Spatio-temporal Similarity Measure for Network Constrained Trajectory Data

YING XIA

School of Information Science and Technology, Southwest Jiaotong University, Chengdu, 610031, China
School of Computer Science and Technology, Chongqing University of Posts and Telecommunications
Chongqing, 400065, China
xiaying@cqupt.edu.cn

GUO-YIN WANG

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications
Chongqing, 400065, China
wanggy@cqupt.edu.cn

XU ZHANG

Department of Computer Science & Engineering, Inha University, Incheon, 402751, South Korea
zhangxu@dblab.inha.ac.kr

GYOUNG-BAE KIM

Department of Computer Education, Seowon University, Cheongju, 361742, South Korea
gbkim@seowon.ac.kr

HAE-YOUNG BAE

Department of Computer Science & Engineering, Inha University, Incheon, 402751, South Korea
hybae@inha.ac.kr

Abstract

Trajectory similarity measure is an important issue for analyzing the behavior of moving objects. In this paper, a similarity measure method for network constrained trajectories is proposed. It considers spatial and temporal features simultaneously in calculating spatio-temporal distance. The crossing points of network and semantic information of trajectory are used to extract the characteristic points for trajectory partition. Experiment results show that the storage space is decreased after trajectory partition and the similarity measure method is valid and efficient for trajectory clustering.

Keywords: constrained trajectory, road network, spatio-temporal similarity measure, trajectory clustering.

1. Introduction

Recent advances in wireless communication, sensor network and location determination technologies, make it possible to gather continuous spatio-temporal data which represent positions and movements of moving objects. Numerous spatio-temporal data with location and time attributes are able to be acquired from GPS devices, RFID sensors, RADAR or satellites. [1]

Generally, it is assumed that a moving object moves in the X-Y plane and the traversed path is a set of line segments in (x, y, t) space. These paths are defined as trajectories. According to the motion environment of the moving objects, there are two kinds of trajectories.[2] One is generated by the free moving objects, which are common in analysis of meteorology and animal behaviors. Objects move freely without any motion restrictions. The other is the constrained trajectories moving on the predefined spatial networks such as road segments. There are rich knowledge could be extracted from trajectory data, such as clusters, sequence patterns, association rules, evolution trends, outliers, etc. Trajectory data mining is very useful in decision making. However, trajectory data have such characteristics as huge volume, high dimension, dynamic growth, domain constraints and spatio-temporal correlation. There are many challenging problems such as data storage, computing efficiency, high performance mining algorithms, et al.

Trajectory similarity measure is an important procedure in trajectory mining for analyzing the behavior of moving objects. By identifying similar trajectories, various data mining techniques like clustering, classification can be applied to discover useful moving patterns. Trajectory similarity measure is originated from the similarity analysis of time series data and a lot of research works have been done. The majority of the research of trajectory similarity analysis is based on the assumption that objects can move freely without any motion restrictions. However, in many fields, e.g. intelligent transport system, most trajectories are constrained in the road network.

It is a common view that similar trajectories may indicate a potential moving pattern. Trajectory similarity analysis is useful for many applications of intelligent transport systems. For example, mining moving pattern for traffic prediction, analyzing trajectory congestion for traffic navigation, etc. In these applications, efficient representation of trajectories and similarity measurement with spatio-temporal constraints satisfying are desired. However, raw position information in trajectories of moving objects can not be stored and maintained for a long time due to the huge volume or privacy reasons.[1] In this paper, we focus on efficient spatio-temporal feature extraction and similarity analysis of network constrained trajectories. Meanwhile, semantics as *stops*[3] is introduced and refined for trajectory partition, which has been proved to be useful for improving the similarity measure effect to some extent.

The rest of the paper is organized as follows. In Section 2, related works are introduced and the research motivation of this paper is discussed. In Section 3, the method of reconstruction and partition of network constrained trajectories according to spatio-temporal feature and semantic information is detailed. Spatio-temporal similarity measure method is introduced in Section 4. In section 5, trajectory pre-processing procedure and complexity are discussed. Section 6 presents the experimental results. Finally, in section 7, we conclude this paper.

2. Related Work and Motivation

Many methods for trajectory similarity measure have been developed such as Euclidean，Dynamic Time Warping (DTW)，Longest Common Subsequence (LCSS)，Edit Distance on Real Sequence (EDR) and Edit distance with Real Penalty (ERP).[4-7] As the raw data in the trajectory is recorded as (x, y, t) without considering the road network constraint, it is nature to introduce the classic Euclidean method. However, the Euclidean distance results in poor performance regarding the quality of the results and suffers from high computation cost dealing with 3-dimensional. The other proposed methods are inapplicable for network constrained trajectories because they only considered the shape-based similarity and did not fully explore the spatio-temporal characteristics between two trajectories. E.g. as we know that the distance of trajectories in road network depends on the connectivity of road segments, two trajectories in the road network may need a great cost to reach each other even though the above distances are short.

