

Research Article

Anomaly Detection and Restoration for AIS Raw Data

Shuguang Chen ¹, Yikun Huang ², and Wei Lu³

¹Department School of Management, Xi'an University of Finance and Economics, Xi'an 710100, China

²Concord University College of Fujian Normal University, Fuzhou 350117, China

³School of Information, Xi'an University of Finance and Economics, Xi'an 710100, China

Correspondence should be addressed to Yikun Huang; fjnuhyk@163.com

Received 18 December 2021; Accepted 19 February 2022; Published 30 March 2022

Academic Editor: Pei-Wei Tsai

Copyright © 2022 Shuguang Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the wide application of location detection sensors in maritime surveillance, a large amount of raw automatic identification system (AIS) data is produced by many moving ships. Anomaly detection and restoration of the big AIS data are important issues in marine data mining, because they offer a reliable support to users to mining the behaviors of ships. This paper develops a novel approach to detect anomaly AIS data based on the ships' maneuverability, such as the maximum acceleration, the minimum acceleration, the maximum distance, and the maximum angular displacement, which were designed to detect the anomaly AIS data. Furthermore, the performance of the developed approach is compared with that of Daiyong-Zhang's method and Behrouz-Haji-Soleimani's method to assess its detection efficiency. The results show that the proposed approach can be applied to easily extract the abnormal data. Finally, based on the developed approach to detect the anomaly data and cubic spline interpolation method to restore the AIS data, experiments are conducted on the AIS data of Xiamen Port of Fujian Province, China, that prove to be effective for marine intelligence research.

1. Introduction

With the rapid development and maturity of positioning technology, communication technology, and network technology, various types of mobile intelligent terminals with positioning and navigation functions are becoming more and more widely used, and the location of mobile objects (including people, vehicles, ships, and animals) relevant information is increasingly accessible and can be collected on a large scale. This type of location data usually contains information such as geographic coordinates, speed, direction, and time, and it continues to increase and update rapidly over time [1]. It is called trajectory big data. Given that trajectory big data records the movement of moving objects over time and can objectively reflect the activities of individuals or groups of moving objects, as well as their impact on the environment, it has led to the concern of scholars in various fields such as natural sciences, social sciences, and environmental science [2–4]. With the rapid increase on international trade, an increasing number of vessels have

been come into service; as a result, the safety and security of marine transportation have become the most dominating attention of marine surveillance. Since 2002, the International Maritime Organization (IMO) requires automatic identification system (AIS) transponders to be aboard vessels that are above 300 gross tonnages on international voyages, cargo ships over 500 gross tonnages in all waters, and all the passenger ships regardless of size [5]. The AIS tracks vessel movement by means of electronic exchange of navigation data between vessels, with onboard transceiver, terrestrial, and satellites. This navigation data is related to the ship itself (including its static parameters and dynamic activity records, such as ship name, ship maritime mobile service identity, ship size, ship type, speed, location, time, heading, and rate of turn) and includes other features such as the sea state. For all reasons mentioned above, a massive amount of AIS data is produced. The use of this massive AIS data is an important part of intelligent marine transportation system and conducts research on ship collision avoidance [6–8], ship behavior analysis [9, 10], ship emission analysis,

trajectory analysis [7, 10–21], maritime surveillance [10, 11, 20, 22–24], accident investigation [8, 25], etc. Anomaly detection and restoration [26, 27]–[20, 23, 24, 28] are the fundamental key research problems in the marine intelligent transport system, which aims to identify and restore the abnormal data in the AIS data generated by the users through multiple aboard transceivers.

The identification and restoration of anomaly AIS data play a vital role in the intelligent analyses of AIS data, because User Datagram Protocol (UDP) is adopted for the AIS data packet transmission, during which packet-disordering and data packet dropouts occur. Another related reason deals with the quality of the raw AIS data, e.g., error and anomaly, and it is well known that the raw AIS data may be tampered to inform false types of movements, such as fishing activity in protected areas. Consequently, the error and lost AIS data will interfere with maritime management due to the misjudgment of the maritime state. Besides, it will decrease the effectiveness of analysis on ship behavior and traffic flow based on the AIS data.

Recently, many studies have focused on using techniques to detect and restore the anomaly AIS data based on an optimal trajectory calculated from classification algorithms, which designed a ranking score inventory considering the difference between the optimal trajectory and the real ones, but lots of existing works ignored the impact of ship speed or just considered the impact of ship's locations on the optimal trajectory [28]. Besides, with regard to the data veracity of ship's movement features, some works [29] focused on the threshold of ship movement features based on the ship's navigation data, such as ship speed, ship location over time, and ship heading. To deal with the anomaly detection of AIS data, the current methods proposed in the literature can mainly be classified into the following two categories as follows:

- (i) One is to design a near-optimal path to evaluate the matching between the real trajectory and the near-optimal path, such as Behrouz-Haji-Soleimani's graph search algorithm [28], classification methods, or clustering algorithm [7, 14, 20, 30]. However, these methods need to predict the main route used by ships
- (ii) Another one is to design the rules to detect the unreasonable track points based on ships' design specifications [23, 24, 29]. Nevertheless, both the unreasonable drift track points with slow speed and the unreasonable acceleration points under certain limit can hardly be detected

In this work, Daiyong-Zhang's method is generalized and extended its unreasonable acceleration from the maximum acceleration to the minimum acceleration. It must be mentioned that this paper aims to compensate the detection of drift track point when its average speed does not exceed the maximum speed, while [29] only considers to detect the drift track point when its average speed exceeds the maximum speed. In order to avoid a misjudgment of the drift track point, the proposed mode uses speed integral to obtain the maximum drift distance in moving. Besides, the detec-

tion of unreasonable turn point is simplified from the detection rate of turn in [29] to the measurement for the angular displacement of turn. Considering characteristic of the packet dropouts and data cleaning, the cubic spline interpolation method is employed to compensate the consecutive of the trajectories, which aims to minimize the accelerate vector along longitude, latitude, and velocity in the AIS data. Case studies over the AIS data sets are carried out to verify the effectiveness of the proposed method.

