8-27-2021

Data-driven based automatic routing planning for MASS

Qingwu Wang

WORLD MARITIME UNIVERSITY

Malmö, Sweden

DATA-DRIVEN BASED AUTOMATIC ROUTING PLANNING FOR MASS

By

Qingwu WANG

The People's Republic of China

A dissertation submitted to the World Maritime University in partial Fulfillment of the requirements for the award of the degree of

MASTER OF SCIENCE
In
MARITIME AFFAIRS
(MARITIME SAFETY AND ENVIRONENT MANAGEMENT)

2021

DECLARATION

I certify that all the material in this dissertation that is not my own work has been identified, and that no material is included for which a degree has previously been conferred on me.

The contents of this dissertation reflect my own personal views, and are not necessarily endorsed by the University.

Signature: ________________________

Date: ________________________

Supervised by: Assit. prof. Changing BAI
Dalian Maritime University

Assessor:

Co-assessor:

ACKNOWLEDGEMENTS

I am sincerely grateful to the World Maritime University and Dalian Maritime University for offering me this opportunity to study.

I will give my great thanks to assistant professor Chunjiang BAI. Bai's rigorous and realistic scientific research attitude and sincere care for students will benefit me for life. I sincerely thank my mentor for the kindness. I have carefully cultivated. Here, I just want to say solemnly, thank you!

Then, I pleased to acknowledge the director of MSEM, Jian ZHAO for invaluable assistance all through the study life of MSEM. He also makes considerable comments and suggestions for this paper.

Many thanks to Mrs. Lu ZHAO, the teaching assistant of MSEM. She's capable of handling mountains of trifles successfully.

Thanks to my wife for her dedication to the family and for her support and understanding during my studies, and thanks so much for provoking the family for me. Thanks to my father, mother, father-in-law, mother-in-law, and four elderly people for their support and help. I deeply owe the family too much. Thank to my son and my daughter, they were very clever and they gave me the motivation to study.

ABSTRACT

Title of Dissertation: **Data-driven Based Automatic Routing planning for MASS**

Degree: **MSc**

With the continuous development of unmanned technologies such as unmanned vehicles, more and more scientists begin to focus on the research of unmanned ships. As one of the key technologies in the field of unmanned ship's research, route planning plays an important role in the development of unmanned ships. This paper proposes a method for automatic route planning using historical AIS data.

The AIS data is generated by the AIS terminal periodically and automatically according to the navigation status of the ship. The broadcast period ranges from 2 seconds to 6 minutes. AIS base stations can receive thousands of AIS messages every second. In order to speed up the processing ability of AIS data, a trajectory data compression method was proposed based on the Douglas - Peucker algorithm. At the same time, a method for determining the compression threshold for the Douglas-Peucker algorithm is proposed based on the smallest ship domain. This algorithm ensured that the compressed trajectory data retains the original trajectory feature points while eliminating the repetitive and redundant information in the original trajectory data. An automatic route planning method based on trajectory data and data-driven technology is proposed. By improving the DBSCAN algorithm and establishing a similarity metric, clustering analysis of the waypoints in the identified trajectory feature points is performed to identify the waypoints on the route. Compared with the previous processing methods for the track section, it has strong anti-noise ability and high processing efficiency. The algorithm can ensure that the route is consistent with the macro situation of the traffic flow described by the original trajectory data.

KEY WORDS: unmanned ships, MASS, AIS, route planning, data-given

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIS	Automatic Identification System
MASS	Maritime Autonomous Surface Ship
IT	Information Technology
DT	Data Technology
IoT	Internet of Things
UAV	Unmanned Aerial Vehicle
USV	Unmanned Surface Vehicle

Chapter 1 Introduction

1.1 Background

In recent years, the intelligence level for ships has been continuously improved, and new concepts and technologies such as big data, cloud computing, and the Internet of Things have emerged in an endless stream. The realization of unmanned ships has a hardware foundation and technological support. With the continuous development of unmanned technologies such as unmanned road vehicles and unmanned aerial vehicles, and the continuous advancement of research and development technologies for unmanned ships, it is not impossible for unmanned ships to sail on the sea in the future. There are three types of unmanned ships, including fully autonomous, semi-autonomous and non-autonomous. Non-autonomous unmanned ships mainly refer to remote controlled unmanned ships. Semi-autonomous unmanned ships mainly rely on pre-set procedures to perform specific tasks. Fully autonomous unmanned ships are the ultimate goal. They have the functions of auto navigation, environmental perception and adaptive control. The research on unmanned ships includes many aspects such as ship design, route planning and navigation, surface object detection and automatic recognition, automatic collision avoidance and obstacle avoidance, and motion control.(Breithaupt et al., 2017)

This paper focus on unmanned ships, with automatic route planning, technology as the main content of the research.(Andersson & Ivehammar, 2017)

Generally speaking, for route planning, it mainly relies on manual analysis of nautical publications, such as "Ocean Passage for the World"(NP136), "Sailing Directions" (NP1-NP72), "Routing Chart".(Christiansen et al., 2013) The traditional route planning procedure has a certain degree of subjectivity. The quality of the route is subject to the professional level and attitude of the crew members. As a result, in recent years, many experts and scholars begin to research the automatic route planning. At the same time, automatic route planning technology is one of the key technologies in the field of unmanned ship's research, and to a certain extent marks the level of intelligence of unmanned ships.(Cheng et al., 2018)

The traditional automatic route planning technology is mainly based on electronic chart platforms, radar or satellite remote sensing images. It uses pattern recognition, data mining and other methods to identify navigable waters and non-navigable areas, and then use intelligent search algorithms to find out the navigable waters. Specify the shortest route from the start point to the end point. The route generated in this mode is only the best route from the perspective of graph theory. There is still a certain gap between the results of theoretical analysis and practical applications. For example, the route extracted by the intelligent search algorithm cannot intelligently identify the Traffic Separation Schemes (TSS). Some relevant navigation rules also should be taken into consideration.(Breithaupt et al., 2017)

The automatic route planning of unmanned ships refers to the automatic planning of a barrier-free path from the start point to the end point based on the voyage order, in accordance with the principles of safety and economy. At present, the automatic route planning platform is generally based on electronic charts, radar images, and large-scale water photos and thermal images taken from satellites. The algorithms include heuristic search algorithms such as Dijkstra algorithm and A* algorithm, ant colony algorithm, Bionic optimization algorithms such as genetic algorithm, as well as visual method, artificial potential field method, simulated annealing algorithm, and AIS data-driven methods. This article will focus on the data-driven method based on AIS.(Matui et al., 1981)

Today, trajectory data is showing an explosive growth trend. It is very important to propose a data-driven route automatic generation method and improve the existing automatic route planning mode with the research the big data strategy and management analysis mode.

1.2 Research significance

The AIS data is broadcasts periodically by terminals equipped on ships, aids to navigation and so on. In particular, the application of low-orbit satellites equipped with AIS receivers has extended the transmission/reception range of AIS signals to several thousand kilometers, which make it possible to establishing a global AIS data network. The AIS data include static data of ships and dynamic data. Static data include the ship name, tonnage information, arrangement of the entire of the AIS and so on while dynamic data include the voyage data, such as the port of departure and port of destination latitude, longitude and time. With the latitude, longitude and time, it shows the trajectory information on the ship. (Shrivastava & Gupta, 2012) So AIS data have spatial-temporal characteristics. In addition, the AIS information on the same ship is related to each other. The continuous AIS information depicts the trajectory of the ship. The AIS information on multiple ships contains information on maritime traffic flow, such as information on the habitual route of the ship, information on the collision avoidance behavior of the ship.

Since 2003, the China Maritime Safety Administration (MSA) has invested 457 million RMB in stages and surgeons to build China's coastal and inland AIS shore-based network systems. In 2009, the AIS National Center began to be established. In May 2010, the AIS National Center realized the cooperation with IALA. Net networking has formed a global networked trajectory database. The establishment of the trajectory database provides a large amount of usable materials for the automatic route planning method based on trajectory data. With the continuous development of new technologies and concepts such as big data, pattern recognition, and cloud computing, under the background of the rise of data science, trajectory analysis and processing technologies have also made great progress. Through trajectory analysis, it is possible to discover the overall characteristics and laws of traffic flow, identify frequently encountered locations and main encounter situations, identify traffic-intensive areas, and provide a decision-making basis for maritime traffic management. The development of trajectory analysis and processing technology provides strong support for the research of data-driven route automatic generation method.(J. Wang et al., 2020)

This research starts with the study of customary routes as a breakthrough, studies the customary routes of ships in advance based on AIS trajectory data, combines relevant content of route planning, and combines with the ECDIS platform to realize the automatic route planning function of ships. According to the requirements of IMO A.893 (21) on the voyage plan, the evaluation model of the pros and cons of the route is established, which comprehensively considers the width of the track zone, the characteristic parameters of the traffic flow, the ship's own motion parameters, the amount of drift under the action of wind, waves, and the yaw limit. (Zhou et al., 2020) Establish a "route field" model based on the safety and the surplus width of the channel. Then, according to the IMO's requirements for route safety inspection, quantitative evaluations are made on the recommended routes of the navigation box, the routes extracted by AIS, and the routes designed based on ECDIS, and the designs are based on different types. The navigation characteristics of ships of different tonnage classes, excavate and analyze the optimal route of the ship under the specific ship parameter model.

1.3 Research content

The research content of each chapter of the thesis is as follows.

Chapter 1 Introduction. It mainly introduces the research background, research significance and organization structure of the thesis. The thesis is to explore the theory and method of automatic route planning of unmanned boat routes based on data-driven under the background of the rapid development of unmanned technology, the gradual implementation of the national big data strategy.

Chapter 2 Related theories of MASS and route planning. It reviews the development history of MASS (Maritime Autonomous Surface Ships) and introduces the key technologies involved in the research of MASS. Also, it makes a looking forward to the application prospects and development trends of MASS. At last, it summarizes the latest research results of trajectory analysis and processing technology for a data-driven perspective and application in maritime affairs.

Chapter 3 Trajectory data compression and threshold determination based on the Douglas-Peucker algorithm. The trajectory data are so big that they have to be compressed. In this chapter, it introduces the compression algorithm called Douglas-Peucker algorithm. Aiming at the problem of determining the compression threshold, it draws the concept of the ship domain. And, the smallest ship domain is obtained based on the AIS trajectory data and used as the standard for determining the compression threshold.

Chapter 4 Trajectory data clustering and automatic route planning based on DBSCAN algorithm. In this chapter, it identifies the waypoints in the compressed trajectory data, with the DBSCAN algorithm performing the cluster analysis. Using the algorithm, it finds out the optimal route according to the specified starting point and ending point.

Chapter 5 Conclusion and Outlook. It summarizes the main research results and the innovations of the paper, and it also pointed out that the research that needs to be improved, including future research directions.

Chapter 2 Related theories of MASS and route planning

This chapter mainly includes two aspects. The first is to review the development history of MASS. In this part, we introduce the key technologies involved in the research of MASS and the application prospects and development trends of MASS. We also research on automatic route planning, technology, and determine the research method which means the data driven trajectory analysis and processing technology to explore and research on automatic route planning. Secondly, we introduced the data-driven maritime application examples and the prospect of trajectory analysis and processing technology.(Ma et al., 2019)

2.1 Research status of unmanned ships

2.2.1 Development and prospects of MASS

In June 2017 the Maritime Safety Committee of the International Maritime Organization (IMO) made the historic decision at MSC 98 to conduct a regulatory scoping exercise to consider the suitability of extant IMO instruments to remote controlled and autonomous ships.(IMO, 2017)

Fast-forward one year to May 2018, MSC 99 was always going to be an important meeting for Maritime Autonomous Surface Ships (MASS), to use the IMO vernacular, and so it proved. (Veal, Robert, 2018)

The Maritime Safety Administration of Singapore launched an initiative called "MASSPorts" at a web conference held on August 4, 2020, which aims to achieve the consistency of MASS sea trials and operating standards. Eight IMO member states, including China, Denmark, Finland, Japan, the Netherlands, Norway, South Korea and Singapore, responded to the initiative and will jointly form the MASSPorts network. Norway has already played a key role in establishing three autonomous solution testing areas in Norwegian waters and establishing the Norwegian Autonomous Ship Forum. South Korea will invest about 130 million U.S. $ between 2020 and 2025 to develop an autonomous navigation system for MASS and apply it to the actual ships.