Several similarity measure methods for network constrained trajectories have been presented also. Jung-Im Won et al. introduced a similarity measure scheme that judges the degree of similarity by considering the total length of matched road segments.[8] However, it only considers spatial similarity of trajectories. Hwang et al. presented a similarity measure method based on network distance,[9,10] which measures spatial and temporal similarity in filtering phase and refinement phase separately, and a set of points need to be selected in advance. Other methods analyze spatial and temporal similarities with a weights selection,[11,12] which splits the

spatio-temporal correlation of trajectories. Another typical trajectory clustering approach proposed by Lee[13] is focusing on the trajectory partition based on characteristic points, which inspired our research. However, the proposed method is a general trajectory method which is not optimized for road network constraint. Furthermore, they do not take account of temporal information during clustering. In fact, spatial and temporal similarity of trajectories should not be measured separately. For example, there are three moving objects A, B and C, and their corresponding trajectories TR_A, TR_B, TR_C. When the object B is moving with A, the object C moves along the same path but possess a time lag to A, the trajectory TR_A and TR_B are considered to be closing in both spatial and temporal dimensions, we regard them as spatio-temporal neighbors. TR_C shows its spatial similarity with TR_A and TR_B, however there is a time lag, so they are not adjacent in temporal dimension as well as in spatio-temporal dimension. To measure the similarity of network constrained trajectories, spatial and temporal features should be considered simultaneously.

3. Reconstruction and Partition of Trajectories

Raw positions of moving objects acquired from mobile devices such as GPS are discrete, which can be expressed by the point P(x, y, t) in (x, y, t) space. Fig.1 shows that it is easy to generate the trajectory of a moving object using linear interpolation or curve interpolation method based on its sample points. It is obvious that the motion of road network constrained trajectory is restricted by the linear road segments. These line segments connect from end to end to form a polyline in (x, y, t) space, their projections in the X-Y plane are the route of moving objects, which exactly overlap with road segments.
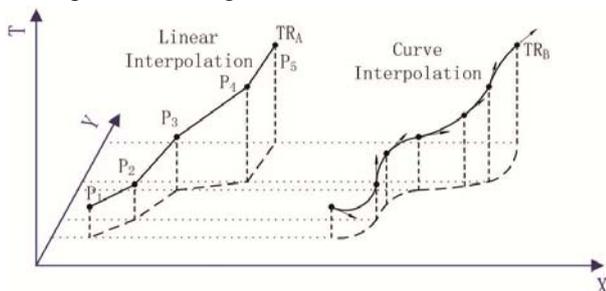

Fig. 1. Interpolation method to describe a trajectory

Raw trajectory consists of all discrete location points gathered from mobile devices, it covers entire spatial, temporal and semantic features of a moving object. The location point where the behavior of a trajectory changes rapidly is defined as *characteristic points*(*CP*).[13] Finding all the characteristic points of a trajectory is helpful to fully explore its spatio-temporal and semantic features. The purpose of trajectory reconstruction and partition is to extract all *characteristic points* and get rid of useless location data. It can reduce the storage usage and improve the efficiency for further similarity measurement.

The exact spatial, temporal and semantic features aggregated to the trajectory are related to the application context. We partition network constrained trajectory based on two kinds of *characteristic points*:

- *Crossings which indicate the joint and split points of trajectory.*
- *Stops and moves which describe the semantic information of trajectory for more meaningful understanding.*

3.1. *Crossings as Characteristic Points*

Unlike the free moving trajectories, network constrained trajectories have more restrictions in spatial aspect. Therefore, trajectory partition lies more on road network rather than the speed and direction changes of moving objects.

It is obvious that the network constrained trajectories always have spatial overlaps with the road segments, and possess a time period delay or intersection. Another observation is that the road network consists of line segments, on which the linear interpolation method is easy to be performed. We propose that to record the location data of moving objects when they across the crossing of the network and get rid of other raw sampling data, because there always be a joining and splitting around here. When the moving objects are moving under the road network constraints, the detail features of their movement seldom affect the analyzing results of trajectories. So, we can choose the crossing points as the *characteristic points* to partition the whole trajectory into subtrajectories.

3.2. *Stops as Characteristic Points*

Generally, raw location points of a trajectory have little or even no semantics, which makes it hard to extract

knowledge and understand moving pattern from the viewpoint of real-world applications. It is necessary to find a way to fill the gap between raw positions and trajectory semantics.

Spaccapietra proposed a conceptual model[3] of trajectories, which divides trajectories into two states called *stops* and *moves*. This model allows users to attach semantic data to specific parts of the trajectory. For example, in the applications of city traffic management, stops could be traffic lights, roundabouts, speed controllers, traffic jams, and so on.

We treat *move* as a sub-trajectory which describes the object position change in a temporal interval, and regard *stop* as a part of a trajectory represents that an object is staying in the same position or a small area within a period, as shown in Fig. 2.

It is important to note that moving object around crossings should not be considered as *stops*. We treat these crossings according to road network as surrounding characteristics of trajectories, while *stops* represent the semantic feature of trajectory itself.

