Stop and Move Semantic Trajectory Clustering with Existing Spatio-Temporal Data Model

Fatima J. Muhdher
Department of Computer Sciences
King Abdul-Aziz University
Jeddah, Kingdom of Saudi Arabia
Email: fmuhdher [AT] stu.kau.edu.sa

Jonathan Cszalas
Department of Computer Scinces
King Abdul-Aziz University
Jeddah, Kingdom of Saudi Arabia

Abstract— **Due to the phenomenal growth of spatio-temporal datasets, trajectory mining, or specifically, trajectory clustering, is used in order to manage, understand and analyze the existing spatio-temporal data of moving objects. The spatio-temporal data model has the ability for processing historical queries, discovering similar patterns or patterns with specific features/properties, discovering the interesting places within a trajectory, and discovering common trajectories and sub-trajectories. In this work, we combine a stop and move semantic trajectory clustering method to an existing spatio-temporal data model. We will use this clustering method, as it is clusters individuals within a single trajectory, not only according to the geometric properties of the trajectory, such as the velocity and direction, but rather for its dependence on the semantics of the trajectory by looking inside the stop points. This helps to understand the semantics of the trajectory and the behavior of the moving objects. The efficacy of this method is then shown using two real, trajectory datasets.**

Keywords-Spatio-Temporal Data Model; Moving Objects; Trajectory Mining; Trajectory Clustering; Trajectory Semantic; Stop And Move Clustering.

I. INTRODUCTION

The usage of positioning devices and applications, in mobiles and vehicles, induces a lot of geospatial applications. Thus, Moving Object Databases (MODs) will greatly impact Geospatial Information Systems (GIS), as they give the user the ability for modeling trajectories and continuous movements. There are massive amounts of moving objects, as well as their trajectories and other information regarding their movements, that need managing and analyzing in order to be ready for spatio-temporal analysis. These databases, often referred to as spatio-temporal databases, are dealing with the geometries of the moving objects that are changing continuously over time.

The phenomenon of rapidly growing geo-reference and spatio-temporal datasets is due to the now ubiquitous, daily transactions, which provide spatio-temporal data, such as the traffic management of the transformation networks. With the availability of large number of trajectories within spatio-temporal data, knowledge discovery is often the most critical (key) step, as these trajectories have their own semantic and detail information. And each point in the trajectory could have its own details as well. Newer works focus on trajectory mining, the knowledge discovery from these spatio-temporal datasets which identifies valuable information such as similar patterns or patterns with specific properties.

Since the trajectory is a collection of sample points/positions, where each position has its own semantic properties, it is difficult to analyze and understand these trajectories from the user's point of view. Accordingly, there are several works related to trajectory data analysis and trajectory mining. In general, trajectory mining, and especially trajectory clustering, is useful in many aspects: in processing of historical queries, discovering similar patterns or patterns with specific features/properties, discovering the interesting places within a trajectory, and discovering common trajectories and sub-trajectories.

There are some important questions that need to be answered in trajectory mining:

1. What are the kinds of patterns that can be inferred from the trajectory dataset?

2. What is the most suitable form of trajectory mining that will be able to extract those patterns?

One of the most important trajectory mining algorithms is trajectory clustering, which can discover groups of similar trajectories according to similarity measure, interesting locations, and places within a single trajectory, moving object behavior, or trajectory semantic. The kind of information and patterns that will be inferred from the trajectory dataset will be determined according to the used clustering method and the kind of data used in said clustering method.

Recently, clustering algorithms are being used as an active research area, specifically with trajectory data models. Some of the clustering method concerns are regarding the similarity of a specific context and in a specific domain. Some of these works give the user the ability to adjust the similarity threshold [1], while others are concerned with discovering sub-trajectories [2].

The other new type of trajectory clustering is related to clusters that stop and move within a single trajectory in order to discover the interesting places in the trajectory. The CB-SMot method [3] is a trajectory clustering method utilizing the velocity of the moving object, which, in turn, clearly depends on the stops and moves of a trajectory. Another work of trajectory clustering is DB-SMoT [4], which is a trajectory clustering method, accounting for the direction of the moving object. The work in [5] is combining both methods (CB-SMoT and DB-SMoT) with a knowledge base in order to understand the semantics of the trajectory and the moving object's behavior inside the stop points. This clustering method is outlined in the following spatio-temporal model.

A. The Spatio-Temporal Data Model

This section presents the spatio-temporal data model that is proposed in [6], through which we would like to enhance the stop and move semantic trajectory mining method.

The target of developing this spatio-temporal model is to extend a spatio data model to be used with constrained transportation networks, where all static and moving objects in the model are based on the Open Geospatial Consortium (OGC) standards. There are many applications of moving objects that are constrained by transportation networks, especially in Geospatial Information Systems (GISs). The constructed network of this spatio-temporal data model conforms to the Geographic Data Files (GDF) standard. Moreover, in this work, they extend a query language in order to manage, retrieve, query, manipulate, and analyze this spatio-temporal data. In this model, they are concerned with the geometrics, shapes and extents, which are not variable during the movement time. Additionally, they are focused on the history of the moving objects in OGC-based ORDBMSs. The benefit of building this model based on the OGC is the ability to analyze the spatio-temporal data.

