Advanced Science and Technology Letters
Vol.31 (ACN 2013), pp.87-92
http://dx.doi.org/10.14257/astl.2013.31.21

A Personalized Trajectory Similarity Evaluation Method in VANETs

Nianhua Yang[1,2], Jiming Zheng[2], Qing Liu[2], and Yuru Cao[2]

[1]Department of Computer Science and Engineering
Shanghai Jiaotong University, Shanghai 200240, China
[2]School of Business Information Management
Shanghai University of International Business and Economics, Shanghai 201620, China
yangnh@sjtu.edu.cn, zhengjiming@suibe.edu.cn, liuq@suibe.edu.cn, caoyuru2003@suibe.edu.cn

Abstract. Finding similar trajectories in massive vehicular trajectory data can benefit emerging novel mobile applications, such as carpooling, friend recommendation and traffic analysis. This paper proposes a novel spatio-temporal based trajectory similarity evaluation method. Significance of each point on the query trajectory can be assigned according to personal preference. Speed factor is also considered in the evaluation approach. Furthermore, the condition of same length in space and time for the compared trajectories is not compulsive in our method.

Keywords: trajectory similarity; network distance; VANET

1 Introduction

With the development and widespread usage of mobile location aware devices, huge geographic location data are captured every day. A trajectory is a sequence of timestamped geographic locations of a mobile user. Finding similar trajectories in massive trajectory data can benefit emerging novel mobile applications, such as carpooling, friend recommendation, traffic analysis and location based services. For example, office workers can find carpooling partners by querying trajectories which are similar with their commuter routes [8].

Much trajectory similarity evaluation approaches appeared in previous researches allow objects to move freely without any motion restrictions in 2D or 3D space [9]. However, in vehicular ad-hoc networks (VANETs), vehicles can move only on pre-defined roads. In such scenarios, Euclidean distance between two moving objects does not reflect their real distance [9]. In other words, similar trajectories measured by Euclidean distance may be dissimilar when considering

This research was partially supported by Shanghai 085 Project for Municipal Universities, the Innovation Program of Shanghai Municipal Education Commission under grant No. 14YZ134 and Shanghai special scientific research funds for selecting and cultivating excellent young teachers in colleges and universities granted in 2012.

ISSN: 2287-1233 ASTL

the network topology. The shortest path distance between two nodes in a road network is defined as network distance [9].

In most of previous studies [1], [2], [5], [3], only spatial similarity is considered in trajectory similarity evaluation. In the recent years, researchers realized that temporal factors are also important and considered in similar trajectories evaluation [9]. Driving parameters (speed, acceleration and direction) of trajectories also affect the similarity between trajectories [7]. The significance of each point on a trajectory may be different for a special mobile user who is requesting similar trajectories. So significances of sampling points on trajectories should be considered for similarity evaluation [8].

Though these before mentioned literatures consider parts of these influence factors respectively, there is no existing trajectory similarity measuring approach consider spatio-temporal factors, driving parameters and significances of sampling points simultaneously, to the best of our knowledge. This paper proposes a trajectory similarity measuring method based on network distance and temporal distance. It not only considers driving parameters but also allows the requestor to assign a significance parameter for each sample point on the querying trajectory.

The rest of this paper is organized as follows. Section 2 introduces preliminaries about road networks and trajectories. Section 3 details our proposed similarity measures. Section 4 discusses differences between our method and previous work. Finally, Section 5 concludes the work.

2 Preliminaries

2.1 Road Network

In VANETs, the mobility of objects is constrained by an underlying road network. The road network connectivity is modeled by using a graph representation, composed by vertices and edges. Each edge is assigned with a cost.

Let a connected and undirected graph $C = (V, E)$ represents a road network, where V is the set of vertices and E is the set of edges. A vertex represents a road intersection or an end of a road. An edge is defined as a connection of a pair of vertices.

Each sampling point of a trajectory should located on an edge.

2.2 Trajectory

We assume that each moving object (taxi or bus) is equipped with wireless communication devices and location aware devices. The moving object will report its geographic location and driving parameters, such as speed, at predefined intervals. This paper assumes that all trajectories have already been matched onto the edges in the corresponding road network according to some map-matching approaches [4], [6]. And the moving object always follows the shortest path connecting two points. Let r be a set of trajectories in a road network. Each trajectory $T \in r$ is defined as:

$$T = (< l_{1},v_{1},t_{1} >,\cdots,< l_{i},v_{i},t_{i} >,\cdots,< l_{m},v_{m},t_{m} >) \quad (1)$$

where m is the sampling point number of the trajectory, $l_i = (lg_i, la_i)$ represents a geographical location, lg_i and la_i denote the longitude and latitude of the point respectively, t_i is the time instance that the moving object reporting the state information, v_i represents the speed of the moving object at the moment of t_i.