Fig. 2 shows how *characteristic points* have been chosen by the above methods. Both *crossings* and *stops* are reserved as *CP* to partition a trajectory. The advantage of this method is that it describes the spatial, temporal and semantic features integrally while reduces the dataset size. The following section introduces the similarity measure method which is based on the partitioned sub-trajectories in the following section.

4. Spatio-temporal Similarity Measure Method

Similar trajectories should be closing in both spatial and temporal dimensions. To define a distance between two trajectories, the commensurability of spatial and temporal measuring metric should be satisfied. The metric of similarity can be normalized to [0, 1], where 0 represents two trajectories are irrelevant and 1 means they are the most similar. We propose to calculate the spatial similarity and temporal similarity first, and then transfer them into the same measuring metric to get the final spatio-temporal similarity.

Trajectories could be regarded as datasets with continuous positions. We note that the effective use of *Jaccard coefficient*[14] and *extend Jaccard coefficient*[15] for similarity measurement on sequence data. Based on the idea of *Jaccard coefficient*, we give the definition of similarity between trajectories as follow.
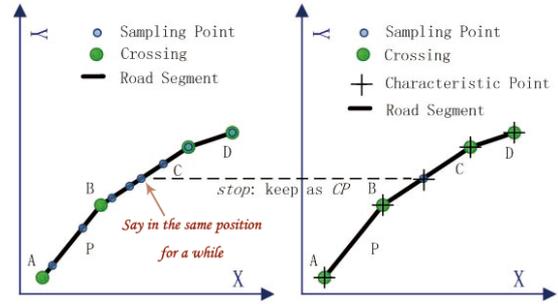

Fig. 2. Trajectory partitioning, find stops in trajectory and keep them as characteristic points

Definition 1. *The similarity of two trajectories is the ratio of the common part to the summation of the common and uncommon parts.*

$$Sim(TR_i, TR_j) = \frac{L_c(TR_i, TR_j)}{L(TR_i) + L(TR_j) - L(TR_i, TR_j)} \quad (1)$$

Where, $L_c(TR_i, TR_j)$ means the total length of the common part between trajectory TR_i and TR_j. We represent the spatial length of sub-trajectory tr_i with $SL(tr_i)$, and $TL(tr_i)$ as its lifespan in temporal dimension. $L_c(TR_i, TR_j)$ can be calculated as $\Sigma \, SL(tr_i)$ in spatial dimension and $\Sigma \, TL(tr_i)$ in temporal dimension. $L(TR_i)$ represents the total length of trajectory TR_i with $SL(TR_i)$ in spatial dimension and $TL(TR_i)$ in temporal dimension.

For example, there are four trajectories in Fig. 3, each of them has several sub-trajectories.
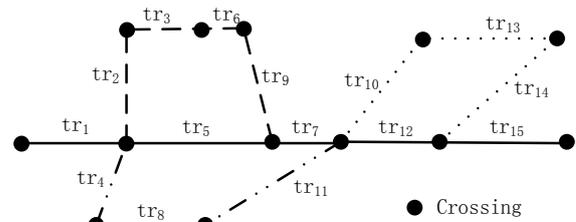

Fig. 3. Trajectory and sub-trajectory in road network

$TR_1 = \{tr_1, tr_5, tr_7, tr_{12}, tr_{15}\}$
$TR_2 = \{tr_1, tr_2, tr_3, tr_6, tr_9, tr_7, tr_{12}\}$
$TR_3 = \{tr_4, tr_8, tr_{11}, tr_{10}, tr_{13}, tr_{14}\}$
$TR_4 = \{tr_5, tr_7, tr_{10}, tr_{13}, tr_{14}, tr_{15}\}$

The similarity between TR_1 and TR_2 could be calculated as:

$$Sim(TR_1, TR_2) = \frac{L(tr_1) + L(tr_7) + L(tr_{12})}{L(tr_1) + L(tr_5) + L(tr_7) + L(tr_{12}) + L(tr_{15}) + L(tr_2) + L(tr_3) + L(tr_6) + L(tr_9)}$$

The proposed definition helps us to acquire a metric in both spatial and temporal dimension of trajectory for similarity measure. It seems that the similarity computation cost between two trajectories is very high. However, the dataset size of each trajectory is sharply reduced after trajectory partition and reconstruction.

According to definition 1, the spatial similarity (*SSim*) is mainly affected by the common segments of trajectories. Obviously, when it covers whole trajectory, then *SSim* is 1, and if there is no any common segment between TR_i and TR_j, *SSim* is 0. Temporal similarity (*TSim*) follows the same definition and its value also ranges from 0 to 1.

Most of the trajectories on road network are heavily overlapping with each other from the spatial view. Some research focus on the trajectory similarity measure only considering the spatial features in order to get shape-based similar trajectories.[8,16] However, the temporal features of a trajectory represent the moving trend varying with time, only considering the spatial features of trajectories and ignoring the temporal feature has an obvious limitation in discovering the evolution trends of trajectories and other hidden spatio-temporal knowledge. We focus on finding similar trajectories with both high spatial similarity and temporal similarity, which can be defined as spatio-temporal neighbors.