The network in this model is constructed as a set of sections, routs, and junctions between the routs. The section is the main, basic element in the network, wherein the network graph structure is stored. The routes can be either unidirectional or bidirectional, which make it, in some cases, necessary to indicate the positions of both sides of the route. Finally, the he junction is the intersection of two or more routes.

The data types on the network can represent static (GPOINT and GLINE) or moving objects (MGPOINT). The GPOINT is related to positions in the network such as gas stations, while GLINE is related to regions on the network, such as regions in the network with a specific speed limit. Additionally, there are some essential operations on GPOINT and GLINE that are used for crating and updating the network structure graph. The other operations are used to give the user the ability to get some information about the network as shown in Figure 1. These operations are categorized by two sets;

1) A set of operation to access the information of the routes:

- LENGTH operation: It is used to get the length of a specific route.
- CURVE operation: It is used to get the LRS geometry of a specific route.
- DUAL operation: It is used to get the type of the route.

2) A set of operations to check the topological relationships between the network and the data types:

- ON_ROUTE: Where A GPOINT can belong to a route.
- INTERSECTS: A GLINE value can intersect a route.
- CONTAINS: A GLINE value can contain a route.
- IS_CONTAINED: A GLINE value can be contained in a route.

LENGTH (int network_id, int route_id)	→ Float
CURVE (int network_id, int route_id)	→ OGC-based-geometry
DUAL (int network_id, int routeid_id)	→ int (kind of the route)
ON_ROUTE (gpoint geom, int route_id)	→ Boolean
INTERSECTS (gline geom, int route_id)	→ Boolean
CONTAINS (gline geom, int route_id)	→ Boolean
IS_CONTAINED (int route_id, geom gline)	→ Boolean

Figure 1. User Operations on static objects

The data type MGPOINT is related to the moving objects in the network. There are also a set of operations on the moving objects as shown in Figure 2. They give the user the ability to get information about the network and the moving object:

- IN_SPACE and IN_NETWORK operations: They are used to convert between the network values and the spatial values.
- INST and VAL operations: They are used to get two components based on an Intime value.
- DEFTIME operation: It is used to get all the time intervals that are defined with an object.
- TRAJECTORY operation: It is used to construct the moving object's path of the network.
- AT_INSTANT operation: It is used to get the position of a moving object in a specific time.
- AT_PERIODS: It is used to return the region of the network where the moving object travels in a specific duration of time.
- AT operation: It is used to restrict a function to specific times when the case is that its value is included in the second argument.
- DIRECTION operation: It is used to get the direction of a movement in at a specific time.
- INSIDE operation: It is used to check if the moving object exists in a specific region of the network at a specific time.
- SIZE operation: It is used to calculate the length of a specific trajectory.
- DURATION operation: It is used to compute spend time during the movement of a moving object.

- CURRENT and NOW: They are used to get the last position and the last stored time that is related to a specific moving object on the network.

IN_SPACE (gpoint or gline geom) → OGC-based-geometry
IN_NETWORK (OGC-based-geometry geom) → gpoint
VAL (intime position) → gpoint
INST (intime position) → timestamp
DEFTIME (Ugpoint positions) → periods
TRAJECTORY (Ugpoint positions) → gline
ATINSTANT (int moid, timestamp time) → intime
ATPERIODS (int moid, periods time) → Ugpoint
DIRECTION (int moid1, int moid2, timestamp time) → Float
SHORTEST_PATH (int moid1, int moid2, timestamp time) → gline
AT (int moid, gline geom) → Ugpoint
INSIDE (int moid, gline geom, timestamp time) → Boolean
SIZE (gline geom) → Float
DURATION (Ugpoint geom) → Float
NOW (int moid) → timestamp
CURRENT (int moid) → gpoint

Figure 2. Operations on moving objects

Because of the space limit, the details of the network model, data types, and operations are detailed in [6].

These operations on static and moving objects will give the user the ability to build complicated algorithms, including, but not limited to, shortest path algorithms and stop and move semantic trajectory clustering algorithms.

The proposed work of this spatio-temporal data model includes the inclusion of additional trajectory mining techniques. We added the stop and move semantic trajectory clustering to enhance the model, thereby allowing it to be analyzed according to the geometric properties of a trajectory. Some examples can be velocity and direction, and according to a knowledge base that consists of a set of rules in order to understand the semantic of the trajectory and the behavior of the moving object. After adding this method to the spatio-temporal data model, it will have the ability to find interesting places within the network and the ability to analyze the behavior of a group of moving objects, as well as the ability to discover similar or specific patterns. The main drawback of this method is that the one stop could satisfy more than one rule in the knowledge base, resulting in ambiguity, otherwise referred to as "noise".

In general, our contributions are as follows:

- We add a clustering technique to follow up the future work of the spatio-temporal data model in order to enhance its ability to analyze the spatio-temporal data.
- Clustering the trajectory on this spatio-temporal model not only by using the geometric properties of the trajectory, but also with knowledge base, which represents the semantic information of the trajectory.
- We use the model in constructing knowledge bases according to the analysis of the trajectory semantics and the moving object's behavior.

The reminder of this paper is organized as follows. Section 2 presents the related works of our research. The details of combining the stop and move semantic trajectory clustering method to the spatio-temporal data model is discussed in Section 3. Section 4 presents the results, and, finally, the conclusion and potential future works are in Section 5.

