Poster Paper

On-Line Mobility Pattern Discovering using Trajectory Data

Ticiana Coelho da Silva
Federal University of Ceará, Brazil
ticianalc@ufc.br

Karine Zeitouni
Université de Versailles-St-Quentin, France
karine.zeitouni@uvsq.fr

José A. F. de Macêdo
Federal University of Ceará, Brazil
jose.macedo@lia.ufc.br

Marco A. Casanova
Department of Informatics - PUC-Rio, Brazil
casanova@inf.puc-rio.br

ABSTRACT

Mobile location tracking becomes ubiquitous in many applications, which raises great interests in trajectory data analysis and mining. Most existing work tackled the problem of offline trajectory pattern mining. Dynamic discovery and updates of patterns in trajectory data streams in (quasi) real time is a more complex task. In this paper, we propose an incremental algorithm to solve this problem, while maintaining the evolution of the patterns as well as the membership of the moving objects to their patterns.

1. INTRODUCTION

The huge volume of collected trajectories opens new opportunities for discovering the hidden patterns about mobility behaviors. These patterns may apply to characterize individual mobility as well as groups sharing similar trajectories for a certain time period. Usually, this analysis is done off-line, i.e., by applying data analysis and mining techniques on the previously collected data [1]. This allows characterizing the past movements of the objects but not the current mobility patterns. Nowadays, many services exist that involve moving objects (e.g., persons, vehicles, animals) to report their trajectory continuously (e.g., every second or every minute). Analyzing these data in real time may bring a real added-value in the comprehension of the city dynamics, and the detection of regularities as well as anomaly, which is essential for decision making. Among these patterns, we consider in this paper the trajectory group constitution and evolution, based on sub-trajectory cluster analysis. Such discovery may help the search for effective re-engineering of traffic, or dynamically detecting events or incidents, e.g., at a city level.

One important property of tracking application is the incremental nature of the data. The data will grow to reach a huge size as time goes. Finding patterns in these data in (quasi) real time is challenging. Since all the tracked moving objects change their positions over time, movement patterns also evolve in time. Furthermore, new moving objects may start sending their positions while others may stop. There exist approaches for online clustering of moving objects position, but they are restricted to instantaneous positions. Subsequently, they fail to capture the displacement behavior along time. By continuously tracking moving objects sub-trajectories at each time window, rather than just the last position, it becomes possible to gain insight on the current behavior, and potentially detect suspicious behaviors in real time. Although analyzing historical (sub)trajectory data is the majority of mobility patterns approaches today (including trajectory clustering, flocks, convoys, swarms, gathering [1]), no solution exist for clustering and maintaining clusters of sub-trajectories in real time. That is why we believe that our study is relevant.

In this paper, we address the problem of online discovery of mobility patterns and their evolution by tracking the sub-trajectories of moving objects at each time window. To solve this problem, we propose a framework discussed on the next section. Since all the objects change their sub-trajectories data from time to time, new moving objects appear as well as others disappear from the system during a time window. We define a new structure, called micro-group, to incrementally maintain the relationship among moving objects.

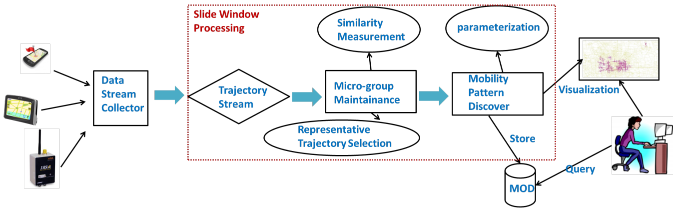

Figure 1: Steps of our proposed framework

2. FRAMEWORK

Our framework (see Figure 1) follows these main steps: (i) collect trajectory data stream at each time window, (ii) apply the similarity measure, (iii) maintain the micro-group(s), and (iv) discover the mobility patterns.

A trajectory is a sequence of the locations of a moving object at each time-stamp and is denoted by $TR_j = p_1p_2...p_r...p_{len_j}$. Here, p_k $(1 \leq k \leq len_j)$ is a point (x_k, y_k, t_k) in a three dimensional space, where (x_k, y_k) indicates the location of the object at time t_k. The length len_j of a trajectory can be different from those of other trajectories.

First step. Consider $i = [t, t + \delta t]$ be the time window observed for the set of moving objects sub-trajectory I_i. Let $I_i = \{(o_1, ST_{1,i}), (o_2, ST_{2,i}),...,(o_n, ST_{n,i})\}$, where $ST_{j,i}$ is the sub-trajectory of the moving object o_j on I_i. Therefore, I_i is the stream at the time window i.

Second step. To measure the similarity between two moving objects sub-trajectories $ST_{k,i}, ST_{j,i}$ at the time window i, we implement the synchronous Euclidean distance, which accounts for time, space, and direction. However, our framework is suitable to any distance function for trajectories. Through this paper, $distance(ST_{k,i}, ST_{j,i})$ denotes the distance function between sub-trajectories.

