Spatio-Temporal Similarity Analysis between Trajectories on Road Networks

Jung-Rae Hwang[1], Hye-Young Kang[2], and Ki-Joune Li[2]

[1] Department of Geographic Information Systems, Pusan National University, Korea
[2] Department of Computer Science, Pusan National University, Korea
{jrhwang,hykang}@isel.cs.pusan.ac.kr, lik@pnu.edu

Abstract. In order to analyze the behavior of moving objects, a measure for determining the similarity of trajectories needs to be defined. Although research has been conducted that retrieved similar trajectories of moving objects in Euclidean space, very little research has been conducted on moving objects in the space defined by road networks. In terms of real applications, most moving objects are located in road network space rather than in Euclidean space. In this paper, we investigate the properties of similar trajectories in road network space. And we propose a method to retrieve similar trajectories based on this observation and similarity measure between trajectories on road network space. Experimental results show that this method provides not only a practical method for searching for similar trajectories but also a clustering method for trajectories.

Keywords: Trajectories, Road Network Space, Similarity between Trajectories

1 Introduction

With the spread of mobile computing, research to efficiently handle moving objects, where their movement is represented by a trajectory as a set of line segments in (x, y, t) space, has become important [1]. Since the trajectory of a moving object contains a lot of information, it is an interesting task to analyze trajectories for several application areas. One of the most important requirements for analyzing trajectories is to search for objects with similar trajectories and cluster them. For example, a query such as "Find all moving objects whose trajectories are similar to a given query trajectory" is typical.

While research has been done regarding locating similar trajectories of moving objects on Euclidean space, very little has been done regarding to moving objects in road network space. For most of real applications, we are interested in moving objects in road network space rather than in Euclidean space. In order to analyze the behavior of moving objects in road network space, a measure for determining the similarity between the trajectories of moving objects needs to be defined. This measurement of similarity allows for the retrieval of similar trajectories and the eventual discovery of their patterns and clusters.

Due to the properties of road network space, the methods that can be useful to search for similar trajectories differ from the methods currently in use [2][3]. The current methods have the following drawbacks. First, they assume Euclidean space, and Euclidean distance is no longer valid in road network space, where the distance is limited to the space adjacent to the roads. Since measuring similar trajectories is highly dependent on the definition of distance, the similarity measurements as defined for Euclidean space are inappropriate for road network space, and consequently the methods based on Euclidean space are not suitable for our purpose.

Second, the previously used methods do not fully exploit the spatiotemporal properties of trajectories and most of them only consider spatial similarities. For example, two trajectories passing through the same area at different times are considered similar, even though they are not similar in spatiotemporal sense. In addition two trajectories that are moving in opposite directions are considered to have similar trajectories according to those previous methods.

Our research is motivated by two requirements. First, our method should be based on the characteristics of moving objects on road network space. Second, we should simultaneously consider a spatiotemporal similarity as well as spatial similarity. Based on these ideas, we propose a search method for similar trajectories of moving objects on road networks. Our method is based on spatiotemporal properties and reflects spatial characteristics on road networks.

This paper is organized as follows. In section 2, we introduce the related work and drawbacks of the previously used methods and investigate the characteristics of similarity for the trajectory of moving objects on road networks. In section 3 we propose a method for searching for similar trajectories on road networks. Experimental results are given in section 4. Finally, we conclude and suggest future work in section 5.

2 Related Work and Motivation

In this section, we introduce related researches on moving objects on road networks. Also, we discuss the problems of existing methods by investigating the trajectory characteristics of moving objects on road networks and present the motivations of this paper.

2.1 Related Work

Searching for similar trajectories of moving objects is closely connected to two research issues: 1) representing the trajectory of moving objects and 2) defining measurements of similarity.

