Illegal Sea Project Discovery Algorithm based on Hash Table and DBSCAN Algorithm

Fengjuan Cui[1], Xiaowei Xu[1], Li Zhang[1], Jie Yin[1*], Meng Tan[1], Feng Chen[1], Yuan Song[2], Yarong Li[2] and Zhen Zhao[2]

{[*]Corresponding author: yinjie@ncs.mnr.gov.cn}
{cfj.08@163.com, xuxiaowei@ncs.mnr.gov.cn, 329022814@qq.com, yinjie@ncs.mnr.gov.cn, 86802378@qq.com, 85531118@qq.com, 490666142@qq.com, 15224314399@163.com, zzqust@126.com}
[1]The North China Sea Data&Information Service, SOA, Qingdao, China

[2]Qingdao University of Science and Technology, Qingdao, China

Abstract: Marine cage culture does not need to be reclaimed, so it has a certain degree of conceal ability, which leads to the situation that some individuals or even enterprises install cages to carry out aquaculture in the open sea without approval. These behaviors not only violate the national legal system of marine natural resources management and cause pollution to the marine environment, but also have an inestimable impact on the safety of maritime shipping and marine tourism. The existing supervision methods, such as manual or simple density statistics of ships in the grid, have low work efficiency, huge labor cost and high misjudgment rate. In view of the above drawbacks, this paper proposes an illegal sea project detection algorithm based on hash table and DBSCAN algorithm. Through DBSCAN clustering analysis of spatial location and time in ais information, the clustering status of ships is separated. Further, hash table is used to calculate and save the relevant latitude and longitude distance data in advance to optimize the calculation time of DBSCAN algorithm. Finally, the spatial scope of the clustered coordinate points is judged, and the operating scope of ships in the open sea is distinguished, so as to narrow the key areas of government ships and drones, and improve the ability of maritime supervision.

Keywords: Illegal Sea Project Discovery, Hash Table, DBSCAN, Sea Area Supervision

1. Introduction

China is a big country of aquaculture and the largest mariculture country in the world. The output of mariculture in 2021 was 22.1114 million tons, up 3.55 percent year on year, and the mariculture area was 2,025.51 thousand hectares, up 1.50 percent year on year. China is the only major fishing country in the world where the total aquaculture volume exceeds the total fishing volume.

Seawater cage is a widely used aquaculture technology at present. As an intensive

ICSETPSD 2023, November 17-18, Coimbatore, India

DOI 10.4108/eai.17-11-2023.2342795

aquaculture method, its application broadens the breeding sea area, expands the breeding capacity, improves the breeding conditions, and is conducive to reducing the pressure of coastal aquaculture and fishery fishing, which is of great significance to the protection of marine ecosystems.

Marine cage culture does not need to be reclaimed, so it has a certain degree of conceal ability, which leads to the situation that some individuals or even enterprises install cages to carry out aquaculture in the open sea without approval. These behaviors not only violate the national legal system of marine natural resources management and cause pollution to the marine environment, but also have an inestimable impact on the safety of maritime shipping and marine tourism. Taking dalian city, liaoning province as an example, the management committee of changxing island has issued notices and notices to the society many times since april 2021, urging illegal cage and floating valve farmers in the south of fengming island,Hulushan bay, lahuo and dongjiakou waters to dismantle illegal facilities by themselves before april 1, 2022 [1].

At present, the methods to identify and detect marine cage farming include various methods such as official ship and UAV cruise, remote sensing image identification and ais identification. The ais identification method is to manually find out the abnormal behavior of the ship in the unapproved waters by analyzing the position of the ais ship, including low-speed sailing, staying and cruising, and then combine the former two methods to confirm. In general, the work efficiency is low, the labor cost is huge, and the misjudgment rate is high. One way to improve efficiency is to draw the electronic fence and analyze the position, trajectory, speed, direction and smoothness of continuous trajectory in the ais information of the ship, which can improve the work efficiency to a limited extent.

