Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230907

Research on the Application of Marine Multi-Modal Big Data in Maritime Law Enforcement

Jinchao LIU[a1] and Li ZHU[b2] and Zhihong DU[c,3] and Chao YU[d4]
[a] *Key Lab of Anti-Terrorism Command Information Engineering, China Coast Guard*
[b] *Key Lab of Anti-Terrorism Command Information Engineering*
[c]*Officers College of PAP*
[d]*Engineering University of PAP*

Abstract. This article first elaborates on the characteristics and classification of big data for maritime law enforcement, and analyzes the current application status of big data in the field of maritime law enforcement. Subsequently, based on the research situation of big data application in the marine field both domestically and internationally, discussions were conducted on the direction of big data application towards maritime law enforcement. Finally, in response to the shortcomings in the application of big data for maritime law enforcement, and drawing on existing research experience at home and abroad, a method based on the Apriori algorithm for data mining of maritime illegal activities is proposed.

Keywords. Ocean; big data; Maritime law enforcement

1. Introduction

China has a vast ocean area, a long coastline, and numerous islands. In recent years, the proportion of the marine economy in the national economy has been increasing. The 20th National Congress put forward the special goal of "developing the marine economy, protecting the marine ecological environment, and accelerating the construction of a maritime power." However, with the acceleration of the international situation, the pressure to safeguard national maritime rights and interests is increasing. The number of foreign fishing disputes is increasing, and the problem of destroying the marine ecological environment is becoming increasingly prominent. Illegal activities such as marine smuggling and illegal immigration have also increased year by year, bringing multiple impacts and challenges to the comprehensive management of the ocean. Currently, the handling of various illegal activities by the maritime police mainly relies on ships and aircraft for on-site verification and disposal. Due to the objective environment of the ocean and the configuration of police forces, enforcement capabilities are very uneven in terms of space and time distribution, and maritime law enforcement

[1] Jinchao LIU, Key Lab of Anti-Terrorism Command Information Engineering; e-mail: hjeljc@163.com
[2] Corresponding author: Li ZHU, Key Lab of Anti-Terrorism Command Information Engineering; e-mail: zhuli_cnrs@foxmail.com
[3] Zhihong DU, Officers College of PAP; e-mail: dzhoec@163.com
[4] Chao YU, Engineering University of PAP; e-mail: 1838923947@qq.com

activities cannot achieve effective coverage at all times and in all areas, resulting in poor law enforcement effectiveness. With the continuous rise of technologies such as "big data, cloud computing, and artificial intelligence," and with the aim of strengthening maritime safety early warning and prevention, real-time situational data generated by key sea areas, islands, various ships, aircraft, and offshore operation platforms can be collected. By integrating large data sets of marine ecology, meteorology and hydrology, security management, maritime security, and fishing production, collecting and organizing information related to sea-related politics, economics, security, and other fields from the internet and media, based on the multi-modal data of the ocean, correlation integration, cluster analysis, and in-depth mining can be carried out to provide scientific basis for assisting in the prediction of marine risks, event warnings, formulation of disposal plans, law enforcement effectiveness evaluation, and other businesses, greatly improving the "scientific and rational management of the ocean, and the use of the ocean in accordance with the law and regulations" for a new pattern of comprehensive management of the ocean.