Concerning the first research issue, many studies have investigated ways that the trajectories of moving objects can be represented [4][5]. In particular, representation models for trajectories have been proposed based on Markovian and non-Markovian probability models in [6], which are effective in extracting useful information from trajectories. Another interesting model has been proposed in

[7] that considers the geospatial lifelines of multiple granularities. These methods deal with moving objects on Euclidean space. However, most moving objects in real applications, such as vehicles or trains, are found in road network space rather than in Euclidean space. There has been some research regarding representing and handling the movement of objects in road network space [5][8][9]. A model for representing and querying moving objects on road networks is clearly presented in [5]. The representation of moving objects along a road network was also presented in [9]. A nearest neighbor search method for moving objects on road networks was introduced in [8].

Regarding the second research issue, the most important studies of search methods for similar trajectories are found in [2] and [10]. A method for finding the most similar trajectory of a given query trajectory within a database using the longest common subsequence model was proposed in [10]. However, this method has two problems when used to search for similar trajectories of moving objects on road networks. First, this method does not take temporal or spatiotemporal variables into consideration. For example, two trajectories passing through the same area at different times are considered similar. Second, since this method is based on Euclidean space, it cannot be used to search for similar trajectories on road networks as discussed in the previous section.

A method for measuring the similarity between trajectories based on shape was defined in [3]. The advantage of this definition is that spatiotemporal aspects are taken into account, unlike [2]. However, since this method assumes Euclidean space, it is difficult to apply it to road network space. A similar method was proposed in [11] but has the same problem of Euclidean distance as [2].

2.2 Similarity of moving object trajectory on road networks

Most moving objects are in road network space rather than in Euclidean space. There are several differences between Euclidean space and road network space. First, figure 1 illustrates the different definitions of distance in Euclidean and road network space. In figure 1, the actual distance from a to b is not 4 km but 9 km. Second, different coordinate systems are employed for road network space. While the (x, y, t) coordinate system is the most popular one in Euclidean space, (Sid, d, t) is more efficient in road network space, where Sid is a road sector identifier, and d is the offset from the starting point of the road sector. Queries are given by specifying the road sector ID rather than an area in Euclidean space. It is easier to calculate distance between two points on road networks by using road network coordinate systems than Euclidean coordinate systems. Finally, road network space requires additional data to describe the connectivity between road sectors. These differences should be carefully examined and considered when analyzing trajectories in road network space.

Let's investigate trajectory properties on road networks. Figure 2 shows an example of trajectories in (x, y, t) space, where t represents the time-axis and (x, y) space, which is the projected space. TR_A, TR_B and TR_C in figure 2 are trajectories in (x, y, t) space, while TR'_A, TR'_B and TR'_C are projected trajectories onto the (x, y) plane. TR_A and TR_B pass the exact same point in the same

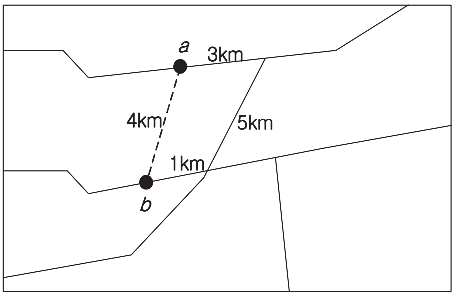


Fig. 1. Example of distances in road network space and Euclidean space

order but at different time intervals. On other side, TR_A and TR_C are a short distance apart from each other, but move at very similar time intervals. When we project these trajectories on the (x, y) plane, TR_A and TR_B are projected on an equal trajectory, while the projected trajectories of TR_A and TR_C are placed at different locations. This shows that if the spatiotemporal variable is considered, then TR_C would have the most similar trajectory to TR_A. This outcome differs from the similarity measures proposed by [2] and [11], which consider only the similarity between projected trajectories.