In conclusion, the main contributions of this paper are summarized as follows:

- (i) This paper proposes a model to detect the anomaly drift track point, which builds a maximum distance and a minimum distance between the drift point and its adjacent track points
- (ii) The minimum acceleration of the ship is modeled by using the design specifications of ships, such as the distance for a ship to decelerate from the design speed to zero. Moreover, the maximum acceleration of the ship is investigated for the detection of unreasonable acceleration
- (iii) An efficient and effective model which is just only based on the difference of heading between the drift point and its adjacent track points is proposed to detect the unreasonable track point of turn
- (iv) The 156-AIS data with a cruise of length 110 m and at interval of 10 s in December 22, 2018, from Xiamen International Cruise Center is employed to verify the effectiveness of the proposed detection models. The simulation results demonstrate that the number of anomaly AIS data by using our method is less than Daiyong-Zhang's method, but our proposed method is superior than Daiyong-Zhang's method when anomaly drift track point has slow speed. Besides, the simulation results also present that Behrouz-Haji-Soleimani's method can also be effective for discriminating the anomaly state of the objected trajectory, but it needs lots of trajectories to build the optimal trajectory
- (v) For a case study of a passenger ship (call sign: 6285, ship name: MIN LONG YU 9 777, MMSI: 412596777, ship's length is 19 m, and transmission time interval is 10 s) in December 22, 2018–January 3, 2019, from Xiamen Port, experiments show that a new smooth and reasonable trajectory is reconstructed based on the combination method of our proposed detection approach with the cubic spline restore algorithm

The remainder of this paper is organized as follows. The detection modeling of anomaly activity, such as stop state, acceleration, drift track point, and turn, is presented in Section 2. Section 3 depicts cubic spline interpolation method for data restoration. In Section 4, case studies are conducted and experimental results are shown. Section 5 concludes the paper.

2. Classification and Determination of Outliers in Ship AIS Trajectory Data

2.1. AIS Trajectory Representation. Trajectory is an important type of spatiotemporal data. It is used to represent the history and continuous state information of a moving object changing with time. It can also be considered as a time-to-state mapping. In other words, given a time $t (t \in R^+)$, the state space of moving target at time t can be obtained by using a continuous function F of time t . For a d -dimensional state space vector, the mapping can be expressed as $F : R^+ \rightarrow S^d$. A trajectory T_r of a ship is a finite set $T_r = \{(s_1, t_1), (s_2, t_2), \dots, (s_n, t_n)\}$, where $s_i \in [1, n]$ is a state at a trajectory point and $t_i \in [1, n]$ is a timestamp at a trajectory point. Yet, traditional studies on state space of ship trajectory only consider position information of the targets. In order to identify the anomaly AIS data accurately, longitude, latitude, speed, and heading are taken into account when ship's state space is studied in this paper.

2.2. Classification and Judgment of Abnormal Points of Ship Trajectory. By analyzing the state space of the raw AIS data sets provided by the VTEExplorer website, such as changes in longitude, latitude, speed, and heading, rules to identify the inaccurate AIS data according to the ship's maneuverability are shown in the following subsection.

2.2.1. Abnormal Stop. According to the message format specified by the International Telecommunication Union communication standard, AIS transponders may accept a duplicate message in the packet forwarding mechanism. In another word, two adjacent AIS trajectory points are almost the same with each other except their timestamps, i.e., $(s_i = s_j, t_i \neq t_j, i \neq j, i, j \in [1, n])$. If these duplicate or abnormal messages are not processed, the false judgement of ship's running states as its stop states will occur. With regard to the detection of these abnormal messages, the determination method of the abnormal stopping point can be given as follows. For the AIS sequence, if the speed of the i -th point is greater than 2 knots (1 knot = 1 nautical mile/hour), but its geographic coordinates (lon_i and lat_i), speed v_i , and heading cog_i are the same as that of the $i+1$ -th point, then the $i+1$ -th point is judged as an abnormal stopping point. The judgement rules are shown in the following:

$$\begin{cases} v_i > 2, \\ |\text{lon}_{i+1} - \text{lon}_i| = 0, \\ |\text{lat}_{i+1} - \text{lat}_i| = 0, \\ |v_{i+1} - v_i| = 0, \\ |\text{cog}_{i+1} - \text{cog}_i| = 0. \end{cases} \quad (1)$$

2.2.2. Abnormal Acceleration Point. According to the current design standards for ships, the distance between the stationary and the design speed trajectory points for a ship in full load condition is about 20 times of the length of the ship, while the distance between the stationary and the design

speed trajectory points for a ship in no load condition is reduced to 1/2~2/3 times of the original distance. The stopping stroke of a ship, which is affected by its displacement, is generally 8-20 times of the length of the ship. As is shown in equation (2), to obtain the maximum and the minimum acceleration, the minimum distance equals to 10 times and 8 times of the length of the ship, respectively. Assuming that the length of the ship is L , the design speed is V_d , the maximum acceleration is a_{\max} , the minimum acceleration is a_{\min} , the time interval from the stationary to the design speed V_d with a uniformly accelerating is $t_{i \max}$, and the time interval from the design speed V_d with a uniformly decelerating to the stationary is $t_{i \min}$.

$$\begin{cases} V_d = a_{\max} \times t_{i \max} = -a_{\min} \times t_{i \min}, \\ 10L = \frac{1}{2} a_{\max} t_{i \max}^2, \\ 8L = \frac{1}{2} a_{\min} t_{i \min}^2. \end{cases} \quad (2)$$

Based on equation (2), the maximum acceleration a_{\max} and the minimum acceleration a_{\min} of the ship can be obtained as follows.

$$\begin{cases} a_{\max} = \frac{V_d^2}{2 \times 10L}, \\ a_{\min} = -\frac{V_d^2}{2 \times 8L}. \end{cases} \quad (3)$$

After the speed and time difference between the i -th point and the $i+1$ -th point, transient acceleration at time $i+1$ -th can be derived by (4). If the calculated transient acceleration is greater than the maximum acceleration or less than the minimum acceleration, then the $i+1$ -th point is an abnormal acceleration point.

$$\begin{cases} a = \frac{v_{i+1} - v_i}{t_{i+1} - t_i}, \\ (a_{\max} - a)(a - a_{\min}) < 0. \end{cases} \quad (4)$$

2.2.3. Anomalous Drift Point. Theoretically, the travelling reachability distance between two trajectory points can be obtained by the integral calculation with the speed change on the route. For a trajectory data sequence, if the distance between two adjacent AIS trajectory points exceeds their maximum reachability distance, the trajectory point is judged as an abnormal drift point. Assume that the timestamp and speed for the i -th trajectory point is t_i and v_i , respectively. At the same time, t_{i+1} and v_{i+1} represent the timestamp and speed for the $i+1$ -th trajectory point, respectively. In theory, the travelling reachability distance between the i -th and $i+1$ -th trajectory points should be $\int_{t_i}^{t_{i+1}} v dt$. But in a fact that the velocity variation pattern cannot be obtained through the measurement, the maximum reachability distance between the two trajectory points is

