Handling Knowledge over Moving Object Trajectories using Formal Concept Analysis

R Wannous, C Vincent, Salah Eddine Boukhetta

Handling Knowledge over Moving Object Trajectories using Formal Concept Analysis

R. Wannous[1], SE. Boukhetta[2], and C. Vincent[1]

[1]La Rochelle University, L3i laboratory, EA 2118, France
[2]Aneo, France
{rouaa.wannous,cecile.vincent}@univ-lr.fr
sboukhetta@aneo.fr

Abstract. We define the term semantic trajectories of moving objects as a sequence of spatio-temporal points with associated semantic information. Spatio-temporal points are directly generated by sensors that capture the position of a moving object in time.
In this paper, we work on a marine mammal trajectory case study. We consider the seals' trajectories to understand their behavior within groups and identify their activities simultaneously in the same place. We define the activities of mobile objects in the form of rules given by the domain expert. To gain knowledge of trajectories, we use the `GALACTIC` platform, which is a new platform based on Formal Concept Analysis (FCA) for computing a concept lattice from heterogeneous and complex data. Data in `GALACTIC` are described by predicates according to their types. Here, we will use interval sequence plugins to analyse the trajectories of seals, where interval sequences are represented by a set of time intervals in which a seal performs an activity in a geographical zone. Finally, the results show the simultaneous activities of a group of seals.

Keywords: Spatio-Temporal data · Formal concept analysis · Trajectory mobile objects · `GALACTIC`.

1 Introduction

Thanks to device networks, GPS-enabled devices, and especially mobile sensors and satellite such as the Global navigation satellite system (GNSS), data on moving objects is collected in various application domains. Moving objects are entities that change their position in space and in time. We can consider these objects as moving points such as vehicles, people, animals or moving regions. Each moving object, either point or region, has a sequence of positions between two or more time-varying stops ordered temporally, characterized by a starting position and an arrival position. We call this sequence of positions a `trajectory` [40]. The trajectory traces a moving object from a departure point to a destination point as sequences of data (sample points captured, time of the capture). As indicated in [43], a trajectory is defined as the evolution of the position of a moving object that changes its location in space over a given time interval. Each moving object has a trajectory.

Moving objects and their trajectories play an essential role in applications dealing with tracking animals, traffic control, tracking environmental phenomena, forecasting based on vehicle paths and so on. However, `raw trajectories` do not contain the goals of journey or the activities performed by the moving object. Large datasets need to be analyzed and modeled to meet the users' requirements. Analyzing the trajectory of a moving object means identifying the components of its trajectories (start, stop, move, end) and determining the different relationships between them.

Despite the improvement from raw to structured trajectories, data is still lacking in the knowledge aspects that are fundamental for their efficient use. The analysis of both raw and structured trajectories does not lead to sufficient and significant information. These analyses can be exploited by different applications that associate the trajectories of the moving objects with other elements of any field of interest. The latter must be retrieved, represented, and interpreted as a domain knowledge to extract semantics (activities) from trajectories. Semantic enrichment processes allows us to find and present the locations of activities from the data collection.

These works [46,36] present and queer semantic trajectories in the maritime domain based on a specific ontology. Indeed, they are based on an ontology modelling approach for semantic trajectories [46]. However, the domain part of the trajectory ontology focuses on the mobile object's characteristics and its trajectory's activities. Their model considers a trajectory as a spatio-temporal concept; then, they map it with other temporal and spatial model resources. We apply our modeling approach to a specific domain: marine mammal tracking. The experimental results addressed time computation and spatial storage problems of the ontology inference. Subsequently, we proposed solutions to reduce the inference complexity by defining time constraints in [47] and inference pass refinements in [45]. These latter studies mainly focus on time computation on single object trajectories.

This paper deals with marine mammal tracking applications, namely `Seal Trajectories`. In this work, we transform the initial low-level real world GNSS tracking data i.e. `Raw Trajectory` data to high-level `Semantic Trajectories`. Our goal is to enrich trajectory data with semantics to gain more knowledge. We propose a clustering method for grouping `Seal Trajectories` based on their activities and zones of interests, using Formal Concept Analysis (FCA for short).

Formal Concept Analysis (FCA) appeared in 1982 [48], then in Ganter and Wille's 1999 work [26], it resulted from a branch of applied lattice theory that first appeared in the book of Barbut and Monjardet in 1970 [6]. The lattice property guarantees a hierarchy of clusters and a complete and consistent navigation structure for interactive approaches [22]. The formalism of pattern structures [25,31] and abstract conceptual navigation [21,24] extend FCA to deal with non-binary data, where data is described by patterns such that the pattern space must be organized as a semi-lattice in order to maintain a Galois connection between objects and their descriptions. By the FCA framework, pattern lattice and bases of rules are defined, where a concept is composed of a subset of objects together with their common patterns, and a rule possesses patterns in premises and conclusions. However, pattern lattices are huge, often intractable [30], and the need for approaches to direct the search towards the most relevant patterns is a current challenge. Logical Concept Analysis [23] is a logical expression that replaces a generalization of FCA in which we use sets of attributes. The power set of attributes mentioned by the Galois connection is replaced by an arbitrary set of formulas associated with a deduction relation (i.e., subsumption), and conjunctive and disjunctive operations, and therefore forms a lattice.

Inspired by pattern structures, the NextPriorityConcept algorithm, introduced in a recent article [19], proposes a user-driven pattern mining approach for heterogeneous and complex data as input. This algorithm allows generic pattern computation through specific *descriptions* of objects by predicates. It also proposes to reduce predecessors of a concept by refining a set of objects into fewer ones through specific user exploration *strategies*, resulting in a reduction of the number of generated patterns.

Some algorithms appear within the FCA framework for analyzing sequence data; we can mention works for mining medical care trajectories using pattern structures [13,14], sequence mining to discover rare patterns [15], and other studies on demographic sequences [27,28]. However, we found fewer papers on the discovery of interval-based sequences using FCA methods. We can cite Kaytoue et al.'s work on gene expression data [32].

The rest of the paper is structured as follows. After introducing the Preliminaries of FCA in Section 2, we discuss the related work in Section 3. We prepare the trajectory data to be analysed by `GALACTIC` in Section 4. Section 5 describes our experiments on the proposed approach using `GALACTIC`. Finally, Section 6 concludes this paper and offers directions for future work.

2 Preliminaries

2.1 Interval-based Sequences

A sequence s is a succession of itemsets from a dictionary Σ, often in the form of $s =< X_i >_{i \leq n}$, where $X_i \subseteq \Sigma$ is a subset of items i.e., itemset. A temporal sequence is a sequence where each itemset X_i must have an associated timestamp t_i. An *Event* (or *Time frame*) E, is a triple $E = (\underline{t}, \overline{t}, X)$ where $X \subseteq \Sigma$ is an *itemset*, $\underline{t}$ is the starting time and $\overline{t}$ is the ending time, $\underline{t} \leq \overline{t}$. For better readability we refer to $(\underline{t}, \overline{t})$ by T.

Interval-based sequence. An *interval-based sequence* (or *Time frame sequence*) $s = \langle (T_i, X_i) \rangle_{i \leq n}$ is a list of events (or time frames), verifying $\overline{t_i} < \underline{t_{i+1}}$, thus an interval-based sequence is a list of separate intervals containing itemsets. The size of the interval-based sequence is the number of its time frames. We refer to the interval-based sequence by *sequence*.