2. Characteristics, Classification and Current Application Status of Maritime Law Enforcement Data

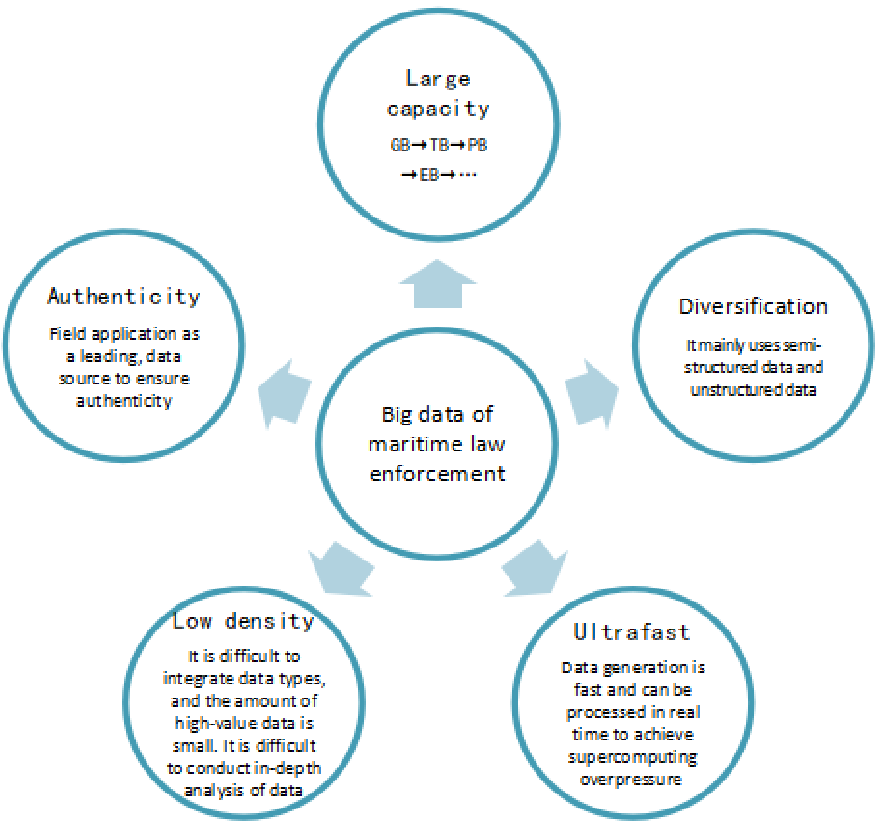

Figure 1. Big data characteristics of maritime law enforcement.

In view of the particularity of the Marine environment, the ability of human beings to acquire knowledge from the sea is very limited compared with the land due to the limitation of communication and other factors. However, with the rapid development of new technologies such as satellite communication and multi-mode sensing, big data in the field of maritime law enforcement is also accumulating exponentially under the lead of demand. It is a collection of videos, picture, voice, sensor data in various application scenarios, infrared, electromagnetic spectrum and other information, which is mainly obtained through observation, monitoring, investigation, statistics and other channels. Most of the data are unstructured or semi-structured, showing the characteristics of a wide range of data sources, diverse modes, complex structure, low density of data value, and obvious spatial and temporal distribution specificity. As shown in Fig. 1.

According to the Coast Guard Law of the People's Republic of China, coast guard agencies at all levels carry out maritime security and crack down on smuggling and smuggling in the areas under the jurisdiction of the People's Republic of China. To supervise and inspect, within the scope of their functions and duties, the exploitation and utilization of Marine resources, Marine ecological environment protection, and Marine fishery production and operations, and to prevent, stop and punish illegal and criminal activities at sea. According to the needs of maritime tasks, the big data of maritime law enforcement can be divided into three categories: basic data, dynamic data and decision data.

2.1. Basic Data

1.Basic data of management objects, such as ship name and number of offshore operating vessels, port of registry, crew information, captain, ship performance parameters, etc. 2. Basic data of our ships. 3. Marine geographic mapping data, such as digital vector sea chart, orthophoto image, digital elevation model, island topography, etc. 4. Meteorological and environmental data, such as Marine meteorological and hydrological historical background data of China's sea areas, ports, etc. 5. Social and cultural data, such as specific fishery resources fisheries, no-fishing areas, Marine wildlife reserves, dangerous navigation areas, uninhabited islands, offshore oil and gas platforms, common port anchorages, etc.6. Literature data, such as laws and regulations in the field of Marine law enforcement, policy documents, historical law enforcement case information, etc.

2.2. Dynamic Data

It refers to the real-time acquisition and perceived monitoring data that can reflect the changes of the electromagnetic environment, meteorology and hydrology, network environment, nuclear biochemistry, social and cultural environment, as well as the changes in the operation status of transportation facilities and important economic facilities. 1. Manage the situation positioning data of objects, such as the real-time trajectory information of ships and aircraft obtained based on AIS, ADS-B, Beidou, radar and other equipment. 2. Real-time monitoring data, such as visible light, infrared, remote sensing images and SAR imaging video image data obtained based on ships, aircraft, satellites and maritime observation points. 3. Marine meteorological observation data, such as Marine meteorological (wind field, temperature, humidity and pressure, etc.), physical ocean (temperature, salinity, ocean current, etc.), Marine biochemistry and other environmental information collected in real time from ship base, fixed point and mobile observation points. 4. Remote sensing observation data, large-scale dynamic monitoring information such as Marine water color, dynamic environment, key target events (sea ice, illegal occupation of islands, sea oil spill pollution, etc.) obtained by satellites and aircraft. 5. Electromagnetic environment data, such as real-time observation data such as electromagnetic spectrum changes in sea areas.

2.3. Decision Data

Decision data mainly refers to the comprehensive analysis and processing of basic data and dynamic data around the needs of law enforcement tasks, and generates high-value data supporting sea situation processing, maritime support, police deployment, maritime law enforcement and other activities. 1. Task prediction and analysis data based on changes in Marine meteorology, hydrology and electromagnetic environment. 2. Early warning and analysis data based on the maritime safety management situation. 3. Resource allocation and analysis data based on maritime mission disposal. 4. Analysis data based on the effects of maritime enforcement actions. As shown in Table 1.

Table 1. Classification of ocean multi-modal big data

Data classification	Main data format	Data sources
Basic data	● text file ● picture	Combine text manual input and machine generation.
Dynamic data	● Positioning information ● video ● speech sounds ● environmental data	Through satellite, radar, AIS, video surveillance, communication equipment, environmental sensors and other situational awareness systems acquisition.
Decision data	● Text file ● Command order	Mining and generating basic data and dynamic data.