2. Research Status

DBSCAN(Density Based Spatial Clustering of Applications with Noise) is a widely studied clustering algorithm in the field of data mining and machine learning. It can identify clusters of arbitrary shapes and deal with noise. And it can find clusters of arbitrary shape in noisy spatial databases, so DBSCAN algorithm has been applied to many fields.

In the field of target location and processing trajectory analysis, Liu Yong improved the DBSCAN algorithm [2] by taking the core point as the origin, dividing the neighboring points of the core point into different quadrants, and then completing the expansion operation of the cluster according to the number of the neighboring points in each quadrant and whether the neighboring points deviated from the core point. On the basis of the classical DBSCAN spatial clustering algorithm, Guo Naikun preprocessed the ship trajectory data, such as cleaning and compression, and divided it into sub-trajectory segments connected by feature points, and then introduced the time distance measurement method to realize the spatio-temporal clustering of ship trajectory [3]. Liu Yu adopted the fusion distance as the distance calculation method of the ship trajectory, and fully considered the heading information and speed information for similarity measurement. The trajectory sub-segments after trajectory segmentation were analyzed by clustering, and the typical motion trajectory of the ship could be obtained through experimental analysis [4]. Dai Shaosheng proposed a local adaptive DBSCAN clustering algorithm based on grid partition idea. According to the characteristics of the data set itself, the grid space is generated and the grid space is mapped, and then the Gaussian kernel function is used to estimate the local density of each grid interval to determine the key parameters of the algorithm, which has high clustering accuracy and low time-consuming [5].

In recent years, the combination of DBSCAN algorithm and other information technologies has developed rapidly, especially the combination of DBSCAN and deep learning technology. The method generally uses deep learning methods to automatically extract data features, and then uses clustering algorithms for clustering, so as to improve the operation efficiency. In the field of combining DBSCAN with deep learning, Zhang Bohan selected DTW similarity measurement algorithm and DBSCAN density clustering algorithm to cluster and extract ship normal trajectory, and improved DTW (Dynamic Time Warping) distance in the similarity measurement of ship trajectory [6]. Zhang Zhiyuan studied and analyzed several commonly used clustering algorithms, such as K-means algorithm and DBSCAN algorithm,and compared the advantages and disadvantages of each algorithm and application scenario s[7]Yu Gang improved the sample trajectory data extraction method by using Geohash coding and DBSCAN spatial clustering to deal with the spatial correlation between spatio-temporal trajectory points [8]Lu Xiao proposed two algorithms, consistent subspace clustering network and relation-guided subspace clustering network, based on subspace consistency assumption and local reservation constraint, and carried out experimental verification, and systematically compared related methods [9].

3. The Disadvantages of The Existing Monitoring Methods such as Manual or Simple Density Statistics of Ships in The Grid

Since Marine cage aquaculture requires the use of various sizes of ships for cage maintenance, these ships are equipped with AIS equipment in order to ensure navigation safety. Through the analysis of the characteristics of AIS trajectory, the centralized occurrence and suspicious navigation behavior of ships in the unapproved sea area can be found, which can effectively reduce the inspection intensity of government ships and drones and other regulatory means. To a certain extent, the supervision ability of the utilization of Marine space resources can be improved, and the order of the utilization of Marine space resources can be strengthened.The existing supervision work mainly divides the unapproved open sea into grids, queries the AIS trajectories of ships in a certain period of time, and makes preliminary judgments through manual visual methods or statistics of the density of ships in the grid, which is time-consuming, inefficient, and difficult to guarantee the accuracy.

