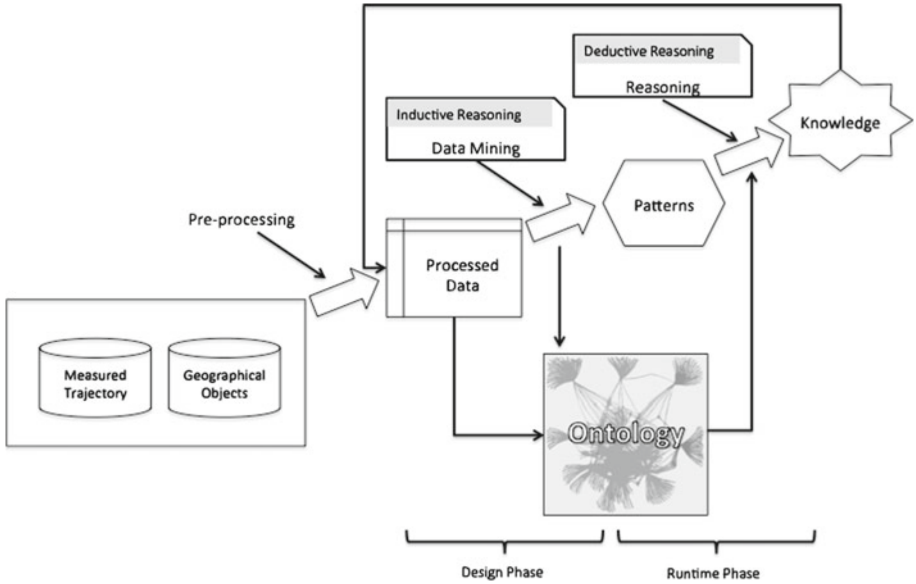


Fig. 2 The conceptual semantic-enriched knowledge discovery process

extracted patterns, which are not focused on the problem at hand. Consider, for example, the extraction of trajectory patterns: most of the methods to extract trajectory patterns are focused on the geometric properties of the trajectory thus discovering geometric trajectory patterns, which might be useless, non-applicable, or not interesting for the final user [10]. A way to close the gap between the "KDD knowledge" and the "business actionable knowledge" is to consider the contextual domain information during the discovery process, thus implicitly adding semantics to the discovered patterns [13,14].

This section introduces the *semantic-enriched KDD process* as an iterative and interactive process involving inductive and deductive reasoning, as illustrated in Fig. 2. The measured trajectories are stored in an **integrated data repository**, which also stores the syntactic trajectories as well as the mined patterns. The first step is the **pre-processing**, where the syntactic trajectories are computed from the given measured trajectories. These are then stored as object-relational tables in the integrated data repository. The **mobility behavior ontology** is designed and the mapping between ontology concepts to data and patterns is defined. The next step runs the **data mining algorithms** on measured, syntactic, or semantic trajectories in order to compute movement patterns. At the reasoning step, the mobility behavior ontology is populated with trajectories and patterns. The ontology inference engine is executed, and the defined axioms are interpreted to classify the ontology instances using the appropriate concepts on human behavior. What we obtain after this step is a classification of the mined movement patterns into a given behavior class. Moreover, following the style of the "inductive database approach" [43], the result of the analysis step can be stored and used as input for a refinement of the previous result in a new iteration of the KDD process.

At this point, it is worth to point out that the semantic-enriched KDD process involves three key users: *the domain expert*, *the ontology builder* , and *the analyst*.

The first two combine their skills to develop both an ontology representing a specific domain and outline the analysis needed—in terms of patterns and behavior definitions—to reach the application objectives: this is called the *conceptual KDD application*. The analyst

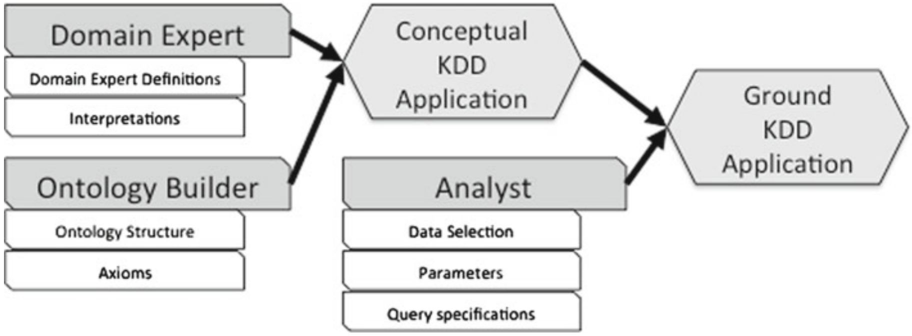


Fig. 3 The three key users of the semantic-enriched KDD process

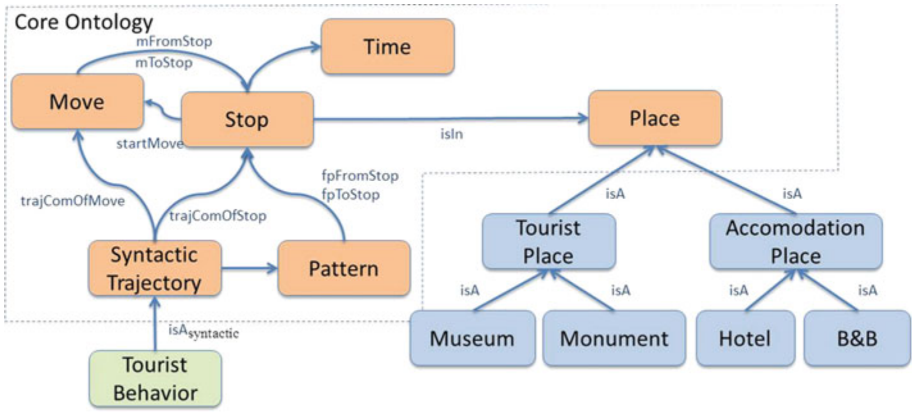


Fig. 4 An example of the mobility behavior ontology for a tourist management domain. The core ontology is indicated by *orange boxes*, while the application ontology is represented by *green* and *blue boxes*. The *green box* indicates the tourist behavior concept defined by axioms (color figure online)

is the user who actually uses the implemented system, realizing the *ground* (or *actual) KDD process* based on the requirements of the domain expert.

In more detail, the analyst manages the data of the KDD application, defining the ontology-relational data mapping and running the appropriate mining algorithms to reach the analysis objectives. This user has the knowledge of the data mining tasks, the preprocessing techniques, the data management languages, and any detail about the data to be mined. Figure 3 illustrates the roles of the key users.

During the *design phase*, the mobility behavior ontology is prepared and the ontology-relational data mapping is defined. During the *runtime phase*, the following steps are performed:

1. **Data pre-processing**: during this step measured trajectories are transformed into syntactic (and semantic) trajectories.
2. **Ontology mapping**: trajectories and geographical objects mapped to the ontology are automatically imported into the mobility behavior ontology as concept instances, using the Ontology-Relational data Mapping.
3. **The data mining step**: this inductive step mines the data stored in the database and generates new movement patterns extracted from data.

4. **The reasoning step:** the result of this deductive step is a new classification of the instances (either patterns or trajectories) into human behavior; the new generated knowledge is stored into the repository.

The integration between the inductive/deductive reasoning is materialized with the sequence of mining and reasoning steps, and a new mining task is possibly executed on the discovered knowledge, starting a new inductive–deductive reasoning cycle. It is worth recalling that both the phases design and runtime are to be executed off-line when it comes to a possible application that exploits the extracted knowledge.

Let's consider an example of the semantic-enriched KDD process for understanding *home-work routine behavior*. In this case, the mobility behavior ontology developed during the off-line phase may include concepts such as trajectory, stop, move, place, home, work, and patterns. Relationships between these concepts may include, for example, "*Home is a place*" or "*Trajectory is composed of stop*." The mobility behavior ontology also includes axioms defining some complex concepts. For example, *"Home is a place where a user frequently stops during the night"*. Then, starting from measured trajectories collected from people moving in a city and stored in the repository, we compute the syntactic trajectories as a sequence of stops and moves; then, we associate the hierarchical structure of the mobility behavior ontology to data; thus, for example, the "Stop" concept is mapped to the "stop" table in the database through the ontology-relational data mapping. From there, the online phase starts and the stops table is mined with a frequent mining algorithm to find the most frequent stops: this is the criterion used to find the home of users as defined in the *Home* axiom. This step also imports patterns (e.g., frequent stops) to the mobility behavior ontology. Finally, the reasoning engine runs to assign to *Home* all the places where users frequently stop during the night and to *Home-Work* the trajectories having both home and work stops.

It is worth noticing that, although the semantic-enriched KDD process includes several tasks, it is not mandatory to run all these tasks to infer human behavior from trajectory data. Fewer tasks can sometimes be sufficient to discover new knowledge. For example, a user can be only interested in a single trajectory behavior such as individual tourist behavior: in this case, the data mining step might not be required. Similarly, a user might be interested in finding patterns using syntactic trajectories. These examples will be further explained in the experiments (Sect. 6).

5 Inferring behavior from trajectory data: The Athena tool

The semantic-enriched knowledge discovery process has been implemented into the system Athena we are describing in this section.

