A Semantic Interpretation of Unusual Behaviors Extracted from Outliers of Moving Objects Trajectories

Sana CHAKRI, Said RAGHAY, Salah EL HADAJ

Laboratory of applied mathematics and informatics,
Cadi Ayyad University
Marrakesh, Morocco

Abstract—The increasing use of location-aware devices has led to generate a huge volume of data from satellite images and mobile sensors; these data can be classified into geographical data. And traces generated by objects moving on geographical territory, these traces are usually modeled as streams of spatiotemporal points called trajectories. Integrating trajectory sample points with geographical and contextual data before applying mining techniques can be more gainful for the application users. It contributes to produce significant knowledge about movements and provide applications with richer and more meaningful patterns. Trajectory Outliers are a sort of patterns that can be extracted from trajectories. However, the majority of algorithms proposed for discovering outliers are based on the geometric side of trajectories; our approach extends these works to produce outliers based on semantic trajectories in order to give meaning to the outliers extracted, and to understand the unusual behaviors that can be detected. To prove the efficiency of the approach proposed we show some experimental results.

Keywords—*Moving objects analysis; spatial databases; data mining; Semantic clustering; semantic trajectories*

I. INTRODUCTION

Researchers from spatial databases, GIS, data mining, and knowledge extraction communities have developed several techniques for mobility analysis. As consequence three research areas have been expended; The first one focuses on data modeling to provide definitions and extensions of trajectory related data types such as moving objects, points, lines, or regions. The second deals with data management to optimize the storage of mobility data with suitable indexing and querying techniques. And the last one that is the main topic of this research deals with the analysis of patterns that can be extracted from stored data like trajectories by using spatiotemporal data mining algorithms. Several data mining methods have been proposed for extracting patterns from trajectories. However, the majority of them use trajectories without looking for any additional information, and yet by considering only the raw trajectory data, discovering why an object followed a different route become very complex since no additional information (called semantic) is given about the moving object. This additional information can hide behind a lot of meanings; in fact it can lead to a better understanding of the patterns extracted. This is can be achieved by combining the raw mobility tracks (e.g., the GPS records) with related contextual data in order to use semantic trajectories instead of

focusing only on the geometric side of trajectories. Therefore, applying mining techniques on semantic trajectories continues to prove success stories in discovering useful and non-trivial behavioral patterns of moving objects. Several data mining methods have been proposed for extracting behaviors from trajectories such as chasing behaviors [1], flocks [2], avoidance [3], etc. In this paper, we focus on trajectory outlier detection. Trajectory outliers are sort of patterns that can be extracted from semantic trajectories of moving objects. The objective in trajectory outlier detection is to find trajectories that do not comply with the general behavior of the trajectory dataset. While most of pattern analysis focuses on patterns that are common in the trajectory dataset, outlier detection focuses on rare patterns such as trajectories that follow a path different from the common path followed by most of the other moving objects, or objects following the same path but behave differently than the other objects (very slow or fast objects compared to the majority of the moving objects). Trajectory outlier detection can be very useful in traffic analysis, it helps understand the flow of people that move between regions, how this flow is distributed and what are the characteristics of the movements. In high traffic routes, outliers can give some alternative paths that can reduce the volume of traffic, or give the best or worst path that links two areas, by extracting outliers, users can easily discover suspicious behaviors like company cars that escape from their normal route. In fact, detecting semantic outliers proves his efficiency, especially to discover suspicious behaviors in a group of people, to find alternative routes in traffic analysis in many applications such as transportation, ecology, animal tracking, health sector, crime sector, and climatology, etc. Indeed, by adding semantics to outliers, the analysis became more performed; we can discover the reasons for each behavior extracted. The interpretation of outliers can provide more information to the decision maker. Thus, many new applications are interested in understanding and using semantic interpretation of the moving object behavior. Semantics refers essentially to additional contextual and geographical information available about the moving object, apart its position. Semantics contain both the geometric properties of the moving object as well as the geographic properties and any other additional information like the moving object's activity, mode of transportation, speed or any data that can help give more meaning to the behavior extracted. The purpose of this research is to find spatial, spatiotemporal and temporal outliers among semantic trajectories, analyzing them

taking into account their semantic data to understand the meaning of the outliers detected, especially to give an answer to the famous question "why an object could deviate from a group?"

The rest of the paper will be organized as follows: in section 2 we will present the related work, in section 3 we will present the semantic outlier detection in which we will discuss the flow to construct semantic trajectories then apply mining algorithms for extracting outliers, section 4 will provide with the methodology used to give meaning to outliers extracted. Section 5 illustrates the algorithms used, section 6 gives case of study and in section 7 we will discuss the work proposed and gives some comparisons.