In Fig.4, trajectories in Group A possess spatial similarity. However, we can refine Group A into Group 1 and Group 2 according to the temporal similarity. Obviously, trajectories in Group 1 are spatio-temporal similar, while they are irrelevant or dissimilar with trajectories in Group 2.

Some research have the most similarity with our method[11,12,17], however, due to different measure metrics in spatial and temporal dimensions, they all need spatio-temporal weight parameters to acquire a combined similarity. The limitation is that weight parameters should be known in advance, as W_{net} and W_{time} shown in Eq. (2).

$$D_{total}(TR_i, TR_j) = W_{net} \times D_{net}(TR_i, TR_j)$$
$$+ W_{time} \times D_{time}(TR_i, TR_j) \quad (2)$$

Acknowledge from previous research that $D_{net}(TR_i, TR_j)$ is the road network costs between two trajectories,[5,16] $D_{time}(TR_i, TR_j)$ is the time costs to reach each other. W_{net} and W_{time} correspond to the weight of spatial and temporal features. The spatio-temporal similarity of trajectories may be calculated by using this
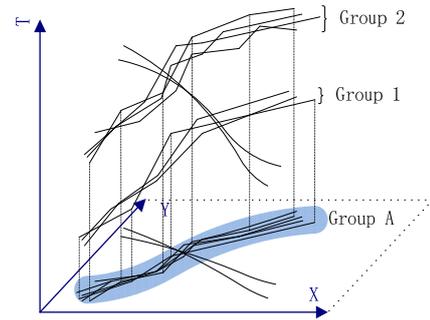

Fig. 4. Temporal feature of trajectories

method. This weight adjustments method is widely employed for similarity measure. However, it needs to predefine the coefficient of spatial and temporal features.

In particular, we aim to find the trajectories which have the most spatio-temporal similarity. The spatial and temporal features have the same impact to trajectories. As a result of this, the coefficient of them is considered to be equivalent. The Eq. (2) can be simplified to:.

$$D_{total}(TR_i, TR_j) = D_{net}(TR_i, TR_j) + D_{time}(TR_i, TR_j) \quad (3)$$

Here we present a simple analysis of function $f(x, y) = x + y$ and function $g(x, y) = x \times y$. When x and y are in [0, 1], $g(x, y) \leq f(x, y)$. So, the similarity measure method which satisfies $g(x, y)$ must satisfy $f(x, y)$. On the other hand, the character of $g(x, y)$ is that when one of the value approaches 0, the other can not influence the final value. It minimizes the impact of any single feature to the spatio-temporal similarity of two trajectories. In this paper, we argue that the trajectories should be considered as spatio-temporal neighbors when the spatial and temporal similarities are both high. Above all, we define spatio-temporal similarity and spatio-temporal dissimilarity as follow:

Definition 2. *Spatio-temporal similarity of two trajectories is the product of spatial similarity and temporal similarity.*

$$STSim(TR_i, TR_j) = SSim(TR_i, TR_j) \times TSim(TR_i, TR_j) \quad (4)$$

Similarity is a non-negative value range from 0 (no similarity) to 1 (complete similarity). The more similar two trajectories, the higher similarity they have. We have known that the spatial similarity and temporal similarity possess the same range of [0, 1], and the value of spatio-temporal similarity calculated with Eq. (4)

makes the result still between [0, 1], where 0 means that there is no spatio-temporal relationship between two trajectories, 1 indicates that two trajectories are completely overlap. The proposed similarity measure method is based on two observations. First, any single-feature similarity could not be considered as spatio-temporal similarity. Second, trajectories are restricted by both spatial and temporal features, and the influence of them are equal to each other. For example, when the spatial similarity of two trajectories is 0, the two trajectories are far from each other in spatial, No matter how the temporal similarity is, the spatio-temporal similarity measure is still 0 which indicates that they are far away from each other in spatio-temporal view, vice versa. Fig. 4 shows that this method can distinguish Group 1 and Group 2 very well. The trajectories in each group are spatio-temporal similar to each other, while trajectories between Group 1 and Group 2 are dissimilar in spatio-temporal view.

According to definition 2, we do not need to measure all the spatial and temporal similarity of the trajectories, and we mentioned that the trajectory dataset size involved in similarity computing is sharply reduced by trajectory partition and reconstruction, which makes the total time cost of trajectory similarity measure decreased. For example, when we calculate spatial similarity of two trajectories, it is unnecessary to calculate the temporal similarity if the spatial similarity value is 0.

Dissimilarity is used to indicate the difference between two objects. Generally, we regard the term of distance as the synonyms of dissimilarity.[15] In reality, similarity and dissimilarity are often interchangeable. The *Jaccard* distance, which measures dissimilarity between sample sets, is complementary to the *Jaccard* coefficient and is obtained by subtracting the *Jaccard* coefficient from 1. In this paper, we use *Jaccard* distance to indicate the spatio-temporal distance between trajectories.

