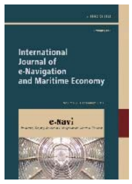
International Journal of e-Navigation and Maritime Economy 8 (2017) 001–010

e-Navi

Available online at www.e-navigation.kr
e-Navigation Journal



Original article

Douglas-Peucker-Based Accelerated Similarity Measures for Massive AIS Trajectories

Kai WANG[a], Ryan Wen LIU[b,*], Yan LI[c], Maohan LIANG[d], Yi LIU[e], Jianhua WU[f], Jingxian LIU[g]

[a] School of Navigation, Wuhan University of Technology, Wuhan, China, kwang@whut.edu.cn
[b] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China, wenliu@whut.edu.cn, Corresponding Author
[c] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China, liyanwhut@163.com
[d] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China, mhliang@whut.edu.cn
[e] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China, liuyi_hy@whut.edu.cn
[f] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China, wujh63@sina.com
[g] Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China, ljxteacher@sohu.com

Abstract

With the rapid development of the global economy, transport safety and security have become the key issues in maritime transportation all over the world. In practical applications, the Automatic Identification System (AIS)-based measurement of similarities between different vessel trajectories plays an important role in improving maritime transportation, e.g., maritime navigation, maritime supervision and management. However, the received AIS datasets are usually composed of a large amount of redundant information which could significantly increase the computational complexity. To deal with this problem, a Douglas-Peucker (DP)-based calculation method is introduced in this paper to accurately compress the spatio-temporal AIS trajectories while preserving the main geometrical structures. Based on the compressed trajectories, it is able to accelerate the Dynamic Time Warping (DTW) algorithm for the measurement of similarities between different vessel trajectories. In particular, the combination of DP and DTW has the capacity of significantly reducing the computational cost and guaranteeing the accuracy of similarity measures. The experimental results have demonstrated the superior performance of the proposed method in terms of computational cost and accuracy of similarity measures.

Keywords: Automatic Identification System, Douglas-Peucker, Trajectory compression, Dynamic time warping, Similarity measures

1. Introduction

1.1 Background and related work

With the rapid development of the global economy, a large amount of commodities are being carried through maritime transportation from place to place around the world (Kaluza *et al.*, 2010). Under this circumstance, the control of transportation safety has attracted increasing attention in maritime transportation. However, these commonly-used facilities often suffer from some potential limitations including human error, wicked time-validity and nearby environmental influences. To overcome these limitations and guarantee safe navigation, the International Maritime Organization (IMO) (2001) promotes the Automatic Identification System (AIS) in maritime transport because of its excellent performance in numerous aspects (Mokhtari *et al.*, 2007). It is a navigational safety system and mandatory vessel communication for the use of collision avoidance, coastal surveillance, and traffic management (Robards *et al.*, 2016). This device can automatically exchange the important and real-time vessel information with ships and AIS base stations nearby, which plays a crucial role in vessel tracking and services. There are mainly two primary message types in the AIS system. The main type is the static information including ship name, ship Maritime Mobile Service Identify (MMSI), ship type, ship size, etc. The other type is the kinematic information including ship location, speed, heading, rate of turn and so on (Mao *et al.*, 2016).

AIS can be utilized as an effective tool for anomaly detection, trajectory prediction, collision avoidance and AIS buoy, etc. Felski *et al.* (2015) pointed out the disadvantages of the commonly-used radar in the field of collision avoidance, since radar is effective only for constant values of ship motion parameters. Through the experiments, Felskiet *et al.* (2015) have demonstrated that AIS can effectively overcome the limitation of radar to decrease the ratio of collisions in practice. Mao *et al.* (2016) constructed a standard AIS database to illustrate the importance of trajectory prediction for automated detection of vessel abnormalities. In addition, the establishment of AIS buoys has great impact in the field of navigational safety. The AIS buoy can not only be used as beacon lights, but also can deliver and receive AIS information. It is able to guarantee the navigation security for one vessel under complex transportation environments. The visualization of vessel trajectories based on AIS data has the capacity of detecting abnormal ship behaviours and assisting maritime management (Li *et al.*, 2016) An AIS-driven partition-and-group framework was proposed in Ref. (Lee *et al.*, 2007) to perform the trajectory clustering, which plays an important role for automatic detection of abnormal vessel behaviours in ocean engineering.