II. RELATED WORK

To our best knowledge, there are a medium number of researches to detect outliers in trajectories. However, only a few of them focus on semantic as they focus on geometric data, so we can split this research area to two essential fields: the first filed focuses only on the geometrical side of outliers like [4] which is an efficient technique to discover spatiotemporal outliers and causal relationships between them. Another one is proposed in [5] used for detecting outlier sequences in precipitation data. A roughest approach is described in [6] for spatiotemporal outlier detection. A survey was presented in [7], in which more approaches for outlier detection in temporal and spatiotemporal data were discussed. The second filed handles semantic data besides than geometrical one, their approaches are closer to our research like [8 9]. For the first work, authors try to find outliers between regions of interest, in the second authors try to find the specific standard path that the outlier deviates and propose to give a meaning to it. In [10], the main objective is to discover outliers among trajectories that have the same goal and move between the same regions and to give a meaning to these outliers extracted. Authors in [11] tries to extract anomalous behaviors in single-trajectory data, in [12], authors propose a method of detecting avoidance behaviors between moving objects, and the paper [13] tries to detect abnormal pedestrian behavior based on a new trajectory model, [14] and [15] are recent works that tries to detect outliers based on vehicle trajectories and multi-factors. Our work extends these works by giving a global approach which starts by merging GPS feeds with semantic data to produce semantic trajectories, then applying the mining algorithm proposed in order to give a very deeper analysis to the outliers extracted, we also try to analyze the outliers extracted according to semantic data to give more precision to the reasons for which some moving objects deviate from the main route.

III. SEMANTIC OUTLIER DETECTION

A. Enriching trajectories with semantics

Trajectories of moving objects present a huge data warehouse where users can extract several information according to the application domain studied, this is can be achieved by applying data mining techniques based on both temporal and spatial data mining algorithms. However, spatiotemporal data mining is only one step between all the knowledge discovery processes. In fact, to extract meaningful

knowledge, the trajectories must follow several steps to be ready to use for data mining, our approach gives the whole process that trajectories pursuit to be structured and enriched before being used. First, it consists of enriching trajectories with semantic data throughout a process where the raw trajectories will be built from GPS feeds, cleaned, well structured, and enriched before applying data mining algorithms. Figure 1 illustrates the process pursued to build semantic trajectories; it is structured in three steps to prepare trajectories for data mining. The first step is raw trajectories building, where we try to prepare trajectories by cleaning and structuring the GPS points which can be defined as: Definition1: A point p is a tuple (x,y,t), where x and y are spatial coordinates and t is the time instant in which the coordinates were collected. The formatted points produce a healthy raw trajectory that is defined as: Definition2: A trajectory T is a list of points $(p1, p2, p3,..., pn)$, where $pi = (xi,yi,ti)$ and $t1< t2 < t3 < ... < tn$.

The second step (Semantic Trajectory Enrichment) takes as input these structured trajectories, and tries to segment it into episodes (sub-trajectories) of stops and moves, then annotated them with related contextual data to product semantic trajectories. A sub-trajectory can be defined as:

Definition3: Let $T = (p1, p2, p3,..., pn)$, be a trajectory. A sub-trajectory S of T is a list of consecutive points $(pk, pk+1, pk+2,..., pm)$, where $p \in T$, $k \geq A$, and $m \leq n$.
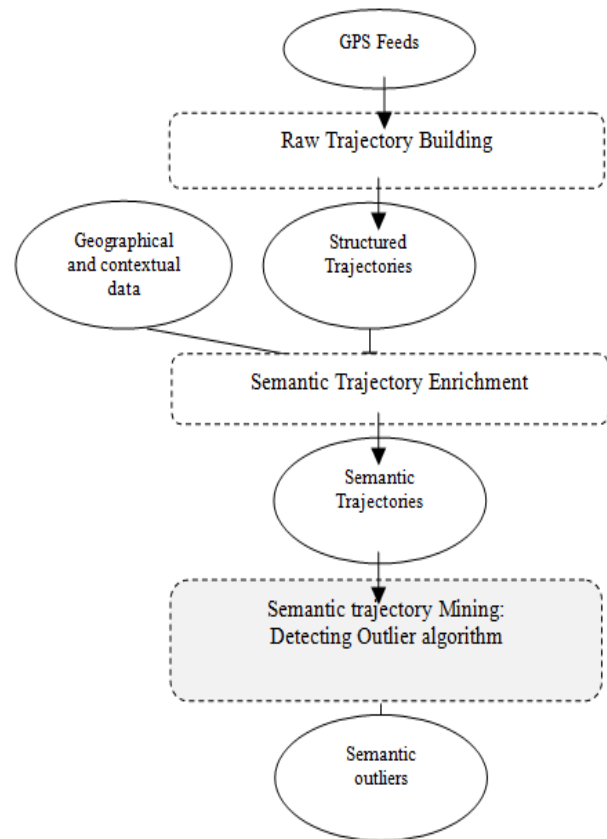

Fig. 1. Trajectory enrichment and extraction process

These semantic trajectories will be the input of the third phase that is semantic trajectory mining, where we will be able to apply mining algorithms to extract suspicious behaviors of moving objects (outliers), more details about the process of enrichment are explained in [16].

B. Extracting Semantic Outliers

Globally, outlier analysis in classical databases reveals odd objects which appear to be inconsistent with the other objects in the database. This definition implies that the object is significantly different from the overall database as a whole. However, in case of spatiotemporal databases, it is possible for an object to appear consistent with the entire database objects, but appear unusual with a local neighborhood [17, 18]. Therefore, we can say that an outlier is a spatiotemporally geo referenced object whose non-spatiotemporal attribute values differ from objects in its spatiotemporal neighborhood. Otherwise, a spatiotemporal outlier is a local shakiness or inconstancy. An outlier can refer either the whole trajectory, or more often it refers parts of trajectories called sub-trajectories, where the moving object chooses to behave differently compared to the rest of the other moving objects trajectories and then becomes suspicious [19].

