Research Progress on Ship Anomaly Detection Based on Big Data

Bohan Zhang, Hongxiang Ren, Pengjie Wang and Delong Wang
Key Laboratory of Marine Dynamic Simulation
Dalian Maritime University
Dalian, Liaoning Province, China
zhangbh.1997@gmail.com

Abstract—**The purpose of ship behavior anomaly detection is to identify and monitor some non-expected behaviors of ships, so as to improve the navigation safety of ships. Its research is of great significance to the safety guarantee of maritime navigation, intelligent monitoring of sea areas and the development of port management. This paper summarizes and evaluates the research progress of ship anomaly detection based on big data and points out the future development trend. First of all, the concept of ship abnormal behavior is introduced, and the process of data-driven ship abnormal detection and its data basis are described in detail. Secondly, the data-driven ship anomaly detection methods are divided into statistical method, machine learning method and neural network method, and their research status and existing problems are reviewed respectively. Finally, focusing on maritime big data, temporal and spatial correlation of scenarios, online real-time anomaly detection and other aspects, the current problems and challenges in the study of ship anomaly detection are discussed, and the future research direction is introduced.**

Keywords-abnormal detection; maritime big data; machine learning; neural network; progress review;

I. Introduction

With the increasing density of ocean traffic, accidents such as shipwreck, smuggling crime, pirate hijacking and collision of ships on inland rivers occur frequently, and the safety of ocean traffic is facing severe challenges. Traditional maritime monitoring is conducted by manual monitoring of shore-based related institutions, which is limited by complex and heterogeneous data, information overload, personnel fatigue and inattention, etc., and cannot achieve good monitoring effect.

In the 19th century, the field of statistics began to study outliers in data sets, and then "anomaly detection" was applied and developed in various fields. At present, under the background of rapid development of the big data science, data-centric smart maritime platforms have emerged in the field of maritime supervision, with ships as related carriers, It can realize the comprehensive perception of ship dynamic information and static information, crew information, cargo information, navigation security information and other elements, provide technical support for maritime decision-making, management. Therefore, how to make use of rich maritime data to mine and analyze maritime traffic information, detect abnormal dynamics of ships, and provide decision-making support for maritime safety guarantee for relevant institutions has become a research hotspot.

II. Abnormal Behavior of Ship

The abnormal behavior of a ship can be summarized as: actions that do not conform to the predefined normal behavior model of a ship in the water area [1], for example, the ship changes speed for abnormal reasons, changes course, deviates from the route, appears in the inappropriate area, etc., these behaviors may often be related to maritime accidents, illegal smuggling, pirate hijacking, terrorist activities, etc [2].Roy defined the abnormal behavior of ships as deviating from the normal route, entering the restricted zone, abnormal ship handling, abnormal ship position and abnormal ship speed [3].

It can be seen from the above relevant researches that the definitions of ship abnormal behavior are different according to different research emphases. After sorting out and summarizing, we can classify the ship's abnormal behavior into three categories: abnormal position; Abnormal movement; abnormalities inconsistent with the ship's current situation. The specific classification of exceptions is shown in Figure 1, which covers all the abnormal behaviors of ships under general circumstances.

III. Ship Anomaly Detection Based on Big Data

A. Maritime Big Data

At present, the data involved in ship abnormal behavior detection mainly includes ship tracking information, attributes and navigation situation related data [4]. These two types of data and their subcategories can be summarized in Table 1：

TABLE I. DATA SOURCE

Category	Sub-Category
Ship attributes and tracking data	Self-reporting positioning data
	Observation-based positioning data
	Registries information
Contexual geogrphic data	Natural environmental data
	Human related activities

Acknowlegment: Thanks to the National Natural Science Foundation of China (grant number: 51679024) and Fundamental Research Funds for the Central Universities (grant numbe: 3132020372 & 3132020131) for supporting this paper.