II. Related Works

This section presents some related works about moving objects in term of modeling, querying, data generators, and some other related research in this field.

A. *Modeling and Querying*

There are two important research areas in the field of moving objects databases: the location management, and the spatio-temporal database [7]. The location management is related to the predictive queries and to the queries that are related to the current position of the moving objects. The most important works in this area are on the Moving Object's Spatio-Temporal (MOST) model [8] and the Future Temporal Logic (FTL) language [9]. The spatio-temporal database related to the trajectories and to the historical queries of the moving objects as in [10][11] where the complete evolution of the moving objects are represented as a database's attribute. The spatio-temporal data model that will be used in this work is related to the second research area which is related to the history of the moving objects in OGC-based ORDBMSs.

The European standard Geographic Data Files (GDF) [12] is one of the most important and major standards of spatial network models where it is describing the road networks and the road data. In this standard, the road network consists of a set of (Roads), where the Roads consists of (Road Elements) which having a (Junction) at each end. The Junction could be located at two or more road centerlines' intersection. According to this standard, the road network model is constructed of a structure of three levels:

Level 0: It is the (Topology) level where everything in the network is represented as (nodes and edges).

Level 1: It is the (Features) level where all simple features like (the Road Elements and the Junctions) and their attributes like (one way, the road width and the number of lanes) that are related to the network have been described.

Level 2: It is the (Complex Features) level where it aggregates the "simple features" together to build a higher-level feature.

There are some important works related to this context. One of those works is (DEDALE) [13] which is a spatial database system that depends on a constraint-based model for representing and manipulating the geometric and the spatio-temporal data. In this work, the objects moves are restricted by an embedded network. Another work [14] is about modeling and querying of moving objects within a road. In this work, a usual graph model is used to model the networks that consist of relations and edges.

Moreover, there are some researches [15][16][17] about the issues of the data modeling in special networks. The work of Jensen and his colleagues was about the real road networks and its complexity, where the using of simple directed graph is not sufficient to modeling those complicated networks.

The data model of moving objects in a network and its query language that is proposed in [10], has some interesting and important features. First, the positions of a moving object is described according to the network, which doesn't mean that this data model is using Linear Referencing System (LRS) to represent the positions, rather than using the embedding space with the (x, y) points. There are two advantages of using the LRS where the geometry of any point in the network will be stored only once. Moreover, it is much simpler to find the connection between a moving object with any part of the network which is done by checking the existence of id of the route id in the description of the movements. Second, the users have the ability to extend some information and controls of the network with standard tools of the DBMS, such as controlling the speed limit and the route type within a part of the network. Besides, the user can define some operations such as the shortest path. Moreover, they have the ability to describe the static or moving objects on the network, such as gas stations and vehicles. The spatio-temporal data model in [6], which will be used in this work, uses the mentioned advantages of that work [10], with some changes according to the common points and standards between the different DBMS models, while the work in [10] is implemented with the SECOND DBMS.

The HERMES system is proposed in [18][19]. To support the Oracle data management, the STOC (Spatio-Temporal Object Cartridge) was proposed in [20] as an Oracle extension.

B. *Data Generators*

In general, the data generators are used to generate a dataset of moving objects within a transportation network. There are some known moving objects data generators such as (GSTD, OPORTO and SUMO) that are proposed in [21][22][23], while the most popular and important data generator is proposed in [24]. It provides data about moving objects that are constrained within a specific transportation network. This work has some advantages. First, it is an open source data generator with GUI of the moving objects within the network. Second, the loading of the network edges controls the speed and the route of the vehicles.

The proposed work in [25] uses the SECOND DBMS in order to generate the data model rather than developing a new one. The problem with this work is that it is related to moving objects that are moving freely within a 2D plane and not constrained by a transportation network. Now, they are working on supporting the moving within a transportation network with the SECOND data model.

C. *Other Research Areas*

This section presents some research areas regarding MODs.

1) *Moving Object Trajectory Uncertainty*

There are many sources that are affecting and generating the uncertainty of the moving object trajectories, such as imprecise positions of the positioning applications and devices. There are some researches in this area such as the works that are proposed in [26][27][28] regarding the uncertainty modeling, managing, and querying of uncertainty of the moving objects.

2) *Spatial Trajectories Indexing*

In order to retrieve the trajectory data, such a moving object's traveling history in MODs (that contains a huge number of trajectories in an efficient way) relies on the use of the appropriate indexing method (Trajectory indexing). In this research area, there are some researched works on developing and implementing the indexing structures, such as the works proposed in [29][30][31][32].

3) *Trajectory Mining and Moving Objects Ontology (MOO)*

Trajectory mining in MODs is an interesting research area. The trajectories have their own semantics as they have some geometric properties. From the user's point of view, it is difficult to analyze and discover the information from a trajectory. Thus, there is a need to develop data models and tools in order to discover meaningful and similar patterns from a trajectory data. There are some applications that have been developed in this area such as the work in [33]. The work that is proposed in [34] is a query language of trajectory data mining. Moreover, there are some other works proposed in [35][36][37] about discovering similar trajectories and trajectory sample points mining. Besides, there are some researches about trajectory similarity such as the works on [38][39][40].

