Predicting User Activities using GPS Collections

Simona Andrieș
Faculty of Computer Science,
"Alexandru Ion Cuza" University of Iasi
General Berthelot, No. 16
simona.andries@info.uaic.ro

Adrian Iftene
Faculty of Computer Science,
"Alexandru Ion Cuza" University of Iasi
General Berthelot, No. 16
adiftene@info.uaic.ro

ABSTRACT

The purpose of this paper is to retrieve as much information as possible from a set of GPS coordinates. The information obtained and presented in this study are the most frequented places of a user and the modes of transport used to get from one place to another. The considered modes of transport are: walking, biking and travelling by car, train or airplane. This kind of information can be used in many types of applications like recommendation systems or it can be studied in order to have a better understanding of the human activity and find new ways to improve it.

Author Keywords

Prediction; DBScan; GPS.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors; Design; Measurement.

INTRODUCTION

The everyday schedule of people gets more and more busy, but their desire to be up to date with the activities of their friends and family doesn't change. Another great factor that should be considered is that since the free time is so short, it is important not to waste it in traffic, searching for an address in a new neighborhood or having unpleasant experiences in a restaurant. Consequently, the need to always improve applications that register and manage routes or recommendation systems applications appears. This type of application can be used by people with an active life or by the companies that need to manage the routes made by the employees in a day.

The GPS coordinates of a user represent an extremely important source of information, which utilized in a good way can bring a lot of benefits. For example, with the help of a recommendation system like Google Traffic [6], the user can choose a route to a certain location that is longer than the usual route, but that doesn't have traffic jams and this way a lot of time is saved. Another example is that if during the city holidays the traffic is busier than usual in certain areas of the city, it can be inferred that the next year, in the same period, the traffic will be busy also. Based on this information a recommendation can be made that could help a lot of users.

Nowadays, mobile phone and especially smartphones are something extremely common among people. These phones are equipped with GPS receptors, they provide wireless access to internet or 3G/4G services and this makes social network applications available to users at any moment of the day.

Intelligent Transport Systems [8] represent a suite of technologies meant to bring improvements to the different modes of transport and how they are managed. They range from navigation systems, traffic lights control systems, automatic plate number recognition and highway surveillance cameras to parking sensors.

One of the main technologies that is part of Intelligent Transport Systems is Global Positioning System (GPS). According to [3], the GPS functionality is based on using space satellites as reference points for localizing points on earth. By using a very precise measurement of distance in a straight line between the receptor and at least dour satellites, the position (latitude, longitude, altitude) of any point of earth can be determined. The distance between the satellite and the receptor [3] is calculated by timing the period needed by the radio signal to get from the satellite to the receptor. The data registered by GPS equipped devices will be studied in order to determine the user's locations of interest and to find the modes of transportations used to get from one place to another.

The application proposed in this paper represents a way to facilitate the way GPS trajectories logged by devices can be analyzed. Based on these trajectories information of interest can be determined that are important for the user, for the companies that monitor the activities performed by employees on the field and last but not the least to public institutions interested in improving traffic in cities. Having this kind of information available, they can take better decisions or consolidate the opinions made by using other kinds of observations or studies.

The main information that can be obtained based on GPS trajectories are: the locations of interest of a user and the amount of time spent there, the modes of transport used to get from one place to another and the routes that are most commonly used by a user. In this paper the first two points will be discussed the most.

SIMILAR APPLICATIONS

In the last years there were developed more and more applications that help the users to be connected with their close ones or recommendation system applications that offer suggestions based on preferences like Geolife.

Geolife [5] a Microsoft project, is a social network based on locations that helps users connect and share between them their experiences. This application analyses the location history of the users in order to determine the most popular routes and locations among them with the purpose of creating recommendations for trips or everyday routes. Based on this information and also on the circles a person is present in, the application measures the similarity between users and makes personalized recommendations (see Figure 1).