978-1-7281-6579-0/20/$31.00©2020 IEEE

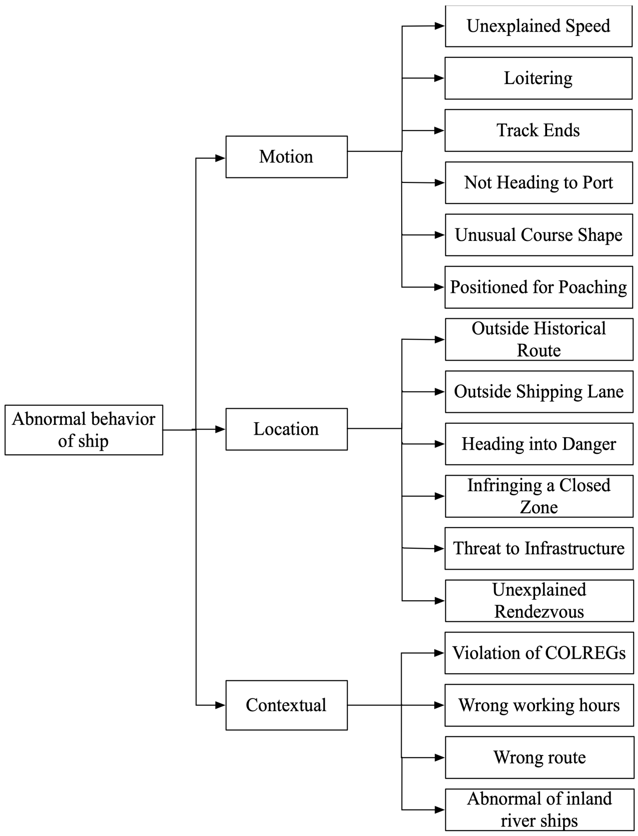

Figure 1. Classification of ship abnormal behavior

To detect ship anomalies, it is necessary to collect ship static data, such as ship name, ship type and size, route, etc., and the dynamic information of the ship, such as ship position, ship speed, course, trajectory, etc. These data belong to ship attributes and tracking information. According to its acquisition methods, it can be divided into three categories: self-reporting positioning data, mainly including the ship's own active transmission for ship collision avoidance (Automatic identification System, AIS data), reporting to regulatory authorities and other types. The self-report positioning data mainly includes the ship's position, speed and other motion information, as well as the static information related to the ship and voyage; observation-based positioning data refers to positioning information collected by various sensors of the ship; ship registration information, including ship construction, management, safety inspection and other information, which is basically inherent and can supplement ship tracking information.

In some cases, the behavior of the ship is obviously abnormal, but it can be reasonably explained in the context of current time and space. Therefore, the situational environmental information of the ship's navigation must be considered: Natural environmental data, including meteorological information and Marine environmental information; Information related to human activities, including port area, traffic management system, risk limitation area and other special sea areas set artificially, specific work information of ships, avoidance behavior of ships manipulated artificially, etc.

B. *Detection Process*

In the study of anomaly detection, a clear normal model should be defined based on the data set. The normal model is generated by predictable repeated events, while the outlier refers to data objects inconsistent with the normal model [5]. Therefore, in the process of anomaly detection, the normal model must be obtained first and used as a standard to detect the abnormal.

The detection of ship abnormal behavior based on data drive consists of two stages: mining maritime big data, learning to build a model representing normal ship behavior, which is generally manifested as statistical distribution or behavior feature class cluster; by comparing the deviation between the normal behavior model and the ship movement data to be detected, the anomaly is detected. If the deviation exceeds a certain threshold, it will be recognized as an anomaly. Data-driven anomaly detection is suitable for large-scale tableless data sets. Outliers are identified by unsupervised methods that do not rely on external information. The anomaly detection process can be represented as Figure 2.

IV. METHODS

From the perspective of technology, the common methods of using big data to detect ship anomalies can be divided into statistically based methods, machine-learning-based methods and neural network-based methods.

A. *Statisitcal Analysis*

Based on the statistical method for anomaly detection is to use the historical data of the ship movement to fit a statistical model to represent the normal behavior of the ship. Using statistical inference determine whether the new observation of ship behavior characteristics belong to the normal model, if the probability of the observed values in the model output is very low, then the observation is considered abnormal.