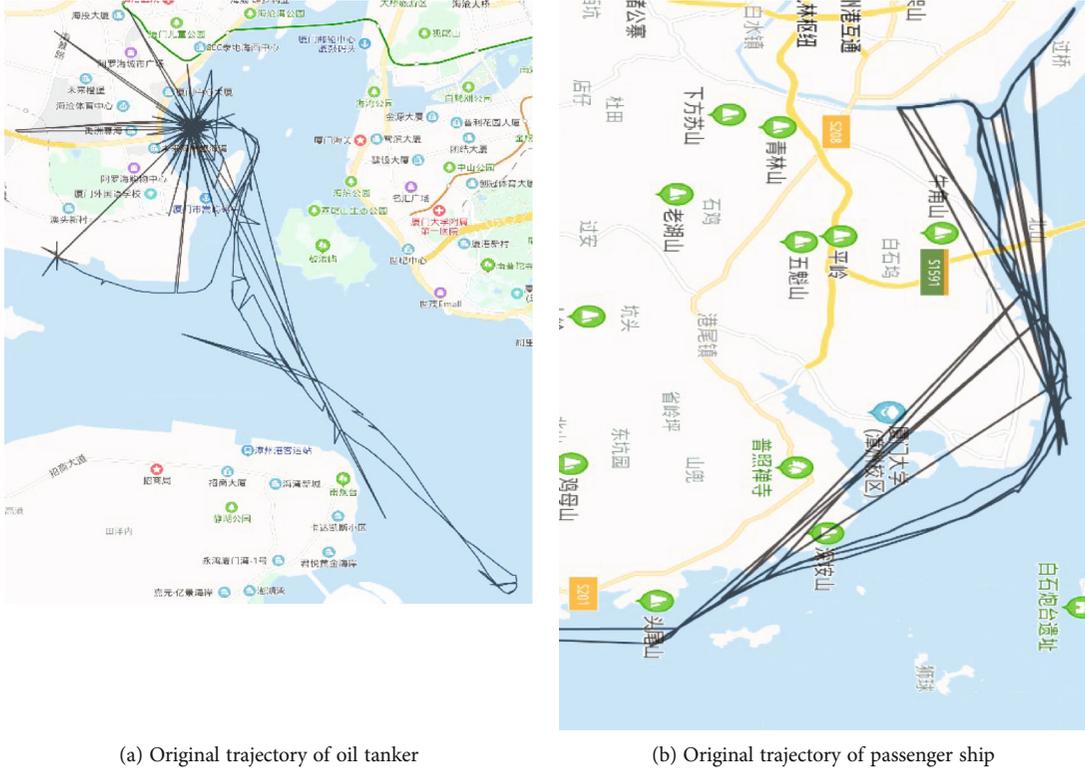


FIGURE 1: Original trajectory diagram.

estimated.

$$s_{\max} = \int_{t_i}^{t_m} (v_i + a_{\max}(t - t_i))dt + \int_{t_m}^{t_{i+1}} (v_{i+1} + a_{\min}(t_{i+1} - t))dt, \quad (5)$$

where t_m satisfies the condition of $v_{i+1} = v_i + a_{\max}(t_m - t_i) + a_{\min}(t_{i+1} - t_m)$, ship is in the uniformly accelerative motion with maximum accelerate a_{\max} in time interval $[t_i, t_m)$, and ship is in the uniformly decelerative motion with minimum accelerate a_{\min} in time interval $[t_m, t_{i+1})$.

After the integral calculation with the speed change on the route, the maximum reachability distance of the i -th trajectory point and the $i + 1$ -th trajectory point can be derived by (5). As is shown in (6), if the calculated spherical distance $dis(i, i + 1)$ between two adjacent AIS trajectory points is greater than the maximum reachability distance, then the $i + 1$ -th point is an abnormal drift point.

$$dis(i, i + 1) > \int_{t_i}^{t_m} (v_i + a_{\max}(t - t_i))dt + \int_{t_m}^{t_{i+1}} (v_{i+1} + a_{\min}(t_{i+1} - t))dt. \quad (6)$$

2.2.4. Anomalous Turning Point. As is well known, the swing diameter is an important parameter for evaluating the steering capability of a ship. Generally speaking, the maximum swing diameter d of a ship can be obtained by the formula $d = k \times l$, $k \in [2, 4]$ based on the design specification of ships, in which l is the length of ship, k is the coefficient to mea-

TABLE 1: Trajectory outlier distribution.

Trajectory point	Ship type	Quantity
Normal trajectory point	Tanker	7800
	Ship	503
Anomalous drift point	Tanker	4035
	Ship	67
Abnormal stop point	Tanker	0
	Ship	0
Abnormal acceleration point	Tanker	12
	Ship	0
Abnormal turning point	Tanker	3118
	Ship	0

sure the ship maneuverability, and usually the value range of k is $[2, 4]$. If the speed of the ship is v and the maximum swing diameter is d , then the maximum rate of turn is $r = 360v/\pi kl$; as shown in (7), the maximum angle of turn ω_{\max} from i -th trajectory point to the $i + 1$ -th trajectory point can be obtained through the integral calculation with the maximum rate of turn.

$$\omega_{\max} = \int_{t_i}^{t_{i+1}} \frac{360v}{\pi kl} dt \leq \frac{360}{\pi kl} s_{\max}. \quad (7)$$

In equation (7), s_{\max} is the maximum reachability distance between the two trajectory points.