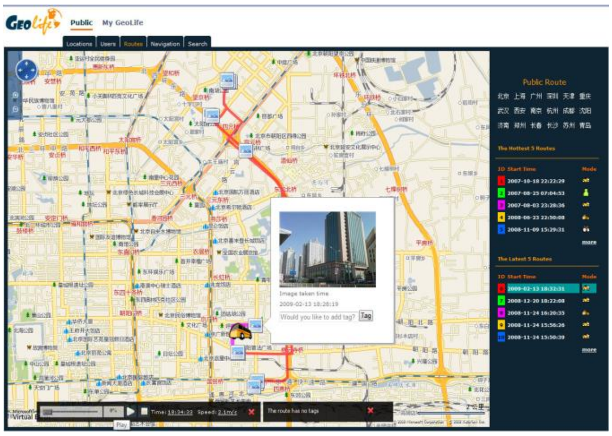

Figure 1. GeoLife application [10]

By analyzing the location history using Data Mining techniques, Geolife determines the most important locations for a user or his frequent travelling routes. Other functionalities offered by this application are the suggestion of locations of interest based on the recommendations made by other users or to suggest itineraries that can be used while travelling. A common problem encountered while developing recommendation systems is to find a reliable way to calculate the similarity between users using their location history and deciding the level of interest a certain location will have for a user.

According to [19], another common difficulty encountered in this research area is to find the mode of transport of a user. People can change frequently the mode of transport in a day and some common problems are given by the always changing traffic conditions from big cities or by the poor quality of the GPS coordinates registered by devices sometimes.

There are a number of studies that have the purpose of determining the mode of transport used on a route based on a set of GPS coordinates. The majority of them identify the segments covered by foot, bike, car and train and are mainly based on the speed calculated between consecutive points. In addition to these studies [2, 14 and 15], this paper brings as a novelty the possibility to determine segments travelled with the airplane. Another difference is that in order to correctly determine the routes travelled by train the true north direction is used.

The studies that identify also the public modes of transport use GIS data or other kinds of algorithms like the one proposed in the study [18]. This algorithm is based on a group of segments that are labeled with the mode of transport organized in a graph as a base. An edge represents the route of the user from a point to another and for each edge a weight is calculated. This weight represents the distribution of probabilities for the possible modes of transport and each new segment is classified based on the previously computed probability for the edge corresponding to the segment. This method has the disadvantage that it requires a very big initial set of data in order to classify new segments with a good precision.

GPS Trackmaker [7] is an application that allows the upload of files containing GPS logs with the purpose of displaying a map with the travelled routes and computes some metrics based on it like acceleration, average speed and maximum speed.

GPS DATA ANALYSIS

In this study a set of GPS Tracks from Microsoft Research Asia Geolife project [4] was analyzed. These GPS tracks [19] were collected from 182 users over a period of five years. The majority of the coordinates are registered in Beijing, but also in other cities from China. This set of data consists of a sequence of GPS coordinates accompanied by time stamps. As specified in the user manual [19] of the GPS tracks, the trajectories were logged using different types of devices equipped with GPS receptors and have different recording times ranging in the interval 1-5 seconds or 5-10 meters. The data set encompasses a variety of movements of the users, ranging from the routine home – work and back to sports activities or relaxation.

The data set is analyzed with the purpose of extracting new information from a big quantity of data by summarizing the information in order to make it easier to analyze. Based on the information gathered from this analysis statistics can be created to determine the most popular mode of transport, the busiest time of the day in a certain place and many others.

Simultaneously, this data set can be used in a wide area of research like location recommendation or to study user trajectories. In this paper it was used to determine the locations of interest, the amount of time spent there and the modes of transport used to get from one place to another by the user.

The paper Applying Data Mining in Prediction and Classification of Urban Traffic [11] proposes a way to classify data in order to understand and predict traffic based on decision trees. A decision tree has the structure of a flowchart. In this kind of structure each internal node is tested for a certain property and each branch of a node represents the result of a test. The leaves of the tree represent a class and the main node is called a root. The resulting tree can be used for classifying information because each new record passes through the root node and after a test it is passed on to the next node. This process goes on until a leaf node is reached that represents the class (decision) and the path to the leaf node is a set of rules.