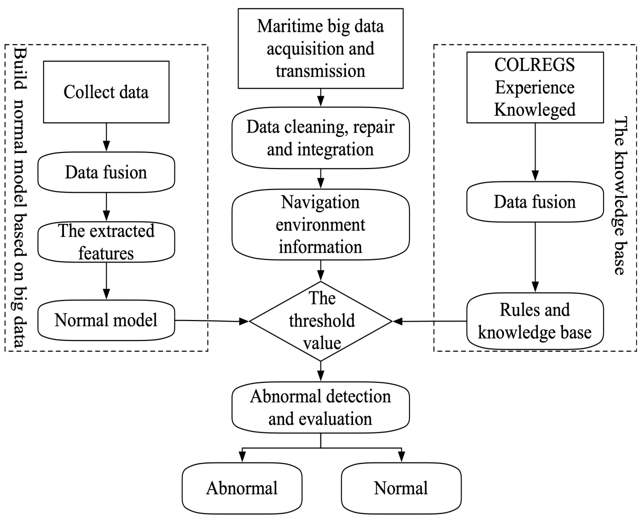

Figure 2. Abnormal detection process

Holst [6] proposed a ship trajectory model based on grid and velocity vector, used two-dimensional Gaussian distribution to establish probability distribution model for normal ship trajectory. Through this model, the probability of input trajectory points was calculated, and the trajectory points whose probability was less than a certain threshold were identified as abnormal trajectory points. The method of Holst was improved by Laxhammar [2] combined with Expectation Maximization and the Gaussian Mixture Model. The algorithm can effectively identify some significant abnormal behaviors, but a large number of training samples are needed to determine the Model parameters. If the data do not meet the Gaussian distribution, the result of abnormal detection is not ideal. Zhen Rong [7] used AIS data and fitted ship trajectory data with the least square method to obtain the ship's normal trajectory model. Then, he calculated whether the distance between the ship trajectory point to be detected and the normal trajectory was greater than the 95% confidence interval of the normal trajectory to determine whether there were abnormalities. The abnormal detection results of this method are shown in Figure 3. Rong [8] proposed a data-driven non-parametric Bayesian probabilistic trajectory prediction model based on Gaussian process, which described the uncertainty of ship position through continuous probability distribution. On this basis, new AIS data were used to update the probabilistic model iteratively to predict ship track in real time.

Statistical methods have a solid theoretical foundation, and anomaly detection using statistics is relatively easy to be realized. When historical data are sufficient and the detection type knowledge adopted is perfect, anomaly detection results are accurate and effective. The disadvantage is that the statistically based exception detection is limited to dealing with specific problems. When the exceptions are uniformly dispersed in the sample, the statistical method is ineffective. In addition, because it is difficult to define the threshold between abnormal and non-abnormal, the statistical method has the possibility of detection error in some cases.

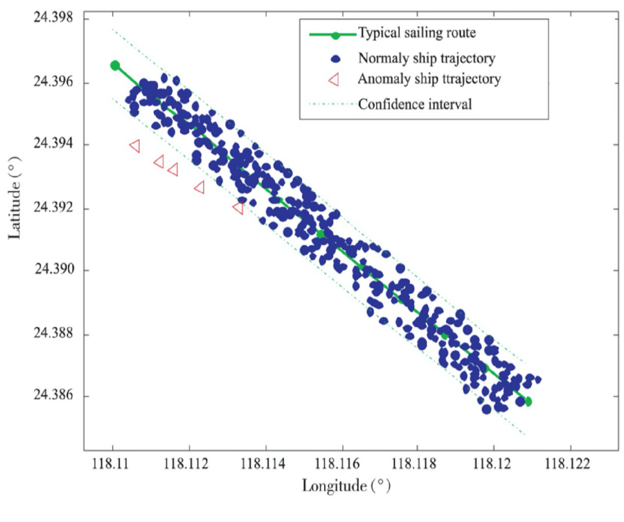

Figure 3. Abnormal detection results based on Statistical Theory

B. Machine Learning

Machine learning has developed rapidly in the field of artificial intelligence in recent years. Its core idea is to train a large amount of data, analyze the hidden rules or structures, continuously improve the performance of the algorithm, and predict or make judgments on events. Abnormal detection based on machine learning means that normal behavior and typical abnormal patterns of ships can be obtained through independent learning from previous maritime data and experience, and the characteristics of normal behavior can be analyzed and extracted, and the learning model can be established through training the most appropriate algorithm, so as to realize the identification of anomalies.