Definition 3. *Spatio-temporal distance of two trajectories is expressed by its spatio-temporal dissimilarity:*

$$STDist(TR_i, TR_j) = 1 - STSim(TR_i, TR_j) \qquad (5)$$

Where, $STSim(TR_i, TR_j)$ represents spatio-temporal similarity between trajectories TR_i and TR_j. The value range of spatio-temporal distance between TR_i and TR_j is still [0, 1]. When *STDist* is 0, it means two trajectories

are full overlapped in both spatial and temporal dimensions. When *STDist* is 1, it means trajectories are far away from each other.

5. Trajectory Pre-processing

Generally, raw position data collected by mobile devices can not be used directly for data mining and knowledge discovery, the procedure of trajectory pre-processing is necessary, see in Fig.5. Firstly, trajectories are constructed based on the raw position data after linear interpolation and outlier filtering. Secondly, after matching with geographic road network, the network *crossings* and semantic *stops* are selected as *characteristic points*, and the trajectories are reconstructed. Finally, spatio-temporal similarity and distance are calculated with *Jaccard coefficient* by processing spatial and temporal features under a uniform metric. The spatio-temporal similarity and distance can be used as parameter for further data mining, e.g. clustering, classification, pattern recognition, etc.

During trajectory reconstruction, all the raw position data are scanned once. The time complexity is $O(n)$, where n is the number of raw position data. The efficiency of *characteristic points* selection is decided
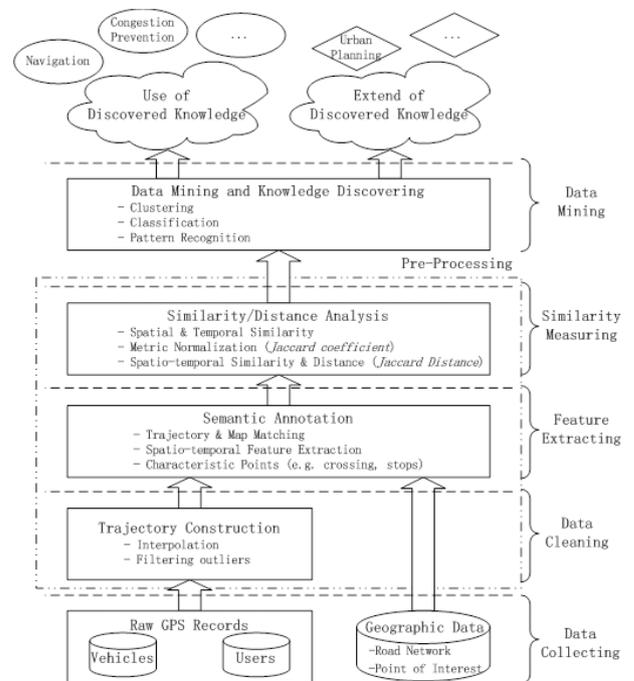

Fig. 5. Trajectory pre-processing procedure

by the complexity of road network. With the contribution of trajectory reconstruction, there are *m characteristic points* selected for trajectory partition (*m*<<*n*), the storage space can be sharply reduced. Spatio-temporal similarity measurement is time-consuming when deal with each pair of trajectories. However, with the benefit of the proposed trajectory partition and spatio-temporal similarity measuring methods, economy time performance can be gained in the experiment. Generally, the time complexity of query spatio-temporal similar trajectories could be $O(m^2)$. The time complexity could be improved to $O(log\ m)$ with the help of trajectory indexing.

6. Experiments

Experiments are implemented on a Intel(R) Pentium(R) Dual CPU T2330 @1.80GHz PC with 3G Bytes of main memory running on Windows XP Professional equipped with Java 1.6 and Matlab 7.0. Road network data generator developed by Brinkhoff is used to get trajectories.[18] Trajectory data are generated on the map of Oldenburg. The generator is modified to acquire trajectories with random *stops* (The time interval is [5ms-15ms]).

Experiment I. Dataset size of the trajectories

Five datasets are analyzed to test the dataset size of trajectories, which consists of 100, 1000, 2000, 5000 and 10000 trajectories, respectively. After performing the trajectory partition, the trajectories are divided into sub-trajectories by *characteristic points*, and most of the raw location points on road segments are erased.

We test the size of each dataset by exporting trajectory dataset into a file located in the secondary storage. The experimental results are shown in Table 1. Obviously, the reconstruction and partition of network constrained trajectories with *characteristic points* can sharply reduce the storage usage. In this experiment, we choose *stops* with a time interval larger than 10ms. *Stops* with time interval less than 10ms in a trajectory are ignored.

Table 1. Secondary storage usage of trajectories (KB)

Trajectories	100	1000	2000	5000	10000
Raw	71	759	1529	3941	8126
Partitioned	40	395	1102	1892	4431

By using the modified Brinkhoff's Generator, objects could be generated and move through various road network segments. When moving object travels on a short road segment, there will be seldom raw sampling position, while there will be more sampling position data across a long road segment.