There are many applications of trajectory clustering in vast scientific research fields, e.g., anomaly detection, trajectory prediction. The important and fundamental step of trajectory clustering is to measure the similarities between different trajectories. In current literature, numerous methods have been proposed to measure the similarities. Roche *et al.* (1998) considered the correlation ratio as a new similarity measure to perform the robust registration of multimodal images. This method is able to provide a good trade-off between accuracy and robustness. To select a better similarity measure to implement functional magnetic resonance image registration, Freire *et al.* (2002) have compared three similarity measure methods (i.e., Geman-McClure (GM) estimator, correlation ratio and Mutual Information (MI) relative) through numerous numerical experiments. Since trajectory clustering is often sensitive to undesirable outliers, Li *et al.* (2017) have proposed a multi-step trajectory clustering method for robust AIS trajectory clustering. In addition, Dynamic Time Warping (DTW) algorithm has been introduced to solve the problem of voice recognition. Experimental results have shown that this method could achieve satisfactory performance for similarity measures in many fields (Muda *et al.*, 2010).This paper tends to robustly measure similarities between different AIS trajectories using the DTW algorithm.

However, in practical applications, the AIS trajectories are usually composed of a huge amount of AIS data, which could increase the computational cost and destroy the time-validity. In order to accelerate the AIS trajectory similarity measures, more attention should be paid to significantly compress the AIS data while preserving the main structural properties. Theoretically, Donoho *et al.* (1998) reviewed the Shannon theory for the lossy data compression. Portniaguine and Zhdanov (2002) introduced the cubic interpolation to the 3-D magnetic anomaly inversion and demonstrated that the method is feasible and could guarantee a high quality.

For the sake of time-validity of Electronic Chart Display and Information System (ECDIS), Douglas-Peucker (DP) algorithm has been successfully applied to reduce the amount of AIS trajectory data (Zhang *et al.*, 2016). Saalfeld (1999) has certified that the DP algorithm can guarantee that the resulting simplified polyline is topologically consistent with itself and with the neighbouring features. Therefore, there is a huge potential to combine DP with DTW to accelerate the similarity measures for massive AIS trajectories.

Due to the massive AIS trajectories, the direct implementation of the original DTW algorithm will suffer from high computational cost in practice. To overcome this limitation, we propose to develop an effective DTW-based similarity measure method based on the DP algorithm. The proposed method has the capacity of guaranteeing the accuracy of trajectory similarity measures while significantly reducing the computational time. Numerous experiments will be conducted to illustrate the good performance of our proposed method.

1.2 Organization

The remainder of this paper is organized into several sections. In Section2, we briefly review the important basic concepts, including the DP and DTW methods. A two-step framework is proposed for the measures of trajectory similarities in Section 3. Section 4 presents numerous experiments to demonstrate the effectiveness of the proposed method. Finally, we conclude this paper by summarizing our contributions in Section 5.

2. Basic Mathematics Theories

2.1. Douglas-Peucker algorithm

The classic DP algorithm belongs to the category of vertex sub-sampling algorithms (Saalfeld, 1999). Massive experiments have shown that the DP algorithm is able to effectively compress the trajectories while guaranteeing the main geometrical structures. In particular, the resulting simplified trajectories are topologically consistent with the original ones, especially for the neighbouring features in trajectories (Saalfeld, 1999). Therefore, the DP algorithm will be introduced in this paper to compress the massive AIS trajectories before the measures of trajectory similarities. This preprocessing step can significantly decrease the computational cost of DTW-based AIS trajectory similarity measures. The aim of the DP algorithm is to reconstruct the trajectory with less but more important points (i.e., characteristic points) extracted from the original trajectory. The reconstructed trajectory is used to approximate the original trajectory, which can maintain the general geometrical structures.

The original trajectory A is defined as follows.

$$A=(A_1, A_2, \cdots, A_i, \cdots, A_n) \tag{1}$$

Where A denotes the trajectory point for $i=1,2,\ldots,n$. When the number of points is large enough, the original trajectory can be replaced by line segments $\overline{A_1A_2}, \overline{A_2A_3}, \ldots, \overline{A_{i-1}A_i}, \ldots, \overline{A_{n-1}A_n}$. In order to decrease the amount of trajectory points, we reconstruct the trajectory with fewer but more important points which are selected from the original point set A, i.e.,

$$A'=(A_{b1}, A_{b2}, \cdots, A_{bj}, \cdots, A_{bm}) \quad A' \subseteq A \tag{2}$$

If the characteristic points can be obtained accurately, the broken lines $\overline{A_{b1}A_{b2}}, \overline{A_{b2}A_{b3}}, \ldots, \overline{A_{b(j-1)}A_{bj}}, \ldots, \overline{A_{b(m-1)}A_{bm}}$ can substitute the original trajectory approximately. Thus, it is of great necessity to introduce the DP algorithm in this paper to compress the original AIS trajectories for the purpose of reducing the computational cost of DTW-based similarity measures.

