

Adapting the User Path through Trajectory Data Mining

João Ramos, César Analide, José Neves, and Paulo Novais

Algoritmi Centre, Department of Informatics
University of Minho
Braga, Portugal
{jramos, analide, jneves, pjon}@di.uminho.pt

Abstract. The pervasiveness of location based services such the GPS on mobile devices enabled the gathering of massive spatial-temporal data. These databases enabled the mining of new data in order to calculate frequent patterns and predict the movement of the objects. In the development of our system for guiding the user with cognitive disabilities (CogHelper) we are applying a trajectory data mining to adapt and adjust the path to user preferences. Indeed, the guiding process may be more useful and it increases the quality of life of the user through this new functionality (conciliated with the speculative computation module). Thus, instead of the user having to adapt to the application we are developing a system that adapts the path to the user. Rather than being guided through the shortest path he may be oriented by a longer but preferred path. The main contribution of this paper is the specification of the trajectory data mining which is incorporated in CogHelper system.

Keywords: Trajectory Data Mining, Frequent Patterns, Sequence Mining, Clustering

1 Introduction

Location-acquisition technologies are becoming more embedded in people's daily life. Examples of such systems are mobile devices with GPS or GSM, and computers that if connected to the Internet may be founded through their IP address, just to name a few. This data acquisition results in larger amounts of location-based information and the need to treat and obtain results in such collected data becomes imperative. By mining the frequent trajectory patterns in the spatial-temporal database one may detect the regularity of moving objects (e.g., people or vehicles). Thus, it is possible to define frequent patterns and, with this data, predict the movement of objects.

Several authors have studied and developed different algorithms/methods for mining spatial-temporal data [16,10,11,8,1]. The different algorithms are described in more detail in Section 2. The spatial-temporal trace generated by a moving object is defined as spatial trajectory [17] and is represented by a set

of chronologically ordered points. Each point is constituted by its coordinates (latitude and longitude) and a timestamp. According to Zheng [17] there are four sources of data: mobility of people, mobility of transportation vehicles, mobility of animals, and mobility of natural phenomena. In our case (as described in Section 4) we are focused on the mobility of people and the data gathering may be subdivided in active and passive recording. The former is related to the GPS acquisition whereas the latter is collected by GSM connection using the cell towers. Despite the method for acquiring the data and before using the trajectory data there is a fundamental step where the data is preprocessed [17]. Here a number of issues are considered in order to prepare the data for the data mining. Thus, it is important to filter the noise (points where the GPS error is bigger than the usual), to segment and to compress the data (without removing its utility), and to do a map matching in order to have a more accurate path.

Considering our previous work (CogHelper, see [14,15]), we intend to implement a module of trajectory data mining where the goal is to adapt the path to the user preferences. Thus, instead of guiding the user through the shortest path, the system may suggest a longer but preferred traveling path.

This paper is organized as follows: in Section 2 it is presented the related work regarding the existing algorithms and their use in trajectory data mining. In Section 3 it is provided a brief description of our system and its main goal. In Section 4 we detail the goal of applying a data mining algorithm in our system considering the adaptation of the path to the user. Finally, in Section 5, the main conclusions are drawn.

2 Related Work

There are several models and methods to efficiently query spatial-temporal databases in order to obtain different types of information like frequent trajectories, the type of trajectory, among others.

Alvares *et al.* [1] changed the usual paradigm of considering a trajectory as a set of ordered points to a semantic point of view. Thus, the authors provided an extraction method to obtain the frequent movements between two places. Instead of knowing the exact trajectory between place A and place B, Alvares *et al.* tried to know to which place the user would travel considering his origin. In their work the authors considered a trajectory as a set of stops and moves, in which stops are important places where the moving object has not moved. The moves are delimited by two stops and by a time interval and are part of the trajectory. Indeed, the authors are able to detect frequent stops during a time interval, frequent moves, duration of stops, among others. The main goal was to find the frequent moves between two stops, for example, detect that if the user is on place A he will travel to place C.