In September 2012, the European Commissions invested 3.8 million U.S. $ to fund Fraunhofer CML, Matorka, MarineSoft, MarinTek, Chalmers University and another 8 confirms to carry out the 3-year "MUNIN" (Maritime Unmanned Navigation through Intelligence in Network) project. The MUNIN project aims to propose the concept of an unmanned ship and verify its feasibility. The operation mode of the unmanned ship proposed by MUNIN is shown in Figure 1.3. The ship mainly relies on the autonomous decision-making system on the ship to achieve autonomous navigation. At the same time, all sensor parameters are transmitted back to the shore-based control center via satellite in real time and are controlled by remote operators. The project takes a 200-meter bulk carrier as the research object and aimed to propose the concept of an unmanned ship and verifies its feasibility. As shown in Figure 2.2, the research content of the project includes autonomous navigation and control such as water object detection, automatic engine control, and shore-based remote control.(Jan, 2013)

In February 2013, China’s first self-designed unmanned survey vessel "Hai Xun 166" achieved its first on-site application on the sea in the South China Sea cruise directed by the China Maritime Safety Administration. The "Hai Xun 166" unmanned survey boat has realized the functions of autonomous navigation/remote control navigation, path planning, automatic collision avoidance and obstacle avoidance, and is equipped with a sonar detection system and a doppler flow velocity monitoring system, which can collect some geographical information in complex waters.(Wang et al., 2018)

In 2013, the "Jinghai 1" unmanned boat developed by Shanghai University (SMU) has navigated in the disputed islands and reefs of China and Vietnam in the South China Sea and obtained a lot of valuable data. Accompanied with the "Xuelong" in 2014, "Jinghai 2" carried out some geographical work in the Antarctic area. At present, the “Jinghai” serial unmanned ships have been integrated with the Beidou system, which can realize marine surveying and mapping, maritime patrol. (Bibuli et al., 2008)

In 2017, the idolphin 38,800-ton smart bulk carrier named "Great Intelligence" developed by the China State Shipbuilding Corporation was awarded the smart ship symbol by CCS and LR. She was the first smart ship certified by the classification society and become the Milestones of smart ship in China.(Liao, 2015)

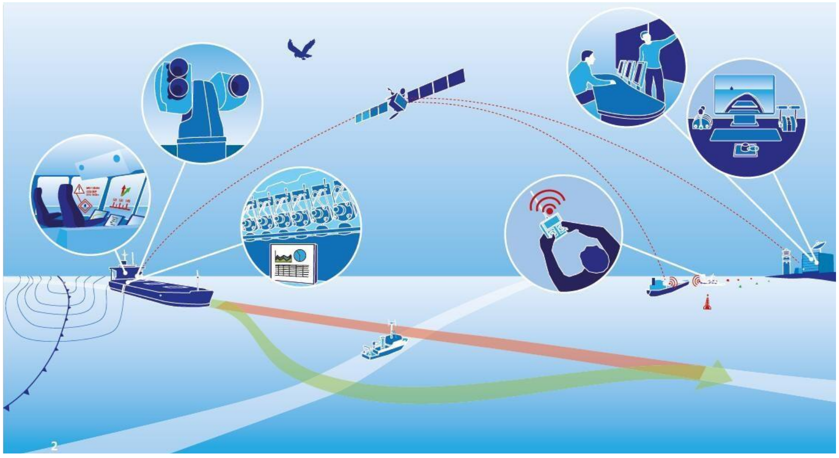


Figure 2.1 the operation mode of the unmanned ship for MUNIN

Source: http://www.unmanned-ship.org/munin

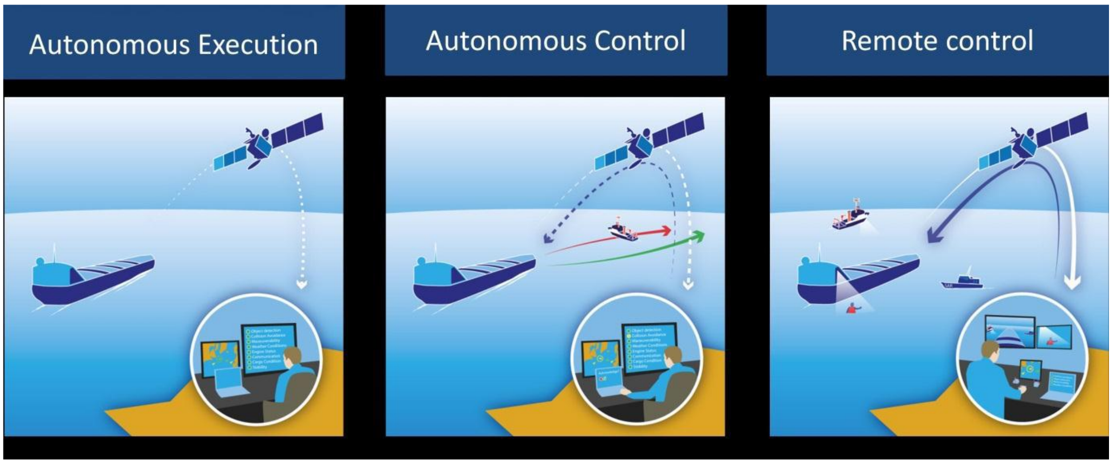


Figure 2.2 “MUNIN” USV Operational Modes
Source: http://www.unmanned-ship.org/munin/

Figure 2.3 Haixun 166
Source: https://www.sohu.com

Figure 2.4 Jinghai

Source: http://www.bjhxhk.cn/

Figure 2.5 Great Intelligece
Source: http://news.takungpao.com/

2.2.2 Key technologies in MASS

Since unmanned ships are required to realize the functions of autonomous path planning, automatic target recognition, and autonomous collision avoidance and obstacle avoidance, the research on unmanned ships involves a wide range of technical fields. Based on the current research status, the current research on the key technologies involved in unmanned ships mainly focuses on the following aspects:(Matui et al., 1981)

(1) Automatic route planning technology

The electronic chart or image is divided into navigable waters and non-navigable waters, and then intelligent search algorithms ,such as Dijkstra algorithm, A* algorithm,(Singh et al., 2018) Artificial Potential Field (APF), and visualization method, are used in the navigable area to search for the shortest path. (Song, 2014) extracted the shape from the electronic chart data and establish an environmental model, and then the Bypass Islands (BI) algorithm is used to find the optimal route from the start point to the end point. (Wu et al., 2015) uses millimeter wave radar and multi-camera stereo vision system to establish a water obstacle map and the underwater obstacle map established by sonar detection is used to establish an environment model after the obstacle is expanded, and then the heuristic search method combining the A* algorithm and the view method is used to find the shortest path. (Ahmed et al., 2015)

The "best route" generated by the above-mentioned research methods are only the best route from the point of view of graph theory. However, they ignored the route rules that the unmanned vessel must comply with in actual navigation, such as a traffic separation scheme (TSS) and the difference between the results of theoretical analysis and practical applications.(Q. W. Wang et al., 2021)

(2) Motion control technology

Because unmanned ships are required to be able to achieve safe and autonomous navigation in various complicated marine environments, which required higher performance of maneuverability and control. Based on the current research status, the motion control problems of the unmanned ships can be divided into two categories: full driver control and underdrive control. Among them, the full drive control mainly aims at the heading and speed control problems, and the under-drive control mainly aims at three aspects: path tracking control, trajectory tracking control and stabilization control.(N. Wang et al., 2021)

a) Full drive control (heading control and speed control)

The heading control and speed control of unmanned ships belong to the research category of full drive control. For the heading control and speed control of unmanned ships, there are many references. The main control methods include sliding mode control, PID control, Linear Quadratic Gaussian (LQG), and Lyapunov direct control, fuzzy control and self-adjusting control.

b) Under drive control (path following control)

The path following control of the unmanned boat requires the unmanned boat to follow a desired path. (Bibuli et al., 2008) takes the "Charlie" unmanned boat as the control object and establishes the basis of Serret. The direct path tracking error motion equation of the Frenet coordinate system reduces the path tracking problem to the speed and heading angular velocity control problem in the inertial coordinate system. (Liao, 2015) also studied the path tracking control of unmanned ships under Serret-Frenet coordinates. (Miskovic et al., 2009) also took the "Charlie" unmanned boat as the control object, and designed the tracking reference heading angular velocity control mode based on the PD controller and the tracking reference heading based on the PDI controller to solve the problem of the straight path tracking of the boat. Angle control mode. (Aguiar & Hespanha, 2004) combined Backstepping technology with logic switch technology for the path tracking control problem of unmanned ships, and designed a self-adjusting supervisory controller under the influence of model parameter uncertainty.(L. P. Wang et al., 2020)

c) Under drive control (trajectory tracking control)

Trajectory tracking control and path tracking control of unmanned ships control the ship to navigate according to the desired trajectory. The difference is that the trajectory tracking control needs to consider time constraints.

d) Under drive control (stabilization control)

The goal of point stabilization control of unmanned ships is to stabilize the unmanned ships to a certain balance point, for example, the automatic berthing and unberthing of unmanned ships. Positioning control applies the principle of stabilization control.

e) Autonomous collision avoidance and obstacle avoidance technology

The autonomous collision avoidance and obstacle avoidance technology of unmanned ships is one of the key technologies to improve the level of autonomy of unmanned ships. At present, the automatic collision avoidance and obstacle avoidance technologies for unmanned ships can be divided into two categories: one is the automatic collision avoidance and obstacle avoidance technology based on path finding, which is mainly used in the hazard avoidance of low-speed unmanned ships. The second is based on the short-range reaction type automatic collision avoidance and obstacle avoidance technology. It mainly used in the complex marine environment for rapid reaction type danger avoidance.(Poikonen & Golden, 2020)

Viewing at the current research status of key technologies of unmanned ships, there are abundant research results in three areas: automatic route planning technology, motion control technology and automatic collision avoidance technology. In order to improve the level of autonomy of unmanned ships, it is necessary to realize the comprehensive improvement and organic combination of all key technologies of unmanned ships.(Lyu & Yin, 2019)

2.2 Trajectory analysis and processing technology based on Data-driven theory

The scale and type of data generated by the convergence of modern information technology and traditional economic society has shown an unprecedented growth trend, and the era of big data has arrived quietly. Data has gradually changed from a simple processing object to a basic strategic resource. How to analyze and use big data has become a common hot issue in modern society. With the continuous development of maritime navigation and communication technology and the continuous improvement of shore-based communication networks, trajectory data has shown an explosive growth trend. At the same time, the data scale and data accuracy have reached unprecedented levels, which provides a large number of data-driven trajectory analysis and processing technologies.(Wu et al., 2015)

In May 2015, Jack Ma, Chairman of the Alibaba Board of Directors, presented at the International Big Data Industry Expo and the Global Big Data Era Guiyang Summit, "China is ushering in the transformation from the IT (Information Technology) era to the DT (Data Technology) era". Jack Ma believes that the transformation of the DT era represented by big data is an improvement of technology in the IT era. The core of the IT era is computers and the Internet, and the core of the DT era is data.

In the era of Wanfang big data, data is no longer just an object in engineering processing. Jim Gray, a well-known database expert and Turing Prize winner, believes that the history of human scientific research has experienced three paradigms including experiment, theory and calculation. In the era of big data, the traditional three scientific research paradigms are in some emerging research fields. It has been unable to function effectively, so a brand-new fourth paradigm is needed to guide current scientific research. The "fourth paradigm" proposed by Jim Gray is a data-exploratory research method. The specific descriptions of the four paradigms are shown in Table 2.1. The "fourth paradigm" actually changes from "computing-centric" to "data-centric".(Meng & Ci, 2013) In short, data-driven is a data-centric processing and application model. The computing-centric era largely stays at the use of past data to explain the past state. The core of big data is to predict and create a quantifiable dimension for human life. The data has changed from the original description of the past state.

Table2.1 Four Science Paradigms

Science Paradigms	Time	Methodology
Empirical	thousands of years ago	descript the natural physical phenomena Briefly
Theoretical	hundreds of years ago	Modeling analysis, inductive reasoning
Computational	several decades ago	Simulate complex natural physical phenomena
Data Exploration	now	Construct open and collaborative research and innovation based on massive data

Generally speaking, the raw data are of low value. So to make it effective during processing the big data, such as the trajectory data, it is necessary to do some reprocessing works, such as cleaning. The cleaning techniques include mean filter, median filter, Kalman filter and particle filter. The cleaning of trajectory data in the era of big data must be more cautious, because a large amount of subtle and useful information is mixed in the huge data. If the granularity of data cleaning is too fine, it is easy to filter out the subtle useful information, while if the granularity of data cleaning is too large, it is hard to achieve the real cleaning effect. So a careful balance between the quality and quantity of data cleaning is required. In chapter 3, an effective algorithm is proposed.(Zhang et al., 2021)

2.3 Summary

This chapter mainly discusses the development history of unmanned ships, the key technologies involved in the research, application prospects, development trends, and the connotation of data-driven trajectory analysis and processing technologies. The key technologies involved in the research, maritime applications, and technology prospects. Unmanned ships have a broad application space in the military and civilian fields. Academia, industry, and the military have all paid great attention to the research of unmanned ships. The research of unmanned ships involves a wide range of technologies, and this paper focus on the automatic route planning technology. Data-driven is an emerging processing and application model. Especially in the context of the implementation of the national big data strategy and the transformation of the DT era. Data-driven has received unprecedented attention and has been successfully applied in the fields of transportation, medical care, health, and education. In the context of the big data explosion and the rise of data science, based on massive trajectory data, data-driven analysis and research methods are applied to explore and propose new data-driven route planning models for unmanned ships.