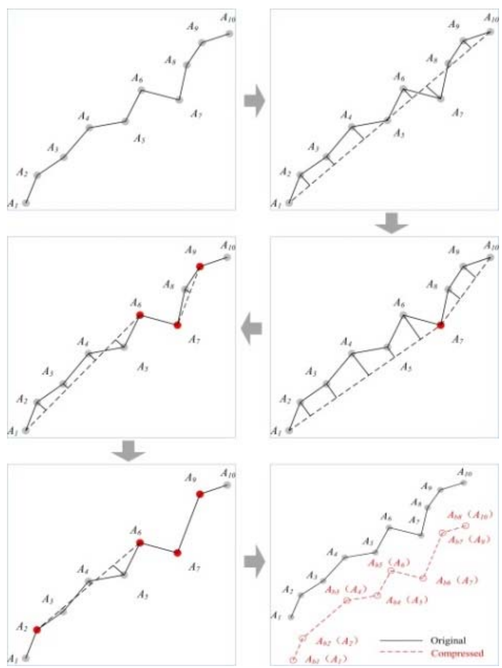


Figure 1: Flowchart of Basic Douglas-Peucker Algorithm

The schematic diagram of the DP algorithm is visually illustrated in Figure 1. The line segments connecting 10 points ($A_1, A_2, \ldots, A_{10}$) construct the original trajectory. To preserve the main geometrical structures of the original trajectory and reduce the number of trajectory points, it is necessary to extract the characteristic points from the original trajectory. First, we must select a pre-defined threshold (i.e., tolerance) λ as a benchmark to simplify the trajectory. The line ($\overline{A_1 A_{10}}$) connecting the first point (A_1) and last point (A_{10}) is taken for the datum line. Then the vertical Euclidean distance of each point to the datum line is calculated in the original trajectory. It can be found that some of the vertical Euclidean distances are larger than the fixed threshold λ, (e.g., A_7), the point related to the maximum vertical Euclidean distance will be selected to divide the original trajectory into two sub-trajectories (e.g., $\overline{A_1 A_7}, \overline{A_7 A_{10}}$) (Zhang *et al.*, 2016). This procedure will be performed iteratively until there is no characteristic point which has larger Euclidean distance than the fixed threshold λ.

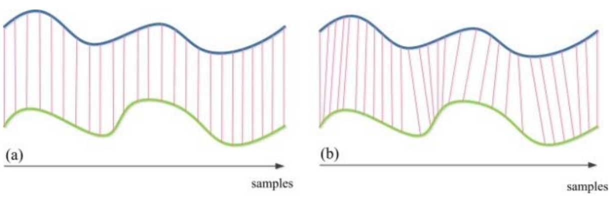


Figure 2: Similarity Measures via Primitive Approach (a) and DTW Algorithm (b)

2.2. Dynamic time warping algorithm

The DTW algorithm is capable of effectively measuring similarity between two time series which may vary in spatio-temporal domain (Muda *et al.*, 2010). This algorithm has been widely utilized in various disciplines including classification, clustering, anomaly detection and so on (Ratanamahatana and Keogh, 2004). The essential purpose of DTW is to find an optimal alignment between two given (time-dependent) sequences under certain restrictions. The traditional approaches, which compute the matching distance between two time sequences, usually extract the sequence sample points and compare the similarity sample-by-sample. The disadvantage is that the approaches could not guarantee the robust similarity measures (Rath and Manmatha, 2003), as shown in Figure 2(a).The basic principle of the DTW algorithm is to compare two time sequences and measure their similarities by computing a minimum distance between these two time sequences (Muda *et al.*, 2010). In Figure 2(b), it can be found that DTW is able to effectively overcome the weakness that exist in traditional methods.

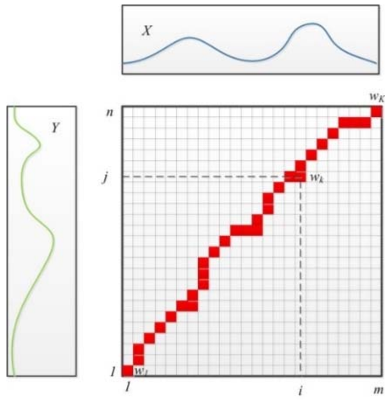


Figure 3: The Flowchart of Basic Dynamic Time Warping Algorithm

Let X and Y denote two time sequences, whose lengths are m and n respectively. These two time sequences are given by.

$$X = x_1, x_2, \cdots, x_i, \cdots, x_m \tag{3}$$

$$Y = y_1, y_2, \cdots, y_j, \cdots, y_n \tag{4}$$

To match the two time sequences based on DTW, we build an m-by-n matrix whose certain element includes the distance $d(x_i, y_j)$ between the two points x_i and y_j . In particular, the Euclidean distance is applied in this work, i.e., $d(x_i, y_i) = \sqrt{(x_i - y_i)^2}$. Each element ($x_i$, y_j) in matrix corresponds to the alignment between the points x_i and y_j. A warping path, W, is a successive set of matrix elements that defines a mapping between X and Y. Thus, we can obtain the following formulation.