Another experiment is performed to show the secondary storage usage when various objects are generated and traveling along road segments randomly. The parameters in Generator are defined as follows: 5 initial moving objects at release time, 7 new moving objects generated per micro-second, and run the Generator for 40ms. This experiment is performed 10 times. 278 trajectories are generated in each Generator run with different count of raw positions, the average number of raw positions is 3929.5. In Fig.6, we can see that the average raw positions of partitioned trajectories are 2205.8. Obviously, partitioned trajectories save secondary storage space in different extent by ignoring some raw positions of given trajectory without losing the original information.

Experiment II. Time cost of querying spatio-temporal similar trajectories

Addition-based Approach (*AA*) and Multiplication-based Approach (*MA*)[12] have the most similarity with our method. They conducted the experiment with Brinkhoff's Generator and showed a better retrieval time for searching similar trajectories when compared to earlier researches. Our scheme was compared with them in the same Generator to test the efficiency of retrieving
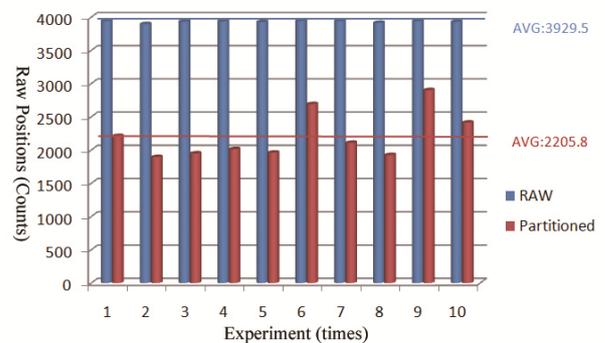

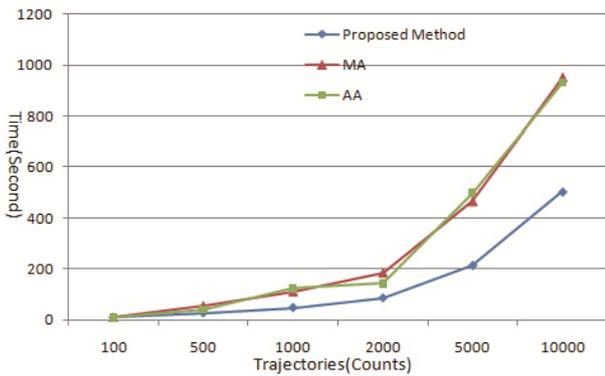Fig. 6. Comparison of storage usage between raw and partitioned trajectories

Fig. 7. Time cost of Similarity search method.

similar trajectories for any given query trajectory.

The simulation result is shown in Fig. 7, which presents the time cost of a given similar query in various trajectory datasets. The proposed method works well when the count of trajectories is below 2000, and more efficiently when the number of trajectories is increased. The proposed method is more efficient, because the *AA* and *MA* algorithm measure temporal similarity with different time period as hours, days and weeks, which are suffered to more time cost for weight selection and calculation. The proposed method can reduce the calculation time cost significantly because it only calculate the similarity when its spatial similarity is not 0. Another reason is that there is only a small count of partitioned sub-trajectories are calculated which has been proved in Experiment I.

Experiment III. Applications of spatio-temporal

similarity measure

To find spatio-temporal patterns of trajectories and predict the moving trends of moving objects, we use a clustering method to gather trajectories together with higher spatio-temporal similarity.

Trajectories in a cluster are represented as spatio-temporal neighbors. Previous researches have shown their success in trajectory clustering[8,13], however, we focus on network constrained trajectories and cluster trajectories by analyzing spatio-temporal similarity. We perform this experiment to prove that the proposed spatio-temporal similarity measure method is useful and efficient in data mining area as a pre-processing procedure. We transform spatio-temporal distance to *Eps* parameter in density-based clustering method DBSCAN[19], and define a cluster should at least possess 3 trajectories.