Chapter 3 Trajectory data compression and threshold determination based on Douglas-Peucker algorithm

AIS (Automatic Identification System) is widely used in ships and Vessel Traffic Services (VTS). It exchanges information with other ships, AIS base stations, satellites and other equipment within the signal coverage area. The International Convention for the Safety of Life at Sea (SOLAS) requires international ships of 300 gross tonnage and above, non-international cargo ships of 500 gross tonnage and above, and all passenger ships to be equipped with AIS. Depending on the state of the ship's movement, the shipborne AIS equipment generally broadcast messages in 2 seconds to 6 minutes. The message includes static information such as the ship's name, Maritime Mobile Service Identify (MMSI), as well as other static information. Dynamic information includes heading, course over ground speed (COG), latitude, longitude time, and so on. With the latitude, longitude time, we can get the trajectory of the ship, so AIS messages include trajectory data. With the gradual improvement of the AIS shore-based network and the gradual application of satellite AIS (Satellite-based AIS), the coverage, continuity, effectiveness and availability of trajectory data have reached unprecedented levels, and the amount of trajectory data has also shown explosiveness Growth trend. In waters with heavy traffic, AIS base stations can receive thousands of AIS messages every second. Massive trajectory data provides a large amount of materials for the automatic route planning of unmanned ships based on data-driven, however, it brings great challenges to data storage and transmission.

Therefore, it is necessary to use a reasonable method to remove redundant information while extracting the feature points of the original trajectory, and to accurately represent the original trajectory while ensuring low distortion. In order to reduce the amount of calculation and speed up data processing, it is necessary to compress the AIS track data. The effect of AIS trajectory data compression largely depends on the characteristics of the AIS trajectory data itself. For example, the trajectory data for anchoring ships can be replaced by the trajectory point data at the time of initial anchoring and at the end of anchoring. During this period, other trajectory data information broadcast by the AIS device is of little significance to the characteristics of the entire trajectory. In addition, in open waters, ships generally sailing with a fixed course and speed, the trajectory data information of a speed-maintaining ship can also be replaced by the trajectory point data of the initial direction-maintaining and speed-maintaining sailing and the trajectory point data of the end of the direction-maintaining and speed-keeping sailing. During this period The trajectory point data of can be obtained through time linear interpolation. Ship motion trajectory data in other motion modes can also be characterized by extracting feature points. Data compression is one of the key technologies for trajectory data preprocessing. The compression algorithm for line elements can be used to compress the trajectory data, and the number of points used to characterize the original trajectory can be reduced on the premise of ensuring the characteristics and topological structure of the original trajectory data. The feature points of the original trajectory are extracted while the redundant information is to be removed. Trajectory data compression provides a high-quality solution for the efficient storage and analysis of massive trajectory data.

Trajectory data compression is one of the important components and key technologies of the data-driven automatic route planning method for unmanned ships. This chapter mainly introduces the trajectory data compression algorithm and the threshold determination based on the Douglas-Peucker algorithm. After preprocessing the trajectory data such as data analysis and data cleaning. The Douglas-Peucker algorithm is used to compress the trajectory data; in view of the problem of determining the compression threshold, the concept of the ship domain is used for reference, and the smallest ship domain is obtained based on the AIS trajectory data as the standard for determining the compression threshold.

3.1 Douglas-Peucker Algorithm

Connecting the AIS track points with the same MMSI in chronological order, it will form the actual track of the ship at that time. The actual track is a linear element rather than a point element or area element. Therefore, the compression algorithm of the line elements can be used to compress the track data. Linear feature compression has been widely used in map synthesis (Map Generalization), cartographic synthesis (Cartographic Generalization), numerical elevation model summary (Digital Elevation Model Data Profile), etc.(Pallero & Lg, 2013) Compression algorithms for linear elements can be divided into two categories, local compression algorithms and global compression algorithms. The local compression algorithm determines the choice of points by analyzing the relationship between every 2 or 3 adjacent points. This method is simple and efficient, but due to its local processing characteristics, it is difficult for the local compression algorithm to obtain the optimal results. In the global compression, the elimination or retention of a certain point is determined according to the relationship between this point and all trajectory points in the trajectory. Well-known global compression algorithms include Visvalingam-Whyatt algorithm, Douglas-Peucker algorithm, etc. . According to the experimental results of the literature, the Douglas-Peucker algorithm has the most accurate compression effect and the smallest deformation comparing to the original trajectory. (Douglas & Peucker, 2011)Therefore, this paper uses the Douglas-Peucker algorithm to compress the trajectory data.

The core idea of Douglas-Peucker algorithm is to extract a point set A' from the point set A that constitutes all the trajectory, which can reflect the overall and local main characteristics of the trajectory. Suppose the point set A describing the trajectory is: A=(Bakker et al., 2019..., An). If the sampling points are dense enough, the trajectory can be approximately replaced by a polyline $\{A_1A_2, A_2A_3, \ldots, A_{n-1}A_n\}$(Lee et al., 2019). When performing trajectory compression, the main task of the Douglas-Peucker algorithm is to extract a feature point set A' from the point set A that can reflect the main shape of the original trajectory: $A' = \{A_{d1}, A_{d2}, A_{d3}, \ldots, A_{dn}\}$. At the same time, the trajectory is approximately replaced by a polyline $\{A_{d1}A_{d2}, A_{d2}A_{d3}, \ldots, A_{dn-1}A_{dn}\}$. For the feature point set A', in addition to requiring a sufficiently small number of feature points included in the subset A', it is more important to reflect the overall and local important features of the original trajectory within the range of accuracy requirements. (Douglas & Peucker, 2011)

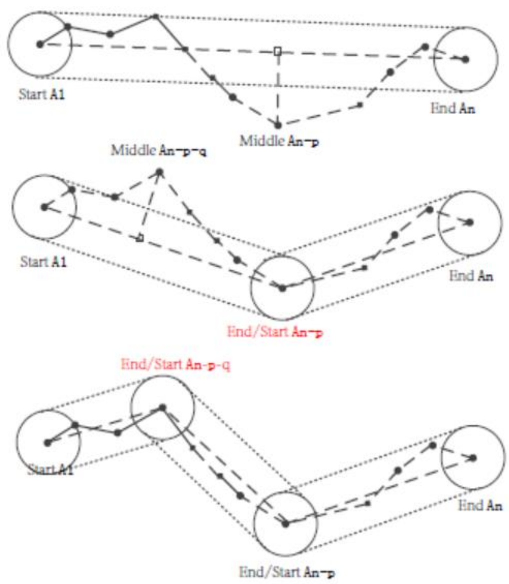


Figure 3.1 Douglas-Peucker Algorithm

Take figure 3.1 as an example to illustrate the principle of Douglas-Peucker algorithm. In the figure, the solid line represents the original trajectory, the long dashed line represents the compressed trajectory, and the radius of the circle is the compression threshold f. Set the first point (start) and the last point (end) in the original track point. set as the initial anchor point and the initial floating point, and the line formed by connecting the anchor point and the floating point is called the baseline. Calculate in turn the vertical distances from the middle points to the baseline except the initial anchor point and the initial floating point, and record the maximum vertical distance and the corresponding point. If the maximum vertical distance is less than the preset compression threshold r, the original trajectory point set can be approximately replaced by the baseline; otherwise, the point corresponding to the maximum vertical distance is defined as the split point and the original trajectory point set is divided into two as the limit At the same time, the split point is used as the floating point of the forward point collection and the anchor point of the backward point collection. The split point detection and segmentation are performed recursively on the split forward point set and the backward point set, until no new split point can be detected in each subset after splitting. The pseudo-code for the execution of Douglas-Peucker algorithm is shown in Table 3.1.

Table3.1 Pseudo Code for Douglas-Peucker Algorithm

Algorithm* 1：*Douglas-Peucker

```
input: Point List (Point 1, 2,    , n), PL for short; τ(threshold)
parameter: d (distance), dmax (maximum distance), index
for i = 2 to i = (n-1)
    //Find the point with the maximum distance
    d = PerpendicularDistance ( PL [i], Line( PL [1], PL [n] ) )
    if d > dmax
      index = i
      dmax = d
    end if
end for
  if dmax >= τ
    //If max distance is greater than threshold, recursively simplify
        RecursiveRlt1[] = Douglas-Peucker(PL [1...index], τ)
        RecursiveRlt2 [] = Douglas-Peucker(PL [index...n], τ)
        ResultList[] = { RecursiveRlt1 [], RecursiveRlt2 [] }
    else
        ResultList[] = {PL [1], …, PL [n]}
    end if
return ResultList[]
```

Now, let's discuss the trajectory data compression framework and process design.

(1) decode AIS Data

According to 61162-1 standard, the valid range of characters that AIS equipment can transmit is between 0X20 and 0X7E , including checksum, reserved characters and undefined characters. Each record transmitted by AIS device is marked with "$" or "!" as the start character and "<CR><LF>" as the end character.

According to the International Telecommunication Union (ITU) standard 1371, the encapsulated AIS data is processed by high-level data link control. Therefore, the first step of the overall framework and process design is to decode the message broadcast by AIS equipment according to IEC61162-1 standard and ITU 1371 standard.

(2) create trajectory database

It is a crucial factor to store retrieve and transfer the ais data fast. It is a key problem to extract the AIS message with the same MMSI code and import them into Douglas-Peucker compression algorithm in time order quickly. SQLite database is an embedded lightweight database management system that abides by ACID (atomicity, consistency, isolation, durability). It occupies very low system resources and has many advantages, such as light weight, fast processing speed, open source, zero configuration and so on. Therefore, this paper uses SQLite database to store the track data after analysis. The internal storage mechanism of SQLite database is shown in Figure 3.2. It should be pointed out that the sample data used in this paper is still lightweight, and SQLite database is one of the best solutions for lightweight data management. When trajectory data reaches the level of big data, SQLite database is no longer applicable. Big data solutions, such as Hadoop big data management platform, are needed, but the overall framework design and implementation process remain unchanged. What is need to do for big data platform is just to change the processing statement of the interface.

A table named "TRACK" in the database is used to store track data. The track data message broadcast by the AIS device contains a wealth of information. In order to distinguish the track data records broadcast by different AIS devices, the MMSI code is used as a field (SHIPID) in the "TRACK" table. In the Douglas-Peucker compression algorithm, the broadcast time (LTM), longitude (Lon), and latitude (Lat) of the trajectory data are respectively used as a field in the single table. In addition, in the decoded AIS record, A+B is equal to the ship length, C+D is equal to the breadth of the ship, and the length and breadth of the ship are respectively regarded as a field in the single table. The single "TRACK" table design in the database management scheme is shown in Figure 3.3.

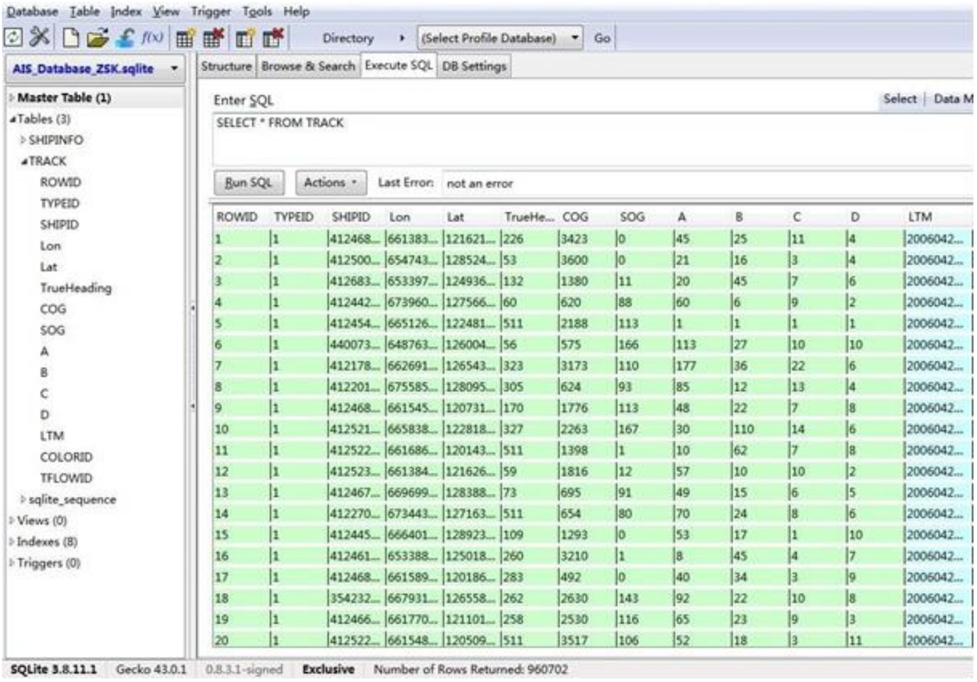


Figure 3.2 “TRACK” Table in Database Management

(2) clean the AIS data

The message broadcast by AIS equipment includes static information, dynamic information and information related to voyages. The amount of dynamic information is generally huge and some of them are invalided. For example, the ship length is over 512m or cog is over 60kn, and so on. Therefore, before compressing the AIS data, it is necessary to clean the information broadcast directly by the AIS device according to the principle of data cleaning. Data cleaning mainly includes two aspects. One is to remove incomplete data, including incomplete dynamic data caused by communication packet loss and incomplete static information caused by manual input, such as information broadcast by AIS equipment with MMSI of 0. Eliminate erroneous data, for example, points with large changes in latitude and longitude in time-continuous records. Commonly filtering methods include mean filtering, median filtering, Kalman filtering and particle filtering. The method of Kalman filter to eliminate abnormal points in trajectory data is relatively simple and can effectively eliminate abnormal points in trajectory data. In this paper, Kalman filter is used to eliminate abnormal points in trajectory data.