$$W = w_1, w_2, \cdots, w_k, \cdots, w_K \tag{5}$$

Where $max(m, n) \leq K \leq m + n - 1$. To guarantee the accuracy of similarity measures, several constraints should be considered to restrict the warping path as follows:

a) Boundary condition: $w_1=(1,1)$ and $w_K=(m,n)$ .It is able to guarantee that the warping path can start at the bottom left and end at the top right in the matrix.

b) Continuity and Monotonicity: Given $w_k=(a,b)$ and $w_{k-1}=(a,b)$, where $0 \leq a - a' \leq 1$and $0 \leq b - b' \leq 1$. This requires that the warping path will not turn back and both i and j indexes are monotonic increasing (Muda *et al.*, 2010). In addition, the points in W should be monotonically spaced in time (Ratanamahatana and Keogh, 2004).

Generally speaking, the warping path which has the minimum warping cost will stay close to the diagonal of the matrix. As a consequence, the minimum cumulative distance $\gamma(i,j)$ can be described as follows:

$$\gamma(i,j) = d(x_i, y_j) + \min \left\{ \begin{array}{l} \gamma(i-1, j-1) \\ \gamma(i, j-1) \\ \gamma(i-1, j) \end{array} \right\} \tag{6}$$

The whole process of the DTW algorithm for time sequence (i.e., AIS trajectory in this work) similarity measures is visually illustrated in Figure 3.

3. The Two-Step Framework for Measuring AIS Trajectory Similarity

A DP-based DTW method is proposed in this work to accelerate the similarity measures between massive AIS trajectories. Although the DTW method has been widely studied in current literature, to the best of our knowledge, no research has been conducted on a DP-based DTW method for measuring trajectory similarities thus far. In particular, the proposed method is able to significantly compress the AIS trajectories while preserving the important structural properties. The DP-based DTW method could correspondingly accelerate the measures of AIS trajectory similarities in practical applications. The main steps of our proposed method can be found in Figure 4.

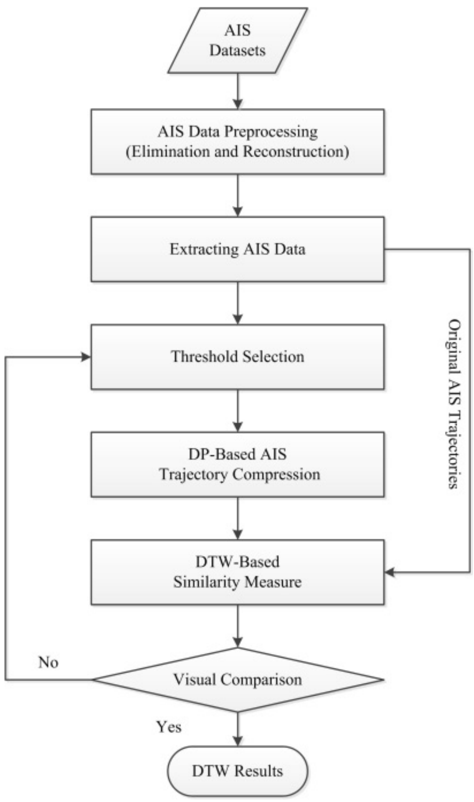


Figure 4: The Flowchart of Our Proposed Method

It is well known that the AIS data often suffers from missing data, undesirable outliers during data acquisition and transmission. To ensure the accuracy of similarity measures, it is necessary to effectively preprocess the original AIS trajectories. In this work, the trajectory points related to the missing data and outliers will be detected and reconstructed using the cubic spline interpolation (Cuche *et al.*, 2000). The introduced interpolation method is able to guarantee high-accurate results even low degree polynomials for the spline are selected. Furthermore, it has the capacity of avoiding the problem of Runge's phenomenon that often exists in high degree polynomial interpolation methods. In addition, the AIS trajectories with stay points will be deleted in this work to guarantee the robust similarity measures. We only consider the appropriate AIS trajectories in our numerical experiments. In practice, the computational time of similarity measures often suffers from long computational time due to the massive AIS trajectories. To deal with this disadvantage, the DP algorithm was introduced in the first step to effectively compress the AIS trajectories. This algorithm was able to significantly reduce the amount of AIS data records while well preserving the main geometrical structures of AIS trajectories. However, the trajectory compression quality is still sensitive to the DP threshold λ in essence. The small threshold could guarantee the quality of compressed trajectories but still lead to high computational cost during similarity measures. In contrast, the large threshold could significantly reduce the computational cost but may result in a low quality of trajectory compression. As a consequence, numerous experiments will be carried out to manually select the optimal compression threshold. Influences of compression thresholds on trajectory similarity measures will be implemented in the second step of our proposed method. To further investigate the importance of compression threshold, the computational time of AIS trajectory similarity measures will be compared under different DP compression thresholds. The optimal threshold selected is able to guarantee a satisfactory balance between the accuracy of trajectory similarity measure and DTW computational cost.