To Giannotti *et al.* [8] a trajectory pattern is a set of trajectories where the moving objects visit the same places in the same sequence and the traveling times are very similar. Indeed, the authors define regions of interest and the typical travel time of the moving objects. Like Alvares *et al.* [1], in Giannotti

et al. [8] the trajectory pattern establishes the pattern that if the object is in A it will most likely travel to B, disregarding the particular route between the two places. In their trajectory method, the authors initially convert the raw data (sequential spatial-temporal data) into a sequence of regions of interest and then, in a second phase, the trajectory data mining algorithm is used to detect the movements between two places. A region of interest represents an area where the number of collected raw data is very significant, meaning that the moving object has not change too much its position.

To Ye *et al.* [16] the goal was to detect people's life patterns. The authors, in a first step, convert the GPS raw data into a sequence of stay points, which are geographic regions where the user stays for an amount of time (pre-established threshold). Then they apply a density-based clustering method (in their case the OPTICS [2] was used) to obtain a sequence of location history, *i.e.*, sequence of clusters that represents points where the user stays for a period of time. With these sequences, the authors are able to treat the data and get results such as the user habits during the week or in the mornings, among others.

WhereNext is a location predictor from Monreale *et al.* [11]. Like the other described works, here the authors convert the GPS log data to trajectory patterns. To obtain these patterns the authors adapted the algorithm proposed in [7] and they represent sequences of dense spatial regions. Then, the authors build a tree considering the previously generated patterns and their connections. Using the tree the authors are able to predict the next movement of an object by applying different prediction strategies. To create the initial trajectory patterns Monreale *et al.* [11] use the total set of points that are within a given region. Thus, instead of using just the historic data of a given object, they use the data of all moving objects to create the patterns in a region.

In [10], Lee *et al.* try to find trajectories that are frequently repeated, being able to analyze and predict the next movement of the objects. The authors use the spatial and temporal attributes of the collected data to mine the frequent trajectories. In their approach the authors initially convert the raw data to create a graph from which they create trajectory lists. These lists (in the form of a tree) indicate to where the moving object usually travels when it is on a specific location (represented by a node in the tree). After this step Lee *et al.* transverse the mapping graph through a developed graph-based algorithm and detect frequent patterns.

3 CogHelper: System Description

CogHelper is an orientation method for people in the earlier stages of cognitive disabilities, *i.e.*, with mild and moderate disabilities. In the advanced stages of this condition (severe and extreme) the person may not be able to use a mobile device, thus the orientation is not viable. Being an on-going work the main features of the system have been described in [13,14,15]. The goal is to guide the user by using an augmented reality interface (Figure 1b) in which we try to minimize the cognitive effort needed to understand the direction that should be

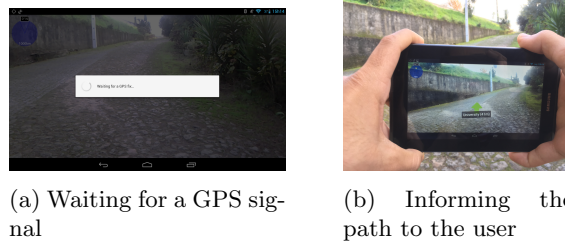


Fig. 1: Orientation method for people with disabilities - Augmented reality interface

followed. While the system is calculating the best traveling path a screen appears informing the user to wait (Figure 1a). Through a speculative computation module the system is able to predict and anticipate possible user mistakes however, as described in Section 4, the adapted traveling path is obtained through a trajectory mining module.

Although not being the main users, caregivers are also another type of people who may need to execute the application. Indeed, we have developed a localization feature in order to let caregivers know, in real time, the current location of the person with cognitive disabilities. This feature is very important since caregivers may develop a different activity while knowing the location of the person under care, not having to physically be with him.