The Athena tool implementation finds its roots in the research carried on during the GeoPKDD project,[6] during which preliminary prototypes of Athena were developed: *Daedalus* [38] was a first data mining query language prototype system embedding trajectory mining algorithms into a unified architecture. This tool evolved into the *GeoPKDD system* [36], the fully developed prototype of a data mining query language enriched with reasoning capabilities. From this system, two development directions took place. The first direction led to the Athena system, described in this paper, which focused on the behavior informatics aspects of KDD. The second direction led to the M-Atlas system, which was more oriented in the improvement of the data mining tools [22] and enrichment of the query language without

6 GeoPKDD Project. Geographic Privacy-aware Knowledge Discovery and Delivery. http://www.geopkdd.eu/.

any automatic deductive capabilities. M-Atlas does not provide reasoning capabilities and it is more oriented to the mining step. On the other hand, Athena provides reasoning capabilities at the price of reduced data mining primitives.

The Athena architecture includes the conceptual components of the semantic-enriched KDD process, namely the mobility behavior ontology, the data mining algorithms, and the integrated data repository. As already pointed out, the first component represents the main concepts in the mobility behavior domain and it is equipped with a reasoning engine for behavior inference. The second component implements the trajectory mining algorithms. The third component stores the input and output of each step of the semantic-enriched KDD process, namely trajectories, patterns, and mobility behavior ontology. We are now introducing the details on how these components are realized in Athena.

5.1 The mobility behavior ontology

As previously introduced, the mobility behavior ontology is built by the ontology engineer supported by the domain expert. This ontology represents the concepts of the behavior domain that are used to classify trajectories and patterns into the appropriate human behavior. To clarify the role of the ontology, let us consider again an example in the tourism management domain. In this case, the mobility behavior ontology explains the movement of tourists in the urban context. The syntactic trajectories are represented in the core ontology by the concepts trajectory, stop, and move and the relationships between them (e.g., `Trajectory trajCompOfStop Stop` is the OWL statement stating that a trajectory has a relation "*composed of*" with the stop concept). The context of the syntactic trajectory consists of possible destinations that are represented at the application ontology level by the concepts `Museum` and `Monument`—subclasses of `Tourist Place` and representing the tourist places—`Hotel` and `B&B`—subclasses of `Accomodation Place` and representing the accommodation places. Both are subconcepts of the general concept `Place` representing the possible points of interest in a city.

The mobility behavior ontology links concepts of the core ontology to the application ontology. For example, the concept `TouristBehavior` specializes the concept `Trajectory` by defining an axiom stating that a tourist behavior is represented by any trajectory stopping in an `AccomodationPlace` (e.g., an hotel) and in a `TouristPlace` (e.g., a museum).

The `TouristBehavior` concept can thus be defined by the following axiom (using the OWL syntax):

$$\begin{aligned}\texttt{TouristBehavior} \equiv\ &(\texttt{trajCompOfStop some AccomodationPlace})\ \texttt{and}\\ &(\texttt{trajCompOfStop some TouristPlace})\end{aligned}$$

where `trajCompOfStop` is a relationship between the `Trajectory` and the `Stop` classes indicating that the tourist trajectory is composed by a number of stops. The `some` OWL operator corresponds to an "exist" logical operator. Therefore, a tourist behavior is a trajectory composed of stops that are accommodation places and stops that are tourist places. The reasoning engine populates the `TouristBehavior` class with all the trajectories instances satisfying the axiom.

5.2 The data mining tools

The Athena system integrates a number of trajectory mining algorithms, as anticipated in Sect. 3.3. Athena allows analyst users to easily select a data mining algorithm through the use of a data mining query language [22,36] where a SQL query expresses the call to the desired algorithm along with the needed parameters. In the following, we show an example of the use of the Athena query language to call a data mining algorithm for the case of clustering. We use an implementation of DBScan for trajectories called optics [3]. In the following, we show an example of SQL-like query generated by the system calling the *T-Optics* algorithm on a set of measured trajectories called *traj_data*:

```
CREATE MODEL clustering AS MINE T-OPTICS
FROM (SELECT id, traj FROM traj_data)
SET T-OPTICS.METHOD = ROUTE_SIMILARITY AND T-OPTICS.EPS = 250m
```

Here, the CREATE MODEL statement indicates the data mining method (T-Optics), while the FROM clause selects the dataset to be mined (here the measured trajectory data table). The parameters of the mining algorithm are METHOD and EPS and they specify the similarity function used to group the trajectories and the maximum distance threshold—in meters—between two trajectories to be considered similar (the definitions of the distance function can be found, for example, in [22,47]). The patterns resulting from this query are stored back in the data storage.

5.3 The data storage

The Athena data storage is the integrated repository at the basis of the semantic-enriched KDD process. Particularly, the data storage component must support the representation and handling of measured/syntactic/semantic trajectories, mobility behavior ontology, and mined movement patterns. Furthermore, this component should be easily connected to both an ontology inference engine and the mining algorithms. To achieve that, the Athena implementation has relied on Oracle DBMS.[7] The advantages of this choice are manifold. First, Oracle provides support for object-relational as well as spatial data needed to manage domain knowledge objects such as buildings and roads (Oracle Spatial Data Cartridge additional component).[8] The second advantage is the temporal data cartridge [40] to handle the temporal information needed to deal with trajectories. Again, Oracle permits the storage of the mined movement patterns via the patterns cartridge [38]. Finally, Oracle supports the representation of ontologies via the semantic technologies cartridge[9] providing an embedded reasoning engine based on the OWLPRIME formalism. It is important to point out that Oracle is optimized to store and handle huge datasets, which is a crucial issue when handling trajectory data.

The Athena data storage architecture is illustrated in Fig. 5. The basic object-relational repository of Oracle is enriched with several components. Here, we see that the combination of the spatial and temporal cartridges allows defining a trajectory data cartridge [40], which in turn defines the (measured) trajectory as a native object (namely a *data type*) in the storage system. The cartridge patterns define the data types for the mined movement patterns integrated in the system, such as T-Clustering [42], T-Patterns [21], flocks [52], frequent patterns, and sequential patterns [2,9] The semantic technology component allows the representation, storage, and manipulation of ontologies.

[7] Oracle DBMS: http://www.oracle.com/us/products/database/index.html.

[8] Oracle Spatial http://www.oracle.com/it/products/database/options/spatial/index.html.

[9] Oracle: Oracle Semantic Technologies: http://www.oracle.com/technology/tech/semantic_technologies/index.html.

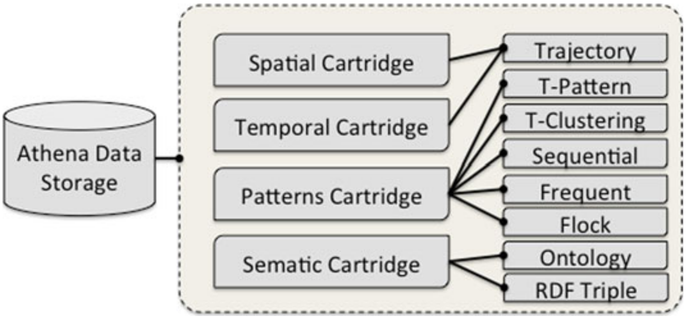


Fig. 5 The Athena data storage, besides storing object-relational data, provides plug-in cartridges for storing spatial and spatio-temporal data, discovered patterns, and ontologies

TID	SID	StartTime	Duration (sec)	Geometry	Semantic
1	1	12:55:00	540		Station
1	2	13:20:10	380		--
1	3	19:07:00	1870		Park

Fig. 6 An example of the stops table (*black columns*) with the semantic information attached during the enrichment operation (*gray column*)

5.4 The Semantic-enriched KDD process in Athena

Given the basic components illustrated in the previous section, we now show how Athena handles each step of the semantic-enriched KDD process: from the trajectory preprocessing, to the mobility behavior ontology definition and database mapping, to the data mining inductive and reasoning tasks.

5.4.1 The data preprocessing step

As described in Sect. 3.1, the preprocessing step is composed of two operations: (i) the geometrical transformation applied on measured trajectories in order to build the syntactic ones and (ii) the semantic enrichment to build the semantic trajectories.
From measured to syntactic trajectories. The stop computation procedure analyses the measured trajectories sample points in order to find maximal segments of trajectories where the object is motionless. An appropriate threshold setting depends on the application and can guarantee the usefulness of all the extracted stops with respect to the analysis requirements. For example, a shot stop of few minutes may represent a traffic light, interesting for a traffic management application, but not in a tourist context.

The syntactic trajectories, represented as sequences of stops and moves are stored in the data storage as a relational database table. Figure 6 shows an example of the stop table, where TID indicates the trajectory identifier, SID the stop identifier, StartTime the timestamp of the beginning of the stop, and finally, the duration in seconds.

After the stops have been computed, the moves are implicitly represented as a relationship between two consecutive stops. Therefore, in the example, a move is created connecting SID 1 and SID 2, another move is defined connecting SID 2 and SID 3, and so on.