In this paper decision trees were used to classify the segments resulted from the splitting of the trajectory based on certain criteria. The attributes of the segments that were used in the classification are: speed, maximum speed, altitude, total time and the distance of the segment. As previously stated, the tree leaf node represents the decision and it can have one of the values: walk, cycle, car, train, and airplane.

Mining Important User Places

One of the most important things that can be discovered based on the GPS coordinates are the most important locations for the user, where he spends the most time, or stops by regularly for a shorter or longer period of time. Figure 2 presents a basic routine covered by a user in a day.

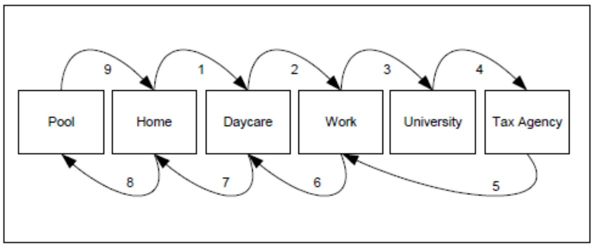

Figure 2. Typical places visited by a person in a day [Zhou, Bhatnagar]

How can we determine these stops? An idea would be to consider a stop a place where the speed is 0 km/h for a minimum period of time. This approach has limitations because GPS devices can have signal shortages, or the speed could be close to 0km/h because the user is stuck in a traffic jam. Moreover, a stop is not so easy to determine, because even if the user is standing still, the captured GPS logs are not 100% precise and can have noises [22].

Other approaches use more sophisticated techniques like splitting a journey in trajectories. This way, a 5 minute break indicates a stop, whereas a 5 minute trajectory at a very low speed indicates that the car is stuck in a traffic jam. Andrienko [1] and Zheng [21] consider in their studies that a location is important if a person spends a lot of time there. Stops are considered as consecutive GPS coordinates such that the distance between the points doesn't exceed a certain threshold (e.g. 50 meters) and the time spent there exceeds a minimum threshold (e.g. 30 minutes).

CB-SMoT [12] considers that a stop is determined based on the variation of speed on a trajectory. This way, the stops are the segments where the speed is lower than the average speed of the trajectory. A stop during a trip indicates that we got to a location of interest or close to one. An interesting area of trajectory study is determining and labeling the points of interest (POI) that motivate the stop and, if none is found, labeling the stop with "unknown".

Determining POI for an User

In order to determine the POI for a user 3 clustering algorithms were studied in order to find the most suitable one. The studied algorithms are: K-means, DBSCAN and DJ-Clustering [16 and 22]. The density based clustering algorithms like DJ-Clustering and DBSCAN need two variables: Eps (distance between nodes) and MinPts (minimum number of points needed to form a cluster). These values must be chosen according to the density of the registered GPS logs.

The main idea of the DJ-Clustering algorithm is to group close objects and represent them using the minimum bounding rectangle. The data structure used to represent the information is an R-Tree where each node stores two things: a way to identify the children nodes and the bounding box of the children nodes. In order to have a better understanding of how this algorithm works a library was used to notice the results. The Figure 3 shows the result of applying the algorithm on the data set.