There are some works [41][42][43] that are emerging in the GIS in order to enhance the analysis and modeling of trajectories of moving objects. Another work [44] was about trajectory pattern discovery. The work in [45], concerning the discovery of the dynamic aspects that are related to a moving object with descripting the real-world occurrence's semantics. And there are other studies in the MOO research domain, which are proposed in [46][47].

III. COMBINING DB-SMOT AND CB-SMOT WITH A KNOWLEDGE BASE

This section is presenting the main parts and components of combining DB-SMoT and CB-SMoT with a knowledge base within the "spatio-temporal data model for moving object database" [6].

A. *Stops and Moves*

The following characteristics of the stops and moves are related to the Spaccapietra [r1] and with the combining of DB-SMoT and CB-SMoT with a knowledge base.

Stop: Which is the most important part of the trajectory where:

- The user defined this part explicitly as a stop [r1].
- The moving object does not move for a period greater than the MinTime threshold that is predefined by the user.
- The changing of the direction of the moving object is greater than the MinDirection threshold.
- The changing of the velocity of the moving object is greater than the MinVelocity threshold.
- The stops in the same trajectory have to be temporally disjointed [r1].

Move: Which is a part of the trajectory where:

- The part of the trajectory that is delimited by [r1]:

a) Two consecutive stops.

b) The beginning time and the first stop.

c) The last stop and the end time.

d) [tbegin, tend], which is the case when there are no stops in the trajectory.

- The (tbegin) is the trajectory duration initial point tend is the end point [r1].

B. Basic Definitions

This section will presents some basic definitions on the stop and move clustering methods. These definitions depend on the context of the application where the method is being used.

Definition 1: *SGPOINT (Stop GPOINT).*

It is the position where the moving object stops at time ti. It is a GPOINT data type but with some geometric properties (VelocityVar, DirectionVar and TimeDur) that are related to the moving objects at the time ti, when it stops at this position. Where, i = 0,...., N. Moreover, the goal of letting the moving object stop at this position will be added to this datatype.

***Definition* 2:** Sub trajectory (SUBTRAJECTORY).

It is a list of the positions where a moving object stops <SGPOINT0, SGPOINT2,....., SGPOINTN>.

Where, i = 0,...., N and t0 < t1 < ... < tN

After determining the target trajectory that the method will cluster, the predefined stops by the user during the building of the network, will be added, along with the SUBTRAJECTORY, if their positions are with this trajectory.

***Definition* 3:** Cluster (CLUSTER).

It is a cluster sub trajectory (SUBTRAJECTORY) of the trajectory T with respect to the system thresholds (MinTime, MinVelocity, MinDirection and MaxTolerance)

C. Thresholds

There are four thresholds within this technique; MinTime, MinVelocity, MinDirection, and MaxTolerance. The user controls those thresholds in order to specify them according to the context of the target application.

- *MinTime:* It is the minimum duration of time for a stop at position pi in order to consider it for stop (SGPOINT).
- *MinVelocity:* It is the minimum change in the velocity of a moving object at a position pi in order to consider it as a candidate position for stop (SGPOINT).
- *MinDirection:* It is the minimum change in the direction of a moving object at a position pi in order to consider it as a candidate position for stop (SGPOINT).
- *MaxTolerance:* It is the maximum number of candidate stops that could be found consecutively in a cluster. The benefit of using this threshold is to determine if the change in the direction of the moving object is a noise or the change in the direction just ends.

Those thresholds are computed by using the operations in the spatio-temporal data model in [6].

D. CB-SMoT Clustering Method

The CB-SMoT is a clustering method that deals with one trajectory, where the clusters are generated according to the variation of the velocity of a moving object within a single trajectory. In CB-SMoT, the main threshold to find the clusters is the velocity of the moving object. Thus, the greater impact is related to the velocity variation compared to the direction variation, like traffic management applications. The condition of clustering is to have a velocity variation lower than the MinVelocity threshold for the amount of time equal to or greater than MinTime. It is has two main phases:

- ***First Phase:*** Generate clusters as sub trajectories by evaluating each trajectory according to the velocity where it has to be less that the MinVelocity threshold for amount of time equal to or greater than MinTime. Then, labeling them as an unknown stop.
- ***Second Phase:*** Checking the intersection of the clusters, which comes from the first phase, where the candidate stops, which is defined by the user. Then, all the clusters that intersects with a geographic object for amount of time equal to or greater than MinTime will be labeled with geographic object names, while the others will remain labeled as unknown stops.

The details of this method are presented in [3].

E. DB-SMoT Clustering Method

The DB-SMoT is also a clustering method that deals with one trajectory, where the clusters are generated according to the variation of the direction of a moving object within a single trajectory. In DB-SMoT, the main threshold to find the clusters is the direction of the moving object. Thus, it is concerned with the applications that have greater impacts related to the direction variation, more than velocity variation, like bird migration applications. The condition of clustering is to have a direction variation lower than the MinDirection threshold for amount of time equal to or greater than MinTime.

It is has two main phases:

- ***First Phase:*** Generate set of stops according to the variation of the direction of a moving object and the MinDirection threshold.
- ***Second Phase:*** Clustering the stops according to the clustering MinTime and the MaxTolerance thresholds. After that, the positions that did not clustered as a stop will be considered as a move.

The details of this method are presented in [5S].

