Using DTW to Measure Trajectory Distance in Grid Space

Yushun Wang, Peng Lei, Hanhai Zhou, Xiaoping Wang, Min Ma, Xiaoyun Chen
School of Information Science and Engineering
Lanzhou University, Lanzhou, China
Email: {*wangysh12, leip12, zhouhh12, wangxp12, mam12, chenxy*}@*lzu.edu.cn*

***Abstract*—With the development of science and technology, we have collected a lot of trajectory data. Obtaining useful pattern from these data is an important direction of the trajectory data mining. In order to find these models, a good distance measurement between trajectories of moving object may be necessary. In this paper, we propose a new method of distance measurement, the Grid-Based DTW (GDTW). The trajectory will be converted into continuous grid cell, then get the distance by Dynamic Time Warping (DTW) measurement. This method can more effectively measure the distance of two trajectory. Experiment results show the effectiveness of our proposed means.**

***Keywords*—moving object trajectories, grid representation, distance measure, Grid-Based DTW, DTW**

I. Introduction

As the sensor, satellite and GPS are widely used, people have got a lot of trajectory data, and the corresponding application for trajectory data mining has become very useful. For example, animal migration pattern can be found through the analysis of the migration path of animals, travel hot line can be got by the analysis of the tourist route of tourists.

In order to discover the useful pattern in trajectory set, the accurate distance or similarity measurement between trajectories is required. To look for a good distance measurement, first of all, we should choose an appropriate trajectory representation, and then, we need a reasonable distance measurement method.

Most of existing distance measure between trajectories based on Euclidean space. Trajectory in Euclidean space is represented by spatial coordinate sequence. These coordinates are discontinuous due to the different sampling frequency and speed of moving object, which will lead to that the same trajectory have many different representation, For example, under the condition of the same sample frequency, a person and a car went by the same path, their representations are different. As shown in Fig.1, trajectory $t1$ is made up of four sample points, and trajectory $t2$ is composed of seven sample points. Although the representation of two trajectories are different, they are actually the same. in other words, their distance should be 0. Unfortunately, some distance measurement methods based on Euclidean space cannot handle the problem. Therefore, in our paper, we make use of grid space representation. The trajectory in grid space is represented by continuous grid cell sequence. Due to lots of trajectory composed of spatial coordinate sequence, we need to transform trajectory to the grid space. For example, trajectories in Fig.1 can be transformed to the grid space, their representations: $t1 = \{g2, g3, g4, g5, g6, g7\}$, $t2 = \{g2, g3, g4, g5, g6, g7\}$. We can see that trajectories in Fig.1 have the same grid sequence. To some extent, this representation can ignore the impact of the different sampling frequency and speed of moving object.

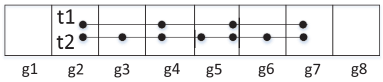


Fig. 1. The different of Euclidean representation and Grid representation

Whether Euclidean space representation or Grid space representation, the trajectory is composed of a sequence. So some sequence measurement techniques such as Euclidean distance (ED) [1], Longest Common Subsequence (LCSS) [2], and Dynamic Time Warping (DTW) [3] can be chosen. The advantages and disadvantages of them will be detailed in the related work.

In our paper, we will put forward to a novel distance measurement between trajectories—Grid-Based DTW (GDTW), the method first transform Euclidean space representation to Grid space representation by Dynamic Grid Partitioning method, and then using the DTW method measure the distance of transformed trajectories. Through experimental analysis, this way can ignore the influence of the different sampling frequency and speed of moving object, give an applicable distance measurement, and be related to the direction.

The rest of the paper is organized as follows. In section II, we introduce the concept connected with trajectory, and recommend the related work. In section III, our method GDTW would be explained. In section IV, according to experiments, we analyze and compare the various measurement methods. Finally, the conclusion is given in section V.

II. Related Work

Previous efforts of distance measurement in trajectory are mostly based on the Euclidean space, which mainly include the ED [1], LCSS [2], DTW [3]. Additionally, there are some studies on the Grid space, such as the Maximum Common Grid (MCG) [4], Grid-Based LCSS [5], etc.

In order to better introduce these methods, we first to define the trajectory and its related concepts.

