

## A Novel Trajectory Similarity Evaluation Method in VANETs

Nianhua Yang<sup>1,2</sup>, Jiming Zheng<sup>2</sup>, Qing Liu<sup>2</sup> and Yuru Cao<sup>2</sup>

<sup>1</sup>*Department of Computer Science & Engineering, Shanghai Jiaotong University,  
Shanghai 200240, China*

<sup>2</sup>*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**

*With the increasing number of vehicles equipped with GPS enabled wireless communication devices, locations and trajectories of vehicles can be collected every time. Finding similar trajectories in massive 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. In this method, the 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, most of existing trajectory similarity evaluation methods hold the precondition that the compared trajectories should have the same space and time length. This is not a compulsive condition in our method.*

**Keywords:** *trajectory similarity, network distance, VANET*

### **1. Introduction**

With the development and widespread usage of mobile location aware devices, huge geographic GPS 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, advertisement pushing and location based services. For example, office workers can find carpooling partners by querying trajectories which are similar with their commuter routes [1]. More and more approaches have appeared for evaluating trajectory similarity. However, some problems in trajectory similarity evaluation still need to be resolved.

On the one hand, most of approaches appeared in previous researches allow objects to move freely without any motion restrictions in 2D or 3D space [2] and using Euclidean distance between two moving objects as similarity measuring. 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 [2]. In other words, similar trajectories measured by Euclidean distance may be dissimilar when considering the network topology.

On the other hand, in most of previous studies [3-6], only spatial similarity is considered in trajectory similarity evaluation. In the recent years, researchers realized that temporal factors are also important and should be considered in similar trajectories evaluation [2]. 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 [1].

Though these previous mentioned literatures have considered 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 framework for trajectory similarity measuring based on spatio-temporal factors, driving parameters and significances of sampling points. It is an extension of our previous work [8] which is included in the proceedings of ACN 2013.

The rest of this paper is organized as follows. In the next section, we present related work. Section 3 introduces preliminaries about road networks and trajectories. Section 4 details our proposed similarity measures. Section 5 discusses the difference between our method and previous work. Finally, Section 6 concludes the work and briefly describes our future work.

## 2. Related work

The problem of trajectory similarity evaluation has been studied extensively in the last years. Several types of trajectory similarity measuring functions have been proposed based on different distance metrics, such as Euclidean distance [9], network distance [2], Edit distance [10] *et al.*

In VANETs, the mobility of vehicles is constrained by road networks. In other words, the position of a vehicle must satisfy the road networks constrains. The road network connectivity can be modeled by using a weighted graph. Several efforts [2] have been performed towards efficient spatial and spatio-temporal query processing evaluation in constrained spatial networks. Jensen *et al.* [11] propose nearest neighbor queries in road networks. The road network is modeled by a graph. Chen *et al.* [12] address the problem of monitoring the  $k$  nearest neighbors to a dynamically changing path in road networks. They propose a three-phase Best-first Network Expansion (BNE) algorithm for monitoring the  $k$ -PNN and the corresponding shortest path in road networks. Zheng *et al.* [13] focus on representing the uncertainty of the objects moving along road networks as time-dependent probability distribution functions. And they propose efficient algorithms for processing spatio-temporal range queries. Shang *et al.* [14] propose and investigate a novel spatial query called Reverse Path Nearest Neighbor (R-PNN) search to find the most accessible locations in road networks. The R-PNN query can be used in many important applications such as urban planning, facility allocation, traffic monitoring, etc. In-route nearest neighbor queries are also studied by Yoo and Shekhar [15].

These previous contributions concentrate on spatial or spatio-temporal query processing for finding nearest-neighbors among a specified range. However the issue of trajectory similarity has not been studied. The similarity between trajectories is defined as the Euclidean distance between directed discrete lines [16]. Laurinen *et al.* [17] propose an efficient algorithm for trajectory similarity calculation based on Euclidean metric and spaces. Buchin *et al.* [18] propose an approach to finding similar subtrajectories using a distance measure that is defined as the average Euclidean distance at corresponding times. These contributions only use Euclidean distance as similarity metric.

Hwang *et al.* [19] propose a trajectory similarity evaluation approach based not on Euclidean distance but on road network distance. The proposed filtering method is based on

spatial similarity and the refining method is based on temporal distance. In that paper, the number 1 and 0 are used to represent similar and dissimilar respectively. The similarity with such definition does not take into consideration any notion of similarity range [2]. Therefore, how similarity between two trajectories can't be determined. In order to solve this problem, Tiakas *et al.* [2] propose a similarity model in spatial networks. The network distance between two trajectories is used for the definition of similarity. So the level of similarity can be determined. The above mentioned studies have assigned equal significance to each sample point on trajectories. Shang *et al.* [1] takes into account different significance of each sample point on the query trajectory.

Pelekis *et al.* [7] argue that driving parameters (speed, acceleration and direction) also affect the similarity between trajectories. Parent *et al.* [20] argue that the data captured from the device, such as the acceleration and direction, are complement elements for a trajectory and may inference the trajectory similarity. Most of previous studies ignore these elements when computing trajectory similarity.

