Journal of Computers Vol. 33 No. 4, August 2022, pp. 169-180
doi: 10.53106/199115992022083304014

Recognition Model and Simulation of Busy Waters in Fishing Area Based on Density Clustering

Tian Chai[1], Han Xue[1], Xia-Teng Zhou[2], Jian-Zhou Liu[1], Wen-Yi Wang[1*]

[1] College of Navigation, Jimei University, Xiamen 361021 China
199561000048@jmu.edu.cn
[2] Harbin Institute of Technology, Harbin 150000 China

Received 25 October 2021; Revised 15 January 2022; Accepted 15 February 2022

Abstract. In this paper, the characteristics of the potential behavior of ships contained in AIS information are considered, in order to ensure the safety of navigation in merchant ships' fishing areas and reduce the occurrence of collisions between commercial fishing boats. Based on the DBSCAN density clustering algorithm, the identification model of busy water in fishing areas is developed and applied to Minnan fishery in the Taiwan Strait. In addition, a real ship simulation is performed based on real traffic flow data. The results demonstrate that the proposed model can accurately identify the spatial distribution and scale of busy fishing area, adjust the algorithm parameters according to the merchant ship scale, and generate the targeted recommended route decision for the fishing area. Finally, the STELLAR EXPRESS is compared with the original route. The obtained results show that the recommended route provides a high safety in the fishing area, while retaining the ship operating costs. The model results can be used as a reference for maritime security departments in order to divide the fishing area warning water, and the identification method provides a novel approach for the safety supervision of maritime department.

Keywords: navigation in fishing areas, AIS information, collision avoidance, density clustering

1 Introduction

With the rapid development of China's import and export trade, shipping, which is the main transport capacity, has led to a sharp increase in the number of coastal ships. In recent years, the use of large ships emerged, and the marine navigation environment became busy and complex. Therefore, the navigation safety of commercial ships cannot be guaranteed. For instance, the navigation complexity is largely reflected in encountering commercial fishing vessels. Some commercial ship drivers have not received trainings related to the avoidance of fishing vessels, are not familiar with the operation rules and navigation characteristics of fishing vessels, lack a systematic understanding of the water characteristics in the fishing area, and rush through the fishing area. Therefore, they cannot accurately respond in case of emergency, which results in collision accidents.

The most difficult part of clustering is not in principle or implementation, but in how to select appropriate parameters to obtain the optimal classification, stable and meaningful classification results. Among them, stable classification is the premise. The results obtained by several clustering can not be very different, so the results will not be credible. Meaningful classification means that the classified data has some practical significance in reality. Therefore, clustering is only a method. The final question to be answered is what are the characteristics and significance of the obtained classes. Consistency clustering determines whether the clustering parameters are appropriate by changing the clustering data set. All the data in it are extracted from the original data, which can also be understood as a subset of the original data considering the consistency of the clustering performance of any data in different samples.

The technical contribution of this work is summarized as follows:

(1) DBSCAN density clustering algorithm is improved considering the characteristics of the potential behavior of ships contained in AIS information. The parameters reflect the operating range radius of the fishing boats. The water area where three fishing vessels intensively operate, is considered as the vigilance water area. The commissioning parameters are those obtained when another fishing vessel appears in the area with a radius R of around one fishing vessel. The area is considered as the vigilance water area, which is captured by the algorithm.

(2) This algorithm is used to recommend new routes. The original route of the ship passed through the busy waters of the fishing area. During the non fishing moratorium period, the centralized operation of fishing vessels affects the navigation safety of ships. The cargo ship is large, the remaining speed is difficult to control, the opera-

---

* Corresponding Author

tion is inconvenient, and it is difficult to avoid the fishing boat at close range.

(3) The identification model of busy water in fishing areas is developed and applied to Minnan fishery in the Taiwan Strait. The proposed model can accurately identify the spatial distribution and scale of busy fishing area, adjust the algorithm parameters according to the merchant ship scale, and generate the targeted recommended route decision for the fishing area.

The rest of this paper is organized as follows: Section 2 reveals related work. Section 3 describes our algorithm for identification and modeling of busy water. Simulation results are reported in Section 4. Finally, Section 5 concludes the paper.

2 Related Works

There are mainly three kinds of clustering algorithms: clustering method based on partition, hierarchical clustering method, density based clustering method. K-means is the most basic clustering algorithm. It is very simple, that is, first specify k points, and divide different sample points into the set of the nearest point, so as to divide all samples into k categories. By classifying the sea state and ship operation state into multiple models, K-means clustering was used to analyze the influence of sea state and ship operation state [1]. The privacy protection in the process of K-means clustering and K-Nearest Neighbor classification was studied [2]. The clustering accuracy of K-means from the perspective of outlier removal and distance measurement was improved [3]. To prevent the K-means clustering algorithm from relying on the initial clustering center, particle swarm optimization algorithm was used to optimize the K-means clustering center [4]. A practical protocol for K-means clustering in a cooperative manner was proposed [5]. A method bootstrap mean value where blocks were generated by substitution in the data set was proposed [6]. Considering the estimation of the average of random variables, if enough blocks were generated, the bootstrap median of the average had a better collapse point than the median of the average. The algorithm performed Lloyd type iteration and used bootstrap mean and median strategy. The integrated K-means Laplacian with 12 different kernel functions was improved to form paired similarity matrices [7]. Their effects on the performance of existing algorithms and proposed algorithms were studied.