4. Experiment Results and Discussion

In this section, we tend to investigate the influences of different compression thresholds for the DP algorithm on DTW-based AIS trajectory similarity measures. The experimental AIS datasets used in this paper were collected from the AIS base stations in the Wuhan section of the Yangtze River, which includes 180 ship trajectories of 28008 pairs of coordinate points. The numerical experiments were carried out using MATLAB R2015a (The Math Works, Natick, Inc., MA) on a machine with 3.30 GHz Intel(R) Pentium(R) CPU G3260 and 4.00GB RAM.

4.1. The influences of DP compression thresholds on AIS trajectories

It is well known that the AIS trajectories are commonly composed of a large amount of coordinate points. The larger the amount of AIS data is, the more accurate the trajectories are. However, with the increasing amount of AIS trajectories, it is becoming more difficult to transmit, store and process the AIS data. As a consequence, how to compress the massive quantities of spatio-temporal trajectories has attached increasing attention. This paper aims to implement the effective compression for massive AIS trajectories while maintaining their main features.

Table 1: The Number of AIS Trajectory Points and Compression Ratio for Different DP Compression Thresholds, i.e., Tol=0, 1.0e-8, 5.0e-8, 1.0e-7, 2.5e-7, 5.0e-7, 1.0e-6, 2.5e-6, 5.0e-6, 7.5e-6, 1.0e-5, 1.0e-4

Thresholds	AIS Trajectory Points	Compression Ratio (%)
0	28008	0.000
1.0e-8	25045	10.58
5.0e-8	25027	10.64
1.0e-7	24931	10.99
2.5e-7	23545	15.94
5.0e-7	21142	24.51
1.0e-6	18713	33.19
2.5e-6	15432	44.90
5.0e-6	8811	68.54
7.5e-6	6211	77.82
1.0e-5	5053	81.96
1.0e-4	996	96.44

In order to investigate the influences of different DP compression thresholds on the qualities of AIS trajectories, we manually select 12 thresholds ranged from 0 to 1.0e-4 to perform the AIS trajectory compression experiments. For the sake of better comparison, the quantitative results for DP-based AIS trajectory compression are detailedly illustrated in Table 1. As shown in this Table, the compression ratios and the number of remaining trajectory points are statistically analysed for different DP compression thresholds. It is well known that the structural properties of remaining trajectory points are critically important for measuring the similarities between different AIS trajectories. Thus, it is necessary to select the optimal compression threshold for simplifying AIS trajectories. From Table 1, it can be observed that the number of AIS trajectory points is decreased and the compression ratio is also increased with the increasing value of the DP compression threshold. The increased compression ratio could significantly reduce the processing time of DTW-based trajectory similarity measures. If the threshold ranges from 1.0e-8 to 2.5e-7, the number of AIS trajectory points and compression ratio are changed slightly as the increase of compression threshold. In contrast, the threshold can obviously change these two assessment criteria if the threshold ranges from 5.0e-7 to 1.0e-4. The visual displays of compressed AIS trajectories under different compression thresholds could be found in Figure 5.

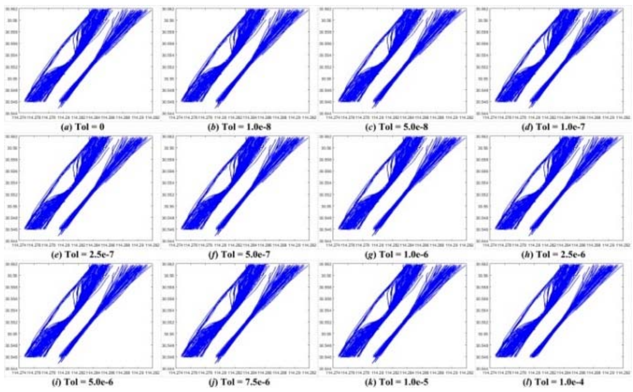


Figure 5: The Visual Displays of Compressed AIS Trajectories Under Different DP Compression Thresholds

Theoretically, the highest compression ratio is able to guarantee the lowest computational cost for measuring the AIS trajectory similarities. However, the large threshold could degrade the geometrical structures of AIS trajectories leading to negative effects on similarity measures, maritime navigation, maritime supervision and management in practice. To further investigate the influences of DP compression thresholds on AIS trajectories, we propose to visually analyse the DTW-based AIS trajectory similarities for different compression thresholds in the next section.