Consider the example in Fig 1 for an alphabet $\Sigma = \{C, M, P, H\}$ (where C stands for Castle, M for Museum, P for Public Garden and H for Historical Place), the sequences represent trajectories of visits of three tourists $s1, s2$ and $s3$. In this example, visitors may be in two or more different locations

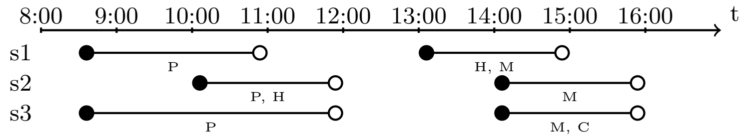


Fig. 1: Example of interval-based sequences

at the same interval as the intervals are large enough and we may don't have the exact interval of each location.

Subinterval. For two intervals, $T = (\underline{t}, \overline{t})$ and $T' = (\underline{t'}, \overline{t'})$, we say that T is subinterval T', if : $\underline{t} \geq \underline{t'}$ and $\overline{t} \leq \overline{t'}$ and we write $T \preceq T'$, that corresponds to the containing relation from Allen's relations [2].

Projections. We introduce the *projection* operator Φ of a sequence s, over a given interval T, that selects all the itemsets of the sequence included in this interval : $\Phi_T(s) = \{X' \ : \ T' \preceq T \text{ and } (T', X') \in s\}$. Dually, the *projection* operator Φ, over an itemset $X \subseteq \Sigma$ selects all the intervals where the items of X may occur: $\Phi_X(s) = \{T' \ : \ X' \subseteq X \text{ and } (T', X') \in s\}$. $\Phi_\Sigma(s)$ represents a set of all the intervals in s.

Subsequence. A sequence s, is *subsequence* of another sequence s', $s \Subset s'$ if for all $(T, X) \in s$, there exists $(T', X') \in s'$ such that $T \preceq T'$ and $X \subseteq X'$. We also say that s' is **supersequence** of s.

Affix. A prefix/suffix of a sequence $s = \langle (T_i, X_i) \rangle_{i \leq n}$ according to a window w, is the subsequence of s composed by the first/last w time frames of s, prefix$(s, w) = \langle (T_i, X_i) \rangle_{1 \leq i \leq w}$, suffix$(s, w) = \langle (T_i, X_i) \rangle_{(n-w) < i \leq n}$.

Cardinality. For a set of sequences A, an item $x \in \Sigma$ and an interval T, the function *card* gives the number of sequences $a \in A$ possessing the item x in the projection of a over T, $x \in \Phi_T(a)$.

$$card(A, T, x) = |\{a \ : \ x \in \Phi_T(a), a \in A\}| \tag{1}$$

When $card(A, T, x)$ is maximal, we denote $card(A, T, x)$ by $card_{max}(A, T)$. We define $card_{min}(A, T)$ in the same maner when $card(A, T, x)$ is minimal.

From example in Fig 1 we have, $\Phi_{(10:00, 11:00)}(s2) = \{P, H\}$, the prefix of $s1$ is $\langle (08{:}30, 11{:}00), P) \rangle$, and all three tourists were in the museum from 14:00 to 15:00, so $\langle (14{:}00, 15{:}00), M) \rangle$ is subsequence of $s1, s2$ and $s3$. For $A = \{s1, s2, s3\}$ and $T = (11{:}00, 12{:}00)$ $card(A, T, P) = card_{\max}(A, T) = 2$.

2.2 Description of the NextPriorityConcept algorithm

The NextPriorityConcept algorithm [19] computes concepts for heterogeneous and complex data for a set of objects G, its main characteristics are:

Heterogeneous data as input, described by specific predicates. The algorithm introduces the notion of *description* δ as an application to provide predicates describing a set of objects $A \subseteq G$. Each concept $(A, \delta(A))$ is composed of a subset of objects A and a set of predicates $\delta(A)$ describing them. Such generic use of predicates makes it possible to consider heterogeneous data as input, i.e., numerical, discrete or more complex data. However, unlike classical pattern structures, predicates are not globally computed in a preprocessing step, but locally for each concept.

Concept lattice generation. The NextPriorityConcept algorithm is inspired by Bordat's algorithm [9], also found in Linding's work [35], that recursively computes the immediate successors of a concept, starting with the bottom concept. It is a dual version that computes the immediate predecessors of a concept, starting with the top concept $(G, \delta(G))$ containing the whole set of objects, until no more concepts can be generated. The use of a priority queue ensures that each concept is generated before its predecessors, and a mechanism of propagation of constraints ensures that meets will be computed. NextPriorityConcept computes a concept lattice and therefore is positioned in FCA framework, with the possibility of extraction of rules, closure computations or navigation in the lattice, that can be useful in many fields of pattern mining and discovery.

Predecessors selection by specific strategies. The algorithm also introduces the notion of *strategy* σ to provide predicates (called *selectors*) describing candidates for an object reduction of a concept $(A, \delta(A))$ i.e., predecessors of $(A, \delta(A))$ in the pattern lattice. A selector proposes a way to refine the description $\delta(A)$ to a reduced set $A' \subset A$ of objects. Several strategies are possible to generate predecessors of a concept, going from the naive strategy classically used in FCA that considers all the possible predecessors, to strategies reducing the number of predecessors in order to obtain smaller lattices. Selectors are only used for the predecessors' generation, they are not kept either in the description or in the final set of predicates. Therefore, choosing or testing several strategies at each iteration in a user-driven pattern discovery approach would be interesting.

The main result in [19] states that the NextPriorityConcept algorithm computes the formal context $\langle G, P, I_P\rangle$ and its concept lattice (where P is the set of predicates describing the objects in G, and $I_P = \{(a, p),\ a \in G, p \in P\ :\ p(a)\}$ is the relation between objects and predicates) if description δ verifies $\delta(A) \sqsubseteq \delta(A')$ for $A' \subseteq A$. The run-time of the NextPriorityConcept algorithm has a complexity $O(|\mathcal{B}|\,|G|\,|P|^2\,(c_\sigma + c_\delta))$ (where $\mathcal{B}$ is the number of concepts, c_σ is the cost of the strategy and c_δ is the cost of the description), and a space memory in $O(w\,|P|^2)$ (where w is the width of the concept lattice).

3 Related Literature

Researchers have actively investigated data management techniques during the last decade, including modeling, indexing, inferencing, and querying large data [51,37,47]. Most of these techniques are only interested in representing and querying moving object trajectories [8,46,5].

A trajectory describes the evolution of the position of moving objects at a given time interval [29]. It is a spatio-temporal path of the ongoing moving object. Consequently a trajectory is a series of points that trace the moving object's path. Each point has a position in the spatial dimension (X, Y) and a time interval in the temporal dimension (T). The point is considered as a stop and the changing position between two successive points is considered as a move.

Spaccapietra et al. [42] proposed a conceptual view of trajectories in which trajectories are a set of stops and moves. Each part contains a set of semantic data. Based on this conceptual model, several studies have been proposed such as [4,52]. Alvares et al. [4] proposed a trajectory data pre-processing method to integrate trajectories with the space. Their application concerned daily trips of employees from home to work and back. However, the scope of their paper is limited to the formal definition of semantic trajectories with space and time without any implementation and evaluation. [16] is a work that proposes a geographic ontology-based conceptual trajectory model called STriDE. It focuses on the representation of moving objects and dynamically changing environments. Actually, STriDE is composed of an object identity, a set of object properties, a geometric spatial representation and the timeslice's valid period. In this model, a semantic trajectory is defined as a set of timeslices having a starting and an ending spatio-temporal point. More recently, MASTER [38] is a conceptual semantic trajectory model considering trajectories as semantic multi-dimensional sequences. This model has been converted to a logical RDF Schema and implemented using a middleware that stores RDF data in multiple NoSQL databases. It focuses on the heterogeneity of the semantic information of trajectories paying particular attention to the relationships between moving objects.