From syntactic to semantic trajectories. This step uses the application information contained in the *MapFile.* Consider the following MapFile example:

```
Monument = Select ID from Buildings where type like 'Monument'
Hotel = Select ID from Hotel
Museum = Select ID from Buildings where type like 'Museum'
...
```

The first row maps the ontology concept *Monument* to a specific query on the table *Buildings*. Assuming that *Buildings* has an attribute "type" describing the kind of building, this query selects the building of type "Monument." Similar is the case of museums, whereas the hotels instances are selected as all the records of the *Hotel* table.

Using this information, the syntactic trajectory stops are mapped to the selected places. This is computed by overlapping the stops geometries with all the geometries of all the interesting places in the geographical area of interest (hotel, restaurants, museums, etc.). An example of semantic trajectory is presented in Fig. 6 where the semantic information is attached to the syntactic trajectory.

5.4.2 *Mobility behavior ontology mapping step*

This step imports the ontology, defined during the design phase, and the entities mapped by the MapFile into the integrated repository. The ontology is stored into the Oracle database through the semantic technologies cartridge. The ontology representation format is OWL-PRIME, a subset of OWL DL based on resource description framework[10] (RDF) syntax. OWLPRIME has the advantage of providing the efficiency of *instance checking* over large datasets at the price of a reduction of the logical operators allowed in the language. Therefore, this language results to be less expressive compared to the more common OWL DL. One of the missing logical operators is *intersection,* but Oracle overcomes this limitation by providing a mechanism to express *user-defined rules*. An example of a user-defined rule is the following:

```
#User_rule: '(?x rdf:type :Cultural) (?x rdf:type :HighSpending)',
('?x rdf:type :HighSpendingTourist)'
```

This is an "IF-THEN" rule and states that "if x is an instance of class `Cultural` and x is an instance of class `HighSpending`, then x is also an instance of the class `HighSpendingTourist"`. A user-defined rule has to be added to the ontology whenever an intersection operator is needed to define an ontology concept through an OWLPRIME axiom.

Once the ontology has been built with any external editor (such as Protégé),[11], the Athena system uses the following modules to import the ontology, the associated user-defined rules, and the instances:

Ontology importer. This module imports the ontology file as exported by the editor into the Athena repository.
User rules importer. This module imports the additional user-defined rules when present.

10 Resource Description Framework RDF: http://www.w3.org/RDF/.

11 Protege: Protégé-OWL editor http://protege.stanford.edu/overview/protege-owl.html.

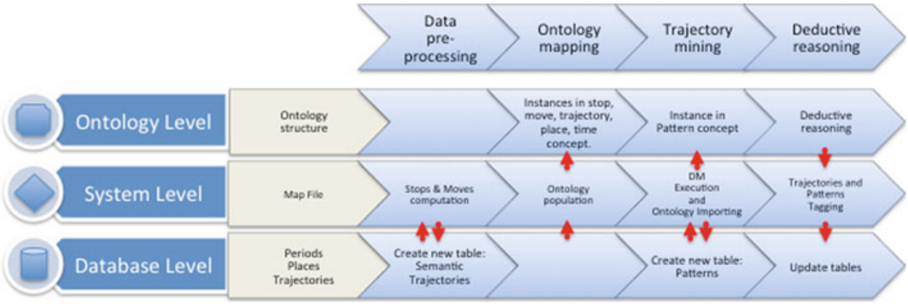


Fig. 7 The process flow of the Athena system considering three different levels: the ontology level, the system level, and the database level

Basic instances importer: This module first imports the basic information (i.e., places and periods) as instances of the ontology. Then, it imports the syntactic trajectories as set of stops and moves, and the semantic trajectories with the relationship between stops and places.

5.4.3 The trajectory mining step

This component accomplishes the inductive mining step. Since the data repository stores both measured and syntactic trajectories, we have integrated algorithms that run on both kinds of trajectories. In particular, frequent patterns and sequential patterns [2,9] can be directly applied to the syntactic trajectories. Frequent patterns can be computed by grouping stops appearing more frequently in trajectories, whereas sequential pattern extracts the most frequent sequences of stops.

On the other hand, other classes of mining algorithms run on measured trajectories. As said before, we have integrated T-Pattern [21], flock [52], and T-Clustering [42] algorithms. All the discovered movement patterns are stored into the data repository and imported into the ontology.

5.4.4 The deductive reasoning step

In this step, both trajectories and patterns are stored in the ontology representation format, which is RDF. The ontology reasoner exploits the axiom definitions and the user rules to classify syntactic trajectories and/or movement patterns into the appropriate behavior classes. This is the last enrichment step where syntactic trajectories and patterns are semantically associated with their appropriate behavior described in the ontology. Given the trajectory data, Athena returns one or more trajectory/pattern *behavior class* (i.e., ontology concepts) that semantically represents the *meaning* of such trajectory/pattern in terms of movement behavior.

5.4.5 Running example

Figure 7 shows a schema of the semantic-enriched KDD process performed by Athena. In this schema, we consider as an example a database with the following tables: periods, places, and trajectories. The periods table contains the temporal discretization such as durations (e.g., long, medium, or short) or absolute intervals (e.g. morning, afternoon, or evening).

The places table contains the geographical object of interest for the application such as buildings, roads, points of interest (POIs), while the trajectories table contains the trajectories measures. The map file, which contains the queries to retrieve this information from the database, expresses the links between tables and core concepts in the ontology. The Athena system starts computing from the measured trajectories the syntactic ones through the stops and moves computation. The result is stored in the table SyntacticTrajectories. Moreover, during this step, a spatial intersection between the geometry stops and the geographical object geometry is performed in order to obtain the semantic trajectories.

In the ontology mapping step, the system populates the ontology, translating each entity (i.e., syntactic trajectories, semantic trajectories, places, and period) into a RDF triplet, which is the basic formalism to represent ontologies in Oracle. The geometric relations computed in the previous step are exploited to populate the *isIn* property between stop and place instances. The trajectory mining step is the data mining algorithm execution, which applies the mining algorithm to trajectories stored in the data repository. This step populates the patterns table. These patterns are then translated into RDF instances to populate core ontology classes. Additionally, the ontology relationships between the trajectories and patterns are populated.

After the mobility behavior ontology is populated, the reasoning phase executes the OWL-PRIME reasoner. This phase results in the classification of the trajectory data and patterns into the new inferred behavior classes.

In this section, we have firstly introduced how the basic components are implemented in Athena, namely (a) the mobility behavior ontology, (b) the data mining tool to infer patterns from data, and (c) the integrated data repository to store trajectories (measured, syntactic, semantic), extracted patterns, the mobility behavior ontology and the inferred behavior. These components are exploited in the semantic-enriched KDD process illustrated by 3 main steps (1) data preprocessing when syntactic and semantic trajectories are built, (2) the mining step where patters are extracted from data, and (3) the deductive step when behaviors are inferred from patterns and trajectories. Finally, we have illustrated the process with the schema of a running example.

6 Experiments

Two different application domains in the context of behavior inference from movement data have been selected for the evaluation of the Athena system. These distinct contexts show that the Athena framework is general enough to cope with different application domain.

In both the experiments described below, we have designed the core and application ontologies using Protégé-OWL. Then, the ontologies have been exported to the format accepted by Oracle (N-TRIPLE) and finally imported to the Oracle semantic data repository.

The case studies presented here have the objective of finding movement behavior in two application domains: traffic management and recreation behavior, as described below.

6.1 Traffic management application

The application requirements and behavior knowledge concepts for this scenario have been obtained from interviews carried out with the Milan Mobility Agency, in Italy. This organization is responsible for providing safe, effective, and efficient traffic control along city roadways while promoting alternative modes of transportation. The agency also promotes and develops safety programs, monitors traffic regulations, and reviews collision statistics.

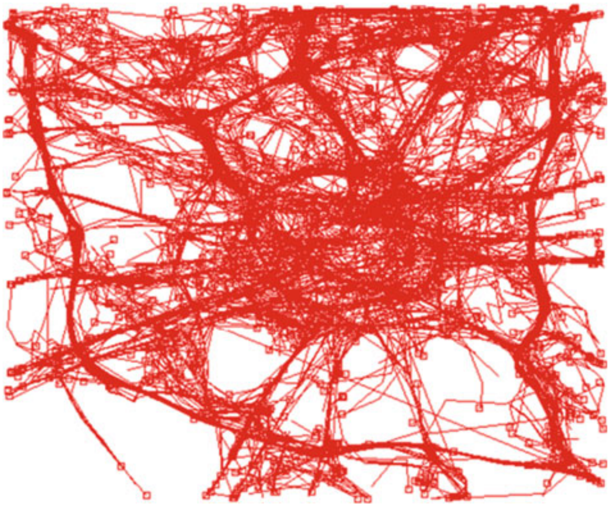


Fig. 8 A visualization of a sample of collected GPS trajectories in Milan

One of their main interests is to understand *home–workplace routine behavior* in order to efficiently manage road closures and detours and to improve public transportation.

Therefore, our application requirement is to infer *home–workplace routine behavior* in the trips of individuals. Currently, traffic managers use questionnaires manually filled in by urban commuters. This procedure has obvious drawbacks: the number of urban commuters actually filling in the questionnaires is quite low in comparison with the number of them daily driving in the city. Furthermore, the questionnaires are filled in once every 5–10 years, thus this information becoming outdated very soon.