In Euclidean space, the *Trajectory* and the *Trajectory Length* are defined as follows:

978-1-4799-4808-6 /14/$31.00 ©2014 IEEE

Definition 1 (Trajectory): In Euclidean space, A trajectory T_r is a finite sequence of locations with timestamps, i.e., $T_r = \{(p_1, t_1), (p_2, t_2), \ldots, (p_n, t_n)\}$ with $t_i < t_{i+1}$ for $i = 1, 2, \ldots, n-1$. p_i is a sampling point that is observed at time t_i.

Definition 2 (Trajectory Length): In Euclidean space, the trajectory length is the number of sampling point of trajectory.

In Grid space, the *Trajectory*, the *Trajectory Length* and the *Grid Size* are defined as follows:

Definition 3 (Trajectory): In Grid space, a trajectory T_r is a finite sequence of grid cell that is passed by T_r.

Definition 4 (Trajectory Length): In Grid space, the trajectory length is the number of grid cell of trajectory.

Definition 5 (Grid Size): In Grid space, Grid is a square, and grid size is the length of square side.

ED [6], [7] of two trajectories is the average l_2-norm between the corresponding sequences of two trajectories. The disadvantage is that ED cannot deal with two trajectories with different Length.

DTW [8], [9] is able to measure the similarity between two sequences which may vary in time or speed. Its basic idea is to extend trajectories by repetition of elements, and then find matching with minimizing total distance by iteration. It can be implemented by Dynamic Programming technique. The DTW measurement is related to the direction, no parameters, and can handle two trajectories with different length.

In order to analysis similarity of string or sequence, [10] firstly propose LCSS measurement. Because the trajectory is composed of a sequence, so [11] naturally apply LCSS to trajectory measurement. Given two trajectories, LCSS find the longest common subsequence of them. LCSS can be achieved by Dynamic Programming technique. Parameters ε are used to check whether two points match each other, when the distance of two points is less than ε, the two points are matched, and otherwise they are not matched. LCSS might be related to direction, sensitive to parameters and time consuming, but it can cope with two trajectories with different *Trajectory Length*. As the use of parameter ε, LCSS measurement is unreasonable.

MCG [4] thinks that the more common region of two trajectories is, the more similarity they have. MCG divides the whole movement area into the grid. Given two trajectories, MCG firstly converts the trajectory to the grid space, and then measures the number of common grid cell. We can see that MCG is unrelated to direction, it can measure the distance of two trajectories with different *Trajectory Length*, but MCG measurement is rough, because it cannot consider similarity of the parts that are not common region.

Grid-Based LCSS [5] and MCG is similar. The different point is that Grid-Based LCSS measure the size of the longest common subsequence and MCG measure the size of common area. We can see that LCSS is connected with the direction and can cope with trajectories with different *Trajectory Length*, but it is rough, because it doesn't analysis the distance of the sequence are not matched.

Both MCG and Grid-Based LCSS measurement divide the whole movement area into the grid, so the region after division

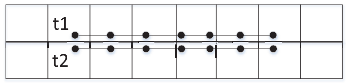


Fig. 2. The lack of Static Grid Partitioning method

is fixed, we call Static Grid Partitioning method. This division method is insufficient, as shown in Fig.2, two trajectories are very close (less than the grid size), but grid sequence after transformation are completely different, and this will lead to unreasonable distance value. Therefore, we adopt Dynamic Grid Partitioning method, which will be introduced in the next section.

III. Grid-based Dynamic Time Warping

In this section, we will explain our proposed approach, Grid-based Dynamic Time Warping (GDTW) method. First, we introduce the *Dynamic Grid Partitioning method*, and then detail the GDTW.

A. Dynamic Grid Partitioning Method

Both MCG and LCSS use the Static Grid Partitioning method, as shown in Fig.2, which is unreasonable. Therefore, in this paper we propose a feasible grid partitioning method, namely *Dynamic Grid Partitioning (DGP)*. Given two trajectories, this method first determines the boundary points of the two trajectories, and then partitions the area passed by the two trajectories from the boundary points, according to the parameter *Grid Size*. Fig.3 shows an example of the DGP method. In Fig.3, the solid triangles present the boundary points.