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.

### 3. Preliminaries

#### 3.1. Road Network

In VANETs, the mobility of objects is constrained by an underlying road network, such as shown in Figure 1. The road network connectivity is modeled by using a graph representation, such as shown in Figure 2, composed by vertices and edges. Each edge is assigned with a cost which represents the distance between the two points.

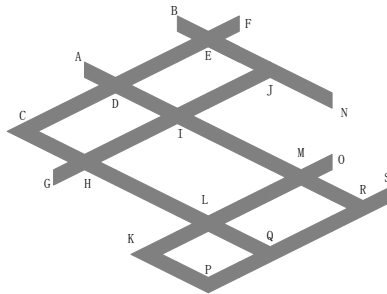


Figure 1. A Road Network Example

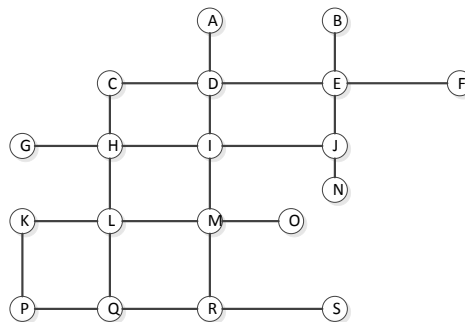


Figure 2. A Road Network Example

Let a connected and undirected graph  $G=(V,E)$  represents a road network, where  $V$  is the set of vertices and  $E$  is the set of edges. A vertex indicates a road intersection or an end of a road. An edge is defined as a connection of a pair of vertices. The cost of each edge represents the road distance between the two vertexes of the edge.

### 3.2. 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 [21, 22]. And the moving object always follows the shortest path connecting two points.

Let  $\Gamma$  be a set of trajectories in a road network. Each trajectory  $T \in \Gamma$  is defined as:

$$T = (\langle l_1, v_1, t_1 \rangle, \dots, \langle l_i, v_i, t_i \rangle, \dots, \langle l_m, v_m, t_m \rangle) \quad (1)$$

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  $t_i$ .

## 4. Trajectory Similarity Measures

Due to restrictions posed by the road network, measuring trajectory proximity by means of the Euclidean distance is not appropriate [2]. This paper use the network distance as similarity metric instead 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  $G$  of the road network and it is globally constant for the application.

### 4.1. 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:

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

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:

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

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.

#### 4.2. DMS with Speed Information (DMSS)

The distance measure defined in the previous section take 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 that 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:

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

#### 4.3. 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:

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

#### 4.4. 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:

$$d_{NWT} = w_{NW} \square d_{NW}(T_a, T_b) + w_T \square d_T \quad (6)$$

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

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

$$d_C = w_{NWS} \square d_{NWS}(T_a, T_b) + w_T \square d_T \quad (7)$$

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

### 5. Discussion

This section details differences between previous studies and our work.

The study in [2] 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 [2] uses the minimum 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 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 significance for different sampling points on the query trajectory for similarity computing comes from [1]. The speed information has been considered in similarity measuring in [7]. This paper integrates these factories into the spatio-temporal similarity evaluation method.

### 6. Conclusion and Future Work

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 trajectory 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.

In the future, we will develop a query processing algorithm based on the proposed similarity evaluation approach. And we will utilize the proposed approach for data mining among the collected traffic GPS data.

## Acknowledgements

This research was partially supported by Shanghai 085 Project for Municipal Universities, the Research 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.