In the studies related to fishery resources in China, some researchers use the method of cluster analysis to study the temporal and spatial characteristics of fishery activities, and the impact of the fishing operations on water areas, based on the vessel monitoring system (VMS) [8]. In the marine transportation field, the main purpose is to cluster the ship trajectories, in order to guard against the abnormal behavior of ships by obtaining their main motion trajectories. In navigation production, with the development of modern navigation aids, the spatio-temporal data generated by ships in the process of operation starts to be recorded and collected. For instance, the automatic identification system (AIS) is the main source of ship dynamic information data. These large, rich and real-time AIS data are widely used in ship navigation law extraction, potential behavior feature mining etc [9].

Due to the difficulty of navigation in merchant shipping fishing areas, combined with the theory of marine traffic engineering [10] and based on the ship AIS data, this paper develops the identification model of busy water in fishing areas using the density clustering algorithm. In addition, it identifies and judges the fishing areas that highly affect the navigation of merchant ships, and determines the spatial distribution and scale of busy water. Finally, it provides decision-making suggestions on route design for merchant ship pilots.

DBSCAN is a clustering algorithm commonly used in the data mining field. Based on the density of the dataset, the algorithm determines the division of categories by the close connection degree of sample data distribution. That is, there should be sample data belonging to the same category, near any sample of a category. The algorithm divides the samples with tight characteristics into one class, so that it calculates a cluster category referred to as “cluster”. The algorithm obtains all the final clustering results, by dividing each group of closely related sample data into different category attributes.

The ship position data is extracted from the received AIS information. The data have good connectivity, large amount and density. The dataset can also be divided according to the region and time [11]. Based on the characteristics of this data, combined with the advantages of regional density connectivity of the DBSCAN algorithm, the running algorithm can better perform the clustering of fishing vessel active ship positions with high density. The clustering results can also be controlled by adjusting the parameters related to the density. This has a high applicability when considering the ship size as the classification criteria of navigation studies in fishing areas.

3 Identification and Modeling of Busy Water

3.1 Solution of Coordinate Point Spacing between Two Ship Positions

The WGS-84 geographic coordinate system is usually used in navigation. It considers the geocenter as the origin of the coordinate system. In the maritime navigation process, the longitude and latitude of the ship's position point are converted from its own three-dimensional coordinates and the coordinates between the fixed-point oriented ellipsoid. Therefore, the distance between the spherical coordinate points should be considered, when using a computer to process the AIS ship position data.

This paper uses the classic haversine formula, which considers the sinusoidal function as the operation core. Even if the distance between two ships is very small, the calculation can remain accurate. It is expressed as:

$$haver\sin\left(\frac{d}{R}\right)=haver\sin\left(\varphi_2-\varphi_1\right)+\cos\left(\varphi_1\right)\cos\left(\varphi_2\right)A, \tag{1}$$

$$A=haver\sin\left(\Delta\lambda\right), \tag{2}$$

$$haver\sin\left(\theta\right)=\sin^2\left(\frac{\theta}{2}\right)=\frac{1-\cos\left(\theta\right)}{2}, \tag{3}$$

where R is the radius of the earth (6371 km), φ_2 and φ_1 are the latitudes of the two ships, and $\Delta\lambda$ is the longitude difference.

3.2 AIS Data Processing and Optimization

The extracted ship AIS data are decoded and stored, while the fishing vessel AIS data including the MMSI, time, ship position, ship speed and heading are selected. After preprocessing, clear errors and data that do not conform to the scientific theory of navigation, are eliminated. For instance, the MMSI of a ship is 0, the latitude exceeds 90° or the longitude is greater than 180°. In this paper, the method proposed in [11] is used to denoise the AIS dynamic data of abnormal fishing vessels.

In recent years, China has strengthened the supervision of fishing vessels and began to promote the Beidou VMS navigation aid system in coastal fishing vessels, in order to ensure the operation safety of the fishing vessels. However, according to the current data, the number of AIS equipment still far exceeds the number of Beidou terminals [12]. Due to the fact that the fishery supervision system is not perfect enough, a considerable number of fishing vessels is usually close the AIS terminal in order to ensure that they are the first to find fish during production and operation. This results in a certain difference in the amount of information fed back by the equipment. Therefore, it is necessary to quantify the extracted fishing vessel AIS data, in order to reduce its impact on the overall data characteristics.