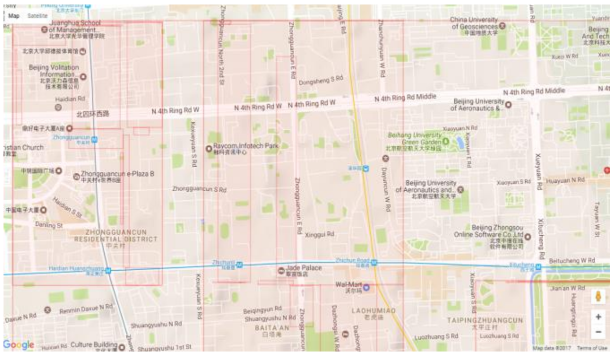

Figure 3. The result of the application of the DJ-Clustering algorithm

After analyzing the three algorithms, the conclusion was that the algorithm suitable for resolving the problem of determining the points of interest is DBSCAN. DJ-Clustering is more useful for resolving other kind of problems like finding all the locations of interest around a certain point and K-Means clustering has the disadvantage that it requires as input the number of clusters that need to be obtained, and this number can vary a lot depending on the size of the data set and the frequency of the GPS coordinates or the period of time over which the coordinates were registered. In our case, if the period is very short, it should be possible to get 0 clusters, so setting a predefined number of clusters is not helpful in this case.

In Figure 4 can be noticed a couple of locations of interest determined for a user using DBSCAN clustering algorithm and some of the tracks of a user from a week. Each line color represents a different day.

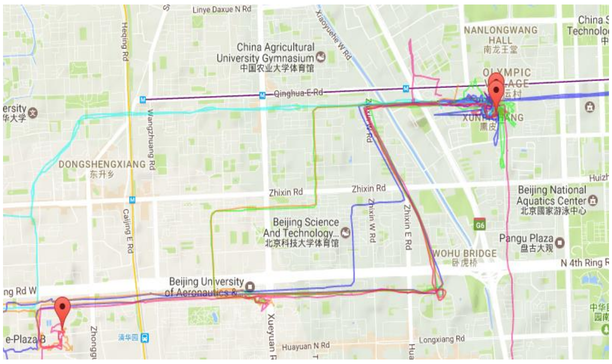

Figure 4. Point of interests for User 052, 14 - 20 July 2008

Determining Segments for a User

The concept of trajectory is associated to the always changing position of an object over a period of time, so a trajectory is spatio temporal concept. The concept of journey means that the movement of an object has the purpose of travelling from one point to another. A journey take a finite amount of time and covers a distance in space (SPACCAPIETRA et al., 2008).

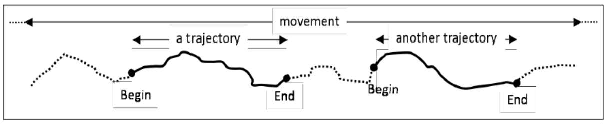

Figure 5. Trajectories extracted from a list of GPS coordinates [13]

A lot of applications are not interested in keeping and analyzing big volumes of data, but only the segments they are interested in. Figure 5 represents with interrupted line a section from the GPS logs of a person and with a full line two sections that were identified as relevant. Each trajectory is defined using spatio temporal coordinates called Begin and End.

Among the information that can be determined based on GPS coordinates there is the possibility to find common features of trajectories. This is based on evaluating similarities that allow defining classes of trajectories. The most important techniques that are based on machine learning are classification and cluster analysis. In this paper, a trajectory is formed from a sequence of GPS coordinates that represent the trip made by a user in order to get from one point to another and can be covered using multiple modes of transport. A trajectory is then split in segments, each segment having associated the transportation mode that was used.

Yu Zheng, in the paper Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web [20] considers two premises in order to do the segmentation of the GPS coordinates. The first one is that the change of transportation mode must be done by walking and the second one is that between segments there must be a break or a stop where the speed is very close to 0 km/h.

The segmentation of GPS coordinates takes place in two steps and it also has a preprocessing and post processing step. The preprocessing has the purpose of reducing the errors that could appear because of wrong GPS coordinates. The preprocessing algorithm is based on the distances computed relative to five points (A, B, C, D, and E). For each of them a sum of the distances to the other four points is made. If the sum reported to C is the biggest, the coordinates of point C are replaced by the point resulted by calculating the midpoint between C and the point given by the minimum between the segments (AC, BC, CD, CE).

The first step is to split the journey in trajectories (separated by longer stops) and the second step is to split the trajectories in segments corresponding to the different means of transportation used. The segmentation process leaves from the premise that every person switches multiple modes of transport while travelling and this is why the segments are split in two types: walking and non-walking. Then, the non-walking segments are classified using the decision tree.