3.2 Determination of compression threshold

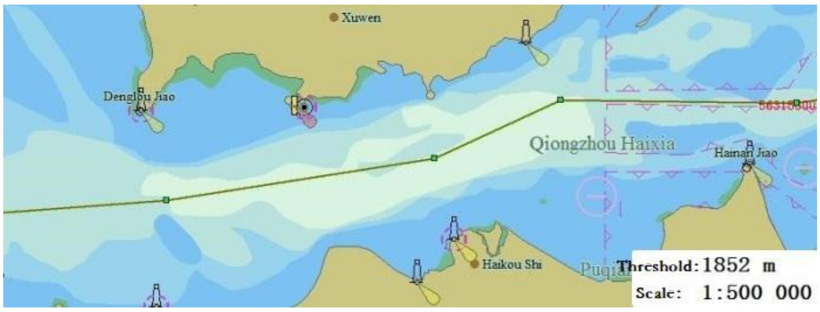
3.2.1 The meaning of the trajectory data compression threshold

In Douglas-Peucker compression algorithm, besides the input trajectory data that needs to be compressed, the only parameter that needs to be determined is the compression threshold r. In the following, two examples are used to illustrate the meaning and significance of the compression threshold. Taking the trajectory data of a ship with MMSI code 563156000 and 636014456 across the Qiongzhou Strait on a certain day in 2011 as examples. According to the overall framework and process design of trajectory compression, the compression thresholds are set to 1.852 m, 18.52m, 185.2 m, and 1852 m. Compress the trajectory data of the two ships respectively, and then the feature points extracted after compression are represented by small squares and displayed on the electronic chart platform, as shown in Figure 3.6. and Figure 3.7.

a. Threshold = 1.852m

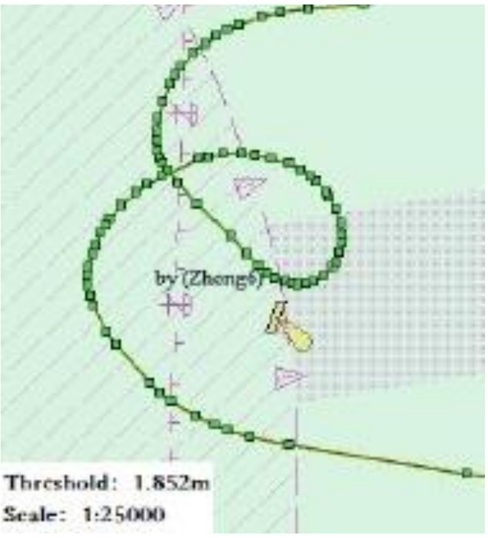
b. Threshold = 18.52m

c. Threshold = 185.2m

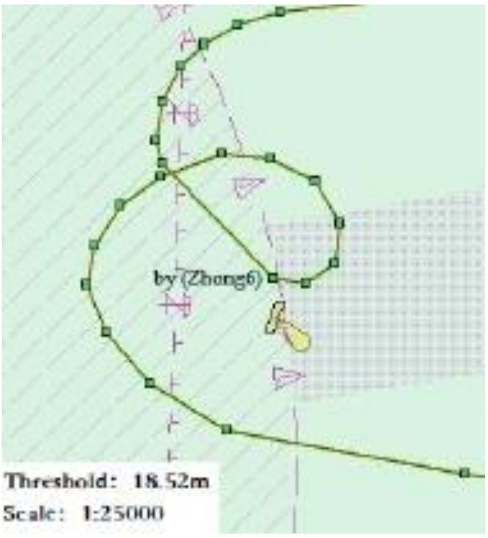


d. Threshold = 1852m

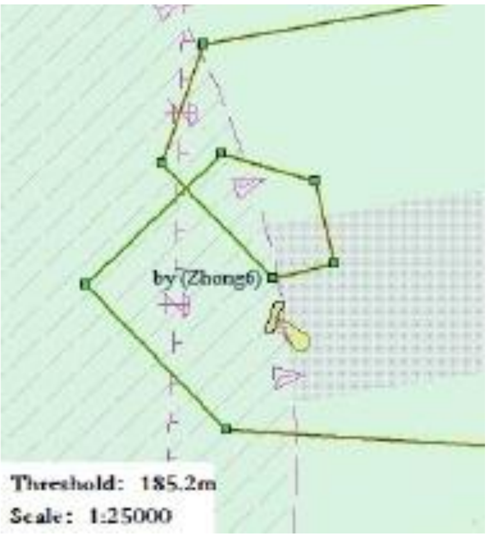
Figure 3.3 Trajectory Data Simplification of Ship Which MMSI is 563156000 in Different Threshold



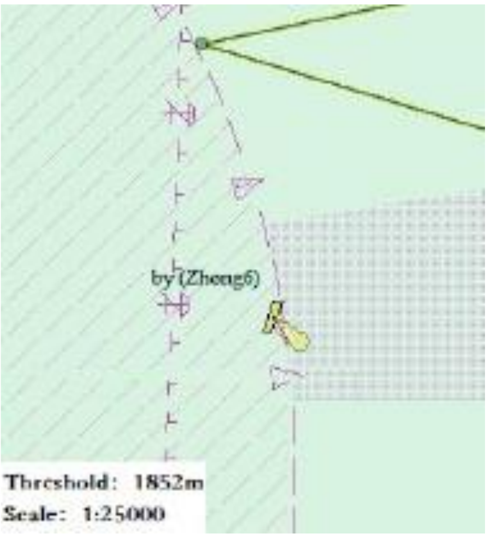
a. Threshold=1.852m



b. Threshold=18.52m



c. Threshold=185.2m



d. Threshold=1852m

Figure 3.4 Trajectory Data Simplification of Ship Which MMSI is 636014456 in Different Threshold

By observing the compression effect pictures of the trajectory data of the ships with the MMSI code of 563156000 and the MMSI code of 636014456 under different thresholds, combined with the compression rate statistics in Table3.2, it can be seen that,

(1) With the increasing of the compression threshold, the number of compressed trajectory points is less and less, and the compression ratio is higher and higher,

(2) When the compression threshold is 1.852 m (0.001 nm), the compressed track points are more dense and have stronger ability to represent the original track features,

(3) When the compression threshold is 1852 m (1 nm), the compressed trajectory points are too sparse to represent the original trajectory features.

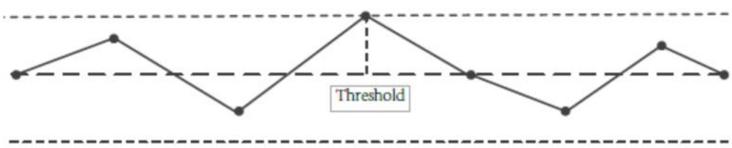
Therefore, the size of the threshold determines the ability of the compressed trajectory to represent the original trajectory features. How to improve the ability of representing the original trajectory features under the premise of ensuring the compression rate is a key problem to be solved. A reasonable threshold should not only improve the compression ratio and ensure the processing efficiency of the subsequent process, but also ensure that the compressed trajectory can accurately represent the characteristics of the original trajectory within the allowable error range.

Table3.2 Compression Rate of Ship Trajectory Data in Different Threshold

Threshold（m）	MMSI 563156000			MMSI 636014456		
	Data no. before compression	Data no. after compression	ratio	Data no. before compression	Data no. after compression	ratio
1.852	8155	1247	84.71	4581	686	85.03
18.52	8155	158	98.06	4581	149	96.75
185.2	8155	47	99.42	4581	39	99.15
1852	8155	11	99.87	4581	12	99.74

3.2.2 Relationship between ship domain and compression threshold

The ship domain refers to the effective water area which is centered on the ship and keeps other ships and obstacles out to ensure the safety. Ship domain model is an important theoretical basis for ship collision avoidance decision-making, navigation safety evaluation, channel planning and design, and marine traffic capacity research. The ship domain model can provide decision-making basis for determining the safe navigable waters of ships. As shown in Figure 3.1 and figure 3.8 (a), the compression threshold f of Douglas-Peucker algorithm refers to the maximum allowable vertical distance from the track point in the original track to the track line (baseline) after compression. Therefore, as shown in Figure 3.8 (a), the solid line represents the track line formed by the connection of the original track points, and the long dotted line represents the compressed track line. Then the track band within the threshold range can be defined as the track band with the compressed track line as the center and extending in two directions perpendicular to the compressed track line. It can be seen from Fig. 3.1 and Figure 3.5 (a) that all the original track points within the track band can be represented by the compressed track line. Similarly, as shown in Figure 3.5 (b), the track band within the ship domain can be defined as the track band with the ship's course as the center and extending the size of the ship domain in two directions perpendicular to the course. By comparing figure (a) and figure (b) in Figure 3.5, in order to ensure that all the original track points corresponding to the compressed track line are within the safe range of the ship domain, the ship domain model is used as the standard to determine the compression threshold.



(a) Compresed track within the threshold

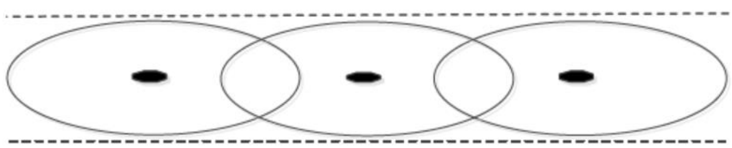


Figure 3.5 The Relationship Between Ship Domain and Simplification Threshold

The concept of ship domain was first introduced by Fuji (Matui et al., 1981), the ship domain is defined as a two-dimensional oval water area with the ship as the center and other ships must avoid entering. The other two frequently cited ship domain models are developed by Goodwin and Coldwell. After that, many experts and scholars have carried out extensive and in-depth research on the shape, size and application of ships from different angles, which can be divided into statistical analysis based ship domain model, analytical expression based ship domain model and intelligent theory based ship domain model. Because different ship domain models consider different factors, the size and shape of ship domain are different. However, in practical application, especially in the heavy traffic waters, the acceptable ship area size is often smaller than the empirical value.

For example, Fuji (1981) established a model of a ship domain as an oval shape with the ship in the center. The major axis of the oval shaped domain reflects a tolerated distance between the stand on vessel and the others, either ahead or aft of herself. The minor axis would be described as the closest range between ships while travelling abeam.(Matui et al., 1981) Similarly, the size of the ship domain model proposed in reference (Z & J, 2009) varies with the navigation environment.

In view of the fact that the empirical value of the ship domain is often too large, the actual navigation environment must be considered when determining the size of the ship domain in practical application. Therefore, this paper proposes a method to determine the minimum feasible ship domain size based on trajectory data, and takes it as the basis to determine the compression threshold. The proposed method is not to propose a new ship domain model, but a fast, convenient and simple method to determine the size of ship domain based on trajectory data.

3.2.3 Determination of compression threshold based on trajectory data

According to the analysis in section 3.2.2, the size of ship domain can be determined as the standard of Douglas-Peucke. But the empirical value of ship domain model size is too large, and it is necessary to determine the minimum acceptable ship domain size according to the actual navigation environment. The steps to determine the minimum feasible ship area size based on trajectory data are as follows.

(1) The SQL statement "select distinct shipid from track order by shipid" is used to search the track database, and the MMSI codes of all ships in the related area are extracted.

(2) The SQL statement "select * from track where shipid =% d order by LTM" is used to extract the track data with the same MMSI code and arrange them in chronological order. Then the coordinate points in the track data are linearly interpolated according to time to ensure that each integral time point corresponds to a coordinate point.

(3) In order to carry out linear interpolation for all trajectory data in the related area according to the method shown in step (2).

(4) Select any ship in the study area and set it as the center ship to calculate the bearing and distance of other ships relative to the center ship at a certain integral time point. In order to facilitate the superposition of the calculation results of different ships, the calculated distance is divided by the ship length of the center ship to get the relative distance, and the data with the relative distance greater than 10 is eliminated. The

bearing and distance of other ships relative to the center ship at other integral time points of the center ship are calculated in turn.

(5) According to the data obtained in step (4), set the center ship at the position of (0,0) in the coordinate system, and draw the scatter diagram of the distance and bearing of other ships relative to the center ship at all integral time points.

(6) Select other ships in the study area as the center ship in turn, and repeat steps (4) and (5) to draw the scatter diagram of distance and bearing of other ships relative to the center ship at all integral time points of other center ships.

(7) The distance and bearing scatter diagrams of all the center ships are superimposed together, and the size of the ship domain can be calculated from the superposition. It can be obtained from the scatter diagram. (Shu-kai, 2016)

It should be pointed out that the above seven steps are the basic steps to determine the minimum feasible ship area size based on the trajectory data, and there are some details to be noted in the actual operation, for example, to eliminate the trajectory position when the ship is at anchor or on berth; To ensure that the position of the two ends of the linear interpolation trajectory has practical significance.

In order to illustrate the scientific rationality and feasibility of determining the minimum feasible ship area size based on trajectory data more vividly, this paper takes 962 ships collected by Qiongzhou Strait VTS center from July 2, 2011 to July 12, 2011 as an example. (Singh et al., 2018)

Taking the 5,902,840 trajectory data of ships as an example, the trajectory data are processed according to the method proposed in this section, and the results are as follows.