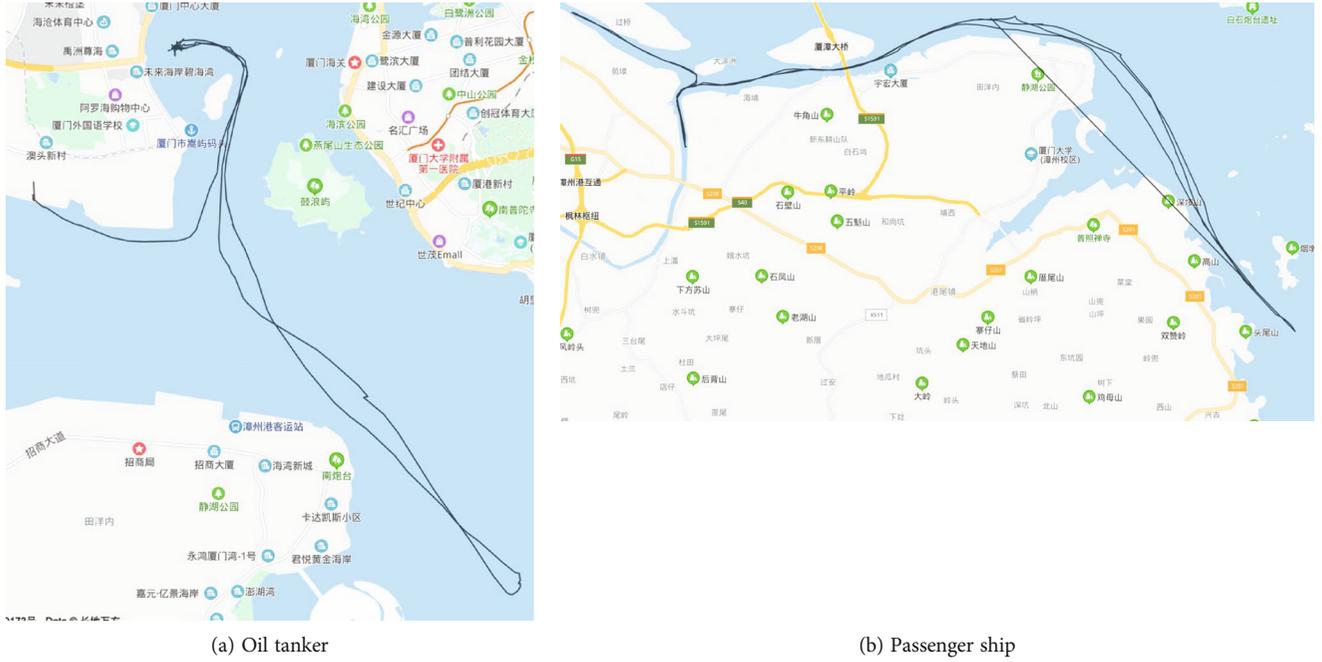


FIGURE 2: The trajectory after clearing the abnormal point.

Note that an abnormal turning point can be easily identified by using the heading difference $|\text{cog}_{i+1} - \text{cog}_i|$ between two adjacent AIS trajectory points, as shown in the following.

$$|\text{cog}_{i+1} - \text{cog}_i| > \omega_{\max} = \int_{t_i}^{t_{i+1}} \frac{360v}{\pi kl} dt \leq \frac{360}{\pi kl} S_{\max}. \quad (8)$$

3. Ship AIS Trajectory Data Repair

The ship trajectory established based on AIS data is a sequence of points including discrete space-time information. In order to satisfy the subsequent research and application based on trajectory, it is necessary to delete the abnormal points in the raw data. However, the deletion of the abnormal points will cause the trajectory sequence to become discontinuous; the loss of AIS messages can also lead to this situation. Therefore, synchronous interpolation is needed to obtain continuous trajectories in practical applications. Cubic spline interpolation is one of the commonly used methods for spatiotemporal trajectory interpolation and synchronization. In the case of spatiotemporal data with few missing or intermittent missing, the cubic spline interpolation method has good repair and synchronization effects.

According to the International Telecommunication Union communication standards, the time interval for sending navigation and location-related messages is related to the type of ship and its speed. For class A ships, the interval for sending AIS messages should be no more than 10 seconds; for class B ships, the interval should be no more than 30 seconds. In the case of normal navigation, the trajectory data that can be obtained is relatively dense. The number of discontinuity trajectory points caused by message loss and out-

lier deletion is relatively small. Therefore, the cubic spline interpolation method is suitable for trajectory point repair and synchronization. But when the ship is berthing, the AIS message sending time interval will become 3 minutes. Interpolation is not necessary in this case. In addition, the ship may also shut down its AIS radio station on its own initiative. This will lead to a long segment of missing trajectory points. In this case, the ship behavior is uncertain, so it is not suitable for interpolation too.

According to the AIS message characteristics mentioned above, we segment the trajectory data after deleting outliers. Based on the time interval and speed of the points, the trajectory is divided into normal navigation section, stop section, and closed radio section. The rules to divide segments are as follows: (1) AIS messages in which interval time is less than 3 minutes are segments of normal navigation; (2) AIS messages in which interval time is greater than or equal to 3 minutes and less than or equal to 5 minutes and speed is less than 1 knot are segments of stop; (3) AIS messages in which interval time is greater than 5 minutes are segments of turning off AIS radio. In the process of trajectory data repairing, we will only use cubic spline interpolation for the first case. For the second and third cases, interpolation is not carried out. In the first case, the interpolation is carried out according to different ship types and speeds, and the specified time interval of AIS message is used as the step size.

Suppose there are m trajectory points in the AIS spatiotemporal sequence and the corresponding times of the m points are $t_i (i = 1, 2, \dots, m)$, x_i , y_i , v_i , and θ_i that represent the longitude, latitude, speed, and heading angle of point i , then the velocity of point i in the longitude direction is $v_{xi} = v_i \cos \theta_i$, and the velocity of point i in the latitude direction is $v_{yi} = v_i \sin \theta_i$. Setting the time starting point of the

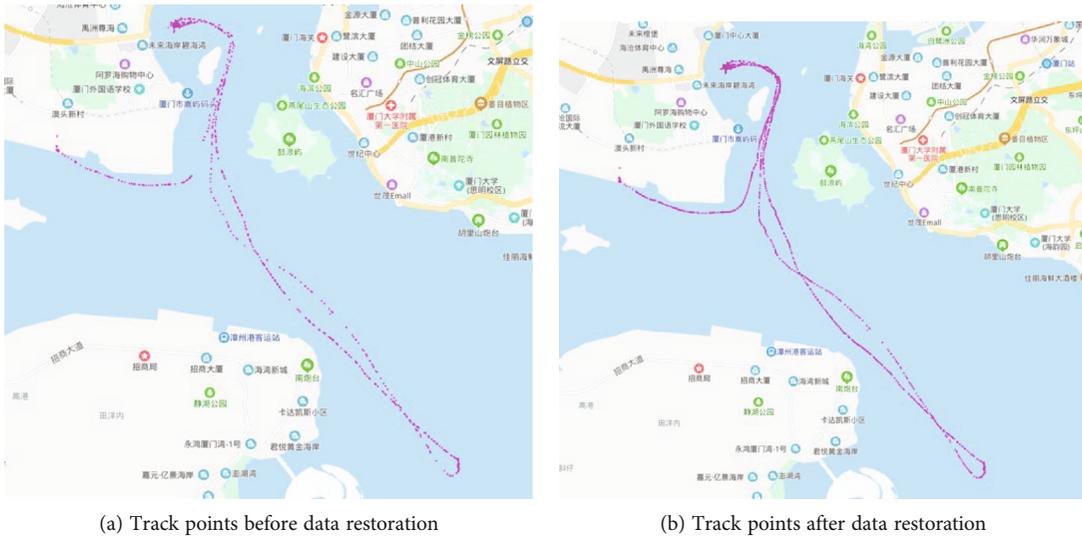


FIGURE 3: Track trajectory diagram of oil tanker.