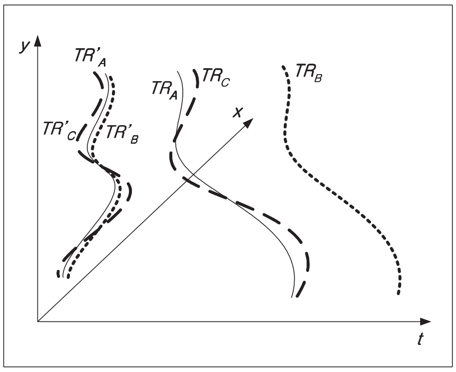
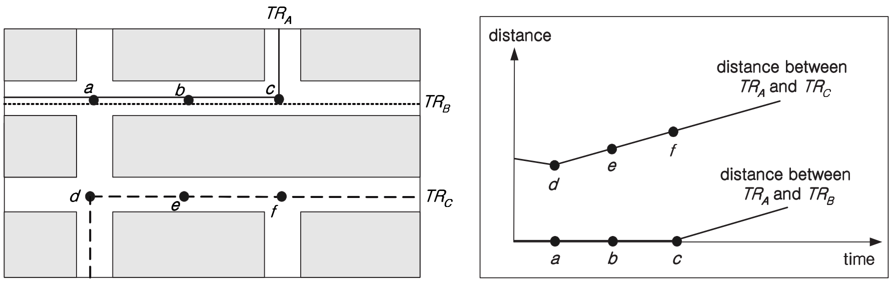


Fig. 2. Trajectories in spatiotemporal space and their projected trajectories

The distance between moving objects in road network space has an interesting property. Suppose that two moving object trajectories, TR_A and TR_B pass through the same points a, b and c on a road at the same time, and they take different road as depicted by figure 3(a). Then, the distance between them rapidly increases after point c according to the road network as shown by figure 3(b). In most cases, two moving objects on two different road sectors result in a relatively large distance between them and exceed the distance threshold used in similarity searches as shown by figure 3(b). This observation implies that it



(a) (b)

Fig. 3. Change of distance between two moving objects on road networks

may be meaningless to compute distance between two moving objects if they are on different road sectors.

3 Searching for Similar Trajectories on Road Networks

In order to retrieve similar trajectories on road networks, we could apply one of the following methods;

- Method 1 : Searching for similar trajectories based on spatiotemporal distance between trajectories.
- Method 2 : Filtering trajectories based on temporal similarity and refining similar trajectories based on spatial distance.
- Method 3 : Filtering trajectories based on spatial similarity and refining similar trajectories based on temporal distance.

We now discuss each of these methods. Method 1 looks simple and therefore attractive. In order to apply this method, we need a robust definition for measuring spatiotemporal distance, which might be the sum of spatial distance and temporal distance. However, it is impossible to define the equivalence between temporal distance and spatial distance. For example, how can the spatial distance equivalent to one minute be defined? We can define the equivalence in a specific situation by considering a parameter such that, for example, dist(1 meter) = dist(α seconds). In most cases, however, the general equivalence cannot be easily defined, and it depends on the context of the application. We leave this issue for further study.

Method 2 requires a definition for temporal similarity or distance in order to filter trajectories. For example, suppose that $[t_s(TR_A), t_e(TR_A)]$, $[t_s(TR_B), t_e(TR_B)]$ are the life spans of the two trajectories, TR_A and TR_B. In practical application, the meaning of distance between two time intervals can hardly be found. It means that the second method is not appropriate for searching for similar trajectories.

Consequently, we propose the third method for searching similar trajectories. For this method, we need a definition of spatial distance between trajectories, which are represented as curves. A widely used definition of distance between curves l and m is Hausdorff distance $dist_H(l, m)$, which is defined as:

$$dist_H(l, m) = \max_{a \in l}\{\min_{b \in m} dist(a, b)\},$$

where $dist(a, b)$ is the distance between two points.

An interesting property of Hausdorff distance is found in the trajectories on road networks. In figure 4(a), the Housdorff distance between TR_A and TR_B $dist_H(TR_A, TR_B)$ is d, determined by the pair of points p and q. And we see that $dist_H(TR_C, TR_D) = d$. The distance between TR_A and TR_B is equal with that between TR_C and TR_D depending on the type of application. It means that we cannot use Hausdorff distance to measure spatial distance.