For the vehicles' trajectories, in [7] their model is customized for ambulances and the process has been validated with real data of connected ambulances belonging to the Indre-et-Loire fire department in France. This raw data identifies how, why and when mobile objects are moving to add semantic information to the trajectories and then use applications such as Global navigation satellite system (GNSS), reporting systems or Remote Maintenance System. The goal is to define a process on how to extract these semantic behaviors from raw data and external database for specific vehicles.

Regarding the trajectory query, [34] studied the problem of searching activity trajectories more effectively by replacing keyword matching with semantic interpretation. The probabilistic topic model is used to interpret the thematic meaning of keywords associated with trajectories and user queries. They developed a hybrid index structure called TP-tree. They proposed practical searching algorithms to prune the search space to support the search for the spatial activity trajectory and thematically in proximity to the user query.

The Geolife project dataset [1], for example, holds 17,621 raw trajectories collected by 78 users while using different modes of transportation. These raw data of trajectories are designated as raw trajectories of moving objects and do not contain semantic annotations and nor metadata. Conversely, the enrichment of this dataset enables users and external applications to request services and semantic trajectories of moving objects data, for example, to make inferences about places of traffic congestion or identify patterns in taxi drivers' routes. The growing availability of raw trajectories of moving objects imposes the need of analysis, semantic enrichment, and sharing with communities that are interested in their application [49,3]. These semantically enriched trajectories correspond to a sequence of spatial-temporal points with associated semantic information (e.g., information on the place visited, type of transport of the moving object, the aim of the movement etc.).

In the domain of tourism, the observation of the trajectory of tourists recorded using a tracking device without any tags is meaningless. It does not give information about his destination, the places they visited, the means of transport they used, etc. However, when adding geographic annotations and sense to the trajectory, several observations and conclusions can be made. For the same tourism example, semantic trajectory provides meaningful information such as the most visited museums, the most used means of transport. Such information can help in improving the tourism sector. Here came the need for semantic trajectories. In fact, they are the evolution of structured trajectories by semantically enriching the trajectory components (begin, stop, move, end) with annotations and by establishing relationships with application-dependent entities [17].

For the trajectory warehouse, authors [39] formulate and design a generic framework for trajectory data warehouse ontological modeling. The proposed approach offers a design platform for knowledge discovery and predictive trend analysis. To validate their approach, they applied the generic model to different case studies such as the tourist movement management, the bird migration movement and the highway traffic and transportation management. their approach needs to incorporate comprehensive optimization measures for faster and more efficient query processing. In addition, the authors highlight the need to formulate and enforce a framework for the modeling and design of the semantic trajectory. In [40], they conducted a systematic survey on the major research into moving object trajectory data and moving object trajectory data warehouse modeling to serve as background for the designers and the research community that are interested in the field of modeling trajectory data and trajectory data warehouses. To model a trajectory data warehouse, they identified the moving object properties and specificities, then modelled moving object trajectory data warehouse model. A comparison was made between them, through which strong and limited contributions were shown.

In the spatial data infrastructures, the authors in [44] presented their definition for documenting and enriching trajectories using spatial data infrastructures. Thus, spatial data infrastructures for semantic trajectories of moving objects can be defined as a set of users, professionals, technologies, standards, policies, data, metadata, Web services for semantic annotation, Maps Application Programming Interfaces (APIs), and different microservices. Their approach enriches trajectories, both automatically and manually.

For the indoor trajectories, authors [33] presented a conceptual model of spatio-temporal indoor trajectories enriched with semantic annotations. They integrated semantic annotations at different levels in order to allow a detailed description of the movements. They derived the discussion on modeling issues that have been overlooked so far and illustrated them with a real-world case study involving a mobility dataset of the Louvre museum visitors. [50] provides the outline of a moving objects database system, aimed at integrating multiple movement data models (e.g. road network models, region-based outdoor models, indoor models) paying attention to the support of semantics and multiple descriptive attributes.

Each approach suffers from whole genericity. Existing spatial trajectory query studies mainly focus on analyzing the spatio-temporal properties of the users' trajectories, while leaving the understanding of their activities largely untouched. Existing approaches could be improved to fulfill this lack.

4 Trajectory analysis using `GALACTIC`

4.1 Description and strategies for Interval-based sequences

`GALACTIC`[1] is a new platform for computing patterns from heterogeneous and complex data that extend the approach of pattern structures [25] and logical concept analysis [23]. It's a development platform for a generic implementation of the NextPriorityConcept [18] algorithm allowing easy integration

[1] https://galactic.univ-lr.fr/

of new plugins for characteristics, descriptions, strategies and meta-strategies. The guides on how to use the platform can be found here[2].

The `GALACTIC` platform allows the analysis of binary, numerical and categorical data. Sequences handling have been added to the platform recently [11] and multiple descriptions and strategies are available for simple, temporal, and interval sequences. Simple Sequences have three descriptions: Maximal Common Subsequences, Prefixed Common Subsequences and K-Common Subsequences. Simple Sequences also have two strategies: Naive Strategy and Augmented Strategy. They aim to extract subsequences from a sequence in various ways.

In order to mine interval-based sequences with NextPriorityConcept algorithm, we have to define specific descriptions and strategies. More descriptions and strategies for interval-based sequences were introduced here [12]. Consider a set G of sequences whose size is smaller than n, defined on an alphabet Σ as input:

A description δ is a mapping $\delta : 2^G \to 2^P$ which defines a set of predicates $\delta(A)$ describing any subset $A \subseteq G$ of sequences. Predicates are of form, *"is subsequence/supersequence of"*.

A strategy σ is a mapping $\sigma : 2^G \to 2^P$ which defines a set of selectors $\sigma(A)$ to select strict subset A' of A as predecessor candidates of any concept $(A, \delta(A))$ in the pattern lattice.

Predicates are computed using the subsequence relation in the form *"is subsequence of "*. For better readability, the sets $\delta(A)$ and $\sigma(A)$ will be treated either as sets of predicates/selectors, or as sets of sequences, they can reciprocally be deduced from each other.

Description for interval sequences The maximal common time frame description `MCTF` refers to the classical maximal common subsequence description [10] and corresponds to the set of maximal subsequences of all sequences in $A \subset G$.

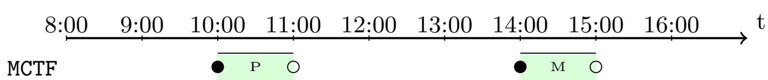


Fig. 2: $\delta_{\text{MCTF}}(\{s1, s2, s3\})$ for $s1, s2$ and $s3$ in Fig 1.

Fig 2 represents $\delta_{\text{MCTF}}(A)$ for $A = \{s1, s2, s3\}$ from Fig 1. We can observe that `MCTF` could be interpreted as "conjunction" where $(14{:}00, 15{:}00, \{M\})$ in δ_{MCTF} means that all $s1, s2$ and $s3$ contain $(14{:}00, 15{:}00, \{M\})$. More formally, `MCTF` is defined for a subset $A \subseteq G$ of sequences by:

Maximal Common Time Frame (MCTF) description.

$$\delta_{\text{MCTF}}(A) = \{\langle (T, X) \rangle \ : \forall a \in A, X \subseteq \Phi_T(a)\} \tag{2}$$

Strategies and selectors for time frame sequences Strategies are used by the NextPriorityConcept algorithm to refine each concept $(A, \delta(A))$ into concepts with fewer objects (sequences) and more specific descriptions.