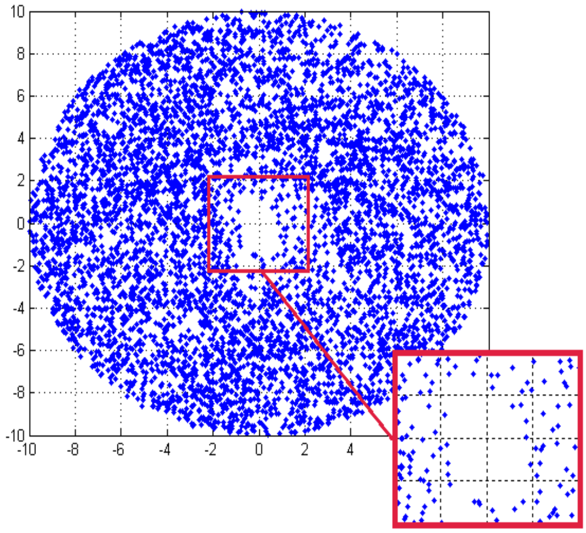


Figure 3.6 Scatter Diagram of Centered Ship

The scatter diagram of the distance and bearing of the center ship is shown in Figure In Figure 3.6, the position of (0, 0) represents the center ship, and the forward direction of longitudinal axis represents the bow direction. The positive direction of the transverse axis represents the right abeam direction of the ship. As can be seen from Figure 3.6, except for the blank area in the middle, other ships are evenly distributed around the center ship. In order to determine the size of the blank area more clearly, the scatter diagram is enlarged, and the side length of the small square is doubled as ship length, as shown in the lower right corner of figure 3.9. As shown in the color box, there is no other ship within the range of one time the ship length in longitudinal direction, and there is no other ship within the range of 0.8 times the ship length in longitudinal direction. According to the analysis in section 3.2.2, **0.8 times the length of the ship** can be used as the maximum threshold for trajectory data compression.

3.3 Summary of this Chapter

Trajectory data compression is one of the important steps and key technologies of data-driven trajectory analysis and processing. The efficiency of subsequent work can be greatly improved by compressing the original trajectory data.

Trajectory data provides great convenience for observing the macro situation of traffic flow and analyzing the micro traffic characteristics. However, when the trajectory data reaches the big data level, it will bring great difficulties and challenges to the data storage, transmission and processing. In this section, Douglas-Peucker algorithm is applied to take data compression, and the overall framework and implementation process of track data compression are designed; The compression threshold is the only parameter in the compression algorithm, and its size directly determines the compression ratio and compression effect. In order to determine the compression threshold, this paper uses the concept of ship domain for reference, and obtained the minimum ship domain based on the trajectory data as the standard to determine the compression threshold.

Chapter 4 AIS Data Clustering and Automatic Route planning Based on Hierarchical DBSCAN Algorithm

Driven by the Internet of things, big data, cloud computing and other new technologies, data-driven smart maritime related research is gradually becoming a research hotspot. With the establishment of Asia Pacific AIS data center in China and the interconnection with IALA net, the global trajectory data becomes readily available, which provides a large number of available materials for data-driven intelligent maritime research.(Wang et al., 2018)

In Chapter 3, after preprocessing massive trajectory data such as data cleaning and data compression, this chapter mainly introduces the trajectory data clustering and automatic route planning method based on hierarchical DBSCAN algorithm: (Wu et al., 2015) Based on the results of track data compression using Douglas-Peucker algorithm, the way points in the track feature points are identified; Then, through the establishment of similarity measurement standard, hierarchical DBSCAN algorithm is used to cluster the way points in different trajectories, so as to identify the way points of the route formed by similar trajectories. Compared with the previous processing methods for track segment, it has strong anti-noise ability and improves the processing efficiency. Then, the connectivity between the route way points is determined by the trajectory data. By comparing the automatically generated route with the traffic flow macro situation described by the original track data, it is proved that the automatic route planning method proposed in this section is effective.

4.1 Design of automatic route planning framework

4.1.1 Overview of route planning

Voyage plan refers to the specific measures and countermeasures related to the safe navigation of ships in the process of sailing from the starting port to the destination port after the voyage task is determined. Route planning is an important part of the navigation plan. Before the voyage, the second mate shall, according to the performance of the ship and the sea area to be navigated, refer to the route guide and other relevant nautical books and materials, carefully study the hydrometeorological factors in the sea area to be navigated, the navigation regulations such as the ship's routing system, the navigation aids such as the dangerous obstacle signs, and the political situation of the navigation area, etc., so as to choose a safe and economic route.(Zhang et al., 2021)

The two principles of route planning are safety and economy. Safety is the first factor. The route planning should first ensure the safe of ships, and then consider improving the economic benefits. In order to improve the quality of route planning, we mainly refer to the recommended route given in navigation books and materials. In addition, by accumulating a lot of sailing experience and referring to the sailing experience of other ships, we can gradually master the rules of some routes.(Peng et al., 2019)

In the era of paperless navigation, the traditional methods of manually consulting navigation books and materials, analyzing paper charts and manually drawing routes have some limitations, such as low efficiency, and the quality of route planning depends on the professional level and working attitude of officers and the master. With the rapid development of information technology and the continuous improvement of ship intelligence, it is necessary to make full use of the automatic calculation function of computer to explore the problem of automatic route planning. On the other hand, the realization of automatic route planning is an important symbol of the intelligent level of the unmanned ship, and it is also a key step for the unmanned ship to move towards full autonomy. In order to realize the full autonomy of the unmanned ship, we must change the current situation that the unmanned ship relies on shore based or mother ship staff to design the route, Reasonable automatic route planning can not only ensure the safety of the unmanned ship, but also reduce fuel consumption and enhance the endurance of the unmanned ship. (Shrivastava & Gupta, 2012)

4.1.2 Data driven automatic route planning

At present, the research mode of automatic route planning technology can be roughly summarized as follows: grid processing of electronic chart or adaptive binary threshold processing of radar image and satellite photo, so as to divide the processed electronic chart or radar image and satellite photo into navigable water area and non-navigable water area, and then use intelligent search algorithm in navigable water area Artificial potential field method and visual graph method are used to search the shortest path. The "best route" generated in this mode is only the best route from the perspective of graph theory, ignoring the route rules that ships must abide by in actual navigation. There is still a certain gap between the theoretical analysis results and practical application, for example, The routes extracted by the intelligent search algorithm cannot intelligently identify the routes or channels that need special attention according to IMO regulations, such as the navigable separation waters and the recommended traffic flow direction, which results in the inability to drive according to the rules in these areas.(UKHO, 2018)

Rich navigation related information is stored in the trajectory data, which is the objective response of the real navigation situation. If we can fully analyze and mine this information, and extract the navigation related information from the historical trajectory data that has been proved feasible by the actual navigation, it can provide important reference for the automatic generation of routes, This is also in line with the idea of "referring to the specific sailing experience of other ships" in route planning.(Wu et al., 2015)

The traditional road traffic map mapping is mainly realized by manual survey, remote sensing image processing and special mapping vehicle. This traditional working mode needs to collect and process all the road information one by one, which is heavy workload, high cost and low cost.

The map can't be updated in time. With the rapid popularization of intelligent mobile terminals and the construction of smart city, rich digital applications and perfect information infrastructure have become one of the basic characteristics of modern digital city, Using data-driven trajectory analysis and processing technology to automatically generate maps has become a new mode of road traffic mapping. Literature (Ahmed et al., 2015) studies the automatic generation and update method of road traffic map based on vehicle GPS trajectory data, and the generated electronic map can intelligently recommend the shortest path from the set starting point to the end point for users. There are similarities between sea traffic and land traffic, so we can learn from the processing mode and method of track data in road traffic mapping; On the other hand, there are also differences between sea traffic and land traffic. On land, the roads that vehicles can travel and their width and direction are fixed. When ships are sailing on the sea, except for the areas where the ship routing system is implemented, such as traffic separation system and recommended routes, the navigable waters of ships are relatively wide. We need to learn from the road traffic research methods and improve them, and design the automatic route planning framework and process based on massive trajectory data.(Xu et al., 2016)

Data driven is still a hot topic and one of the most challenging research directions. Based on data-driven trajectory analysis and processing technology, massive trajectory data are analyzed and processed, and routes are automatically generated according to the set starting point and end point. It can not only be a successful application example of data-driven in navigation, but also enrich and improve the theory and method of automatic route planning.(Zhang et al., 2015)

4.1.3 Framework design of automatic route planning

(1) Preprocessing and compression of trajectory data (which has been completed in Chapter 3) after preprocessing the trajectory data, such as data analysis and data cleaning, Douglas Peucker algorithm is used to compress the trajectory data, which can reduce the amount of data, improve the processing efficiency and identify the feature points in the trajectory data.

(2) Trajectory way point identification

The heading of the constant line formed by every two adjacent feature points in the compressed trajectory data is obtained, and then the compressed trajectory data is obtained in the trajectory data, every three adjacent feature points constitute the angle, and the way point in each trajectory is identified by setting the turning threshold.

(3) By establishing the similarity measurement standard of way points, the way points in different tracks are clustered by hierarchical DBSCAN algorithm to identify the way points of similar tracks.

(4) Route way point connectivity identification

In the process of clustering the way points, all the way point information of each route way point cluster is recorded, and the connectivity between the route way points is judged by the real track.(Zhao et al., 2018)

The overall framework design of automatic route planning proposed in this chapter is shown in Figure 4.1.

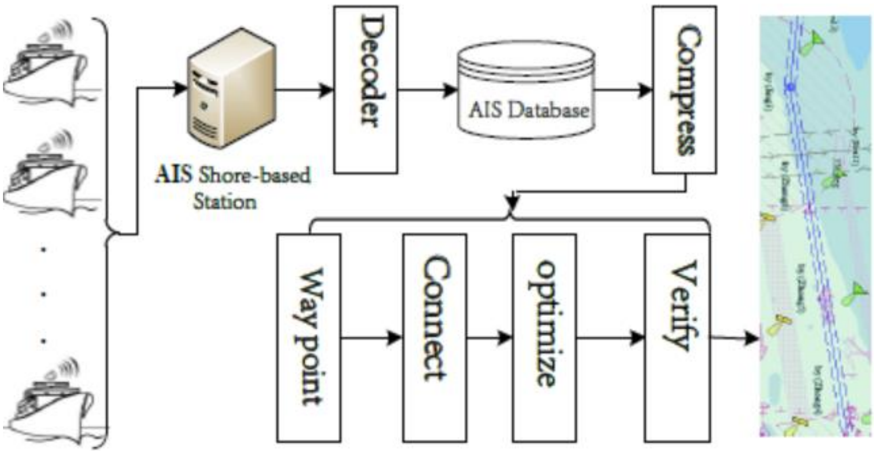


Figure 4.1 Overall Framework Design of Automatic Routing

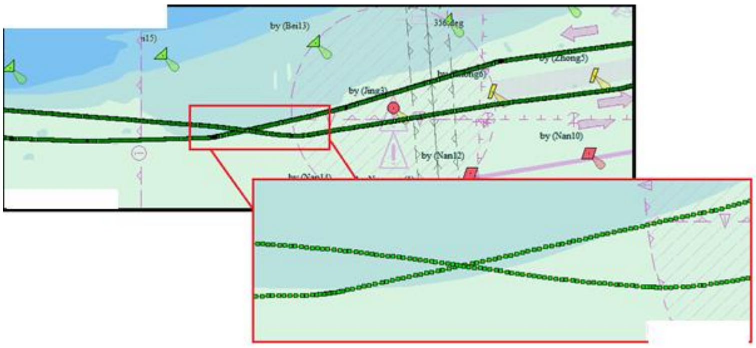
4.2 Trajectory data clustering based on hierarchical DBSCAN algorithm

4.2.1 Trajectory way point identification

According to the track data compression based on Douglas Peucker algorithm in Chapter 3, after the original track data is compressed, the feature points in the original track data can be extracted, and the track line formed by connecting the feature points in time order, that is, the compressed track data, can accurately represent the characteristics of the original track data within the allowable error range. On the one hand, the amount of compressed trajectory data is greatly reduced, which can greatly improve the efficiency of subsequent processing; On the other hand, because Douglas Peucker algorithm is a global compression algorithm, it has a strong ability to identify the overall motion characteristics of trajectory data. As shown in Figure 4.2, on July 7, 2011, the ship with MMSI code 636092238 crossed the Qiongzhou Strait from latitude, and imported the original trajectory data and the compressed trajectory data into the electronic chart platform respectively. Through observation, it can be seen that on the one hand, the compressed trajectory data is significantly reduced, And it can accurately characterize the characteristics of the original trajectory data within the allowable error range; On the other hand, the compressed trajectory data shows a more obvious single ship motion mode, especially when the ship makes continuous slow steering in the red box in the lower right corner of the figure, the compressed trajectory data can show the way point in the trajectory very clearly.(Zhen et al., 2017)