Based on China's fishery Statistical Yearbook in 2014, the number of motorized fishing vessels that are engaged in fishing operations near the coastline, is almost 283000. However, the total number of coastal fishing vessels that are efficiently equipped with guidance AIDS, recorded in Beidou database in 2014, is almost 86411, which accounts for almost 30.5% of all the marine motorized fishing vessels [12]. It can be deduced that in the following sample data time around 2016, almost 1 / 3 of the total number of marine operation fishing vessels, is installed with navigation aids and standardized registration information.

In summary, in this study, based on the data ratio of 1 / 3, the AIS data accuracy is considered as 1 / 3 of the actual accuracy. That is, assuming that there are almost two ships that do not open AIS, or that are not equipped with AIS according to the rules around the AIS position point of one captured fishing boat, the unit area is treated as three fishing boats. This analysis has a strong supporting significance for the parameter value 3 of the DBSCAN algorithm.

3.3 Pseudocode of the Clustering Algorithm

Algorithm 1. Pseudocode of the clustering

```
Input: fishing vessel AIS dataset in space-time range;
Output: distribution of busy water in this space time
```

```
for a size W of sea area, for time interval T
do
  Extract AIS dataset $D=(x_1,x_2,\dots x_m)$
  Perform data processing to obtain the moving position data of the target fishing boat;
  Calculate and measure the distance ( $Dis$ ) between the target active ship position points;
  perform the determination algorithm neighborhood parameters $(\in, MinPts)$
  Start output: cluster division of busy water $C$
  1) Initialize the core fishing boat set $\Omega=\theta$ , number of clusters $K=0$ , and initial busy area cluster $C=\theta$ of the fishing boat set $\Gamma=0$ that is not accessed.
  2) For the fishing vessel sequence $j=1,2,\dots m$ , perform the following steps to find all the core fishing vessel objects:
  a) Continue to use distance measurement in order to find the $\in-$ neighborhood subset $N_{\in}(x_j)$ of fishing vessels $x_j$ ;
  b) In the operation process, as long as the number of sub fishing vessel sets $x_j$ is judged to be satisfied $|N_{\in}(x_j)| \geq MinPts$ , the fishing vessels $x_j$ can be added to the core fishing vessel set $\Omega=\Omega\cup\{x_j\}$
  3) During the operation, the algorithm is ended once the core fishing boat set is satisfied $\Omega=\theta$ . Otherwise, it continues to perform step 4;
  4) In the core fishing vessel set $\Omega$ , a fishing vessel object $o$ is randomly selected. In addition, the core fishing vessel queue $\Omega_{cur}=\{o\}$ of the current busy water area cluster $k=k+1$ , the category serial number of the busy water area cluster $C_k=\{o\}$ , and the busy water area cluster set, are initialized. The unreachable fishing vessel set is finally updated as $\Gamma=\Gamma-\{o\}$ .
  5) If the core fishing boat object meets the queue $\Omega_{cur}=\theta$ in the current busy water area cluster, the current busy water area cluster $C_k$ is generated. The division $C=\{C_1,C_2,\dots C_k\}$ of the cluster and the core fishing boat set $\Omega=\Omega-C_k$ are then updated. Afterwards, the algorithm turns to step 3 for judgment;
  6) The current busy water area cluster is considered as the target cluster. A fishing boat object is taken from its core fishing boat object $o'$ queue $\Omega_{cur}$ . All the $\in$ neighborhood sub fishing boat sets $N_{\in}(o')$ are determined through parameters $\in$ . $\Delta=N_{\in}(o')\cap\Gamma$ is set. The busy water area cluster $C_k=C_k\cup\Delta$ , the unreached fishing boat set $\Gamma=\Gamma-\Delta$ , and the core fishing boat object queue $\Omega_{cur}=\Omega_{cur}\cup(\Delta\cap\Omega)-o'$ are updated. The algorithm then turns to step 5 for verification;
  7) Output: busy water $C=\{C_1,C_2,\dots C_k\}$
  end while
end for.
```

3.4 Algorithm Parameter Adjustment

The definitions of the core parameters of the DBSCAN algorithm are as follows: $\in$ denotes the neighborhood distance threshold of a sample, and $Minpts$ represents the threshold of the number of samples in the neighborhood. Assuming that there are $Minpts$ data in the dataset and the distance between other sample data and the current target sample data is less than or equal to $\in$, it is considered that the current sample is the core sample captured by the algorithm. In addition, the other sample data are determined as the neighbors of the core sample data that continue to be captured by the algorithm. A cluster in the execution result is composed of several core sample data. All its neighbors are recursively determined by the core sample data. The set composed of all the neighbors is the final sample set of the study [13], as shown in Fig. 1.

