

Big data-driven automatic generation of ship route planning in complex maritime environments

Peng Han¹, Xiaoxia Yang^{2*}

¹Administrative Center for China's Agenda 21, Beijing 100038, China

²College of Earth Science, Chengdu University of Technology, Chengdu 610059, China

Received 19 December 2019; accepted 21 January 2020

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

With the rapid development of the global economy, maritime transportation has become much more convenient due to large capacities and low freight. However, this means the sea lanes are becoming more and more crowded, leading to high probabilities of marine accidents in complex maritime environments. According to relevant historical statistics, a large number of accidents have happened in water areas that lack high precision navigation data, which can be utilized to enhance navigation safety. The purpose of this work was to carry out ship route planning automatically, by mining historical big automatic identification system (AIS) data. It is well-known that experiential navigation information hidden in maritime big data could be automatically extracted using advanced data mining techniques; assisting in the generation of safe and reliable ship planning routes for complex maritime environments. In this paper, a novel method is proposed to construct a big data-driven framework for generating ship planning routes automatically, under varying navigation conditions. The method performs density-based spatial clustering of applications with noise first on a large number of ship trajectories to form different trajectory vector clusters. Then, it iteratively calculates its centerline in the trajectory vector cluster, and constructs the waterway network from the node-arc topology relationship among these centerlines. The generation of shipping route could be based on the waterway network and conducted by rasterizing the marine environment risks for the sea area not covered by the waterway network. Numerous experiments have been conducted on different AIS data sets in different water areas, and the experimental results have demonstrated the effectiveness of the framework of the ship route planning proposed in this paper.

Key words: ship route planning, AIS, big data, trajectory data mining, electronic chart

Citation: Han Peng, Yang Xiaoxia. 2020. Big data-driven automatic generation of ship route planning in complex maritime environments. *Acta Oceanologica Sinica*, 39(8): 113–120, doi: 10.1007/s13131-020-1638-5

1 Introduction

With the inevitable rapid development of economic globalization, especially the implementation of China's national strategy *Maritime Silk Road*, international trade for China is increasing at an unprecedented rate. As a result, an increasing number of ships are actively involved in maritime transport within the coastal areas of China, to meet the urgent demand for cross-continent trades. Due to the large volume of traffic flow and relatively congested sea lanes, there is a high probability of maritime accidents in these sea areas. Accidents such as collision and grounding incidents, pose threats to the safety and security of human life, property and environmental protection. The high likelihood of maritime accidents means there is a high demand for search and rescue resources from rescue ships; further requiring the generation of safer and more reliable shipping routes for rescue ships to arrive at destinations as soon as possible. As well as marine accidents, a fast response for incidents involving offshore drilling platforms and aviation crashes that happen in sea areas should also be taken seriously. Therefore, it is of great importance that we formulate safe and reliable routes for rescue ships to arrive at the sites of maritime accidents. Maritime accidents are most likely to occur in offshore areas far away from the shore; currently relevant maritime information is provided broadly and without detailed information for specific areas, which poses a

severe challenge for maritime rescue (Futch and Allen, 2019).

Traditional ship planning routes are usually generated by manually plotting data on nautical charts. However, with the rapid advancement of information technology, a ship route planning can be generated automatically for safe navigation conditions such as water depth, and navigability (Li et al., 2007; Zhang et al., 2007, 2008). In terms of bypassing obstacles, data structures including binary trees (Wang et al., 2010) and stack (Lyu et al., 2018) are adopted to improve the efficiency of computational methods on electronic nautical charts. The width of sailing routes is taken into consideration when navigating through sea areas with more obstacles (Wang et al., 2016). Nevertheless, formulating ship planning routes by looking only at nautical charts will lead to potential safety hazards, especially when maritime accidents happen in sea areas without complete, detailed and comprehensive maritime information.

From 2002, Automatic Identification System (AIS) has become a compulsory facility for ships of over 300 gross tonnage (GT) on international voyages, cargo ships of over 500 GT in all waters, and all passenger ships regardless of the dimensions of the ship. The ships' trajectories contain rich maritime information and have been recorded in detail by the AIS. Correlative information can be deeply excavated by correlative data mining techniques, such as density-based spatial clustering of applica-

*Corresponding author, E-mail: yangxx2003@126.com

tions with noise (DBSCAN) (Ester et al., 1996; Schubert et al., 2017) and its variants (Campello et al., 2015; Kumar and Reddy, 2016; Bryant and Cios, 2018; Li et al., 2018), thus providing significant assistance for the routes formulation (Altan and Otay, 2017; Le Tixerant et al., 2018; Svanberg et al., 2019). The principal fairways were extracted by Chen (Chen et al., 2015), which are based on AIS data to provide a navigation basis for deck officers in the Taiwan Strait. Massive turning points were accurately obtained in the Qiongzhou Strait by analyzing huge amounts of AIS trajectory data, and then the Ant Colony Algorithm (Zhang et al., 2018) was applied to optimize the generated planning routes. These studies primarily target at mining of the specific maritime information in particular sea areas with relatively integrated information, like point (e.g., turning point) and line (e.g., sailing traces) information, and do not comprehensively dig out the empirical navigation as a whole. Maritime networks can also be generated with the use of AIS data, but generating ship planning routes with such information has not been researched to date. The waterway network based on grid data fails to offer comprehensive navigation information (Wang et al., 2019). AIS data are simply the record of historical trajectories of a large volume of vessels, and the trajectories composed of points and lines constitute the traffic information network. Such empirical information can serve as a perfect reference for voyage formulation with the assumption that the routes traversed by most vessels are considered fairly safe and reliable. Therefore, this paper aims to employ the information hidden in maritime big data to generate empirical ship paths. We are particularly interest in how to implement the maritime traffic network information excavated from the massive AIS trajectory into the automatic generation of ship routes, how much difference lies in between the waterway information obtained from this empirical trajectory and the route information in the electronic chart, and whether it can be used as a navigation reference path.