1) Methodology

The purpose is to find spatiotemporal and temporal outliers between regions of interest [20], Analyzing them with semantic data to understand the meaning of the outliers detected. Spatiotemporal outliers refers to sub-trajectories that have spatial and temporal difference compared to common trajectories, while temporal outliers refers to moving objects that behave spatially like the majority of the other moving objects, but temporally they are different; for instance moving objects that took the same route but they accelerate or they mark an important number of stops which make them seen as suspicious moving objects. The analysis presented in this paper are made on sub-trajectories that rely regions of interest which are shapes that have different size and format, depending on the application, they can be regions ROI, lines LOI, or even points POI, they can be districts, dense areas, hotspots, important places, etc. generally a region of interest can be a pre-defined important place or computed by an algorithm that finds dense areas. In our case we consider a region as a point, line or polygon, which is a well-known concept in GIS community. The use of regions allows filtering from the whole dataset only the sub-trajectories that move between the same regions, outliers will be searched among these sets what significantly reduces the search space for outliers. Among the trajectories that cross all regions, we are only interested in the part of trajectories (sub-trajectories) that move between specific regions, we call these sub-trajectories Nominees. After defining the set of nominees, we start looking for temporal outliers, and spatial outliers in which we extract from them spatiotemporal outliers. A nominee will be a spatial outlier when it follows a different path in relation to the majority of the sub-trajectories from its group, and it can be a temporal outlier if it follows the same path, but shows different behaviors compared to the other moving objects. In general, we have two types of path: Populated path that have many trajectories in its proximity. And depopulated Path, it has less trajectories around. The spatial and the spatiotemporal outliers

will be extracted from depopulated paths, while the temporal outliers will be extracted from the populated paths.

To detect if the nominee is in the populated or the depopulated path, we introduce the concept of proximity; A nominee is in proximity to a point if it is close to the point, if a point has a few nominees in its proximity, then at that time the moving object was following a path different from the majority of Nominees, it is in a depopulated path. The maximal distance for a nominee to be in proximity to a point is called PD (proximity distance). In general, there is at least one main route used between two regions, which is more frequent than alternative ways. The minimal amount of a nominee (MA) is the minimal number of points that each point of a nominee should have in its proximity to be part of this main route. The nominee that has all its points in a populated path is considered common trajectory. The nominee that has at least one point where the cardinality of its proximity is less than MA is called expected outlier. So the nominee will always be either Common trajectory or expected outlier. The spatial outliers will be extracted from expected outliers, and the temporal outliers will be extracted from common trajectories. When two nominees leave the start region at a time interval inferior to Maximal Tolerance MT. we can say that they are synchronized.

2) The process

The general process, as shown by images in figure 2, starts by looking at sub-trajectories that have the same arrival, in order to define the nominees (figure 2.a). Like said before; a nominee will be an outlier when he follows different path compared to the majority of the sub-trajectories from its group, or when it behaves differently even if he follows the same
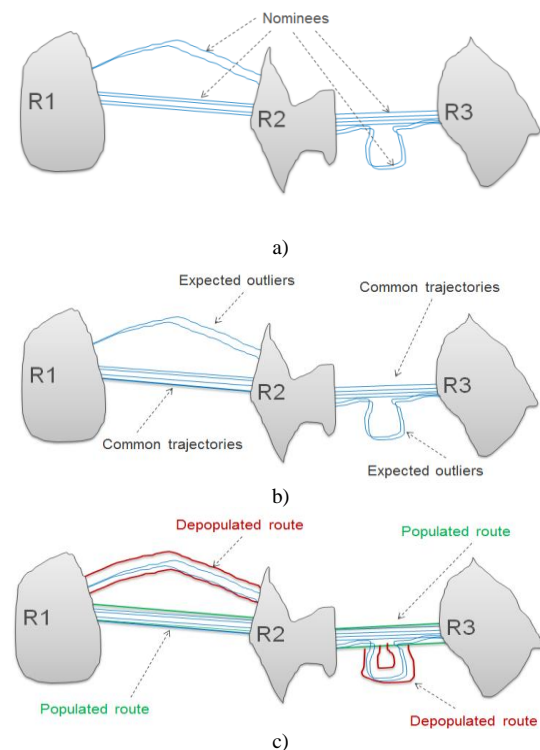
Fig. 2. Steps for detecting outliers

route. So we try to define the expected outliers which are sub-trajectories that have a few neighbors in their proximities, and the common trajectories which have a lot of neighbors in their proximities (figure 2.b). The route followed by the expected outliers is considered as depopulated route, while the route followed by common trajectories is a populated route (figure 2.c).

After grouping the expected outliers and to select from them the spatial outliers, we verify two conditions; the first one is that the expected outlier connects two regions? If yes, we move to the second one that is for these two regions, is there any populated route detected? Because if we want to discuss the existence of spatial outliers, it should be at first a populated route that the majority of moving objects follows, then the deviation can be seen as spatial outlier, if there is no populated route, we can't discuss spatial outliers. When all nominees between two regions are expected outliers, which means there is no common trajectories; there is no populated path that an object could avoid or deviate. Contrariwise, if there is at least one populated path, then the expected outlier did really perform a detour, and becomes spatial outlier. No spatial outlier will exist if there is no common trajectories, as assumption to define a spatial outlier, It should move between two regions of interest, and there must be a populated path that connects the regions such that the spatial outlier should deviate from it, therefore, any sub-trajectory that uses a path different from the populated path is a spatial outlier. For the temporal outliers, they will be extracted from common trajectories. As said before; the temporal outliers are sub-trajectories that follow the same path used by the most of moving objects, but behave differently than the other objects; for example some moving object can make several stops in his way, so it can be seen as a very slow object compared to the majority of the other moving objects, or contrariwise, it can be seen more fast. After extracting the outliers detected, we classify them first according to their speed, and then we try to analyze each group of outliers classified by proposing a meaning to their deviations by looking for the reasons of deviation.