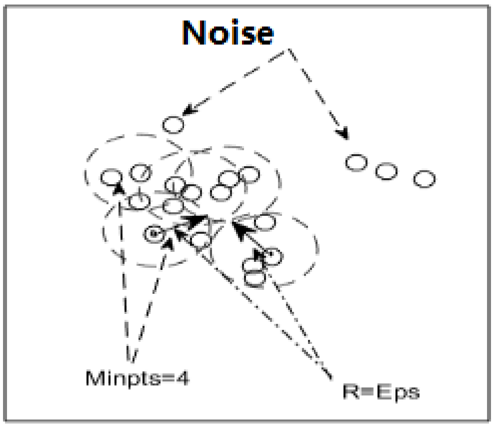


Fig. 1. DBSCAN algorithm parameter adjustment

In this study, the core parameters of the algorithm are provided the practical significance of navigation production. They are then adjusted and analyzed. The specific division is given by:

(1) The parameters $\in$ reflect the operating range radius of the fishing boats.

(2) Based on the previously conducted discussion on the accuracy optimization of AIS data, the water area where three fishing vessels intensively operate, is considered as the vigilance water area. The commissioning parameters $[Minpts = 3]$ are those obtained when another fishing vessel appears in the area with a radius R of around one fishing vessel. The area is considered as the vigilance water area, which is captured by the algorithm. The algorithm ends when the vigilance water area is not established.

(3) The accuracy of the algorithm coordinate system is considered as 0.001°latitude. The nautical calculations are usually given by:

$$1'=1nm = 1852m, \tag{4}$$

$$0.001^{\circ} = 0.06' = 0.06 * 1852 = 111m, \tag{5}$$

$$R=\frac{\in}{0.001} * 111. \tag{6}$$

It can be deduced from this analysis that the larger the parameter value $\in$, the larger the alert operation radius R, and the larger the spatial range of the busy water area identified by the final algorithm. This value can reflect the scope demand of commercial ships to avoid in the fishing area, and their own ship size restrictions. Therefore, the alert operation radius can be used as a reference factor, in order to make targeted decision-making on merchant ships of different scales crossing the fishing area.

4 Simulations

4.1 Determination of the Studied Water

More than 80% of the collision locations of commercial fishing vessels in the coastal water of Fujian and

Guangdong, are located in water farther than 30 nautical miles from the coast [14]. The collision points are scattered, while there is no rule to follow. Therefore, it is impossible to visually alert the range of risk water of different levels, in order to attract the attention of commercial ship drivers. Therefore, it is difficult for merchant ship drivers to find the reference materials for route design in this water area. To deal with this situation, this paper locates in the typical fishing areas of these sea areas, and makes a more accurate discrimination for the busy water in the fishing area, using the model identification method. The safety and economic avoidance of fishing boats are achieved by providing the merchant ship drivers a decision-making reference in route design.

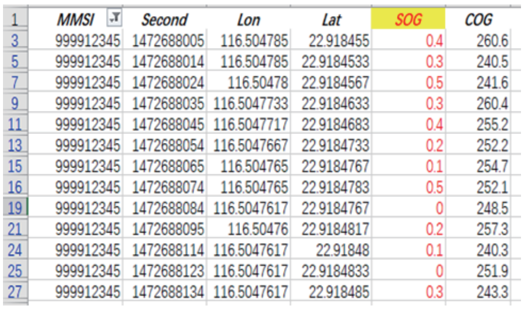
Based on the design concept, the support of relevant material data and the profundity of the significance of the study, the Minnan fishing ground located in the coastal water of Fujian and Guangdong is considered, and the location diagram of the studied water area is drawn. The specific location of the water area has a latitude of 22 ° 47.691 ′ ～ 22 ° 59.82 ′ n, and a longitude of 116 ° 30.19 ′ ～ 116 ° 45.02 ′ E, as shown in Fig. 2.

Fig. 2. Location of the studied water area

4.2 Determination of the Data

Considering that the broadcasting and receiving distance of the AIS equipment is limited, the water close to the coastal base station is preferred, that is, the water on the west side of the Taiwan Strait is selected for fishing vessel data sampling. In this paper, the fishing vessels affecting the passage of merchant ships near the middle route of the Taiwan Strait, are analyzed. In addition, the individual dredging is performed for a long time. Small ships with special operation types are also included in the analysis, in order to further reduce the impact of insufficient data. In terms of sample data time nodes, the seasonal regularity of the fishing operations of fishing vessels is followed, and the data of one month in the fishing period is taken to ensure the data capacity and adaptability of the model.

(1) The AIS data of fishing vessels in the target water from September 1 to October 1, 2016, are retrieved from the AIS repository, with a total of 35050 pieces from 57 ships. After preprocessing, a group of noise data is eliminated, as shown in Fig. 3.