## References

- [1] S. Shang, R. Ding, K. Zheng, C. Jensen, P. Kalnis and X. Zhou, "Personalized trajectory matching in spatial networks", *The VLDB Journal*, (2013), pp. 1-20.
- [2] E. Tiakas, A. N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic and S. Djordjevic-Kajan, "Searching for similar trajectories in spatial networks", *Journal of Systems and Software*, vol. 82, no. 5, (2009), pp. 772-788.
- [3] L. Chen and R. Ng, "On the marriage of Lp-norms and edit distance", *Proceedings of the Thirtieth international conference on Very large data bases*, (2004); Toronto, Canada: VLDB Endowment, pp. 792-803.
- [4] L. Chen, M. T. Özsu and V. Oria, "Robust and fast similarity search for moving object trajectories", *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, (2005); Baltimore, Maryland: ACM, pp. 491-502.
- [5] B. Lin and J. Su, "Shapes based trajectory queries for moving objects", *Proceedings of the 13th annual ACM international workshop on Geographic information systems*, (2005), Bremen, Germany: ACM, pp. 21-30.
- [6] E. Frenzos, K. Gratsias, N. Pelekis and Y. Theodoridis, "Algorithms for Nearest Neighbor Search on Moving Object Trajectories", *Geoinformatica*, vol. 11, no. 2, (2007), pp. 159-193.
- [7] N. Pelekis, G. Andrienko, N. Andrienko, I. Kopanakis, G. Marketos and Y. Theodoridis, "Visually exploring movement data via similarity-based analysis", *Journal of Intelligent Information Systems*, vol. 38, no. 2, (2012), pp. 343-391.
- [8] N. yang, J. Zheng, Q. Liu and Y. Cao, "A Personalized Trajectory Similarity Evaluation Method in VANETs", *Proceedings of The 5th International Conference on Advanced Communication and Networking (ACN 2013)*, Science and Engineering Research Support Society (SERSC).
- [9] R. Agrawal, C. Faloutsos and A. N. Swami, "Efficient Similarity Search In Sequence Databases", *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, (1993), London, UK, Springer-Verlag, pp. 69-84.
- [10] Z. Chen, H. T. Shen, X. Zhou, Y. Zheng and X. Xie, "Searching trajectories by locations: an efficiency study", *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, (2010); Indianapolis, Indiana, USA: ACM, pp. 255-266.
- [11] C. S. Jensen, J. Kolářvr, T. B. Pedersen and I. Timko, "Nearest neighbor queries in road networks", *Proceedings of the 11th ACM international symposium on Advances in geographic information systems*, (2003); New Orleans, Louisiana, USA: ACM, pp. 1-8.
- [12] Z. Chen, H. T. Shen, X. Zhou and J. X. Yu, "Monitoring path nearest neighbor in road networks", *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, (2009); Providence, Rhode Island, USA: ACM, pp. 591-602.
- [13] K. Zheng, G. Trajcevski, X. Zhou and P. Scheuermann, "Probabilistic range queries for uncertain trajectories on road networks", *Proceedings of the 14th International Conference on Extending Database Technology*, (2011), Uppsala, Sweden: ACM, pp. 283-294.
- [14] S. Shang, B. Yuan, K. Deng, K. Xie and X. Zhou, "Finding the most accessible locations: reverse path nearest neighbor query in road networks", *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, (2011); Chicago, Illinois: ACM, pp. 181-190.
- [15] J. S. Yoo and S. Shekhar, "In-Route Nearest Neighbor Queries", *Geoinformatica*, vol. 9, no. 2, (2005), pp. 117-137.
- [16] Y. Yanagisawa, J.-i. Akahani and T. Satoh, "Shape-Based Similarity Query for Trajectory of Mobile Objects", *Proceedings of the 4th International Conference on Mobile Data Management*, (2003); Springer-Verlag, pp. 63-77.

- [17] P. Laurinen, P. Siirtola and J. Röning, "Efficient algorithm for calculating similarity between trajectories containing an increasing dimension. Proceedings of the 24th IASTED international conference on Artificial intelligence and applications, (2006); Innsbruck, Austria: ACTA Press, pp. 392-399.
- [18] K. Buchin, M. Buchin, M.v. Kreveld and J. Luo, "Finding long and similar parts of trajectories", *Computational Geometry*, vol. 44, no. 9, (2011), pp. 465-476.
- [19] J.-R. Hwang, H.-Y. Kang and K.-J. Li, "Searching for similar trajectories on road networks using spatio-temporal similarity", *Proceedings of the 10th East European conference on Advances in Databases and Information Systems*, (2006); Thessaloniki, Greece: Springer-Verlag, pp. 282-295.
- [20] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis and Z. Yan, "Semantic trajectories modeling and analysis", *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, (2013), pp. 1-32.
- [21] H. Huang, D. Zhang, Y. Zhu, M. Li and M.-Y. Wu, "A Metropolitan Taxi Mobility Model from Real GPS Traces. *Journal of Universal Computer Science*, vol. 18, no. 9, (2012), pp. 1072-1092.
- [22] K. Liu, Y. Li, F. He, J. Xu and Z. Ding, "Effective map-matching on the most simplified road network", *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, (2012); Redondo Beach, California: ACM, pp. 609-612.

## Authors



**Nianhua Yang**, he received his BSc in Management Information Systems from Beijing Information Technology Institute, China, in 2001, MSc and PhD in Computer Application Technology from East China University of Science and Technology in 2004 and 2011 respectively. He is now a postdoctor in department of Computer Science & Engineering at Shanghai Jiaotong University. His current research interests include wireless ad hoc networks, location privacy preserving, reputation management, VANETs et al.



**Jiming Zheng**, he received his BSc in Foundry from Shanghai University of Engineering Science, China, in 1991, MSc in Management Engineering from Shanghai University, China, in 1998. He is now an associate professor at Shanghai University of International Business and Economics. His current research interests include data management and data mining.



**Qing Liu**, she received his BSc and MSc in Computer Application Technology from Central China Normal University, in 1995 and 2001 respectively. She is now an associate professor at Shanghai University of International Business and Economics. Her current research interests include data management, Software engineering, Petri nets, *et al.*





**Yuru Cao**, she received his BSc in Applied Mathematics from Anqing Normal University, China, in 2000, MSc in Applied Mathematics from Anhui Normal University, China, in 2003 and PhD in Computer Application Technology from Anhui University, China, in 2006. She is now an associate professor at Shanghai University of International Business and Economics. Her current research interests include data management, privacy preserving, *et al.*