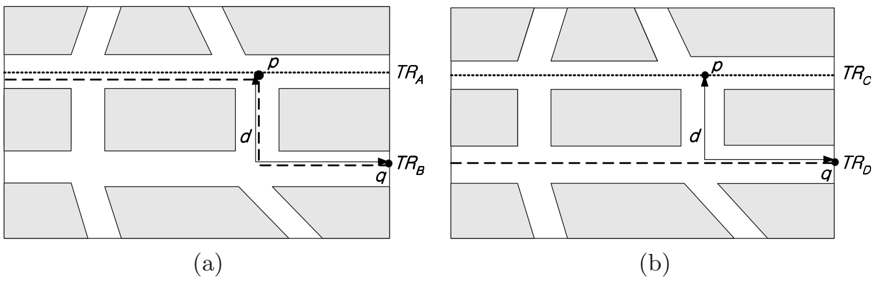


Fig. 4. Hausdorff distance on road networks

Instead of Hausdorff distance, we propose a practical method to determine the spatial similarity between trajectories based on POI(Point of Interest). For example, important intersections of roads or places can be POIs. If two trajectories pass through the same POIs, they are considered similar by the following definition.

Definition 1. *Spatial Similarity between Trajectories on road network space Suppose that P is a set of POIs on a given road networks. Then spatial similarity between two trajectories TR_A and TR_B is defined as*

$$Sim_{POI}(TR_A, TR_B, P) = \begin{cases} 1, & if\ \forall p \in P,\ p\ is\ on\ TR_A\ and\ TR_B \\ 0, & otherwise \end{cases}$$

In order to apply Method 3 for searching for similar trajectories, in addition to a spatial similarity, we also need a measure for temporal similarity. Temporal similarity can be defined as the inverse of temporal distance. In contrast with the discussion of temporal distance in Method 2, temporal distance can be defined, when a POI is given, as the difference between the times two objects passed the same POI as follows:

Definition 2. *Temporal Distance between Trajectories for one POI*
Suppose that $p \in P$*, and* P *is the set of POI. Then the temporal distance between two trajectories* TR_A *and* TR_B *is*

$$dist_T(TR_A, TR_B, p) = |t(TR_A, p) - t(TR_B, p)|$$

If neither TR_A *nor* TR_B *pass through p, the temporal distance is considered as infinity.*

If we consider $t(TR, p_i)$ as the time the i-th POI, was passed each trajectory, TR, is plotted as a point $t(TR) = (t(TR, p_1), t(TR, p_2), ..., t(TR, p_k))$ in a k-dimensional space where k is the number of POIs. Then the temporal distance between two trajectories for a set of POIs is defined as the L_P distance of this k-dimensional space as follows:

Definition 3. *Temporal Distance between Trajectories for a set of POIs*
Suppose that P *is a set of POI and* TR_A *and* TR_B *are two trajectories. Then the temporal distance between* TR_A *and* TR_B *is*

$$dist_T(TR_A, TR_B, P) = L_p(TR_A, TR_B, p) = (\sum_{i=1}^{k} |(p_i(TR_A) - p_i(TR_B)|^p)^{\frac{1}{p}}$$

Algorithm 1. Searching Similar Trajectories

Input. input trajectories TR_{IN}, threshold δ, query trajectory tr_Q, POI set P
Output. similar trajectories TR_{OUT}
Begin
$TR_{Candidate} \leftarrow \phi$
$TR_{OUT} \leftarrow \phi$
For each $tr \in TR_{IN}$
If $\forall p \in P$, p is on tr
then $TR_{Candidate} \leftarrow TR_{Candidate} \cup \{tr\}$
For each $tr \in TR_{Candidate}$
If $dist_T(tr_Q, tr, P) < \delta$
then $TR_{OUT} \leftarrow TR_{OUT} \cup \{tr\}$
return TR_{OUT}
End

Algorithm 1 summaries the search procedure explained in this section. It consists of two steps, the filtering step based on spatial similarity step and the refinement step for searching for similar trajectories based on temporal distance.

Note that we use L_2 distance for the reason of simplicity in this paper, but other types of distance can be employed. Since each trajectory is represented as a point in a multi-dimensional space, we can apply a clustering method.

We call this multi-dimensional space *Temporal Trajectory Space*. A number of methods have been proposed for clustering points in this temporal trajectory space [12][13][14].