1	MMSI	Second	Lon	Lat	SOG	COG
3	999912345	1472688005	116.504785	22.918455	0.4	260.6
5	999912345	1472688014	116.504785	22.9184533	0.3	240.5
7	999912345	1472688024	116.50478	22.9184567	0.5	241.6
9	999912345	1472688035	116.5047733	22.9184633	0.3	260.4
11	999912345	1472688045	116.5047717	22.9184683	0.4	255.2
13	999912345	1472688054	116.5047667	22.9184733	0.2	252.2
15	999912345	1472688065	116.504765	22.9184767	0.1	254.7
16	999912345	1472688074	116.504765	22.9184783	0.5	252.1
19	999912345	1472688084	116.5047617	22.9184767	0	248.5
21	999912345	1472688095	116.50476	22.9184817	0.2	257.3
24	999912345	1472688114	116.5047617	22.91848	0.1	240.3
25	999912345	1472688123	116.5047617	22.9184833	0	251.9
27	999912345	1472688134	116.5047617	22.918485	0.3	243.3

Fig. 3. Schematic diagram of noise data

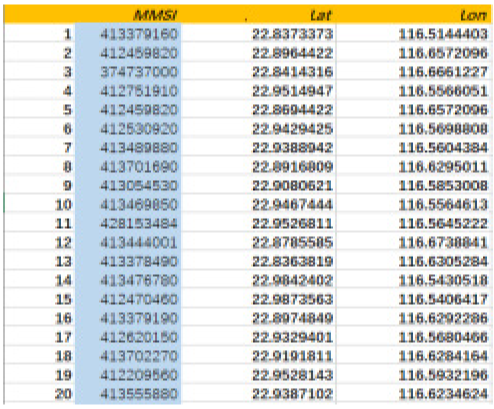
(2) Computation expression of activity probability of a single fishing vessel.

The activity latitude of a single fishing vessel is expressed as:

$$\overline{Lat} = \frac{\sum_{i=1}^{n} Lat_i}{n} \tag{6}$$

$$\overline{Lon} = \frac{\sum_{i=1}^{n} Lon_i}{n} \tag{7}$$

Operation result diagram of fishing vessel activity probability is shown in Fig. 4.



	MMSI	Lat	Lon
1	413379160	22.8373373	116.5144403
2	412459820	22.8964422	116.6572096
3	374737000	22.8414316	116.6661227
4	412751910	22.9514947	116.5566051
5	412459820	22.8694422	116.6572096
6	412530920	22.9429425	116.5698808
7	413489880	22.9388942	116.5604384
8	413701690	22.8916809	116.6295011
9	413054530	22.9080621	116.5853008
10	413469850	22.9467444	116.5564613
11	428153484	22.9526811	116.5645222
12	413444001	22.8785585	116.6738841
13	413378490	22.8363819	116.6305284
14	413476780	22.9842402	116.5430518
15	412470460	22.9873563	116.5406417
16	413379190	22.8974849	116.6292286
17	412620150	22.9329401	116.5680466
18	413702270	22.9191811	116.6284164
19	412209560	22.9528143	116.5932196
20	413555880	22.9387102	116.6234624

Fig. 4. Operation result diagram of fishing vessel activity probability

(3) The visual overview of fishing boat activities is presented in Fig. 5.

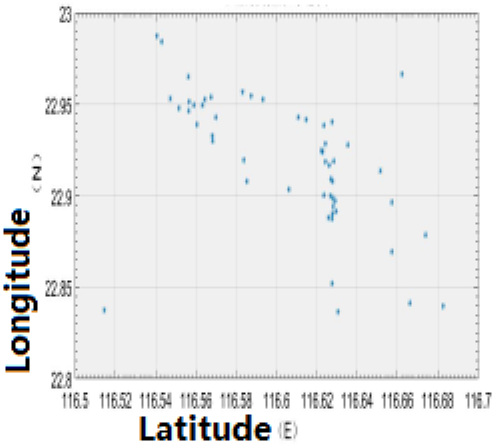


Fig. 5. Schematic diagram of fishing vessel activity

The ship position visualization results demonstrate that the fishing boat operation area is very extensive and there is no obvious law to follow. In addition, the driver cannot intuitively judge the ship operation and avoidance time when passing through the water area. An improper avoidance behavior may lead to an incomplete or impossible avoidance [15]. Due to the fact that the accurate operation of avoiding fishing vessels requires a very high ship operation experience of commercial ship drivers, when commercial ships encounter a group of fishing vessels with intensive fishing operations during navigation in the fishing area, it is recommended to consider them as busy waters and take measures to bypass them in order to avoid direct crossing.