The trajectory data for this experiment were obtained from 17,000 private vehicles with on-board GPS receivers under a specific car insurance contract. These vehicles were tracked during a week in the urban area in the city of Milan. The dataset contains more than 2 million observations yielding more than 200,000 trajectories.[12] A sample extracted from this dataset is visualized in Fig. 8.

Data Preprocessing. The first step of the data preprocessing is the extraction of the stops as defined in Sect. 3.1. In this case due the size of the data we have applied, an approximation defining the stop as: *a portion of a trajectory that stays inside a grid cell for at least a given amount of time* [1]. The obvious limitation of this definition for stop, compared to the one given in Sect. 3 is that a vehicle moving inside a cell may be wrongly considered as a stop. But, due to the huge size of the measured trajectory analyzed the stop computation may be exceptionally heavy. Therefore, this heuristic allows a faster computation, although introducing some degree of uncertainty. Performance results on stop computation are reported in Sect. 7. Moreover for the specific question of this case of study—finding the home-work behavior—the places of interests can be any place—from a home to a work place. For this reason, the places of interest are represented as a partition of the Milano area in cells defined by a 100×100 spatial grid, thus defining cells of 200 square meters. Each of these cells represents a place stored in the database that is linked, through the Map File, to the Place class in the mobility behavior ontology.

[12] This dataset has been donated by Octotelematics for its use within the GeoPKDD project (http://www.geopkdd.eu.).

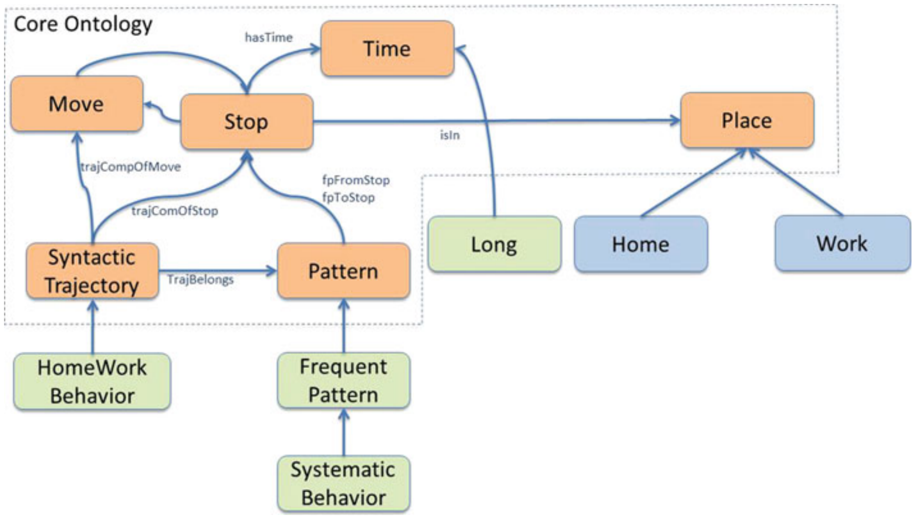


Fig. 9 The mobility behavior ontology for the traffic management experiment

Ontology. In Fig. 9, we show the mobility behavior ontology used for this experiment. The application ontology consists of six classes:

- **Home:** is defined as the place (grid cell) where a vehicle frequently starts its trajectory in the morning.
- **Workplace**: is defined as the place (grid cell) where a vehicle often arrives—at the end of a move—and stops there for a *long* period of time.
- **Long**: A temporal discretization that identifies a stop with a long period of stay. Here long is 4 h or more.
- **Frequent Pattern:** consists of frequent movement patterns computed on all the syntactic trajectories of a single vehicle. This is computed by a frequent pattern algorithm.
- **HomeWork behavior**: this is a syntactic trajectory belonging to a frequent movement pattern beginning at home and ending at a workplace, or beginning at a workplace and ending at home. This definition of a Home-Workplace trajectory has been driven by the interviews with the Mobility Agency.
- **Systematic Behavior**: this behavior is defined as a frequent pattern of the daily trips taken by urban commuters.

In more details, the **HomeWork Behavior** is defined through the following OWL[13] axiom:

```
HomeWork Behavior ≡ (TrajBelongs some
          (systematic Behavior and
                (fpFromStop some (Stop isIn Home)) and
                (fpToStop some (Stop and ((isIn some Workplace) and
                (hasTime some Long))))
        and (trajCompOfMove some (toStop some Home))
```

13 For the sake of readability, we adopted the OWL-DL version of the axiom. OWLPRIME results more complex for the need to encode intersection into user-defined rules (see Sect. 5.4.2)

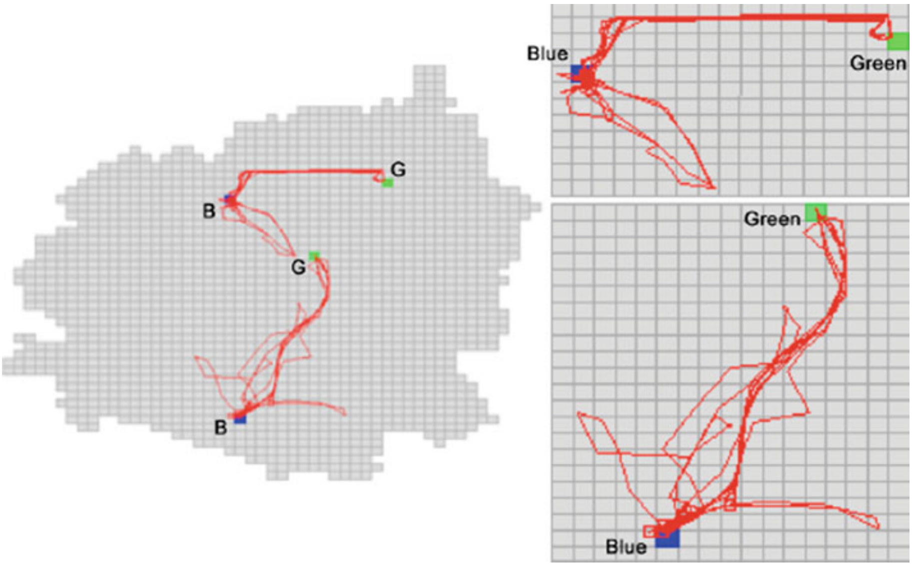


Fig. 10 Two examples of home-work trajectories with their places considered as home (*blue*) and work (*green*) (color figure online)

This means that each syntactic trajectory which belong to a systematic behavior beginning at home, ending at a workplace, and having a last move coming back home is classified as a HomeWork behavior. However, in the literature, there are other definitions such as the one in [44] where *home* is considered the most frequent location visited by a person and *work* is the second most visited location. In case a difference definition is needed, only this axiom needs to be rewritten, while the rest of the process remains untouched.

Trajectory Mining. In the data mining step, the frequent patterns algorithm was run on the syntactic trajectories of each vehicle, the results stored in the repository and then imported into the ontology class frequent pattern.

Deductive Reasoning. After the frequent patterns have been discovered and the core ontology has been populated with the syntactic trajectories and the frequent patterns, the reasoning step runs the ontology reasoner in order to classify these trajectories and patterns into the home-workplace behavior and systematic behavior classes. Figure 10 visualizes the resulting behavior classification of two vehicles. The blue cells on the left represent the homes while the green cells on the right represent the workplaces.

It is interesting to notice that the knowledge discovery process may iterate the steps analyzing the home-work trajectories to scrutinize the movement habits of the home–workplace trips, thus allowing a new human behavior inference. For example, we can identify the individuals that commute during the day or during the night or toward given city areas. Furthermore, Fig. 11 shows the distribution of all syntactic trajectories (green) and the home-work behaviors (dark red) by highlighting the more prominent peeks of the home–workplace trajectories during the rush hours compared to the all-purpose movement of the syntactic trajectories.

It is worth noticing that the data analysis performed by the Athena system on the measured trajectory data is clearly faster and cheaper than the manual questionnaire usually filled in by urban commuters. Furthermore, Athena gives a good degree of flexibility to the traffic manager in changing the different parameters of the analysis, such as the period

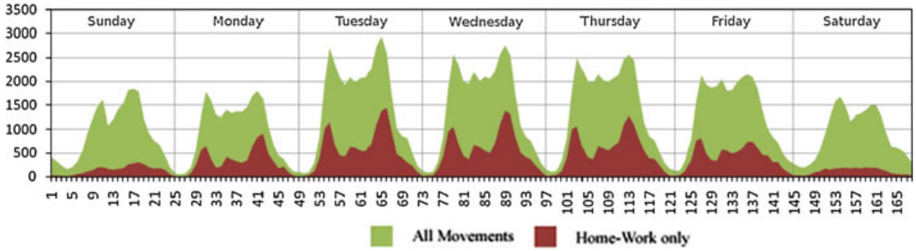


Fig. 11 The distribution of the home-workplace trajectories during the week

of the analysis, the spatial granularity (i.e., using larger urban districts instead of grid cells), the re-definition of concepts (e.g., the home-work behavior) and the possibility to define new movement behavior classes. Further details on this experiment can be found in [36].