Although coast guard agencies at all levels have accumulated a large amount of original data in their daily work, they are only in the initial stage of Marine big data utilization. Compared with the application research of foreign maritime law enforcement agencies in the field of big data, there are still obvious deficiencies. Firstly, the data quality and quantity are not complete enough. Affected by the objective environment, compared with the land, the construction of maritime target detection equipment system is not perfect, and it is difficult to guarantee the full-dimensional perception requirements for maritime law enforcement. Moreover, there is a lack of large-scale law enforcement and remote detection means, such as Marine monitoring satellites, hollow long-endurance drones, unmanned ships, carrier-borne over-of-sight detection equipment, and effective communication means that can meet the needs of high bandwidth and low delay transmission. Secondly, the lack of data mining and analysis ability. At present, the use of Marine big data is mainly reflected in data summary management, statistical analysis and visual display, but it has not been effectively applied in cross-scale fusion analysis, multi-field collaborative processing, and task regularity mining for Marine multimodal big data. Thirdly, the application concept needs to be updated. There are many management departments of Marine big data, and the cross-network, cross-field and cross-department data sharing mechanism is not sound. Moreover, limited by the security and confidentiality mechanism and business management system, a large amount of data cannot be effectively integrated, and the value and benefits cannot be fully played.

With the gradual attention of countries to maritime rights and interests and the rapid development of Marine economy, the pressure of comprehensive maritime management is increasing. Key breakthroughs need to be made in the application of Marine big data according to the characteristics of Marine law enforcement. To provide full-dimensional, three-dimensional and continuous information guarantee services for realizing accurate and efficient maritime law enforcement.

3. Reflections on the Application of Marine Data Research to Maritime Law Enforcement at Home and Abroad

In recent years, many scholars at home and abroad have conducted a lot of research on the ocean multi-modal big data mining and analysis, and carried out small-scale pilot applications in the fields of Marine scientific research and maritime management. Relevant research has achieved certain results, which provides a reference for improving maritime law enforcement efficiency by using big data.

3.1. Governance of Maritime Security Order

The governance of maritime security order has always been a challenging issue in the world. Due to the lack of intelligence clues and on-site evidence, law enforcement in ocean often faces many difficulties. In 2022, Chinese maritime police investigated and dealt with 471 cases of suspected smuggling and over 2000 cases of various criminal offenses related to public security. The trend of illegal activities is clearly on the rise. In the field of marine big data mining, Song et al. proposed a comprehensive model based on Analytic Network Process and Superiority Weakness Opportunity Threats, which can combine environmental changes and policy adjustments related to the ocean domain. By setting reasonable weights and strategy rankings, High-value marine-related information is mined and generated [1]. Yue et al. proposed an improved FP-growth algorithm for data mining of ships in ocean, using the method based on spatial-temporal segmentation to divide the target area, and by mining frequent item sets, the information about the grouping of ships in ocean can be obtained [2]. Wei designed an abnormal target recognition system based on historical data mining, which integrates multiple monitoring methods and knowledge-driven methods. By extracting multi-dimensional feature information of the target, the target information map was constructed, and the target historical knowledge base was established. This system can detect abnormal behaviors of the target by combining semantic rules with the target's historical behavior background based on the fused target information [3]. However, the above researches mainly focus on the application of maritime traffic and marine environment monitoring, and there are few researches on the application of maritime security order governance. If we can integrate network extraction, on-the-spot collection and other methods, based on the big data association analysis algorithm, we can comprehensively and timely mine the maritime intelligence information. By using big data visualization technology, various hot events and illegal activities occurring in China's sea areas can be classified and statistically analyzed. These results can then be presented in a GIS platform for intuitive display, which is conducive to maritime police’s comprehensive grasp of the overall situation of maritime security. Based on the anomaly detection and correlation algorithm in the mining technology, the general trend and abnormal behavior can be identified in the marine big data, and the distribution range and behavioral characteristics of historical events can be deeply analyzed. This can help prevent mass maritime emergencies, early warning of potential illegal activities, and identify high-risk areas, which assists law enforcement personnel in making judgments and disposing of cases.

