

Article

Ship Anomalous Behavior Detection in Port Waterways Based on Text Similarity and Kernel Density Estimation

Gaocai Li ¹, Xinyu Zhang ^{1,*}, Yaqing Shu ², Chengbo Wang ³, Wenqiang Guo ¹ and Jiawei Wang ¹

¹ Maritime Intelligent Transportation Research Team, Navigation College, Dalian Maritime University, Dalian 116026, China; haley1gc@163.com (G.L.); guowenqiang@dmlu.edu.cn (W.G.); wangjiawei1068@dmlu.edu.cn (J.W.)

² School of Navigation, Wuhan University of Technology, Wuhan 430063, China; y.shu@whut.edu.cn

³ Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei 230052, China; wangcb23@ustc.edu.cn

* Correspondence: zhangxy@dmlu.edu.cn

Abstract: The navigational safety of ships on waterways plays a crucial role in ensuring the operational efficiency of ports. Ship anomalous behavior detection is an important method of water traffic surveillance that can effectively identify abnormal ship behavior, such as sudden acceleration or deceleration. In order to detect potential anomalous ship behavior in real time, a method for ship anomalous behavior detection in waterways is proposed based on text similarity and kernel density estimation. Under the assumption of known traffic patterns entering and leaving the port, this method can identify ship behaviors that violate traffic patterns in real time. Firstly, kernel density estimation is applied to construct a traffic pattern density model for ship trajectories entering and leaving the port, used to estimate the density values of ship motion states. Simultaneously, a semantic transformation method is used to convert traffic pattern trajectory into pattern trajectory text, which is used to identify the ship's traffic pattern. Subsequently, the historical trajectory data of the target ship are transformed into textual trajectories, and text similarity is used to identify ship inbound and outbound traffic patterns. Furthermore, the constructed traffic pattern density model is used to estimate real-time density values of the state of ship motion, and the trajectory points that exceed the threshold of the anomaly factor are marked as anomalies. Finally, the effectiveness of the proposed method is validated using simulation data, and the results indicate an accuracy of more than 90% for the comprehensive detection of anomalous behavior. This study, approaching the detection of potential ship anomalous behavior from the perspective of port traffic patterns, enriches the methods of ship anomalous behavior detection in port waterways.

Keywords: port waterways; traffic patterns; text similarity; kernel density estimation; ship anomalous behavior detection



Citation: Li, G.; Zhang, X.; Shu, Y.; Wang, C.; Guo, W.; Wang, J. Ship Anomalous Behavior Detection in Port Waterways Based on Text Similarity and Kernel Density Estimation. *J. Mar. Sci. Eng.* **2024**, *12*, 968. <https://doi.org/10.3390/jmse12060968>

Academic Editor: Sergei Chernyi

Received: 22 April 2024

Revised: 29 May 2024

Accepted: 5 June 2024

Published: 8 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Navigation safety in port waterways, as the main waterway for ships entering and leaving ports, directly affects the operational activities of ports. The European Maritime Safety Agency analyzed 21,173 maritime accidents that occurred in the past decade, and the results indicate that the proportion of maritime accidents in port waters exceeds 40% [1]. Therefore, port waters are identified as areas with frequent maritime accidents. Furthermore, according to the United Nations Conference on Trade and Development, maritime trade volume is expected to continue to grow at an annual rate of 2.1% over the next five years [2]. This is expected to lead to an increasing traffic density of ships, making the navigational environment more complex in port waterways, posing greater challenges for the safety supervision of ship traffic in waterways [3,4]. Once a maritime disaster occurs, it can deliver a devastating blow to the environment essential for the survival of marine life [5,6]. Ship anomaly behavior detection is an essential maritime surveillance method

that utilizes anomaly detection algorithms to identify deviations in the ship's movements from traffic patterns. This approach can effectively reduce the rate of maritime accidents, improve waterway traffic safety, and optimize the capacity to raise early warnings of potential risks. It serves as a crucial means for achieving automation and intelligence in maritime supervision [7,8].

When navigating within harbor waters, ships are required to adhere to port navigation rules, including following designated courses and speeds within specific navigational waterways. These actions create a distinct pattern of vessel movements in and out of the port, known as the port traffic model. Deviations from this normal traffic pattern, such as sudden changes in course and speed, resulting in veering off the navigational route or stopping within the channel, are considered abnormal behaviors [9]. Such deviations often indicate potential navigation risks, such as loss of vessel control, collisions, and groundings [10]. In this paper, we consider behaviors such as abrupt acceleration or deceleration and course deviations from the established port traffic model as abnormal behaviors. Ship anomaly detection typically involves two parts: constructing an anomaly detection model and using the detection model to identify anomalous ship behavior. Currently, ship anomaly detection methods based on numerical calculations can be categorized into three types: based on statistical analysis, based on neural networks, and based on clustering [11]. Statistical methods assume that ship movements follow a certain probability distribution, thereby establishing a corresponding probability distribution model to detect abnormal behavior. When the ship movements fall within a high-probability region, normal ship behavior is considered. Rong et al. [12] assumed that the ship's motion along the ship's route follows a normal distribution. They defined the 95% probability interval of the route as the waterway boundary and detected deviations from the historical trajectory by calculating the Gaussian function values of the ship's lateral distribution. Laxhammar et al. [13] assumed that the ship's motion follows a Gaussian mixture model (GMM). They trained and learned from historical trajectory samples to obtain parameter estimates for the GMM model. The anomalies in the new trajectory points were determined by evaluating the GMM probability values. Mascaro et al. [14] integrated automatic identification system (AIS) data, weather, and time data. They used Bayesian networks to generate dynamic and static Bayesian network models for detecting anomalies in a given trajectory. For harbor waters, ships' trajectories exhibit concentrated distributions within navigational channels but with uneven density, and distinct directional differences in inbound and outbound traffic. It is challenging to employ a unified probability distribution model to construct ship anomaly detection algorithms. Furthermore, current research on ship anomaly detection primarily focuses on coastal and inland waterways, with relatively fewer studies conducted in harbor waters.

Cluster-based methods are among the most commonly used approaches for ship anomaly behavior detection [15–17]. Clustering is based on the similarity of ship trajectories in terms of location or motion states (course, speed, etc.). Clustering methods are employed to identify trajectories with ship behaviors similar to those of the same motion pattern. The recognized cluster motion patterns are considered normal, and other ship behaviors deviating from these patterns are deemed abnormal. Pallotta et al. [18] proposed a ship anomaly detection algorithm called Traffic Route Extraction and Anomaly Detection (TREAD). This method uses clustering algorithms to extract normal traffic patterns and establishes an anomaly detector to detect anomalous behavior in ships. Zhen et al. [19] introduced an abnormal ship behavior detection method combining ship trajectory clustering and a naive Bayes classifier. They designed a trajectory similarity measurement method that considers position and course features in AIS data. Then, hierarchical clustering (HCA) and k-medoids clustering methods were applied to model ship navigation patterns. A naive Bayes-based ship abnormal detection classifier is established to identify abnormal ship behaviors. Botts et al. [20] applied the DBSCAN algorithm to cluster ship trajectory points, allowing the identification of anomalous behavior in the given ship AIS trajectory data. Liu et al. [21] used density clustering methods to group historical AIS data, extract

typical trajectories as cluster centers, and set an abnormal threshold for ship speed. When the ship speed exceeds the set threshold, it is considered a speed anomaly. Radon et al. [22] applied the DBSCAN clustering algorithm to extract ship motion patterns from historical trajectory points. They then used these motion patterns to detect potential anomalies, and finally, through contextual verification, filtered out true anomalies from potential ones. In clustering-based methods, there is a lack of consideration of information beyond the influences on vessel behavior, such as vessel static information and navigation rules, resulting in the incomplete characterization of vessel behavior. Therefore, there is a need for further research to incorporate the factors influencing vessel behavior into trajectory similarity measurements and clustering methods to obtain more reasonable, accurate, and comprehensive representations of vessel navigation behavior [19].

In recent years, with the rapid development of artificial intelligence technology, algorithms based on neural networks have garnered increasing attention in ship anomaly behavior detection [23,24]. The neural network-based approach involves establishing a neural network model, training it with historical data to obtain a normal ship behavior model, and using the trained model to detect new behaviors. If ship behavior deviates beyond a certain range from the normal behavior model, it is considered abnormal. Rhodes et al. [25] used a neural network to learn the current position, speed, and course information of the ship to predict its future positions. If the predicted position deviates from the actual position, the ship behavior is deemed abnormal. Huang et al. [26] proposed a ship trajectory anomaly detection method based on an LSTM model that incorporates ship size, environmental information, and time interval features. Hu et al. [27] used variational autoencoders to discover potential connections between each dimension of normal trajectories and spatial similarities between normal trajectories, and used deep reinforcement learning algorithms to train trajectory anomaly detection models. Nguyen et al. [28] proposed a deep learning approach for ship anomaly behavior detection using AIS data streams. They combined recursive neural networks with latent variable modeling to extract useful information from AIS data streams and construct a normal model for ship anomaly behavior detection. Eljabu et al. [29] trained a graph convolutional network model for the spatiotemporal maritime traffic network. They constructed a graph network bias detector using graph embedding and context embedding techniques to identify ships deviating from their routes. However, the neural network-based approach also has its drawbacks, such as the large demand for pre-training trajectory data and sensitivity to initial values and parameters. It is necessary to consider the practical maritime application scenarios comprehensively.

However, these methods often fail to consider external factors, such as maritime traffic rules, when constructing anomaly detection models. Port navigation rules are usually formulated and enforced by each country's port management agencies or authorities. The primary purpose of these rules is to ensure port security, the safe navigation of ships, and efficient port operations. The rules involving the safe navigation of ships include rules for ships entering and leaving ports, rules for waterways and anchorages, and rules for using port facilities. These navigation rules will jointly affect the ship's navigation behavior when entering and leaving the port, such as selecting anchorages, channels, routes, headings, and speeds [30]. Ships of the same type will follow the navigation rules of the port to form a specific inbound/outbound traffic pattern. The ship will enter specific port waters from anchorage water along a specific waterway and berth at a specific terminal operating area. For port waters with complex traffic, it is difficult for traditional methods to accurately identify the traffic patterns of ships entering and leaving the port, so it is easy for a high rate of false positives or false negatives to arise in anomaly detection. Hence, for ship anomaly detection in port waters, two main challenges need to be addressed, as follows:

1. How to accurately identify the traffic patterns of inbound and outbound ships based on relevant maritime information;
2. How to effectively detect potential anomalous ship behavior based on inbound and outbound traffic patterns.

With the rapid development of artificial intelligence technology, natural language processing techniques have found widespread and effective applications in fields such as information retrieval, text classification, and speech recognition [31–33]. Recently, some scholars have begun applying natural language processing techniques to extract knowledge in the traffic domain. Hughes et al. [34] used segmentation and word frequency statistics to identify and classify different accident risks in road traffic accident data. Huang et al. [35] and Li et al. [36] used semantic transformation methods to convert ship trajectories into trajectory texts. They utilized topic models to identify ship motion pattern topics from the trajectory texts. Kernel density estimation is a non-parametric statistical method used to estimate the probability density of data. Due to its nonparametric and unsupervised characteristics, this method is widely used in the field of anomaly detection [37].

Building on previous research, this paper proposes a method for ship anomalous behavior detection in port waterways based on text similarity and kernel density estimation, named TS-KDE. The method first employs a kernel density estimate function to construct a pattern trajectory kernel density estimation model for ship inbound and outbound traffic patterns. Simultaneously, a semantic transformation method is used to convert and generate textual representations of inbound and outbound traffic pattern trajectories. Subsequently, the historical trajectory data of the target ship are semantically transformed into textual trajectories, and text similarity is applied to identify ship inbound and outbound traffic patterns. Finally, using the constructed kernel density estimation model, real-time density values of ship motion states are estimated, and trajectory points exceeding an anomaly factor threshold are flagged as anomalous. The framework of the method is illustrated in Figure 1. Through verification and analysis in multiple simulation scenarios, the method demonstrates good real-time detection performance. The research contributions of this study are summarized as follows.

- (1) A novel method for detecting abnormal ship behavior in waterways is proposed. This paper is dedicated to studying the abnormal behavior of ships in waterways. When designing the anomaly detection algorithm, the impacts of ship static attributes, port traffic rules, etc., on ship movement were considered, and the dense historical trajectories of ships in the waterways were divided into different inbound and outbound traffic patterns, which enabled the inbound and outbound traffic patterns to represent ships' typical patterns.
- (2) It is proposed to use the multiple attributes of ship trajectory to perform semantic transformation on the ship trajectory points to improve the accuracy of traffic pattern identification. This paper converts ship type, course, speed, and geospatial attribute semantics into ship trajectory text, ensuring the completeness of ship movement information, eliminating the problem of only considering the distance between trajectories in traditional ship traffic pattern identification, and improving ship traffic pattern identification results' accuracy.
- (3) This paper proposes using the cosine similarity measurement method to identify ship inbound and outbound traffic patterns, improving traffic pattern identification efficiency. The cosine similarity measure is a method based on vector calculation. Based on the vectorized representation of target trajectory text and traffic mode trajectory text, the traffic mode can be identified by simply calculating the cosine similarity between vectors, eliminating the need to calculate different trajectories. Therefore, the method proposed in this paper will significantly improve the efficiency of traffic mode identification compared to traditional methods.
- (4) A method for detecting abnormal ship behavior based on kernel density estimation, which can detect abnormal ship behavior in a timely and effective manner, is proposed. A local density and abnormal factor calculation method was constructed to ensure the abnormal ship behavior detection method's accuracy and real-time performance. Simulation experiments have verified that abnormal ship behavior can be detected promptly, and the accuracy of the detection results can reach more than 90%.