6.2 Recreation management application

In this application, the recreation managers are interested in understanding visitor behavior to create and deliver recreation and fitness programs in a variety of settings in the Dwingelderveld National Park, in the Netherlands. They are particularly interested in understanding exploring, socializing, and disturbing types of behavior of different types of visitors in this park [32].

The data to be analyzed were obtained from three different information sources. The first source was a questionnaire containing records about visitor characteristics, preferences, and motivations for visiting the national park. The questionnaire was manually filled in by all visitors in the experiment. The raw trajectory data have been collected by the visitors carrying a GPS receiver during their movement in the park. And, finally, the park map containing the path network and access points (i.e., parking lots) of the national park. This experiment was carried out during 7 days (weekend and weekdays) in spring and summer 2006 for a total of 461 visitors [32]. A visualization of the trajectories is depicted in Fig. 12.

Data Preprocessing. During the data preprocessing step, the stops and moves were computed. The spatial and temporal thresholds used to identify the stops were 10 minutes and 20 meters, respectively. Notice that values of these thresholds are quite low since the visitors were predominantly walking in the park. Furthermore, we extracted the interesting places (e.g., radio telescope, and café) from a questionnaire associated with the dataset and filled in by the visitors.

Ontology. The mobility behavior ontology used in this experiment is shown in Fig. 13. The application ontology consists of nine concepts:

- **Interesting Places:** the places in the park where a visitor usually stops for recreational activity - such as eating and bird watching. Some examples include the Café or the Radio telescope located in the Dwingelderveld National Park.
- **Forbidden Areas:** the areas where a visitor should not stop at any time during his/her visit to the park to avoid disturbing animals.
- **Intersection Path:** the path intersections, where a visitor stops for orientation purposes.
- **Long**: The period of time which identifies a stop with a long period of stay.
- **Flock Pattern:** consists of flock movement patterns computed on all the syntactic trajectories.

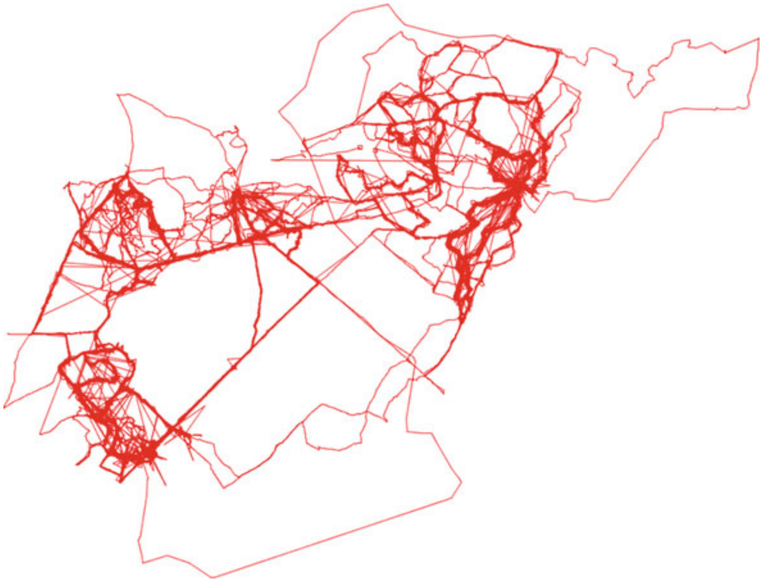


Fig. 12 The trajectories of visitors in the Dwingelderveld National Park

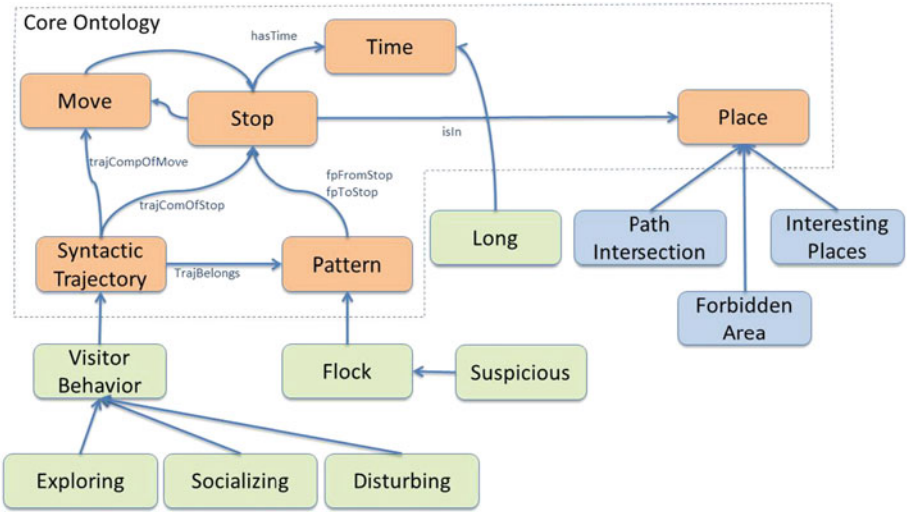


Fig. 13 A fragment of the behavior ontology used for the Park experiment

- **Visitor Behavior**: is a semantic trajectory where the stops occur at the predefined places such as interesting places, path intersections, and forbidden areas.
- **Exploring Behavior**: is the movement of visitors in the park carefully exploring the features of the park: here is defined as stopping at the path intersection and stopping at interesting places.
- **Socializing Behavior**: is defined as visitors encountering and staying together for a given period of time.

- **Disturbing Behavior**: a group of people who have stopped in a forbidden zone of the park for a given period of time.
- **Suspicious Behavior**: any individual belonging to a disturbing group.

The exploring behavior has been modeled in terms of having at least one stop in a path intersection place for a long period of time and after having other stops located at the interesting places. The exploring behavior is encoded in the following OWL axiom:

```
Exploring Behavior = (trajCompOfStop some
    (VisitorTrajectory and
        (Stop and ((isIn some PathIntersection) and (hasTime some
        Long)))) and (trajCompOfStop some(Stop and (isIn some
        InterestingPlaces)))
```

Since the axiom exploits the "*and*" operator, we used the Oracle user-defined rule described below:

```
((?x :trajCompOfStop ?y) (?y :hasTime ?t)(?t :rdfType :Long) (?y :isIn ?p) (?p
:rdfType :PathIntersection)(?x :trajCompOfStop ?z) (?z :isIn ?p) (?p :rdfType
:InterestingPoints), NULL, (?x :rdfType :Exploring))
```

The *Socializing* behavior has been modeled by exploiting the flock pattern algorithm: this algorithm extracts a minimum number of visitor trajectories moving together for a given period of time [29,30,52]. Therefore, the *Socializing* behavior can be defined by the following axiom:

```
Socializing = trajBelongssomeFlock
```

A further interesting feature of the mobility behavior ontology is the possibility of defining new axioms by exploiting already defined axioms and concepts, thus expressing a "complex" behavior. This is the case of *Suspicious* behavior defined as a trajectory belonging to a *Disturbing* behavior, which in turn use the flock pattern to model a group of visitors staying together and stopping in a forbidden area.

```
Suspicious Behavior = trajBelongs some Disturbing
```

and `Disturbing` is defined as

```
Disturbing = Flock and fpFromStop some (Stop isIn ForbiddenArea)
```

It is worth noticing that the disturbing or suspicious axioms—as any other axiom in these experiments—do not represent the *intention* of the visitors to be disturbing. Indeed, it may include visitors that simply got lost in a forbidden area. However, the point is that the behavior expressed by the movement of the person has to be considered disturbing according to the expert definitions.

Trajectory Mining. During this step, we run the flock algorithm [52] on the measured trajectories to find clusters of people moving closely at the same time.

Figure 14 shows the visitors trajectories (dark gray) belonging to the discovered flock patterns where the stops are depicted by one of the following:

- Interesting places: such as *camp* (green), Radio telescope (purple) or cafe (yellow)
- Path intersections (blue)
- Forbidden areas (red)

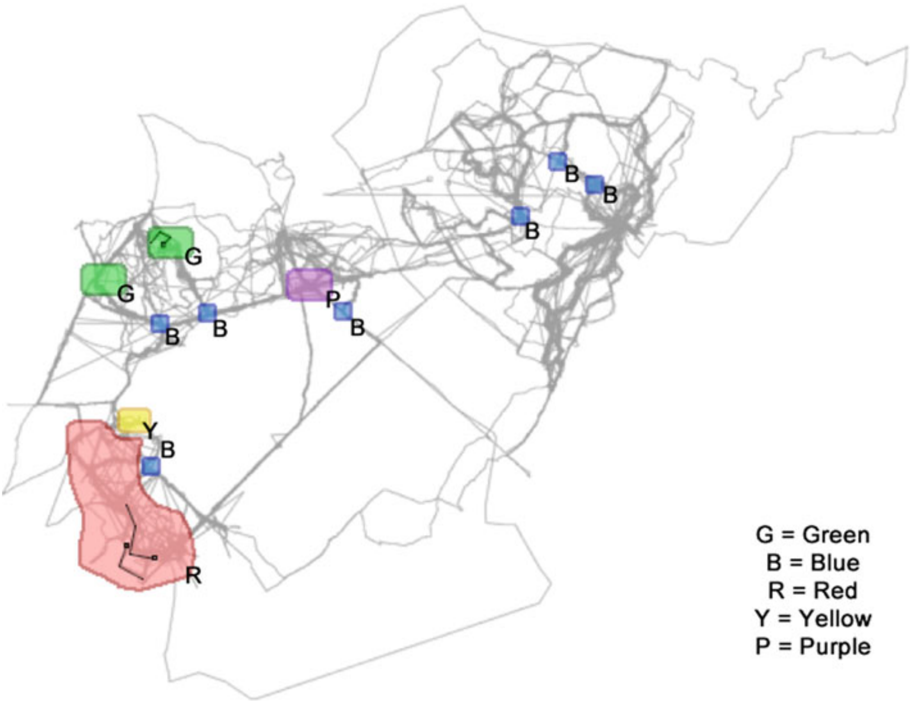


Fig. 14 The visitor trajectories that belong to the discovered flock patterns. *Colored areas* represent the camp (*green*), the Radio telescope (*purple*), the cafe (*yellow*), the path intersections (*blue*), and the forbidden areas (*red*) (color figure online)