3.2. Guarantee of Maritime Navigation Safety

Maritime navigation is greatly influenced by oceanic climate and lacks sufficient traffic signs compared to land, making maritime navigation riskier. With the application of various navigation and positioning devices such as AIS, GPS, Beidou, as well as meteorological warning devices, safe maritime navigation has become possible. However, with the rise of the ocean economy, maritime traffic has become increasingly busy, and traffic accidents occur frequently due to catastrophic weather conditions such as typhoons and non-compliant operations. As in the "2020.8.30" traffic accident case investigated by the maritime police, a Liberian freighter "SBI PERSEUS" collided with a Chinese fishing vessel "MinJinYu 05119" in the Taiwan Strait, causing the sinking of "MinJinYu 05119" and the disappearance of 12 crew members. In the "2022.6.12" maritime traffic accident case, a speedboat collided with the "YueLuYu 13253" fishing vessel in a certain sea area of Shanwei City, Guangdong Province, causing the fishing vessel to capsize, resulting in two deaths and one injury. All these accidents caused huge loss of people and property. Many scholars have conducted extensive research on using big data for navigation safety warnings. Li et al. used maritime VTS data as the basis, studied the main characteristics of maritime traffic, and used ST-DBSCAN and Apriori algorithms to mine the spatial-temporal patterns of maritime traffic, such as co-occurrence, multi-ship encounter, and spatial-temporal correlation. They analyzed the macro features, spatial-temporal patterns, and complexity of maritime traffic flow, which can provide guidance for maritime traffic safety management and risk avoidance [4]. Geng et al. proposed a meteorological routing algorithm for typhoon avoidance based on marine meteorological big data. The algorithm is a hybrid of genetic algorithm and RHC strategy, which can quickly and safely avoid the large wind and wave areas and static obstacle areas affected by typhoons based on real-time meteorological data. At the same time, a dynamic programming model is constructed according to the impact of wind, waves, and currents on ship stalling, which can achieve a balance between safe navigation and the shortest sailing time [5]. Dai et al. developed an early warning system for maritime traffic risk under poor visibility meteorological conditions. They used the short-term ship traffic flow density prediction algorithm based on the extreme learning machine theory in the ANN network. Through the learning and training of big data samples, the continuous prediction accuracy could reach more than 80%, which could realize the visualization and intelligent monitoring of maritime navigation risks under bad weather conditions [6]. By relying on the maritime 110 emergency response platform and conducting feature association research based on maritime traffic and meteorological data, it can help guide emergency avoidance for vessels engaged in marine production operations and enhance the emergency rescue capabilities of maritime law enforcement vessels under severe weather conditions. After a major disaster occurs, tracing and analyzing the original data can assist law enforcement personnel in conducting cause investigations, fixing illegal evidence, etc., which will greatly improve the efficiency of maritime management.

3.3. Management of Ocean Resources and Environment

With the vigorous rise of the ocean economy, human production and living spaces are constantly expanding to the ocean, and the protection of marine ecological environment also faces great challenges. In the daily law enforcement of maritime police, problems such as illegal sand mining, unauthorized use of sea areas, illegal occupation of uninhabited islands, damage to submarine cables, and marine oil spills often occur. For example, in 2022, Shanghai Maritime Police Bureau investigated a major change in the construction scale of the Ab section seawall and site filling project of Xiaoyangshan Reclamation Phase I Project. Jiangsu Maritime Police Bureau investigated the "2022.5.19" illegal sand mining case. The maritime police bureau in Beihai discovered through aerial law enforcement that basic infrastructure was illegally built on the uninhabited island of Monkey Island in Yantai. These actions have caused serious damage to marine resources and the environment, and the ecological balance of the ocean urgently needs to be reconstructed. In terms of using big data to govern the marine environment, many scholars at home and abroad have conducted many beneficial explorations. For example, in the area of marine oil spill monitoring, many teams have developed a lot of synthetic aperture radar oil spill detection and classification algorithms based on traditional machine learning using marine remote sensing data, such as linear models, artificial neural networks, bayesian classifiers, etc.

In terms of coastline monitoring, Wang Houjun et al. proposed an optical imaging map obtained by airborne SAR radar, which can be compared with the area, shape and other spatial geographic information obtained at different periods of reclamation. This method can complete the extraction of the coastline and automatically output the analysis results of reclaimed land and geological features by detecting changes in spatio-temporal distribution characteristics [7]. In terms of marine resource law enforcement, due to limited coverage of maritime law enforcement forces, it is difficult to ensure full-time and all-round law enforcement in the jurisdictional waters. Based on Marine multi-modal big data, similar detection algorithms are used to carry out intelligent real-time monitoring and analysis of large-scale offshore oil spills, illegal sea reclamation, illegal occupation of uninhabited islands, illegal dumping and other acts of damaging marine resources and environment. This helps to monitor and remotely identify various illegal activities without human intervention, allowing for timely discovery and disposal, thereby further improving the efficiency of marine governance. In addition, continuous observation and evaluation of relevant data can also predict the effectiveness of control measures, effectively reducing environmental pollution hazards.

3.4. Integrated Management of Fisheries

At present, Marine fishery has become an important pillar of Marine economic development. In order to accelerate the protection of Marine fishery resources and ecological restoration, China has implemented the policy of summer fishing ban in most of the Bohai Sea, Yellow Sea, East China Sea and South China Sea since 1995. However, in recent years, problems such as overfishing, illegal fishing and ecological destruction of resources have emerged one after another, seriously restricting the development of