Johansson [9] proposed a ship trajectory anomaly detection method based on bayesian networks, with six ship trajectory point movement information: latitude and longitude, heading, course, speed, ground speed, ship type, as a network node, each node associated probability model is established, through the historical data to complete training, calculation of the input variable trajectory point all nodes joint probability, and according to the continuous input trajectory point average abnormal joint probability for judgment. Lane [10] takes five specific ship trajectory anomalies： AIS signal anomaly, deviation from route, ship inbound and outbound anomaly, embarkation zone and distance anomaly between ships as node variables of Bayesian network to conduct probability assessment of maritime threats. Vespe [11] uses AIS data and adopts unsupervised learning algorithm to automatically learn the ship motion pattern. This algorithm can realize incremental learning, so as to dynamically adjust the model to adapt to environmental changes and realize ship trajectory anomaly detection. Soleimani [12] uses the A * algorithm for the optimal trajectory planning out of the ship, for the ship trajectory length, latitude and longitude gradient and curve area, unsupervised incremental learning method is used to compare planning path and characteristics of ship actual trajectory deviation, obtain trajectory abnormal value, abnormal detection results as shown in Figure 4, -0.75 as outliers threshold to tag of outlier, this method does not need training, widely applicable scene. Lei [13] proposed an anomaly detection that uses machine learning to build normal trajectories model. The space, sequence and behavior of ship motion are taken as the eigenvalues of ship anomaly patterns. According to the normal model established by the algorithm, the ship trajectory is evaluated for the abnormal eigenvalues. Toloue [14] proposed the detection of ship trajectory anomaly based on hidden Markov model, and the velocity, position and course of ship trajectory point were used as parameters to significantly reduce the false alarm rate of anomalies, and the experiment showed that the accuracy of anomaly detection was more than 96%.

At present, the main techniques of detecting ship abnormal behavior using machine learning include decision tree, nearest neighbor algorithm, naive Bayes, support vector machine, etc. Although machine learning can improve the adaptability of anomaly detection, in practical application, machine learning is faced with challenges such as insufficient high-dimensional feature space and sample size, model overfitting, local optimization and poor interpretability.

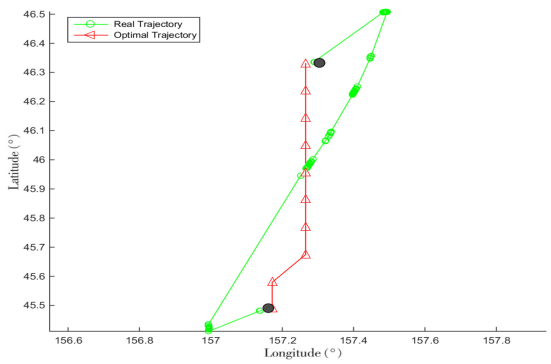

Figure 4. Abnormal detection results based on Machine Learning

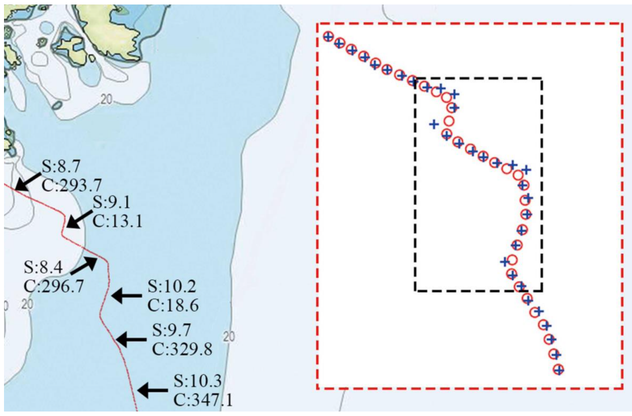

Figure 5. Abnormal detection results based on Neural Network

C. The Neural Network

Neural network simulates the neuron network of human brain to build abstract model, and then forms different networks through different connections. Neural network learning rules are very simple, with strong autonomous learning ability, can map extremely complex nonlinear relations, and has strong robustness and memory functions. At present, there are many researches on using neural network to detect ship abnormal.