Fig. 15 Two types of inferred human behavior: disturbing (*left*) and exploring (*right*)

Deductive Reasoning. The Athena system imports all the patterns into the mobility behavior ontology and then the reasoning engine is run to obtain the classification of visitors' behavior based on the ontology axioms. Once the reasoning step is completed, the Athena graphical interface visualizes the different types of trajectories. For example, the user can visualize the visitors' trajectories belonging to disturbing and exploring behaviors as shown Fig. 15. The left part depicts the disturbing behavior, whereas the right part shows the exploring behavior.

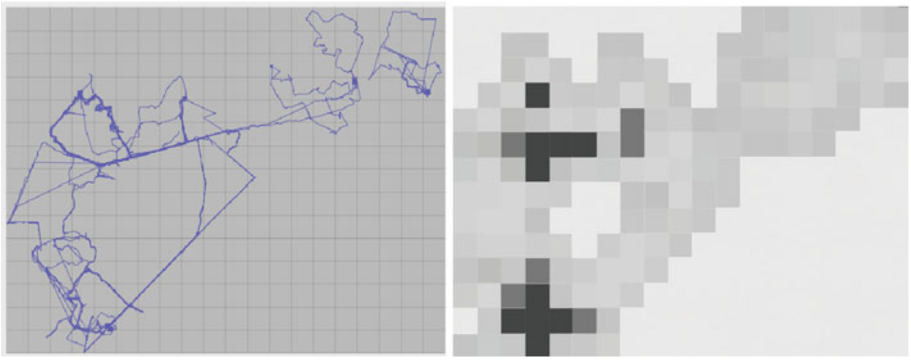
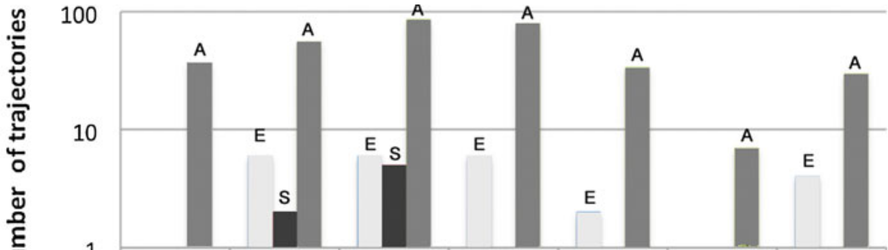


Fig. 16 The exploring behavior trajectories (*left*) and their spatial distribution (*right*). *Darker regions* represent the denser areas



	19/05/2010	26/05/2010	29/05/2010	07/08/2010	10/08/2010	19/08/2010	20/08/2010
Explorers	1	6	6	6	2	1	4
Suspicious	0	2	5	0	0	0	0
All	37	56	87	80	34	7	30

Fig. 17 The daily distribution of visitor trajectories classified as exploring and suspicious behavior

We can notice how the trajectories expressing the exploring behavior tend to be sparser and reach further areas of the park. To the contrary, the disturbing trajectories tend to stay around the forbidden area.

Given these two types of behavior inferred on trajectories, we can iterate the analysis—similarly to the previous experiment. An example of analysis is to study the space distribution of human behavior in the park using a spatial grid. Figure 16 illustrates the spatial grid associated with the spatial distribution of the visitor trajectories that belong to *Exploring* behavior.

We can notice how the exploring behavior, although being sparse, is denser in the west part of the park. This is probably due to the fact that most of the interesting places are found in that area, thus inducing visitors to carefully explore all of them. In particular, exploring visitors are more concentrated in the area surrounding the camps and the forbidden areas. Furthermore, we can analyze the different behavior using the distribution in time. In this case, we have analyzed the daily distribution of the semantic trajectories and, particularly, the exploring and suspicious behavior. The result is illustrated in Fig. 17.

This analysis provides evidence that the suspicious behavior has taken place during two particular days. This is an example on how the classification of suspicious behavior can lead the way for a deeper analysis of the data in order to discover other behaviors. Although this experiment has been run on a small dataset, it highlights the main features of the Athena system, showing not only how to classify mined patterns into human behavior, but also how the semantic-enriched KDD iterative process on measured/syntactic/semantic trajectories lead to the discovery of new knowledge.

It is important to point out that in both applications, we have not explicitly dealt with privacy issues, although the privacy of the individual under observation is a crucial concern when dealing with personal location data. In the case of these two datasets, tracked users have signed a disclosure agreement for the analysis of their data. However, this is a general problem and several techniques have been proposed in the literature to make the measured and semantic trajectories anonymous [16,33]. Athena may be easily extended to embed these techniques in the data preprocessing step.

7 Complexity analysis and system performance

This section presents the complexity analysis of the system together with the runtime study for applications shown above. Then, a brief discussion on accuracy issues is given.

As described in Sect. 5, the system performs four main steps: *data-preprocessing, ontology mapping, data mining* , and*reasoning*. In the following, we analyze each step in order to detect the critical complexity factors for the entire system.

The *data-preprocessing* step essentially consists in the stops and moves computation of the syntactic trajectories. Each point of each raw trajectory is scanned in order to detect the set of stops. The moves are created between two consecutive stops, thus creating a syntactic trajectory. The computational cost of this task is $O(|P|)$ where P is the set of points in the raw measured trajectory dataset. The second task is the *ontology mapping*, which imports the instances from the repository into the mapped concept classes of the mobility behavior ontology. The data to be loaded can be divided into three subsets: the set of stop and moves computed in the preprocessing step, called M (i.e., stops and moves), the contextual information called X (in our case, populating the place concept with geographical feature names) and the temporal information called T (populating the time concept). The complexity of the loading step is therefore $O(|M| + |X| + |T|)$ since each entry in the database becomes an instance in the ontology. During the instances loading task, the relationship between them are created, for example the case of *Stop IsIn Place*. These relationships are determined using the spatial and temporal intersections between the stops/moves and the context and temporal instances: $O(|M| \times |X|)$ and $O(|M| \times |T|)$.[14] Therefore, the complexity of this step is $O(|M| + |X| + |T| + |M| \times |X| + |M| \times |T|) = O(|M| \times |X| + |M| \times |T|)$.

The *data mining* step is strongly dependent on the application because it is realized by different algorithms that are selected by the analyst according to the application requirements. The notation O(Mining(P)) is used to represent the complexity of a generic data mining algorithm applied on a dataset P. After running the mining algorithm, a set of patterns K is discovered. Each of these patterns is stored in the database and afterward they are imported into the mobility behavior ontology. To create the relationships between them, for example the *Syntactic Trajectories BelongsTo Pattern* and the set M (stops, moves, and trajectories), the Athena system performs an additional operation. The complexity of this task is then $O(|M| \times |K|) + O(\text{Mining}(P))$.

[14] We consider X and T as the minimal set of places and time intervals used in the application.

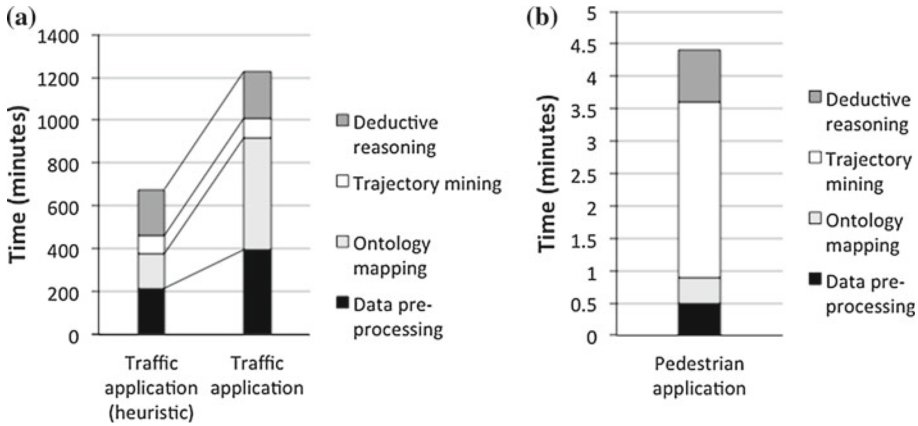


Fig. 18 The running time of the two experiments during the different tasks of the semantic-enriched KDD process. **a** The traffic management application with and without the heuristic. **b** The running time of the recreation management experiment

The *reasoning* step is realized by the Oracle semantic cartridge and the complexity upper bound is $O(|C| \times (|M|+|X|+|T|+|K|))$ where C indicated the number of concepts multiplied by the number of all the instances in the ontology.