F. Combining DB-SMoT and CB-SMoT with a Knowledge Base

This method proposes to look inside the stops that are considered as an identification of the important part of a single trajectory. This method is used to analyze the behavior of the moving object within a stop in order to infer the goal of that stop besides the geometric properties of a moving object; Direction and velocity. It has two main phases:

- ***First Phase:*** Evaluating the trajectory according to the geometric properties (velocity and direction)
- ***Second phase:*** Checking the stops within the domain knowledge in order to infer the goal of a specific stop on patterns of changing in velocity and direction.

There are some basic definition related to this method:

Definition 4: (Trajectory Context).

It is a set of the conditions and influences that are used in order to identify the reason (why) behind the stop of the moving object in this position during a specific interval [tbegin, tend].

Definition 5: (Contextualized Stop).

It represents the important (SGPOINTS) of a specific trajectory where the mobile object has been stopped, because of a specific reason, for while where the duration is greater than or equal to MinTime threshold.

Definition 6: (Contextualized Sub-stop).

It is a stop within a sub-trajectory where:

- The goal of this sub-stop is inferred from the set of rules of the knowledge base.
- This goal is considered as a sub-goal of the main goal of the contextualized stop which is representing the sub-trajectory goal.

The conceptual representation of the semantic trajectory is shown in Figure 3 [5].

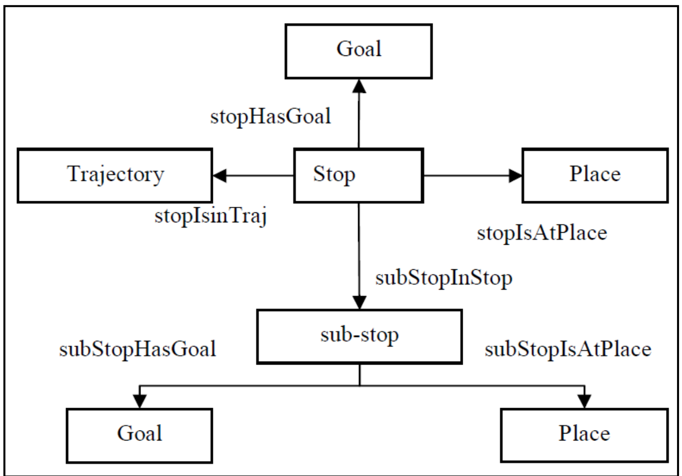

Figure 3. Semantic Trajectory Conceptual Representation

G. Knowledge Base

The usage of the knowledge base is in order to look inside each stop, to understand the behavior of the moving object, along with the objective measures and the geometric properties of the moving object. The semantic information that is stored in the knowledge base is related to the domain of the application where the method is used. The knowledge base is useful to understand the trace of a moving object in order to use that information with decision-making applications such as marketing and urban planning. The knowledge base is a set of (rules and checks) in order to determine if the sub-stop is satisfying one or more of the rules, where each rule is representing a specific goal. The goal of contextualized stop is represented from a summary of a set of goals that are inferred for all the sub-tops of that sub trajectory. An example of a knowledge base that is describing the pedestrians' behavior inside the shopping center is shown in the following Figure 4 [5].

Figure 4. Example of a knowledge base

H. The Method Algorithms

This section presents the algorithms that are used with this method according to the "spatio-temporal data model for moving object database" [6]. Thus, those algorithms will be related to the data types and operations that had been defined in this spatio-temporal data mode.

minTime	maxSpeed	maxDirection	goal
2 hours	0	0	cinema
1 hour	1,5 km/h	20 degrees	shopping
8 hours	0,5 km/h	10 degrees	working

1) ***Algorithm (1):*** *The proposed approach.*

Inputs:

- A set of stops (STRAJECTORY).
- A knowledge base of the application domain (kbase).

Output: A set of the contextualized sub-stops (CLUSTER).

Method:

- Computing the Sub-stops according to the velocity threshold (MinVelocity) by using the CB-SMoT method in order to compute the MinVelocity clusters (sub-stops) for each stop in S.
- Computing the Sub-stops according to the direction threshold (MinDirection) by using the DB-SMoT method. In order to compute the MinDirection clusters (sub-stops) for each stop in S.

- If the list of the sub-stops is not empty, Then
- analyze the sub-stops by using the knowledge base (kBase), by calling the method executeInference.

If the stop does not have sub-stops, the stop will be the input of the executeInference method as a set of sub-stops. The pseudo code of this algorithm is shown in Listing 1.

2) ***Algorithm (2):*** *The executeInference method.*

Input:

- The set of sub-stops (STRAJECTORY).
- The knowledge base (kBase).

Output: A set of the contextualized sub-stops (CLUSTER).

Method:

- Compute for each stop s included in sub-stops:

a) The Time duration of that stop as (TimeStop).

b) The Velocity of that sub-stop, which had been computed previously by CB-SMoT.

c) The Direction Variation of that sub-stop, which had been computed previously by DB-SMoT.

- Compute for each rule within the knowledge base:

a) The maximum speed of the rule,

b) The maximum direction variation of the rule,

c) The minimum time of the rule.