From a group of moving object sub-trajectories $S_i = \{(o_1, ST_{1,i}), (o_2, ST_{2,i}), ..., (o_m, ST_{m,i})\}$ at the time window i, we define the representative trajectory as a pair composed by one moving object and its sub-trajectory which is similar to the major behavior of S_i. To choose $(o_j, ST_{j,i})$ as the representative, our approach checks the number of moving objects that have their sub-trajectory similar to $ST_{j,i}$ and also uses a Gaussian kernel function (our voting function) to estimate the representativeness of $ST_{j,i}$. This voting function is also used on [2], and it has been widely used in a variety of applications of pattern recognition.

Definition 2.1. For a set of moving object sub-trajectory $S_i = \{(o_1, ST_{1,i}), (o_2, ST_{2,i}), ..., (o_m, ST_{m,i})\}$ at the time window i, ρ be a representativeness threshold, ϵ be a given distance threshold and τ be a size/density minimum threshold, $(o_j, ST_{j,i})$ is a **representative trajectory** of S_i if and only if:

1. $\forall (o_k, ST_{k,i}) \in S_i$,
   $\text{voting}(ST_{j,i}, ST_{k,i}) = e^{-\frac{distance^2(ST_{j,i}, ST_{k,i})}{2\sigma^2}} > \rho$
2. $N_\epsilon(o_j) = \{(o_k, ST_{k,i}) \in S_i | distance(ST_{j,i}, ST_{k,i}) \leq \epsilon\}, then\, |N_\epsilon(o_j)| \geq \tau$

The parameter σ shows how fast the function "voting" decreases with the distance. The intuition behind the relationship of "distance" and "voting" function is: If "distance" is close to zero, the "voting" is close to its maximum value. This means that if $ST_{j,i}$ is very close (in time, space and direction, for example) to $ST_{k,i}$, then $(o_j, ST_{j,i})$ is a candidate to be the representative. Otherwise, if the distance is high, the "voting" function is close to its minimum value, meaning that $ST_{j,i}$ is very far away from $ST_{k,i}$, so $ST_{k,i}$ is ill-represented by $ST_{j,i}$.

We define a new structure called micro-group, based on the concept of representative trajectory.

Definition 2.2. For a set of moving object sub-trajectory $S_i = \{(o_1, ST_{1,i}), (o_2, ST_{2,i}), ..., (o_m, ST_{m,i})\}$ at the time window i, let O_i be the set of moving objects in S_i, ϵ be a distance threshold, τ be a size/density minimum threshold, ρ be a representativeness threshold. A **micro-group** g is defined as a set of objects satisfying:

1. $g \subseteq O_i$
2. $\exists o_j \in g$, such that $R_g^{traj} = (o_j, ST_{j,i})$ is a representative trajectory of g w.r.t. ϵ, τ and ρ.

The ρ value can be chosen according to the maximum allowed distance between the representative sub-trajectory and the farthest member of a micro-group. The ϵ and τ thresholds captures the density around the representative trajectory, similarly to DBSCAN.

Third step. We propose an algorithm to incrementally maintain each micro-group (since all the moving objects update their sub-trajectories, some moving object may leave or join a micro-group) and to capture its evolution patterns. At the initialization phase (i.e., a cold-start of the clustering) the representative trajectories are randomly chosen among the core objects (as defined in DBSCAN). Micro-groups are first derived as the objects that vote for these representatives. The most important contribution is the maintenance phase. The intuition behind is: (i) to check for each micro-group whether the representative is still valid in the next time window (it survives), otherwise, the micro-group disappears or splits, (ii) to track moving object sub-trajectories that are likely to join the micro-group (e.g., outliers, new objects, and other objects migrating from another micro-group), (iii) for the remaining objects, a similar process to the initialization allows creating new micro-groups.

When a micro-group g_i survives, the algorithm checks for each moving object $o_k \in g_i$ if it is still well represented by $(o_j, ST_{j,i})$. If it is not, o_k is deleted from g_i and either it migrates to another micro-group, or it becomes an outlier, or it forms a new micro-group with other outliers. A micro-group g_i splits or disappears when it changes its representative trajectory. If g_i splits, new micro-groups have to be computed using the g_i data. However, if g_i is not dense enough to generate micro-group(s), its moving objects become outliers and g_i disappears.

Fourth step. We use the maintained micro-groups to discover mobility patterns, by capturing the evolution of micro-groups over time. Since each micro-group is density based, it is suitable to find sub-trajectory density based clustering (for example, merging micro-groups results in density based sub-trajectory clusters). Furthermore, this paves the way for online discovery of more complex patterns, such as flocks, convoys, leadership. Indeed, flocks could be derived from micro-groups by a light post-processing since it is a subset of the later. The convoys are also similar to the density based clusters generated by our algorithm. The representative is a close notion to leadership. Hence, this information could enrich a Moving Object Database, allowing new query and visualization types.

3. CONCLUSION

In this paper, we have present a framework to track and discovery mobility patterns in moving objects trajectory data streams. We also proposed an incremental algorithm to maintain the patterns evolution from time to time. It is noteworthy that we evaluated our approach on real data sets, which shows its effectiveness and its efficiency.

4. REFERENCES

[1] Yu Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology*, page 29, 2015.

[2] Costas Panagiotakis, Nikos Pelekis, Ioannis Kopanakis, Emmanuel Ramasso, and Yannis Theodoridis. Segmentation and sampling of moving object trajectories based on representativeness. *KDE, IEEE Transactions on*, pages 1328–1343, 2012.