The final overall complexity of the Athena system is therefore the sum of all the steps. Furthermore, we can observe that it is reasonable to assume that in several applications, the number of stops and moves is larger than the size of the other sets (usually the set M is one or more orders of magnitude larger than the others), and, as a result, the following assumption can be made:

$$|M| > |C|\,, \quad |M| > |X|\,, \quad |M| > |T|\,, \quad |M| > |K|\,.$$

Thus, the resulting complexity becomes:

$$O\left(|P| + |M|^2\right) + O\,(\text{Mining}\,(P))$$

The factor O(Mining(P)) is the complexity of one of the integrated algorithm (see [2,9,21,42,52]), thus we can simplify the complexity to:

$$O(|M|^2 + \text{Mining}(P))$$

This highlights the two critical points of the Athena system from the complexity point of view:

- The number of stops and moves detected in the trajectory data. In other words, the size of the relevant data needs to be considered in the process.
- The data mining algorithm executed by the system. The computational cost of analyzing the measured trajectory data in order to discover interesting patterns could be computationally expensive and could become the predominant factor.

Figure 18 shows the running times of the experiments presented above. The predominant factor of the traffic ma application is the database size, while the one of the park application is the data mining algorithm. Even if the running time is not compatible with an *on-line* system, it remains still acceptable in the case of an *off-line* scenario that is the one where Athena works.

Since the results of the enriched KDD process are based on two main steps, inductive and deductive reasoning, the accuracy of Athena is based on the accuracy of these two steps. The mining step is based on the accuracy of the specific algorithms, while the accuracy of the deductive step is based on the OWL reasoning engine, which in turn is based on description logics. Therefore, the accuracy in this case is straightforward, since all the trajectories/patterns are classified by the ontology reasoning engine that rely on the description logics formal semantics and represents the domain expert definitions.

However, an important point is the validation of the results with an established ground truth. Unfortunately, these datasets are not accompanied with a ground truth—such as an annotation of the user on the trajectories. Therefore, we have done an empirical evaluation of the results with domain experts in the two fields.

8 Conclusions

This paper presents a semantic-enriched knowledge discovery process for inferring human behavior from the interpretation of movement patterns and exploits the interpretation of movement patterns to infer human behaviors. The main steps of this process are described with a detailed description of the implemented system called Athena. The system has been evaluated into different application domains: traffic management and recreation behavior.

Some open issues still remain to be investigated. First, the inductive–deductive reasoning cycle could be further exploited by automatically transforming the mining patterns into new domain knowledge. For example, a previously unknown human behavior may emerge from mining the semantic trajectory data and this new behavior may become part of the ontology definitions. Finally, a compelling need is to develop a method to formally measure the accuracy of the resulting inferred human behavior. The accuracy of the results is a critical point not easy to cope. Indeed, the lack of a ground truth in the two application case studies—and in general in these kinds of applications—in terms of inferred behavior, makes this task not obvious and more research is needed to give a proper answer to this issue.

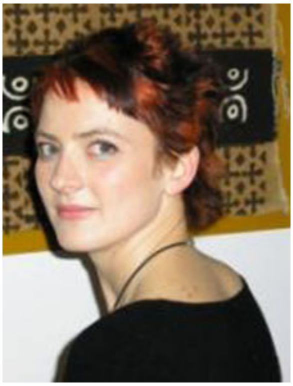
References

1. Agrawal R, Gunopulos D, Leymann F (1998) Mining Process Models from Workflow Logs. In: Ramos I, Alonso G, Schek H-J, Saltor F (eds) Advances in database technology—EDBT'98: sixth international conference on extending database technology. Springer, Valencia, pp 469–483
2. Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Buneman P, Jajodia S (eds) Proceedings of the 1993 ACM SIGMOD international conference on management of data. ACM, Washington, D.C., pp 207–216
3. Ankerst M, Breunig MM, Kriegel H et al (1999) OPTICS: ordering points to identify the clustering structure. In: Delis A, Faloutsos C, Ghandeharizadeh S (eds) SIGMOD 1999, proceedings ACM SIGMOD international conference on management of data. ACM Press, Philadelphia, pp 49–60
4. Antunes C (2007) Onto4AR: a framework for mining association rules. In: Nijssen S, De Raedt L (eds) International workshop on constraint-based mining and learning (CMILE—ECML/PKDD 2007). Warsaw, September 2007
5. Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF (eds) (2003) The description logic handbook: theory, implementation, applications. Cambridge University Press, Cambridge
6. Baglioni M, de Macedo J, Renso C et al (2008) An ontology-based approach for the semantic modelling and reasoning on trajectories. In: Il-Yeol Sang et al (eds) Advances in conceptual modeling-challenges and opportunities ER 2008 workshops proceedings (SeCoGIS 2008). Springer, Barcelona, Spain, October 20–23, pp 344–353

7. Baglioni M, de Macêdo JAF, Renso C et al (2009) Towards Semantic Interpretation of Movement Behavior. In: Sester M, Bernard L, Paelke V (eds) Advances in GIScience, proceedings of the 12th AGILE conference. Springer, Hannover, Germany, pp 271–288
8. Bellandi A, Furletti B, Grossi V et al (2007) Ontology-driven association rules extraction: a case of study. In: Bouquet P, Euzenat J, Ghidini C, McGuinness D, Serafini L, Shvaiko P, Wache H (eds) Proceedings of the international workshop on context and ontologies representation and reasoning (C&O:RR) collocated with the 6th international and interdisciplinary conference on modelling and using context (CONTEXT-2007). Roskilde, Denmark, August 21st, 2007
9. Bogorny V, Kuijper B, Alvaresz LO (2009) ST-DMQL: a semantic trajectory data mining query language. IJGIS 23(10):1245–1276
10. Bogorny V, Wachowicz M (2009) A framework for context-aware trajectory. In: Cao L, Yu PS, Zhang C, Zhang H (eds) Data mining for business applications. Springer, USA, pp 225–239
11. Cao L, Huang JZ, Bailey J et al. (eds) (2011) New frontiers in applied data mining PAKDD 2011 international workshops (behavior informatics 2011 (BI2011) in conjunction with the 15th pacific-asia conference on knowledge discovery and data mining (PAKDD2011)). Springer, LNAI 7104
12. Cao L, Yu PS (eds) (2012) Behavior computing modeling, analysis, mining and decision. Springer, Berlin
13. Cao L, Yu PS, Zhang C (eds) et al (2009) Data mining for business applications. Springer, Berlin
14. Cao L, Yu PS, Zhang C et al (2010) Domain driven data mining. Springer Hardcover, Berlin
15. Cavus O, Aksoy S (2008) Semantic Scene Classification for Image Annotation and Retrieval. In: da Vitoria Lobo N, Roli F, Kwok JT, Anagnostopoulos GC, Loog M (eds) Proceeding of the 2008 joint IAPR international workshop on structural, syntactic, and statistical pattern recognition, (also in Lecture Notes in Computer Science, volume 5342). Springer, Orlando, pp 402–410
16. Chen K, Liu L (2010) Geometric data perturbation for privacy preserving outsourced data mining. Knowl Inf Syst 29:3:657–695
17. Cui J, Liu H, He J et al (2011) TagClus: a random walk-based method for tag clustering. Knowl Inf Syst 27:2:193–225
18. Dodge S, Weibel R, Lautenschultz A-K (2008) Towards a taxonomy of movement patterns. Inf Visalizat 7(3–4):240–252
19. Fayyad UM, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery: an overview. In: Fayyad UM, Piatesky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in knowledge discovery and data mining. AAAI/MIT Press, Cambridge, pp 1–34
20. Feng S, Wang D, Yu G et al (2011) Extracting common emotions from blogs based on fine-grained sentiment clustering. Knowl Inf Syst 27(2):281–302
21. Giannotti F, Nanni M, Pinelli F et al (2007) Trajectory pattern mining. In: Berkhin P, Caruano R, Wu X (eds) Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining. ACM 2007, San Jose California, pp 330–339
22. Giannotti F, Nanni M, Pedreschi D et al (2011) Unveiling the complexity of human mobility by querying and mining massive trajectory data. VLDB Journal Special issue on Data Management for Mobile Services
23. Giannotti F, Pedreschi D (eds) (2008) Mobility, data mining, and privacy: geographic knowledge discovery. Springer, Berlin
24. Gottgtroy P, Modaini R, Kasabov N et al (2003) Building evolving ontology maps for data mining and knowledge discovery in biomedical informatics. In: Proceedings of the third Brazilian symposium on mathematical and computational biology (BIOMATIII), Rio de Janeiro, Brazil, vol 1, pp 309–328
25. Gruber TR (2008) Ontology. In: Ling L, Tamer Özsu M (eds) Entry in the encyclopedia of database systems. Springer, Berlin
26. Guarino N, Oberle D, Staab S (2009) What is an ontology? In: Staab S, Studer R (eds) Handbook on ontologies. Springer, Berlin, pp 1–17
27. Güting R, Schneider M (2005) Moving objects databases. Morgan Kaufmann, Los Altos
28. Hägerstrand T (1970) What about people in regional science? Pap Reg Sci 24(1):6–21
29. Laube P, Imfeld S (2002) Analyzing relative motion within groups of trackable moving point objects. In: Egenhofer MJ, Mark DM (eds) Proceedings of the second international conference on geographic information science LNCS 2478. Springer, Boulder, pp 132–144
30. Laube P, van Kreveld M, Imfeld S (2004) Finding REMO—detecting relative motion patterns in geospatial lifelines. In: Fisher P (ed) Developments in spatial data handling, 11th international symposium on spatial data handling. Springer, Berlin, pp 201–214
31. Kietz J, Serban F, Bernstein A et al (2010) Data Mining Workflow Templates for Intelligent Discovery Assistance and Auto-Experimentation. In: Hilario M, Lavrac N, Kok JN (eds) Proceedings of the ECML/PKDD10 workshop on third generation data mining: towards service-oriented knowledge discovery (SoKD10). Barcelona, Spain, pp 1–12