Marine fishery. It also poses a threat to the safety of fishermen and the order management of maritime production and operation. For example, in 2022, the Coast Guard cracked 214 cases of illegal fishing of aquatic products, investigated and dealt with 525 fishery administrative cases, and broke up several illegal fishing gangs. Domestic and foreign scholars have carried out some research on the use of big data to assist the identification of fishing boats. For example, Geng Jiali proposed an improved MSC-FBI algorithm. Based on the clustering of VMS fishing boat tracks, feature extraction is carried out on the track features corresponding to fishing boats in different fishing behaviors, and then a fishing boat behavior recognition model is established according to each type of track features. The algorithm has good robustness, but only classifies tracks into anchorage, fishing and sailing. It cannot cover all operational behaviors [8]. Kai Sheng et al. established an analysis model for classification of fishing boats and cargo ships by using a logistic regression algorithm based on AIS ship track features and multidimensional features such as sailing speed and turning action [9]. Erico N. de Souza et al., based on historical AIS data, adopted three different machine learning, data mining and filtering methods to detect the potential fishing activities of trawlers, Seine fishing vessels, etc., so as to realize the large-space and high-precision identification of the operation conditions of fishing vessels at sea [10]. In terms of Marine fishery law enforcement, risk prediction of illegal activities in Marine fishery based on Marine big data will be of great significance to the maintenance of Marine fishery resources. For example, by analyzing the characteristics of illegal fishing and other behaviors, focusing on the classification of data such as prohibited fishing areas, prohibited fishing periods, prohibited fishing tools and prohibited fishing vessels, it can assist in the investigation and collection of evidence of behaviors damaging fishery resources. By tracking and analyzing the long-term data of key operating areas and fishing vessels, it can help the law enforcement officers of the coast guard to establish thermal maps of maritime control, predict potential risk areas, and provide a basis for formulating fishery protection policies reasonably and adjusting law enforcement areas scientifically [11].

4. A mining and Analysis Method for Illegal Acts at Sea Based on Apriori Algorithm

The above analysis results show that the volume of Marine big data is large, which contains high value data to be mined; There are many researches on offshore data algorithms, but few practical applications. There is a lot of demand for maritime law enforcement applications, but key technologies need to be broken through.

The data mining technology based on association rules is mainly applied to mining the internal correlation in big data. Since it was first proposed in 1993, there have emerged some classical algorithms such as Apriori and FP-Growth, which aim to find out some hidden rules and objective knowledge by finding out all frequent item sets. Based on the statistics of maritime law enforcement cases over the years, it can be seen that they are characterized by rich data types and complex structural levels, while the traditional single-dimensional association rule algorithm cannot meet the needs of data mining. In view of the characteristics of maritime law enforcement, adopting multidimensional association rule algorithm based on Apriori for data mining analysis will help to explore the correlation and trend between illegal behaviors. To prevent and stop maritime crimes to provide auxiliary decision analysis.

4.1. Introduction to Association Rules

Basic concepts of association rules [12]

I={I1, I2,... , Ii} is a set of terms, D is the set of all transactions, transaction T is a subset of I, there are TÍ I.

A is a set of items, when and only when AÍ T, transaction T contains A.

Association rules is a similar expression, such as A Þ B. In the expression, AÌ T, BÌ I. And A∩B=Φ.

Rule A B has support s in transaction set D, where s is the percentage of D transaction containing A∪ B, expressed by probability as P(A∪ B)=s.

Rule A B has a confidence c in the transaction set D, where c is the percentage of transactions in D containing both A and B transactions, equal to the conditional probability P (B / A). It is expressed as:

$$\text{support}(A\ B)= P(A \cup B) \tag{1}$$

$$\text{confidence}(A\ B)=P(B/A) \tag{2}$$

If a rule meets the minimum support (min-sup) and minimum confidence threshold (min-conf) at the same time, this rule is frequent and is called a strong rule, in which the threshold can be set by the user.

Multidimensional association rules:

If the rule involves two or more dimensions, such as age (m, "35... 39 "), income (m, " 42K ... 48K ") , derive the commodity (m, " electric vehicle ") , then called multidimensional association rules.

4.2. Classic Apriori algorithm

- Step 1, input N-dimensional data set, minimum support (min-sup) threshold and minimum confidence (min-conf) threshold.
- Step 2, the N-dimensional data set is scanned to calculate the support degree of a single item (k=1). The set of items meeting the minimum support threshold is defined as frequent 1-item set, denoted as L1.
- Step 3, continue to scan the n-dimensional data set, frequent 1-item sets in L1 are connected, and candidate 2-item set C2 is generated. The set of candidate 2-item set C2 satisfying min-sup is defined as frequent 2-item set, denoted as L2.
- Step 4, continue to scan the n-dimensional data set, iterate and connect layer by layer according to the above rules, and obtain the set of frequent k item sets according to Lk-1 item, which is denoted as Lk. Due to the continuous elimination of unqualified candidate item sets, frequent item sets cannot be obtained eventually. When the resulting frequent item set Lk+1 is an empty set, the frequent k- item set L is output until. Otherwise, repeat this step.
- Step 5, all the strong association rules R can be output by sifting out the frequent item sets satisfying min-sup and meeting min_conf at the same time. At this time, the algorithm ends. As shown in Fig. 2