4 Experimental Results

In order to examine the feasibility of our method, we performed experiments with real trajectory data set gathered from taxis in Seoul. We clustered the trajectories based on the search method proposed in the preceding section. For the first step, we found all of the trajectories that passed through four given POIs. For the second step we clustered the points in the temporal trajectory space by means of a shifted Hilbert curve [14].

Figure 6 shows the results of the first step, which are the trajectories passing through the given POIs, where the x-axis is the offsets of the trajectories and the y-axis is the time they passed the POIs. The descending curves represent the trajectories with opposite directions. The clusters marked on the right were obtained by the second step. We see that this result corresponds with our intuitive clustering.

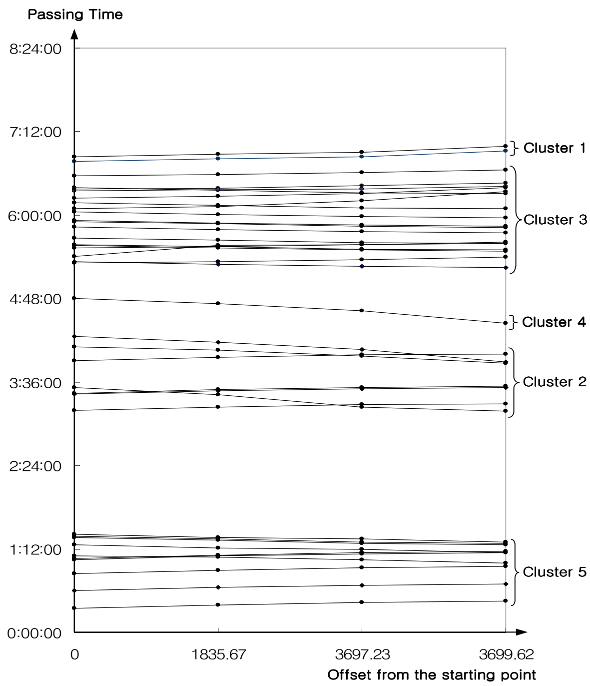


Fig. 5. Result of the first step of clustering

These trajectories were then clustered by means of temporal distance as shown in figure 7, where only the starting and ending POIs are depicted as the x-axis and y-axis, respectively, for the reason of simplicity. Note that the x-axis and y-axis are normalized to [0,1]. An interesting fact is observed that most points in the temporal trajectory space were found around a linear graph. In fact, the slope of this line represents the speed of the moving objects, which were similar in our experiment.

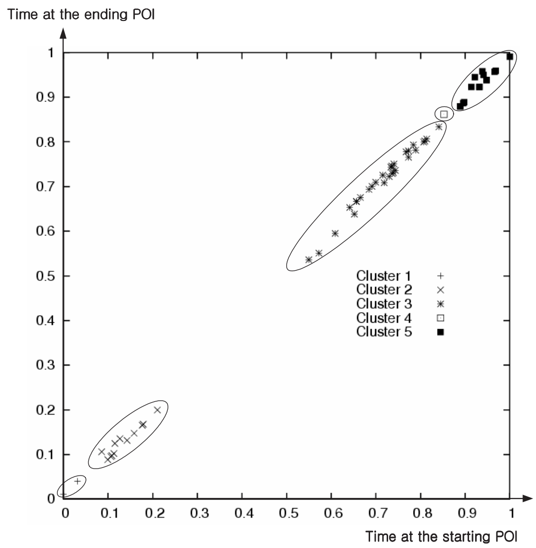


Fig. 6. Clustering of trajectories by temporal distance

5 Conclusion and Future work

Analysis of the similarity between trajectories on road networks has many potential applications. For example, it is helpful to analyze the trajectories of cars used for commercial purposes on a road for making marketing strategies. In this paper, we present the important properties of trajectories on road networks and propose a method for searching for similar trajectories on road networks.

Our method differs from the previous methods in two aspects. Firstly, our method fully exploits the properties of road network space, whereas the previous approaches assume Euclidean space. To the best of our knowledge, it is the first method for searching for similar trajectories in road network space. Secondly, spatial and temporal similarities are considered by our method, while the previous methods only took spatial similarity into account. And our method can be used to cluster trajectories as shown by experiments.