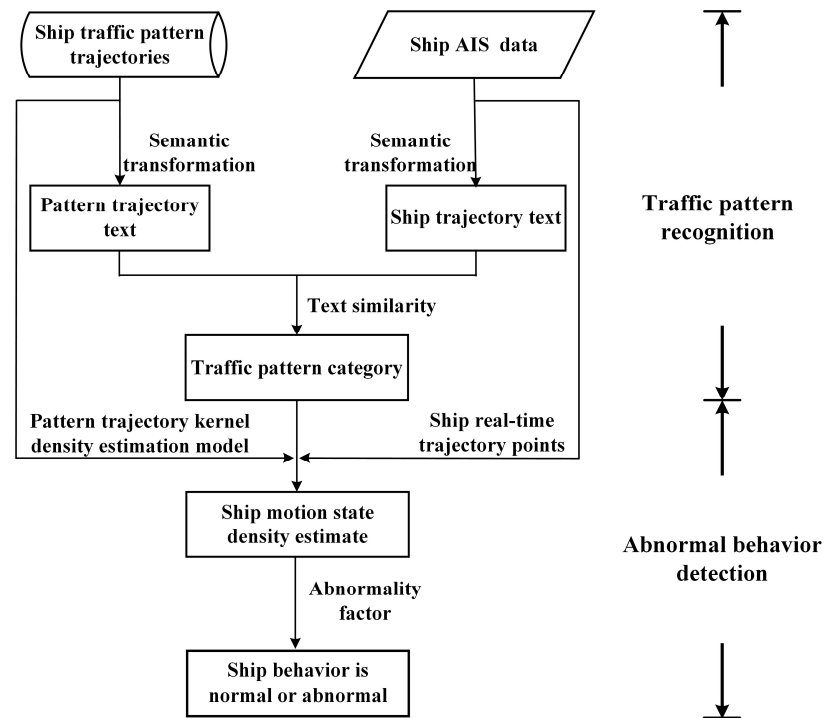


Figure 1. Ship anomaly behavior detection in port waterways based on text similarity and kernel density estimation.

The remainder of this paper is organized as follows. Section 2 presents the principles of ship anomaly detection methods. The experiments and results analysis are discussed in Section 3. The advantages and limitations of the proposed method are addressed in Section 4. Section 5 provides conclusions and future works.

2. Methods

The TS-KDE method framework first identifies the traffic patterns of ships entering and leaving the port, and then detects abnormal behaviors of ships that deviate from these patterns. Text cosine similarity is used to calculate the similarity between the traffic pattern text, after semantic transformation, and the text of the ship trajectory to be tested, identifying the traffic pattern followed by the ship. Potential abnormal behaviors deviating from the identified patterns are detected using a kernel density estimation-based ship anomaly detection factor.

2.1. Ship Traffic Pattern Recognition

Traditional ship traffic pattern recognition methods typically use trajectory similarity metrics to calculate the distances between trajectories. Such methods require calculating similarity values between a trajectory and all other trajectories, leading to inefficiency. Additionally, in narrow waterways with concentrated ship trajectories, it is challenging to accurately classify ship traffic patterns. In this study, we draw inspiration from text similarity methods, transforming ship trajectories into trajectory text and comparing the similarity between trajectory texts to achieve ship traffic pattern recognition.

2.1.1. Trajectory Text Generation

In order to precisely identify the traffic patterns of the ships entering and leaving ports, this paper combines four types of spatio-temporal trajectory information, including ship position, course, speed, and geographical spatial information, to represent the motion characteristic of the ship. First, the four types of spatio-temporal trajectory information mentioned above are transformed into semantically comprehensible data. Subsequently,

the transformed semantic information is organized into a trajectory dictionary to generate textual documents representing ship trajectories.

The semantic representation of ship position conveys information about a ship’s location within navigable waters. To better understand ships’ position information, this study assigns geographic codes to water areas that describe the meaning of the characteristics of the ship position. As illustrated in Figure 2b, the water areas are divided into $M \times N$ square grids, each grid denoted as $G_{m,n}$, where m and n represent the row and column codes of the grid. Therefore, the semantic information of the position can be represented by the row and column codes, such as $ST(lon_k, lat_k) = Row\#Col$.

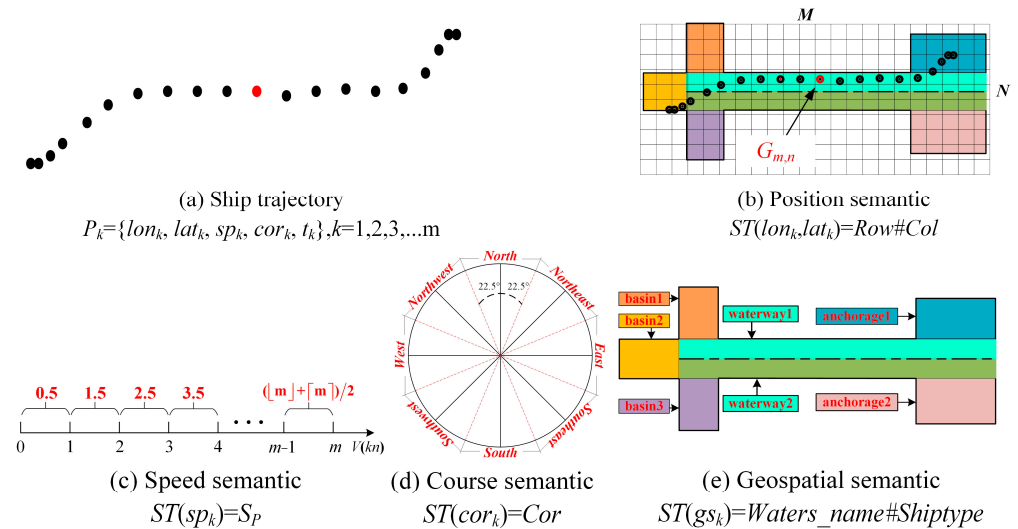


Figure 2. Semantic transformation of ship motion characteristic.

The semantic representation of the course of ship conveys the course of a ship’s navigation. To better describe the semantic information of course, a model based on a conical direction derived from geographic information theory is used. As shown in Figure 2d, course is uniformly divided into eight conical directions, namely, $Dir = \{North, Northeast, East, Southeast, South, Southwest, West, Northwest\}$. Its semantics can be described as $ST(dir_k) = Dir$.

The semantic representation of ship speed describes the magnitude of a ship’s speed in sailing waters. This paper adopts a segmented approach with 1-knot intervals to partition ship speeds. As shown in Figure 2c, the ship speed distribution is obtained as $V = \{(0, 1], (1, 2], \dots, ([m], [m])\}$, where m is the maximum speed of ship in the study water area. In this case, the average value of each speed segment is taken as the semantic feature of speed, such as $Sp = \{0.5, 1.5, \dots, ([m] + [m])/2\}$. Its semantics can be described as $ST(sp_k) = Sp$.

The semantic representation of the geographic space of the ship describes the waters through which the ship passes when entering and leaving the ports. Geographical spatial semantics can also reflect the influence of port traffic rules [36]. This paper combines the type of ship and the names of the sailing waters traversed to characterize the semantic information of port traffic rules, that is, $ST(gs_k) = Waters_name\#Shiptype$. The type of ship is determined based on the category of the ship, and the water names are related to the division of the port waters, as shown in Figure 2e.

Semantic conversion can transform the numerical information from the trajectory into easily understandable semantic information. To obtain a textual representation of the trajectories, this paper constructs a trajectory dictionary to represent ship trajectories in text form. The trajectory dictionary expresses the ship’s motion state through a series of motion words in writing. In this paper, each motion word is a comprehensive representation of ship position semantics, geographical space semantics, speed semantics, and course semantics, i.e., $mw = Row\#Col\#Waters_name\#Shiptype\#Dir\#Sp$. Therefore, the trajectory dictionary

contains a total of $N_p \times N_g \times N_c \times N_s$ motion words. Here, N_p is the number of words describing ship position features, determined by the size of the study water area and the dimensions of the spatial grid division; N_g is the number of words describing geographical space, determined by water type and ship classification; N_c and N_s describe the number of words representing ship course and speed features.

A trajectory document is composed of a set of motion words that describe the continuous movement of ships entering and leaving ports. Assuming that the study water area contains M trajectories, i.e., M trajectory documents, it can be represented as $TD = \{td_1, td_2, \dots, td_M\}$. For each trajectory document, td_i , it can be represented by a finite, continuous sequence of N motion words, $td_i = \{mw_1, mw_2, \dots, mw_N\}$.

2.1.2. Trajectory Text Similarity Measurement

For evaluating text similarity, common methods include cosine similarity, Jaccard similarity, Jensen–Shannon divergence, and so on [38,39]. Due to its good performance in calculating the similarity between texts of different lengths and its insensitivity to vector dimensions, cosine similarity is widely used. The text words are encoded and vectorized, and the cosine value between vectors is measured to quantify the similarity of the text. This paper utilizes the cosine similarity method to measure the similarity of trajectory texts. Motion words are vectorized, where each motion word is assigned a different dimension, and the values on each dimension correspond to the frequency of that motion word appearing in the trajectory document. Cosine similarity provides a measure of similarity between two documents in terms of their topics. The specific method steps are as follows:

Step 1—Tokenize the trajectory text to obtain a list of motion words corresponding to the trajectory text, i.e., $T_l = [mw_1, mw_2, \dots, mw_m]$, where m represents the number of motion words;

Step 2—Summarize and de-duplicate the motion words in the set of trajectory texts to obtain a vocabulary of motion words. The number of words in this vocabulary is the dimensionality of the trajectory text vector, i.e., $V_m = [mw_1, mw_2, \dots, mw_n]$, where n represents the number of non-repeated motion words;

Step 3—Calculate the word frequency for each trajectory text by constructing a word frequency vector corresponding to the trajectory text;

Step 4—For two trajectory texts with n -dimensional word frequency vectors \vec{A} and \vec{B} , where \vec{A} is $[a_1, a_2, \dots, a_n]$ and \vec{B} is $[b_1, b_2, \dots, b_n]$, the cosine value of the angle θ between vectors \vec{A} and \vec{B} can be obtained by Equation (1) [39]

$$\cos(\theta) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|} = \frac{\sum_{i=1}^n a_i \times b_i}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}} \quad (1)$$

In this paper, semantic conversion is employed to transform pattern trajectories into pattern trajectory text. Historical trajectory data of the target ship are converted, generating the corresponding trajectory text. The cosine similarity between the pattern trajectory text and the target trajectory text is measured to ultimately identify the traffic pattern of the target ship entering and leaving the port.

2.2. Ship Anomaly Detection

To enable the real-time detection of abnormal ship behavior, an assessment can be made of the extent to which the ship’s movements conform to overall traffic patterns. The density estimation-based outlier detection method detects anomalies by comparing the density of a sample with the density of its neighborhood, where the neighborhood density refers to the average local density within its vicinity. A novel approach is proposed that calculates an anomaly factor for each point on the trajectory based on the mean density of the neighborhood, which is used to detect abnormal ship behavior.

2.2.1. Definition of Neighborhood

Suppose X is a dataset, and n is the number of data points in dataset X , where x is a d -dimensional data object, that is, $x = [x_1, x_1, \dots, x_d]$. For any positive integer $m (m \leq n)$, the m -distance of data point x is represented as $\text{dis}_m(x)$. $\text{dis}_m(x)$ must simultaneously satisfy the following two conditions [40]:

- (1) There are at least m sample points $Q \in X \setminus \{x\}$ in the dataset X , such that $\text{dis}(x, Q) \leq \text{dis}_m(x)$
- (2) There are at most $m - 1$ sample points $Q \in X \setminus \{x\}$ in the dataset X , such that $\text{dis}(x, Q) < \text{dis}_m(x)$

$\{x, Q\} \in X$, and $\text{dis}(x, Q)$ is the distance between the data points x and Q in the dataset X . For ease of calculation, this paper adopts the Euclidean distance. That is, for data points $x =$

(x_1, x_2, \dots, x_d) , $Q = (q_1, q_2, \dots, q_d)$, $\text{dis}(x, Q) = \sqrt{\sum_{i=1}^d (x_i - q_i)^2}$. The m -distance neighborhood of data point x can be represented as the set of sample points whose distance to x is not greater than x 's m -distance, denoted by $\text{Domain}_m(x) = \{Q \in X | \text{dis}(x, Q) \leq \text{dis}_m(x)\}$.

Each sample point in the m -distance neighborhood is an m -distance neighbor of x . Figure 3 illustrates the 4-distance neighborhoods of nodes p_1 and p_2 . It is obvious from Figure 3 that the 4-distance neighborhood of p_2 is larger than that of p_1 . Therefore, within the unit area, the number of sample points for p_1 (i.e., the local density of p_1) is greater than that for p_2 (i.e., the local density of p_2). Consequently, the size of the m -distance neighborhood in spatial samples is inversely proportional to the local density of the samples. Thus, with a constant m -value, the smaller the local density of a sample point and the larger its m -distance, the greater the likelihood of it being an outlier. In this scenario, the level of the anomaly of the ship's point p_2 is higher than that of point p_1 .

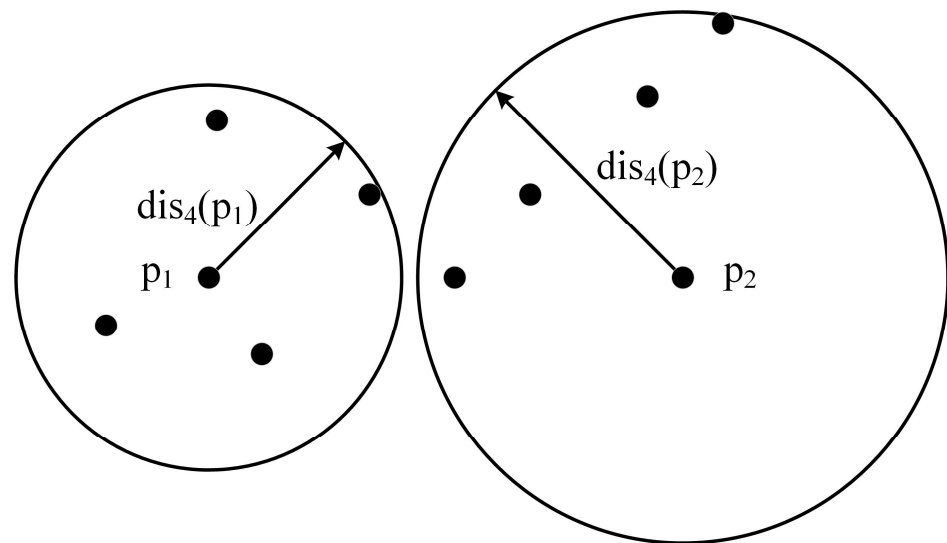


Figure 3. Four-distance neighborhood.

2.2.2. Kernel Density Estimation

The density value of the sample points is closely related to the characteristics of the outliers, so the density value of sample points can be calculated to detect anomalies. Kernel density estimation (KDE) is an important method for studying the distribution characteristics of random variables in samples, and can be used to calculate the domain values of sample points [41].