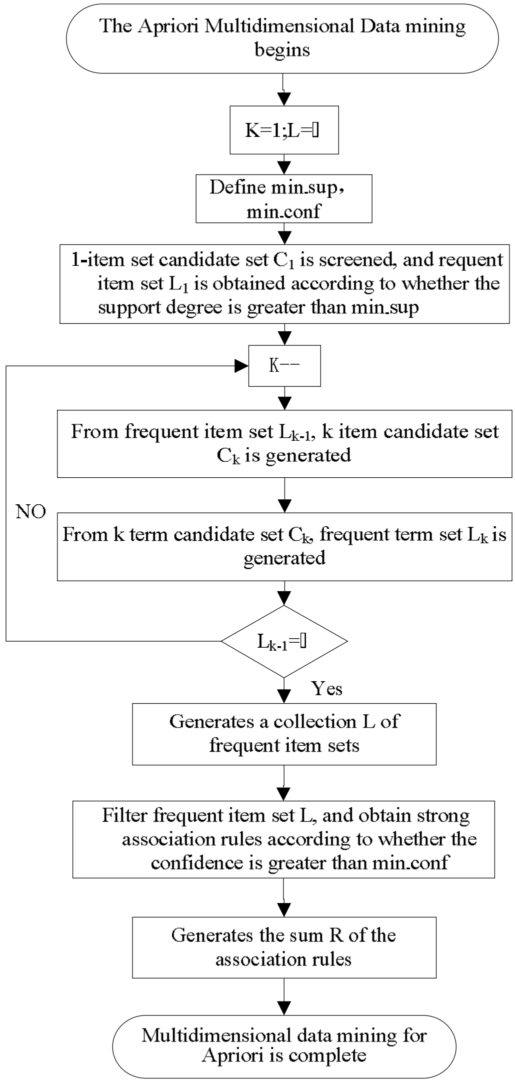

Figure 2. Multi-dimensional association rule algorithm flow for Apriori

4.3. Maritime Law Enforcement Data Mining Application Based on Association Rules

Based on the information of maritime law enforcement cases in recent years, the characteristics and rules of typical illegal behaviors at sea are analyzed by data mining based on Apriori algorithm.

- Data preprocessing

Due to the characteristics of incomplete, inaccurate and weak consistency of the information of maritime law enforcement cases, effective mining information cannot be obtained without data preprocessing. In order to improve the efficiency of mining and analyzing the law of illegal acts at sea, multidimensional data cube should be established first and data should be governed. Through the analysis of key factors affecting the order of maritime security, combined with the nature and characteristics of maritime illegal cases, the attributes of data are divided into nine categories, such as case type, time, incident sea area, ship type, port of registration (by province), temperature, weather, wind power and value involved, etc. 9-dimensional data cube is established, as shown in Fig. 3.

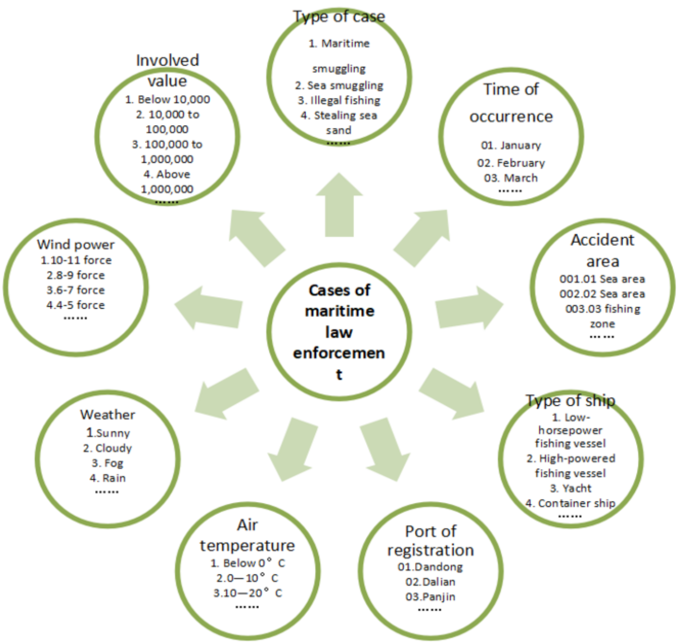

Figure 3. Schematic diagram of the data cube

There are many data types in the case, so it is necessary to define appropriate numeric types to facilitate the processing of various elements. For example, factors such as case type, event sea area, ship type, port of registry, and weather can be set to enumerate data. Converts the attribute values to a set of continuous integers by setting the mapping relationship. The time, wind, air temperature, and involved value can be set as numerical data. In order to facilitate statistical consistency, the data can be generalized according to the discrete characteristics of the data and converted into a set of continuous integers according to certain rules. Part of the records are selected from the law enforcement case database over the years, and the data format after processing is shown in Table 2.

