Applied Mathematical Sciences, Vol. 7, 2013, no. 144, 7173 - 7177
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2013.311652

Mining Relationships in Proximity Movements

Dmitry Namiot

Lomonosov Moscow State University
Faculty of Computantional Mathematics and Cybernetics
119899, Moscow, Russian Federation

Abstract

With the development of location detection devices (for example, smartphones), movement data has become widely available nowadays. Many modern applications collect large datasets in the form of trajectories. Many researchers have devoted their works to trajectory-based queries. We can point to a continued interest to similarity queries, to getting various forms for aggregating behavior, etc. In this paper we target relationships among moving objects for context-aware applications, where location data could be replaced by proximity information. This paper extends our previous works devoted to usage of proximity in geospatial applications.

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

1 Problem statement

In this paper we deal with relationships among moving objects. There are two basic points. Firstly, as a base for relationships among two objects we will use their spatiotemporal interactions only. Secondly, we will use wireless proximity information instead of traditional geo data. And objects in our models are smartphones (or other trackable wireless devices). So, we will simply associate our objects with mobile users.

As per [1], relationships between two moving objects can be classified as attraction, avoidance or neutral. In an attraction relationship, the presence of one mobile user causes the other to approach (i.e., appear or reduce the distance between them). As a result, the users should have a higher probability to be spatially close than expected based on chance. In an avoidance relationship, the presence of one user causes the other to move away (increase the distance or even leave the trackable area completely). So, the users have a lower probability to be spatially close than expected. And the definition for neutral relationship is obvious. In this case the probability that users are being spatially close is what would be expected based on independent movements.

Our proximity measurements are based on the passive Wi-Fi monitoring [2]. We can cite here our own papers where proximity information (wireless proximity, in particular) replaces geo data in trajectories databases [3][4][5]. It can be said that the above mentioned relationships are a weaker type of links compared with convoys or moving clusters (flocks).

2 Related works

The above mentioned paper [1] discusses relationships for the moved objects. Of course, there are many papers discussed trajectories for the moved object. Actually, we can treat patterns on trajectories (e.g., Flock [6]) as some form of attraction relationship. Some of the papers discuss the special form of trajectories-based patterns – following. The follower is this pattern has the similar trajectory as the leader, but arrives at a location with some (constant or varied) time lag [7].

As it is mentioned in [1], the relationship type is often not enough. The degree of the relationship is very important question too. In other words, we need to calculate the confidence in a given type of relationship. For example, two mobile users (mobile phones) may be discovered to be close each other (in proximity terms) for 5 out of 20 timestamps during the observation time. Can we make the conclusion about the relationship of attraction type? For our practical tasks (retail area) the observation time could be fixed for one day (24 hours – it is well connected to the business life cycle). Can we suggest some metric that lets make conclusions about relationships? Finally, this information will be used in some service like Geo Messages [8], where messaging service will be offered to the attractors.

The principles of replacement location data with network proximity could be found in papers [2] or [9], for example. It is analogue of Euclidian distance for network fingerprints.

3 Attraction relationships

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: $[t_i, E_i]$.

Suppose we have two trajectories (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_1 is the first mobile, T_2 is the secont phone, t_{ij} describes a timestamp and E_{ij} describes Wi-Fi environment.

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:

$$MinMax(m) = \min(f_1(m),f_2(m))/\max(f_1(m), f_2(m))$$
$$S = \sum_m (f_1(m)+f_2(m)) * MinMax(m) \text{ [9]}$$

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. Fingerprinting is based on the assumption that the Wi-Fi devices measure signal strengths in the same way. Actually, it is not so (due to differences caused by manufacturing variations, antennas, orientation, batteries, etc.). To account for this, we can use a variation of fingerprinting called ranking. Instead of comparing absolute signal strengths, this method compares lists of access points sorted by signal strength. For example, if the positioning scan discovered (SSA; SSB; SSC) = (-60; -80; -40), then we replace this set of signal strengths by their relative ranking, that is, (RA; RB; RC) = (3; 1; 2) [2]. Finally, in our system we used the following definition for similar Wi-Fi fingerptints: they have the same set of visible Access Points (the same set of MAC addresses) and they have the same relative ranking.

In general, we cannot warranty, that data will be collected for the same time slot for the each device. It is simply doe to fact that each phone records data (fingerprints) independently.