Suppose X_1, X_2, \dots, X_n are samples in the dataset X , and x_1, x_2, \dots, x_n are the corresponding observed values. For a data object with d -dimensional features, the KDE based on the observation value distribution can be approximated as:

$$f_n(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \tag{2}$$

In general, the kernel function is set as the Gaussian kernel function, resulting in multidimensional Gaussian kernel density.

$$f_n(x) = \frac{1}{nh^d(2\pi)^{d/2}} \sum_{i=1}^n e^{-\frac{\|x-X_i\|^2}{2h^2}} \tag{3}$$

The kernel density function can fit the distribution characteristics of the observation values, thereby obtaining a distribution model for the observation values. The bandwidth parameter h is a crucial hyperparameter that controls the smoothness of the distribution model. When h is too large, the probability density curve becomes smoother, but it may obscure data structures; when h is too small, it may introduce excessive data noise. The optimal fixed bandwidth for the Gaussian kernel function can be calculated using Equation (4) [42]:

$$h^* = An^{-\frac{1}{d+4}} \tag{4}$$

where $A = [4/(d + 2)]^{1/(d+4)}$. In response to the uneven distribution of traffic flow in port waters, aiming to improve the accuracy of abnormal detection results for ship behavior (position, speed, and course), a strategy is used in constructing the KDE model. It involves using a small bandwidth in areas where trajectory points are densely distributed and a large bandwidth in regions with sparse data distribution.

This study proposes the use of adaptive KDE for calculating the density values of the ship behavior data. The adaptive kernel density function can be obtained through a two-stage process [43]. The first stage is typically the fixed KDE. The second stage initially calculates the adaptive window width $\tilde{h}_k, k = 1, 2, \dots, n$ as $\tilde{h}_k = h^* \lambda_k$, where h^* is given in Equation (4), and

$$\lambda_k = \left(\frac{f_n(x)}{l}\right)^{-\gamma} \tag{5}$$

where l is the geometric mean of a sequence $\{f_n(x)\}$, that is,

$$\log l = \frac{1}{n} \sum_{i=1}^n \log f_n(x) \tag{6}$$

where γ is the sensitivity parameter such that $0 \leq \gamma \leq 1$, typically set to 0.5. Finally, the adaptive KDE can be obtained, as in Equation (7).

$$f_n(x) = \frac{1}{n \left(\tilde{h}_k\right)^d (2\pi)^{d/2}} \sum_{i=1}^n e^{-\frac{\|x-X_i\|^2}{2\tilde{h}_k^2}} \tag{7}$$

2.2.3. Neighborhood Density

Anomaly detection based on the local kernel is heavily dependent on the selection of the neighborhood parameter m . To enhance the robustness of the impact of parameter m , a mean-based neighborhood density is proposed. In contrast to traditional neighborhood density, mean-based neighborhood density refers to the mean of the local kernel densities

of all sample points in the neighborhood. It is highly sensitive to outlier data in the neighborhood, and its calculation formula is given in Equation (8):

$$Density(x) = \frac{\sum_{Q \in Domain_m(x)} f_n(x)}{m} \tag{8}$$

Using a mean-based local kernel density estimation can reduce the error impact caused by a single domain point, thus enhancing robustness. To minimize the error in the estimation of the kernel density based on the mean and improve the efficiency of subsequent anomaly detection, this study sets m to 3. This means calculating the mean of the kernel densities of the 3 nearest historical trajectory points as the neighborhood density for the test trajectory point.

2.2.4. Anomaly Factor

The anomaly factor is used to estimate the abnormality of ship behavior and is a decisive factor in determining whether ship behavior is abnormal. In general, normal samples are located in dense areas with higher local kernel density and neighborhood density. Outlying samples are typically distributed in sparse regions, where sample points have lower local kernel density, and the distance between sample points and their neighbors is relatively far. Therefore, using the local kernel density and its neighborhood density estimation of sample points as the anomaly factor helps to determine whether the sample point belongs to an abnormal object. Therefore, the abnormality factor can be obtained using Formula (9). Substituting Equation (8) into Equation (9), we can derive the final formula for calculating the abnormality factor, as shown in Equation (10) [40].

$$Abnor_factor(x) = \frac{Density(x)}{f_n(x)} \tag{9}$$

$$Abnor_factor(x) = \frac{\sum_{Q \in Domain_m(x)} f_n(x)}{mf_n(x)} \tag{10}$$

As indicated by Formula (10), when the local kernel density of a sample is smaller, and the mean density of the neighborhood is higher, the abnormality factor is greater, indicating a higher likelihood that the sample is anomalous. Generally, for anomalous data, their local kernel density and mean neighborhood density differ significantly, leading to an abnormality factor much greater than 1 [40]. In the context of ship movements, where behaviors change continuously and gradually, the timely detection of potential anomalies can reduce the risk of maritime accidents and improve the response time of regulatory authorities, enhancing the efficiency of maritime emergency rescue. Therefore, this study sets the threshold for the abnormality factor at 3 as the criterion for the evaluation of the abnormalities.

$$\begin{cases} \text{Abnormal, if } Abnor_factor(x) \geq 3 \\ \text{Normal, else} \end{cases} \tag{11}$$

When inputting the current state of the ship's behavior, including ship position (longitude, latitude), ship course, and ship speed, the detection of anomalies in the behavior of the ship is achieved by calculating the kernel density value and the neighborhood density of the ship's behavior. This process ultimately establishes a relationship between the anomaly factor and the threshold.

3. Experiments and Results

3.1. Experimental Setup

The test scenario for anomaly behavior detection is the main port area of Tianjin Port. The main port area of Tianjin Port includes one main waterway, two small boat waterways, and three warning waterways. Eleven terminals are distributed on both sides of the West and North Basin. Three anchorages are distributed on both sides of the main waterway, as

shown in Figure 4. To ensure the safe and efficient passage of ships entering and leaving the port, the competent authority of Tianjin Port has formulated navigation rules for ships entering and leaving the port. This study selected AIS data for the Tianjin Port water area from 1 September to 30 November 2016, and after data preprocessing, obtained the complete 1956 trajectories of ships entering and leaving the port. The accuracy of ship traffic patterns has a significant impact on the subsequent detection of abnormal ship behavior. Based on previous research [44], we have identified 40 traffic patterns for ships entering and leaving the port that comply with the navigation rules of Tianjin Port, as shown in Appendix A. To identify the traffic patterns of the test ship entering and leaving the port, the trajectories of the 40 traffic patterns were converted into 40 different pattern trajectory texts.

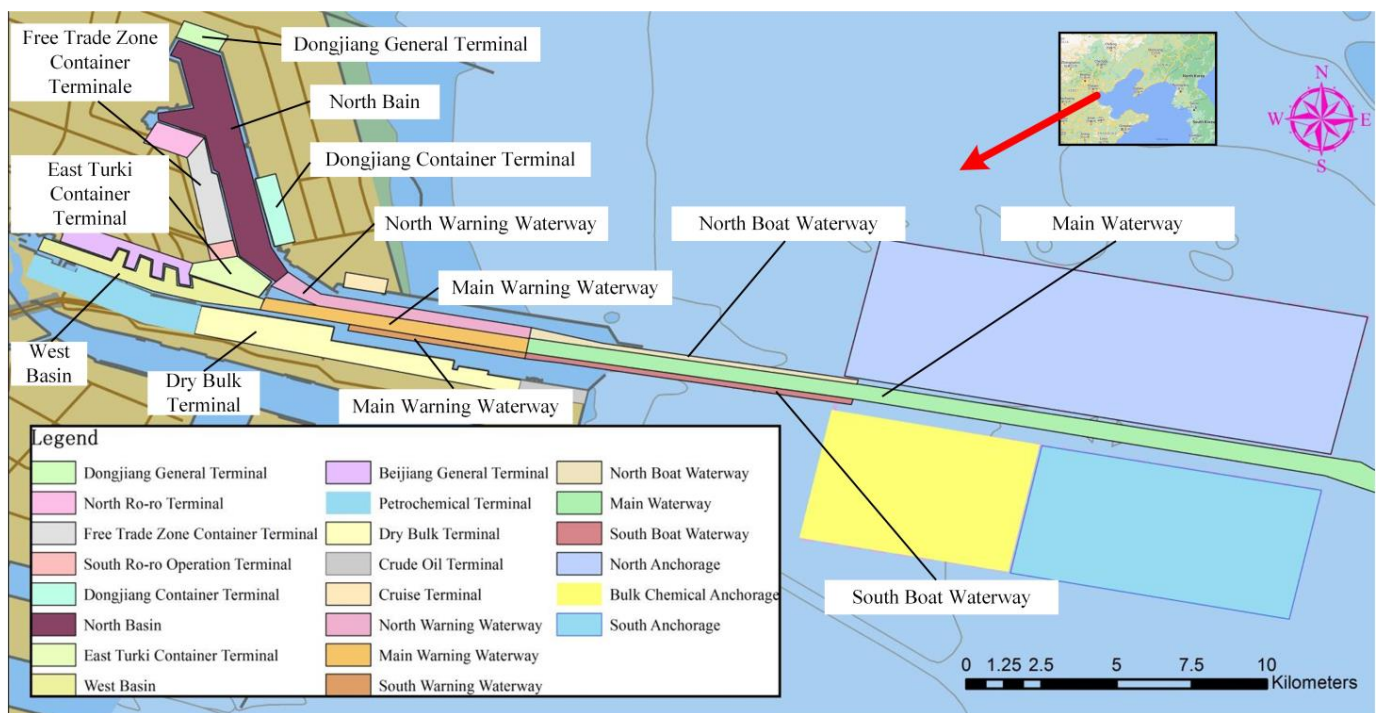


Figure 4. Layout of water and land facilities in Tianjin Port.

To validate the effectiveness of the proposed method, three sets of ship maneuvering experiments were simulated using a maritime simulator platform. Each set of simulation experiments was designed for specific scenarios of ship entry and exit from the port. Continuous motion dynamics data, including ship position, speed, and course information, were obtained by manipulating ships of different types and scales. The ship data used in the simulation experiments are presented in Table 1. Based on this, the TS-KDE method proposed in this paper was employed for the real-time detection of ship anomalous behavior.

Table 1. Experimental data for simulated ship maneuvering.

| ID | Ship Type | Ship Length (m) | Ship Width (m) | Abnormal Type |
|----|-----------|-----------------|----------------|-----------------------------|
| 1 | Cargo | 125 | 20 | Deviation from the waterway |
| 2 | Container | 170 | 28 | U-turn Within the waterway |
| 3 | Container | 231 | 35 | Stopping in the waterway |

3.2. Evaluation Index

Ship abnormal behavior detection is a typical binary classification method, whereby the detection results can be categorized as normal or abnormal. The simulated trajectory data are labeled as normal and abnormal, and then the TS-KDE method is utilized for

abnormal behavior detection. The performance of the TS-KDE method can be evaluated using various performance metrics, such as accuracy, precision, recall, et al. [45]. These evaluation indexes are related with the confusion matrix [46], as shown in Table 2. The three metrics chosen to evaluate the proposed TS-KDE are shown below.

- (1) Accuracy represents the proportion of samples with correct abnormal detection results among the total number of samples, as shown in Equation (12).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

- (2) Precision, also known as positive predictive value, represents the proportion of true positive samples among the samples detected as positive by abnormal behavior detection, as shown in Equation (13).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

- (3) Recall, also known as sensitivity or true positive rate, represents the proportion of actual positive samples among the positive samples detected by abnormal behavior detection in the entire set of positive samples, as shown in Equation (14).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

Table 2. Confusion matrix of abnormal behavior detection results.

| Confusion | Detection Results | |
|--------------|--------------------|--|
| | Abnormal | Normal |
| Ture results | Abnormal Normal | True Positive (TP) False Positive (FP) False Negative (FN) True Negative (TN) |

3.3. Experimental Results and Analysis

3.3.1. Deviation from the Waterway

Experiment 1: Simulating the navigation process of a cargo ship from the North Anchorage to the Main Waterway, reaching the Dry Bulk Terminal. The ship gradually accelerates from a stationary state at the North anchorage, and after entering the Main Waterway, it begins to violate the implemented navigation rules of the port.

Continuously generating new trajectory texts based on the time series of ship trajectories and performing real-time cosine similarity measurements with 40 types of pattern trajectory texts, the ship’s movement was found to match the inbound traffic pattern category 23, as shown in Figure 5a. Figure 5b displays the trajectories of traffic pattern category 23 and experimental ship 1. It can be seen from Figure 5b that the traffic pattern recognition method proposed in this paper can accurately identify the ship’s traffic pattern. Using the proposed ship abnormal behavior detection method to obtain the abnormal factor values of the ship’s behavior states, this paper identifies abnormal behavior when $Abnor_factor \geq 3$. To facilitate visual display, when the abnormal factor value is greater than 3, it is marked as 3, as shown in Figure 6a. Meanwhile, by comparing the course and speed information of experimental ship 1 with the speed and course of the corresponding traffic pattern, it was found that the ship violated the port’s navigational rules by diagonally entering the South Boat Waterway (a waterway dedicated to small ships leaving the port) after entering the waterway. Furthermore, the ship changed its course multiple times, accelerating to around 20 knots, significantly exceeding the speed and course distribution of historical pattern trajectories and the speed limits stipulated by the port. This process was marked as an anomaly, as shown in Figure 6b,c. In addition, the values of the three metrics accuracy, precision and recall of experimental ship 1 are shown in Figure 7.