First, 100 trajectories on Oldenburg map are generated, and the spatio-temporal distances of trajectories are calculated with the proposed method. However, trajectory partition is performed with and without stops separately. Then, we perform trajectory clustering with spatio-temporal distance. Experiment results are shown in Fig. 8 and Fig. 9. We find that spatio-temporal similarity analysis with trajectory semantics helps to discover more interest clusters. As shown in Fig. 8, we do not consider *stops* as *CP* during spatio-temporal similarity analysis. The result represents trajectories moving in the X-Y-T space (A) and the X-Y space (B) respectively. Among them, we can see that there are three spatio-temporally closing trajectory
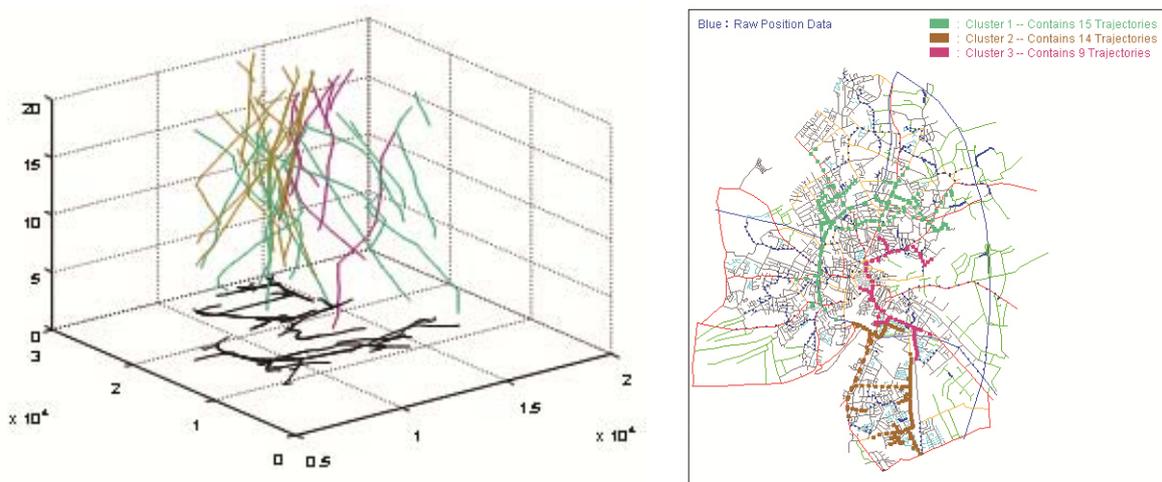

Fig. 8. Trajectory clustering results with proposed spatio-temporal similarity, *stops* characteristic points are **NOT** considered during spatio-temporal similarity analysis. (A) is clusters showed in MATLAB, (B) shows clusters in road network map.
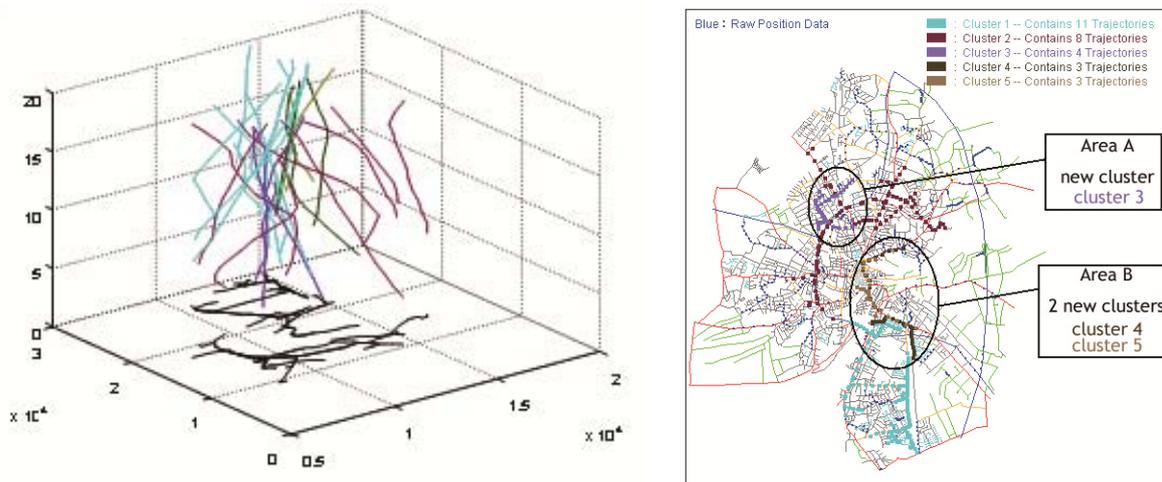
Fig. 9. Trajectory clustering results with proposed spatio-temporal similarity, *stops* characteristic points are considered during spatio-temporal similarity analysis. (A) is clusters showed in MATLAB, (B) shows clusters in road network map.

groups with their corresponding route projections on the X-Y plane in (A). Apparently, the three clustered trajectory groups have obvious similar spatio-temporal moving patterns. The experimental result shows that spatio-temporal similar trajectories can be retrieved effectively by transfer the spatio-temporal distance to a clustering parameter.

When clustering trajectories with *stops*, see Fig. 9, we analyze trajectories with *stops* by the time interval larger than 10ms. Fig. 9(A) shows 5 clusters in MATLAB with obvious moving trend and pattern. Compared with Fig.8, although we perform analysis on the same dataset, the result is different: there are 3 clusters in Fig. 8(B) while 5 clusters in Fig. 9(B). In Area A shown in Fig. 9(B), there is a new cluster deprived from *cluster 1* in Fig. 8(B). And *cluster 3* in Fig. 8(B) has been divided into two separate clusters at Area B in Fig. 9(B). From the semantic view, there should be *stops* between trajectories in *cluster 2* and *cluster 3* which indicate a significant time gap.

The experiment shows that the proposed trajectory partition and similarity measure methods can be used as pre-processing procedure for trajectory clustering, spatio-temporal semantic analysis and other knowledge discovery tasks.