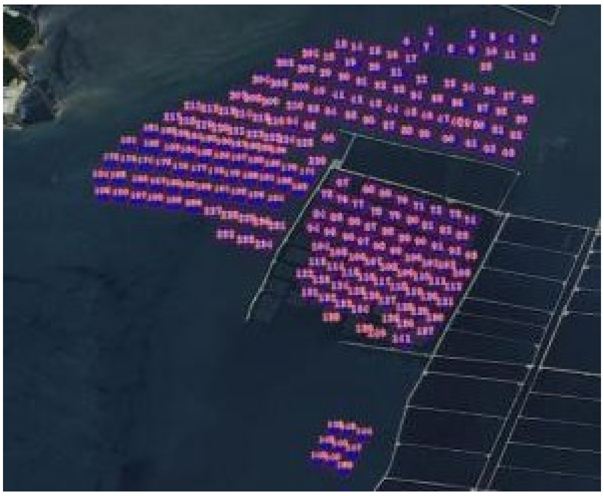

Fig. 1. An orthophoto of illegally occupied open sea for farming taken by a UAV

4. Illegal Sea Item Discovery Algorithm based on Hash table and DBSCAN Algorithm

In view of the above drawbacks, this paper proposes an illegal sea project detection algorithm based on Hash table and DBSCAN algorithm. Through clustering analysis of spatial location and time in AIS information, the clustering state of the ship is separated. Further, the spatial scope of the clustered coordinate points is judged to distinguish the scope of ship operations in the open sea, so as to narrow the key areas of government ships and drones, and improve the ability of maritime supervision.

From the analysis of actual data, the spatial scope of seawater cage culture is within 5 kilometers. Because the spatial range is not large, the error caused by the latitude and longitude span can be ignored when calculating the spatial distance, and the plane distance of latitude and longitude is directly used as the basis for spatial distance in clustering calculation.

Due to the limited travel distance of the ship operating in the current space, the spatial distance between the front and back points in the series of ship longitude and latitude information is not more than 1 minute (1 nautical mile, 1.851 km). Calculate the longitude and latitude difference between the two points before and after the longitude and latitude, and form a Key string like "0.01-0.01". The corresponding Key value is searched in the Hash Table calculated in advance and the corresponding latitude and longitude distance is obtained. If the corresponding Key value is not found in the table, the distance between the two points is beyond the predetermined spatial range, and it can be directly determined that the spatial distance is too large to belong to the same cluster.

DBSCAN clustering process can be used to directly calculate the spatial distance, its formula is the conventional triangular distance calculation method. At the same time, in order to improve the efficiency of spatial distance calculation, some scholars use R-Tree, Kd-Tree or M-Tree to improve the spatial search speed of spatial points in the clustering process, which solve the problem of large amount of data to varying degrees. In the improved algorithm proposed in this paper, the Hash Table can calculate and save the relevant latitude and longitude distance data in advance. In the actual use process, the calculation time is less, and its efficiency can be comparable with the spatial adjacent point search method of R-Tree, Kd-Tree and M-Tree, and the use process is simpler (the algorithm is shown in Figure 2).