The *Augmented Minimum Cardinality (AMC)* strategy computes all the possible refinements of a concept $(A, \delta_{\text{MCTF}}(A))$ by adding in the events of sequences of δ_{MCTF} any item with a minimal cardinality $card(A, T, x)$ for each time frame T. More formally, σ_{AMC} is defined for $A \subseteq G$ by:

Augmented Minimum Cardinality (AMC) strategy.

$$\begin{aligned}\sigma_{\text{AMC}}(A) = \{\langle (T, X) \rangle :\ & \forall a \in A, \Phi_T(a) \subseteq X \text{ and } \forall x \in X \\ & card(A, T, x) = |A| \ \vee \ card(A, T, x) = card_{min}(A, T)\}\end{aligned} \tag{3}$$

Fig 3, represents the generated Hasse diagram of the concept lattice generated for sequences in Fig 1 using the `MCTF` description and the `AMC` strategy, where in each concept the symbol $ represents

[2] https://galactic.univ-lr.fr/guides/

the identifier of the concept, and the symbol # represents the number of sequences inside the concept, i.e., its support.

The concept lattice can be red as a hierarchy of concepts (groups/clusters). The first concept in top contains all the objects (the 3 visits) and their description which is general. Then predecessor concepts contains less objects and more specific description (concepts $1 and $2) and so on to the bottom concept that doesn't contain any object. The concept $0 contains the description of the 3 visits. The concept $1 describes the two visits, *s*1 and *s*3 as they were in the Public Garden from 08:30 to 11:00 then in the Museum from 14:00 to 15:00.

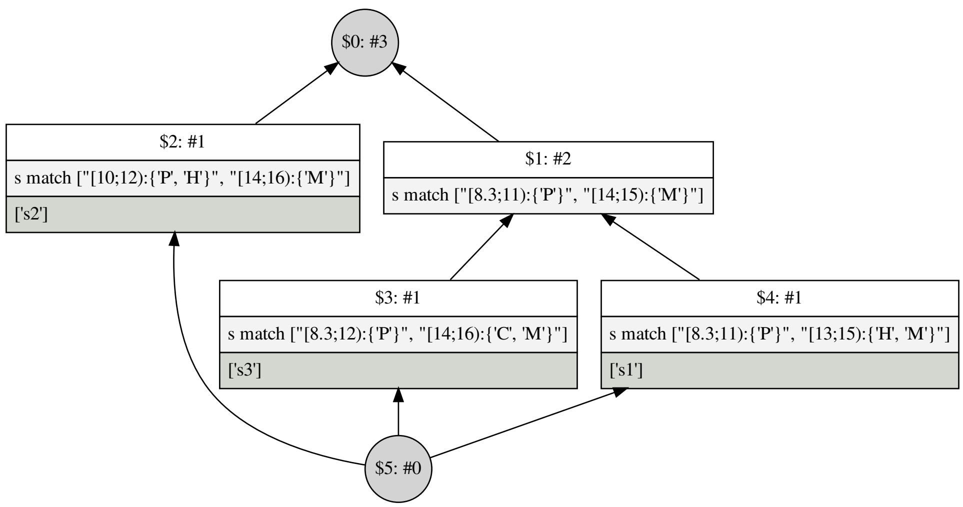


Fig. 3: Hasse diagram of the reduced concept lattice for the AMC strategy and the MCTF description

4.2 Raw data to semantic trajectories

We deal with seal trajectories. The data comes from the LIENSs[3] (CNRS/University of La Rochelle) in collaboration with SMRU[4]. These laboratories work on marine mammals. Captured data of seals between their haulout sites along the coasts of the English Channel or in the Celtic and Irish seas are captured using GNSS[5] systems provided by SMRU. These sensors, in a tag glued to the fur of the animal behind the head, capture trajectories consisting of spatio-temporal data. Trajectories data can also contain some meta-data about the moving objects. These datasets are organized into sequences of points. Every sequence, mapped to a temporal interval, characterizes a defined state of the animal. In our application, we consider three main states of a seal: *hauling out, diving* and *cruising*. Fig 4 shows the three states, the transitions and the states guard conditions:

- GPS locations are captured every 20 minutes when the seal is at the surface;
- Diving data contains maximum depth, total duration, surface duration (time spent at the surface after a dive and before the next one) and the TAD index. The TAD (Time Allocation at Depth) index defines the shape of a seal's dive, as mentioned in [20]. A dive starts when the tag goes below a chosen "depth threshold". For most GPS/GSM deployments, the chosen depth threshold was 1.5 meters;
- A haulout is a period of time spent on land by the seal. It starts when the tag is dry for at least 10 minutes and ends when the tag is wet for at least 40 seconds.

[3] http://lienss.univ-larochelle.fr
[4] SMRU: Sea Mammal Research Unit- http://www.smru.st-and.ac.uk
[5] GNSS : Global Navigation Satellite System

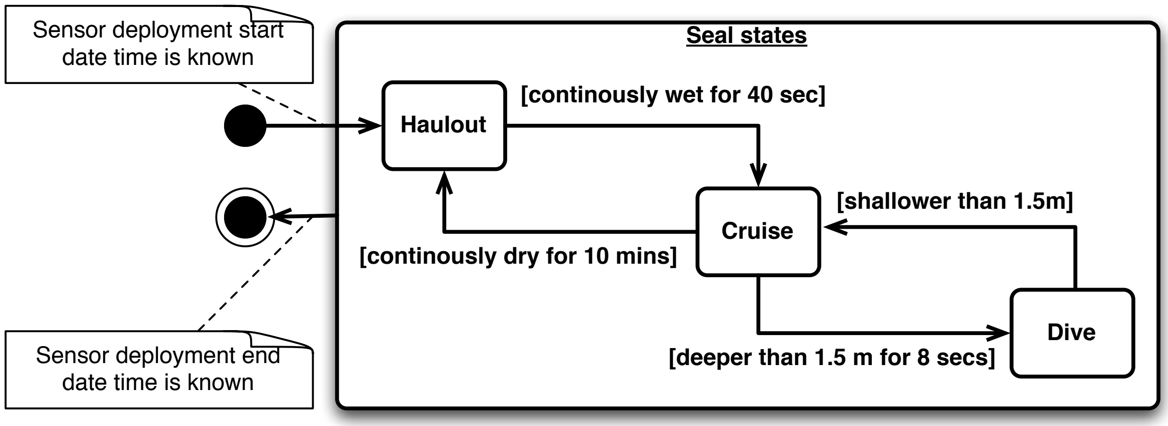


Fig. 4: The three states of seal trajectory

We consider the *diving* state in this work. We compute the `TAD` index over a data set to classify geometric shapes of dives. For this classification, we can distinguish three patterns as shown in Fig. 5, presented in [41]:

- dive geometric shaped in form of V: if $0 <= TAD < 0.7$
- dive geometric shaped in form of U+V: if $0.7 <= TAD < 0.9$
- dive geometric shaped in form of U: if $0.9 <= TAD < 1$

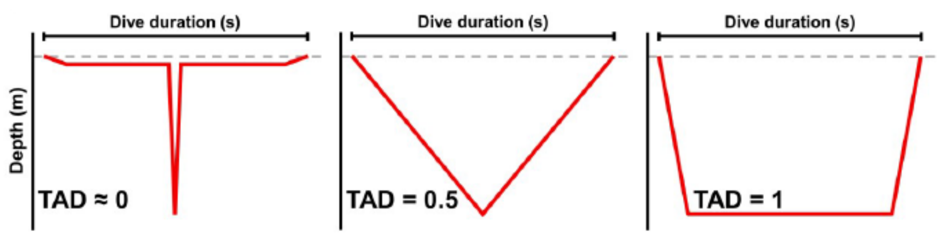


Fig. 5: The dive shape or TAD [41]

According to the domain expert, there is a correlation between the geometrical shape of the dives and the seal activities. In addition to the geometric shape of the dives, we take into account the maximum dive depth and surface ratio which is the ratio between surface duration and dive duration. The decision Table 1 summarizes conditions of the IF parts of rules associated with seals' activities. In the raw data, we have trajectories for 14 seals. Table 2 shows a snapshot of the raw data of seals with their properties.