4.2. DTW-based AIS trajectory similarity measures

The DTW-based similarity measures are carried out for both original and compressed AIS trajectories in this subsection. First, the compressed AIS trajectories are respectively generated for 12 different DP compression thresholds (i.e., Tol=0, 1.0e-8, 5.0e-8, 1.0e-7, 2.5e-7, 5.0e-7, 1.0e-6, 2.5e-6, 5.0e-6, 7.5e-6, 1.0e-5, 1.0e-4). The influences of different compression thresholds on DTW-based trajectory similarity measures will be further investigated. In our experiments, we tend to take the trajectory similarity measure with Tol=0 (i.e., uncompressed AIS trajectories) as the benchmark. This original version will be compared with the compressed trajectories under the other 11 different DP compression thresholds. The numerous experiments on AIS trajectory similarity measures are visually summarized in Figure 6.

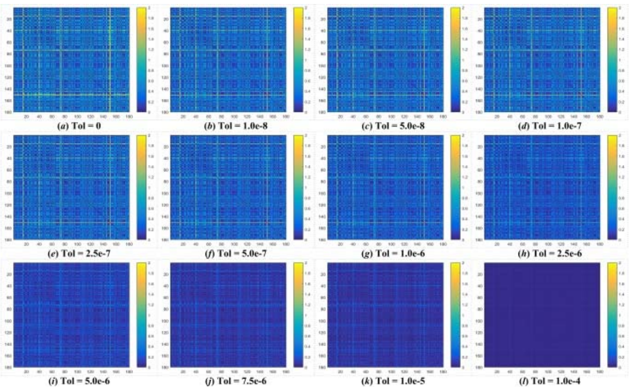


Figure 6: The Visualization of DTW-Based Similarity Measures for Compressed AIS Trajectories Under Different DP Thresholds

It can be found that the results of similarity measures with the compressed AIS trajectories are somewhat different from the result with uncompressed AIS trajectories. It is worth pointing out that if the thresholds are within a certain range, the results with compressed data will be similar to the original version. Under this condition, unobvious influences of different thresholds on DTW-based trajectory similarity measures are found in experiments. In particular, if the threshold ranges between 1.0e-8 and 2.5e-6, the DP algorithm is able to well preserve the main geometrical structures of AIS trajectories. The accuracy of trajectory similarity measurement could be guaranteed correspondingly. However, when the threshold is within a range between 5.0e-6 and 1.0e-4, the captured trajectory similarities are obviously different from the similarities calculated from the original AIS data, especially for the large compression threshold. More significant differences between original and compressed versions could be easily found as the threshold increases. It means that the structural properties of the AIS trajectories have been destroyed after DP compression with large thresholds. To guarantee the accuracy of similarity measures, more attention should be paid to the optimal selection of DP compression thresholds.

To improve the accuracy of DTW-based trajectory similarity measures, small DP compression threshold should be selected in theory. However, the small threshold in practice would lead to high computational cost. In general, with less AIS trajectory data, the processing time of DTW-based similarity measures will also be decreased accordingly. We tend to investigate the computational cost of similarity computation for compressed AIS trajectories under different compression thresholds (including original AIS trajectories). The computational time of our proposed method is summarized in Table 2. To reduce randomness, each experiment for one threshold ran 20 times to obtain the final average results.

Table 2: CPU Computational Time (mean±std) for Measuring AIS Trajectory Similarities Under Different DP Compression Thresholds (unit: s)

Thresholds	CPU Time	Similarity
0	2.8687±0.0093	—
1.0e-8	2.5549±0.0011	High
5.0e-8	2.5527±0.0020	High
1.0e-7	2.5366±0.0044	High
2.5e-7	2.3323±0.0242	High
5.0e-7	1.9923±0.0069	High
1.0e-6	1.6655±0.0011	High
2.5e-6	1.2696±0.0116	High
5.0e-6	0.5700±0.0003	Low
7.5e-6	0.3496±0.0009	Low
1.0e-5	0.2716±0.0003	Low
1.0e-4	0.0887±0.0002	Low

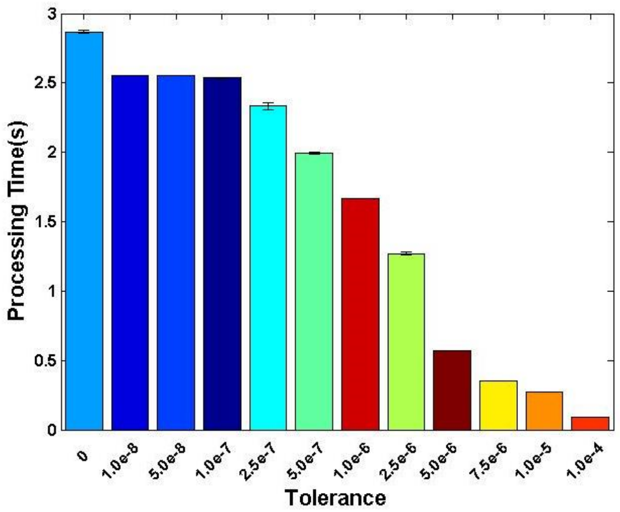


Figure 7: The Statistical Results of Computational Time for DTW-Based Trajectory Similarity Measures Under Different DP Compression Thresholds