IV. GIVING MEANING TO UNUSUAL BEHAVIORS

After extracting outliers from semantic trajectories, the main goal of the next step is to add meaning to the outliers extracted. The next step is about splitting the outliers extracted to several types according to their semantic interpretation;

A. Spatiotemporal outliers

Figure 3 illustrates the classification of spatiotemporal outliers extracted from spatial outliers.

1) Stop outliers

It occurs when the moving object made a stop for some time during the deviation, for instance the moving object had an appointment, a meeting, go shopping after work, pick up the children at school, go with friends, pass by a market, or something to do somewhere else that was not in the standard path. This is an intentional detour with a reason. To discover if an outlier has a stop we need to look for stops not in the complete outlier trajectory, but only in the sub-trajectory that

corresponds to the outlier (deviation), i.e., the outlier segment. We consider as a stop a sub-trajectory that its speed is close to zero for a minimal amount of time (MT).

2) Emergency outliers

It occurred when the moving object took an alternative route and shows an important acceleration of its speed, the reasons can be almost about an emergency case like an ambulance transporting patient, or someone trying to escape from police, etc. to detect if there is an emergency we need to compare the speed of the fast outlier with the speed of the synchronized outliers that took the same deviation. We consider that there is an emergency outlier if the speed of the fast outlier is higher than the double of the average speed of the synchronized outliers detected in the same derived route.

3) Regular outliers

It occurs when the moving object deviates from the populated route without an important change of speed, or with a degradation of speed. This may reveal that the populated route is temporarily busy or is under reconstructions, or there is an accident, or even there is an event that block the path, so the moving object is forced to deviate from the populated route, Which can cause a big traffic on the alternative ways, and as consequence, the speed of the moving object may decrease. Our algorithm assembles all these reasons in three types of outliers: the blocked route outlier, the avoided route outlier, and the traffic jam outlier.

a) Blocked route outliers

Expresses any deviation because something happens close to the populated route which causes some blockage, for instance, an accident, route reconstructions, or some artistic events like a carnival or a concert. The challenge is how to discover the case that blocked the populated route; we start by analyzing only the part of the closest populated route deviated by the outlier (we call it the main segments), then we look if there is an activity around the main segments, if yes, we verify the time of this activity to be sure that the outlier was generated in the moment of the action. And finally we verify that at the time of the activity, there are no synchronized segments in the populated route, to prove that the path was blocked by the event, so the moving objects were forced to take an alternative route. Thus, a blocked route outlier is an outlier that deviates from the populated route because a blocking activity is happening close to the populated route at the same time of the deviation.

b) Avoided route outliers

This type of outliers is similar of the first type, the only difference is that there is an activity in the populated route, but this activity doesn't cause any blockage, an example could be a police checkpoint; In this case, the majority of moving objects will take the populated route normally, but some of them choose to avoid this event. For discovering such type of outliers we verify if there is an activity in the populated route. If yes we verify if there are some trajectories which Travers the populated route in the time of the activity to prove that the activity doesn't block the route. At this time we can say that this outlier is of type avoided route outlier.
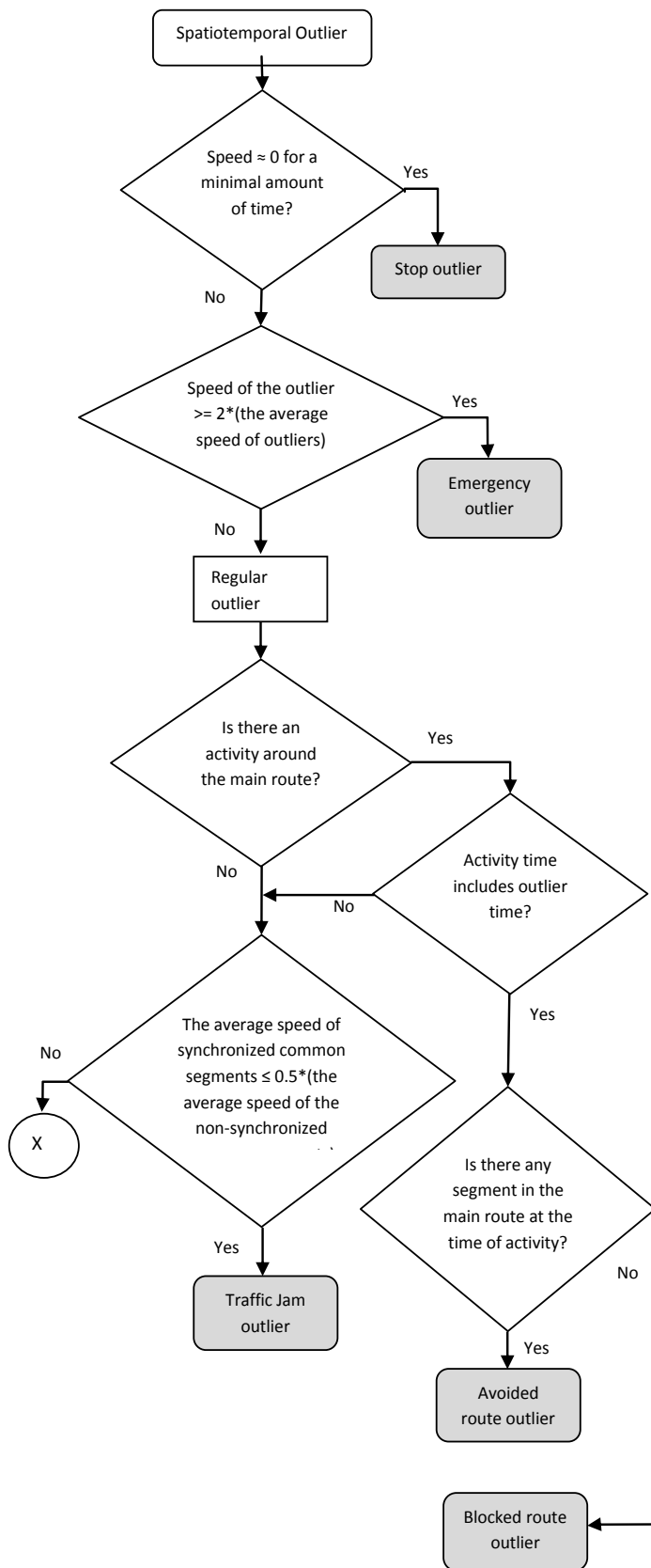
Fig. 3. Operating logic schema for giving meaning to spatiotemporal outliers extracted