3 Trajectory Similarity Measures

Due to restrictions posed by the road network, measuring trajectory proximity by means of the Euclidean distance is not appropriate [9]. This paper use the network distance as similarity metric insteading of the Euclidean distance. The smaller the network distance is, the higher the trajectory similar is. This section will follow a step-by-step construction of the similarity measure. It firstly takes into account only spatial factors with personalized significances assigned to sampling points. Then, the speed information will be added in the following step. After that, time distance will be proposed. Finally, a combined distance measuring function which considering above mentioned factors is constructed.

Let $d_a(l_{ai}, T_b)$ denotes the shortest path distance function from a sampling node l_{ai} in a trajectory T_a to another trajectory T_b. The shortest path distance between two nodes is considered as network distance. Let D_G represents the diameter of the graph C of the road network and it is globally constant for the application.

3.1 Distance Measured by Spatial Information (DMS)

Definition 1 *The network distance $d_N(T_a, T_b)$ between two trajectories T_a and T_b is defined as follows:*

$$d_N(T_a,T_b) = \frac{1}{m}\sum_{i=1}^{m} \frac{d_a(l_{ai}, T_b)}{D_G} \qquad (2)$$

where m is the sampling node numbers of trajectory T_a which is named as query trajectory.

According to the Def. 1, sampling node numbers on each trajectory could be difference.

A mobile user may have different interests to the location point on the query trajectory in applications such as location base services. So the query trajectory is a weighted data trajectory. Each sampling position l_{ai} on the query trajectory holds the weight w_{ai} which represents the significance of the point among the query processing.

Definition 2 *The network distance for two trajectories with weighted sampling points on the query trajectory is defined as follows:*

$$d_{NW}(T_a,T_b) = \frac{1}{m}\sum_{i=1}^{m} \frac{w_{ai} d_a(l_{ai}, T_b)}{D_G} \qquad (3)$$

where $\sum_{i=1}^{m} w_{ai} = 1$.

The weight w_{ai} for the point l_{ai} on the query trajectory T_a is assigned by the requestor.

3.2 DMS with Speed Information (DMSS)

The distance measure defined in the previous section takes into consideration only the network distance and personalized significance for different points on the query trajectory. In applications such as carpooling and traffic analysis, the speed information is very important.

Let S_G represents the maximum speed limitation of the discussed road network. And the limitation is globally constant for the application. Let $d_s(l_{ai}, l_{bi})$ represents the speed difference between those in the location l_{ai} and l_{bi}. l_{ai} is a sampling location on the query trajectory T_a, l_{bi} is the nearest point on the trajectory T_b from T_b to l_{ai}.

Definition 3 *The network distance for two trajectories with speed consideration and weighted sampling points on the query trajectory is defined as follows:*

$$d_{NWS} = \frac{1}{m}\sum_{i=1}^{m} \frac{w_{ai} \cdot d_a(l_{ai}, T_b) \cdot d_s(l_{ai}, l_{bi})}{D_G \cdot S_G} \tag{4}$$

3.3 Distance Measured by Temporal Information (DMT)

The similarity measures defined in the previous sections do not take into consideration the time information. In applications such as carpooling and traffic analysis, time information is important.

Let $|d_t(l_{ai}, l_{bi})|$ represents the time distance between the report time of the position l_{ai} which is on the query trajectory T_a and the report time of the position l_{bi} which is on the another trajectory T_b. l_{ai} is a sampling location on the query trajectory T_a, l_{bi} is the nearest point on the trajectory T_b from T_b to l_{ai}. l_{am} is the last sampling point of the trajectory T_a. l_{bm} is the nearest point on the trajectory T_b from T_b to l_{am}. l_{a1} is the first sampling point of the trajectory T_a. l_{b1} is the nearest point on the trajectory T_b from T_b to l_{a1}.

Definition 4 *The time distance between a query trajectory* T_a *and another trajectory* T_b *is defined as follows:*

$$d_T = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{\max\{|d_t(l_{am}, l_{b1})|, |d_t(l_{a1}, l_{bm})|, |d_t(l_{am}, l_{a1})|, |d_t(l_{bm}, l_{b1})|\}}{|d_t(l_{ai}, l_{bi})|} \tag{5}$$

3.4 Combined Distance Measure

Now, we have different distance measures that can be used to query similar trajectories from different metrics for different length in time, space and speed. Several applications may require some combined measures for similarity querying.