This work mainly focuses on investigating the influence of trajectory compression on similarity measures. The implementation of similarity measures for trajectories is performed only based on compressed AIS trajectories. Before the AIS trajectories are stored in the database, the trajectories have been compressed via the introduced DP algorithm. Essentially, DP compression can be regarded as an off-line process without urgency in practice. In contrast, it has more to implement on-line similarity measure to meet the requirement of efficiency. Thus, the on-line similarity measure is considered to be more important than off-line DP compression in this work. To make an intuitionistic contrastive analysis, the statistical results of computational time for trajectory similarity computation under different compression thresholds are visually illustrated in Figure 7. It can be found that the largest threshold could generate the lowest computational time, but the resulting accuracy of similarity measures would be degraded due to the loss of structural properties of original AIS trajectories. In contrast, although the main structural properties could be satisfactorily preserved with the smallest threshold, the related computational time will be still high and constrain the practical application. Thus, there is a great potential to develop an automatic method to adaptively select an optimal threshold for achieving a good balance between the accuracy of trajectory similarity measure and DTW computational cost

4.3. MATLAB-based software for AIS trajectory similarity measures

For the sake of convenience to prospective readers, as shown in Figure 8, a MATLAB-based computer software is developed in this paper to measure the similarities between different AIS trajectories. Based on the developed computing platform, it will be convenient for us to better understanding and evaluate the proposed Douglas-Peucker-based DTW method. Our developed computing platform is mainly composed of three parts. The first part of our platform is a control panel, which contains “Load AIS Datasets” and “Threshold Selection”. The second part is about the visual illustrations of AIS points and trajectories before and after implementation of DP compression. The users could manually select the optimal threshold in the first part to investigate the influences of DP compression thresholds on AIS trajectories in the second part. The visualization of similarities for original and compressed AIS trajectories is displayed in the third part. The qualitative influences of compression thresholds on visualization performance could be investigated correspondingly. In particular, the interested users could easily upload the specific AIS datasets and perform the numerical experiments with trajectory similarity measures through the MATLAB Graphical User Interface (GUI). The resulting experimental results will be visually illustrated in our proposed computing platform. For the sake of evaluating experiments, the final source code will be freely available.

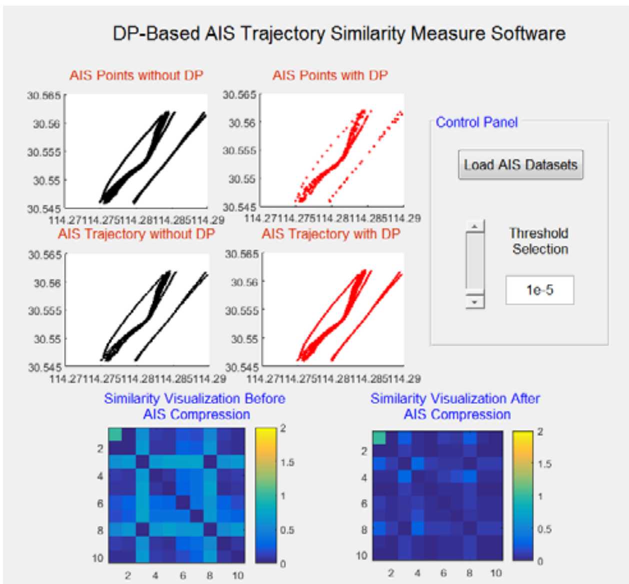


Figure 8: The MATLAB GUI of Similarity Measure Software for Massive AIS Trajectories

5. Conclusion

In this paper, we proposed a Douglas-Peucker-based DTW method to accelerate the similarity measures between massive AIS trajectories. This proposed method could significantly compress the AIS trajectories while maintaining the main geometrical structures. Therefore, the DTW-based trajectory similarity measures could be obviously accelerated in practical applications. There is a great potential to extend the proposed method to reduce the computational time for AIS trajectory clustering, which plays an important role in maritime navigation, maritime supervision and management. Honestly, the trajectory compression quality and the accuracy of similarity measures are sensitive to the DP compression threshold. In particular, the large threshold is able to significantly accelerate the DTW-based trajectory similarity measures. However, it is difficult to guarantee the important structural features and degrade the compression quality. Although it is available to enhance the compression quality through the manual selection of a small threshold, this small one could not reduce the computational cost for measuring the similarities between massive AIS trajectories. In practice, the selection of the DP compression threshold was always carried using the method of exhaustion. This manual method could achieve a good balance between the trajectory compression quality and the number of remaining trajectory points but at the expense of consuming more time. To overcome this limitation, more attention will be paid to develop an adaptive method for automatically selecting the DP compression threshold in our future work. The proposed method will have the capacity of accurately compressing the AIS trajectories while guaranteeing the quality of DTW-based similarity measures.