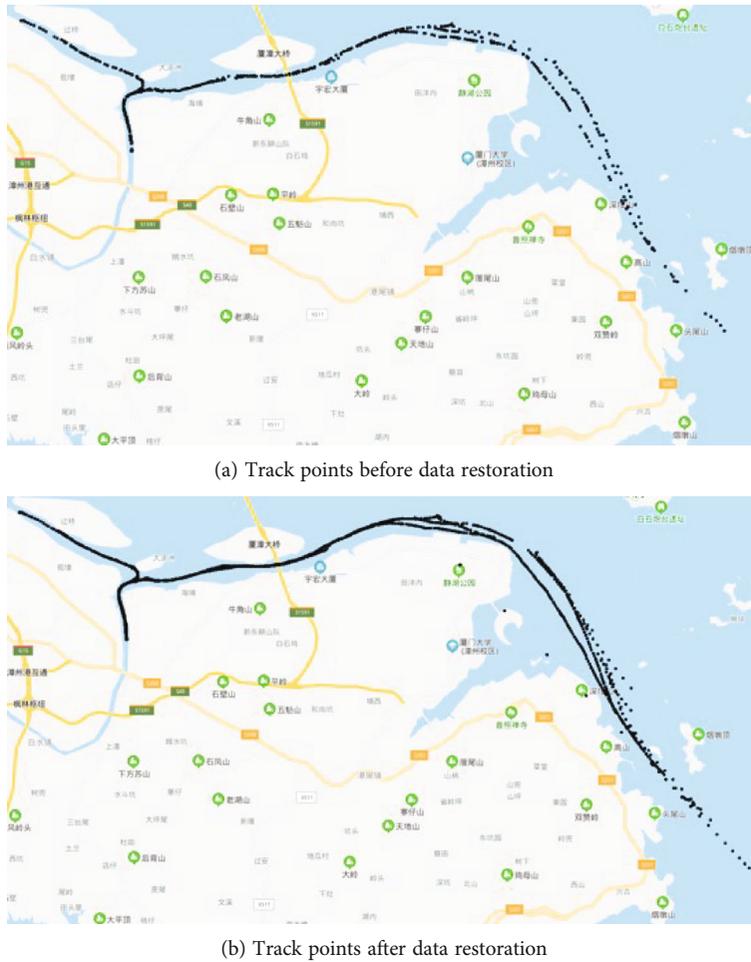


FIGURE 4: Track trajectory diagram of passenger ship.

sequence to be interpolated as zero time, and using the corresponding latitude and longitude coordinates as the origin of the coordinates, for the space-time sequence to be inter-

polated, the derivatives at the endpoints of the latitude and longitude directions are v_{xi} and v_{yi} . The coefficient matrix can be obtained by substituting v_{xi} and v_{yi} into the cubic

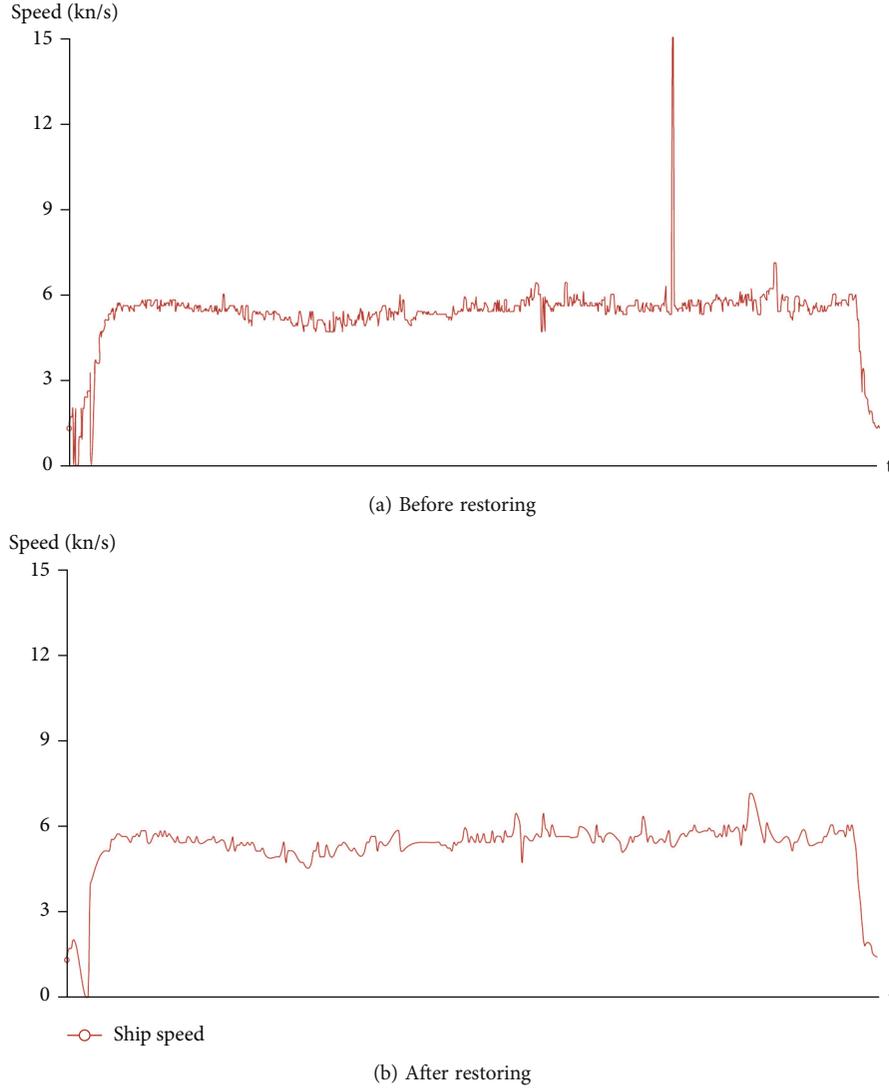


FIGURE 5: Comparison of speed changes.

spline function for each segment. Then, the coordinates of longitude and latitude corresponding to any time t can be obtained by cubic spline function in each segment.

Take longitude calculation of ships as an example. Let the longitude coordinate of ship $y(t)$ be a function of time, and satisfy the cubic spline function $y(t) = a\Delta t^3 + b\Delta t^2 + c\Delta t + d$ in the time period $[t_i, t_j]$; then, $v_y = dy(t)/dt = 3a\Delta t^2 + 2b\Delta t + c$, and $a_y = dv_y(t)/dt = 6a\Delta t + 2b$.

According to the boundary conditions, for the specific time t_1 and t_2 , it should be satisfied:

$$\begin{cases} y(t_1) = at_1^3 + bt_1^2 + ct_1 + d, \\ y(t_2) = a(t_2 - t_1)^3 + b(t_2 - t_1)^2 + c(t_2 - t_1) + d, \\ v_y(t_1) = 3at_1^2 + 2bt_1 + c, \\ v_y(t_2) = 3a(t_2 - t_1)^2 + 2b(t_2 - t_1) + c. \end{cases} \quad (9)$$

According to the boundary values $y(t_1), y(t_2), v(t_1), v(t_2)$

of the sequence to be interpolated, the spline function coefficients can be obtained. Therefore, we can obtain the function expression of longitude coordinate $y(t)$ with respect to time t . According to the interpolation time interval, we can further obtain the longitude coordinates of the corresponding trajectory points using the function $y(t)$. The interpolation methods of other parameters of the trajectory sequence (latitude, speed, and heading) are the same and will not be described again.