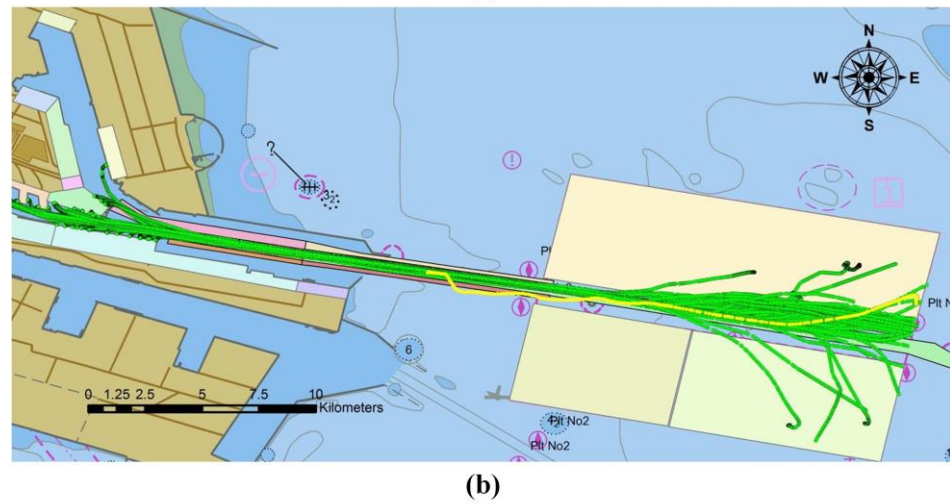
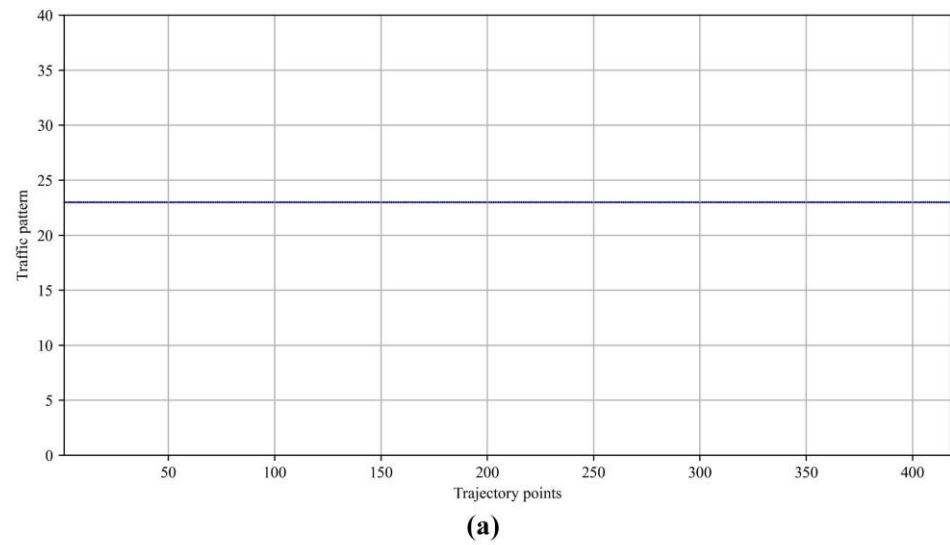


Figure 5. Recognition of inbound and outbound traffic patterns for experimental ship 1. (a) Traffic pattern followed by the experimental ship 1 when entering the port. (b) Trajectory of experimental ship 1 (yellow markers) and trajectory of traffic pattern category 23 (green markers).

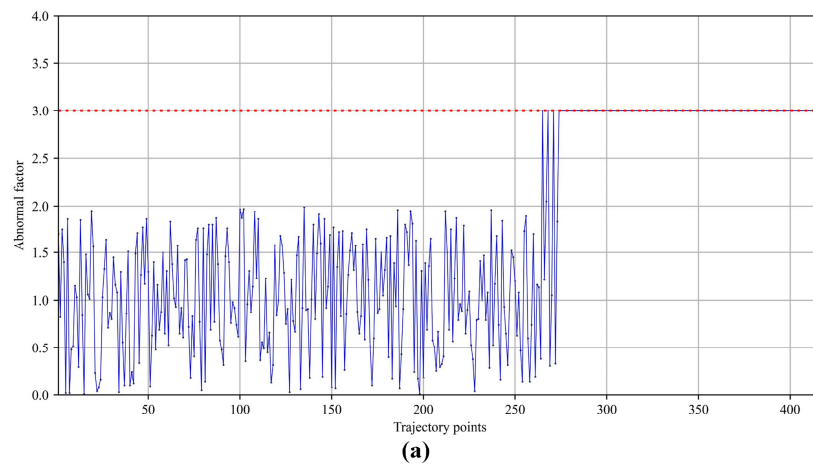


Figure 6. Cont.

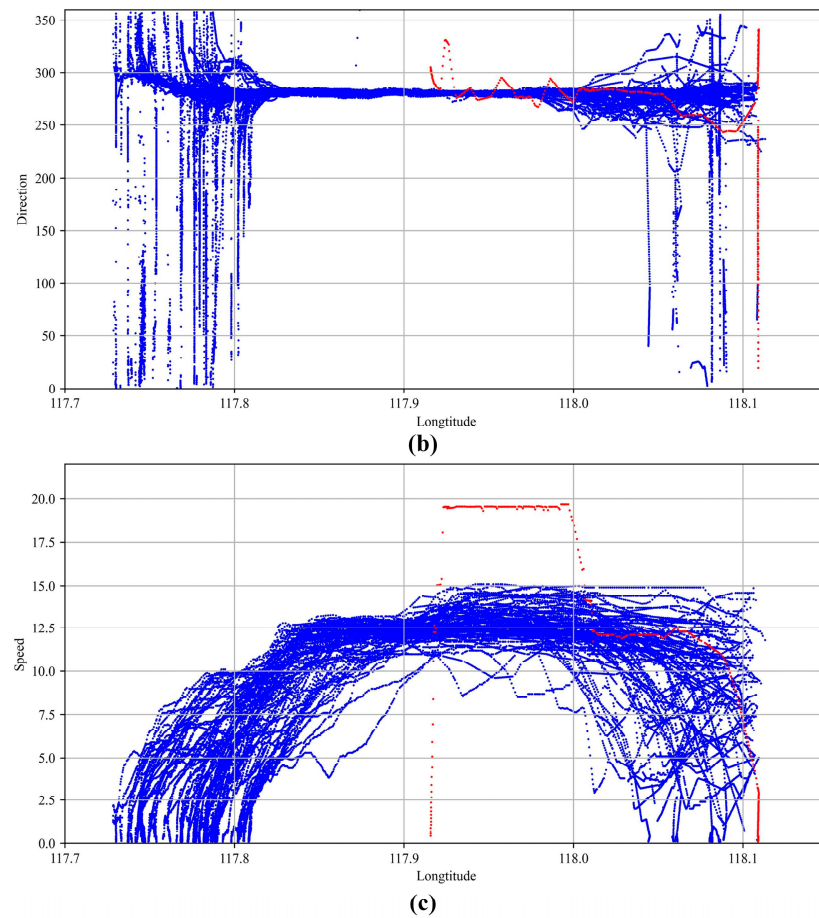


Figure 6. The results of anomaly behavior detection for experimental ship 1. (a) Values of the abnormal factor of the state of the motion of experimental ship 1. (b) Variation in the course of experimental ship 1 and traffic pattern category 23 with longitude. (c) Variation in the speed of experimental ship 1 and traffic pattern category 23 with longitude.

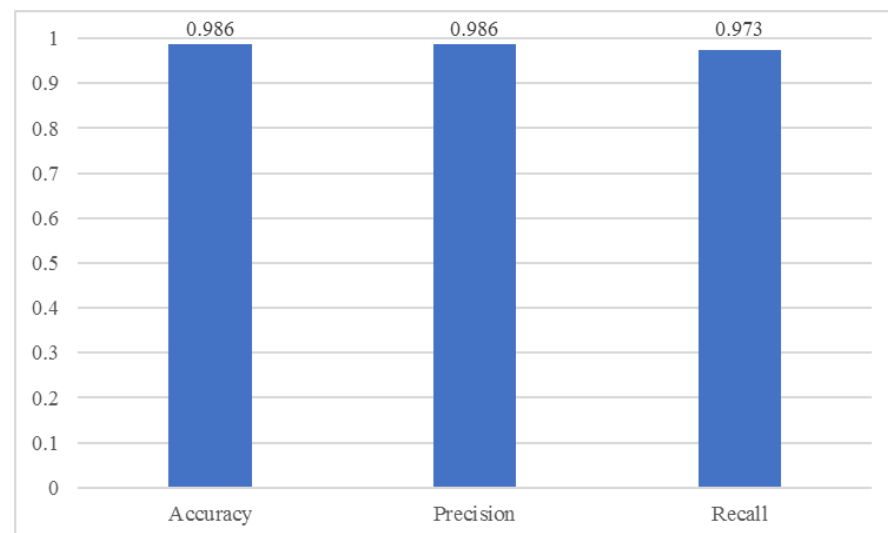


Figure 7. The values of the three metrics accuracy, precision and recall.

3.3.2. U-Turn within the Waterway

Experiment 2: Simulating the navigation process of a container ship from the North Anchorage to the operation area of the Dongjiang Container Terminal in the North Basin. In this process, the ship first enters the Main Waterway from the North Anchorage, then

crosses the Main Waterway to enter the South Boat Waterway, and finally performs a turning maneuver to enter the Main Waterway.

Continuously generating new trajectory texts based on the time series of ship trajectories and performing real-time cosine similarity measurements with 40 types of pattern trajectory texts, the ship’s movement is found to match inbound traffic pattern category 8, as shown in Figure 8a. Figure 8b shows the trajectory of traffic pattern category 8 and experimental ship 2. It can be seen in Figure 8b that the traffic pattern recognition method proposed in this paper can accurately identify the traffic pattern of the ship. Using the proposed ship abnormal behavior detection method to obtain the abnormal factor values of the ship behavior states, this paper identifies abnormal behavior when $Abnor_factor \geq 3$. To facilitate visual display, when the abnormal factor value is greater than 3, it is marked as 3, as shown in Figure 9a. Meanwhile, a comparison of the course and speed information of experimental ship 2 with the historical traffic pattern trajectory in terms of speed and course showed that, after entering the Main Waterway, the ship suddenly made a significant left turn to enter the South Boat Waterway (a waterway dedicated to small ship leaving the port). This action, until re-entering the Main Waterway, violated the port’s navigational rules. Additionally, the ship’s speed and course were below the average values of the historical pattern trajectory, affecting the normal navigation of the port. This process was marked as an anomaly, as shown in Figure 9b,c. In addition, the values of the three metrics of accuracy, precision and recall for experimental ship 2 are shown in Figure 10.

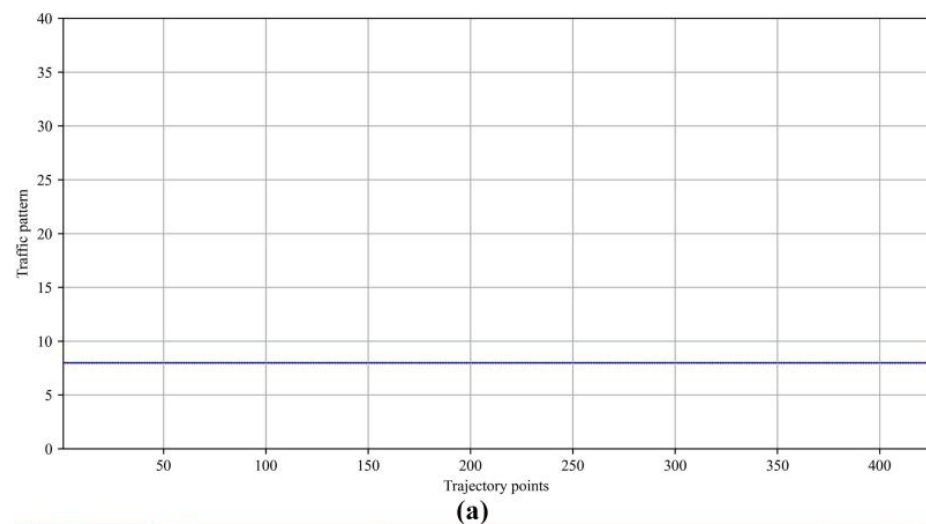


Figure 8. Recognition of inbound and outbound traffic patterns for experimental ship 2. (a) Traffic pattern followed by the experimental ship 2 when entering the port. (b) Trajectory of experimental ship 2 (yellow markers) and trajectory of traffic pattern category 8 (green markers).

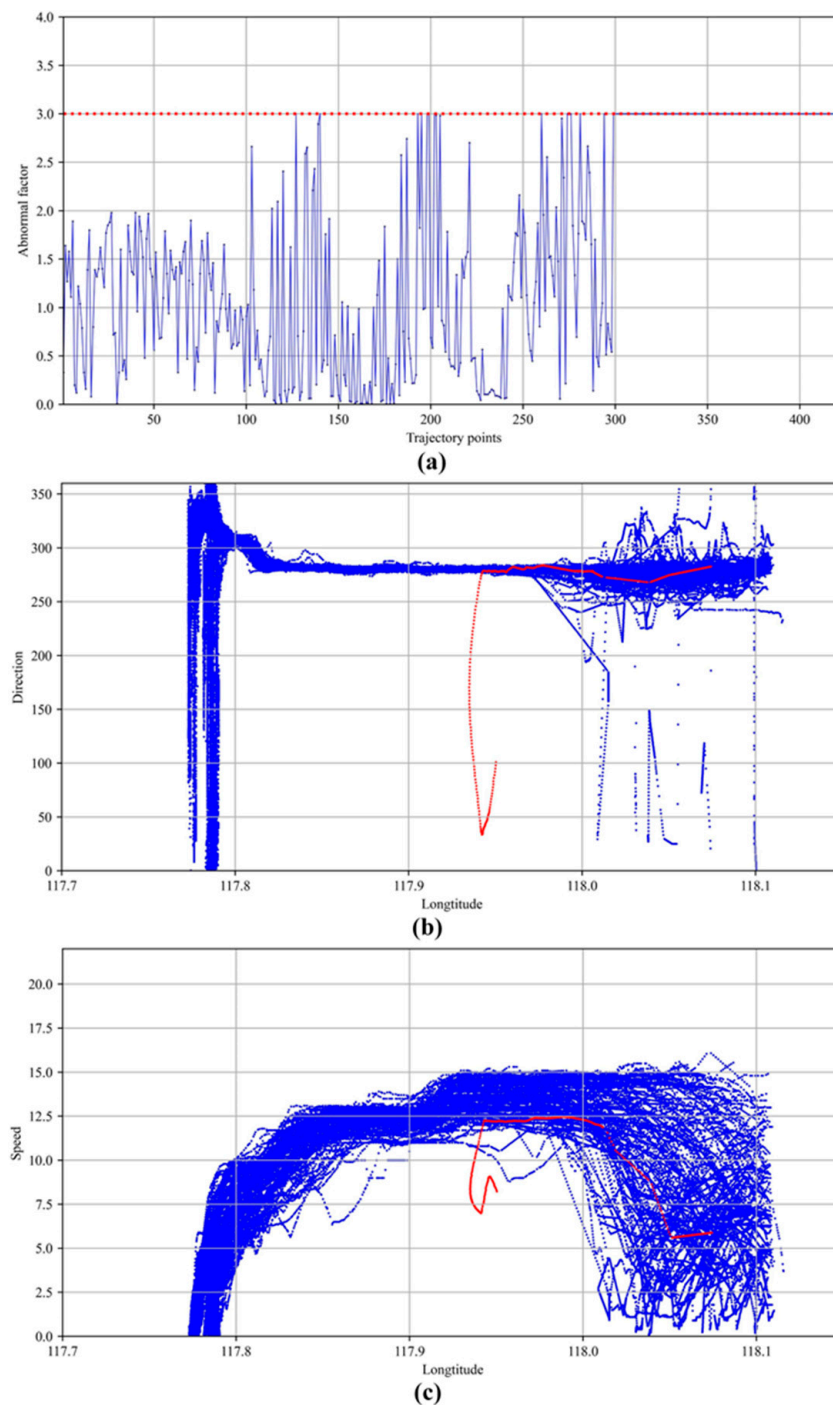


Figure 9. The results of anomaly behavior detection for experimental ship 2. (a) Values of the abnormal factor of the state of the motion of experimental ship 2. (b) Variation of course of experimental ship 2 and traffic pattern category 8 with longitude. (c) Variation of speed of experimental ship 2 and traffic pattern category 8 with longitude.