When identifying the way points in the trajectory data, it is necessary to find every two adjacent trajectory points in the trajectory data in turn
Then the angle of every three adjacent feature points in the trajectory is calculated, and the way point in the trajectory data is identified by setting the turning threshold. If the way point in the trajectory is identified based on the original trajectory data, not only the data is large, time-consuming and labor-consuming, but also when the ship makes continuous slow turning, as shown in Figure 4.2 (a), the included angle of each adjacent three trajectory points in the trajectory data is very small, and the way point in the trajectory data cannot be accurately identified, Therefore, it is impossible to identify the movement mode of a single ship accurately, which will affect the intensive reading and validity of the next step based on way point clustering. Firstly, the track data based on Douglas-Peucker algorithm has high reduction rate, small amount of data and high efficiency; Second, Douglas Peucker algorithm is a global compression algorithm. The extracted feature points can better maintain and reflect the overall motion characteristics of a single ship. As shown in Figure 4.2(b), the compressed trajectory data can clearly show the way point of the ship. Therefore, the track data

compression based on Douglas-Peucker algorithm not only has the advantages of high compression rate and high efficiency of subsequent processing, but also has the advantages of high accuracy in motion pattern identification and steering point identification.(Zou et al., 2018)



(a) track with original AIS data

(b) track with compressed AIS data

Figure 4.2 comparation of track of a single ship between original data and compressed data

According to the international regulations for collision avoidance and the requirements of good ship craft, if the conditions permit, the steering range should be large enough for other ships to be able to clearly detect through visual or radar observation, and a series of continuous small range steering should be avoided. In this section, set 5°as the way point threshold in the trajectory. The selection of this parameter has a certain robustness. In the subsequent clustering operation, the similarity measure can be established to measure the similarity

The similar way points are clustered, and the track points that reach the turning threshold but are not way points are defined as noise points to be eliminated.(Yanıkoğlu et al., 2019)

Suppose that the compressed trajectory data is expressed as: T = {P_0, P_1, P_2,..., P_n}, where p=[φ，λ]。 Then the course C_{P0P1} of the constant direction line connected by the trajectory point P_0 and P_1 under the Mercator projection can be calculated by equation (4.1).

$$\begin{cases} MP(\varphi) = 7915.70447 \lg[\tan(\frac{\pi}{4} + \frac{\varphi}{2})(\frac{1-e\sin\varphi}{1+e\sin\varphi})^{e/2}] \\ C_{P_0P_1} = \arctan(\frac{\lambda_2 - \lambda_1}{MP(\varphi_2) - MP(\varphi_1)}) \end{cases} \quad (4.1)$$

In formula (4.1), *e* is the first eccentricity of ellipsoid.

Similarly, under Mercator projection, the heading C is the constant direction line connected by the track points P1 and P2 can be calculated by formula (4.1), and then the steering angle T_{P0P1P2} composed of line segment P0P1 and line segment P1P2 at the track point P1 can be calculated by C_{P0P1} and C_{P1P2}. If the steering angle T_{P0P1P2} is greater than 5 °， Then the track point P1 is marked as the way point. For each way point, in addition to its position information, longitude (lon) and latitude (lat), MMSI code, turning angle T, and direction to the way point should also be recorded _In), such as C_{p0p1}, and direction out of the way point _Out), such as C_{P1P2} . Therefore, an effective way point (WP) consists of six elements: WP = {MMSI, lon, lat, t, direction_In，Direction_Out}。 Based on 962 compressed trajectory data collected by Qiongzhou Strait VTS center from July 2, 2011 to July 12, 2011, the way points are identified in turn, and the six elements information of each way point is recorded. The steering points of 962 identified trajectory data are displayed on the electronic chart platform, as shown in Figure 4.3, in which each green triangle represents a steering point.

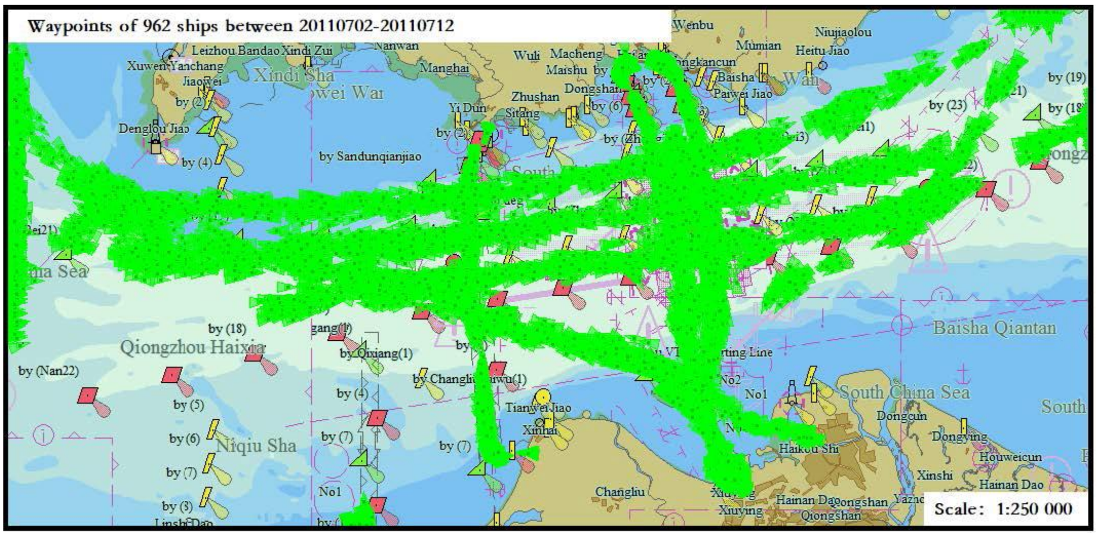


Figure 4.3 All Way points in Qiongzhou strain

Note: This is all the waypoints recognized according to the definition in section 4.2.1.a green triangle vessel in the figure denotes one of way points in a trajectory passing through Qiongzhou strait.

4.2.2 The Principle of DBSCAN algorithm

The density-based spatial clustering algorithm with noise points (DBSCAN) has the advantages of high processing efficiency, strong anti-noise ability, and ability to find clusters of arbitrary shapes when processing high-dimensional spatiotemporal data (Spation-temporal Data). The basis of incremental clustering algorithm (Chakraborty & K, 2014). The core idea is to cluster high-density regions of spatiotemporal data into clusters, and each object in the cluster is required to reach a certain threshold (MinimumPoints, the number of other objects within its given radius (Eps-Neighborhood)). MinPts) (Shrivastava & Gupta, 2012). The result of clustering is to divide the similar spatio-temporal data in the spatio-temporal point set S into the same group according to the similarity metric, so that the difference in data similarity in the same group is as small as possible and the difference between groups is as large as possible. Several important definitions in the DBSCAN algorithm are as follows:

Eps-Neighborhood: The Eps-Neighborhood object of a point P refers to the set of all points in the point set S whose distance to the point p is less than Eps, that is,,

NHEps(p)={q∈S｜D(p,q)≤Eps}；

CorePoint: If the number of points in the Eps-neighborhood object of point P is greater than or equal to MinPts, then point P is the core point;

DirectlyDensity-reachable If point Q is corepoint and p∈NHEps(q)， It is said that the density of point P can be reached directly from point Q.

Density-reachable： For the point set {p1, p2,..., pn}, and p1=q, pn=p, if the point pi+1 is directly accessible from the point pi , then the point P is said to be accessible from the point Q;

Density-connected: If there is a point r(r∈S)that makes the density of both points p and q reachable from point r, it is said that the density of points p and q can be connected;

OutlierPointorNoisePoint： If the density of point p is not reachable from any point in the point set, then point p is a noise point;

When DBSCAN clustering algorithm is used to cluster spatiotemporal data, all data are marked as not visited at first, and then cluster randomly from any point in the point set. For any selected point P, first search the EPS neighborhood object of point P and mark point P as access. If the number of EPS neighborhood objects of point P is greater than or equal to minpts, that is, point P is the core point, then point P forms the rudiment of cluster; Then, for all the points marked as not visited in the EPS neighborhood object of point P, the EPS neighborhood object is searched in turn, and the above steps are executed circularly until all the density connected points in the cluster are found. In the new round, the point marked as not accessed in any selected point set, repeat the above steps. The pseudo code of DBSCAN clustering algorithm is shown in table 4.1.

Table4.1 Pseudo Code for DBSCAN Algorithm

```
algorithm 1：DBSCAN
input：points
param：Eps， MinPts
output：result C
DBSCAN(S, Eps, MinPts)
  C=0
  For each point P in S
     If P is visited
        Continue next point
     End if
     Mark P as visited
     Eps-Neighborhood=regionQuery(P, Eps)
     If sizeof(Eps-Neighborhood)<MinPts
        Mark P as NOISE
     else
        C=nextcluster
        expandCluster(P, Eps-Neighborhood, C, Eps, MinPts)
     end if
   end for
 END
   regionQuery(P,Eps)
        return Eps-Neighborhood(P)={Q∈S｜D(P,Q)≤Eps}
   expandCluster(P, Eps-Neighborhood, C, Eps, MinPts)
     add P tocluster C
     for each point P' in Eps-Neighborhood
        if P' is not visited
          mark P' as visited
          Eps-Neighborhood'=regionQuery(P',Eps)
          If sizeof(Eps-Neighborhood')>=MinPts
             Eps-Neighborhood=Eps-Neighborhood joined with Eps -
             Neighborhood'
          endif
        endif
        if P' is not yet member of anycluster
          add P' to cluster C
        end if
     end for
```

Take figure 4.4 as an example to illustrate the clustering analysis and results of DBSCAN algorithm. In Figure 4.4, the dotted circle represents the EPS clustering radius of the point, and minpts is set to 3; Points 2 to 7 marked in blue are core points, because there are at least three points in their EPS neighborhood; Points {2,3,4}, {3,4,5}, {5,6,7} are all directly accessible to each other; The red marked points 1 and 8 are not core points, because there are only two points in their EPS neighborhood; Point 1 and point 8 can be connected by point chain < 4, 5, 6 > density, so the eight points from point 1 to point 8 belong to the same cluster; Point 9 marked in orange is a noise point because it is not as dense as any point.

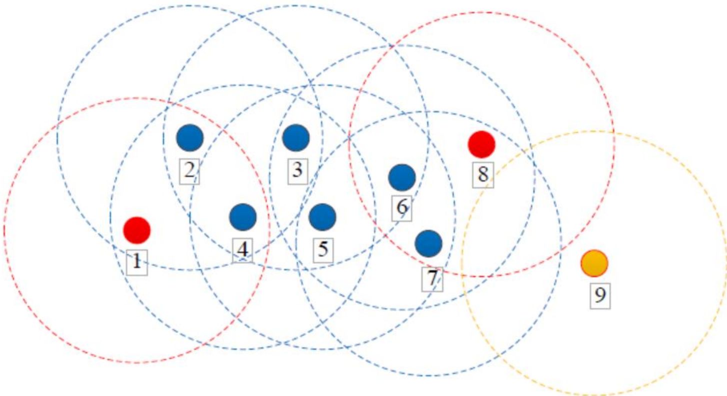


Figure 4.4 An Illustration of DBSCAN Cluster Algorithm

4.2.3 Trajectory data clustering and way point identification

This section mainly uses hierarchical DBSCAN algorithm to cluster the identified way points in Section 4.2.1, so as to identify the way points in the route. As shown in Figure 3.10, the trajectory data collected by the VTS center of Qiongzhou Strait in one day on July 7, 2011 is imported into the electronic chart platform for observation, which can not only be used to analyze the macro characteristics of traffic flow, but also show clearly the route formed by the superposition of trajectory data. In the same way, similar way points in trajectory data can be clustered by setting up similarity metrics, and way points can also be extracted from them.

In the original DBSCAN algorithm, there are two parameters, clustering radius EPS and neighborhood density threshold minpts. When using DBSCAN algorithm to cluster the way points, domain knowledge should be combined to ensure that the way points formed by clustering are real and effective. When determining the similarity measurement standard of way point, the following two factors are mainly considered: spatial proximity and turning similarity. Spatial proximity mainly means that the way points that can be clustered in the same cluster should be limited in a certain geographical space; Steering similarity mainly refers to the steering angle and direction that can be clustered among the steering points of the same cluster_ In、Direction_ The out difference should also be limited to a certain range. Due to the different track density of different routes, it is difficult to determine a set of clustering parameters to ensure the accurate extraction of the way point information of different routes. Compared with Figure 3.10 and figure 4.3, it can be seen that in the heavy traffic waters, the way points in the trajectory data are also densely gathered. In these waters, more accurate clustering parameters need to be used, and the measurement standards of spatial proximity and turning similarity need to be strict to ensure that the waypoints in different routes can be distinguished; In the areas with sparse traffic, the maneuverable space of the ship is large, and the motion patterns vary greatly. The way points in the trajectory data are sparse and clustered together. In these waters, rough clustering parameters are needed, and the measurement standards of spatial proximity and turning similarity are relatively broad, In order to avoid defining the way point in the same route as the noise point. Therefore, this section proposes a hierarchical DBSCAN clustering algorithm. The hierarchical DBSCAN clustering algorithm does not improve the algorithm itself to ensure the simplicity and efficiency of the original algorithm. Instead, in the first layer clustering, a set of clustering parameters are determined, and the DBSCAN algorithm is used to cluster the way points. In the second layer clustering, according to the results of the first layer clustering, the top 10% of the larger clusters are selected, A set of strict clustering parameters is determined, and then DBSCAN algorithm is used to cluster the way points in the cluster. After many experiments, the final hierarchical DBSCAN clustering parameters are shown in table 4.2. The hierarchical DBSCAN algorithm is used to cluster the steering points identified in Section 4.2.1. The results are shown in Figure 4.5. 173 different clusters are formed. Different clusters are marked with different colors in the electronic chart platform.