32. van Marwijk R, Pitt DG (2008) Where Dutch recreationists walk: Path design, physical features and walker usage. In: Raschi A, Tamperi S (eds) Proceedings fourth international conference on monitoring and management of visitor flows in recreational and protected areas. Management for Protection and Sustainable Development, Montecatini Terme, Italy, pp 428–432
33. Monreale A, Trasarti R, Renso C et al (2011) C-safety: a framework for the anonymization of semantic trajectories. Trans Data Privacy 4:2:73–101
34. Nanni M, Kuijpers B, Korner C et al (2008) Spatiotemporal data mining. In: Giannotti F, Pedreschi D (eds) Mobility, data mining, and privacy: geographic knowledge discovery. Springer, Berlin
35. Nanni M, Trasarti R (2009) K-BestMatch reconstruction and comparison of trajectory data. In: Saygin Y, Yu JX, Kargupta H, Wang W et al (eds) ICDM workshops 2009 international workshop on spatio and spatio temporal data mining in cooperation with IEEE—ICDM 2009. IEEE Computer Society 2009, Miami, pp 610–615
36. Nanni M, Trasarti R, Renso C et al (2010) Advanced knowledge discovery on movement data with the GeoPKDD system. In: Manolescu I, Spaccapietra S, Teubner J, Kitsuregawa M, Léger A, Naumann F, Ailamaki A, Ozcan F (eds) EDBT 2010. ACM, Lausanne, pp 693–696
37. Nigro HO, Gonzalez Cisaro SE, Xodo DH (eds) (2008) Data mining with ontologies: implementations, findings and frameworks. IGI Global, Hershey
38. Ortale R, Ritacco E, Pelekis N et al (2008) The DAEDALUS framework: progressive querying and mining of movement data. In: Aref WG, Mokbel MF, Schneider M (eds) Proceedings of the 16th ACM SIGSPATIAL international symposium on advances in geographic information systems. ACM-GIS, Irvine, pp 52:1–52:4
39. Pelekis N, Kopanakis I, Kotsifakos E et al (2011) Clustering uncertain trajectories. Knowl Inf Syst 28(1):117–147
40. Pelekis N, Theodoridis Y (2006) Boosting location-based services with a moving object database engine. In: Chrysanthis PK, Jensen CS, Kumar V, Labrinidis A (eds) Proceedings of the 5th ACM international workshop on data engineering for wireless and mobile access. ACM, Chicago, pp 3–10
41. Reeve L, Han H (2005) Survey of semantic annotation platforms. In: Haddad H, Liebrock LM, Omicini A, Wainwright RL (eds) Proceedings of the 2005 ACM symposium on applied computing, SAC '05. ACM, New York, pp 1634–1638
42. Rinzivillo S, Pedreschi D, Nanni M et al (2008) Visually-driven analysis of movement data by progressive clustering. Inf Visual 7(3/4):225–239
43. Romei A, Turini F (2011) Inductive database languages: requirements and examples. Knowl Inf Syst 26:3:351–384
44. Song C, Qu Z, Blumm N et al (2010) Limits of predictability in human mobility. Science 19 327(5968): 1018–1021
45. Spaccapietra S, Parent C, Damiani ML et al (2008) A conceptual view on trajectories. Data Knowl Eng J 65:1:126–146
46. Spinsanti L, Celli F, Renso C (2010) Where you stop is who you are: understanding peoples' activities. In: Gottfried B, Aghajan H (eds) Proceedings of the 5th workshop on behaviour monitoring and interpretation—user modelling. Karlsruhe, Germany
47. Trasarti R, Pinelli F, Nanni M et al (2011) Mining mobility user profiles for car pooling. In: Aptè C, Ghosh J, Smyth P (eds) Proceedings of the 17th ACM SIGKDD conference on knowledge discovery and data mining. ACM, San Diego, pp 1190–1198
48. Uren V, Cimiano P, Iria J et al (2006) Semantic annotation for knowledge management: requirements and a survey of the state of the art. J Web Semant 4(1):14–28
49. Valente A, Breuker J (1996) Towards principled core ontologies. In: Gaines BR, Mussen M (eds) Proceedings of the KAW-96. Banff
50. Yan Z (2009) Towards semantic trajectory data analysis: a conceptual and computational approach. In: Gigaux P, Senellrt P (eds) Proceedings of the VLDB 2009 PhD workshop co-located with the 35th international conference on very large data bases (VLDB 2009). VLDB Endowment, Lyone, France
51. Yan Z, Parent C, Spaccapietra S et al (2010) A hybrid model and computing platform for spatio-semantic trajectories. In: Aroyo L, Antoniou G, Hyvönen E, Teije A, Stuckenschmidt H, Cabral L, Tudorache T (eds) The semantic web: research and applications. Springer, Berlin, pp 60–75
52. Wachowicz M, Ong R, Renso C et al (2011) Finding moving flock patterns among pedestrians through spatio-temporal coherence. Int J GIS 25(11):1849–1864

Author Biographies



Chiara Renso holds a Ph.D. and M.Sc. degree in Computer Science from University of Pisa (1992, 1997). She is currently a KDDLab member and permanent researcher at ISTI Institute of CNR, Italy. Her research interests are related to spatio-temporal reasoning, data mining query languages, semantic data mining, mobility data mining, and ontologies. She is the author of more than 60 publications and has been reviewer for several data mining and GIS conferences. She has been co-chair of the First, Second, and Third International Workshop on Semantic Aspects of Data Mining in conjunction with IEEE ICDM conference. She is co-editor for the special issue of journal Knowledge And Information System on "Context-aware data mining". She is a scientific coordinator of the EU-FP7-PEOPLE project SEEK.

Miriam Baglioni graduated in Computer Science with honors, in 2001, from the University of Pisa. Her thesis was entitled "DMQL: a proposal for a Data Mining Query Language". She received her Ph.D. in Computer Science, in July 2005 from the University of Pisa. Her thesis was entitled "Representation and exploitation of domain knowledge in a KDD environment". She was a part of the Knowledge Discovery and Delivery Laboratory. She is an associate researcher at IIT-CNR. She is currently working in the field of social networks.

Jose António F. de Macedo holds an M.Sc. in Computer Science and a Ph.D. degree in Computing, both from the Pontifical Catholic University of Rio de Janeiro, Brazil. During his Ph.D. course, he worked at ISI Laboratory at ENST-Bretagne, France, and at Database Laboratory at EPFL, Switzerland. From 2006 to 2009, he was a Senior Researcher at Database Laboratory at EPFL, Switzerland. He joined the Computer Science Department of the Federal University of Ceará in 2009, where he is currently a Associate Professor. In 2010, he was granted with a research fellowship from the National Counsel of Technological and Scientific Development (CNPq) of Brazil. He is the author of more than 30 refereed journal and conference papers and his research interests include conceptual modeling and database design, cloud computing data management, and semantic web.

Roberto Trasarti was born in 1979 in Italy. He graduated in Computer Science in 2006, at the University of Pisa. He discussed his thesis on ConQueSt: a Constraint-based Query System aimed at supporting frequent patterns discovery. He started the Ph.D. in Computer Science at the School for Graduate Studies "Galileo Galilei", (University of Pisa). In June 2010, he received his Ph.D. presenting the thesis entitled "Mastering the Spatio-Temporal Knowledge Discovery Process". Currently, he is a researcher at ISTI-CNR as member of Knowledge Discovery and Delivery Laboratory. His interests regard data mining, mobility data analysis, artificial intelligence, and automatic Reasoning.

Monica Wachowicz is currently working at the University of New Brunswick, Canada. Her research is focussed on discovering mobility patterns evolving from social interactions and developing the next generation of services that can translate mobility patterns into actionable information about societal lifestyle expectations and needs. Her research interests are in the areas of geographic knowledge discovery, space-time representation and reasoning, and visual analytics.