An integrated maritime traffic network is proposed in this paper with the navigation experience excavated from the AIS trajectories; the ship planning routes were generated automatically based on the constructed framework. First, the trajectories were clustered based on the density-based clustering method, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and then the experiential shipping routes were extracted on the basis of the clustered trajectories. Next, the maritime navigation network was constructed based on the experiential routes to build the node-arc network topology information. Finally, shipping routes were generated by a network optimization method based on empirical navigation networks in sea areas covered by the navigation network. In other areas, the planning routes

were generated by incorporating external conditions such as hydrometeorological information. The method proposed in this paper specialized in generating ship route planning for rescue ships in a lower timeframe, while providing safe and reliable routes for rescue ships to reach sites of maritime accidents as soon as possible.

2 The proposed framework

An advanced method for generating ship planning routes is proposed in this paper. The implementation block diagram is shown in Fig. 1. The method consists of two parts: the construction of a maritime empirical route network, and the planning route generation (the first part consists of three primary procedures). First, an increased number of AIS trajectories are grouped into different clusters according to spatial location and geometric characteristics. Then, centerlines can be extracted in each different trajectory cluster. Finally, the network topology is constructed based on the centerlines extracted from the trajectories. Take the point T in the sea area as the destination, where the destination is generally not on the grid of the route that has been generated by the empirical route network. Thus, in the route generation, the position M on the route network (including edges and nodes) that is closest to the destination T in a straight line needs to be determined first. Then, an optimal path could be calculated from the departure port to the position M on the route network. Next, according to the water depth information of the specific sea areas and by avoiding the stranded areas, another optimal path from position M to destination T is also generated based on the raster method. Eventually, the two optimal paths mentioned above are integrated to generate the final ship planning routes.

3 Construction of the experiential ship route network

3.1 AIS trajectory clustering

The basic idea of the trajectory clustering algorithm is to cluster the historical AIS trajectories using the DBSCAN clustering algorithm based on the structural similarity of the trajectories (Ester et al., 1996; Wang et al., 2004; Yuan et al., 2011; Wei et al., 2016). To obtain better clustering results on the AIS historical trajectories, it is necessary to preprocess the ships' trajectories. Three types of trajectories that may hinder the extraction of ship sailing routes should be removed. The first type is a trajectory that has a low sampling rate for the number of trajectory points; they are too small to represent the actual traces of the vessels. A threshold N is set beforehand to delete the trajectory if the number of trajectory points is less than N . Trajectories with vessels

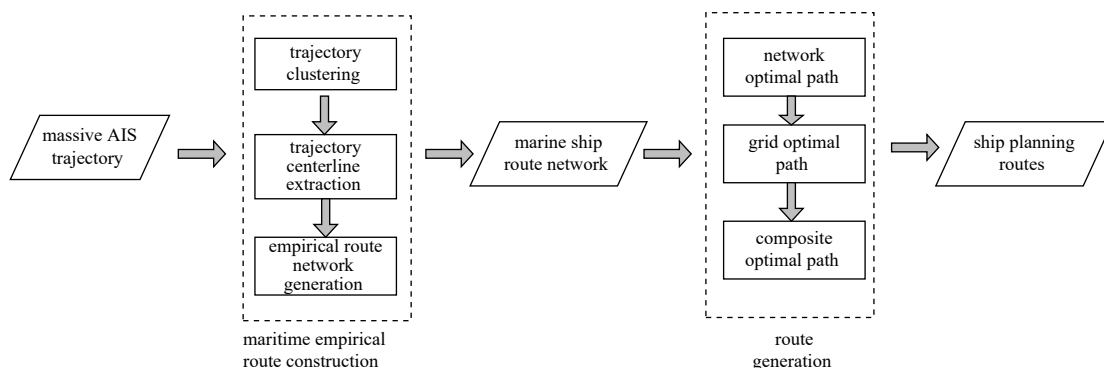


Fig. 1. Block diagram of ship planning routes based on empirical navigation information.

drifting or moving within a small area range should also be cut off. This type of trajectory is identified based on the relationship between the cumulative length d of the trajectory, and the linear distance D from the start point to the end point of the trajectory. If d is greater than five times D , d is considered to be much larger than D ; such trajectories fall into this category of trajectory and should be removed. Likewise, trajectories with an abnormal speed cannot be considered. Generally, the speed v of a ship is in (v_1, v_2) , where v_1 is a low-speed threshold and v_2 is a high-speed threshold. Trajectories with a speed outside of the defined range should also be removed. After removing these three representative types of trajectory, the DBSCAN clustering algorithm can be applied based on the structural similarity of the ships' trajectories.

The ship's trajectory structure is a collection of maritime attributes reflecting the characteristics of the ship's trajectory, such as tracks, position, velocity. The similarity of ship's trajectory measured in this paper is based on two attributes, i.e., direction and position. The measurement of the trajectory direction is called the structural distance in trajectory direction, and the measurement of the trajectory position is called the structural distance in trajectory position, and finally the structural distance can be obtained by weighted sum.

A complete trajectory consists of many sub-trajectories. Considering the trajectory may influence the clustering efficiency, a threshold is set for the controlling angle of two adjacent sub-trajectories. Turning angle of a trajectory refers to the difference between the trajectory directions of two adjacent sub-trajectories. If the difference between the trajectory directions of two adjacent sub-trajectories is less than the defined threshold, it means that the angles of the two trajectories are similar, then the two adjacent sub-trajectories are integrated into one high-level sub-trajectory; if the direction of the two adjacent sub-trajectories varies significantly, that is, the angle difference is greater than the defined threshold, then the two sub-trajectories are retained.