c) Traffic jam outliers

Expresses deviations due to a heavy charge at the rush hour, it occurs if we found an outlier, but no activity is blocking the populated route, so we start looking if there is a traffic jam. For that we look for the slow traffic in the populated route at the time of the outlier. To measure the speed on the populated route at the same moment that the outlier deviated from it, we need to look only at the segments of the synchronized common trajectories. The average speed of all synchronized common segments in the same populated route is compared to the speed of the non-synchronized common segments in the same route. We consider that there is a traffic jam when the average speed of those who are synchronized is less than half of the average speed of the non-synchronized.

B. Temporal outliers

Temporal outliers are common trajectories that follow the populated route, but with an important difference of the speed compared to the other common trajectories. For extracting such type of outliers, we make use of the average speed used by the moving objects in the populated route, we make a comparison between each sub-trajectory from the common trajectories and the average speed for all common trajectories that traverse the same route with some tolerance, and we extract two essential types; temporal emergency outliers, and temporal stop outliers.

1) Temporal emergency outliers

This type of outliers is extracted from fast common trajectories that traverse the populated route. It occurred when the moving object stay in the populated route but shows an important acceleration of its speed, the reasons can be almost about an emergency case. To detect if there is a temporal emergency, we need to compare the speed of the fast common trajectory detected with the speed of the synchronized common trajectories that took the same populated route. We consider that there is a temporal emergency outlier if the speed of the fast common trajectory is higher than the double of the average speed of the synchronized common trajectories detected in the same populated route.

2) Temporal stop outliers

The temporal stop outliers are common trajectories that Travers the populated route with a very slow speed compared to the synchronized common trajectories in the same route, it occurs when the moving object made a stop for some time in the populated route. To discover if the common trajectory has a stop we need to look for stops in the sub-trajectory that corresponds to the common trajectory. We consider as a temporal stop outlier a sub-trajectory that its speed is close to zero for a minimal amount of time (MT).

V. ALGORITHM

In this section we present the algorithms used to detect and interpret the outliers extracted. Figure 4 shows the pseudo-code of the main algorithm.

The algorithm starts by computing the nominees that move between two regions, with the function detectNominee. This function checks for every trajectory if it intersects the pair of regions. Once the nominees are computed, the algorithm searches for the common trajectories (trajectories that follow

the populated route) with the function findCommon, considering the parameters PD and MA, this function checks for all points of a nominee in the set if the number of points in proximity is greater than MA.

```
1   INPUT:
2   T; // set of trajectories
3   R; // set of regions
4   PD; // Proximity distance
5   MA; // Minimal amount
6   MT; // Maximal tolerance
7   MS; // minimal speed
8   MD; // minimal duration
9   ML; // minimal lenth
10  MSA; // minimal Stops allowed
11  OUTPUT:
12  ClassifiedOutliers; // set of outliers
13  for each (reg1,reg2)in R
14  {   N = detectNominees(T,reg1,reg2); // find nominees
15      CommonSet = FindCommon(N,PD,MA); // find common trajectories
16      if(not_empty(CommonSet)) then
17      {   ClassifiedOutliers.TemporalEmergencyO(CommonSet); // find temporal emergency outliers
18          ClassifiedOutliers.TemporalStopO(CommonSet,MS,MD,MSA); // find temporal outliers
19          OutlierSet = N  MINUS  CommonSet; // Set of spatial outliers
20          For each outlier in OutlierSet
21          {   outlierSegments = getOutSeg(outlier,ML); // all segments of the outlier
22              ClassifiedOutliers.StopOutlier  = findStops(outlierSegments, MS,MD);
23              O = outlierSegments - ClassifiedOutliers.StopOutliers;
24              ClassifiedOutliers.EOutlier = findEmergencyOutlier(O);
25              O = O - ClassifiedOutliers.EOutlier;
26              ClassifiedOutliers.ROutlier = findRegularOuteOutlier(O);
27          }
28      }
29      return ClassifiedOutliers;
30  }
```

Fig. 4.   Pseudo Code of the main algorithm

If this is the case, then the nominee is considered as common trajectory. If the set of common trajectories is not empty, the algorithm tries to extract temporal emergency outliers and temporal stop outliers, and then it goes for finding the spatial outliers, since there is a common path that connects both regions.