Since this work is only a staring point of research regarding searching for similar trajectories on road networks, there are a number of related issues. First of all, studies comparing other methods presented in section 3 should be carried out. And the method for selecting POIs should be studied, while we assume in this paper that they are given by users. And integration of trajectories and the attributes of drivers will be interesting and practical for real applications such as geo-marketing or insurance industries.

Acknowledgements

This research was partially supported by the Program for the Training of Graduate Students in Regional Innovation which was conducted by the Ministry of Commerce, Industry and Energy of the Korean Government and by the Internet information Retrieval Research Center(IRC) in Hankuk Aviation University. IRC is a Regional Research Center of Kyounggi Province, designated by ITEP and Ministry of Commerce, Industry and Energy.

References

1. Dieter Pfoser, Christian S. Jensen, and Yannis Theodoridis. Novel Approaches in Query Processing for Moving Object Trajectories. In *Proceedings of the 26th International Conference on Very Large Data Bases*, pages 395–406, 2000.
2. Michail Vlachos, George Kollios, and Dimitrios Gunopulos. Discovering Similar Multidimensional Trajectories. In *Proceedings of the Eighteenth International Conference on Data Engineering*, pages 673–684. IEEE Computer Society, 2002.
3. Yutaka Yanagisawa, Jun ichi Akahani, and Tetsuji Satoh. Shape-Based Similarity Query for Trajectory of Mobile Objects. In *Proceedings of the Fourth International Conference on Mobile Data Management*, pages 63–77. Springer-Verlag, 2003.
4. Laurynas Speicys, Christian S. Jensen, and Augustas Kligys. Computational Data Modeling for Network-constrained Moving Objects. In *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems*, pages 118–125, 2003.
5. Michalis Vazirgiannis and Ouri Wolfson. A Spatiotemporal Model and Language for Moving Objects on Road Networks. In *Proceedings of the Seventh International Symposium on Spatial and Temporal Databases*, pages 20–35. Springer-Verlag, 2001.
6. Ramaswamy Hariharan and Kentaro Toyama. Project Lachesis: Parsing and Modeling Location Histories. In *Third International Conference, GIScience*, pages 106–124. Springer-Verlag, 2004.
7. K. Hornsby and M. Egenhofer. Modeling Moving Objects over Multiple Granularities. *Annals of Mathematics and Artificial Intelligence*, 36:177–194, 2002.
8. Christian S. Jensen, Jan Kolárvr, Torben Bach Pedersen, and Igor Timko. Nearest Neighbor Queries in Road Networks. In *Proceedings of the Eleventh ACM International Symposium on Advances in Geographic Information Systems*, pages 1–8, 2003.
9. Nico Van de Weghe, Anthony G. Cohn, , Peter Bogaert, and Philippe De Maeyer. Representation of Moving Objects along a Road Network. In *Proceedings of the twelfth International Conference on Geoinformatics.*

10. Michail Vlachos, Dimitrios Gunopulos, and George Kollios. Robust Similarity Measures for Mobile Object Trajectories. In *Proceedings of the Thirteenth International Workshop on Database and Expert Systems Applications*, pages 721–728. IEEE Computer Society, 2002.
11. Choon-Bo Shim and Jae-Woo Chang. Similar Sub-Trajectory Retrieval for Moving Objects in Spatio-temporal Databases. In *Proceedings of the Seventh East-European Conference on Advances in Databases and Informations Systems*, pages 308–322. Springer-Verlag, 2003.
12. Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pages 103–114. ACM Press, 1996.
13. Haixun Wang, Wei Wang, Jiong Yang, and Philip S. Yu. Clustering by Pattern Similarity in Large Data Sets. In *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, pages 394–405. ACM Press, 2002.
14. Swanwa Liao, Mario A. Lopez, and Scott T. Leutenegger. High Dimensional Similartity Search With Space Filling Curves. In *Proceedings of the seventeenth International Conference on Data Engineering*, pages 615–622. IEEE Computer Society, 2001.