The split of the journey in trajectories is made determining the places where the user spends at least 30 minutes. If such a place is found, it means that the previous point is the end of a trajectory and the next point represents the beginning of a new trajectory. If no such place is found it means that the whole set of data represents a single trajectory.

One of the difficulties encountered in the segmentation process is given by what wrongly seems to be a very frequent change or transportation mode that is impossible. This is caused by the poor traffic conditions or by the stops at traffic lights. To solve this issue, at the end of the classification a post processing is made to merge the segments that have a very short duration.

Determining the means of transport used

One way of segmenting the trajectories is using the means of transport used by a person. This information is useful in planning the public transport system. This type of segmentation is based on the main features of each transport means that can be used. Usually, a walker has a maximum speed of 5 km/h, while using a bicycle a person can reach around 10 to 15 km/h and the buses have clear paths.

Based on speed, acceleration, and gear change rate, and using a technique such as supervised learning or decision trees, we can determine for each trajectory the means of transport used. This information is important because, for many people, the route to the office involves walking, one or more means of public transport and walking again. A library was used to build the decision tree, and the algorithm that classifies segments based on the built-in tree was written for this application.

Identifying the means of public transport used is not trivial because many aspects need to be considered. Not just the use of recorded speed is sufficient, as traffic conditions can cause unusual variations, but a machine may have a speed similar to that recorded by a bicycle. And often, people often use more means to get from one point to another.

One important thing to keep in mind is that most public transports have unique features that can help us identify the means of transport used in a particular segment of a journey. A common point for all means of transport is the average speed of each. For this work, we have set some intervals specific to each transport mode based on the data set available. Because some users have annotated their trajectories with very important information, the means of transport used over time, we used this information to deduce the typical speeds at the locations where the GPS coordinates were recorded. Thus, we determined all the GPS coordinates specific to the annotated time interval, and we calculated the average speed for each record based on them.

Some means of transport are easier to identify than others because they have some defining features that eliminate any other possible variation. The *plane* is the easiest to identify because of the flying altitude. The cruising altitude for a line jet is about 10,000 m. And another important feature is the speed of about 850 km/h.

The next means of transport that can be relatively easily identified is *walking*. This is because speed is very small and constant throughout the route. The range used in this work for the average speed is 2-7 km/h.

The *bicycle* is the next transport vehicle analyzed. Beginning with this means of transport, precise identification of the means of transport becomes very difficult, and this is due to the fact that there may be periods of time when the user seems to walk or stop when he is actually at the traffic light. The average speed considered for this means of transport is in the range of 7 to 20 km/h.

The means of transport that can be found among users' annotations are *car*, *bus*, *taxi*. These were treated in this paper as one, namely *car*. In order to identify the means of public transport, it is necessary to use GIS data, which is not part of the scope of this paper, but which may be an improvement that is worth considering for the future.

The next mode of transport is the *car*, which typically has average speeds between 20 and 120 km/h. This means of transport is the most complex because it can "borrow" characteristics from the means of transport described above in situations that are very common. These situations are stopping and starting from traffic lights and last but not least traffic jams. Very often, track sequences such as car-stop-car or car-bicycle-car or combinations of these two have a very short duration for the intermediate route. Starting from a logical premise, namely that a person can not have such a route in real life, we have combined these sequences from the paths into one. This step of combining the segments of the route was done at a later stage of their classification.

The *train* is the last means of transport analyzed in this paper. The minimum speed considered for this means of transport is 120 km/h and the maximum speed is 250 km/h. As anyone can tell, the train and the car have many similarities due to the wide range of journeys that can be made using these means of transport. Therefore, the determination of this means of transport takes place in two steps. The first takes place in the segment classification using a decision tree, this classification is based on the average and maximum speed, the distance traveled and the duration of the route. Practically, the annotated results following classification may restrict the type of trains determined on high-speed or interurban trains. The second step occurs after segment classification and is based on direction. Bearing [17] is the angle between the North Pole and the trajectory line of a moving object. What distinguishes the car from the train is the angle with which it can make curves. A car can at any time make a curve that has an angle of 90 degrees, while a train can make this curve, but much wider and at a greater distance. In this step, the direction is analyzed, and if there is no sudden change of direction, the labeling of a segment previously tagged with a car on the train changes.