Table 1: Decision table of `IF` parts of seal activities

Rules	**Max dive depth** (meter)	**Dive shape** or TAD	**Surface ratio**= surface dur/dive dur
Resting	< 10	>0.9	> 0.5
Foraging	> 3	> 0.9	< 0.5
Travelling	> 3	< 0.7	$<$ all
TravellingForaging	> 3	> 0.7 & < 0.9	< 0.5

Table 2: Snapshot of raw seals' trajectories

SEAL	DE_DATE	LAT	LON	SURF_DUR	DIVE_DUR	MAX_DEP	TAD
V13	2020-10-13 16:24:56	49.377413	-1.156923	4	56	1.5	0.040984
V13	2020-10-13 16:25:12	49.377399	-1.156932	36	12	1.6	0.364491
V13	2020-10-13 16:26:32	49.377327	-1.156976	84	44	1.6	0.452500
V13	2020-10-13 16:28:08	49.377240	-1.157029	12	12	1.5	0.067568
V13	2020-10-13 16:28:36	49.377215	-1.157044	12	16	1.5	0.044204

4.3 Analyzing trajectory with GALACTIC

In order to analyse trajectory data using GALACTIC, we transform the raw data to interval-based sequences defined by $s = \langle (T_i, X_i) \rangle_{i \leq n}$ where $T_i = (\overline{t_i}, \underline{t_{i+1}})$, is an interval of time, and X_i is a set with two elements: the activity of the seal and its geographical location.

For simplicity, we aggregate geographical data (LAT, LON) to one decimal point, in which we calculate the activity according to rules in Table 1, data are taken in an interval of 24 hours. A snapshot of the results of the aggregating data are shown in Table 3.

Table 3: Aggregating data

SEAL	DE_DATE	LAT	LON	ACTIVITY
V13	2020-10-13 16:24:56	49.3	-1.1	Foraging
V13	2020-10-13 16:25:12	49.3	-1.1	Foraging
V13	2020-10-13 16:26:32	49.3	-1.1	Foraging
V13	2020-10-13 16:28:08	49.3	-1.1	Foraging
V13	2020-10-13 16:28:36	49.3	-1.1	Foraging

We iterate then over data to determine the interval of time in which a SEAL was doing one activity in a specific time. With this process, we managed to get trajectories of dozens of seals with an average length of 25 Time Frame. The results for SEAL V13 is shown in Sequence 1.1:

```
'V13': {'interval': { (1602599096, 1602789332) : {Foraging, 49.4-1.1},
                      (1602789332, 1603096448) : {Resting, 49.3-1.1},
                      (1603096448, 1603561500) : {Foraging, 49.4-1.1},
                      (1603561500, 1603651028) : {Resting, 49.3-1.1},
                      (1603651028, 1603942524) : {Foraging, 49.4-1.1}, ...
       }}
```

Sequence 1.1: Interval-based sequence of seal V13

For the time frames, we used timestamps: the time frame (1602599096, 1602789332) is in fact ('2020-10-13 16:24:56', '2020-10-15 21:15:32').

Consider the example in Sequence 1.2, the sequences represent trajectories of three seals $V21, V23$ and $V24$. In this example, every seal has an interval attribute, this attribute is an interval-based sequence or a trajectory. Seal $V21$ has two time frames, in which he performs the activity *Foraging* in the location (49.4;-1.0) and the activity *Resting* in the location (49.3;-1.1).

```
'V21': {'interval': { (1615400056, 1622243688): {'Foraging', '49.4-1.0'},
                      (1622243688, 1622448596): {'Resting', '49.3-1.1'}
       }}
'V23': {'interval': { (1622448596, 1622719184): {'Foraging', '49.4-0.8'},
                      (1622719184, 1623369484): {'Resting', '49.3-1.1'},
                      (1623369484, 1623548956): {'Foraging', '49.4-0.9'},
                      (1623548956, 1624510800): {'Resting', '49.3-1.1'},
                      (1624510800, 1624659724): {'Foraging', '49.4-1.0'}
       }}
'V24': {'interval': { (1615442280, 1625712540): {'Resting', '49.3-1.1'}
       }}
```

Sequence 1.2: Interval-based sequence of seal V21, V23 and V24

Fig. 6 represents the results of the analysis using GALACTIC of seals $V21, V23$ and $V24$. The result is of form of a concept lattice, starting with the top concept \$0 containing all three seals, and which have two predecessors, \$1 and \$2. \$1 contains two seals ($V21$ and $V24$), and it describes these two seals by: they both have activity *Resting* in the location (49.3;-1.1) between 2021-05-29 at 01:14:28 and 2021-05-31 at 10:09:56.

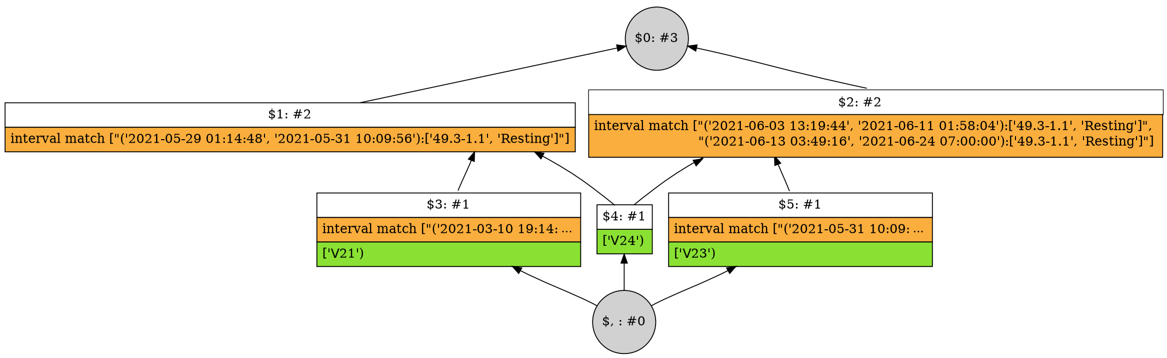


Fig. 6: Results of `GALACTIC` for SEALS V21, V23 and V24

5 Experiments

Exploiting moving object trajectories is the key to discovering knowledge about behaviors of moving objects, which helps in decision-making. To exploit moving object trajectories, we determine the properties of the moving object. The significant difference between these properties is the change of the dive shape (TAD) which help to define the behavior of moving objects.

The main objective of our system is to automatically classify seals as groups doing activities in the same place at the same time. In our experiment, we consider one real seal's trajectory data captured in 2020 and 2021. We have 10 000 captured data in form of points available as CSV files. Related to the experts' interests, we focus only on the seal $Foraging, Resting$ activities in this experiment.

Using the procedure described in Subsection 4.3, we managed to transform the raw data to seal trajectories that can be analysed using `GALACTIC`. We used for this analysis the Maximal Common Time Frame (MCTF) description and the Augmented Minimum Cardinality (AMC) strategy. Fig 7 shows the resulting concept lattice generated by `GALACTIC`. The concept lattice shows many groups where seals performed same activities in the same places, here the results are shown in a compact mode because the length of the image needs more space. We will show results in details using map figures. The concept lattice contains 24 concepts. As we are interested in descriptions with location and activity, we limit the display of predicates to only those with two elements (location, activity). This is why concepts \$1, \$2, \$3, \$4, \$5, \$6 and \$7 contains `interval match []`, these concepts contains in fact a long list of predicates that contains either a location or an activity but not both. Consider concept \$16, here we present a zoom, this concept describes two seals, V23 and V24, these two seals performed *Resting* activity at location (49.3;-1.1) between 03/06/2021 and 11/06/2021 and between 13/06/2021 and 24/06/2021. Another fact we can visualize here is that there are two major groups: (V21, 23, 24 and 26) and (V13, V14, V16, V17 and V19). The difference between these two groups was confirmed later by domain experts.