Table 2. Statistical table of maritime illegal cases

ID	Type of case	Time	area	Type of ship	Port of registration	Air temperature	Weather	Wind power	Involved value
1	1(Maritime smuggling)	3 (March)	203(203 fishing zone)	1(Low-horsepower fishing vessel)	07 (Zhejiang)	3(10-20°C)	2 (Cloudy)	6(1-2 force)	3(100000 to 1000000)

2	2(Sea smuggling)	5 (May)	076(76 fishing zone)	2(High-powered fishing vessel)	04 (Shandong)	3(10-20°C)	3 (Fog)	4(4-5 force)	2(10000 to 100000)
3	3(Illegal fishing)	7 (July)	114(114 fishing zone)	2(High-powered fishing vessel)	05 (Jiangsu)	5(Above30°C)	1 (Sunny)	5(2-3 force)	3(100000 to 1000000)
4	5(Commit traffic offence)	11 (November)	092(92 fishing zone)	4(Container ship)	04 (Shandong)	2(0-10°C)	5 (Rain)	4(4-5 force)	4(Above 1000000)
5	4(Stealing sea sand)	2 (February)	100(100 fishing zone)	5(Cargo ship)	04 (Shandong)	2(0-10°C)	3 (Fog)	6(1-2 force)	3(100000 to 1,000,000)
6	3(Illegal fishing)	6 (June)	545(545 fishing zone)	2(High-powered fishing vessel)	11 (Hainan)	5(Above30°C)	2 (Cloudy)	6(1-2 force)	2(10000 to 100,000)
7	1(Maritime smuggling)	2 (February)	255(255 fishing zone)	4(Container ship)	08 (Fujian)	3(10-20°C)	5 (Rain)	5(2-3 force)	3(100000 to 1,000,000)
…	…	…	…	…	…	…	…	…	…

- Data mining

Taking the case of stealing sea sand as an example, there is a strong correlation between it and the time of occurrence, the sea area of occurrence, the type of vessel and the port of registration of vessel. In order to simplify the introduction of the data mining process, 5-dimensional data is screened from 100 records in the historical database for algorithm display only. Due to the small number of samples, the obtained results do not represent the real rules. The data of illegal behaviors in Table 1 are denoted as T. ID indicates the case record. A, B, C, D and E are the impact factors of five dimensions related to illegal behaviors. As shown in Table 3.

Table 3. Records of illegal incidents at sea

Record TID	Impact Factor
T1	A1、B3、C203、D1、E7
T2	A2、B5、C076、D2、E4
T3	A3、B7、C114、D2、E5
T4	A5、B11、C092、D4、E4
T 5	A4、B2、C100、D5、E4
T6	A3、B6、C545、D2、E11
T7	A1、B2、C255、D4、E8
T8	A4、B3、C091、D5、E7
T9	A3、B6、C164、D2、E5
T10	A2、B8、C220、D4、E8
…	…
T100	A4、B12、C091、D5、E7

Assuming minimum support 0.04 and minimum confidence 0.5. The first step calculates the support of the candidate 1-item set, the candidate set less than the minimum support threshold will be pruned, then determine the set of frequent 1-item sets, the calculation process is shown in Fig. 4.

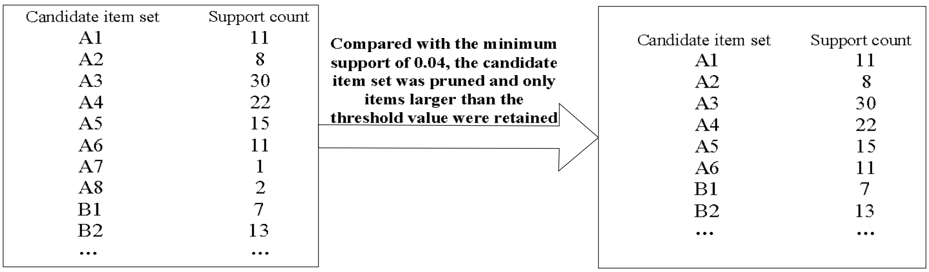

Figure 4. The generation process of frequent 1- item set

In the next step, candidate 2-item set is generated by connecting frequent 1-item set with itself, and then the above calculation process is repeated. Candidate item sets less than the minimum support threshold will be pruned, as shown in Fig. 5.

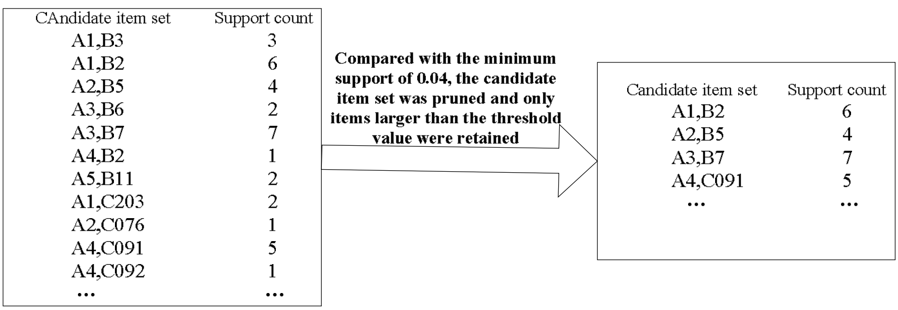

Figure 5. The generation process of frequent 2- item set

Similarly, frequent 3-item set can be generated by frequent 2-item set connecting with themselves and pruning processes. The specific calculation process is similar to the previous one and will not be shown. Through continuous recursive mining, it is finally obtained that the frequent 4-item set of mining is {A3,B7,C114,D2} and {A4,C091,D5,E7}, and the support counts are all greater than or equal to 4. The mining process of frequent item set is over.