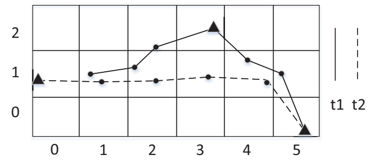


Fig. 3. An example of Dynamic Grid Partitioning method

After partitioning, the two trajectories t_1, t_2 in Fig.3 can be represent as follows:

$t_1 = \{g(1,1), g(2,1), g(2,2), g(3,2), g(4,2), g(4,1), g(5,1), g(5,0)\}$

$t_2 = \{g(0,1), g(1,1), g(2,1), g(3,1), g(4,1), g(5,1), g(5,0)\}$

B. GDTW for Trajectories

We have learned about DTW measurement in section II, and now give the following explanation for our situation. Given two trajectories, DTW measurement extends trajectories by repeating grid cell and makes grid pairs, computes the distance between grid pairs, then sums up these distance. But the trajectory after transformed have a lot of categories, the purpose of the DTW is to find the shortest distance matching under the condition of the holding time order. So-called keep order of time is the correspondence between grids can't cross.

The iterative equation of DTW measurement is given as $DTW_{t_1,t_2}(m,n)$ where t_1 and t_2 are two moving objects and m and n are the numbers of cells visited by t_1 and t_2 respectively, and

$$DTW_{t_1,t_2}(i,j) = \begin{cases} \sum_{k=1}^{i} d(k,1) & if\ j=1 \\ \sum_{k=1}^{j} d(1,k) & if\ i=1 \\ d(i,j)+ & \\ \min\left\{ \begin{array}{c} DTW_{t_1,t_2}(i-1,j-1), \\ DTW_{t_1,t_2}(i-1,j), \\ DTW_{t_1,t_2}(i,j-1) \end{array} \right\} & otherwise \end{cases} \tag{1}$$

$$d(g(i_1,j_1), g(i_2,j_2)) = \sqrt{(i_1-i_2)^2 + (j_1-j_2)^2} \tag{2}$$

In Eq.(1), $DTW_{t_1,t_2}(i,j)$ stands for the DTW distance between the first i grid cells of t_1 and the first j grid cells of t_2. $d(i,j)$ is the distance between the i-th grid cell of t_1 and the j-th grid cell of t_2. Eq.(2) demonstrates the equation of the distance between grid cells.

The equation of the DTW is given in the form of recursion. DTW is a typical Dynamic Programming problem, which can be solved by breaking the problem into several sub-problems.

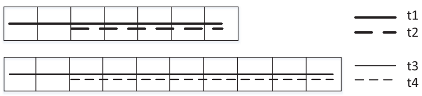


Fig. 4. The lack of DTW

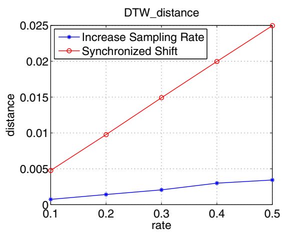
Now, we can get the DTW distance between the two trajectories, but the result may be inappropriate. Assuming the situation as shown in Fig.4, the different part of $t1, t2$ and $t3, t4$ is the same, and the same part is different. But their distance is the same, this is not reasonable, so we will make the following processing to get the final result.

$$dis(t_1,t_2) = \frac{DTW_{t_1,t_2}(i,j)}{\max\{t_1.length, t_2.length\}} \tag{3}$$

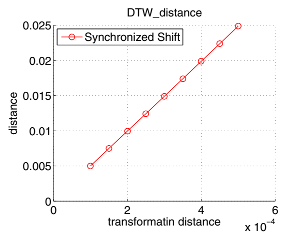
IV. Experiments

This part will show comparison of above distance or similarity methods. In order to test the reasonability of these methods, we define two trajectory transformation operations: Increase Sampling Rate Transformation and Synchronized Shift Transformation. For Increase Sampling Rate Transformation, parameter *rate* ($0 \sim 1$) control the proportion of the sample point that increase. For Synchronized Shift Transformation, parameter *rate* control the proportion of the points that are moved, and parameter *distance* control the distance of that point shifted. The detailed information about two transformations can be learn from [12]:

Our experiments is designed as follows: we employ the dataset of Beijing taxi trajectories [13]. The first group experiments, we fix the parameter *distance*($distance = 0.0003$), and then vary the parameter *rate* from 0.1 to 0.5 with the step of 0.1, to observe change of distance between the original trajectory and transformed trajectory. However, EDR and LCSS measurement use another threshold ε to determine the matched pairs of points. The relationship between *distance* and ε will heavily affect the results. Therefore, we conduct two sets of experiments for LCSS and EDR, i.e. with ε = 0.0004 being greater than distance, and ε = 0.0001 being less than distance. The second group experiments, we fix the parameter *rate*($rate = 0.3$), and change the value of parameter *distance* from 0.0001 to 0.0005 (Euclidean distance in spatial space) with the step of 0.00005, observing change of distance between the original trajectory and transformed trajectory. For EDR, LCSS, we set ε = 0.0002, between 0.0001 and 0.0005. For measurement methods with Grid space representation, we set the parameter *grid size* = 0.0003.



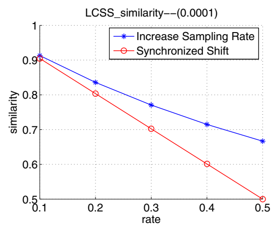
(a) Result of DTW Distance with different *rate*



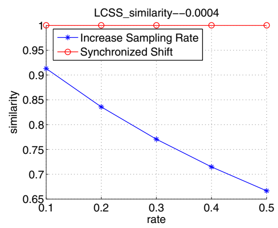
(b) Result of DTW Distance with different *distance*

Fig. 5. The result of DTW measure

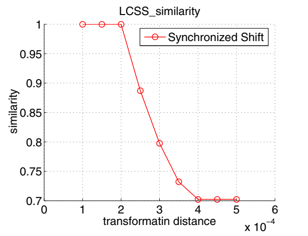
Fig.5 shows the result of the DTW measurement. For Increase Sampling Rate Transformation, the DTW is susceptible; For Synchronized Shift Transformation, the result of DTW is reasonable.



(a) Result of LCSS Distance with different *rate* and ε less than *distance* of transformation



(b) Result of LCSS Distance with different *rate* and ε larger than *distance* of transformation

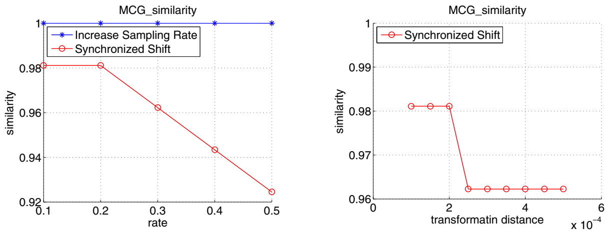


(c) Result of LCSS Distance with different *distance*

Fig. 6. The result of LCSS measure

Fig.6 displays the result of LCSS similarity measurement. We can clearly see that LCSS depends on parameters ε. Therefore, since the existence of parameters ε, we think LCSS do not give reasonable result in according to the location relation of two trajectories.

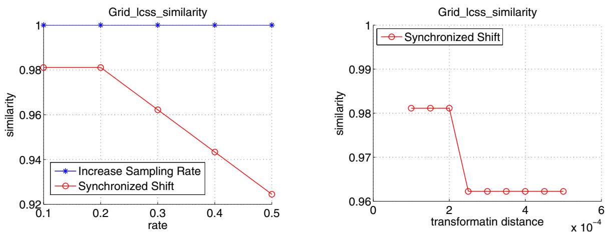
Fig.7 exhibits the result of MCG similarity measurement. Due to the use of the grid, For Increase Sampling Rate Transformation, the MCG is robust; For Synchronized Shift



(a) Result of MCG Distance with different *rate* (b) Result of MCG Distance with different *distance*

Fig. 7. The result of MCG measure

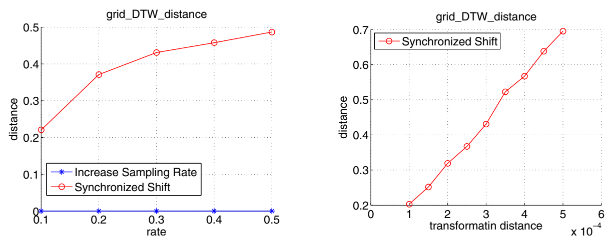
Transformation, such as Fig.7(b), as the result of MCG only measures the number of the common grid cell, without taking into account the influence of the rest part. Therefore, we think MCG do not achieve reasonable result in accord to the location relation of two trajectories.