Submitted : June 10, 2017
Accepted : November 15, 2017

6. Acknowledgements

This work was supported by NSFC (No.: 51609195), Independent Innovation Research Funding for Undergraduates (No.: 2017-HY-A1-03), and National Training Program of Innovation and Entrepreneurship for Undergraduates (No.: 20171049712002).

References

Cuche, E., Marquet, P., and Depeursinge, C. (2000), Aperture apodization using cubic spline interpolation: Application in digital holographic microscopy, *Optics Communications*, Vol. 182, No. 1, pp. 59-69.

Donoho, D. L., Vetterli, M., DeVore, R. A. and Daubechies, I. (1998), Data compression and harmonic analysis, *IEEE Transactions on Information Theory*, Vol. 44, No. 6, pp. 2435-2476.

Felski, A., Jaskólski, K. and Banyś, P. (2015), Comprehensive assessment of automatic identification system (AIS) data application to anti-collision manoeuvring, *Journal of Navigation*, Vol. 68, No. 4, pp. 697-717.

Freire, L., Roche, A. and Mangin, J. F. (2002), What is the best similarity measure for motion correction in fMRI time series? *IEEE Transactions on Medical Imaging*, Vol. 21, No. 5, pp. 470-484.

Kaluza, P., Kolzsch, A., Gastner, M.T. and Blasius, B. (2010), The complex network of global cargo ship movements, *Journal of the Royal Society Interface*, Vol. 7, No. 48, pp. 1093-1103.

Lee, J. G., Han, J. W. and Whang, K. Y. (2007), Trajectory clustering: A partition-and-group framework, *Proceedings of the 2007 ACM SIGMOD international conference on Management of data, Beijing, China, 11-14 June 2007*, pp. 593-604.

Li, H. H., Liu, J. X., Liu, R. W., Xiong, N. X., Wu, K. F. and Kim, T. (2017), A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis, *Sensors*, Vol. 17, No, 8, pp. 1792.

Li, Y., Liu, R. W., Liu, J. X., Huang, Y., Hu, B. and Wang, K. (2016), Trajectory compression-guided visualization of spatio-temporal AIS vessel density, *Proceedings of the 8th International Conference on Wireless Communications & Signal Processing, Yangzhou, China, 13-15 October 2016*, pp. 1-5.

Mao, S. B., Tu, E., Zhang, G. H., Rachmawati, L., Rajabally, E. and Huang, G. B. (2016), An automatic identification system (AIS) database for maritime trajectory prediction and data mining, *arXiv preprint arXiv:1607.03306*.

Mokhtari, A. H., Wall, A., Brooks, P. and Wang, J. (2007), Automatic Identification System (AIS): Data reliability and human error implications, *Journal of Navigation*, Vol. 69, No. 3, pp. 373-389.

Muda, L., Begam, M. and Elamvazuthi, I. (2010), Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques, *arXiv preprint arXiv:1003.4083*.

Portniaguine, O. and Zhdanov, M. S. (2002), 3-D magnetic inversion with data compression and image focusing, *Geophysics*, Vol. 67, No. 5, pp. 1532-1541.

Ratanamahatana, C. A. and Keogh, E. (2004), Everything you know about dynamic time warping is wrong, *Proceedings of the 3th Workshop on Mining Temporal and Sequential Data, Seattle, Washington, USA, 22-25 August 2004*, pp. 1-11.

Rath, T. M. and Manmatha, R. (2003), Word image matching using dynamic time warping, *Proceedings of the International Conference on Computer Vision and Pattern Recognition, IEEE, Madison, Wisconsin, USA, 16-22 June 2003*, pp. 521-527.

Robards, M. D., Silber, G. K., Adams, J. D., Arroyo, J. Lorenzini, D., Schwehr, K. and Amos, J. (2016), Conservation science and policy applications of the marine vessel Automatic Identification System (AIS)-A review, *Bulletin of Marine Science*, Vol. 92, No. 1, pp. 75-103.

Roche, A., Malandain, G., Pennec, X. and Ayache, N. (1998), The correlation ratio as a new similarity measure for multimodal image registration, *Proceedings of the 1th International Conference on Medical Image Computing and Computer-Assisted Intervention, Cambridge, Massachusetts, USA, 11-13 October 1998*, pp. 1115-1124.

Saalfeld, A. (1999), Topologically consistent line simplification with the Douglas-Peucker algorithm, *Cartography and Geographic Information Science*, Vol. 26, No. 1, pp. 7-18.

Zhang, S. K., Liu, Z. J., Cai, Y. Wu, Z. L. and Shi, G.Y. (2016), AIS trajectory simplification and threshold determination, *Journal of Navigation*, Vol. 69, No. 4, pp. 729-744.

There is no conflict of interest for all authors.