4. Experiment Analysis

In order to verify the effect of the method for processing and repairing the abnormal points of ship trajectory data proposed in this paper, some original AIS data were downloaded through the VTE Explorer website for experiments. We select the part of data about Xiamen Port and surrounding waters. The time range is from 11:46:37 on December 21, 2018, to 7:30:22 on January 3, 2019; the spatial range is 117.7737° E and 24.08784° N to 118.63037° E and

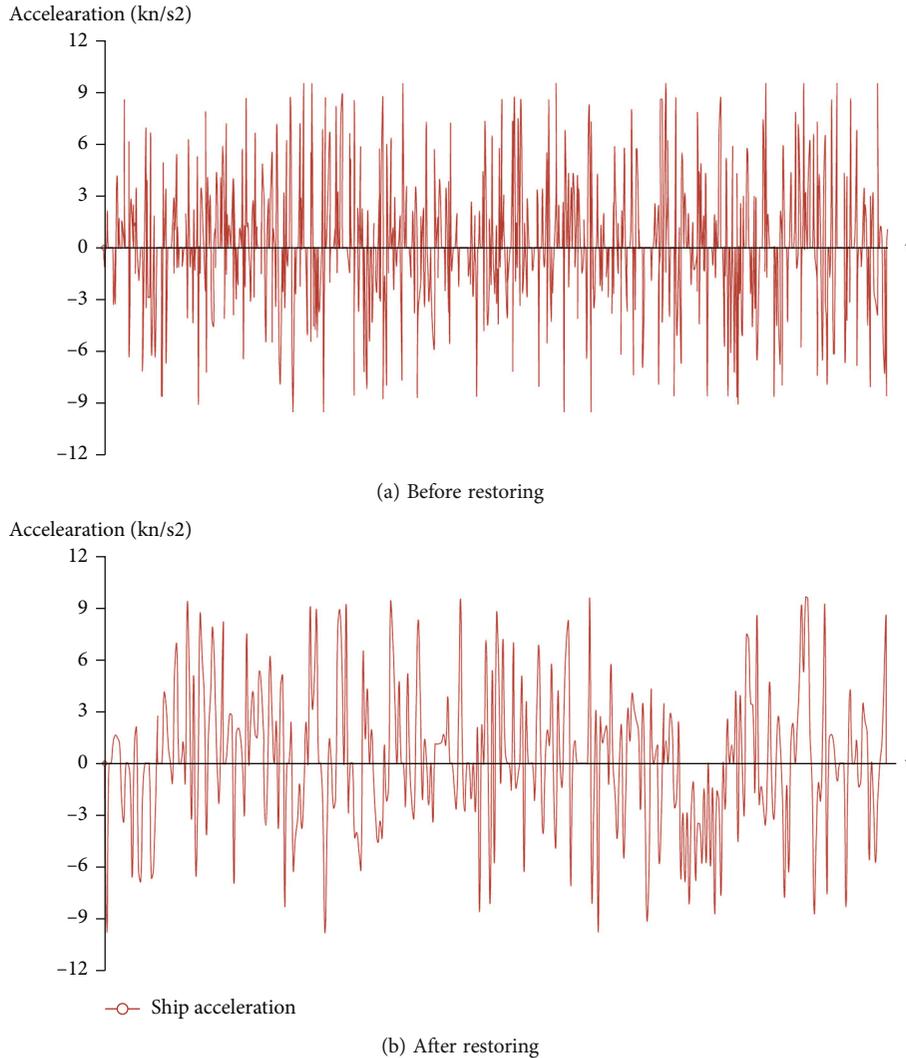


FIGURE 6: Comparison of acceleration changes.

24.691° N, including 12158622 pieces of position data and 387745 pieces of static data. The experiment selects the trajectory data of an oil tanker and a passenger ship for analysis, processing, and comparison. The MMSI number of the oil tanker is 413698470, the call sign is BVHW8, the name is HAI GONG 167, the length of the ship is 32 m, and the time span of the track point is from December 21, 2018, hours 46 minutes 41 seconds to December 22, 2019 23:59 minutes 50 seconds, a total of 14,965 position data; passenger ship MMSI number is 412596777, call sign is 6285, the name is MIN LONG YU 9 777, the length of the ship is 19 m, the track point time span is from 12:52:57 on December 22, 2018, to 3:15:32 on January 3, 2019, and there are 570 pieces of location data.

4.1. Outlier Identification and Elimination. Before processing of the AIS original position data, the trajectories of the two ships are shown in Figures 1(a) and 1(b) within the set time span.

As can be seen from Figure 1, the trajectories established based on the original position data sequence of the two ships are somewhat messy, and some abnormal drift points can be

seen intuitively. The existence of anomalous drift points makes some trajectory segments even cross land, which is obviously not credible. If these ship's original position data are used as the data source of statistical analysis system, it will cause the statistical analysis results to deviate from the actual situation. For the ship management, monitoring, and analysis system, it will cause erroneous alarms due to abnormal ship behavior frequently. Using the discriminate method for abnormal trajectory points proposed in this paper, we can find that the different types of abnormal points and their number in the original position data of the two ships are shown in Table 1.

It can be seen from the distribution of the abnormal point types of the two ships in Table 1 that the abnormal points of oil tankers account for a relatively large amount, exceeding 50%, while the abnormal points of passenger ships account for a relatively small amount, around 10%; anomalous drift points account for the highest proportion of all types of abnormal points. In addition to the abnormal drift point, the oil tanker also has some abnormal turning points and a few abnormal acceleration points, while the passenger