(a) Result of Grid_LCSS Distance with different *rate* (b) Result of Grid_LCSS Distance with different *distance*

Fig. 8. The result of Grid_LCSS measure

Fig.8 reveals the result of Grid_LCSS similarity measurement. Similar to MCG, Grid_LCSS measures only the longest common subsequence, without considering the effect of the part that are not match, so it cannot give reasonable similarity.



(a) Result of GDTW Distance with different *rate* (b) Result of GDTW Distance with different *distance*

Fig. 9. The result of GDTW measure

Fig.9 shows the result of GDTW, since the use of Grid space representation, GDTW is not affected by Increase Sampling Rate Transformation; For Synchronized Shift Transformation with different *rate* or different *distance*, GDTW can win reasonable distance value, and reflect the relative location relationship of two trajectories correctly.

V. Conclusion

TABLE I demonstrates the comparison results of the different distance or similarity measurement. Y represents robustness for this method, otherwise N. It can be learn from TABLE I that GDTW we proposed is reasonable, it cannot be affected by sampling frequency and speed of moving object; Given two trajectories, it can give the reasonable distance value according to the relative location relationship.

TABLE I. Comparative Results of Measurement Methods

	DTW	**LCSS**	**MCG**
Increase Sampling Rate	N	N	Y
Synchronized shift	Y	N	N
directionality	Y	Y	N
	Grid_LCSS	**GDTW**	
Increase Sampling Rate	Y	Y	
Synchronized shift	N	Y	
directionality	Y	Y	

References

[1] A. C. Sanderson and A. K. Wong, "Pattern trajectory analysis of nonstationary multivariate data," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 10, no. 7, pp. 384–392, 1980.

[2] L. Bergroth, H. Hakonen, and T. Raita, "A survey of longest common subsequence algorithms." in *SPIRE*, 2000, pp. 39–48.

[3] F. K. Soong and A. E. Rosenberg, "On the use of instantaneous and transitional spectral information in speaker recognition," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 6, pp. 871–879, 1988.

[4] H. Zhang, R. Hu, Y. Wang, Q. Leng, and Q. Chen, "A novel method of similarity search for moving object trajectories," in *Automatic Control and Artificial Intelligence (ACAI 2012), International Conference on*. IET, 2012, pp. 235–238.

[5] H.-Y. Kang, J.-S. Kim, and K.-J. Li, "Similarity measures for trajectory of moving objects in cellular space," in *Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 2009, pp. 1325–1330.

[6] J. Richalet, A. Rault, J. Testud, and J. Papon, "Model predictive heuristic control: Applications to industrial processes," *Automatica*, vol. 14, no. 5, pp. 413–428, 1978.

[7] F. Takens, "Motion under the influence of a strong constraining force," in *Global theory of dynamical systems*. Springer, 1980, pp. 425–445.

[8] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000, pp. 285–289.

[9] Y. Sakurai, M. Yoshikawa, and C. Faloutsos, "Ftw: fast similarity search under the time warping distance," in *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2005, pp. 326–337.

[10] D. S. Hirschberg, "Algorithms for the longest common subsequence problem," *Journal of the ACM (JACM)*, vol. 24, no. 4, pp. 664–675, 1977.

[11] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "Indexing multi-dimensional time-series with support for multiple distance measures," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 216–225.

[12] H. Wang, H. Su, K. Zheng, S. Sadiq, and X. Zhou, "An effectiveness study on trajectory similarity measures," in *Proceedings of the Twenty-Fourth Australasian Database Conference-Volume 137*. Australian Computer Society, Inc., 2013, pp. 13–22.

[13] Y. Zheng, X. Xie, and W.-Y. Ma, "Geolife: A collaborative social networking service among user, location and trajectory." *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–39, 2010.