(1) Structural distance in trajectory direction

The structural distance in trajectory direction represents the change of the trajectory direction of the ship trajectory. T_i and T_j are two trajectories, respectively. Then the structural distance in trajectory direction $D(T_i, T_j)$ between the two trajectories can be calculated as

$$D(T_i, T_j) = \begin{cases} \min(\|T_i\|, \|T_j\|)\sin\theta, & 0^\circ \leq \theta < 90^\circ, \\ \infty, & 90^\circ \leq \theta \leq 180^\circ, \end{cases} \quad (1)$$

where θ is the angle between T_i and T_j .

If the angle between the two trajectories is less than 90° , it means that the trajectories are on the same direction, while the angle between the two trajectories is more than 90° , it is considered to be on the opposite direction. If the two trajectories have opposite directions, the structural distance in direction is set to ∞ , so as to distinguish the trajectories with opposite directions during clustering process.

(2) Structural distance in trajectory position

The structural distance in trajectory position represents the difference in ship trajectory position. This difference is measured by the Hausdroff distance method as

$$\begin{cases} L(T_i, T_j) = \max(h(T_i, T_j), h(T_j, T_i)), \\ h(T_i, T_j) = \max_{a \in T_i} (\min_{b \in T_j} d(a, b)), \end{cases} \quad (2)$$

where $L(T_i, T_j)$ is the Hausdroff distance between T_i and T_j ; $h(T_i, T_j)$ is the one-way Hausdroff distance from T_i to T_j ; $d(a, b)$ is the Euclidean distance between two points a and b , a is the point in T_i , and b is the point in T_j .

To calculate the ship's trajectory structural distance, various structural distances need to be given weights. In general, the weights are divided equally. Assuming that the weight of the structural distance in trajectory direction is W_D and the weight of the structure distance in trajectory position is W_L , then the ship's trajectory structural distance SD can be computed as:

$$SD = DW_D + LW_L. \quad (3)$$

3.2 Centerline extraction of shipping waterway

The shipping waterway is defined as the belt area formed by the sailing routes between different ports, where the main characteristics of the shipping waterway can be reflected by the centerline of the belt areas. Therefore, the centerline of the belt area formed by the ships' tracks in this paper is represented by the centerline of the empirical shipping waterway. From the result of trajectory clustering, the centerline of shipping waterway is the characteristic line of each cluster, which represents the broken line of the length, direction, position and shape of the whole cluster. By solving the pair by pair trajectory feature lines method, the extraction of trajectory feature line (i.e., centerline of shipping waterway) is transformed into finding two trajectory feature lines.

The basic idea of extracting the centerline of the shipping waterway is as follows: First, two trajectory lines T_1 and T_2 are selected in the same kind of trajectory set to find the trajectory centerline C_1 . Second, another trajectory T_3 is selected in the same trajectory cluster. Then the centerline C_2 between C_1 and T_3 is calculated once more, and then the previous centerline C_1 is replaced by the current centerline C_2 . The calculation is carried out iteratively in the same trajectory cluster. Finally, a centerline of the trajectory will be obtained in each category, which is also the centerline of the shipping waterway.

As shown in Fig. 2, lines A and B in the figure are two trajectories in the same category, trajectory A and B have 6 nodes, respectively. The specific process of centerline extraction contains the following procedures:

- (1) Connect A_1, B_1 and A_1, B_2 to form triangle $A_1B_1B_2$.
- (2) Calculate the midpoint C_1 of the vertical line from A_1 to line segment B_1B_2 as the starting point of the track centerline.
- (3) Select point A_2 on trajectory A , and find the two points B_2 and B_3 on trajectory B which are the shortest distance from A_2 .
- (4) Calculate the midpoint C_2 of the vertical line from A_2 to line segment B_2B_3 .
- (5) By analogy, other midpoints C_3, C_4, C_5 , and C_6 can be calculated, and finally the centerline C can be obtained by connecting these midpoints.

3.3 Generation for the empirical route network

Nodes are generated at the intersections of the centerlines extracted from trajectories, and the node arc segment topology relationship is formed to construct a waterway network, as shown in Fig. 3. Suppose two centerlines D and E extracted from two trajectories are intersected, node N is generated at the intersection, and centerline D and E are broken into arc segments D_1, D_2 and E_1, E_2 at node N respectively, as shown in Fig. 3b. In this way, D_1, D_2, E_1 and E_2 are recorded as the connected arc segments at node

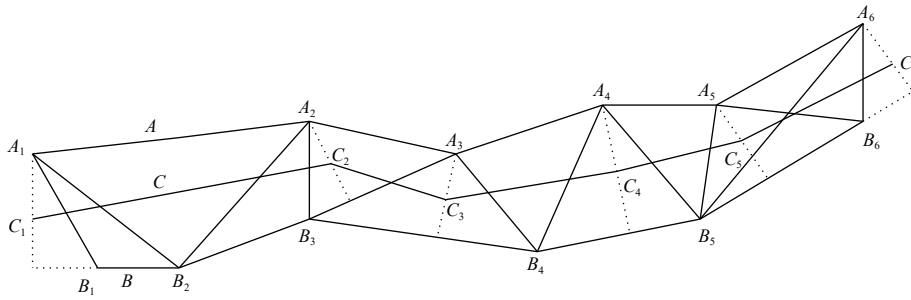


Fig. 2. Schematic diagram of trajectory centerline extraction.

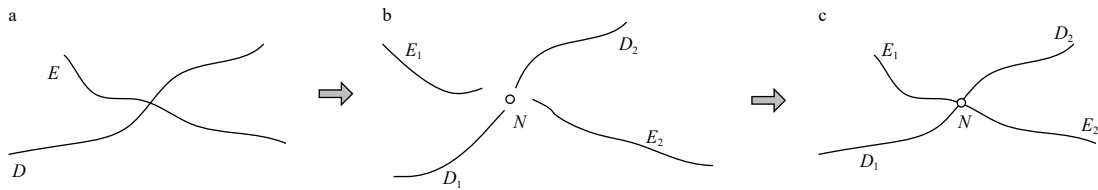


Fig. 3. Generation for the empirical route network: two AIS trajectories (a), node N at the intersection (b), and generated node-arc topology (c).