4.3 Model Result Display

In order to analyze the decision-making scheme produced by the model, a hypothetical merchant ship decision-making point is set near the middle route of the Taiwan Strait (22.86 ° n, 116.6 ° E). This is defined as a merchant ship going north along the traditional route of the middle route.

Based on the merchant ship decision point, different algorithm parameters are set for different ship sizes, while different algorithm decision sequences are performed. The corresponding values of the specific parameters are presented in Table 1.

Table 1. Parameters of the merchant ship scale corresponding algorithm

Length（m）	L ∈ (0,90]	L ∈ (90,200]	L > 200
∈	0.005	0.01	0.018
Minpts	3	3	3
R	500	1000	2000

The results of the algorithm are presented in Fig. 6 to Fig. 8, where the black “O” and red “+” denote areas 1 and 2, respectively.

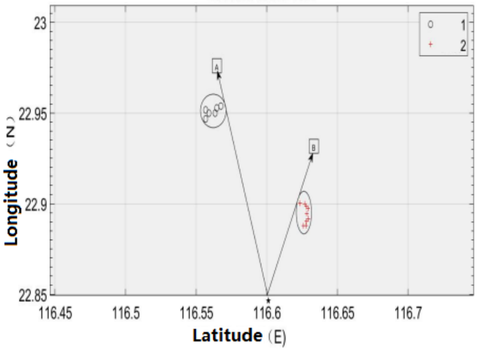


Fig. 6. Execution effect of decision I

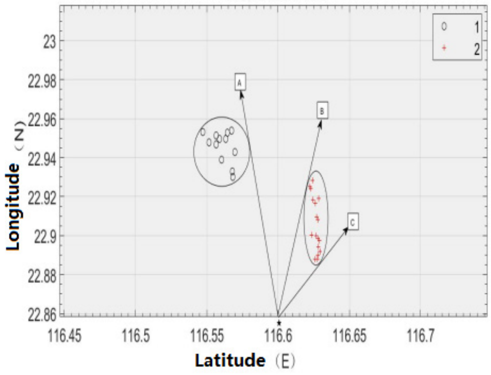


Fig. 7. Execution effect of decision II

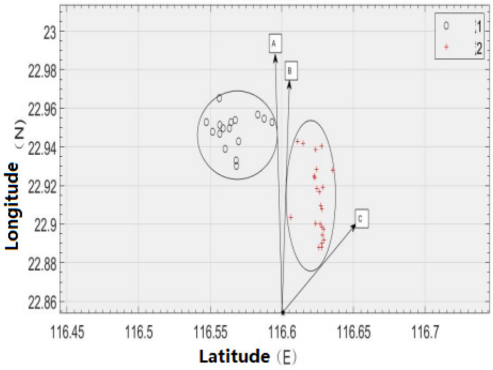


Fig. 8. Execution effect diagram of decision III

It can be seen from the renderings that the ship size parameters indirectly reflect the amplitude demand and self-limitation of the merchant ships when avoiding in the fishing area. For small merchant ships with flexible mobility, the distribution of busy water to be avoided is small. As for large merchant ships where it is difficult to control the peed, the distribution of busy water to be avoided is relatively large. These results accurately reflect the pertinence and adaptability of the decision-making model.

4.4 Real Ship Application and Verification

Based on the theoretical navigation decision, a real ship is selected for the simulation verification of the model. The "stellar express" ship is captured at 1200 hours on April 3, 2020. The detailed information of the ship is consulted, as shown in Table 2.

Table 2. The "STELLAR EXPRESS" ship

MMSI	353787000	Length	209 m
Call sign	3EGZ8	width	32 m
State	Sailing	Draught	9.8 m

(1) It can be seen from the AIS information table that the ship is large. It has a length of 209 meters. It is then more appropriate to perform three iterations of the decision-making algorithm. In order to strengthen the contrast, the route in the Strait is embedded into the model for display. The implementation effect is illustrated in Fig. 9.

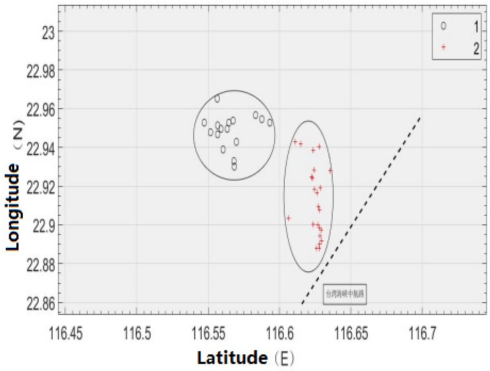


Fig. 9. Execution effect diagram of decision III

(2) Based on the execution results of the decision III algorithm and the basic judgment made in combination with the navigation experience, the route of the large cargo ship tends to the offshore area. Therefore, it is only necessary to present area 2 (referred to as No. 4 busy water area) in the identification result on the chart, as well as the navigation track points of the ship from 1200 to 1306, as shown in Fig. 10.