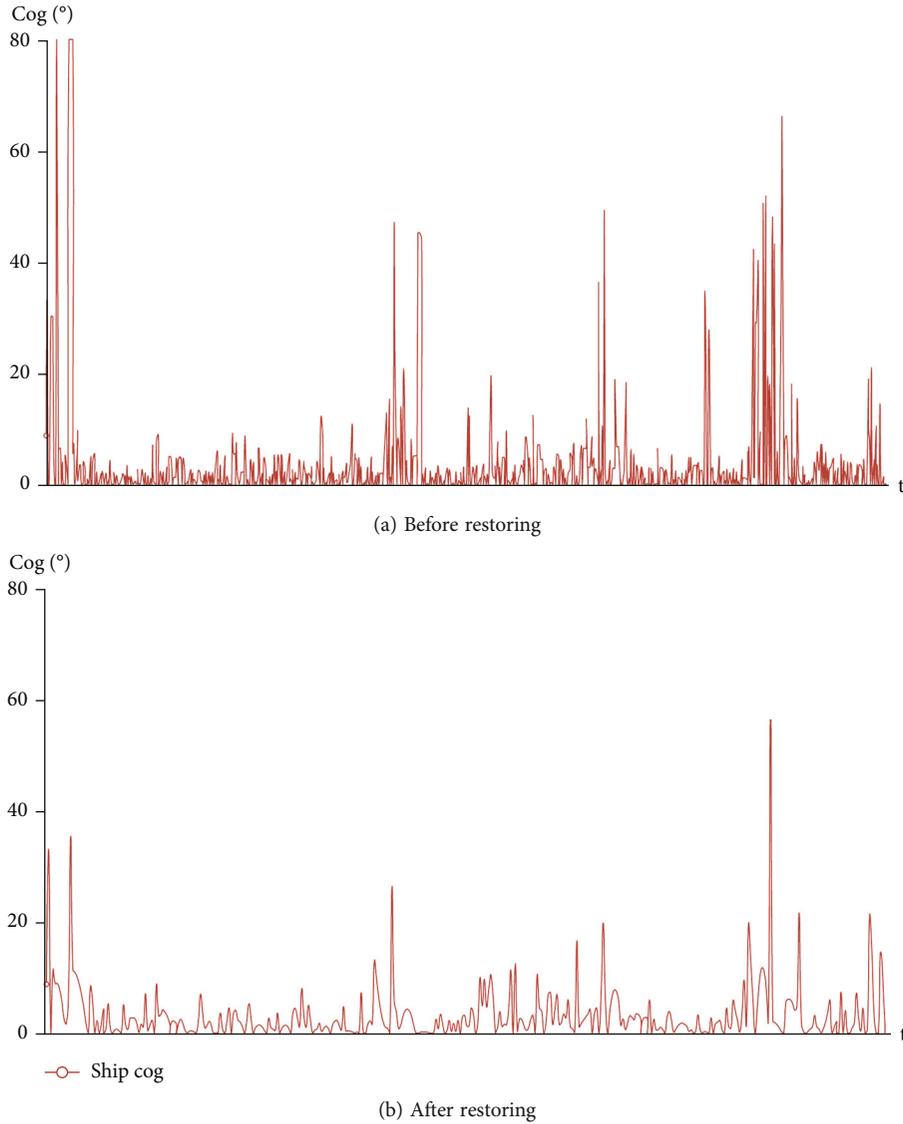


FIGURE 7: Comparison of heading changes.

ship has not found any other abnormal points; no abnormal stopping point has been found in two ships' trajectories. The main reason for the large difference in the distribution of abnormal points between two different types of ships may be that the passenger ship we selected has a short voyage period and basically sailing along the coast, so AIS data transmission is relatively standardized and normal; the voyage period of the oil tanker is relative long and includes the process of entering and leaving the port, which results in abnormalities in AIS data transmission and reception. In addition, there is no abnormal stopping point for both ships. The main reason may be that there is no repeated message forwarding.

According to the abnormal point processing method we proposed in this paper, we can remove all kinds of abnormal points found in the original AIS data. After removing all the abnormal points, the trajectories of the two ships are shown in Figures 2(a) and 2(b). It can be seen from Figure 2 that after clearing the abnormal points, the trajectories of both

ships became clear and identifiable. But for the passenger ship trajectory, there is still a trajectory line across the land. By querying AIS data, we can find that the timestamps of the two points are 13:27:59 on December 22, 2018, and 13:32:19 on December 22, 2018. The time span of the two points does not exceed the normal range, but the ship speed is faster and the track point span is somewhat large, so the track line passes through the land. This problem will not exist after the subsequent track point repairing that we will discuss in the next part.

4.2. Trajectory Missing Point Repair. The overall trajectory of the two ships becomes clear after identifying and deleting the abnormal trajectory points, but the time interval between the trajectory points is somewhat large and uneven, which sometimes cannot meet the requirements of local trajectory analysis and application. It can be seen from the static data that these two ships belong to class B vessels. Therefore, we can repair the trajectories using the cubic spline

interpolation method given in Section 2 of this paper and set the AIS message transmission interval to 30 seconds for interpolation. For the oil tanker, a total of 5,419 trajectory points were inserted, and the ratio of the inserted trajectory points to the total trajectory points is 36.21%. The comparison of the scatter points of the ship trajectory before and after data restoration is shown in Figures 3(a) and 3(b). For passenger ship, a total of 654 trajectory points are inserted, and the ratio of the inserted trajectory points to the total trajectory points is 57.93%. The comparison of scatter points of ship trajectory before and after data restoration is shown in Figures 4(a) and 4(b).

As can be seen from Figures 3 and 4, the density of the ship's trajectory points increased significantly after the data was repaired, so the trajectory became more continuous. However, some segments in trajectories are still in discontinuous state after repairing. As shown in the rectangular marking part of Figures 3(b) and 4(b), there is an obvious gap in each trajectory. The main reason why the trajectory segments have not been repaired is that the time interval between two adjacent track points in the segment is too long, more than 5 minutes for this case. It indicates that the ship is in the state of shutting down AIS equipment during this period, and its behavior is uncertain, so it will not be repaired.

In addition to restoring the abnormal points of position data, we also restore the abnormal points of data such as speed anomalies, acceleration anomalies, and heading anomalies. In the two ships we selected, only the oil tanker has acceleration and heading abnormal points, so only AIS data of the oil tanker was processed for speed, acceleration, and heading anomalies. Before and after clearing and repairing the abnormal points, the comparison of the ship's speed, acceleration, and heading changes can be illustrated by a part of trajectory data, as shown in Figures 5–7 (data range is from 2018-12-22 8:22:51 to 2018-12-22 9:37:45).

As can be seen from Figures 5–7, the abnormal changes of speed, acceleration, and direction beyond the range of ship's maneuverability in the original data have been eliminated. The ship's speed, acceleration, and direction changes tend to be continuous, smooth, and all within a reasonable range after restoration.

5. Conclusion

With the application and popularization of AIS equipment on ship, AIS data has become one of the important data sources for ship traffic flow analysis, maritime supervision, and accident analysis. However, it is difficult for upper-layer applications to apply these AIS data directly because of its unreliability. Aiming at the problem of ship trajectory construction based on AIS data, a method of abnormal point detection and repair in AIS data is proposed in this paper. The proposed method classifies AIS abnormal point and processes them separately according to the longitude and latitude, speed, acceleration, and direction information in AIS data. It is worth noting that the proposed method only needs the AIS data of the ship itself and does not need the support of the historical track data. In addition, the cubic

spline interpolation method is used to repair the trajectory after eliminating the abnormal points, which further improves the continuity and integrity of the trajectory.

The results of processing actual ship trajectories show that the method proposed in this paper can identify all kinds of trajectory abnormal points in AIS data effectively. The interpolation processing method after removing abnormal points can effectively eliminate the sudden changes in position, speed, acceleration, and heading. The trajectory data after being restored are in a reasonable range in terms of latitude and longitude, speed, acceleration, and heading, and the changes are continuous and smooth.