3.3.3. Stopping in the Waterway

Experiment 3: Simulating the navigation process of a container ship from the North Anchorage to the East Tuki Container Terminal in the West Basin. When the ship enters the Main Waterway from the North Anchorage, the ship suddenly decelerates until the speed drops to 0 knots.

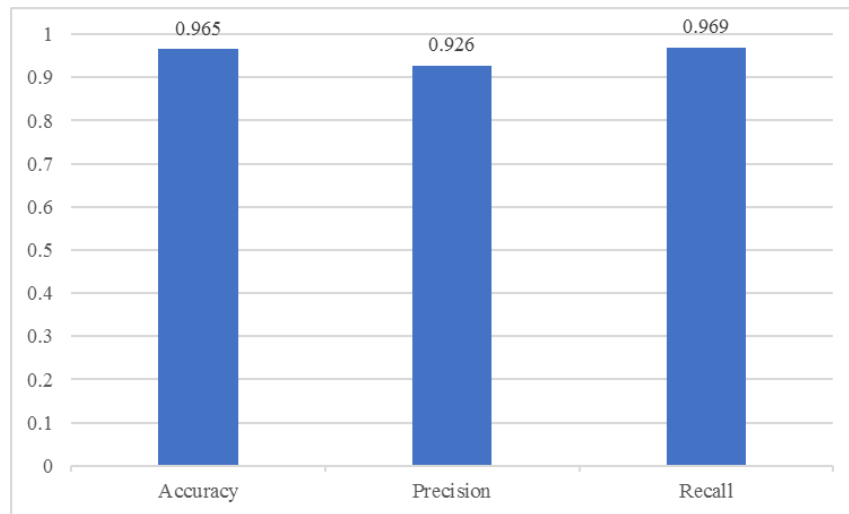


Figure 10. The values of the three metrics accuracy, precision and recall for experimental ship 2.

Continuously generating new trajectory texts based on the time series of ship trajectories and performing real-time cosine similarity measurements with 40 types of pattern trajectory texts, the ship’s movement is found to match inbound traffic pattern category 26, as shown in Figure 11a. Figure 11b shows the trajectory of traffic pattern category 26 and experimental ship 3. It can be seen from Figure 11b that the traffic pattern recognition method proposed in this paper can accurately identify the traffic pattern of the ship. Using the proposed ship abnormal behavior detection method to obtain the abnormal factor values of the ship’s behavior state, this paper identifies abnormal behavior when $Abnor_factor \geq 3$. To facilitate visual display, when the abnormal factor value is greater than 3, it is marked as 3, as shown in Figure 12a. Meanwhile, a comparison of the speed information of experimental ship 3 with the historical traffic pattern trajectory in terms of speed shows that the ship starts to decelerate when entering the main channel, slowing down until the speed drops to 0, affecting the normal navigation of the port, which clearly violates the port’s navigation rules. This process is marked as an anomaly, as shown in Figure 12b,c. In addition, the values of the three metrics accuracy, precision and recall of experimental ship 3 are shown in Figure 13.

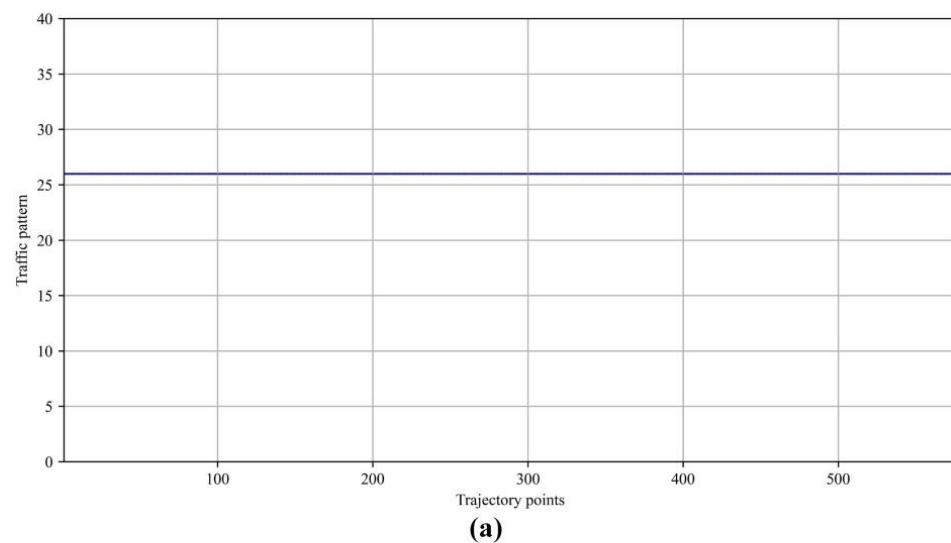


Figure 11. Cont.

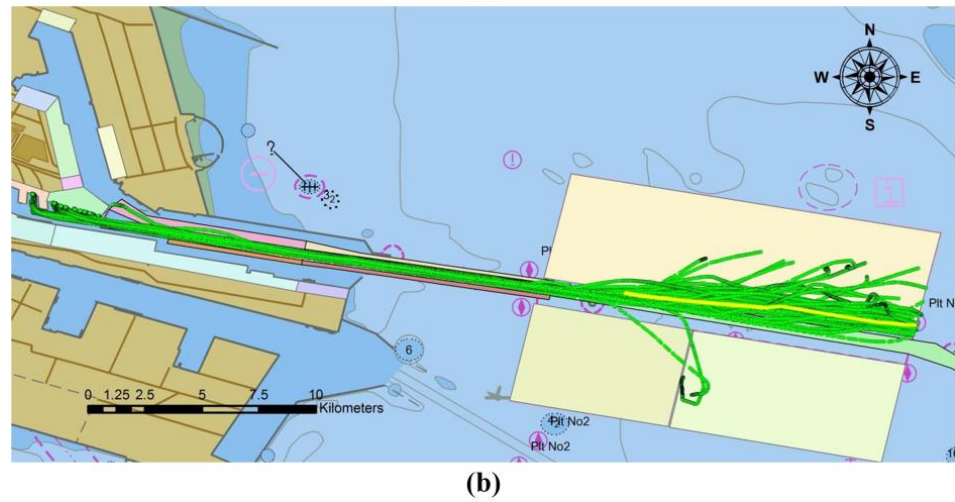


Figure 11. Recognition of inbound and outbound traffic patterns for experimental ship 3. (a) Traffic pattern followed by the experimental ship 3 when entering the port. (b) Trajectory of experimental ship 3 (yellow markers) and trajectory of traffic pattern category 26 (green markers).

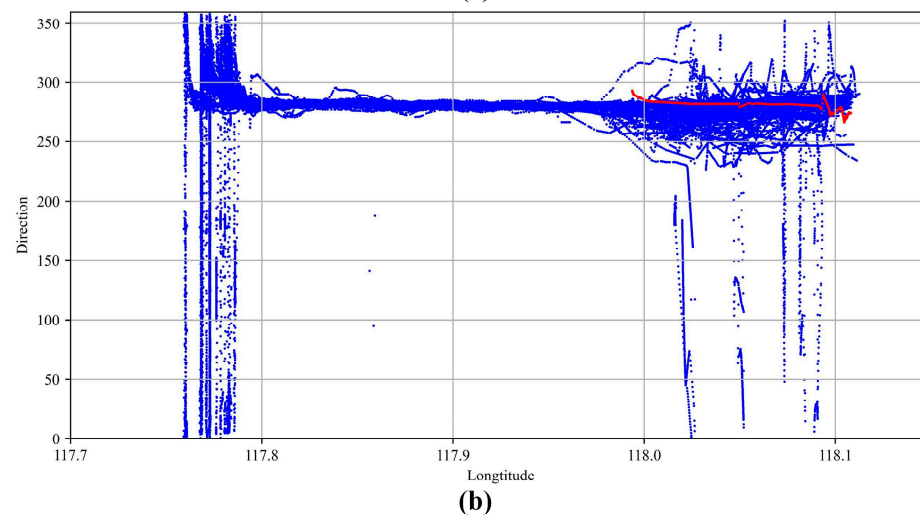
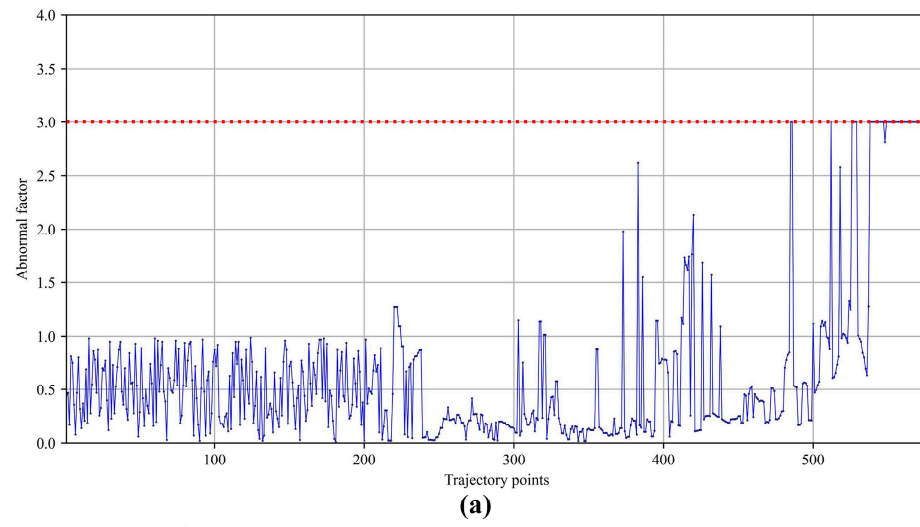


Figure 12. Cont.

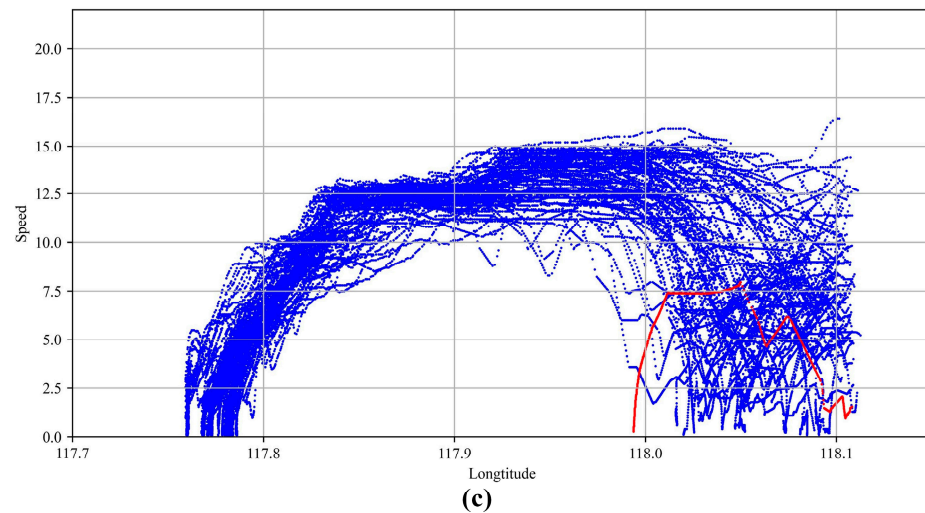


Figure 12. The results of anomaly behavior detection for experimental ship 3. (a) Values of the abnormal factor of the state of the motion of experimental ship 3. (b) Variation of course of experimental ship 3 and traffic pattern category 26 with longitude. (c) Variation of speed of experimental ship 3 and traffic pattern category 26 with longitude.

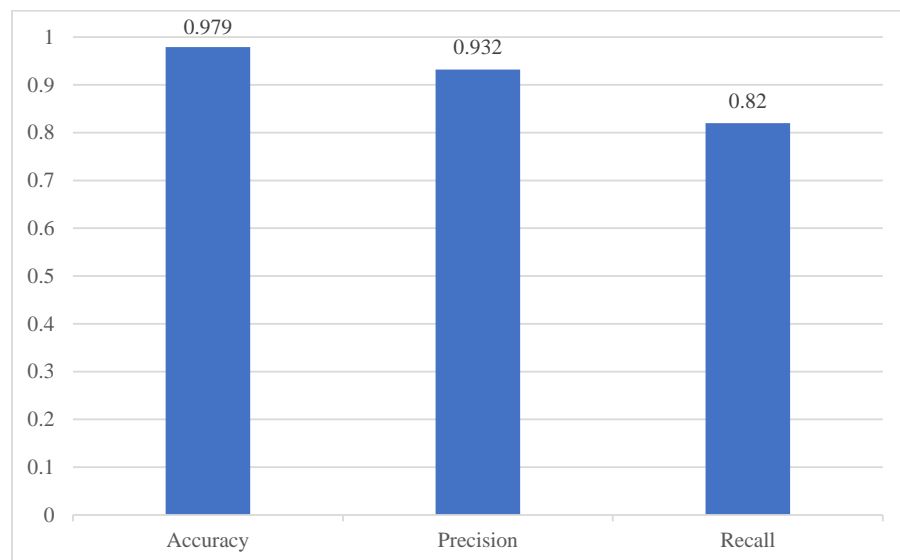


Figure 13. The values of the three metrics accuracy, precision, and recall of experimental ship 3 are shown in Figure 13.