CASE STUDIES

Due to the presence in the data set of tracers accompanied by labels, we have the possibility to evaluate the accuracy of the results obtained in this paper. Here are some examples of the data set.

Use Case 1 – Train (User 010, 28 March 2008)

This case aims to analyze the accuracy with which trajectories traveled by train are identified. Figure 6 shows the user-labeled segments, and Figure 7 shows the result obtained by the proposed application.

Figure 6. Representing data labeled by the user (Use Case 1)

For this user, there are periods where no GPS coordinates have been recorded, which are labeled automatically with missing data. As can be seen from Figures 6 and 7, a difference from the data labeled by the user is given by the identified stops. It is very likely that these are stations where the train has stopped, and the proximity of a walk towards the stop can be an extra argument for this hypothesis.

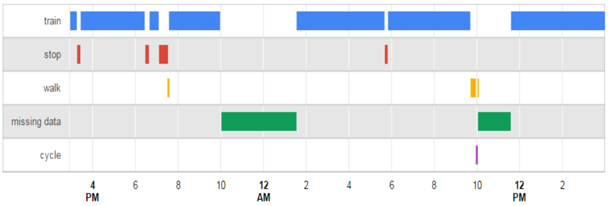

Figure 7. The result obtained by the application (Use Case 1)

Use Case 2 – Airplane (User 128, 30 September 2008)

The purpose of this case is to demonstrate the accuracy with which the airplane is identified. As can be seen in Figures 8 and 9, the airplane is identified with very good accuracy. The same can be said about the train. We can also notice that some segments have been mistakenly labeled with the bike because of the speed too low (see Figure 10) with which the vehicle is moving and that there were generally high traffic fluctuations in the city.

As in the previous example, we can see labeled segments with missing data, the lack of GPS coordinates being a common problem.

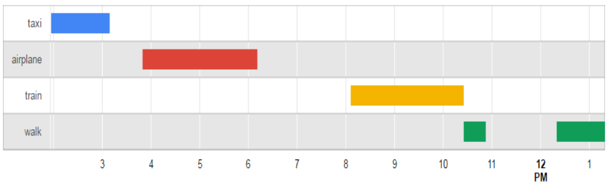

Figure 8. Representing data labeled by the user (Use Case 2)

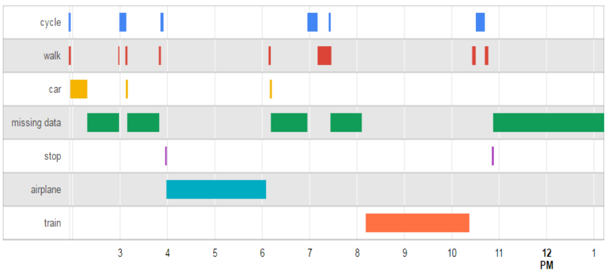

Figure 9. The result obtained by the application (Use Case 2)

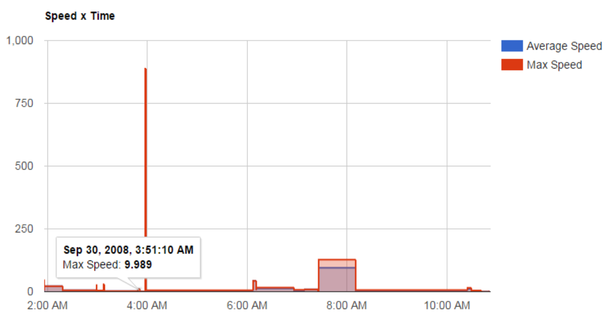

Figure 10. Speed variation over time

Use Case 3 – Walk-car-walk-car (User 179, 21 August 2009)

The purpose of this case is to illustrate the behavior of the application for the alternative walk-car-walk-car segments.