Fig. 10. The “STELLAR EXPRESS” ship track

It can be seen from Fig. 10 that the original route of the ship crosses the busy water of the fishing area. During the non fishing moratorium, the centralized operation of the fishing vessels affects the navigation safety of the ship. The cargo ship is large, the residual speed is difficult to control, the manipulation is not convenient, and some difficulties in avoiding fishing vessels at a close range exist. Therefore, in order to ensure the safety of the ship operation, the recommendation decision is given based on the original planned route of the ship:

(1) The pilot of this ship steered the ship at 1226: new CA1 = 070.76, with a range of almost 3.5 Nm to the next turning point H4.

(2) After reaching the H4 turning point, the pilot steered the ship again: new Ca2 = 053.71, with a range of almost 2 nm, and then resumed the original route.

Whole schematic diagram of the recommended route decision is shown in Fig. 11.

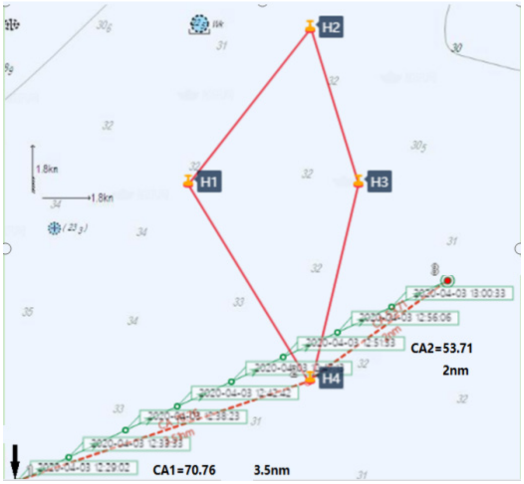


Fig. 11. Whole schematic diagram of the recommended route decision

4.5 Evaluation and Extension of the Model Results

It can be seen by the comprehensive comparison with the original route of “stellar express” that, when taking the recommended route decision, the ship’s deviation distance is not large. However, it can efficiently avoid the busy water in the fishing area, while the avoidance approach is still the most used diversion operation of commercial ship drivers [10].

This example is a decision-making scheme generated by identifying a water area near the middle route. Compared with the traditional route, it greatly ensures the safety of merchant ships sailing in the fishing area, when the operation cost is almost retained. This model is used to identify multiple continuous busy water, and set decision points in the fishing area that highly affects the merchant ships navigation. The different decision-making schemes $\left[T_1, T_2, \cdots, T_n\right]$ generated by the set of decision points $\left[P_1, P_2, \cdots, P_n\right]$, are then associated and integrated, which can be modified according to the year, fishing season and merchant ship scale, in order to form a complete set of decision-making schemes for recommended routes in fishing areas.

The limitation of this algorithm is to rely on AIS. AIS has the problems of insufficient positioning accuracy and multi ship common MMSI, which may affect the accuracy of this algorithm.

4.6 Comparison of Different Algorithms

There are many evaluation indexes of clustering effect, such as accuracy, recall, mean absolute error (MAE), etc. This paper uses MAE to evaluate this experiment. MAE can reflect the difference between the classification of the test set and the real classification. Table 3 lists best cost comparison of different algorithms.

Table 3. Best cost comparison of different algorithms

Bay	Algorithms	Best	Worst	Average
Nanmen	DBSCAN	1.1875	1.2360	1.1904
	Ours	0.9546	0.9962	0.9745
Maluan	DBSCAN	1.3450	1.3907	1.3670
	Ours	1.1166	1.1570	1.1296
Jinluan	DBSCAN	1.2307	1.3029	1.2903
	Ours	0.7906	0.8236	0.8034

The results shows that this algorithm has smaller time complexity than DBSCAN. Compared with DBSCAN, the new algorithm needs additional operation. However, the new algorithm convergences faster and can reach the optimal solution in less searching iteration times. Therefore, it still takes less time to obtain the results.

4.7 Discussion

It is illustrated from above figures that encountering positions in Dongshan Bay waters are mainly distributed in Dongshan Bay estuary waters, including Tongling nearshore waters on the northeast side of Dongshan Island, Chengan operation area waters, Dongshan port area and west waterway of Tasmania island. The channel on the west side of Gulei Peninsula and many berths along the shore are also areas where ships may encounter. Dongshan Bay and its adjacent waters mainly have the following ship traffic flows: The ship traffic flow in and out of Tongling water area through the west waterway of Tayu and the south channel of Dongshan port area, the ship traffic flow in and out of the channel of Dongshan port area, the ship traffic flow in and out of the Gulei port area along the Gulei channel, and the ship traffic flow from Dongding island to Dongshan Island along the recommended internal route of Fujian.