N , and N is recorded as the connected node at D_1 , D_2 , E_1 and E_2 arc segments to complete the establishment of the topology relationship of the node-arc segments, as shown in Fig. 3c. The network is established by the node-arc topology relationship on the centerlines of all intersecting trajectories, and therefore the empirical route network can be generated.

4 Automatic generation of ship route planning

Taking a certain point in the sea area as the destination T , M is the nearest position that has the minimum linear distance to T on the waterway network (Fig. 4). First, the planning route generation method yields the optimal path to M from the departure port based on the waterway network. Then, according to information of water depth in the sea area, an optimal path from M to T is generated based on the raster method by evading stranded areas. Eventually, we integrate the optimal path on the channel network with respect to the path from M to T generated in the sea area of destination T , which forms the final ship planning routes.

4.1 Network optimization route generation

From a certain port to the position M on the waterway network, the optimal path is searched by the network path optimization

method. As the scale of arc nodes on the waterway network is relatively small, the classic shortest path algorithm (i.e., Dijkstra) is adopted (Dijkstra, 1959; Zhan, 1997). The Dijkstra algorithm is primarily used to calculate the shortest distance from one vertex to other vertices in the network, characterized by expanding from the starting point to the outer layers until the target point is reached. For a directed graph with all positive weight vertices, Dijkstra firstly sets the starting point and target point, and then all vertices of the directed graph are divided into two groups represented by S and U , respectively. Specifically, S and U store the vertices whose shortest distances have been calculated and uncalculated, respectively. Next, the distance values of each vertex of the directed graph are defined. The distance values of each vertex in S are the distances of the shortest path of the starting vertex to each vertex. The values of each vertex in U are the distances of a path from the starting point to this vertex. At the beginning of Dijkstra algorithm, there is only one starting vertex in S , whose distance is zero. The values of the vertices in U are updated when the least cost path is found, and the vertices in U are then added to S in increasing order of distance until U is empty. This must satisfy the requirement that the distance of each vertex in S is lower than that in U .

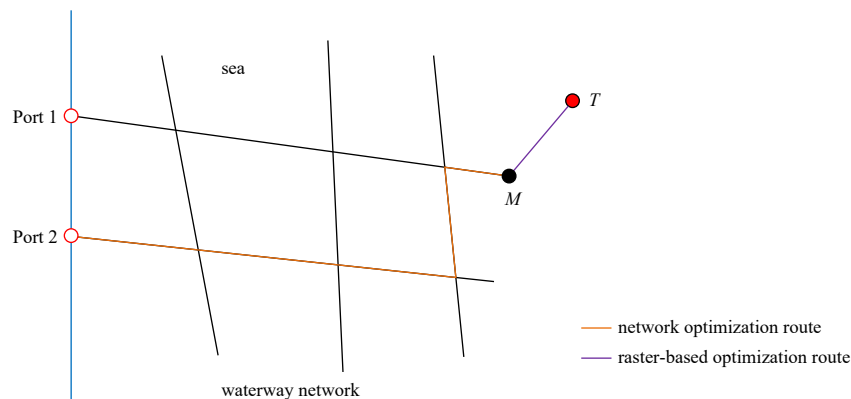


Fig. 4. Schematic diagram for ship route planning.

4.2 Raster-based generation of ship sailing routes

As shown in Fig. 4, from M in the waterway network to destination T in the sea area, the areas where shallow waters may make ships stranded are set as obstacle areas in terms of the water depth information, and the sea areas are set as a grid. The Basic theta* algorithm (Nash et al., 2007) is utilized to achieve the calculation of the optimal path from the position M to the destination T . Basic theta* is an improved method of A*, whose biggest characteristic is that it can move towards any angle without constraining the raster edge, so as to make the path smoother (Muñoz and Rodríguez-Moreno, 2012; Garcia et al., 2013). The difference between Basic theta* and A* is the procedure of the extension of nodes. On the basis of A*, Basic theta* makes a line of sight check between the adjacent node N of the current node P and the parent node P_{parent} of P . If visible, the costs of these two paths (where one is from P_{parent} directly to N and another is from P_{parent} through P to N) are compared, and then the path with least cost is selected. The implementation of the Basic theta* algorithm is as follows:

If S and E denote the starting node and the target node, respectively, for any node P , the total movement consumption from S to the current node P is represented as $g(P)$, the diagonal distance from node P to the target node E is denoted as $h(P)$, the movement consumption from node P to its adjacent node N is recorded as $d(P, N)$, and the value $f(P)=g(P)+h(P)$ is recorded for priority sorting.

(1) The starting node S and the target node E are selected, and putting $(S, 0)$ whose general form should be $(node, value, f)$ in *openList*, which is a priority queue in which the smaller the node f value is, the higher the priority is.

(2) Determining the *openList* is empty or not. If it is empty, the search fails and the target node is unreachable. Otherwise, the highest priority node P in the *openList* is taken out. The node P is determined whether it is the target node E . If so, the search is successful as the parent node of node P is obtained, and the process is iterating (continue to obtain the parent node of the parent node) until the initial node S is found, so as to obtain a path from P to S . Otherwise, repeat Step (2) to Step (5).

(3) Traversing eight adjacent nodes N_1-N_8 of P , and doing sight check between the adjacent nodes N of the current node P and the parent node P_{parent} of P . If it is visible, the costs $g(N)$ of

this two paths which are from P_{parent} directly to N and from P_{parent} through P to N , are compared. Then, the path with less cost is chosen, the $g(N)$ value and the parent node of N are updated simultaneously.