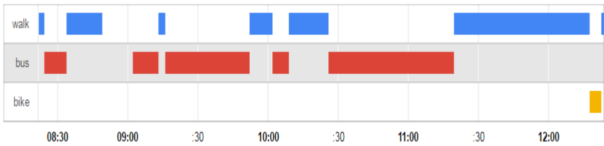

Figure 11. Representing data labeled by the user (Use Case 3)

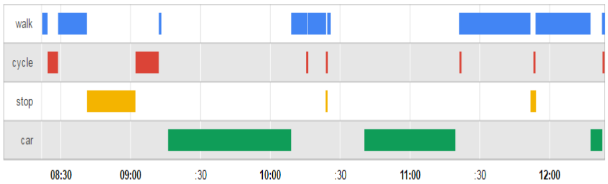

Figure 12. The result obtained by the application (Use Case 3)

As you can see in Figures 11 and 12, the first two walk segments have been identified with good precision. The first bus segment was misidentified as a bicycle because of its very low speed (14 km/h). We can also see that a walk segment is not identified, but is embedded in a car segment. Among the segments labeled by the user, one can also see a labeled bike that was not correctly identified due to the very high speed recorded (38 km/h).

Although the segment identification algorithm has a pre-processing stage in which noise has been reduced, it may not be completely reduced. The third and fourth cycle segments were identified differently due to high speeds. At the same time, it is possible for the user not to explicitly label all means of transport. He could, for example, run when he wants to catch the green color at a traffic light, thus influencing the outcome of the algorithm. Running is not part of the possible means of transport, users who tagged the segments did not have this option.

Use Case 4 – Bicycle (User 020, 28 October 2011)

This case aims at analyzing the accuracy with which the bicycle routes are identified and analyzing a variant that would improve the accuracy of the determination of the means of transport used.

Figure 13 shows a segment labeled by the user that has been wholly completed by bicycle. As can be seen, Figure 14 shows two variants of classification of segments traversed. The first variant uses the maximum speed of all the speeds calculated between each two points. The second variant uses the maximum speed identified based on an index (0.95 * number of points) between determined speeds (ordering upwards), thus ignoring the noise points or at least part of them. For the case studies presented above, it is worth mentioning that there is no difference between segment labels in these two cases.

An alternative to the 95th percentile could be the use of an algorithm that eliminates noises such as the Kalman Filter [9] that estimates and corrects the coordinate position.

Figure 13. Representing data labeled by the user (Use Case 4)

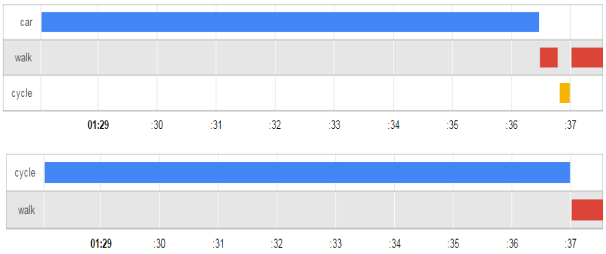

Figure 14. The result obtained by the application before (up) and after (down) application of 95th percentile (Use Case 4)

CONCLUSION

From a large amount of information, we had originally obtained some easy-to-read information by anyone. For good viewing, interest locations have been marked on a map, and the result of segment classification has been displayed in a graph. As we saw, we used a clustering algorithm to determine where the user spends the most time. One of the limitations of clustering algorithms is that there is no exact method to determine if a cluster is correct. The user of the algorithm must decide whether the data resulting from the clustering satisfies the problem it wants to solve. Abnormal data obtained by applying clustering algorithms may correspond to data of major interest, or may not be representative and removed prior to clustering or post-processing.

One of the advantages of the proposed solution is that it can be easily improved in the future by adding new classification criteria to the decision tree. One of the criteria that may be added is acceleration, and adding new classification criteria improves the precision with which segments are classified. Also, in order to get a better result, it would be worth analyzing to see if there are algorithms that completely remove the noise from the dataset.