Data Availability

In order to verify the effect of the method for processing and repairing the abnormal points of ship trajectory data proposed in this paper, some original AIS data were downloaded through the VTE Explorer website for experiments. We select the part of data about Xiamen Port and surrounding waters. The time range is from 11:46:37 on December 21, 2018, to 7:30:22 on January 3, 2019; the spatial range is 117.7737° E and 24.08784° N to 118.63037° E and 24.691° N, including 12158622 pieces of position data and 387745 pieces of static data.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Fujian Province young and middle-aged teachers education research project (Nos. JAT210651 and JAT210647), the Scientific Research Project and Research Innovation Team of Concord University College of Fujian Normal University in 2020 (Nos. KY20200203 and 2020-TD-001).

References

- [1] X. Xue and C. Jiang, "Matching sensor ontologies with multi-context similarity measure and parallel compact differential evolution algorithm," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24570–24578, 2021.
- [2] X. Xue, P.-W. Tsai, and Y. Zhuang, "Matching biomedical ontologies through adaptive multi-modal multi-objective evolutionary algorithm," *Biology*, vol. 10, no. 12, pp. 1216–1287, 2021.
- [3] X. Xue and J. Zhang, "Matching large-scale biomedical ontologies with central concept based partitioning algorithm and adaptive compact evolutionary algorithm," *Applied Soft Computing*, vol. 106, p. 107343, 2021.
- [4] X. Xue and J. Chen, "Matching biomedical ontologies through compact differential evolution algorithm with compact adaption schemes on control parameters," *Neurocomputing*, vol. 458, pp. 526–534, 2021.
- [5] R. Connolly, "Safety of Life at Sea," *Radiouser*, vol. 7, no. 4, 2012.

- [6] P. Chen, G. Shi, S. Liu, and Y. Zhang, "Decision support based on artificial fish swarm for ship collision avoidance from AIS data," in *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*, Chengdu, China, 2018.
- [7] B. Murray and L. P. Perera, "A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data," *Ocean Engineering*, vol. 209, p. 107478, 2020.
- [8] L. Du, F. Goerlandt, and P. Kujala, "Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data," *Reliability Engineering and System Safety*, vol. 200, pp. 1–23, 2020.
- [9] J. L. Shepperson, N. T. Hintzen, C. L. Szostek, E. Bell, L. G. Murray, and M. J. Kaiser, "A comparison of VMS and AIS data: the effect of data coverage and vessel position recording frequency on estimates of fishing footprints," *ICES Journal of Marine Science*, vol. 75, no. 3, pp. 988–998, 2018.
- [10] F. Wang, Y. Lei, Z. Liu, X. Wang, S. Ji, and A. K. Tung, "Fast and parameter-light rare behavior detection in maritime trajectories," *Information Processing and Management*, vol. 57, pp. 1–10, 2020.
- [11] Q. Lu and K. D. Kim, "Autonomous and connected intersection cross-sing traffic management using discrete-time occupancies trajectory," *Applied Intelligence*, vol. 49, pp. 1621–1635, 2019.
- [12] Y. Gong, E. Chen, and X. Zhang, "Antmapper: an ant colony-based map matching approach for trajectory-based applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 390–401, 2018.
- [13] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G. B. Huang, "Exploiting AIS data for intelligent maritime navigation: a comprehensive survey from data to methodology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1559–1582, 2018.
- [14] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 123–144, 2017.
- [15] C. Anagnostopoulos and S. Hadjiefthymiades, "Intelligent trajectory classification for improved movement prediction," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 10, pp. 1301–1314, 2014.
- [16] S. Qiao, N. Han, W. Zhu, and L. A. Gutierrez, "TraPlan: an effective three-in-one trajectory-prediction model in transportation networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1188–1198, 2015.
- [17] M. Schreier, V. Willert, and J. Adamy, "An integrated approach to maneuver-based trajectory prediction and criticality assessment in arbitrary road environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2751–2766, 2016.
- [18] G. Yuan, J. Zhao, S. Xia, Y. Zhang, and W. Li, "Multi-granularity periodic activity discovery for moving objects," *International Journal of Geographical Information Science*, vol. 31, no. 3, pp. 435–462, 2017.
- [19] X. Yin and Q. Chen, "Trajectory generation with spatio-temporal templates learned from demonstrations," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3442–3451, 2017.
- [20] H. Rong, A. P. Teixeira, and G. G. Soares, "Data mining approach to shipping route characterization and anomaly detection based on AIS data," *Ocean Engineering*, vol. 198, pp. 1–12, 2020.
- [21] Z. Yan, Y. Xiao, L. Cheng et al., "Exploring AIS data for intelligent maritime routes extraction," *Applied. Ocean. Research*, vol. 101, pp. 1–10, 2020.
- [22] V. F. Arguedas, G. Pallotta, and M. Vespe, "Maritime traffic networks: from historical positioning data to unsupervised maritime traffic monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 722–732, 2018.
- [23] C. Iphar, C. Ray, and A. Napoli, "Data integrity assessment for maritime anomaly detection," *Expert Systems with Applications*, vol. 147, no. 113219, pp. 113219–113219, 2020.
- [24] X. Pan, H. Wang, X. Cheng, X. Peng, and Y. He, "Online detection of anomaly behaviors based on multidimensional trajectories," *Information Fusion*, vol. 58, pp. 40–51, 2020.
- [25] T. Watawana and A. Caldera, "Analyse near collision situations of ships using automatic identification system dataset," in *In 2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, Nairobi, Kenya, 2018.
- [26] J. Mao, C. Jin, Z. Zhang, and A. Zhou, "Anomaly detection for trajectory big data: advancements and framework," *Journal of Software*, vol. 28, no. 1, pp. 17–34, 2017.
- [27] T. Zhang, S. Zhao, and J. Chen, "Ship trajectory outlier detection service system based on collaborative computing," in *In 2018 IEEE World Congress on Services (SERVICES)*, pp. 15–16, San Francisco, CA, 2018.
- [28] T. Watawana and A. Caldera, "Anomaly detection in maritime data based on geometrical analysis of trajectories," in *In 2015 18th International Conference on Information Fusion (Fusion)*, pp. 1100–1105, Washington DC, USA, 2015.
- [29] D. Zhang, J. Li, Q. Wu, X. Liu, X. Chu, and W. He, "Enhance the AIS data availability by screening and interpolation," in *In 2017 4th International Conference on Transportation Information and Safety (ICTIS)*, pp. 981–986, Banff, Canada, 2017.
- [30] X. Zhang, Y. He, R. Tang, J. Mou, and S. Gong, "A novel method for reconstruct ship trajectory using raw AIS data," in *In 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, pp. 192–198, Singapore, 2018.