(4) If N has already in *closeList*, the following two cases can be ignored. One case is that once N is not in *openList*, then the diagonal distance $h(N)$ from N to E is calculated. Next, let $f(N)=g(N)+h(N)$, $(N, f(N))$ is put into the *openList*. Another case is that if N is in *openList*. Then, $f(N)$ is recalculated by $g(N)+h(N)$. The old $(N, f(N))$ in *openList* is also replaced with a new $(N, f(N))$. Otherwise, it is not treated.

(5) Put the node P into *closeList*.

4.3 Composite ship sailing routes

The method of the sailing route generation in its entirety combines the two procedures mentioned above. From shore-based ports to a waterway network point M , the sailing route is obtained using the shortest path in the waterway network. From point M to the destination point T , the path can be calculated according to the raster method. One complete sailing route from the port to the destination T is yielded by combining these two paths at point M .

5 Experimental results and analysis

5.1 Experimental settings

The Changjiang River (Yangtze) Estuary and Zhoushan Islands are taken as the experimental areas in this paper. The estuary of the Changjiang River (Area 1) includes the Chongming Island and several small islands around it. Several ports on the Chongming Island are the main ports for ships' docking. One month of AIS trajectory data (Table 1) near the waters is taken as the experimental data, as shown in Fig. 5a. Figure 5b displays the East China Sea area of the Zhoushan Islands (Area 2) with a month of AIS trajectory data (Table 1).

The AIS data used in this paper is in a raw data format. There are two types of AIS information, AIVDM (AIS VHF data-link message) and AIVDO (AIS VHF data-link Own-vessel report). The binary decoder is utilized to transform the raw AIS into the ASCII format. The ship type, MMSI (Maritime Mobile Service Identity), latitude and longitude, heading, and speed informa-

Table 1. Detailed information of the AIS experimental raw data

	Number of ships	Number of points	Time range
Trajectories in Area 1	9 000	16 183 696	Feb. 10 to Mar. 9, 2016
Trajectories in Area 2	7 273	14 194 746	Aug. 10 to Sep. 9, 2016

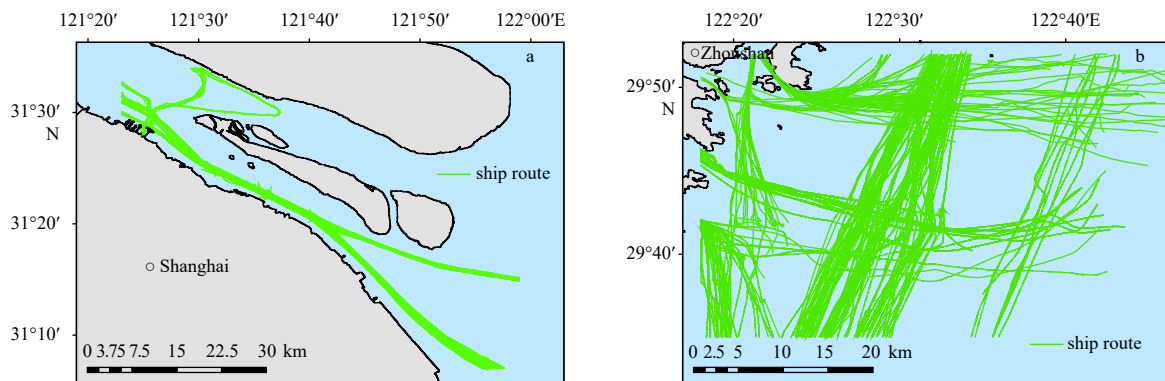


Fig. 5. The experimental areas with AIS trajectory data: the Changjiang River Estuary (Area 1) (a) and Zhoushan sea area (Area 2) (b).

tion are the main target attribute information obtained by AIS decoding.

5.2 Results and analysis on ship planning route network generation

Using the AIS trajectory data of two experimental areas, the waterway network was constructed based on the method proposed in this paper. First, the structural distance was applied to implement the DBSCAN clustering algorithm on the AIS trajectories of the two experimental areas. Then, the centerline was extracted based on the clustered trajectories, and the waterway network was established by the topology of the centerline. The network was generated by the nodes and the waterway between the nodes. The DBSCAN-based clustering results of the two experimental areas are shown in Figs 6a and b respectively, where different trajectory clusters are formed. The centerlines of the waterway were obtained by the methods introduced in this paper, and a topological relationship was established to form the waterway network, as shown in Figs 6c and d.

The generated centerlines of the waterway were compared with the centerlines of the area in the electronic chart to evaluate the effectiveness of the route generation algorithm proposed in this paper. By comparing the route (purple) generated by the AIS empirical trajectories with the centerlines (yellow) of the waterway area in the electronic chart (Figs 7b and d, respectively), ten points (including turning points and the points uniformly distributed along the centerline) were extracted, then these points were projected to the centerline of the channel of the electronic chart. The distance was calculated from the point to its projection point as l_i , then the route similarity was measured as:

$$MEA = \frac{1}{10} \sum_{i=1}^{10} l_i. \quad (4)$$

To compare the differences between the generated ship planning routes and the waterway centerline in the electronic chart, according to the selected points, the distances between the points in the ship routes and the waterway centerline were applied to represent the differences (including the maximum distances, the minimum distances, the total distances, and the average distances). The detailed results can be found in Table 2, and we found that the average distances in Areas 1 and 2 were 1.2592 km and 0.2057 km, respectively.

5.3 Generation of composite ship planning routes

The results of the composite ship planning routes based on the empirical AIS trajectories are displayed in Fig. 8. The generated ship planning routes were mainly composed of two parts, i.e., waterway network- and raster-based ship planning routes. In the water areas that were covered by a waterway network, the ship route labeled in red was calculated by selecting the optimal navigation path. In the water areas without a waterway network, the ship route labeled in green was generated according to the proposed raster method. The risk of water depth has also been considered in our proposed method, to improve the accuracy of planning route generation.