In the next step, the algorithm goal is to add meaning to the outliers extracted. So we go further in semantics by extracting the types of outliers; temporal stop outliers, Temporal Emergency outliers, stop Outliers, Emergency Outliers, Blocked Route Outliers, Avoided Route Outliers and Traffic Jam Outliers. Figure 5 illustrate the algorithms used. The temporal stop outliers and Stop Outliers are classical types that the majority of data mining algorithms use to detect stops of moving objects, The Emergency outliers are extracted from fast outliers, and the temporal emergency outliers are extracted from fast common trajectories. The Regular Outlier captures all outliers that keep almost the same or less speed, and then the algorithm tries to detect from this type the blocked Route Outliers, the avoided Route Outliers and the traffic Jam Outliers.

```
1   INPUT:
2   OutlierSegments outSegs; // set of outlier Segments
3
4   OUTPUT:
5   EmergencyOutlier  EO; // set of emergency outliers
6
7   for each Out in outSegs
8   {
9       if(out.speed) >= (2*(avgSpeed(outSegs)))
10      {
11          EO.add(out);
12      }
13  }
14  Return EO;
```
A

```
1   INPUT:
2   CommonSet commonSet; // set of common trajectories
3
4   OUTPUT:
5   TemporalEmergencyOutlier  TEO; // set of temporal emergency outliers
6
7   for each ComnT in commonSet
8   {
9       if(ComnT.speed) >= (2*(avgSpeed(commonSet)))
10      {
11          TEO.add(ComnT);
12      }
13  }
14  Return TEO;
```
B

```
1   INPUT:
2   OutlierSegments outSegs; // set of outlier Segments
3
4   OUTPUT:
5   RegularOutlier RO; // set of Blocked route outliers
6
7   for each out in outSegs
8   {
9       ComnSeg CS = getComnSeg(Out);
10      Regions R = getIntersection(Cs); // get all regions that intersect the common segment
11      Activities A = R.getActivities(out.time) // get the activities that
12                                        //appnehed at the same time of the outlier
13      if (not_empty(Activities))
14      {
15          SynCmnSeg SCS = getsynchronizedCommon(out);
16          if(empty(SCS))
17              RO.addBlockedRouteOutlier(out);
18          else
19              RO.addAvoidedRouteOutlier(out);
20      }
21      else
22      {
23          SynCmnSeg SCS = getsynchronizedCommon(out);
24          NSynCmnSeg NCS = getNNsynchronizedCommon(out);
25          if(avg(SCS.speed) <= (2*(avg(NCS.time)))
26              RO.addTrafficJam(out);
27      }
28  }
29  Return RO;
```
C

Fig. 5.   Pseudo code of semantics outliers; A : Pseudo Code of emergency outliers, B : Pseudo code of temporal emergency outliers, C : Pseudo code of regular outliers

VI.   EXPERIMENTAL RESULTS

In this section we present the results of experiments with real data, before that we provide with a presentation of the general architecture of our approach in the figure 6. Our approach contains tree main phases in the general architecture, the first one concerns the data preprocessing where the GPS

feeds will be treated to become sample trajectories, then they will be able to be structured in the enrichment process [21]. In the second phase we make use of the Weka-STPM toolkit [22] which is a java toolkit for semantic trajectory data mining and visualization, we have used the CB-SMot algorithm to create Stops and Moves [23].

Fig. 6.   General architecture

After the Semantic process we move to the last phase when we apply the Semantic Outlier Analysis algorithm in which we extract the outliers then add meanings.

For the experimental results we try to analyze two data sets to prove the efficiency of our method, these datasets rare taken from [24 25 26 27 28]. The first one contains trajectories of School Buses dataset which consists of 145 trajectories of two school buses collecting and delivering students around Athens metropolitan area in Greece for 108 distinct days. Notice that we analyzed only trajectories from Monday to Friday. The second are Trucks dataset which consists of 276 trajectories of 50 trucks delivering concrete to several construction places around Athens metropolitan area in Greece for 33 distinct days. The structure of each record is as follows: {obj-id, traj-id, date(dd/mm/yyyy), time(hh:mm:ss), lat, lon, x, y} where (lat, lon) is in WGS84 reference system and (x, y) is in GGRS87 reference system. These datasets are interesting for analyzing outliers because this type of drivers, in general, knows different routes to reach the same place. Therefore, we can find the alternative routes (outliers) in relation to the standard path. In this experiment we consider as interesting regions the districts around Athens metropolitan area. The application domain data are all about information about drivers, the number of students for the school buses, the type and the number of products for

the trucks, the noun of the districts and the activities of the drivers and regions in this period.