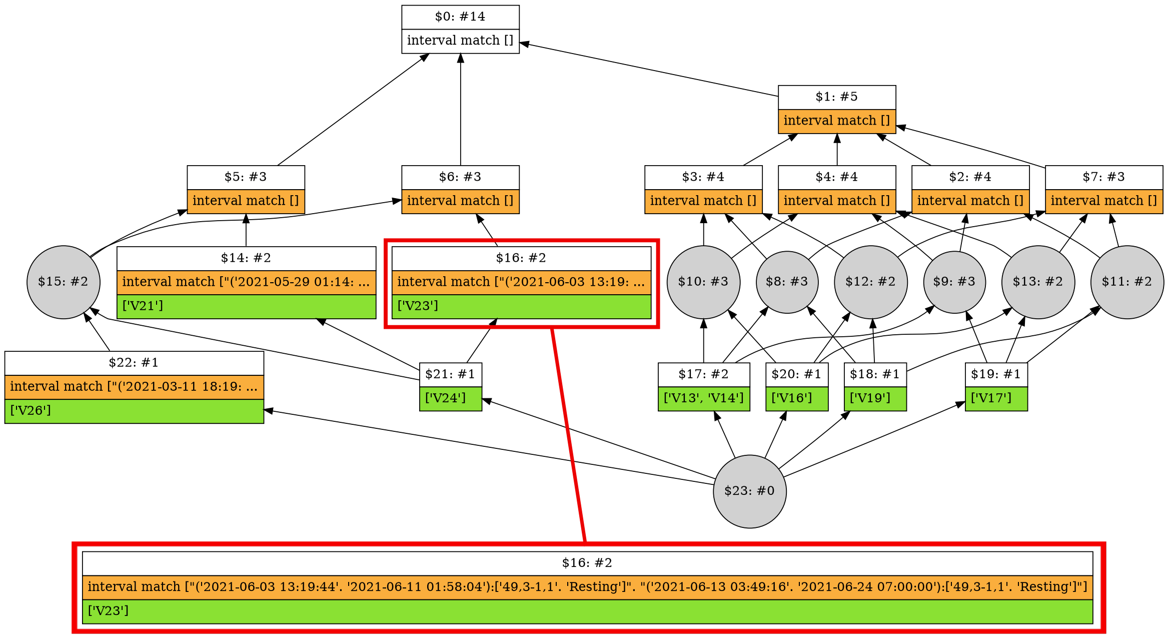
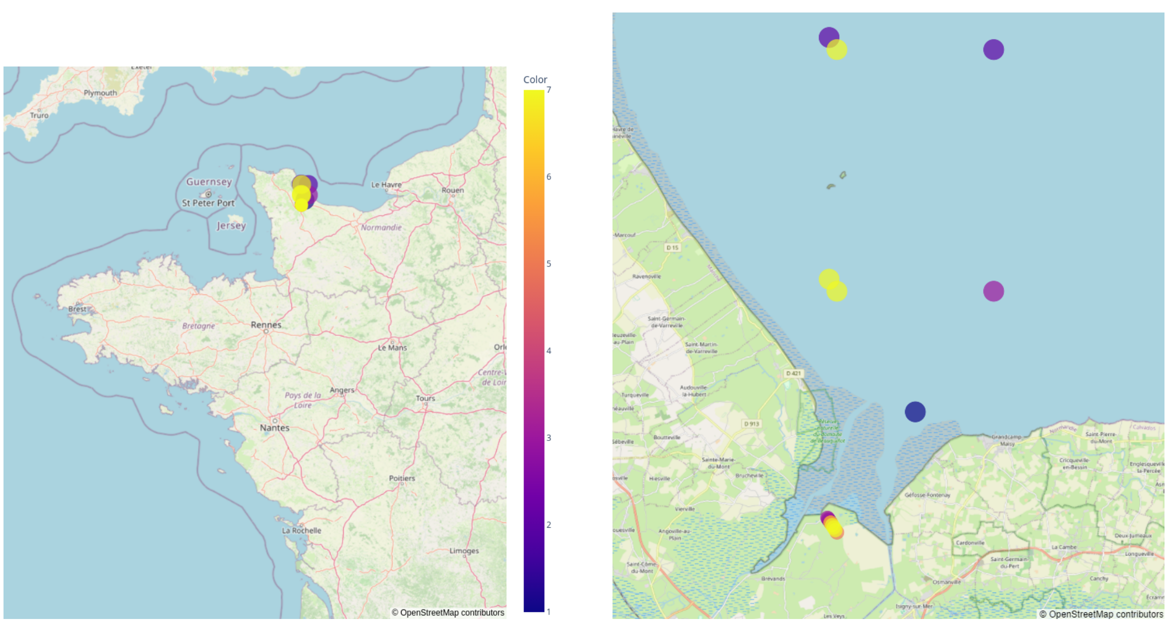


Fig. 7: The concept lattice for the seals trajectories using the AMC strategy and the MCTF description

Fig 8 (a) and its zooming (b) show the geographical map of the seals' activities in Normandie, France. Figs 9 zooms on these activities of the seals during their movements at sea. All activities shown are *Foraging*, which correspond to their feeding activity seat sea. The group of seals doing the activity foraging or resting at the same time in the same place are shown in Table 4.