To add more functionalities to the system it is also possible to connect to external services such as iGenda [4]. Through this external system whenever the user has an appointment the guidance process may be remotely started in order to guarantee that he will not arrive late. Thus, by applying the concept of Ambient Intelligence (AmI) it is possible to provide better assistance to the patients [5]. Indeed, sometimes the care may be more effective if the patient is in his home environment instead of being relocated to a nursery home [3].

4 CogHelper: the Trajectory Mining Module

CogHelper [15,14] is a system where the main goal is to guide people with mild or moderate cognitive disabilities. In more extreme cases of this diagnosis the person may not learn how to use a mobile device, thus his orientation is not possible. The development of the system has also a parallel goal that is to provide a localization feature for caregivers, which let these type of users to know in real time the current location of the person with disabilities.

To guide the user the system has a speculative computation feature that is able to predict a user mistake and alert him in advance. Thus, instead of turning into the wrong path the system may ensure that the user is traveling in the correct one. The speculative computation module, in order to anticipate user mistakes and assure the correct traveling path, needs to be fed with the traveling route. With the goal of adapting the traveling path to the user preferences we

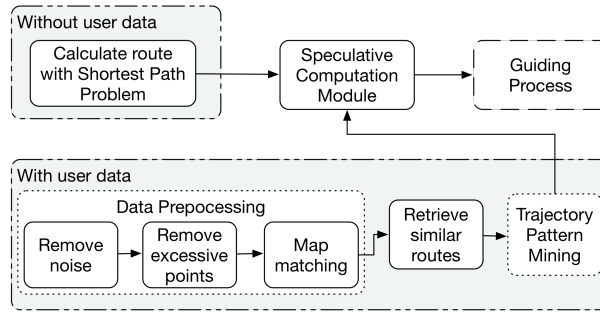


Fig. 2: Process for generating adapted routes

are implementing and testing a trajectory data mining. Thus, instead of guiding the user through the shortest path he may be oriented by a longer but preferred (and more familiar) trajectory.

A trajectory of a person with cognitive disabilities (also defined as user) is a sequence of time-stamped locations obtained by the GPS of the mobile device. The location is composed by Cartesian coordinates, *i.e.* latitude and longitude, and the timestamp. The active recording of the user location (from the GPS) is done whenever the user is receiving the necessary aid from the application. The process for mining the trajectories are depicted in Figure 2. When there is no information about the user (*e.g.*, the person is using the application for the first time) the system calculates the shortest path and retrieves it to the speculative computation module. The shortest path is calculated using an external tool named GraphHopper [9]. When there is previous information about the user, *i.e.*, there is spatial-temporal data from previous trajectories the system uses the available data in order to provide the best trajectory to the user.

To adapt the path it is necessary to do a pre-processing step in order to prepare the available raw data for the mining process. Indeed, and as depicted in Figure 2, it is necessary to remove the noise of the data, remove points that may not increase the detail of the trajectory and do the map matching. The first step of this pre-processing phase (noise removal) eliminates locations whose GPS error was notorious, *i.e.*, points that for some reason (like the reflection of the GPS signal in a building) have an error bigger than the usual GPS error. Then locations that may not have added value to the mining are also removed, *e.g.*, locations that are in the middle of a straight line. Finally, the map matching is the process of getting the remaining data and match it to the roads defined, in our case, in Open Street Map¹ database. Figure 3a represents the raw data recorded from the GPS of the mobile device whereas Figure 3b is the final step of the pre-processing stage (including noise removal and map matching). For this pre-processing phase we use GraphHopper Directions Api [9] which has a map matching feature and is an open source tool.