6 Discussion and conclusions

In this study, a big data-driven mathematical framework was proposed to automatically generate ship planning routes under different navigation environments. This method extracts the waterway network by AIS trajectory data and generates the ship's

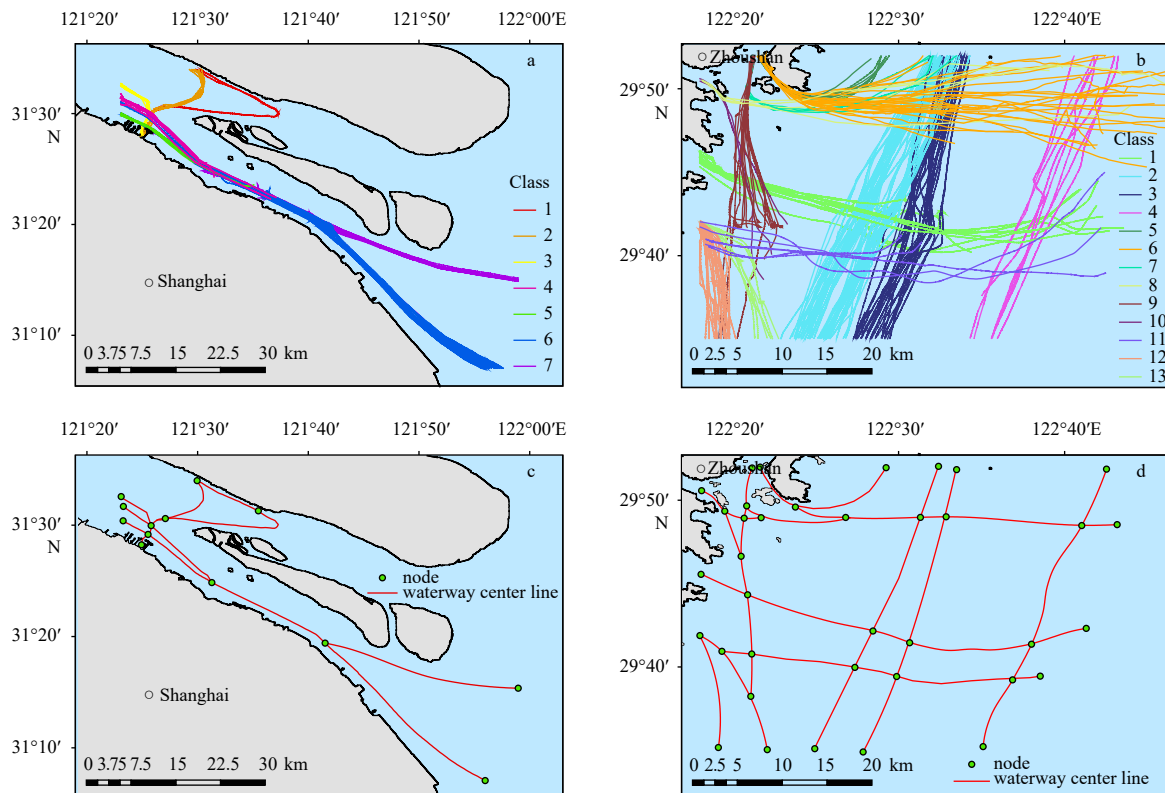


Fig. 6. Waterway network constructed in the experimental area: the DBSCAN clustering results for Area 1 (a); the DBSCAN clustering results for Area 2 (b); the generated waterway network for Area 1 (c); and the generated waterway network for Area 2 (d).