For calculating the similarity for two trajectories we can map measurements from the first track to the second one and calculate the degree of similarity. It is simply the counter for similar nodes (measurements) divided to legth of the track (total amount of nodes). So, it ranges from 0 (no similarity at all) to 1 (full similarity). Note, that this mapping should keep the natural time sequence. This statement is important for trajectories. So, for example, if we map a pair [t_{11}, E_{11}] (first track) to [t_{21}, E_{21}] (second track), then the next pair [t_{12}, E_{12}] (first track) could be mapped to the fingerprint from the second track with time $t \geq t_{21}$.

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 for the function calculates the similarity for two trajectories. Let us denote:

T_1 – a first track, T_2 – a second track, Δ - a time threshold, A_{min} – a minimal degree for similarity for declaring attraction, C_{min} – a minimal amount of measurements for declaring attraction, T_{min} – start time T_{max} – end time in our calculation

1. **Set** $A = 0$; $C = 0$;
2. Collect measurements from T_1 within the time ($T_{min} – T_{max}$) $\rightarrow$ R_1;
3. Collect measurements from T_2 within the time ($T_{min} – T_{max}$) $\rightarrow$ R_2;
4. **If** R_1 is empty **then** Output *false;*
5. **If** R_2 is empty **then** Output *false;*
6. **For each** $r \in R_1$
    10. **Set** $t = r$.time;
    11. **Set** C=C+1
    12. Collect the measurement from R_2 within the time (t, $t \pm \Delta$) $\rightarrow$ R_3;
    13. **For each** *candidate* $\in R_3$
        14. **If** isSimilar (r, *candidate*) **Then**
            15. **Set** A=A+1;
            16. Remove from R_2 records where *time* < *candidate*.time
            17. **End Loop**;
        18. **End If**
    19. **End for**
20. **End for**
21. **If** C< C_{min} **Then**
    22. Output *false*;
23. **End If**
24. **Set** level = A/C;
25. **If** level < A_{min} **Then**
    26. Output *false*;
27. **End If**
28. Output *true*;

4 Results

This paper presents the next attempt in modifications of well-known algorithms from geo-informatics to network proximity area. Our experiments with proximity data logs show satisfactory results. The most obvious next step is to implement this approach for mining Wi-Fi probe requests log.

References

[1] Li, Z., Ding, B., Wu, F., Lei, T. K. H., Kays, R., & Crofoot, M. C. (2013). Attraction and Avoidance Detection from Movements. Proceedings of the VLDB Endowment, 7(3).

[2] D. Namiot and M. Sneps-Sneppe. “Geofence and Network Proximity”, Internet of Things, Smart Spaces, and Next Generation Networking Lecture Notes in Computer Science Volume 8121, 2013, pp. 117-127, DOI: 10.1007/978-3-642-40316-3_11.

[3] D. Namiot and M. Sneps-Sneppe. “Discovery of Convoys in Network Proximity Log”, Internet of Things, Smart Spaces, and Next Generation Networking Lecture Notes in Computer Science Volume 8121, 2013, pp 139-150, DOI: 10.1007/978-3-642-40316-3_13.

[4] Namiot, Dmitry. "Flock Patterns and Context." Applied Mathematical Sciences 7.90 (2013): 4493-4497.

[5] Namiot, D., and M. Shneps-Shneppe. "Analysis of trajectories in mobile networks based on data about the network proximity." Automatic Control and Computer Sciences 47.3 (2013): 147-155.

[6] Vieira, Marcos R., and Vassilis J. Tsotras. "Flock Pattern Queries." Spatio-Temporal Databases. Springer International Publishing, 2013. 61-83.

[7] M. Andersson, J. Gudmundsson, P. Laube, and T. Wolle. Reporting leaders and followers among trajectories of moving point objects. GeoInformatica, 12(4):497–528, 2008.

[8] Namiot, D. (2010, October). Geo messages. In Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2010 International Congress on (pp. 14-19). IEEE. DOI: 10.1109/ICUMT.2010.5676665

[9] M. Azizyan, I. Constandache, and R.Roy, “SurroundSense: mobile phone localization via ambience fingerprinting”, MobiCom '09 Proceedings of the 15th annual international conference on Mobile computing and networking, pp. 261-272, DOI: 10.1145/1614320.1614350

Received: November 24, 2013