7. Conclusions

Although there are many research works performed on similarity analysis for road network constrained trajectories, most of them process trajectory with spatial and temporal features in different measuring metrics. In this paper, we fist reconstruct network constrained trajectories with spatial, temporal and semantic features and partition them into sub-trajectories. Then we calculate the spatio-temporal distance and analyze trajectory similarity through considering spatial and temporal dimensions simultaneously.

Experimental evaluations show that the proposed methods are reasonable. The partition based on spatial, temporal and semantic features brings more explicit and meaningful understanding of moving pattern. It reduces storage usage and decreases time cost of retrieving spatial-temporal similar trajectories. Finally, the experiments have shown that the proposed similarity measure method could be an efficient pre-processing procedure for clustering and other knowledge discovery tasks.

Acknowledgements

References

1. L. Leonardi, S. Orlando, A. Raffaetà, A. Roncato and C. Silvestri, Frequent Spatio-Temporal Patterns in Trajectory Data Warehouses, in *Proc. of the 2009 ACM symposium on Applied Computing*, (ACM New York, NY, USA, 2009), pp.1433–1440.

2. X. F. Meng and Z. M. Ding, *Management of Mobility Data: Concepts and Technology*. (Tsinghua University Press, Beijing, 2009).

3. S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto and C. Vangenot, A conceptual view on trajectories, in *Data and Knowledge Engineering*, 65(1), (Elsevier Science Publishers B. V. Amsterdam, The Netherlands, 2008), pp. 126–146.

4. L. Chen and R. Ng, On the marriage of lp-norms and edit distance, in *Proc. of the 30th Int. Conf. on Very Large Databases,* (VLDB Endowment, 2004), pp. 1040−1049.

5. L. Chen, M. Ozsu and V. Oria, Robust and fast similarity search for moving object trajectories, in *Proc. of the 2005 ACM SIGMOD Int. Conf. on Management of data*, (ACM New York, NY, USA, 2005), pp. 491−502.

6. S. Chu, E. Keogh, D. Hart and M. Pazzani, Iterative deepening dynamic time warping for time series, in *Proc. of the 2nd SIAM Int. Conf. on Data Mining*, (Arlington, VA, USA , 2002)

7. M. Vlachos, G. Kollios and D. Gunopulos, Discovering similar multidimensional trajectories, in *Proc. of the 18th Int. Conf. on Data Engineering*, (IEEE Computer Society Washington, DC, USA, 2002), pp. 673−684.

8. J. I. Won, S. W. Kim, J. H.Baek and Junghoon Lee, Trajectory clustering in road network environment, in *Proc. of the IEEE Symposium on Computational Intelligence and Data Mining*, (2009), pp. 299−305.

9. J. R. Hwang, H. Y. Kang and K. J. Li. Spatio-temporal similarity analysis between trajectories on Road Networks, in *ER (Workshops)*, (2005), pp. 280−289.

10. J. R. Hwang, H. Y. Kang and K. J. Li, Searching for similar trajectories on road networks using spatio-temporal similarity, in *Advances in Databases and Information Systems*. (2006), pp.282−295.

11. E. Tiakas, A. N. Papadopoulos, A. Nanopoulos and Y. Manolopoulos, Trajectory similarity search in spatial networks, in *Proc. of the 10th Int. Database Engineering and Applications Symposium*, (IEEE Computer Society, Washington, DC, USA, 2006), pp. 185−192.

12. J. W. Chang, R. Bista, Y. C. Kim and Y. K. Kim. Spatio-temporal similarity measure algorithm for moving objects on spatial networks, in *Proc. of the 2007 Int. Conf. on Computational Science and its Applications*, (Springer-Verlag Berlin, Heidelberg, 2007), pp.1165−1178.

13. J. G. Lee, J. W. Han and K. Y. Whang, Trajectory clustering: a partition-and-group framework, in *Proc. of the 2007 ACM SIGMOD Int. Conf. on management of data*, (ACM New York, NY, USA, 2007), pp.593−604.

14. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd. (Morgan Kaufmann Publishers , Pittsburgh, PA, USA , 2006)

15. P. N. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*(Addison Wesley, Boston, MA, USA, 2006).

16. Y. Yanagisawa, J. I. Akahani and T. Satoh, Shape-based similarity query for trajectory of mobile objects, in *Proc. of the 4th Int. Conf. on Mobile Data Management,* (Springer-Verlag, London, UK, 2003), pp.63−77.

17. E. Tiakas, A. N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, Dragan Stojanovic and Slobodanka Djordjevic-Kajan, Searching for similar trajectories in spatial networks, *Journal of Systems and Software*, 82(5), (Elsevier Science Inc. New York, NY, USA, 2009), pp.772−788.

18. T. Brinkhoff, A framework for generating network-based moving objects, *GeoInformatica*, 6(2), (Kluwer Academic Publishers Hingham, MA, USA, 2002), pp.153−180.

19. M. Ester, H. P. Kriegel, J. Sander and X. W. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining*, (Portland, OR, 1996), pp. 226−231.