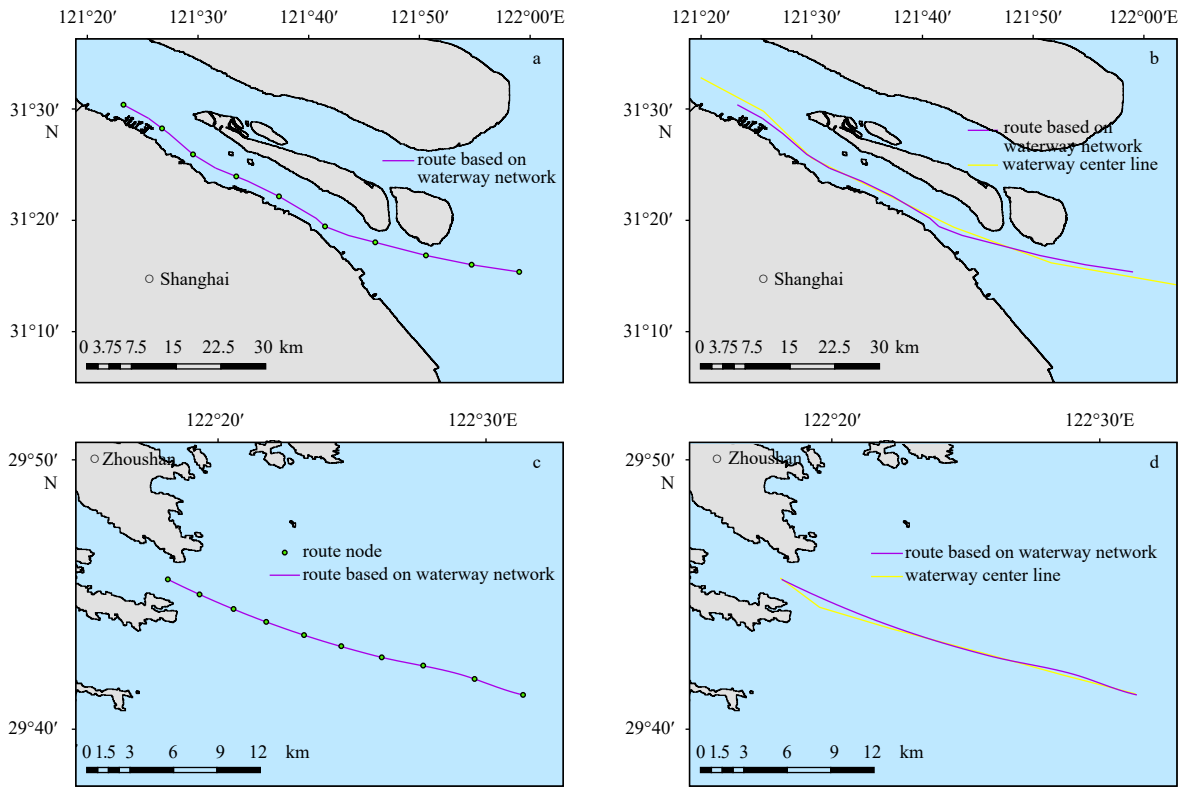


Fig. 7. Evaluation of the generated ship planning routes: the waterway centerline for Area 1 using the waterway extraction method (a); comparison of the waterway centerline in electronic charts in Area 1 (b); the waterway centerline for Area 2 using the waterway extraction method (c); and comparison of the waterway centerline in electronic charts in Area 2 (d).

Table 2. Distance deviation between the extracted route and the centerline of the waterway in the electronic chart

	Max distance deviation/km	Min distance deviation/km	Sum distance deviation/km	MEA (distance deviation average)/km
Area 1	5.212 7	0.146 4	11.33	1.259 2
Area 2	0.67	0.071	2.057	0.205 7

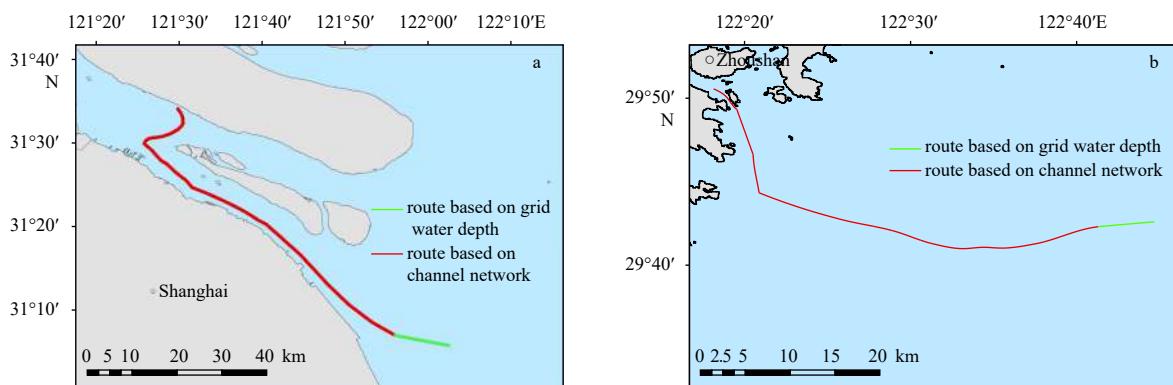


Fig. 8. The results for composite ship planning routes in Area 1 (a) and Area 2 (b).

path based on this. As a result, the shipping routes are derived from trajectory data that contains historical experiences. All tracks are involved in analysis for a certain period of time in a given sea area, rather than based on analysis of a particular shipping path or route.

In existing studies, the automatic generation of routes is usually based on (1) information of electronic charts, (2) clustering of turning points of historical trajectories, (3) the primary fairway of specific sea areas extracted from historical AIS trajectories.

These methods automatically generate ship paths depending mainly on traditional data such as electronic charts or local features such as turning points and lines in navigation.

This paper aims to mine navigation experience from a large amount of historical trajectory data, and to build a waterway network topology based on retaining the geometric characteristics of the fairway. This can be directly used for the automatic generation of ship routes. The proposed method specialized in generating optimal ship planning routes for a timely maritime emer-

gency search and rescue, reducing the execution time for maritime search and rescue services in practical applications. The main benefit of our method is that it takes into full consideration the sufficient empirical navigation information hidden in the massive historical AIS data. Experimental results have been carried out to demonstrate the effectiveness of the method proposed in this paper. It could obtain a high precision ship route that was similar to the waterway centerline in the electronic chart. Corresponding computational errors ranging from 0.2 km to 1 km in different experiments were considered to be acceptable. In maritime applications, the proposed method could provide a tremendous amount of information for maritime emergency search and rescue without the assistance of the widely-used electronic chart.