Definition 5 *The combined spatio-temporal distance measure considering different significance of sampling points can be expressed as follows:*

$$dNWT = wNW \cdot d_{NW}(Ta, Tb) + wT \cdot dT \tag{6}$$

where w_{NW} and wT are weight parameters for corresponding sub-measures. In addition, w_{NW} + wT = 1.

Definition 6 *The combined spatio-temporal distance measure considering speed and different significance of sampling points can be expressed as follows:*

$$dC = wNWS \cdot d_{NWS}(Ta, Tb) + wT \cdot dT \tag{7}$$

where w_{NWS} and wT are weight parameters for corresponding sub-measures. In addition, w_{NWS} + wT = 1.

4 Discussion

This section details differences between previous studies and our work.

The study in [9] requires that two trajectories should contain the same number of sampling points. And they are compared according to the sequence order. So the result will be influenced by distribution of sampling points. Our method is not constrained by these conditions.

The study in [9] uses the minimum network distance between two sampling points on the two trajectories respectively to compute similarity. It does not distinguish the query trajectory and the object trajectory. But our method uses the minimum network distance from a sampling point on the query trajectory to the object trajectory. So the number or sampling points on the object trajectory can be arbitrary.

The idea of assign personalized significances for different sampling points on the query trajectory for similarity computing comes from [8]. The speed information has been considered in similarity measuring in *[7]*. This paper integrates these factories into the spatio-temporal similarity evaluation method.

5 Conclusion

Although there are significant contributions achieved on trajectory similarity evaluation, the vast majority of the proposed approaches assume that the compared two trajectories hold the same number of sampling points. Most of previous approaches compute distance through comparing the distance between two sampling points on the different trajectories following the same sequence order. This paper defines several similarity distance measures through adding influence factors step-by-step. It firstly proposes a distance measure based on the network distance between the sampling point on the query trajectory and the object trajectory. Then the significance factor on different sampling point is added into the measuring algorithm. After that, speed information is considered in the measuring method. Time distance between two trajectories is measured independently. Last but not least, time distance and other distance evaluation formula are combined with different weights.

References

1. Chen, L., Ng, R.: On the marriage of lp-norms and edit distance. In: Proceedings of the Thirtieth international conference on Very large data bases - Volume 30. pp. 792–803. VLDB ’04, VLDB Endowment (2004)
2. Chen, L., ¨Ozsu, M.T., Oria, V.: Robust and fast similarity search for moving object trajectories. In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data. pp. 491–502. SIGMOD ’05, ACM, New York, NY, USA (2005)
3. Frentzos, E., Gratsias, K., Pelekis, N., Theodoridis, Y.: Algorithms for nearest neighbor search on moving object trajectories. Geoinformatica 11(2), 159–193 (Jun 2007)
4. Huang, H., Zhang, D., Zhu, Y., Li, M., Wu, M.Y.: A metropolitan taxi mobility model from real GPS traces. Journal of Universal Computer Science 18(9), 1072– 1092 (May 2012)
5. Lin, B., Su, J.: Shapes based trajectory queries for moving objects. In: Proceedings of the 13th annual ACM international workshop on Geographic information systems. pp. 21–30. GIS ’05, ACM, New York, NY, USA (2005)
6. Liu, K., Li, Y., He, F., Xu, J., Ding, Z.: Effective map-matching on the most simplified road network. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. pp. 609–612. SIGSPATIAL ’12, ACM, New York, NY, USA (2012)
7. Pelekis, N., Andrienko, G., Andrienko, N., Kopanakis, I., Marketos, G., Theodoridis, Y.: Visually exploring movement data via similarity-based analysis. J. Intell. Inf. Syst. 38(2), 343–391 (Apr 2012)
8. Shang, S., Ding, R., Zheng, K., Jensen, C., Kalnis, P., Zhou, X.: Personalized trajectory matching in spatial networks. The VLDB Journal pp. 1–20 (2013), http://dx.doi.org/10.1007/s00778-013-0331-0
9. Tiakas, E., Papadopoulos, A.N., Nanopoulos, A., Manolopoulos, Y., Stojanovic, D., Djordjevic-Kajan, S.: Searching for similar trajectories in spatial networks. J. Syst. Softw. 82(5), 772–788 (May 2009)