We also look into the future and based on the locations of interest determined and the routes used to get from one point to another one can build a chart that could be used to analyze frequently visited routes by the user.

ACKNOWLEDGMENT

This work is partially supported by POC-A1-A1.2.3-G-2015 program, as part of the PrivateSky project (P_40_371/13/01.09.2016).

REFERENCES

1. Andrienko, N., Andrienko, G. Designing visual analytics methods for massive collections of movement. *Cartographica*, 42(2), (2007), pp. 117-138.
2. Biljecki, F., Ledoux, H., Van Oosterom, P. Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 27(2), (2013), 385-407.
3. El-Rabbany, A. *Introduction to GPS: The Global Positioning System*. Artech House, Boston, London. www.artechhouse.com, 194 pages, (2002).
4. GeoLife GPS Trajectories: http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/ (Accesed last time on 1 June 2017)
5. Geolife: http://research.microsoft.com/en-us/projects/geolife/ (Accesed last time on 1 June 2017)
6. Google Traffic: https://www.google.ro/maps (Accesed last time on 1 June 2017)
7. GPS Trackmaker: http://www.trackmaker.com (Accesed last time on 1 June 2017)
8. Intelligent Transport Systems: http://www.etsi.org/technologies-clusters/technologies/intelligent-transport (Accesed last time on 1 June 2017)
9. Kalman, R.E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*. 82 (1), (1960), 35–45.
10. Microsoft UseCase: https://www.microsoft.com/en-us/research/project/geolife-building-social- networks-using-human-location-history/ (Accesed last time on 1 June 2017)
11. Nejad, S. K., Seifi, F., Ahmadi, H., Seifi, N. Applying Data Mining in Prediction and Classification of Urban Traffic. *World Congress on Computer Science and Information Engineering,* (2009).
12. Palma, A. T., Bogorny, V., Kuijpers, B., Alvares, L. O. A clustering-based approach for discovering interesting places in trajectories. In Proceedings of the 2008 ACM symposium on Applied computing (SAC '08). ACM, New York, NY, USA, (2008), 863-868.
13. Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., Theodoridis, Y., Yan, Z. Semantic trajectories modeling and analysis. *ACM Comput. Surv*. 45, 4, Article 42 (2013), 32 pages.
14. Schuessler, N., Axhausen, K. W. *Processing GPS Raw Data without Additional Information*. 15 pages, (2008).
15. Stopher, P., Clifford, E., Zhang, J., Fitzgerald, C. Deducing Mode and Purpose from GPS Data. *Working paper ITLS-WP-08-06*. Institute of Transport and

Logistic Studies, the Australian Key Center in Transport and Logistic Management, the Univiersity of Sydney, (2008).

16. Tan, P. N., Steinbach, M., Kumar, V. *Introduction to Data Mining*. Addison – Wesley, (2006).

17. Rutstrum, C. The Wilderness Route Finder. *University of Minnesota Press*, (2000).

18. Zheng, Y., Chen, Y., Li, Q., Xie, X., Ma, W. Y. Understanding transportation modes based on GPS data for web applications. *ACM Trans*. Web 4, 1, Article 1 (January 2010), 36 pages, (2010).

19. Zheng, Y., Fu, F. Xie, X., Ma, W. Y., Li, Q. *Geolife GPS trajectory dataset* - *User Guide*, 26 July 2011.

20. Zheng, Y., Liu, L., Wang, L., Xie, X. Learning transportation mode from raw gps data for geographic applications on the web. In *Proceedings of the 17th international conference on World Wide Web* (WWW '08). ACM, New York, NY, USA, (2008), 247-256.

21. Zheng, Y., Zhou, X. *Computing with Spatial Trajectories*. Springer, 2011.

22. Zhou, C., Bhatnagar, N., Shekhar, S., Terveen, L. Mining Personally Important Places from GPS Tracks. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*, Istanbul, Turkey, (2007).