References

- Altan Y C, Otay E N. 2017. Maritime traffic analysis of the strait of Istanbul based on AIS data. *The Journal of Navigation*, 70(6): 1367–1382, doi: [10.1017/S0373463317000431](https://doi.org/10.1017/S0373463317000431)
- Bryant A, Cios K. 2018. RNN-DBSCAN: a density-based clustering algorithm using reverse nearest neighbor density estimates. *IEEE Transactions on Knowledge and Data Engineering*, 30(6): 1109–1121, doi: [10.1109/TKDE.2017.2787640](https://doi.org/10.1109/TKDE.2017.2787640)
- Campello R J G B, Moulavi D, Zimek A, et al. 2015. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data*, 10(1): 5, doi: [10.1145/2733381](https://doi.org/10.1145/2733381)
- Chen Jinhai, Lu Feng, Peng Guojun. 2015. A quantitative approach for delineating principal fairways of ship passages through a strait. *Ocean Engineering*, 103: 188–197, doi: [10.1016/j.oceaneng.2015.04.077](https://doi.org/10.1016/j.oceaneng.2015.04.077)
- Dijkstra E W. 1959. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1): 269–271, doi: [10.1007/BF01386390](https://doi.org/10.1007/BF01386390)
- Ester M, Kriegel H P, Sander J, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. Portland, Oregon, USA: ACM, 226–231
- Futch V, Allen A. 2019. Search and rescue applications: on the need to improve ocean observing data systems in offshore or remote locations. *Frontiers in Marine Science*, 6: 301, doi: [10.3389/fmars.2019.00301](https://doi.org/10.3389/fmars.2019.00301)
- Garcia M, Viguria A, Ollero A. 2013. Dynamic graph-search algorithm for global path planning in presence of hazardous weather. *Journal of Intelligent & Robotic Systems*, 69(1-4): 285–295, doi: [10.1007/s10846-012-9704-7](https://doi.org/10.1007/s10846-012-9704-7)
- Kumar K M, Reddy A R M. 2016. A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method. *Pattern Recognition*, 58: 39–48, doi: [10.1016/j.patcog.2016.03.008](https://doi.org/10.1016/j.patcog.2016.03.008)
- Le Tixerant M, Le Guyader D, Gourmelon F, et al. 2018. How can automatic identification system (AIS) data be used for maritime spatial planning?. *Ocean & Coastal Management*, 166: 18–30, doi: [10.31230/osf.io/4sfcd](https://doi.org/10.31230/osf.io/4sfcd)
- Li Huanhuan, Liu Jingxian, Wu Kefeng, et al. 2018. Spatio-temporal vessel trajectory clustering based on data mapping and density. *IEEE Access*, 6: 58939–58954, doi: [10.1109/ACCESS.2018.2866364](https://doi.org/10.1109/ACCESS.2018.2866364)
- Li Yuanhui, Pan Mingyang, Wu Xian. 2007. Automatic creating algorithm of route based on dynamic grid model. *Journal of Traffic and Transportation Engineering (in Chinese)*, 7(3): 34–39, doi: [10.3321/j.issn:1671-1637.2007.03.008](https://doi.org/10.3321/j.issn:1671-1637.2007.03.008)
- Lyu Jintao, Liu Zhiqiang, Wang Na. 2018. Method for automatic ship routing based on route stack. *Journal of Computer Applications (in Chinese)*, 38(S1): 16–19
- Muñoz P, Rodriguez-Moreno M. 2012. Improving efficiency in any-angle path-planning algorithms. In: *Proceedings of the 6th IEEE International Conference Intelligent Systems*. Sofia, Bulgaria: IEEE, doi: [10.1109/IS.2012.6335138](https://doi.org/10.1109/IS.2012.6335138)
- Nash A, Daniel K, Koenig S, et al. 2007. Theta*: any-angle path planning on grids. In: *Proceedings of the 22nd National Conference on Artificial Intelligence*. Vancouver, British Columbia, Canada: AAAI
- Schubert E, Sander J, Ester M, et al. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, 42(3): 19, doi: [10.1145/3068335](https://doi.org/10.1145/3068335)
- Svanberg M, Santén V, Hörteborn A, et al. 2019. AIS in maritime research. *Marine Policy*, 106: 103520, doi: [10.1016/j.marpol.2019.103520](https://doi.org/10.1016/j.marpol.2019.103520)
- Wang Zhou, Bovik A C, Sheikh H R, et al. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861)
- Wang Zhu, Li Shujun, Zhang Lihua, et al. 2010. A method for automatic routing based on route binary tree. *Geomatics and Information Science of Wuhan University (in Chinese)*, 35(4): 407–410, doi: [10.13203/j.whugis2010.04.015](https://doi.org/10.13203/j.whugis2010.04.015)
- Wang Guiling, Meng Jinlong, Han Yanbo. 2019. Extraction of maritime road networks from large-scale AIS data. *IEEE Access*, 7: 123035–123048, doi: [10.1109/ACCESS.2019.2935794](https://doi.org/10.1109/ACCESS.2019.2935794)
- Wang Tao, Zhang Lihua, Peng Renan, et al. 2016. A method for automatically generating the shortest distance route based on electronic navigational chart considering channel width. *Hydrographic Surveying and Charting (in Chinese)*, 36(3): 29–31, 36, doi: [10.3969/j.issn.1671-3044.2016.03.007](https://doi.org/10.3969/j.issn.1671-3044.2016.03.007)
- Wei Zhaokun, Zhou Kang, Wei Ming, et al. 2016. Ship motion pattern recognition and application based on AIS data. *Journal of Shanghai Maritime University (in Chinese)*, 37(2): 17–22, 71, doi: [10.13340/j.jsmu.2016.02.004](https://doi.org/10.13340/j.jsmu.2016.02.004)
- Yuan Guan, Xia Shixiong, Zhang Lei, et al. 2011. Trajectory clustering algorithm based on structural similarity. *Journal on Communications (in Chinese)*, 32(9): 103–110, doi: [10.3969/j.issn.1000-436X.2011.09.015](https://doi.org/10.3969/j.issn.1000-436X.2011.09.015)
- Zhan F B. 1997. Three fastest shortest path algorithms on real road networks: data structures and procedures. *Journal of Geographic Information and Decision Analysis*, 1(1): 70–82
- Zhang Shukai, Shi Guoyou, Liu Zhengjiang, et al. 2018. Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. *Ocean Engineering*, 155: 240–250, doi: [10.1016/j.oceaneng.2018.02.060](https://doi.org/10.1016/j.oceaneng.2018.02.060)
- Zhang Lihua, Zhu Qing, Liu Yanchun, et al. 2007. A method for automatic routing based on ECDIS. *Journal of Dalian Maritime University (in Chinese)*, 33(3): 109–112, doi: [10.3969/j.issn.1006-7736.2007.03.024](https://doi.org/10.3969/j.issn.1006-7736.2007.03.024)
- Zhang Lihua, Zhu Qing, Zhang Anmin, et al. 2008. An intelligent method for the shortest routing. *Acta Geodaetica et Cartographica Sinica (in Chinese)*, 37(1): 114–120, doi: [10.3321/j.issn:1001-1595.2008.01.020](https://doi.org/10.3321/j.issn:1001-1595.2008.01.020)