Applied Mathematical Sciences, Vol. 7, 2013, no. 90, 4493 - 4497
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2013.36306

Flock Patterns and Context

Dmitry Namiot

Lomonosov Moscow State University
Faculty of Computantional Mathematics and Cybernetics
119899, Moscow, Russian Federation
dnamiot@gmail.com

Abstract

The wide deployment of location detection devices (for example, smartphones) leads to collecting of large datasets in the form of trajectories. There are a whole set of papers devoted to trajectory-based queries. Mostly, they are concentrated on similarity queries. In the same time, there is a constantly groving interest in getting various forms for aggregating behavior of trajectories as groups. The typical task, for example, is find all groups of moving objects that move together. For example, we can find convoys of vehicles, groups of people, etc. In this paper we discuss the task of flocks discovery for context-aware applications, where location data could be replaced by proximity information. We propose a framework and several strategies to discover such patterns in streaming context-related data. Our experiments with real datasets show that the proposed algorithms are scalable and efficient.

Keywords: trajectories, geo-spatial data, location, context aware.

1 Problem statement

In this paper we are dealing with so called convoys in trajectories databases. Intuitively, we can present convoy as a group of moving objects that move together, within a predefined distance to each other and for a certain continuous period of time. There are several definitions associated with this term. Convoy is a group of moving object where included objects are in density connection the consecutive time points. Objects are density-connected if a sequence of objects exists that connects the two objects and the distance between consecutive objects does not exceed the given value [1]. The next definition often used in this context is moving cluster (or cluster of moving objects) [2]. The moving cluster exists if a shared set of objects exists across some finite time, but objects may leave and join a cluster during the cluster's life time. Another acromin in this area is flock. Flock is a set of objects that travel within a range while keeping the same motion.

The classical existing methods for querying trajectories are tradionally focused on answering simple single predicate range. For example, we can mention queries like "find all moving objects that were in area A (near the shop) today about 12 p.m." [3]. Some of the reports may use similarity querying methods. The result of a similarity search query is a trajectory closest to the query trajectory according to some metric distance (e.g. Euclidean distance). In spatio-temporal joins [4] a trajectory is reported to the user if its individual behavior satisfies the query predicates.

In our paper we will extend the trajectories databases. Our approach will be based on the recorded data for the wireless context. In other words, we target indoor installations mostly. GPS data for getting location info could be unvailable, but smartphones can record Wi-Fi information (so called fingerprints) [5]. And our intention is to extract flocks from the above-mentioned context log.

2 Related works

Flock pattern queries could be based on the minimul lasting time [6] or the minimum duration as a parameter of the pattern [7]. Unlike the convoy patterns in a flock the cluster has a predefined shape. It is a disk with some predefined radius. A set of moving objects is considered a flock if there is a disk with radius *R* which covers all of them and there are at least some predefined number of objects in the disk. As per [7], the discovery of the "longest" duration flock pattern is an NP-hard problem. In interesting paper [3] authors present polynomial time solution can be found through identifying a discrete number of locations to place the center of the flock disk inside the spatial universe.

For context related data location info is replaced by fingerprints. Each fingerprint is a vector of triples $T = \{ N, M, S \}$, where *N* describes a name for Wi-Fi network, *M* – its MAC-address, *S* – signal level (RSSI). The whole network environment could be described as a vector of triples $E = \{T_1, T_2, \ldots, T_n\}$. Our fingerprint is just a time stamped environment: $[t_i, E_i]$.

The comparison between two fingerprints, f_1 and f_2, is performed as follows. Denote *M* as the union of MAC addresses in f_1 and f_2. For a MAC address $m \in M$, let $f_1(m)$ and $f_2(m)$ be the fractions computed as above. Then the similarity *S* of f_1 and f_2 is computed as follows:

$$MinMax(m) = \min(f_1(m),f_2(m))/\max(f_1(m), f_2(m))$$

$$S = \sum_m (f_1(m)+f_2(m)) * MinMax(m)$$

The intuition behind this metric is to add a large value to S when a MAC address occurs frequently in both f_1 and f_2.
The main difference introduced by the replacement location with network proximity statement is the lack of orientation. We can assume that the two objects are approximately the same distance from the access point, but they may be far enough from each other. For example, in space, they can be on different sides of the access point. So, we should use ring our objects belong to, rather than disk.
Why do we need in general this calculation for Wi-Fi proximity data? It could help us to extend existing approaches for delivering local information to mobile suscribers based on their proximity to Wi-Fi access points [9,10]. The fact of being in a group can help us to personalize the delivered information.