Table 3 presents the statistical results of anomaly detection for the four sets of simulated experimental data. Each experiment corresponds to different types and scales of ship adopting different traffic patterns to enter and leave the port. The model designed in this paper can detect abnormal ship behavior in all four experiments, demonstrating the model’s versatility. In addition, with the increase in the number of samples, the time for anomaly detection will increase accordingly. However, the average time spent identifying each ship’s behavior state remains relatively constant. The anomaly detection model performs well in real time and can identify potential abnormal behaviors. The model comprehensively evaluates whether ship behavior is abnormal based on the ship’s motion state (position, direction, and speed). Compared with the labeled abnormal trajectory points, the proposed method achieves an average accuracy of over 90% in identifying abnormal behaviors.

Table 3. Detection results of abnormal ship behavior in three groups of simulation experiments.

| Experiment | Total Number of Trajectories | True Positive Anomalies | Detecting Anomalous Points | Accuracy (%) | Precision (%) | Recall (%) | Average Time (s) |
|------------|------------------------------|-------------------------|----------------------------|--------------|---------------|------------|------------------|
| 1 | 417 | 148 | 144 | 98.6 | 98.6 | 97.3 | 0.085 |
| 2 | 424 | 130 | 126 | 96.5 | 92.6 | 96.9 | 0.095 |
| 3 | 576 | 50 | 41 | 97.9 | 93.2 | 82 | 0.091 |
| Average | / | / | / | 97.6 | 94.8 | 92.1 | 0.090 |

4. Discussion

This paper introduces a novel method for detecting abnormal ship behavior in port waterways based on text similarity and kernel density estimation. Given the assumption of obtaining ship traffic patterns when entering and leaving the port, a semantic transformation method is employed to convert each traffic pattern into pattern trajectory text. The method utilizes text similarity measurement to identify the traffic pattern of the target ship and uses kernel density estimation to detect abnormal ship behavior. Compared to previous research, the proposed method has several advantages:

- (1) Compared with traditional ship traffic pattern recognition methods, this paper converts ship type, course, speed and geospatial attribute semantics into ship trajectory text, ensuring the integrity of ship movement information. By comparing the similarity of texts, ship traffic patterns can be improved in terms of the efficiency and accuracy of identification results. The traditional ship traffic pattern identification method calculates the similarity based on the distance between trajectories, which can easily lead to deviations in ship traffic pattern recognition, leading to false alarms in detecting abnormal ship behavior. In addition, the traditional ship traffic pattern identification method measures the similarity between the target ship trajectory and all trajectories in each traffic pattern, identifies its traffic pattern based on the similarity measurement value, or finds out the typical trajectories in each traffic pattern, and calculates the similarity value between target trajectories and typical trajectories to identify traffic patterns. The above similarity measurement method requires calculating the distance between trajectory points, which is inefficient. The cosine similarity measure is a method based on vector calculation. Based on the vector representation of the target trajectory text and the traffic pattern trajectory text, the traffic pattern can be identified by simply calculating the cosine similarity between vectors;
- (2) Combined with kernel density estimation to construct a ship abnormal behavior detection method, the abnormal factor value can be set according to the tolerance of abnormal behavior. This paper constructs a calculation method for ship behavior abnormality factors and sets corresponding abnormality thresholds according to the actual needs of the port. For example, to identify potential or suspicious ship behavior, the threshold should be set to a lower value to identify and take response actions quickly. The threshold can be set to a higher value to prevent false alarms and increase the burden on drivers on duty to reduce work pressure.

However, the methods proposed in this paper also have the following limitations:

- (1) This method is based on the assumption that ships enter and leave ports according to port navigation rules. Therefore, it mainly considers the detection of abnormal behaviors of single vessels entering and leaving ports, lacking consideration of detecting ship abnormal behaviors in multi-vessel interaction scenarios. In typical situations, ships navigate within specific channels, making it difficult to encounter situations between vessels.
- (2) Since this method is data-driven, its results may be influenced by the quality of ship trajectory data, especially in ship traffic pattern recognition. In future research, consideration could be given to integrating other maritime data, such as port water depth data, to improve the accuracy of ship abnormal behavior detection. Additionally,

the selection of threshold values for abnormal factors significantly affects the results of abnormal detection. Therefore, appropriate thresholds should be set based on specific scenarios or environments.

5. Conclusions and Future Work

In this paper, we proposed a method for detecting anomalous ship behavior in port waters based on text similarity and kernel density estimation. Taking into account the impact of port traffic rules on ship movements, a text similarity measurement method is used to identify ship traffic patterns, and kernel density estimation is used to detect anomalous ship behavior violating port traffic rules in waterways. This study used natural language processing methods to address the challenge of recognizing traffic patterns in complex waterways. Additionally, we construct a ship behavior anomaly factor based on kernel density estimation, effectively detecting abnormal ship behavior against port traffic rules, with experimental verification results demonstrating an average accuracy exceeding 90%. From the perspective of inbound and outbound traffic patterns, this research detects potential anomalous ship behavior in waterways, enriching methods for detecting abnormal ship behavior in port waters and demonstrating engineering application potential for ship supervision.

In future research, employing trajectory data with a longer time span could enhance the accuracy of ship traffic pattern recognition. Additionally, it is advisable to consider integrating satellite imagery, radar data, cargo composition, and contextual information about the navigational environment to enrich the maritime traffic information, thereby further improving the accuracy of anomaly behavior detection [47]. In addition, when abnormal ship behavior is identified, supervisory personnel need to make further confirmations to clarify the cause of the ship's abnormal behavior so as to prevent false alarms due to the ship's normal navigation adjustments [48].

Author Contributions: Conceptualization, G.L., X.Z. and Y.S.; methodology, G.L., X.Z. and Y.S.; software, G.L. and X.Z.; validation, G.L. and X.Z.; formal analysis, G.L., X.Z. and Y.S.; investigation, G.L. and X.Z.; resources, X.Z.; data curation, G.L. and X.Z.; writing—original draft preparation, G.L., X.Z. and C.W.; writing—review and editing, G.L., X.Z., Y.S., C.W., W.G. and J.W.; visualization, G.L., C.W., W.G. and J.W.; supervision, X.Z. and Y.S.; project administration, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (52371359) and the Dalian Science and Technology Innovation Fund (2022JJ12GX015).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

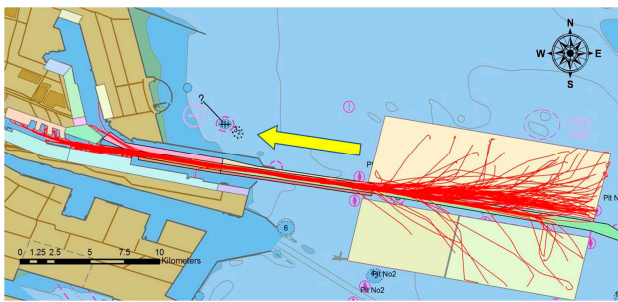
Data Availability Statement: Data are contained within the article.

Acknowledgments: This research was supported by the Navigation College of Dalian Maritime University.

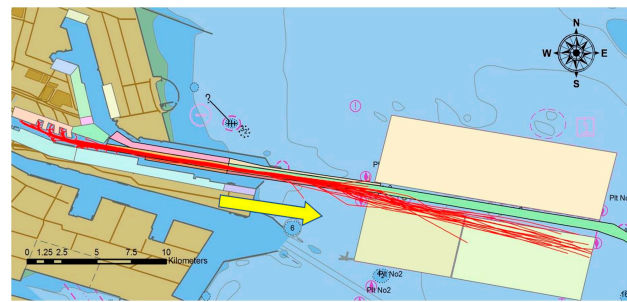
Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

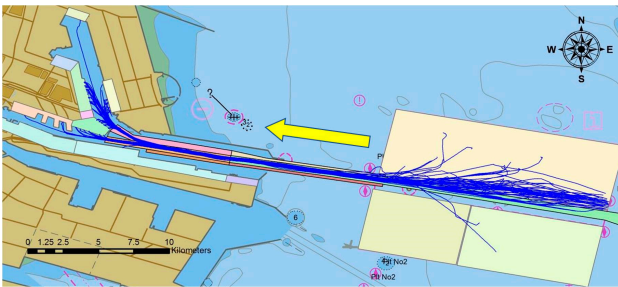
In this paper, we visualized ship traffic patterns based on ArcGIS with Electronic Navigational Charts (ENCs) [44]. Cargo ship trajectories are shown as red lines, and container ship trajectories are shown in blue. We use yellow arrows to indicate the direction of motion of the trajectory cluster.



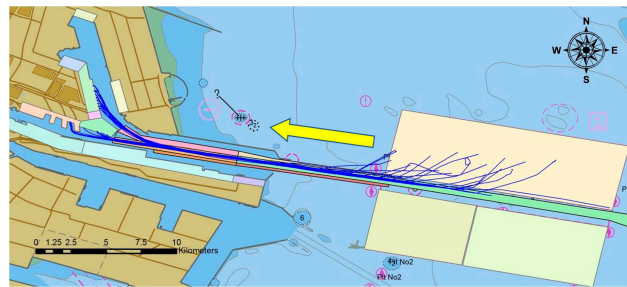
(1) Traffic pattern 1



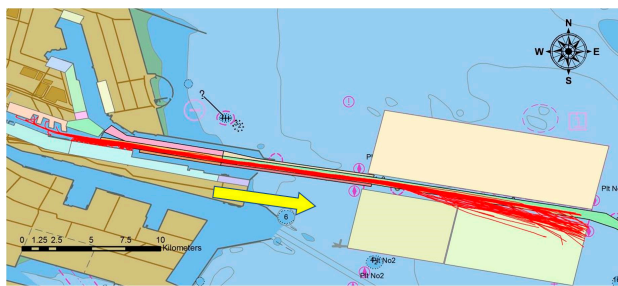
(2) Traffic pattern 2



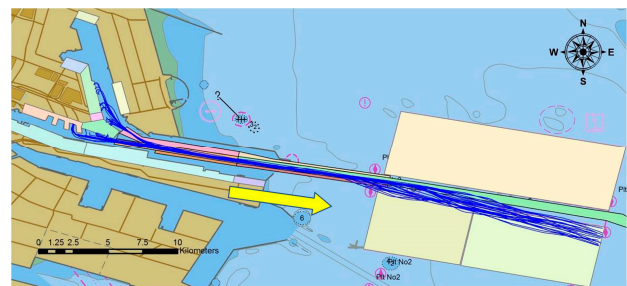
(3) Traffic pattern 3



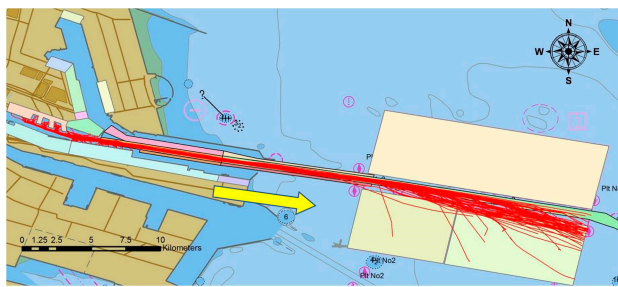
(4) Traffic pattern 4



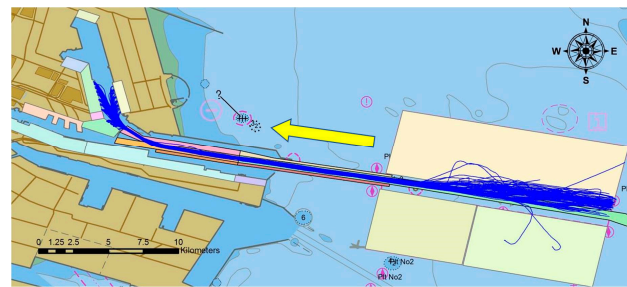
(5) Traffic pattern 5



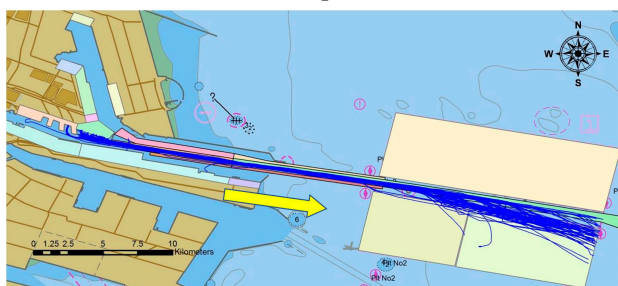
(6) Traffic pattern 6



(7) Traffic pattern 7



(8) Traffic pattern 8

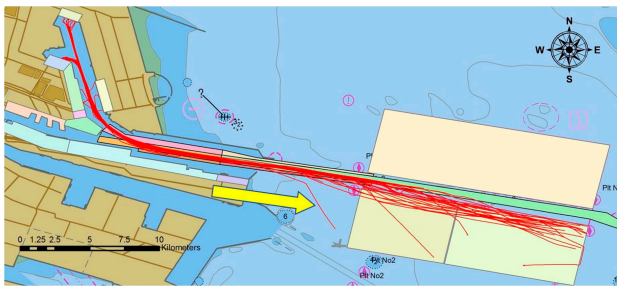


(9) Traffic pattern 9

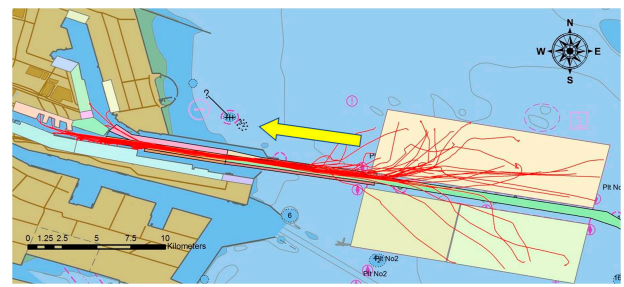


(10) Traffic pattern 10

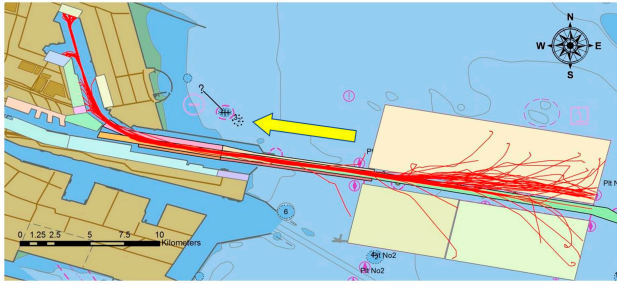
Figure A1. Cont.