- Comparing the Velocity and the Direction of the sub-stop with maximum direction and maximum speed with the rule.
- Then, if the velocity and direction variations of the sub-stop are lower than or equal to the speed and direction variations of the rule, respectively. The method will test the time of the sub-stop and the rule.
- Then, if the time duration of that stop is greater than or equal to the minimal time of the rule, then the method found the goal of that sub-stop in the knowledge base.
- After that, the contextualized sub-stop will be added to the set of the contextualized sub-stops.

The complexity of the algorithm depends on the number of stops as (P), sub-stops as (SP) and the number of rules in the knowledge base as (R). The complexity of the CB-SMoT and the DB-SMoT is the O(P) and the complexity of the executeInference method is O(SP*R). Thus the Complexity of the proposed method is O((2*P) + (SP*R)).

The pseudo code of this algorithm is shown in Listing 2.

```
INPUT:
SUBTRAJECTORY //set of stops (SGPOINT)
kBase //Knowledge Base

OUTPUT: CLUSTER // a set of contextualized substops

METHOD:
SUBTRAJECTORY_1 -> CB-SMoT (SUBTRAJECTORY, MinVelocity, MinTime);
SUBTRAJECTORY_1 -> DB-SMoT (SUBTRAJECTORY, MinDirection, MinTime);
IF (SUBTRAJECTORY_1 != {})
CLUSTER -> executeInference (SUBTRAJECTORY_1, kBase);
END
```

Listing 1. Algorithm 1 (The Proposed Method)

Listing 2. Algorithm 2 (The executeInference method)

IV. RESULTS AND DISCUSSION

This section will present some experiments and evaluation results of the proposed method. The proposed method proved its ability of clustering by performing some experiments on two real datasets; Bird migration and Pedestrian dataset [5]. In this paper, we will present the Pedestrian dataset because it is more related and similar to context of the spatio-temporal data

```
INPUT:
SUBTRAJECTORY //set of SUBTRAJECTORY
kBase //Knowledge Base

OUTPUT: CLUSTER //Set of contextualized substops

METHOD:

FOR each stop s in substops DO
  timeStop = endTime(SGPOINT) - startTime(SGPOINT); //stop duration
  directionStop = getDirectionVariation(SGPOINT); //average dir. of the stop
  speedStop = getSpeedVariation(SGPOINT); //average speed of the stop
  FOR each rule r in kBase DO
    maxDirectionOfRule = getMaxDirection(r); //min direction of this rule
    maxSpeedOfRule = getMaxSpeed(r); //min speed of this rule
    minTimeRule = getMinTime(r); //min time of this rule
    IF (speedStop<=maxSpeedOfRule AND directionStop<=maxDirectionRule)
      IF (timeStop >= minTimeRule)
         s.addGoal(r.getGoal);//add the goal of rule r as goal of s
         CLUSTER -> C + s; //adds s to list of contextualized stops
      ENDIF
    ENDIF
```

model in [6]. In general, with this type of trajectory clustering techniques, the objective of these experiments is to prove that the proposed method has the ability to provide meaningful trajectories and interesting places according to a prior knowledge.

This dataset is generated within a park in Netherlands [m21] where:

- A set of persons in the park had GPS devices and will reports their activity that they would like to do at this park.
- According to those activities, they built the Pedestrian knowledge as shown in Figure 5.
- The behavior of the moving objects in this dataset is characterized by the velocity and the direction of the moving objects.
- The MinTime Threshold is 15 minutes in both cases; walking and cycling).
- The MaxSpeed is adjusted to 7km/h in the case of walking and 36 km/h in the case of cycling.

Figure 6 shows two contextualized trajectories whereas Figure 6 (left) shows a trajectory with one big stop as an input and Figure 6 (right) shows a trajectory with 5 generated sub-stops, where only 4 sub-stops were contextualized in the cases of taking photos and walking. Figure 7 (left) related a single trajectory and Figure 7 (right) shows a trajectory with 3 small sub-stops where only one sub-stop was contextualized.

To validate a semantic trajectory, it has to be evaluated according to the semantic point of view by an expert of the application domain. Moreover, the proposed method shows its effectiveness in clustering a trajectory semantically and its ability to infer the goal and activities of moving objects in individual trajectories by analyzing the behavior of those moving objects through the knowledge base of the application domain.

minTime (min)	maxSpeed (km/h)	maxDirection	Goal
15	7	50	walking
15	36	80	cycling
5	15	90	dog letting
5	4	45	photo
60	2	20	picnic
30	2	20	relaxing
30	20	50	runnig

Figure 5. Knowledge base of possible activities in a park

Figure 6. (left) Stops of one trajectory and (right) Contextualized sub-stops of the same trajectory

Figure 7. (left) Stops of one trajectory and (right) Contextualized sub-stops of the same trajectory

V. CONCLUSION AND FUTURE WORK

The main objective of this paper is to add a trajectory mining, more specifically, a trajectory clustering method to the spatio-temporal data model. We chose the stop and move semantic trajectory clustering method because it is using the sematic of the trajectory with it geometric properties (Velocity and direction). This method uses a knowledge base in order to evaluate the semantics. This method has two main steps. First, it clusters the stops according to the velocity, by using the CB-SMoT method, and the direction, by using the DB-SMoT method. Second, it will clusters all the candidate stops by evaluating them with knowledge base, which is related to the application domain. This clustering method proved its ability as an effective method in clustering two real data sets.