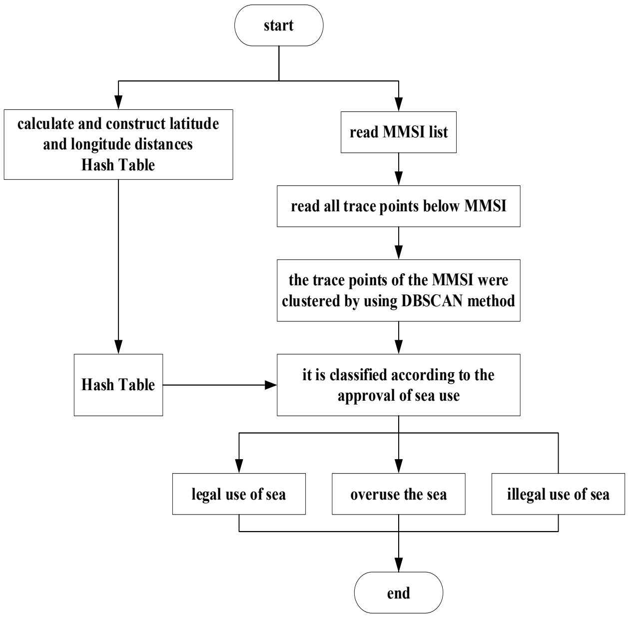

Fig. 2. Data processing flow chart of Hash Table and DBSCAN method combined

5. Experiment and Analysis

Taking the data of a total of 31 days from October 1, 2021 to October 31, 2021 as an example, the AIS data acquired by other projects have been saved in the data table in the previous data processing work. In the process of reading data through the data interface, the system saves two types of data tables at the same time. One is the data table saved by month, and a data table named "AISyyyymm" is created every month (such as "AIS202110" represents all AIS data in October 2021) to save the received data in a predetermined format. The second type is an index table established with the name of the first four digits of MMSI. For example, in the "AIS3456" table, a series of points with MMSI number "345678901" and "345666666" with "3456" as the first four characters of MMSI are saved. MMSI numbers with less than 4 digits are all uniformly reserved in the data table named "AIS0000" for query.

Firstly, the algorithm constructs a Hash Table for querying spatial distance according to the spatial distance difference between latitude and longitude. As mentioned above, the calculation process ignores the actual change of spatial position caused by the change of latitude and longitude, so the spatial table can be constructed only half of the predetermined number, that is, the query value with key value of "0.1-0.2" is consistent with the query value with key value of "0.2-0.1". Table 1 Spatial latitude and longitude distance table calculated according to latitude and longitude difference1 is the spatial distance table of latitude and longitude constructed according to the difference between latitude and longitude, and the unit is "minutes".

There are three solutions to construct and use this Hash Table. The first solution does not consider duplicate values and directly constructs a complete spatial longitude and latitude distance Hash Table; The second scheme considers duplicate values and constructs a spatial longitude and latitude distance Hash Table with half the length. When querying, the matching query is performed twice. If the first query fails, the longitude and latitude differences are swapped and the second query is performed, that is, the first query is "0.1-0.2", and then the second query is "0.2-0.1". If the query fails twice, it is considered that the space distance is too large to cluster. In the third scheme, the Hash Table is the same as the second one, but the latitude and longitude differences are sorted before querying, that is, whether it is "0.1-0.2" or "0.2-0.1", it is converted to "0.2-0.1" query.

Based on the analysis of data characteristics, it is considered that the AIS of small ships generally uses GPS (Global Positioning System) rather than DGPS (differential global Positioning System). Differential Global Position System), its position accuracy is not high, the error is usually 3-5 meters or more.Therefore, we set the granularity of Key values for constructing Hash Table to 0.1 seconds, that is, the spatial accuracy range is about 3 meters, and the spatial span is set to 0.2min *0.2min (the spatial distance is 360m *36 m),so as to construct a Hash Table with a length of 120*120. The cost of space occupancy is completely acceptable.

Table 1. Table of spatial latitude and longitude distance calculated according to the difference of longitude and latitude

Latitude / Longitude	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0.10	0.141	0.224	0.316	0.412	0.510	0.608	0.707	0.806	0.906	1.005
0.20	0.224	0.283	0.361	0.447	0.539	0.632	0.728	0.825	0.922	1.020
0.30	0.316	0.361	0.424	0.500	0.583	0.671	0.762	0.854	0.949	1.044
0.40	0.412	0.447	0.500	0.566	0.640	0.721	0.806	0.894	0.985	1.077
0.50	0.510	0.539	0.583	0.640	0.707	0.781	0.860	0.943	1.030	1.118

Next, the proposed algorithm is applied to the first type of AIS data table. The five fields of ID (sequence number automatically generated by the database), MMSI, longitude, latitude and time are mainly used to construct an entity class with five attribute values, and AIS data containing 22,912,791 records and 30,742 MMSI tracks is read into memory.