The results for school buses are displayed below;

Fig. 7.   School bus trajectories

Fig. 8.   Common trajectories

Fig. 9.   Outliers extracted

TABLE. I.   SCHOOL BUS OUTLIERS EXTRACTED

Nominees	Expected outliers	Spatiotemporal outliers	Common trajectories	Temporal outliers
54598	2402	1778	52196	54

TABLE. II. SEMANTIC SPATIOTEMPORAL OUTLIERS FROM SCHOOL BUS TRAJECTORIES

Spatiotemporal outliers					
Stop	*Emergency*	*Regular*			
448	-	1330			
		Blocked route	*Avoided route*	*Traffic Jam*	*Others*
		21	706	454	149

TABLE. III. SEMANTIC TEMPORAL OUTLIERS FROM SCHOOL BUS TRAJECTORIES

Temporal outliers	
Stop	*Emergency*
54	-

The experimental results for school buses outliers show that the trajectories contain 1778 spatiotemporal outliers from 2402 expected outliers, and contain 54 temporal outliers from 52196 common trajectories, the spatiotemporal outliers contain 448 stop outliers and 1330 regular outliers, in which there are 21 blocked route outliers, 706 avoided route outliers, 454 traffic jam outliers, and 149 outliers none defined.
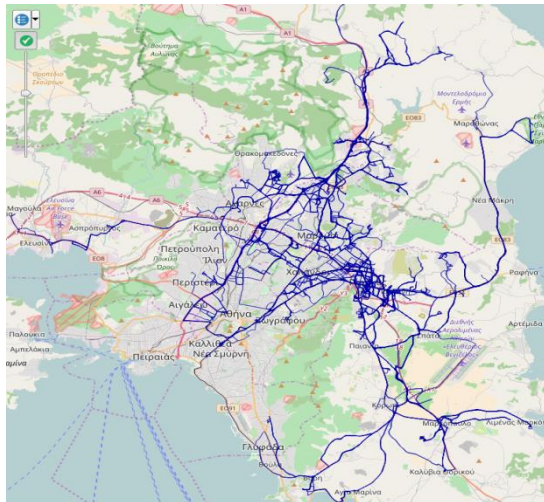
The results for trucks are displayed below

Fig. 10. Trucks trajectories

Fig. 11. Common trajectories

Fig. 12. Outliers extracted

TABLE. IV. TRUCKS OUTLIERS EXTRACTED

Nominees	Expected outliers	Spatiotemporal outliers	Common trajectories	Temporal outliers
35750	9402	1157	26348	421

TABLE. V. SEMANTIC SPATIOTEMPORAL OUTLIERS TRUCKS TRAJECTORIES

Spatiotemporal outliers					
Stop	*Emergency*	*Regular*			
512	14	631			
		Blocked route	*Avoided route*	*Traffic Jam*	*Others*
		14	345	225	47

TABLE. VI. SEMANTIC TEMPORAL OUTLIERS FROM TRUCKS TRAJECTORIES

Temporal outliers	
Stop	*Emergency*
387	43

The experimental results for trucks outliers show that the trajectories contain 1157 spatiotemporal outliers from 9402 expected outliers, and contain 421 temporal outliers from 26348 common trajectories, the spatiotemporal outliers contain 512 stops, 14 emergency outliers, and 631 as regular outliers, in which we have 14 blocked route outliers, 345 avoided route outliers, 223 traffic jam outliers, and 47 other outliers none defined.

VII. CONCLUSION

Several algorithms have been proposed for trajectory data mining, but only a few consider semantics, and very few of them deal with semantics on trajectory outlier detection. In this paper, we gave importance to outliers extracted from semantic trajectories, for that we have proposed a conceptual approach that consist to build trajectories from GPS points, enrich them with semantic data, then apply mining algorithm to detect semantic outliers from moving objects, the algorithm shown in this experiment discovers the populated route that the majority of trajectories followed, then detect all other deviations that trajectories can follow to reach the same place, after that the algorithm divided the results to spatiotemporal outliers and just temporal outliers. The spatiotemporal outliers are extracted

from spatial outliers, and they contain stop outliers, emergency outliers, and regular outliers in which three types are discussed; blocked route outliers, avoided route outliers and traffic jam outliers. The temporal outliers contain stops and emergency outliers that can exist in the populated route. The next step will be the introduction of the direction of outliers extracted, and the introduction of mode of transportation to distinguish the types of moving objects that can use the routes [30, 31], giving more details about results, and studying the parameters of the algorithm.

REFERENCES

[1] Fernando de Lucca Siqueira and Vania Bogorny, "Discovering chasing behavior in moving object trajectories", T. GIS, vol. 15, no. 5, pp. 667–688, 2011.

[2] Monica Wachowicz, Rebecca Ong, Chiara Renso, and Mirco Nanni, "Finding moving flock patterns among pedestrians through collective coherence", International Journal of Geographical Information Science, vol. 25, no. 11, pp. 1849–1864, 2011.

[3] Luis Otávio Alvares, Alisson Moscato Loy, Chiara Renso, and Vania Bogorny, "An algorithm to identify avoidance behavior in moving object trajectories", J. Braz. Comp. Soc., vol. 17, no. 3, pp. 193–203, 2011.

[4] Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan & Xing Xie, (2011) "Discovering SpatioTemporal Causal Interactions in Traffic Data Streams", KDD'11, August 21 -24, San Diego, California, USA.