Bomberger [15] proposed an unsupervised incremental learning algorithm based on correlation neural network to predict the normal trajectory of the ship at the future moment. When the actual trajectory of the ship deviates from the predicted trajectory, the ship is considered to have abnormal movement. Rhodes [16] [17] mainly studied the relationship between ship position and speed. Based on AIS data, he built artificial neural network and fuzzy neural network respectively to learn ship behavior and applied them to ship position prediction and abnormal behavior detection. Zhen Rong [18] proposed a method of ship navigation behavior prediction based on AIS data and BP neural network. Taking the characteristic value of ship navigation behavior at the historical continuous moment as input, the BP neural network was trained and the characteristic value of ship navigation behavior at the next moment was output, so as to realize the prediction of ship navigation behavior in the future. According to the characteristics of ship trajectory data, Yang Bochen [19] combined with genetic algorithm improved the trajectory prediction method using BP neural network. It has proposed a Recurrent Neural Nework - Long Short-Term Memory (RNN-LSTM) model for predicting ship trajectory, which is applied to ship monitoring, anomaly detection, route planning and other aspects of the target sea area. Zhao [20] proposed a method based on Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and recurrent neural network for ship trajectory anomaly detection, clustering results as the training sample set to train a recursive neural network composed of LSTM unit, the neural network as the ship trajectory predictor, abnormal for the ship heading, speed, an abnormal trajectory real-time detection. The detection results are shown in Figure 5.

The biggest problem of the abnormal detection method based on neural network is that the training process is too closed and the reasoning process and basis cannot be explained reasonably. Moreover, compared with traditional machine learning algorithms, neural network training usually requires more data and is more expensive to operate.

V. Problems and Development Trends

In recent years, with the rapid development of big data science, the maritime supervision system and maritime database are increasingly improved, providing a rich data foundation, and the research on ship anomaly detection has made great progress. However, due to the multi-dimensional complexity of maritime data and the changeable sailing environment of ships, the study of ship anomaly detection still faces many challenges. These unresolved problems are the future development direction of the study of ship anomaly detection.

As maritime data becomes easier to access, data sets are getting bigger. At present, most abnormal detection methods are based on a small number of data sets. The large amount of data directly affects the efficiency of online learning and real-time abnormal detection of the normal model. Existing processing methods are difficult to extract ship-related features from heterogeneous and high-dimensional maritime data. The original maritime data has some quality problems, such as noise data, data missing and so on, after being collected, transmitted and stored. In the future research, it is necessary to establish a complete maritime data processing framework around the above problems. While ensuring data quality, it is also necessary to clean, repair and compress data, and introduce a data processing method with higher performance, which can process a large number of heterogeneous information and data from different sensors and fuse them together.

Ship behavior has a great correlation with the current spatiotemporal background of ships. Factors such as meteorological environment, sea state of navigation area, channel structure, steering and collision avoidance will directly affect ship behavior, and abnormal detection without context has a great chance of false alarm. So far, most researches on abnormal detection of ships lack consideration of the environment of ships. In future studies, the ship's spatiotemporal information and maritime navigation rules can be combined to improve the situational awareness ability and improve the accuracy of ship anomaly detection.

At present, in relevant studies on ship anomaly detection, the normal model is mostly based on historical data. The parameter scalability in the model is poor and cannot be adaptive, and the new data cannot be updated and fused effectively. Once the ship's sailing environment changes, it is likely that the anomaly detection cannot be carried out. Most of the application level is still in the offline detection, but the real-time online detection of ship abnormal behavior is more research and practical value. In the face of the challenge of real-time ship abnormal behavior detection, the adaptive ability of normal model should be improved and the ability of online abnormal behavior detection should be improved by using unsupervised learning technology.

In the application of the ship anomaly detection system, the process of anomaly detection is too closed, and users often cannot get a reasonable explanation for the occurrence of anomalies. And in the process of using, all kinds of parameter initialization, threshold setting and so on do not take into account the user's different level of knowledge level. In future studies, the anomaly detection results can be analyzed in combination with maritime knowledge and presented visually, so that the anomaly detection process can be more intuitively understood by users and provide users with good interactive experience.

VI. Conclusion

The study of ship anomaly detection identifies the anomalies of ship behavior from the massive maritime data, providing support for ensuring marine traffic safety and intelligent supervision of maritime traffic. In this paper, the definition of ship anomaly behavior, the data source for anomaly detection and the process of ship anomaly detection are summarized. It Reviews the research progress of ship anomaly detection by statistical analysis, machine learning and neural network, and finally discusses the challenges still faced by ship anomaly detection, and puts forward the development trend of ship anomaly detection research in the future.

References

[1] A. Sidibé, G. Shu, "Study of automatic anomalous behaviour detection techniques for maritime vessels," The journal of Navigation, 2017, pp.847-858,in press.