For future works, we intend to implement the stop and move semantic trajectory clustering method with the spatio-temporal data model in order to evaluate our algorithms. Also, we intend to discover if, according to a set of trajectories, the probability of stops that is satisfying more than one rule will help in solving the drawback of this method. Moreover, we intend to add more trajectory mining and trajectory clustering methods to the spatio-temporal data model. Finally, we intend to build a knowledge base from the combining clustering methods with the spatio-temporal data model.

REFERENCES

[1] D. Guo, and W. Cui, "Mining moving objects trajectories in Location-based services for spatio-temporal database update," Geoinformatics, GIS and Built Environment: Geo-Simulation and Virtual GIS EnvironmentsGeoinformatics, GIS and Built Environment: Geo-Simulation and Virtual GIS Environmentsc, 2008.

[2] K. Buchin, M. Buchin, J. Gudmundsson, M. Löffler, and J. Luo, "Detecting Commuting Patterns by Clustering Subtrajectories," Algorithms and Computatio, pp.644-655, 2008.

[3] M. Nanni, and D. Pedreschi, "Time-focused clustering of trajectories of moving objects," Intelligent Information Systems archive, pp. 267 – 289, 2006.

[4] Rocha J. A. M. R., Oliveira G., Alvares L. O. and Bogorny V., "DB-SMoT: A Direction-Based Spatio-Temporal Clustering Method" In Intelligent Systems (IEEE), page 114-119, 2010.

[5] B. Moreno, V. Times, C. Renso, and V. Bogorny, "Looking inside the Stops of Trajectories of Moving Objects," Proceedings: GeoInfo , pp. 9-20, 2010.

[6] H. Hajari, and F. Hakimpour, "A SPATIAL DATA MODEL FOR MOVING OBJECT DATABASES," Database Management Systems (IJDMS), Vol.6, No.1, 1999.

[7] R. H. Güting, and M. Schneider, "Moving objects databases. Academic Press," 2005.

[8] O. Wolfson, B. Xu, S. Chamberlain, and L. Jiang, "Moving Objects Databases: Issues and Solutions," Statistical and Scientific Database Management, pp. 111-122, 1998.

[9] O. Wolfson, A. P. Sistla, S. Chamberlain, and Y. Yesha, "Updating and Querying Databases that Track Mobile Units," Distributed and Parallel Databases, 7(3):257-387, 1999.

[10] R. H. Güting, V. T. de Almeida, and Z., Ding, "Modeling and Querying Moving Objects in Networks," The VLDB Journal, 15(2):165-190, 2006.

[11] J. Lema., L. Forlizzi, R. H. Güting, E. Nardelli, and M. Schneider, "Algorithms for Moving Objects Databases," The Computer Journal 46(6): pp. 680-712, 2003.

[12] CEN European Committee for Standardisation 1995 Geographic Data Files. Working Group 7.2 of CEN Technical Committee 278. Available at http://www.ertico.com/gdf-geographic-data-files. Accessed August 2012.

[13] P. Rigaux, M. Scholl, L. Segoufin, and S. Grumbach, "Building a Constraint-Based Spatial Database System: Model, Languages, and Implementation," Information Systems, 28(6): 563-595, 2003.

[14] M. Vazirgiannis, and O. Wolfson, "A Spatiotemporal Query Language for Moving Objects on Road Networks," Proc. Of the 7th Intl. Symp. On Spatial and Temporal Databases (SSDT), 20-35, 2001.

[15] C. S. Jensen, T. B. Pedersen, L. Speicys, and I. Timko, "Data Modeling for Mobile Services in the Real World," Proc. Of the 8th Intl. Symp. On Spatial and Temporal Databases (SSDT), 1-9, 2003.

[16] C. Hage, C. S. Jensen, T. B. Pedersen, L. Speicys, and I. Timko, "Integrated Data Management for Mobile Services in the Real World,"

Proc. Of the 29th Intl. Conf. On Very Large Databases (VLDB, Berlin, Germany), 1019-1030, 2003.

[17] L. Speicys, C. S. Jensen, and A. Kligys, "Computational Data Modeling for Network- Constrained Moving Objects," Proc. Of the 11th ACM Symp. On Advances in Geographic Information Systems (ACM-GIS, New Orleans, Louisiana), 118-125, 2003.

[18] N. Pelekis, and Y. Theodoridis, "An Oracle Data Cartridge for Moving Objects," 2006, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.101.9971. Accessed September 2010.

[19] N. Pelekis, E. Frentzos, N. Giatrakos, and Y. Theodoridis, "Supporting Movement in ORDBMS," http://www.softnet.tuc.gr/~ngiatrakos/papers/UNIPI-INFOLAB-TR-2010-01.pdf, 2010.

[20] L. Zhao, P. Jin, X. Zhang, L. Zhang, and H. Wang, "STOC: Extending Oracle to Support Spatiotemporal Data Management," presented at the APWeb, pp. 393–397, 2011.

[21] Y. Theodoridis, J. R. O. Silva, and M. A. Nascimento, "On the generation of spatiotemporal datasets," SSD, pp. 147-164, 1999.

[22] J. M. Sagilo, and J. Moreira, "Oporto: A Realistic Scenario Generator for Moving Objects," Geoinformatica, 5(1):71-93, 2001.