[5] Elizabeth Wu, Wei Liu & Sanjay Chawla, (2010) "Spatio-temporal outlier Detection in Precipitation Data", Knowledge discovery from sensor data, Volume 5840, pp 115-133.

[6] Alepsia Albanese, Sankar K Pal & Alfredo Petrosino, (2011) "A rough set approach to spatiotemporal outlier detection", Proceedings of 9th international conference on Fuzzy logic and applications, Springer-verilog, LNCS Volume 6857, pp 67-74.

[7] Gupta, M., Gao, J., Aggarwal, C. C., and Han, J. (2013). Outlier detection for temporal data: A survey. TKDE, 25.

[8] Fontes, V. C., de Alencar, L. A., Renso, C., and Bogorny, V. (2013). Discovering trajectory outliers between regions of interest. In GeoInfo.

[9] AR Aquino. Alvares L. O., Renso C. and Bogorny V .Towards Semantic Trajectory Outlier Detection. GeoInfo. (2013).

[10] Fontes, V. C, Bogorny V. (2013). Discovering Semantic Spatial and Spatio-Temporal Outliers from Moving Object Trajectories.CoRR, abs/1303.5132.

[11] Huang, H. Anomalous behavior detection in single-trajectory data. International Journal of Geographical Information Science, 29 (12), 2015 .

[12] Lettich, F., et al . Detecting avoidance behaviors between moving object trajectories. Data & Knowledge Engineering . 2016 .

[13] Li, X., et al . . A method of abnormal pedestrian behavior detection based on the trajectory model. In : ICTE 2013 : proceedings of the fourth international conference on transportation engineering, , 2013 .

[14] Shen, M., Liu, D.-R., and Shann, S.-H. Outlier detection from vehicle trajectories to discover roaming events.Information Sciences, 2015 .

[15] Zhang,L.,Hu,Z.,andYang,G. Trajectory outlier detection based on multi-factors. IEICE TRANSACTIONS on Information and Systems, 2014.

[16] Sana Chakri, Said Raghay and Salah El Hadaj. Modeling, Mining, and Analyzing Semantic Trajectories: The Process to Extract Meaningful Behaviors of Moving Objects. International Journal of Computer Applications 124(8):15-21, August 2015. Published by Foundation of Computer Science (FCS), NY, USA.

[17] Lee, J.-G., Han, J., and Li, X. (2008). Trajectory outlier detection: A partition-and-detect framework. In ICDE, pages 140–149. IEEE.

[18] Mateus Barragana, Luis Otavio Alvares & Vania Bogorny, Unusual behavior detection and object ranking from movement trajectories in target regions, International Journal of Geographical Information Science, 2016.

[19] Zheng, Y., Zhou, X.: Computing with spatial trajectories. Springer (2011).

[20] de Graaff, V., de By, R.A., van Keulen, M., Flokstra, J.: Point of interest to region of interest conversion. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL GIS 2013), Orlando, FL, USA, (New York), pp. 378–381. ACM, November 2013.

[21] Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., et al.: Semantic trajectories modeling and analysis. ACM Computing Surveys (CSUR) 45(4), 42 (2013).

[22] Alvares L, Palma A, Oliveira G, Bogorny V Weka-STPM: from trajectory samples to semantic trajectories. In: Proceedings of the Workshop Open Source Code.2010.

[23] Nanni, M., Pedreschi, D. 2006. Time-focused clustering of trajectories of moving objects. Journal of Intelligent Information Systems 27(3) (2006) 267–289.

[24] C. Panagiotakis, N. Pelekis, I. Kopanakis, E. Ramasso, Y. Theodoridis, "Segmentation and Sampling of Moving Object Trajectories based on Representativeness", IEEE Transactions on Knowledge and Data Engineering, 07 Feb. 2011. IEEE computer Society Digital Library. IEEE Computer Society.

[25] N. Pelekis, I. Kopanakis, E. Kotsifakos, E. Frentzos and Y. Theodoridis. "Clustering Uncertain Trajectories", Knowledge and Information Systems (KAIS), DOI 10.1007/s10115-010-0316-x, 2010.

[26] E. Frentzos, K. Gratsias, N. Pelekis and Y. Theodoridis. "Algorithms for Nearest Neighbor Search on Moving Object Trajectories". Geoinformatica, 11:159–193, 2007.

[27] N. Pelekis, I. Kopanakis, E. Kotsifakos, E. Frentzos and Y. Theodoridis. "Clustering Trajectories of Moving Objects in an Uncertain World", In the Proceedings of the IEEE International Conference on Data Mining (ICDM'09), Miami, U.S.A., 2009.

[28] O. Abul, F. Bonchi, M. Nanni, Never Walk Alone: uncertainty for anonymity in moving objects databases, in: Proceedings of the 24nd IEEE International Conference on Data Engineering (ICDE'08), 2008.

[29] Lee, J.-G., Han, J., and Li, X. (2008). Trajectory outlier detection: A partition-and-detect framework. In ICDE, pages 140–149. IEEE.

[30] Prelipcean, A.C., Gidofalvi, G., and Susilo, Y.O. Measures of transport mode segmentation of trajectories. International Journal of Geographical Information Science, 2016.

[31] Ahmad, M., Karagiorgou, S., Pfoser, D., Wenk, C.: A comparison and evaluation of map construction algorithms using vehicle tracking data. GeoInformatica Journal (2015).