Dongshan is the main fishing port of Zhangzhou. Fishing boats often carry out fishing operations in Dongshan Bay and its nearby waters. There is a risk of collision between merchant ships, fishing boats and yachts. There are many aquaculture areas in Dongshan Bay and nearby waters, affecting the navigation safety of ships entering and leaving the port. Meanwhile, the waters around Dongshan Bay are close to the Taiwan Strait. When there are strong winds and waves in the open sea, small ships sail near the shore, resulting in high ship traffic density in the waters around Dongshan Island and increasing the risk of ship encountering.

5 Conclusions

Based on the AIS data of fishing vessels in 2016, this paper develops a busy water area identification model using the DBSCAN clustering algorithm as the core. This model is applied to the water of the Minnan fishing ground in the Taiwan Strait. The fishing area between the fishing ground and the route in the Strait is identified and judged from September to October 2016, based on the simulation of real traffic flow. The water mainly has two busy operation areas, while the central zone is located in area 1 (22 ° 55 ′ N ~ 22 ° 58 ′ n, 116 ° 32 ′ e ~ 116 ° 36 ′ E) and area 2 (22 ° 53 ′ N ~ 22 ° 56.5 ′ n, 116 ° 367 ′ e ~ 116 ° 38 ′ E).

Based on the pilot’s experience in ship operation through the fishing area, the micro avoidance behavior of ship is raised to the macro avoidance decision of ship to region. Using the proposed modeling method, the distribution and scale of busy water in the fishing area are accurately identified, and targeted navigation recommendation decisions in the fishing area are determined for commercial ships of different scales. However, in this study, it is difficult to obtain the real-time AIS data of fishing vessels. In future work, we expect to focus on embedding the influencing factors, such as the meteorological and sea conditions, in order to improve the recognition accuracy of the model.

With the expansion of the scale of ship trajectory data, the time consumption of the algorithm will increase sharply. Therefore, how to optimize the algorithm and effectively reduce the clustering overhead of the algorithm is the later research direction.

6 Acknowledgement

The work is supported by Natural Science Foundation of Fujian Province (No. 2021J01819).

References

[1] J. Lee, M. Roh, K. Kim, Prediction of ship power based on variation in deep feed-forward neural network, International Journal of Naval Architecture and Ocean Engineering 13(2021) 641-649.
[2] D. Zhao, X. Hu, S. Xiong, J. Tian, J. Xiang, J. Zhou, H. Li, K-means clustering and kNN classification based on negative databases, Applied Soft Computing 110(2021) 107732.
[3] N.H.M.M. Shrifan, M.F. Akbar, N.A.M. Isa, An adaptive outlier removal aided k-means clustering algorithm, Journal of King Saud University- Computer and Information Sciences (2021).
[4] Y. Li, X. Chu, D. Tian, J. Feng, W. Mu, Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm, Applied Soft Computing 113(B)(2021) 107924.
[5] E. Zhang, H. Li, Y. Huang, S. Hong, L. Zhao, C. Ji, Practical multi-party private collaborative k-means clustering, Neurocomputing 467(2022) 256-265.
[6] C. Brunet-Saumard, E. Genetay, A. Saumard, K-bMOM: A robust Lloyd-type clustering algorithm based on bootstrap median-of-means, Computational Statistics & Data Analysis 167(2022) 107370.
[7] S. Rengasamy, P. Murugesan, K-means – Laplacian clustering revisited, Engineering Applications of Artificial Intelligence 107(2022) 104535.
[8] Y. Wang, The analysis on trajectory simulation of single otter trawl vessels and its production estimates as well as spatio-temporal characteristics in Zhoushan fishing ground in autumn and winter, [dissertation] Zhoushan: Zhejiang Ocean University, 2014.
[9] X. Xiao, Z. Shao, J. Pan, X. Ji, Ship trajectory clustering model based on AIS information and its application, China navigation 38(2)(2015) 82-86.
[10]Z. Wu, J. Zhu, Marine traffic engineering, 2nd Edition, Dalian Maritime University Press, Dalian, 2004.
[11]L. Xu, L. Xiang, T. Chai, Z. Xiong, Modeling and application of channel capacity of Xiamen port, China navigation 42(2)(2019) 32-36.
[12]G. Guo, W. Fan, S. Zhang, Q. Zheng, X. Wang, Research progress on data mining and application of ship position monitoring system, Marine fisheries 38(2)(2016) 217-224.
[13]J. Song, Y. Guo, B. Wang, Research on parameter allocation method of DBSCAN clustering algorithm, Computer technology and development 29(5)(2019) 44-48.
[14]Q. Li, S. Zeng, Study on the collision between merchant ships and fishing vessels, China fisheries economy 30(2)(2012) 5-11.
[15]H. Tang, J. Huang, Main causes and preventive measures of merchant ships colliding with fishing vessels, China navigation (4)(2007) 24-28.