(a) Map of seals' activities

(b) Zooming map of seals' activities

Fig. 8: Geographical map of the seals' activities in Normandie, France

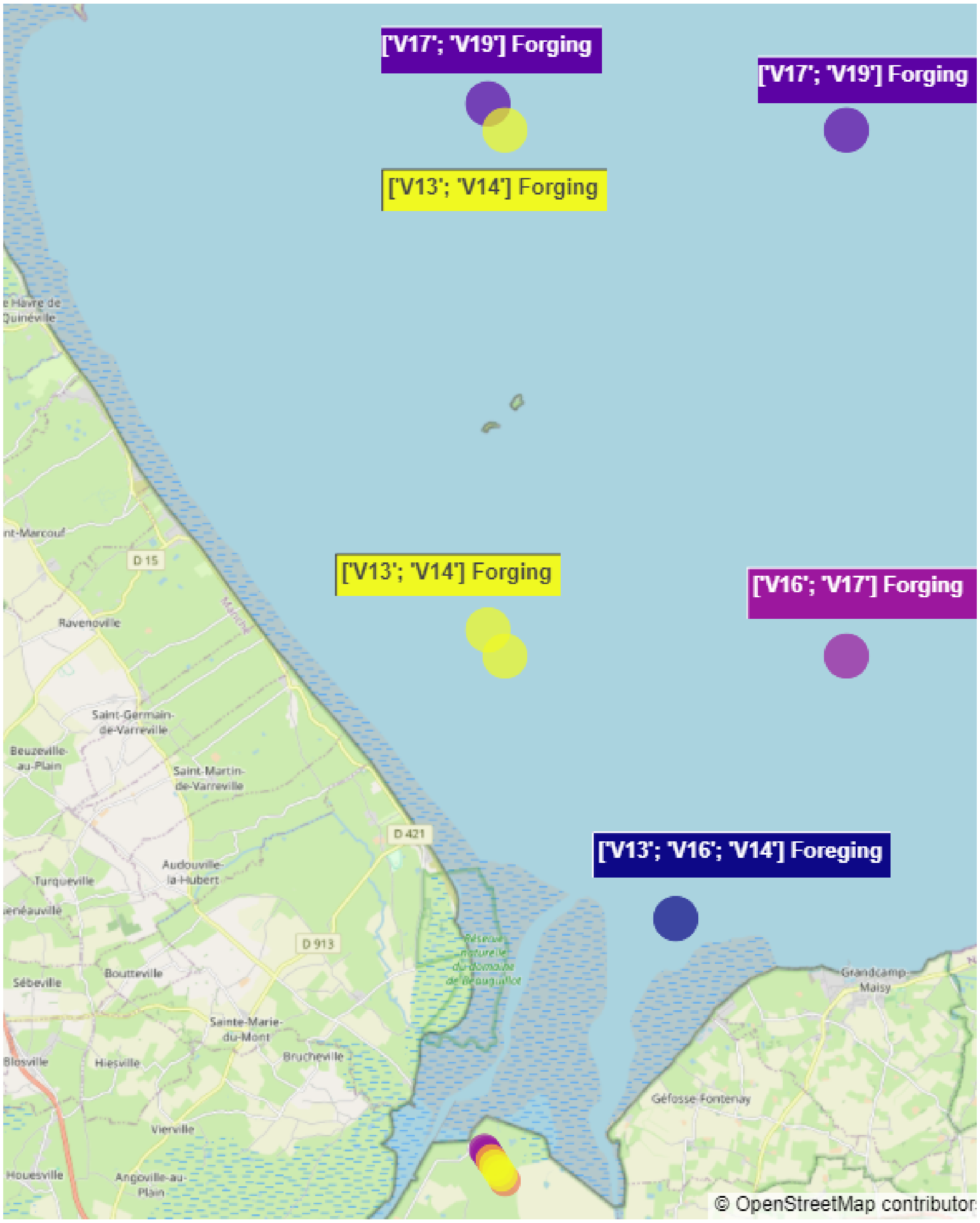


Fig. 9: The activity of the group of seals

The aim of this analysis is to provide a relation between a group of seals, their common activity in a given location at a specific time. The detailed results are show in Table 4 organized as groups, their activities, their location and the activity' time. We have three groups of seals that performed the two activities in different time windows (V17, V19), (V16, V17) and (V13, V14). Three other groups performed *Resting* activity. And one group (V13, V14, V16) performed *Foraging*.

One striking results is that the data are temporally distinct. In fact, for these 14 seals, the domain expert said that *half of the data were fitted with tags in October 2020, and the other half were fitted in March 2021, so the seasonal differences show differences in monitoring periods, and not necessarily in individual rhythms.*

Table 4: Results of extracting subgroups using GALACTIC

Group of seal	**Activity**	**Location**	**Time**
V13, V14, V16	Foraging	49.4-1.1	2020/11/29
V17, V19	Foraging	49.5-1.1	2020/10/17, 30
	Resting	49.3-1.1	2020/10/29
V16, V17	Resting	49.3-1.1	2020/10/22, 24 2020/11/19, 21
	Foraging	49.4-1.0	2020/10/26
V21, V24	Resting	49.3-1.1	2021/05/31
V24, V26	Resting	49.3-1.1	2021/04/08 2021/06/11, 24
V23, V24	Resting	49.3-1.1	2021/04/08, 17
V13, V14	Resting	49.3-1.1	2020/11/27 2020/12/15 2021/01/23, 25, 30 2021/02/06, 25
	Foraging	49.4-1.1 49.5-1.1	2020/10/15, 29, 31 2020/11/07, 14, 29 2021/02/11

6 Conclusion

In this paper, we work on marine mammal (seals) trajectories over the GALACTIC platform. We transformed low-level real world tracking data to high-level semantic trajectories. We defined mobile objects' activities using rules given by the domain expert. The experiment presents semantic enrichment of trajectories with activity rules, then analysing behaviors by extracting subgroups using GALACTIC. We focused in this work for two main activities, `Foraging` and `Resting`.

For future work, we aim at performing more analysis using more data (including more tracked seal data), and may be other moving object's trajectories. One other important aspect we aim to work on is to supply GALACTIC with descriptions and strategies that work directly with real world tracking data without the need to transform, a description in that case may describe the location with a convex hull of the geographical points rather than rounding longitude and latitude.

References

1. Microsoft geolife gps trajectory dataset https://www.microsoft.com/en-us/download/details.aspx?id=52367
2. Allen, J.F.: An interval-based representation of temporal knowledge. In: IJCAI. vol. 81, pp. 221–226 (1981)
3. Ribeiro de Almeida, D., de Souza Baptista, C., Gomes de Andrade, F., Soares, A.: A survey on big data for trajectory analytics. ISPRS International Journal of Geo-Information **9**(2) (2020). https://doi.org/10.3390/ijgi9020088, https://www.mdpi.com/2220-9964/9/2/88
4. Alvares, L.O., Bogorny, V., Kuijpers, B., Macedo, A.F., Moelans, B., Vaisman, A.: A model for enriching trajectories with semantic geographical information. In: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems. pp. 22:1–22:8. ACM (2007)
5. Baglioni, M., Macedo, J., Renso, C., Wachowicz, M.: An ontology-based approach for the semantic modelling and reasoning on trajectories. In: Advances in Conceptual Modeling - Challenges and Opportunities, vol. 5232, pp. 344–353. Springer Berlin/Heidelberg (2008)
6. Barbut, M., Monjardet, B.: Ordres et classifications : Algèbre et combinatoire. Hachette, Paris (1970), 2 tomes
7. Bisone, F., Devogele, T., Etienne, L.: From raw sensor data to semantic trajectories. In: Proceedings of the 5th ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management. EM-GIS '19, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3356998.3365777, https://doi.org/10.1145/3356998.3365777
8. Bogorny, V., Heuser, C., Alvares, L.: A conceptual data model for trajectory data mining. In: Geographic Information Sci., vol. 6292, pp. 1–15. Springer Berlin/Heidelberg (2010)
9. Bordat, J.P.: Calcul pratique du treillis de Galois d'une correspondance. Mathématiques et Sciences humaines **96**, 31–47 (1986)