¹ Open Street Map, available at <http://www.openstreetmap.org/>

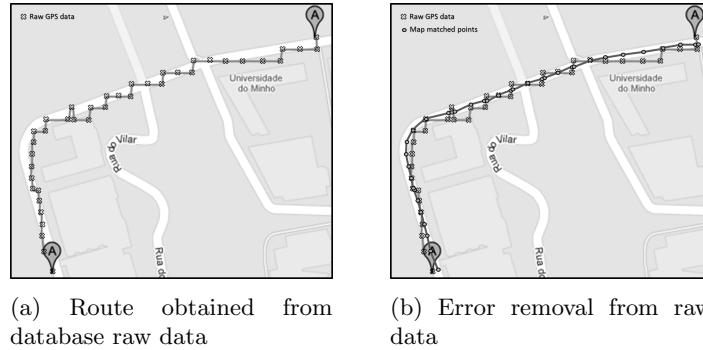


Fig. 3: Trajectory data pre-processing [15]

After the pre-processing step we are able to apply the trajectory data mining method in order to extract the frequent trajectories (patterns) from the data. For this process and in our preliminary tests we are experimenting, with good results, two different methods: the PrefixSpan algorithm [12], which mines frequent sequential patterns; and OPTICS [2], which is a clustering algorithm that groups the locations in clusters. Both methods are based on an open-source API named SPMF [6], which provides an implementation of such algorithms. Through PrefixSpan algorithm we may extract frequent trajectories from the entire traveling paths that the user has done. The trajectories obtained from the execution of the algorithm (frequent trajectory) consider an input value that defines the support of a sequential pattern (indicating the minimum percentage of sequences where the pattern occurs, considering the total patterns). In our case we use the mining method with the data from the user that the system is going to guide. Thus, the path is adjusted to the user preferences. The execution of the algorithm searches for patterns in the dataset according to the defined threshold. The returned patterns may represent an entire trajectory, *i.e.*, may start near the user current location and end at the intended destination; or be part of a trajectory. In this case we use these trajectories and we try to create new traveling routes using the mined patterns. If the recently created trajectory is much longer than the shortest path, the system uses the shortest one, using the other option otherwise. Using the clustering method (*i.e.*, the OPTICS) the data is grouped into clusters considering some input parameters such as the minimum number of points that a core point need to have in its neighborhood and the radius ϵ that the recorded locations may be from the core of the cluster, which defines the neighborhood of a point. In a rough comparison the patterns detected by PrefixSpan are equivalent to the clusters defined by the execution of the OPTICS algorithm. However, by using this mining method it is necessary to create connections between the clusters in order to obtain a trajectory that guides the user between his current location to the intended pre-selected destination.

In the case of using the shortest path or in the case of applying the trajectory data mining method the obtained trajectory needs to be converted in order to be used in the execution of the speculative computation.

5 Conclusions

Location based acquisition methods, due to their pervasiveness, have allowed an easier access to the user location (through, for example, the smartphone) which enabled the access to new data about the user. Through this information one could know the user frequent behaviors and frequent trajectories, *e.g.*, the usual traveling path from home to work. We are developing a system to guide people with cognitive disabilities (the main users of the application) and, at the same time, providing a localization feature to caregivers. Considering the specificities of the target population of the system, it is important to develop methods that adapt the system to the user preferences and not otherwise. Thus, using a speculative computation method the system is able to predict the next movement of the user and anticipate possible mistakes. However, to use this module it is necessary to have input values which are obtained by applying a mining method. Through this paper we describe how the trajectories are adapted to the user through two different trajectory data mining methods. This process is very important since a longer but preferred (or more familiar) path is better than the shortest one, as the user may feel more comfortable traveling along the longer path and the risk of getting lost is minimized.

As future work we intend to define which of the mining methods have better results in adapting the traveling path by doing more exhaustive tests and by enlarging the spatial-temporal data about the users under test, since the preliminary results do not allow the definition of the most suitable method.

Acknowledgments

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT - Fundação para a Ciência e a Tecnologia within the Project Scope: UID/CEC/00319/2013. The work of João Ramos is supported by a FCT doctoral grant with the reference SFRH/BD/89530/2012.