[23] D. Krajzewicz , G. Hertkorn, C. Rössel, and P. Wagner, "SUMO (Simulation of Urban MObility); an open-source traffic simulation," proceedings of the 4th Middle East Symposium on Simulation and Modeling (MESM2002), SCS European Publishing House, pp. 183-187, 2002.

[24] T. Brinkhoff, "A framework for generating network-based moving objects," Geoinformatica, 6(2): pp. 153-180, 2002.

[25] C. Düntgen, T. Behr, and R. H. Güting, "BerlinMOD: a benchmark for moving object databases," The VLDB Journal, vol. 18, pp. 1335-1368, 2009.

[26] Y. K. Huang, C. Chen , and C. Lee, "Continuous k -nearest neighbor query for moving objects with uncertain velocity," GeoInformatica 13(1), 2009.

[27] R. Lange, H. Weinschrott, L. Geiger, A. Blessing, F. Dürr, K., Rothermel, and H. Schütze, "On a generic uncertainty model for position information," In: QuaCon, pp. 76–87, 2009.

[28] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," In: ICDE, pp. 70–79, 2008.

[29] P. Zhou, D. Zhang, B. Salzberg, G. Cooperman, and G. Kollios, "Close pair queries in moving object databases," Proceedings of ACM GIS, 2005.

[30] V. P. Chakka, A. Everspaugh, and J. M. Patel, "Indexing large trajectory data sets with seti," Proc. of the Conf. on Innovative Data Systems Research (CIDR), 2003.

[31] Y. Tao, and D. Papadias, "Efficient historical r-trees," In: ssdbm, pp. 223 – 232, Published by the IEEE Computer Society, 2003.

[32] D. Pfoser, C. Jensen, and Y. Theodoridis, "Novel approaches to the indexing of moving object trajectories," VLDB, 2000.

[33] Y. Zheng, and X. Zhou, Computing with spatial trajectories, Springer Science and Business Media, 2011.

[34] V. Bogorny, B. Kuijpers, and L. O. Alvares, "A Spatio-temporal Data Mining Query Language for Moving Object Trajectories," Instituto de Informatica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil Technical Report TR-357, 22, 2008.

[35] F. Verhein, and S. Chawla, "Mining Spatio-temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases," Proceedings of the DASFAA, M.L. Lee, K.L. Tan and V. Wuwongse (Eds), 3882 of Lecture Notes in Computer Science (Springer), pp. 187-201, 2006.

[36] J. G. Lee, J. Han, and K. Y. Whang, "Trajectory clustering: a partition-and-group framework," Proceedings of the SIGMOD Conference, C.Y. Chan, B.C. Ooi and A. Zhou (Eds) (ACM), pp.593- 604, 2007.

[37] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," Proceedings of the KDD, P. Berkhin, R. Caruana and X. Wu (Eds) (ACM), pp. 330-339.

[38] J. R. Hwang, H. Y. Kang, and K. J. Li, "Searching for similar trajectories on road networks using spatio-temporal similarity," Proceedings of the 10th East European Conference on Advances in Databases and Information Systems (ADBIS), pp. 282–295, 2006.

[39] A., Kharrat, I. S. Popa, K. Zeitouni, and S. Faiz, "Clustering algorithm for network constraint trajectories," Headway in Spatial Data Handling, Springer Berlin Heidelberg, pp. 631-647, 2008.

[40] E. Tiakas, A. N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic, and S. Djordjevic-Kajan, "Searching for similar trajectories in spatial networks," Journal of Systems and Software, 82(5), pp. 772-788, 2009.

[41] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "Semantic trajectories: Mobility data computation and annotation," ACM Transactions on Intelligent Systems and Technology 9(4), 2012.

[42] B. Guc, M. May, Y. Sayigin, and C. Koerner, "Semantic Annotation of GPS Trajectories," In: 11th AGILE International Conference on Geographic Information Science, 2008.

[43] L. Alvares, V. Bogorny, B. Kuijpers, J. F. de Macedo, B. Moelans, S. Spaccapietra, and A. Vaisman, "A model for enriching trajectories with semantic geographical information," Proceedings of the 15th ACM international symposium on Advances in Geographic Information Systems (GIS'07), ACM, New York, NY, USA, pp. 1–8, 2007.

[44] E. Camossi, P. Villa, and L. Mazzola, "Semantic-based Anomalous Pattern Discovery in Moving Object Trajectories," arXiv preprint arXiv:1305.1946, 2013.

[45] K. S. Hornsby, and S. Cole, "Modeling Moving Geospatial Objects from an Event-based Perspective," Transactions in GIS, 11(4), pp. 555-573, 2007.

[46] A. Galton, and M. Worboys, "Processes and events in dynamic geo-networks," Rodriguez M A, Cruz I, Levashkin S, and Egenhofer M (eds) Proceedings of the First International Conference on Geospatial Semantics (GeoS 2005). Berlin, Springer-Verlag Lecture Notes in Computer Science No 3799, pp. 45–59, 2005.

[47] P. Villa, and E. Camossi, "A Description Logic Approach to Discover Suspicious Itineraries from Maritime Container Trajectories," In: Claramunt C, Levashkin S, Bertolotto M (eds) Proceedings of the 4th International Conference on GeoSpatial Semantics, (GeoS 2011), Brest, France, Springer Berlin Heidelberg, vol 6631, pp. 182–199, 2011.