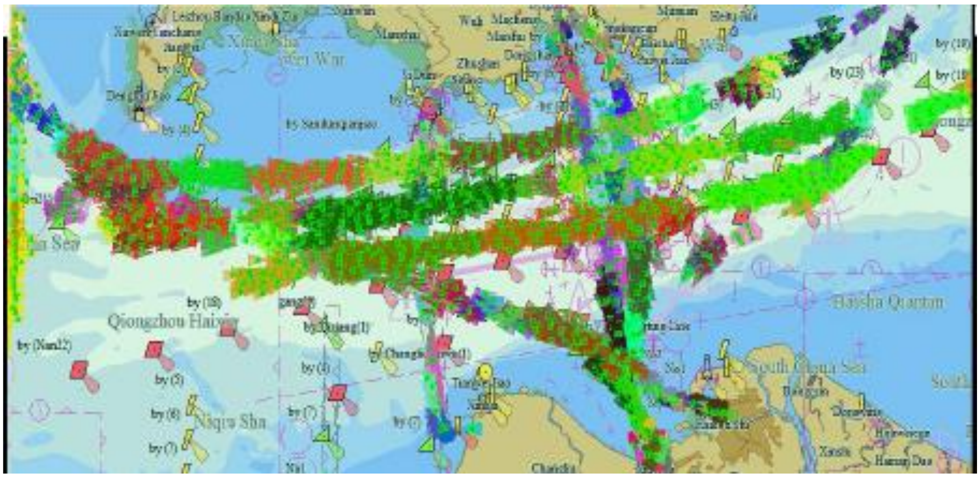


Figure 4.5 Clustered Way points

Note: These are the clutered Using Hierarchical DBSCAN Algorithm. The way points that belong to the same cluster are clustered together and marked in different color.

In section 4.2.1 knuckle steering point identification, six parameters are used to uniquely determine a steering point. For the route turning node (TN) formed by clustering the way points with hierarchical DBSCAN algorithm, five parameters are used to represent it.

TN={Constitution_MMSI, Lon, Lat, Direction_ In, Direction_ Out}。 Among them, construction MMSI refers to the trajectory way points contained in the route way points, that is, which trajectory way points are clustered into the route way points; *Lon* and *lat* refer to the geographical position of the route way point, which is calculated from the position of the included trajectory way point; Direction_In and direction_Out refers to the heading to the waypoint and the heading out of the waypoint, which is also determined by the direction of the included trajectory way point direction_in and direction_out It's worth it. By calculating the mean value of all trajectory way points in the same cluster, we can get the position of route way points, heading, heading and other information, and mark the obtained route way points on the electronic chart platform with a blue triangle, as shown in Figure 4.6.

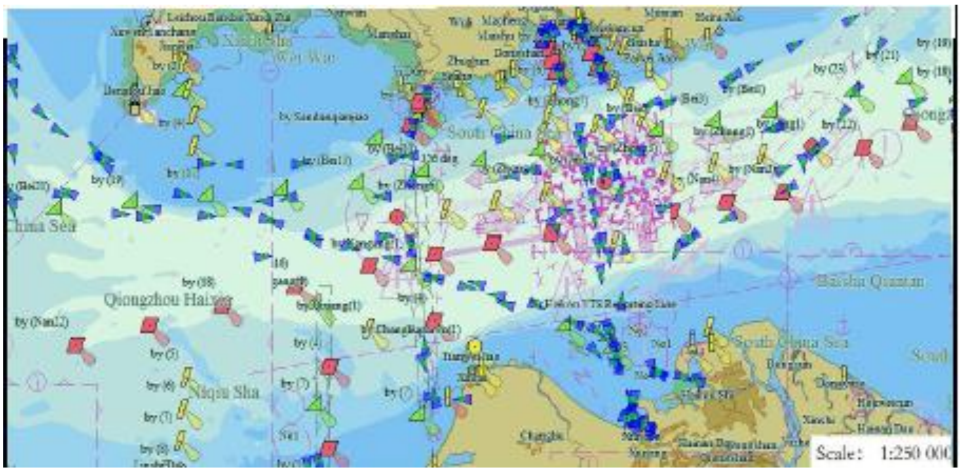


Figure 4.6 Clustered Turning Nodes in Qiongzhou Strait.

Note：A turning node denotes a cluster of way points and its position is derived by averaging all the positions' corresponding information

4.3 Application Example of Automatic Route Planning

The isolated route way points can be obtained by adopting the route way points with hierarchical DBSCAN algorithm. The main content of this section is to determine the connectivity between the two route way points by using the compressed trajectory, and then to determine the directed graph composed of the route way points and the connectivity between them.

In essence, the connectivity between the two route way points is determined by the union of the trajectory way points that make up the respective route way points_ If the union of MMSI is not empty, it means that there are two route way points, and the two route way points may be connected. Construction of two waypoint_ The union of MMSI is not empty, which is only a necessary and insufficient condition for the connectivity of the two waypoints. In addition, it is necessary to determine whether the two waypoints can be directly connected.

In the actual operation of route turn point connectivity identification, all route turn points are identified by construction MMSI searches for the index. SQL is used to extract all the route way points of a specified MMSI and store them in a list or array container. Then, starting from any one of the way points, another route way point nearest to it in space is found for connectivity identification. Assuming that the distribution of the two selected route way points is TN1 and tn2, and the heading of the constant direction line composed of TN1 and tn2 is C_{TN}, the connectivity between TN1 and tn2 is determined by the following criteria:

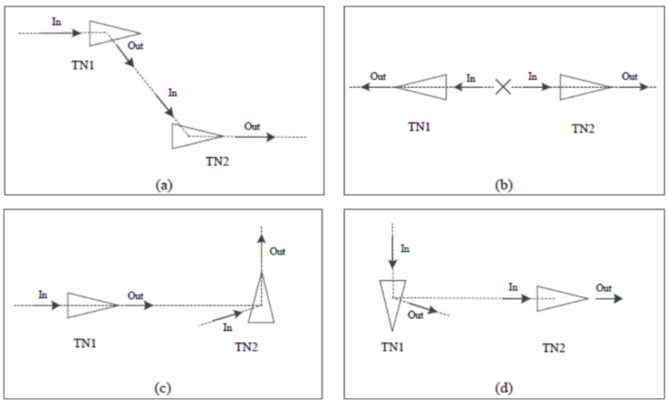


Figure 4.7 Identification of the connection between two turning nodes

（1） *TN1. Direction_Out is similar to TN2.Direction_In,* as shown in 4.7（a）；

or *TN1. Direction_Out* is similar to C_{TN}， as shown in 4.7（c）； or *TN2.*

Direction_In is similar to CTN， as shown in 4.7（d， we need to guarantee |*TN1.Direction_out–TN2.Direction_Out*|≤180°； otherwise， as shown in 4.7（b）， alyhough *TN2.Direction_In is similar to* C_{TN}. But in fact, there is no connectivity between the two waypoints.

If any of the above criteria is met, the two waypoints are considered to be connected. The connectivity of other route way points in the container is identified in turn; After that, SQL is used to extract all route way points including another specified MMSI, and the above steps are repeated to identify the connectivity between route way points until all MMSI in the trajectory database are extracted and identified. After identifying the connectivity between the route way points, a directed graph composed of the route way points and their connectivity can be obtained: g = (TN, l), where TN represents the isolated route way points and l represents the connectivity between the route way points. The example routes generated by the automatic route planning method proposed in this section are shown in Figure 4.8 and Figure 4.9. 4.10. As shown in Figure 4.8 and Figure 4.9, it can be seen that the automatically generated routes conform to the basic principles and requirements of route planning, and meet the requirements of ships using the traffic separation system. In Figure 4.8, the route way points of automatically generated routes are all in the turning area; In Figure 4.9, the automatically generated route is consistent with the approach channel and can be extended to the berth.

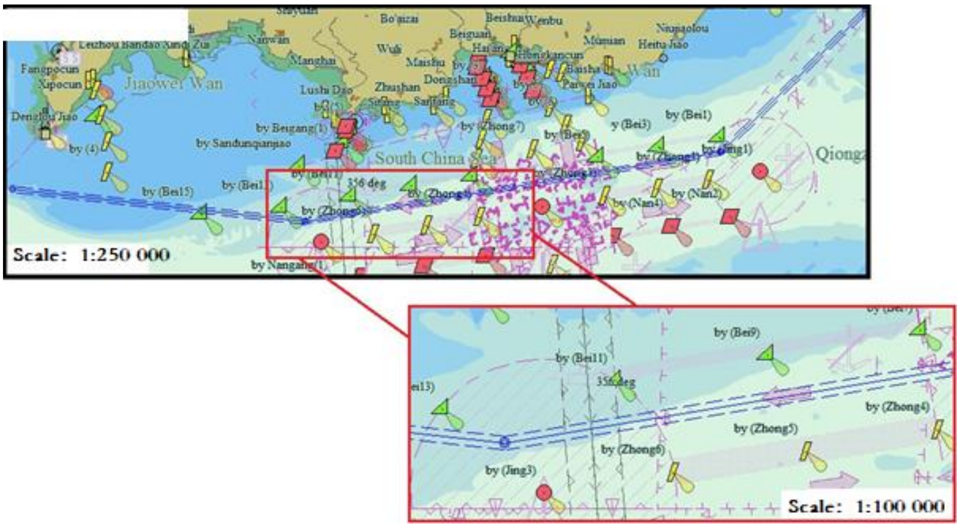


Figure 4.8 Example Route1 Generated Using the Automatic Routing Algorithm

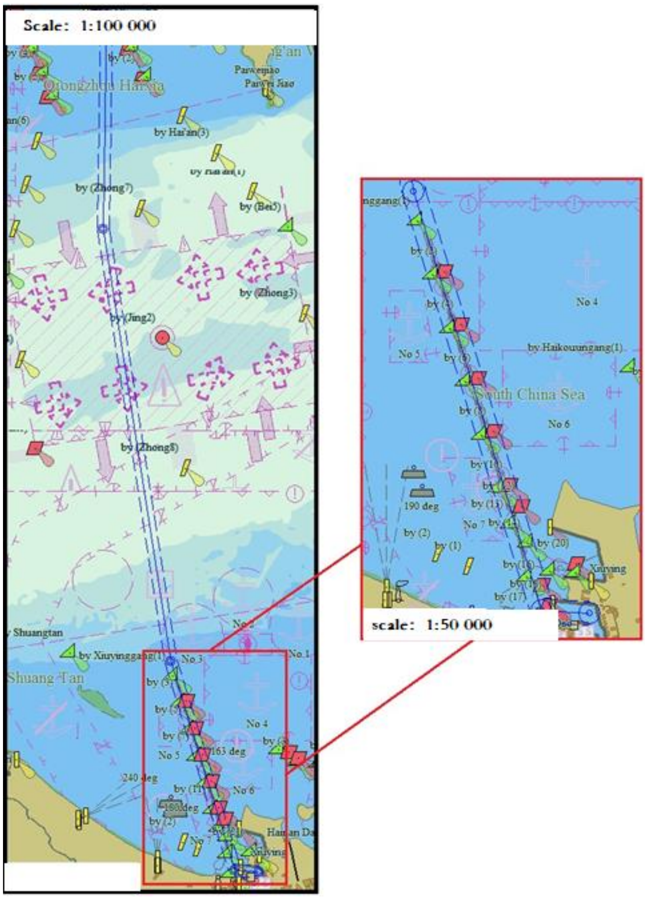


Figure 4.9 Example of route planning using the automatic routing algorithm

4.4 Summary of this Chapter

This chapter mainly proposes a trajectory data clustering algorithm based on hierarchical DBSCAN algorithm and an automatic route planning method. The main work of this chapter includes the following four aspects.

(1) According to the International Regulations for Collision Avoidance and the requirements of good ship craft, the threshold standard of trajectory way point is established, and the identification method of trajectory way point is proposed. According to the proposed method, 962 compressed trajectory data collected by Qiongzhou Strait VTS center from July 2, 2011 to July 12, 2011 are identified in turn;

(2) Considering the spatial proximity and steering similarity, the phase of trajectory way point of similarity measure standard is established. According to the different track density and way point density of different routes, a hierarchical DBSCAN algorithm is proposed based on the simplicity and efficiency of the original DBSCAN algorithm, and the algorithm is used to cluster the way points of different routes, so as to identify the way points of similar routes;

(3) Based on the fact that the connectivity between the route way points is determined by the union of the trajectory way points which constitute the respective route way points, the compressed trajectory is used to determine the connectivity between the two route way points, and the directed graph G = (TN, l) composed of the route way points and the connectivity between them is determined, where TN represents the isolated route way points, L is the connectivity between the route way points;

(4) According to the characteristics of digraph, the framework and process of optimal

path search based on ACO are designed; Finally, two example routes are used to illustrate that the route automatic generation method proposed in this section abides by complex route planning principles, and the macroscopic situation of traffic flow depicted by route and original trajectory data is consistent, which proves that the route automatic generation method proposed in this section is effective.

Chapter 5 Conclusion and Prospects

5.1 Conclusion

In this paper, we considered the unmanned ship as the research object, and we adopted the trajectory analysis technology and processing method based on data. We also studied the problem of automatic route planning. The main research contents can be summarized as following.