References

1. Alvares, L.O., Bogorny, V., Antonio, J., Fernandes De Macedo, J.A., Moelans, B., Spaccapietra, S.: Dynamic Modeling of Trajectory Patterns using Data Mining and Reverse Engineering. ER 2007 pp. 149–154 (2007)
2. Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander, J.: Optics: Ordering points to identify the clustering structure. ACM Sigmod Record pp. 49–60 (1999), <http://dl.acm.org/citation.cfm?id=304187>

3. Carneiro, D., Novais, P., Costa, R., Gomes, P., Neves, J., Tscheligi, M., De Ruyter, B., Markopoulos, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W.: EMon: Embodied Monitorization. Proceedings of the European Conference on Ambient Intelligence 5859, 133–142 (2009), <http://www.springerlink.com/content/57w2233474k7171w/>
4. Costa, Á., Novais, P., Corchado, J.M., Neves, J.: Increased performance and better patient attendance in an hospital with the use of smart agendas. Logic Journal of IGPL 20(4), 689–698 (2011), <http://jigpal.oxfordjournals.org/content/early/2011/02/08/jigpal.jzr021.abstract>
<http://jigpal.oxfordjournals.org/cgi/doi/10.1093/jigpal/jzr021>
5. Costa, R., Neves, J., Novais, P., Machado, J., Lima, L., Alberto, C.: Intelligent Mixed Reality for the Creation of Ambient Assisted Living. Progress in Artificial Intelligence pp. 323–331 (2007), <http://www.springerlink.com/index/6t5110rv6w7q4387.pdf>
6. Fournier-Viger, P.: SPMF - open-source data mining library (2017), <http://www.philippe-fournier-viger.com/spmf/index.php>
7. Giannotti, F., Nanni, M., Pedreschi, D.: Efficient Mining of Temporally Annotated Sequences. In: Proceedings of the 2006 SIAM International Conference on Data Mining. pp. 348–359. Society for Industrial and Applied Mathematics (2006), <http://epubs.siam.org/doi/abs/10.1137/1.9781611972764.31>
8. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory Pattern Mining. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 330–339. KDD '07, ACM (2007), <http://doi.acm.org/10.1145/1281192.1281230>
9. Karich, P., Schroder, S.: GraphHopper Directions API with Route Optimization (2017), <https://graphhopper.com>
10. Lee, A.J.T., Chen, Y.A., Ip, W.C.: Mining frequent trajectory patterns in spatial-temporal databases. Information Sciences 179(13), 2218–2231 (2009)
11. Monreale, A., Pinelli, F., Trasarti, R.: WhereNext : a Location Predictor on Trajectory Pattern Mining. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09. pp. 637–645 (2009)
12. Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q., Dayal, U., Hsu, M.C.: Mining sequential patterns by pattern-growth: the PrefixSpan approach. IEEE Transactions on Knowledge and Data Engineering 16(11), 1424–1440 (nov 2004), <http://ieeexplore.ieee.org/document/1339268/>
13. Ramos, J., Costa, A., Novais, P., Neves, J.: Interactive Guiding and Localization Platform. International Journal of Artificial Intelligence (IJAI) 12(1), 63–78 (2014), <http://www.ceser.in/ceserp/index.php/ijai/article/view/2732>
14. Ramos, J., Novais, P., Satoh, K., Oliveira, T., Neves, J.: Speculative Orientation and Tracking System. International Journal of Artificial Intelligence (IJAI) 13(1), 94–119 (2015), <http://www.ceser.in/ceserp/index.php/ijai/article/view/3529>
15. Ramos, J., Oliveira, T., Satoh, K., Neves, J., Novais, P.: An Orientation Method with Prediction and Anticipation Features. Iberoamerican Journal of Artificial Intelligence 20(59), 82–95 (2017)
16. Ye, Y., Zheng, Y., Chen, Y., Feng, J., Xie, X.: Mining individual life pattern based on location history. In: Proceedings - IEEE International Conference on Mobile Data Management. pp. 1–10 (2009)
17. Zheng, Y.U.: Trajectory Data Mining : An Overview. ACM Transactions on Intelligent Systems and Technology 6(3), 1–41 (2015), <http://dl.acm.org/citation.cfm?id=2743025>