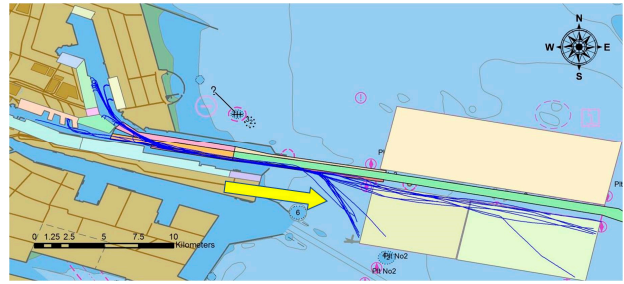
(11) Traffic pattern 11



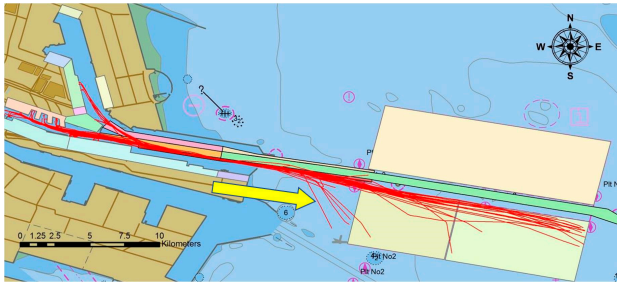
(12) Traffic pattern 12



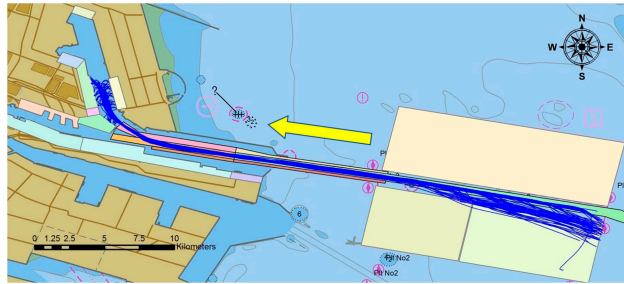
(13) Traffic pattern 13



(14) Traffic pattern 14



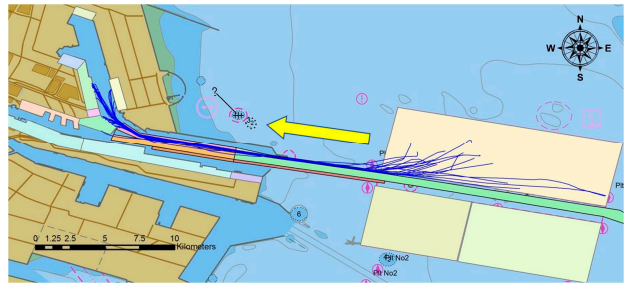
(15) Traffic pattern 15



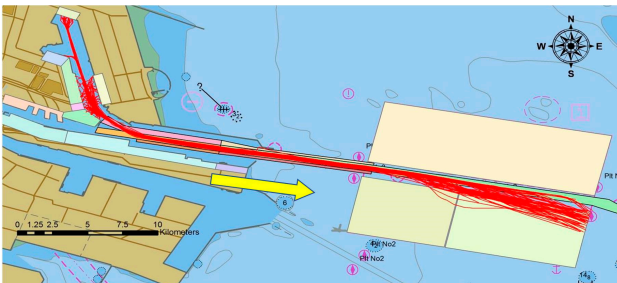
(16) Traffic pattern 16



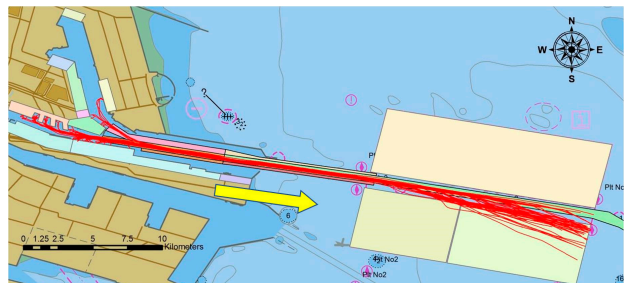
(17) Traffic patte17



(18) Traffic pattern 18

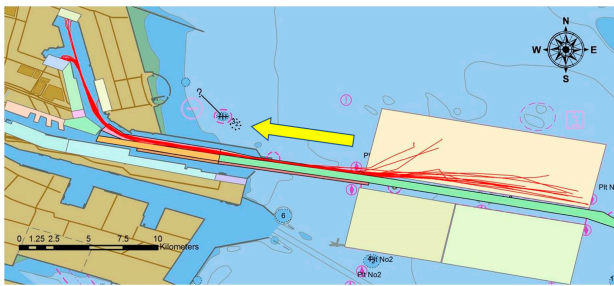


(19) Traffic pattern 19

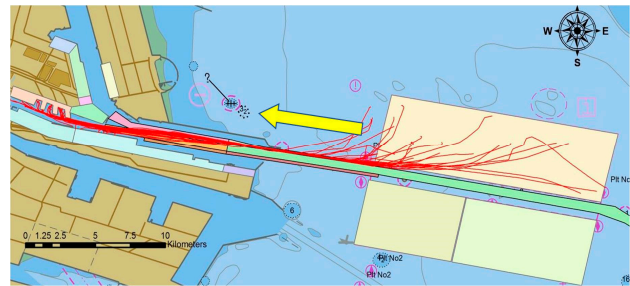


(20) Traffic pattern 20

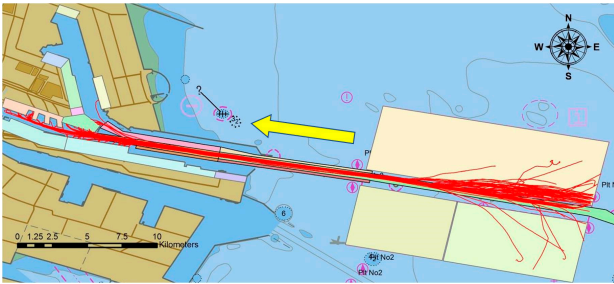
Figure A1. Cont.



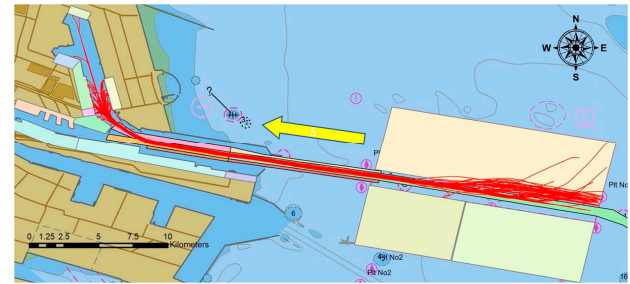
(21) Traffic pattern 21



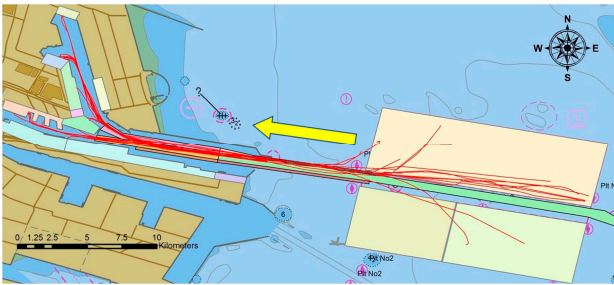
(22) Traffic pattern 22



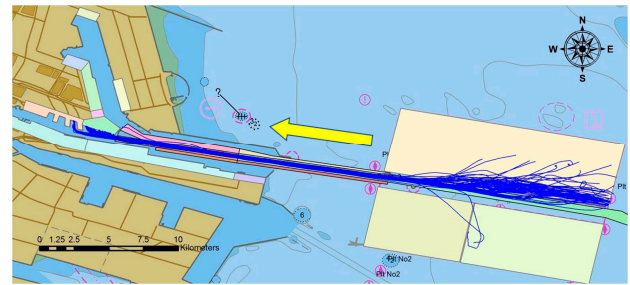
(23) Traffic pattern 23



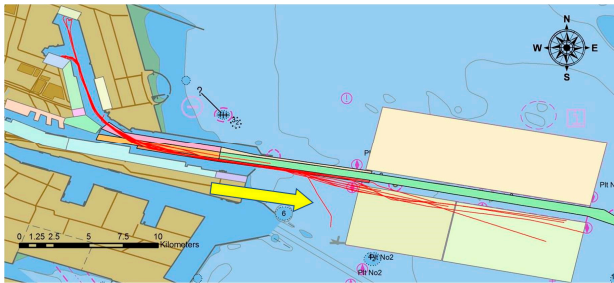
(24) Traffic pattern 24



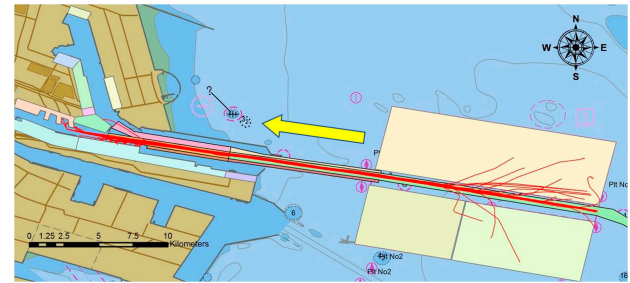
(25) Traffic pattern 25



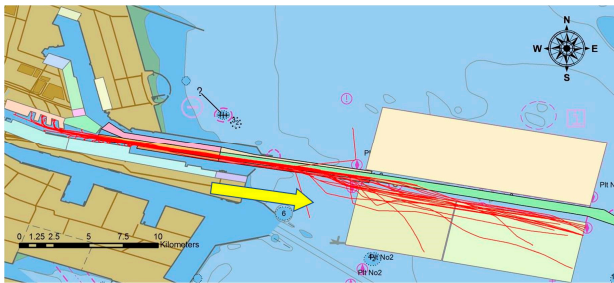
(26) Traffic pattern 26



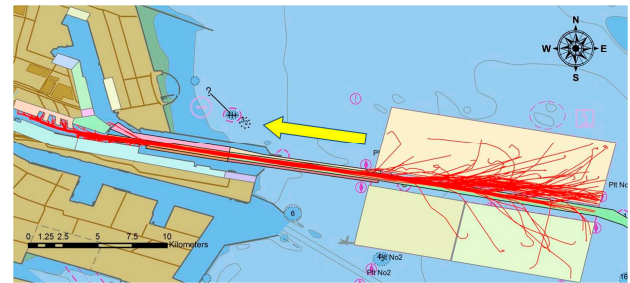
(27) Traffic pattern 27



(28) Traffic pattern 28

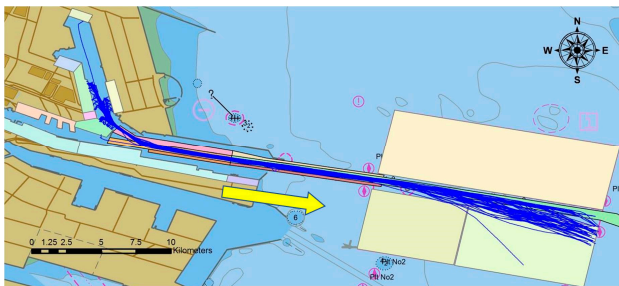


(29) Traffic pattern 29

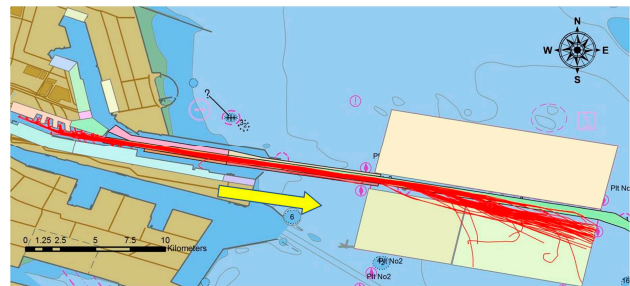


(30) Traffic pattern 30

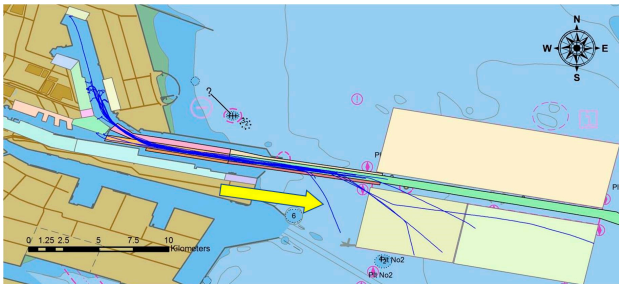
Figure A1. Cont.