[2] R. Laxhammar, "Anomaly detection for sea surveillance," IEEE 11th international conference on information fusion., Cologne, Germany, 2008, pp. 1-8, in press.

[3] J. Roy, " Anomaly detection in the maritime domain," Proceedings of Spie the International Society for Optical Engineering, 2008, 6945, pp. 69450W-69450W-14, in press.

[4] M. Riveiro, G. Pallotta, M. Vespe, "Maritime anomaly detection: A review," Data Mining and Knowledge Discovery. 2018, vol. 8(5). e1266, in press.

[5] V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey," Acm Computing Surveys, 2009, vol. 41(3):75-79, in press.

[6] A. Holst, J. Ekman, "Anomaly detection in vessel motion," Internal Report Saab Systems, Järfälla, Sweden, 2003, in press.

[7] R. Zhen, Z. Shao, J. Pan, "A Study on the Identification of Abnormal Ship Trajectory Based on Statistic Theories," Journal of JIMEI University : Natural Science, 2015, vol. 20(3) : 193-197, in press.
甄荣，邵哲平，潘家财，基于统计学理论的船舶轨迹异常识别[J]. 集美大学学报：自然科学版， 2015，20(3) : 193-197.

[8] H. Rong, A. Teixeira, C. Soares, "Ship trajectory uncertainty prediction based on a Gaussian Process model," Ocean Engineering, 2019, vol. 182, pp. 499-511, in press.

[9] F. Johansson, G. Falkman, "Detection of vessel anomalies-a bayesian network approach," Proceeding of the 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information. Melbourne, Australia, 2007 , pp. 395-400, in press.

[10] R. O. Lane, D. A. Nevell, S. D. Hayward, T. W. Beaney, "Maritime anomaly detection and threat assessment," Proceeding of 2010 13th International Conference on Information Fusion. Edinburgh. IEEE, UK, 2010, pp. 1-8, in press.

[11] M. Vespe, I. Visentini, K. Bryan, P. Braca, " Unsupervised learning of maritime traffic patterns for anomaly detection," 9th IET Data Fusion & Target Tracking Conference: Algorithms & Applications. London, UK, 2012, pp. 14-14, in press.

[12] B. H. Soleimani, E. N. D. Souza, C. Hilliard, S. Matwin, "Anomaly detection in maritime data based on geometrical analysis of trajectories," International Conference on Information Fusion. IEEE, Washington, DC, 2015, pp. 1100-1105, in press.

[13] P. R. Lei, "A framework for anomaly detection in maritime trajectory behavior," Knowledge and Information Systems, 2016, 47(1): 189-214, in press

[14] K. F. Toloue, M. V. Jahan, "Anomalous behavior detection of marine vessels based on Hidden Markov Model," 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Kerman, 2018, pp. 10-12, in press.

[15] N. A. Bomberger, B. J. Rhodes, M. Seibert, A.Waxman, "Associative learning of vessel motion patterns for maritime situation awareness," 2006 9th International Conference on Information Fusion. IEEE, Washington, DC, 2007, pp. 1-8, in press.

[16] B. J. Rhodes, N. A. Bomberger, M. Zandipour, "Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness," International Conference on Information Fusion. IEEE, Washington, DC, 2007, pp. 1-8, in press.

[17] B. J. Rhodes, N. A. Bomberger, M. Seibert, A. Waxman, "Maritime situation monitoring and awareness using learning mechanisms," Military Communications Conference. IEEE, Atlantic City, NJ, 2005, pp. 646-652, in press.

[18] R. Zhen, Y. Jin, Q. Hu, "Vessel behavior prediction based on AIS data and BP neural network," Navigation of China, 2017, vol. 40(02), pp. 6-10, in press.
甄荣，金永兴，胡勤友，等.基于 AIS 信息和 BP 神经网络的船舶航行行为预测[J].中国航海，2017，40(02):6-10.

[19] B. Yang, "Research and application of the trajectory analysis based on AIS,"School of Information and Software Engineering, 2018, in press.
杨博辰. 基于 AIS 的船舶轨迹分析的研究与应用[D].电子科技大学，2018.

[20] L. Zhao, G. Shi, "Maritime anomaly detection using density-based clustering and recurrent neural network," Journal of Navigation. 2019, vol. 72(4), pp. 894-916, in press.