10. Boukhetta, D.C., Bertet K., R.J.: Sequence mining using FCA and NextPriorityConcept algorithm. In: The 15th International Conference on Concept Lattices and Their Applications (CLA 2020) (2020)
11. Boukhetta, S.E., Demko, Ch., Richard, J., Bertet, K.: Sequence mining using fca and the NextPriorityConcept algorithm. In: Concept Lattices and Their Applications 2020. vol. 2668, pp. 209–222 (2020)
12. Boukhetta, S.E., Richard, J., Demko, C., Bertet, K.: Interval-based sequence mining using fca and the nextpriorityconcept algorithm. In: FCA4AI@ ECAI. pp. 91–102 (2020)
13. Buzmakov, A., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A., Raïssi, C.: Fca and pattern structures for mining care trajectories (2013)
14. Buzmakov, A., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A., Raïssi, C.: On projections of sequential pattern structures (with an application on care trajectories) (2013)
15. Codocedo, V., Bosc, G., Kaytoue, M., Boulicaut, J.F., Napoli, A.: A proposition for sequence mining using pattern structures. In: International Conference on Formal Concept Analysis. pp. 106–121. Springer (2017)
16. Cruz, C.: Semantic trajectory modeling for dynamic built environments. pp. 468–476 (10 2017). https://doi.org/10.1109/DSAA.2017.79
17. Damiani, M.L.: Semantic trajectories data models. Handbook of Big Geospatial Data (2020), https://api.semanticscholar.org/CorpusID:215752560
18. Demko, Ch., Bertet, K., Faucher, C., Viaud, J.F., Kuznetsov, S.O.: NextPriorityConcept: A new and generic algorithm computing concepts from complex and heterogeneous data. Theoretical Computer Science **845**, 1 – 20 (2020)
19. Demko, C., Bertet, K., Faucher, C., Viaud, J.F., Kuznetsov, S.: Next priority concept: A new and generic algorithm computing concepts from complex and heterogeneous data. arXiv preprint arXiv:1912.11038 (2019)
20. Fedak, M.A., Lovell, P., Grant, S.M.: Two approaches to compressing and interpreting time-depth information as collected by time-depth recorders and satellite-linked data recorders. Mar Mamm Sci pp. 94–110 (2001)
21. Ferré, S.: Reconciling Expressivity and Usability in Information Access - From Filesystems to the Semantic Web. Habilitation, Univ. of Rennes 1, France (november 2014)
22. Ferré, S.: Reconciling expressivity and usability in information access from file systems to the semantic web (2014)
23. Ferré, S., Ridoux, O.: A logical generalization of formal concept analysis. vol. 1867, pp. 371–384 (03 2000)
24. Ferré, S.: Systèmes d'information logiques : un paradigme logico-contextuel pour interroger, naviguer et apprendre. Doctorat, Univ. of Rennes 1, France (Oct 2002)
25. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In: LNCS of International Conference on Conceptual Structures (ICCS'01). pp. 129–142 (2001)
26. Ganter, B., Wille, R.: Formal Concept Analysis, Mathematical foundations. Springer Verlag, Berlin (1999)
27. Gizdatullin, D., Baixeries, J., Ignatov, D.I., Mitrofanova, E., Muratova, A., Espy, T.H.: Learning interpretable prefix-based patterns from demographic sequences. In: International Conference on Intelligent Data Processing: Theory and Applications. pp. 74–91. Springer (2016)
28. Gizdatullin, D., Ignatov, D., Mitrofanova, E., Muratova, A.: Classification of demographic sequences based on pattern structures and emerging patterns. In: Supplementary Proceedings of 14th International Conference on Formal Concept Analysis, ICFCA. pp. 49–66 (2017)
29. Güting, R., Schneider, M.: Moving Objects Databases. Morgan Kaufmann (2005)
30. Kaytoue, M.: Contributions to Pattern Discovery. Habilitation, University of Lyon, France (february 2020)
31. Kaytoue, M., Codocedo, V., Buzmakov, A., Baixeries, J., Kuznetsov, S.O., Napoli, A.: Pattern structures and concept lattices for data mining and knowledge processing. In: In Proceedings of ECML-PKDDl (2015)
32. Kaytoue, M., Duplessis, S., Kuznetsov, S.O., Napoli, A.: Two fca-based methods for mining gene expression data. In: International Conference on Formal Concept Analysis. pp. 251–266. Springer (2009)
33. Kontarinis, A., Zeitouni, K., Marinica, C., Vodislav, D., Kotzinos, D.: Towards a semantic indoor trajectory model: application to museum visits. GeoInformatica **25** (04 2021). https://doi.org/10.1007/s10707-020-00430-x
34. "Li-Hua Yin", H.L.: Searching activity trajectories with semantics (2019). https://doi.org/10.1007/s11390-019-1942-8, https://jcst.ict.ac.cn/en/article/doi/10.1007/s11390-019-1942-8
35. Linding, C.: Fast concept analysis. In: Working with Conceptual Structures-Contributions to ICC. pp. 235–248 (2002)
36. M. Santipantakis, G., Doulkeridis, C., A. Vouros, G.: An ontology for representing and querying semantic trajectories in the maritime domain. In: Advances in Databases and Information Systems: 27th European Conference, ADBIS 2023, Barcelona, Spain, September 4–7, 2023, Proceedings. p. 224–237. Springer-Verlag, Berlin, Heidelberg (2023)
37. Matthew, P.: A framework to support spatial, temporal and thematic analytics over semantic web data. Ph.D. thesis, Wright State University (2008)
38. Mello, R., Bogorny, V., Alvares, L., Santana, L., Ferrero, C., Frozza, A.A., Schreiner, G., Renso, C.: Master: A multiple aspect view on trajectories. Transactions in GIS (05 2019). https://doi.org/10.1111/tgis.12526

39. Mireku Kwakye, M.: Conceptual model and design of semantic trajectory data warehouse. International Journal of Data Warehousing and Mining **16**, 108–131 (07 2020). https://doi.org/10.4018/IJDWM.2020070106
40. Oueslati, W., Tahri, S., Limam, H., Akaichi, J.: A systematic review on moving objects' trajectory data and trajectory data warehouse modeling. Computer Science Review **47**, 100516 (2023). https://doi.org/https://doi.org/10.1016/j.cosrev.2022.100516, https://www.sciencedirect.com/science/article/pii/S1574013722000508
41. Planque, Y.: Écologie trophique de deux espèces sympatriques de phoques en périphérie de leur aire de répartition. Theses, Université de La Rochelle (Nov 2021), https://theses.hal.science/tel-03720984
42. Spaccapietra, S., Parent, C., Damiani, M., Demacedo, J., Porto, F., Vangenot, C.: A conceptual view on trajectories. Data and Knowledge Engineering pp. 126–146 (2008)
43. Spaccapietra, S., Parent, C., Macedo, J., Porto, F., Vangenot, C.: A conceptual view on trajectories. Data Knowledge Engineering **65**, 126–146 (04 2008). https://doi.org/10.1016/j.datak.2007.10.008
44. Vidal-Filho, J.N., Times, V.C., Lisboa-Filho, J., Renso, C.: Towards the semantic enrichment of trajectories using spatial data infrastructures. ISPRS International Journal of Geo-Information **10**(12) (2021). https://doi.org/10.3390/ijgi10120825, https://www.mdpi.com/2220-9964/10/12/825
45. Wannous, R., Malki, J., Bouju, A., Vincent, C.: Modelling mobile object activities based on trajectory ontology rules considering spatial relationship rules. In: Modeling Approaches and Algorithms for Advanced Computer Applications, Studies in Computational Intelligence, vol. 488, pp. 249–258. Springer International Publishing (2013)
46. Wannous, R., Malki, J., Bouju, A., Vincent, C.: Time integration in semantic trajectories using an ontological modelling approach. In: New Trends in Databases and Information Systems, Advances in Intelligent Systems and Computing, vol. 185, pp. 187–198. Springer Berlin Heidelberg (2013)
47. Wannous, R., Malki, J., Bouju, A., Vincent, C.: Trajectory ontology inference considering domain and temporal dimensions—application to marine mammals. Future Generation Computer Systems **68**, 491–499 (2017). https://doi.org/https://doi.org/10.1016/j.future.2016.01.012, https://www.sciencedirect.com/science/article/pii/S0167739X1600025X
48. Wille, R.: Restructuring lattice theory : an approach based on hierarchies of concepts. Ordered sets pp. 445–470 (1982), i. Rival (ed.), Dordrecht-Boston, Reidel.
49. Wu, T., Qin, J., Wan, Y.: Tost: A topological semantic model for gps trajectories inside road networks. ISPRS International Journal of Geo-Information **8**(9) (2019). https://doi.org/10.3390/ijgi8090410, https://www.mdpi.com/2220-9964/8/9/410
50. Xu, J., Lu, H., Güting, R.: Understanding human mobility: A multi-modal and intelligent moving objects database. pp. 222–225 (08 2019). https://doi.org/10.1145/3340964.3340975
51. Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K.: SeMiTri: A framework for semantic annotation of heterogeneous trajectories. In: Proceedings of the 14th International Conference on Extending Database Technology. pp. 259–270. ACM (2011)
52. Yan, Z., Parent, C., Spaccapietra, S., Chakraborty, D.: A hybrid model and computing platform for spatio-semantic trajectories. In: The Semantic Web: Research and Applications, pp. 60–75. Springer Berlin/Heidelberg (2010)