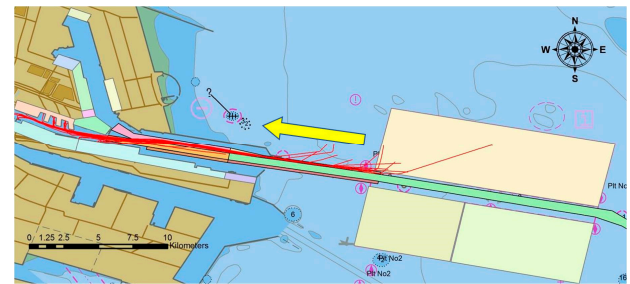
(31) Traffic pattern 31



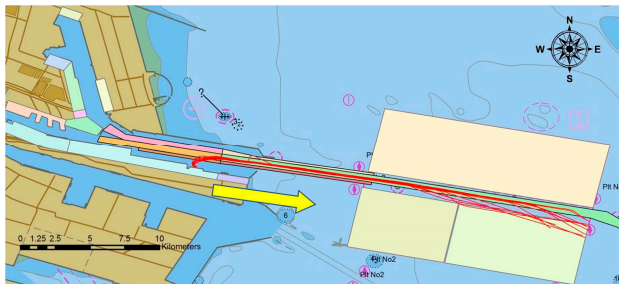
(32) Traffic pattern 32



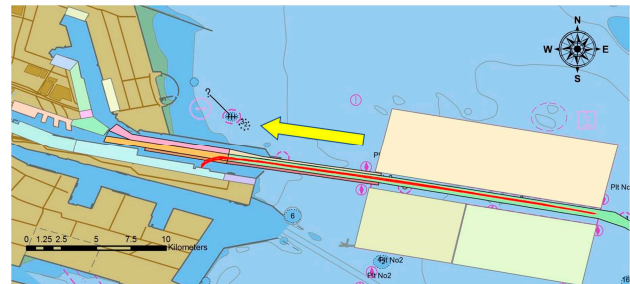
(33) Traffic pattern 33



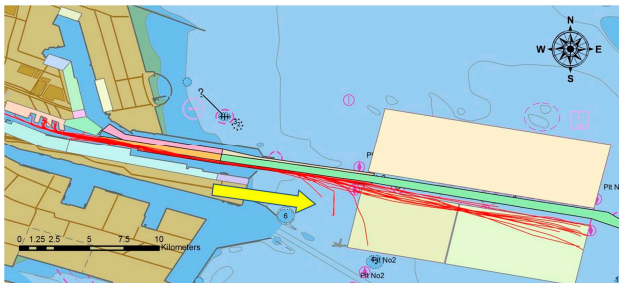
(34) Traffic pattern 34



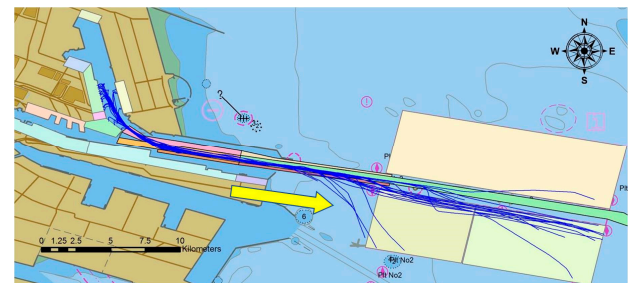
(35) Traffic pattern 35



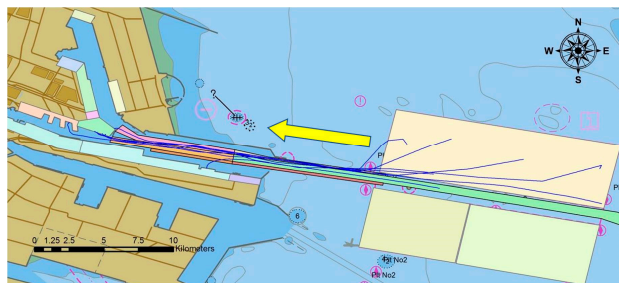
(36) Traffic pattern 36



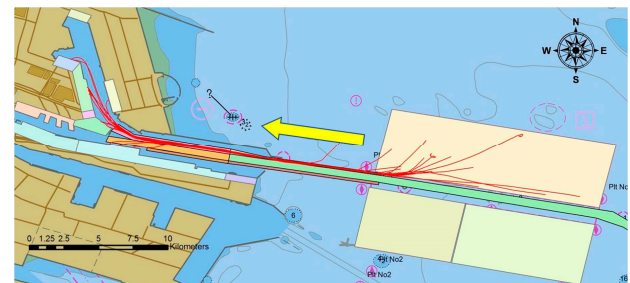
(37) Traffic pattern 37



(38) Traffic pattern 38



(39) Traffic pattern 39



(40) Traffic pattern 40

Figure A1. The traffic patterns of ships entering and leaving Tianjin Port.

References

1. EMSA (European Maritime Safety Agency). Annual Overview of Marine Casualties and Incidents. 2022. Available online: <https://www.emsa.europa.eu/newsroom/latest-news/item/4867-annual-overview-of-marine-casualties-and-incidents-2021.html> (accessed on 1 March 2024).
2. UNCTAD. Review of Maritime Transportation 2022. Available online: https://unctad.org/system/files/official-document/rmt2022_en.pdf (accessed on 1 March 2024).
3. Bai, X.; Cheng, L.; Iris, Ç. Data-driven financial and operational risk management: Empirical evidence from the global tramp shipping industry. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *158*, 102617. [[CrossRef](#)]
4. Zhang, X.; Li, R.; Wang, C.; Xue, B.; Guo, W. Robust optimization for a class of ship traffic scheduling problem with uncertain arrival and departure times. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108257. [[CrossRef](#)]
5. Wang, C.; Zhang, X.; Gao, H.; Bashir, M.; Li, H.; Yang, Z. Optimizing Anti-collision Strategy for MASS: A Safe Reinforcement Learning Approach to Improve Maritime Traffic Safety. *Ocean Coast. Manag.* **2024**, *253*, 107161. [[CrossRef](#)]
6. Zheng, K.; Zhang, X.; Wang, C.; Li, Y.; Cui, J.; Jiang, L. Adaptive collision avoidance decisions in autonomous ship encounter scenarios through rule-guided vision supervised learning. *Ocean Eng.* **2024**, *297*, 117096. [[CrossRef](#)]
7. Shu, Y.; Han, B.; Song, L.; Yan, T.; Gan, L.; Zhu, Y.; Zheng, C. Analyzing the spatio-temporal correlation between tide and shipping behavior at estuarine port for energy-saving purposes. *Appl. Energy.* **2024**, *367*, 123382. [[CrossRef](#)]
8. Liang, M.; Weng, L.; Gao, R.; Li, Y.; Du, L. Unsupervised maritime anomaly detection for intelligent situational awareness using AIS data. *Knowl.-Based Syst.* **2024**, *284*, 111313. [[CrossRef](#)]
9. Sidibé, A.; Gao, S. Study of automatic anomalous behaviour detection techniques for maritime vessels. *J. Navig.* **2017**, *70*, 847–858. [[CrossRef](#)]
10. Laxhammar, R. Anomaly detection for sea surveillance. In Proceedings of the 2008 11th International Conference on Information Fusion, Cologne, Germany, 30 June–3 July 2008; pp. 1–8.
11. Zhang, B.; Ren, H.; Wang, P.; Wang, D. Research Progress on Ship Anomaly Detection Based on Big Data. In Proceedings of the 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 16–18 October 2020; pp. 316–320. [[CrossRef](#)]
12. Rong, H.; Teixeira, A.P.; Soares, C.G. Data mining approach to shipping route characterization and anomaly detection based on AIS data. *Ocean Eng.* **2020**, *198*, 106936. [[CrossRef](#)]
13. Laxhammar, R.; Falkman, G.; Sviestins, E. Anomaly detection in sea traffic—a comparison of the gaussian mixture model and the kernel density estimator. In Proceedings of the 2009 12th International Conference on Information Fusion, Seattle, WA, USA, 6–9 July 2009; pp. 756–763.
14. Mascaro, S.; Nicholso, A.E.; Korb, K.B. Anomaly detection in vessel tracks using Bayesian networks. *Int. J. Approx. Reason.* **2014**, *55*, 84–98. [[CrossRef](#)]
15. Farahnakian, F.; Nicolas, F.; Farahnakian, F.; Nevalainen, P.; Sheikh, J.; Heikkonen, J.; Raduly-Baka, C. A Comprehensive Study of Clustering-Based Techniques for Detecting Abnormal Vessel Behaviour. *Remote Sens.* **2023**, *15*, 1477. [[CrossRef](#)]
16. Zhao, L.; Shi, G. Maritime anomaly detection using density-based clustering and recurrent neural network. *J. Navig.* **2019**, *72*, 894–916. [[CrossRef](#)]
17. Karatas, G.B.; Karagoz, P.; Ayran, O. Trajectory pattern extraction and anomaly detection for maritime vessels. *Internet Things* **2021**, *16*, 100436. [[CrossRef](#)]
18. Pallotta, G.; Vespe, M.; Bryan, K. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy* **2013**, *15*, 2218–2245. [[CrossRef](#)]
19. Zhen, R.; Jin, Y.; Hu, Q.; Shao, Z.; Nikitakos, N. Maritime anomaly detection within coastal waters based on vessel trajectory clustering and Naïve Bayes Classifier. *J. Navig.* **2017**, *70*, 648–670. [[CrossRef](#)]
20. Botts, C.H. A novel metric for detecting anomalous ship behaviour using a variation of the DBSCAN clustering algorithm. *SN Comput. Sci.* **2021**, *2*, 412. [[CrossRef](#)]
21. Liu, B.; de Souza, E.N.; Matwin, S.; Sydow, M. Knowledge-based clustering of ship trajectories using density-based approach. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014; pp. 603–608. [[CrossRef](#)]
22. Radon, A.N.; Wang, K.; Glässer, U.; Wehn, H.; Westwell-Roper, A. Contextual verification for false alarm reduction in maritime anomaly detection. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015; pp. 1123–1133. [[CrossRef](#)]
23. Gamage, C.; Dinalankara, R.; Samarabandu, J.; Subasinghe, A. A comprehensive survey on the applications of machine learning techniques on maritime surveillance to detect abnormal maritime vessel behaviours. *WMU J. Marit. Aff.* **2023**, *22*, 447–477. [[CrossRef](#)]
24. Wang, S.; Zhang, X.; Qin, Y.; Song, W.; Li, B. Marine Target Magnetic Anomaly Detection Based on Multi-Task Deep Transfer Learning. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [[CrossRef](#)]
25. Rhodes, B.J.; Bomberger, N.A.; Zandipour, M. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. In Proceedings of the 2007 10th International Conference on Information Fusion, Quebec, QC, Canada, 9–12 July 2007; pp. 1–8. [[CrossRef](#)]

26. Huang, G.; Lai, S.; Ye, C.; Zhou, H. Ship trajectory anomaly detection based on multi-feature fusion. In Proceedings of the 2021 IEEE International Conference on Smart Data Services (SMDS), Chicago, IL, USA, 5–10 September 2021; pp. 72–81. [\[CrossRef\]](#)
27. Hu, J.; Kaur, K.; Lin, H.; Wang, X.; Hassan, M.M.; Razzak, I.; Hammoudeh, M. Intelligent anomaly detection of trajectories for IoT empowered maritime transportation systems. *IEEE Trans. Intell. Transp. Syst.* **2022**, *24*, 2382–2391. [\[CrossRef\]](#)
28. Nguyen, D.; Vadaine, R.; Hajduch, G.; Garello, R.; Fablet, R. GeoTrackNet—A maritime anomaly detector using probabilistic neural network representation of AIS tracks and a contrario detection. *IEEE trans Intell. Transp. Syst.* **2021**, *23*, 5655–5667. [\[CrossRef\]](#)
29. Eljabu, L.; Etemad, M.; Matwin, S. Anomaly detection in maritime domain based on spatio-temporal analysis of ais data using graph neural networks. In Proceedings of the 2021 5th International Conference on Vision, Image and Signal Processing (ICVISIP), Kuala Lumpur, Malaysia, 18–20 December 2021; pp. 142–147. [\[CrossRef\]](#)
30. Zhou, Y.; Daamen, W.; Vellinga, T.; Hoogendoorn, S. Review of maritime traffic models from vessel behaviour modeling perspective. *Transp. Res. Part C Emerg. Technol.* **2019**, *105*, 323–345. [\[CrossRef\]](#)
31. Dorsey, L.C.; Wang, B.; Grabowski, M.; Merrick, J.; Harrald, J.R. Self healing databases for predictive risk analytics in safety-critical systems. *J. Loss Prev. Process Ind.* **2020**, *63*, 104014. [\[CrossRef\]](#)
32. Rawson, A.; Brito, M. A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis. *Transp. Rev.* **2023**, *43*, 108–130. [\[CrossRef\]](#)
33. Bai, X.; Zhang, X.; Li, K.X.; Zhou, Y.; Yuen, K.F. Research topics and trends in the maritime transport: A structural topic model. *Transp. Policy.* **2021**, *102*, 11–24. [\[CrossRef\]](#)
34. Hughes, P.; Shipp, D.; Figueres-Esteban, M.; Van Gulijk, C. From free-text to structured safety management: Introduction of a semi-automated classification method of railway hazard reports to elements on a bow-tie diagram. *Saf. Sci.* **2018**, *110*, 11–19. [\[CrossRef\]](#)
35. Huang, L.; Wen, Y.; Guo, W.; Zhu, X.; Zhou, C.; Zhang, F.; Zhu, M. Mobility pattern analysis of ship trajectories based on semantic transformation and topic model. *Ocean Eng.* **2020**, *201*, 107092. [\[CrossRef\]](#)
36. Li, G.; Liu, M.; Zhang, X.; Wang, C.; Lai, K.H.; Qian, W. Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model. *J. Mar. Sci. Eng.* **2022**, *10*, 2012. [\[CrossRef\]](#)
37. Hu, W.; Gao, J.; Li, B.; Wu, O.; Du, J.; Maybank, S. Anomaly detection using local kernel density estimation and context-based regression. *IEEE Trans. Knowl. Data Eng.* **2018**, *32*, 218–233. [\[CrossRef\]](#)
38. Prakoso, D.W.; Abdi, A.; Amrit, C. Short text similarity measurement methods: A review. *Soft Comput.* **2021**, *25*, 4699–4723. [\[CrossRef\]](#)
39. Wang, J.; Dong, Y. Measurement of text similarity: A survey. *Information.* **2020**, *11*, 421. [\[CrossRef\]](#)
40. Ma, Y.; Zhao, J.; Su, J.; Xi, T. Outlier mining method based on kernel density estimation. *J. Taiyuan Univ. Sci. Tech.* **2020**, *41*, 456–462+469.
41. Sheather, S.J.; Jones, M.C. A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc.* **1991**, *53*, 683–690. [\[CrossRef\]](#)
42. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*, 1st ed.; Routledge: New York, NY, USA, 2018. [\[CrossRef\]](#)
43. Ristic, B.; La Scala, B.; Morelande, M.; Gordon, N. Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In Proceedings of the 2008 11th International Conference on Information Fusion, Cologne, Germany, 30 June–3 July 2008; pp. 1–7.
44. Li, G.; Zhang, X.; Jiang, L.; Wang, C.; Huang, R.; Liu, Z. An approach for traffic pattern recognition integration of ship AIS data and port geospatial features. *Geo-Spat. Inf. Sci.* **2024**, 1–28. [\[CrossRef\]](#)
45. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process Manag.* **2009**, *45*, 427–437. [\[CrossRef\]](#)
46. Visa, S.; Ramsay, B.; Ralescu, A.; Van Der Knaap, E. Confusion Matrix-Based Feature Selection. In Proceedings of the Twenty Second Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, OH, USA, 16–17 April 2011; Volume 710, pp. 120–127.
47. Bilican, M.S.; Iris, Ç.; Karatas, M.A. collaborative decision support framework for sustainable cargo composition in container shipping services. *Ann. Oper. Res.* **2024**, 1–33. [\[CrossRef\]](#)
48. Venturini, G.; Iris, Ç.; Kontovas, C.A.; Larsen, A. The multi-port berth allocation problem with speed optimization and emission considerations. *Transp. Res. Part D Transp. Environ.* **2017**, *54*, 142–159. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.