- Generate strong rules

Taking the frequent item set{A4,C091,D5,E7}as an example, the non-loophole set of the frequent item set has {A4} , {C091} , {D5} , {E7} , {A4, C091} , {A4, D5} , {A4, E7} , {C091, D5} , {C091, E7} , {D5, E7} , {A4, C091, D5} , {A4, C091, E7} , {A4, D5, E7} and {C091, D5, E7} .First, the confidence conf of the association rule will be calculated, for example in {A4, C091, D5} => {E7},conf= 60% ; and in{C091} => {A4, D5, E7},conf =25% .Then, association rules that do not meet the minimum confidence threshold were removed, min_conf≤50% . Finally, the strong association rule was obtained {A4, C091,D5} => {E7}, and the end of the algorithm

However, it should be noted that the traditional Apriori algorithm defaults to the same degree of importance. In fact, some factors occur less frequently but play a significant role. Therefore, it is necessary to determine the relevant standards and regulations. Then weighted according to expert experience and other methods to ensure that the real situation can be objectively reflected by the mining results.

5. Conclusions

To sum up, Marine multi-modal big data has great potential in obtaining intelligence clues and assisting law enforcement in case handling. However, ocean big data presents obvious various characteristics, such as uneven spatial distribution, large surface data scale, small underwater data scale, large offshore data scale, small offshore data scale, high value data mining difficulty. Therefore, in the application of big data in the field of maritime law enforcement, the following four directions need to be broken through.First, the deployment of Marine multi-modal information sensing and acquisition technologies should be accelerated to improve data scale and quality. Second, relevant statistical analysis based on historical maritime law enforcement data should be carried out in depth to improve the cognitive ability of maritime security control situation. Third, the knowledge mining technology of "artificial intelligence + Marine big data" should be vigorously developed, and the predictive research ability of future maritime illegal trends will be improved through the use of small sample learning and independent learning. Fourth, the deep integration of big data products and maritime law enforcement should be promoted. Data will be introduced into the command and decision-making chain to enhance the capability of coast guard agencies at all levels to manage and use the sea scientifically.

References

[1] Song Shanshan, Wang Jinping, Ji Wanqing. The Research on the Development Strategy of Marine science and technology information field in our country based on the ANP-SWOT model[J]. Science and Technology Management Research, 2020, 40(22): 48-55.

[2] Yue Jiancheng, Wang Yumei, Wu Yafei, et al. Marine group Target Mining Based on Improved FP-growth[J]. Computer and Modernization, 2022, (02): 33-37.

[3] Wei Xuan. Design and Implementation of Abnormal Object Recognition System Based on Historical Data Mining[D]. Beijing University of Posts and Telecommunications, 2019.

[4] Li Yongpan. Study on Main Characteristics of Maritime Traffic Based on AIS Data[D]. Dalian Maritime University, 2019.

[5] Geng Xu. Research on the Method of Ship Fine-track Planning Based on Meteorological Elements[D]. Harbin Engineering University, 2019.

[6] Dai Houxing, Wu Zhaolin. Marine Traffic Safety risk warning system in bad visibility weather [J]. Journal of Traffic and Transportation Engineering, 2018, 18(05): 195-206.

[7] Wang Houjun, Li Ming, Ding Ning, et al. Application and Exploration of airborne MiniSAR of UAV in Reclamation monitoring[J]. Surveying, Mapping and Spatial Geographic Information, 2019, 42 (3): 97-99.

[8] Geng Jiali. Research on Big Data Storage Optimization and Behavior Recognition Technology of Fishing boat track[D]. Hangzhou Dianzi University, 2018.

[9] Sheng K, Liu Z, Zhou D, et al. Research on ship classification based on trajectory features[J]. The Journal of Navigation, 2018, 71(1): 100-116.

[10] De Souza E N, Boerder K, Matwin S, et al. Improving fishing pattern detection from satellite AIS using data mining and machine learning[J]. PloS one, 2016, 11(7): 8248-8261.

[11] Orofino S, McDonald G, Mayorga J, et al. Opportunities and challenges for improving fisheries management through greater transparency in vessel tracking[J]. ICES Journal of Marine Science, 2023, (03): 8-19.

[12] Han Jiawei, Fan Ming, Meng Xiaofeng. Data Mining: Concept and Technology[M]. Beijing: China Machine Press, 2012: 160-165