3 Flocks in context-related data

Suppose we have two tracks of Wi-Fi proximity data:

T_1 as {[t_{11}, E_{11}], [t_{12}, E_{12}], [t_{13}, E_{13}], …}
and
T_2 as {[t_{21}, E_{21}], [t_{22}, E_{22}], [t_{23}, E_{23}], …}

Here t_{ij} describes a time stamp and E_{ij} describes Wi-Fi environment. The similarity for two tracks means that we can map measurements from the first track to the second one. Mapping for Wi-Fi measurements means some form of the similarity – see the above-mentioned comparison between two fingerprints. And our mapping should keep the time sequence. This statement is important for trajectories. So, for example, if we map a pair [t_{11}, E_{11}] to {[t_{21}, E_{21}], then the next pair [t_{12}, E_{12}] could be mapped to the time $t \geq t_{21}$. In general, we cannot warranty, that data will be collected for the same time for the each device.

Because each application (each mobile phone) executes and collects data independently, we can not warranty that for the given timestamp t_{1i} we will find exactly the same value t_{2j} in the second track. We will try to find approximately the same timestamp $t_{1i} \pm \Delta$ where Δ is some constantly selected threshold.

Lets us present the algorithm:

Δ - time threshold, Ω - similarity threshold, T_0 – an original current time, T_{max} – an interval for our calculation, t_1 – time step ($t_1 > \Delta$)

1. Initialize a new candidate set R_1
2. Collect measurements within the time $T_0 - \Delta \rightarrow R_1$;
3. **If** R_1 is empty **then** output *false;*
4. Remove from R_1 all measurements that are not density connected in respect of similarity calculation S;
5. **If** R_1 is empty **then** output *false;*

6. **Set** $t = T_0 + t_1;$
7. **While** $t > T_0 - T_{max}$
   10. Collect the measurement within $t \pm \Delta$ (update measurements with new data);
   10. Remove from R_1 elements without new data (not updated elements) ;
   11. Remove from R_1 elements that are not not density connected in respect of similarity calculation S with new data
   12. **If** R_1 is empty **then** break;
   13. **Set** $t = T_0 + t_1;$
14**. End while**

The finally, R_1 presents the group we are looking for.

4 Results

This papers presents one of our first attempts to use well-known algorithms from geo-informatics for network proximity data. The future work with provide a new set of algorithms as well as our experimental data. The most obvious next step is to deploy signal level measurements in the similarity statement.

References

[1] H. Jeung, M. Yiu, X. Zhou, C.Jensen, and H.Shen Discovery of convoys in trajectory databases Journal Proceedings of the VLDB Endowment, Vol. 1 Num. 1, August 2008, pp. 1068-1080.

[2] P. Kalnis, N. Mamoulis, and S. Bakiras. On discovering moving clusters in spatio-temporal data, SSTD, pp. 364–381, 2005.

[3] Vieira, M. R., Bakalov, P., & Tsotras, V. J. (2009, November). On-line discovery of flock patterns in spatio-temporal data. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 286-295). ACM.

[4] S. Arumugam and C. Jermaine. Closest-point-of-approach join for moving object histories. In IEEE ICDE, pp. 86–86, 2006.

[5] D.Namiot and M.Sneps-Sneppe. Context-aware data discovery. In Intelligence in Next Generation Networks (ICIN), 2012 16th International Conference on (pp. 134-141), 2012. IEEE.
DOI: 10.1109/ICIN.2012.6376016

[6] M. Benkert, J. Gudmundsson, F. Hubner, and T. Wolle. Reporting flock patterns. Computer Geometry Theory Applications, 41(3), pp.111–125, 2008.

[7] J. Gudmundsson and M. van Kreveld. Computing longest duration flocks in trajectory data. In ACM GIS , pp. 35–42, 2006.

[8] Sneps-Sneppe, M., & Namiot, D. (2013). Spotique: A New Approach to Local Messaging. In Wired/Wireless Internet Communication (pp. 192-203). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-38401-1_15

[9] Namiot, D., & Schneps-Schneppe, M. (2011, September). About Location-aware Mobile Messages: Expert System Based on WiFi Spots. In Next Generation Mobile Applications, Services and Technologies (NGMAST), 2011 5th International Conference on (pp. 48-53). IEEE. DOI: 10.1109/NGMAST.2011.19

[10] Namiot, D. (2012, September). Context-Aware Browsing--A Practical Approach. In Next Generation Mobile Applications, Services and Technologies (NGMAST), 2012 6th International Conference on (pp. 18-23). IEEE. DOI: 10.1109/NGMAST.2012.13

Received: June 9, 2013