(1) This paper summarizes the research on unmanned ship. With the continuous improvement of the level of ship intelligence and the rapid development of science and technology, the realization of unmanned ship has the hardware foundation and scientific and technological support. Especially in the background of the rapid development of unmanned technology such as unmanned vehicle and UAV, the research and development technology of unmanned ship is constantly promoted, and the unmanned ship has a realistic possibility of sailing in the sea in the future. By reviewing the development process of unmanned ships, this paper introduces the key technologies involved in the research of unmanned ships. We look forward to the application prospect and possible development trend of unmanned ships.

We comprehensively studied data-driven trajectory analysis and processing technology. Data-driven is a data-centric processing and application mode. The convergence of modern information technology and traditional economic society has produced massive amounts of data. Under the background of the rising of data science, data-driven trajectory analysis and processing technologies are gradually improved. Data-driven analysis and research methods are profoundly changing the exploration methods of traditional scientific research. By introducing the connotation of data-driven, this paper summarizes the data-driven trajectory analysis and processing, and then introduces the application of data-driven trajectory analysis and processing technology in the maritime field with three examples. Finally, it looks forward to the potential breakthrough fields and possible research directions of data-driven trajectory analysis and processing technology in the future.

(3) The compression algorithm and threshold of trajectory data are studied in this paper. Trajectory data compression is one of the important steps and key technologies of data-driven automatic route planning for unmanned ships. By designing the framework and process of trajectory data compression, Douglas-Peucker algorithm, which simplifies the linear elements, is applied to the trajectory data compression. In order to determine the compression threshold, the concept of ship domain is used for reference, and the minimum ship domain is obtained based on the trajectory data as the standard to determine the compression threshold.

(4) Trajectory data clustering and automatic route planning based on ACO algorithm are studied. After preprocessing the massive trajectory data, the way points in the trajectory feature points are identified based on the results of trajectory data compression. Then, the similarity measurement standard of trajectory way points is established from the two aspects of spatial proximity and steering similarity, and the improved DBSCAN clustering method is used to cluster the way points in different tracks, so as to identify the route way points formed by similar tracks. Further more, the connectivity between the route way points is judged by the real track, and the directed graph G = (TN, l) composed of the way points and the connectivity between them is determined. It is proved that the route generated by the automatic route planning method proposed in this paper meets the basic principles and requirements of route planning, and the macroscopic situation of traffic flow described by the route and the original trajectory data is consistent, which proves that the automatic route planning method proposed in this paper is effective.

5.2 Research prospects

In this paper, the automatic route planning method of unmanned ship based on data-driven has been studied, and some research results have been obtained. In future research, the following aspects need to be comprehensively considered:

(1) The data-driven automatic route planning method for unmanned ships has been further improved. This paper proposes an automatic route planning method for unmanned ships, which is based on data-driven and massive trajectory data. In this paper, we also verify the feasibility of the method. However, more factors need to be considered in the practical application of the automatic route planning method proposed in this paper. For example, combined with the performance parameters of the UAV itself, integrated wind, wave, current and other meteorological factors.

(2) The method of data-driven automatic route planning for unmanned ship is further developed. In this paper, we use hierarchical DBSCAN algorithm to collect trajectory data. The two parameters of this method are cluster radius, threshold of domain density and similarity measurement standard of way point. These three parameters are the most optimal parameters selected through a large number of experiments. In order to improve the autonomy and robustness of the automatic route planning method proposed in this paper, the next research direction is to determine the setting standard of parameters according to different trajectory data and select the optimal parameters.

REFERENCES

Aguiar, A. P., & Hespanha, J. P. (2004, 2004). Logic-based switching control for trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty.

Ahmed, M., Fasy, B. T., Hickmann, K. S., & Wenk, C. (2015, 2015-08-13). A Path-Based Distance for Street Map Comparison. *ACM Transactions on Spatial Algorithms and Systems, 1*(1), 1-28. https://doi.org/10.1145/2729977

Andersson, P., & Ivehammar, P. (2017, Dec). Dynamic route planning in the Baltic Sea Region - A cost-benefit analysis based on AIS data [Article]. *Maritime Economics & Logistics, 19*(4), 631-649. https://doi.org/10.1057/mel.2016.18

Bakker, H., Dunke, F., & Nickel, S. (2019). A structuring review on multi-stage optimization under uncertainty: Aligning concepts from theory and practice. *Omega*. https://doi.org/10.1016/j.omega.2019.06.006

Bibuli, M., Bruzzone, G., Caccia, M., Indiveri, G., & Zizzari, A. A. (2008). Line following guidance control: Application to the Charlie unmanned surface vehicle. IEEE/RSJ International Conference on Intelligent Robots & Systems,

Breithaupt, S. A., Copping, A., Tagestad, J., & Whiting, J. (2017, Mar). Maritime Route Delineation using AIS Data from the Atlantic Coast of the US [Article]. *Journal of Navigation, 70*(2), 379-394. https://doi.org/10.1017/s0373463316000606

Chakraborty, S., & K, N. K. N. N. (2014). Analysis and Study of Incremental DBSCAN Clustering Algorithm. *Eprint Arxiv, 1*(2), 2011.

Cheng, L., Yan, Z., Xiao, Y., Chen, Y., Zhang, F., & Li, M. (2018). Using big data to track marine oil transportation along the 21st-century Maritime Silk Road. *Science China Technological Sciences, 62*(4), 677-686. https://doi.org/10.1007/s11431-018-9335-1

Christiansen, M., Fagerholt, K., Nygreen, B., & Ronen, D. (2013). Ship routing and scheduling in the new millennium. *European Journal of Operational Research, 228*(3), 467-483. https://doi.org/10.1016/j.ejor.2012.12.002

Douglas, D. H., & Peucker, T. K. (2011). *Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature*. Classics in Cartography.

IMO. (2017). International Convention for Safety of Life at Sea.

Jan, R. (2013). Maritime unmanned navigation through intelligence in networks: The MUMIN project. *Unmanned Ship*, 177-183.

Lee, E., Mokashi, A. J., Moon, S. Y., & Kim, G. (2019). The Maturity of Automatic Identification Systems (AIS) and Its Implications for Innovation. *Journal of Marine Science and Engineering, 7*(9). https://doi.org/10.3390/jmse7090287

Liao, Y. L. (2015). Serret-Frenet frame based on path following control for underactuated unmanned surface vehicles with dynamic uncertainties. *Journal of Central South University*(22), 223.

Lyu, H. G., & Yin, Y. (2019, May). COLREGS-Constrained Real-time Path Planning for Autonomous Ships Using Modified Artificial Potential Fields [Article]. *Journal of Navigation, 72*(3), 588-608. https://doi.org/10.1017/s0373463318000796

Ma, W. H., Lu, T. F., Ma, D. F., Wang, D. H., & Qu, F. Z. (2019). Ship route and speed multi-objective optimization considering weather conditions and emission control area regulations [Article; Early Access]. *Maritime Policy & Management*, 16. https://doi.org/10.1080/03088839.2020.1825853

Matui, T., Fujii, Y., & Yamanouchi, H. (1981). SURVEY ON VESSEL TRAFFIC MANAGEMENT SYSTEM. *Surveys*.

Meng, X., & Ci, X. (2013). Big Data Management: Concepts, Techniques and Challenges. *Journal of Computer Research And Development, 50*(1), 146-169. https://d.wanfangdata.com.cn/periodical/ChlQZXJpb2RpY2FsQ0hJTmV3UzIwMjEwNjE2EhFqc2p5anlmejIwMTMwMTAxNBoIYmxrNm04a3g%3D

Miskovic, N., Vukic, Z., Bibuli, M., Caccia, M., & Bruzzone, G. (2009, 2009). Marine vehicles' line following controller tuning through self-oscillation experiments.

Pallero, & Lg, J. (2013). Robust line simplification on the plane. *COMPUTERS AND GEOSCIENCES*.

Peng, P., Yang, Y., Cheng, S., Lu, F., & Yuan, Z. (2019). Hub-and-spoke structure: Characterizing the global crude oil transport network with mass vessel trajectories. *Energy, 168*, 966-974. https://doi.org/10.1016/j.energy.2018.11.049

Poikonen, S., & Golden, B. (2020, Spr). The Mothership and Drone Routing Problem. *Informs Journal on Computing, 32*(2), 249-262. https://doi.org/10.1287/ijoc.2018.0879

Shrivastava, P., & Gupta, H. (2012). A Review of Density-Based clustering in Spatial Data. *international journal of advanced computer research*.

Shu-kai, Z. (2016). *Data-driven Based Automatic Routing for Unmanned Ship and Surface Vehicle* Dalian Maritime University].

Singh, Y., Sharma, S., Sutton, R., Hatton, D., & Khan, A. (2018, 2018-12-01). A constrained A* approach towards optimal path planning for an unmanned surface vehicle in a maritime environment containing dynamic obstacles and ocean currents. *Ocean Engineering, 169*, 187-201. https://doi.org/10.1016/j.oceaneng.2018.09.016

Song, L. (2014). Path Planning Research of Unmanned Surface Vehicle Based on Electronic Chart. *Journal of Information and Computational Science, 11*(17), 6245-6254.

UKHO. (2018). Admiralty IALA Maritime Buoyage System.

Wang, J., Xiao, Y., Li, T. S., & Chen, C. L. P. (2020). A Survey of Technologies for Unmanned Merchant Ships. *Ieee Access, 8*, 224461-224486. https://doi.org/10.1109/access.2020.3044040

Wang, L. P., Zhang, Z., Zhu, Q. D., & Ma, S. (2020). Ship Route Planning Based on Double-Cycling Genetic Algorithm Considering Ship Maneuverability Constraint. *Ieee Access, 8*, 190746-190759. https://doi.org/10.1109/access.2020.3031739

Wang, N., Zhu, Z. B., Qin, H. D., Deng, Z. C., & Sun, Y. C. (2021, Mar). Finite-time extended state observer-based exact tracking control of an unmanned surface vehicle. *International Journal of Robust and Nonlinear Control, 31*(5), 1704-1719. https://doi.org/10.1002/rnc.5369

Wang, Q. W., Zhong, M., Shi, G. Y., Zhao, J., & Bai, C. J. (2021). Route Planning and Tracking for Ships Based on the ECDIS Platform [Article]. *Ieee Access, 9*, 71754-71762. https://doi.org/10.1109/access.2021.3078899

Wang, Y., Liang, X., Li, B., & Yu, X. (2018). Research and Implementation of Global Path Planning for Unmanned Surface Vehicle Based on Electronic Chart. In *Advances in Intelligent Systems and Computing* (pp. 534-539). Springer International Publishing. https://doi.org/10.1007/978-3-319-65978-7_80

Wu, P., Xie, S., Luo, J., Qu, D., & Li, Q. (2015). The USV path planning based on the combinatorial algorithm. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia, 38*(1), 62-70.

Xu, J., Li, Y., Zhou, Y., Bai, C., & Li, B. (2016). Analysis on specification of S-101 electronic navigational chart products [S-101 电子海图产品规范解析] [Article]. *Science of Surveying and Mapping, 41*(3), 150-155, Article 1009-2307(2016)41:3<150:S1dzht>2.0.Tx;2-q. <Go to ISI>://CSCD:5673416

Yanıkoğlu, İ., Gorissen, B. L., & den Hertog, D. (2019). A survey of adjustable robust optimization. *European Journal of Operational Research, 277*(3), 799-813. https://doi.org/10.1016/j.ejor.2018.08.031

Z, P., & J, U. (2009). The Ship Domain ? A Criterion of Navigational Safety Assessment in an Open Sea Area. *Journal of Navigation*.

Zhang, W., Yan, C., Lyu, H., Wang, P., Xue, Z., Li, Z., & Xiao, B. (2021, 2021). COLREGS-based Path Planning for Ships at Sea Using Velocity Obstacles. *Ieee Access, 9*, 32613-32626.

https://doi.org/10.1109/access.2021.3060150

Zhang, X. Y., Ji, M. J., Yao, S., & Chen, X. (2015). Optimising Feeder Routing for Container Ships through an Electronic Chart Display and Information System. *Journal of Navigation, 68*(5), 1-21.

Zhao, L. B., Shi, G. Y., & Yang, J. X. (2018, Sep). Ship Trajectories Pre-processing Based on AIS Data [Article]. *Journal of Navigation, 71*(5), 1210-1230. https://doi.org/10.1017/s0373463318000188

Zhen, L., Liang, Z., Zhuge, D., Lee, L. H., & Chew, E. P. (2017, Dec). Daily berth planning in a tidal port with channel flow control [Article]. *Transportation Research Part B-Methodological, 106*, 193-217. https://doi.org/10.1016/j.trb.2017.10.008

Zhou, C. H., Gu, S. D., Wen, Y. Q., Du, Z., Xiao, C. S., Huang, L., & Zhu, M. (2020, Mar). The review unmanned surface vehicle path planning: Based on multi-modality constraint. *Ocean Engineering, 200*, Article 107043. https://doi.org/10.1016/j.oceaneng.2020.107043

Zou, C., Tsui, J., & Peterson, J. B. (2018). Correction to: The publication trajectory of graduate students, post-doctoral fellows, and new professors in psychology [Correction Notice]. *Scientometrics, 117*(2), 1311-1311.