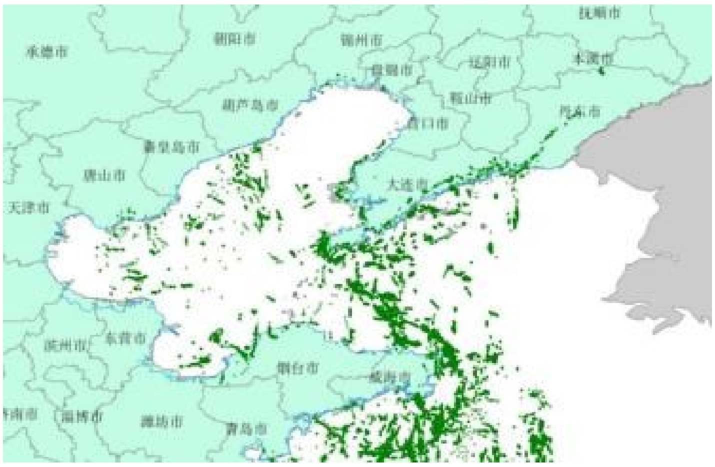

Fig. 3. AIS distribution map in Bohai Sea region in October 2021

Finally, the DBSCAN method was applied to perform spatial clustering. Firstly, the longitude difference and latitude difference of the two spatial locations are calculated. After the key value is formed, the Hash Table is queried to obtain the spatial distance. It can determine whether the corresponding clustering can be carried out or the next cycle can be started. When the spatial clustering within a month is completed, the MMSI number that needs to be focused on can be found out, and all the tracks of the MMSI number in a certain long time (such as within a year) can be completely obtained through the second type table, so as to further determine whether it meets other characteristics of the suspected ship illegally using the sea through manual judgment.

Figure 3 is the AIS distribution map of the Bohai Sea coastal area (116-127 E, 36-42 N) in October 2021; Figure 4 is the spatial clustering result of a single MMSI found, the spatial range of which does not have sea approval; Figure A is the distribution location of the track in the whole North Sea region. b is the clustering result of the two-month tracks in September and October 2021. When the clustering results of ships appear in an unapproved sea use area for several consecutive months, it can be basically concluded that there are different degrees of illegal sea use in the space area.

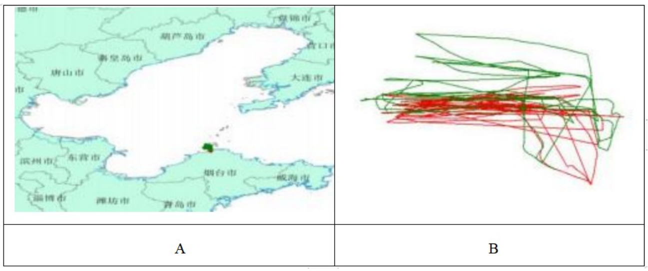

Fig. 4. Spatial clustering results of a single MMSI in the unapproved sea area

Through the processing of the above data, in the Bohai Sea, through the analysis of five consecutive months from June 2021 to October 2022, the spatial clustering ranges of 563 ship AIS trajectories can be obtained, of which 498 can basically coincide with the approved sea range for seawater cage aquaculture, accounting for 88.5%. There are 15 overlapping areas with the approved sea use scope, but the actual operation scope greatly exceeds the approved scope; Four are operating in the sea use areas that have not been approved or applied for at all, and there are relatively certain illegal sea use behaviors. In addition, there were 46 calculated spatial clustering areas, which were neither found to be approved by the sea nor actual ship operations by on-site investigation and UAV inspection. The overall accuracy of the data analysis is about 91.8%. From the perspective of regulatory efficiency, the ship trajectory clustering analysis method using this method can reduce the original ship and UAV inspection scope by about 40%, so as to effectively improve the regulatory efficiency and reduce costs.

6. Conclusion

Through the combination of Hash Table and DBSCAN algorithm, clustering analysis is carried out on the characteristics of ship cruising and operation in the Marine cage aquaculture area, which can effectively identify the operation process of Marine aquaculture. Combined with the approval results of various national, provincial and municipal sea use projects, the hidden illegal use of sea can be preliminarily confirmed. And these areas are regarded as key areas to combine UAV cruise and law enforcement and supervision ship cruise, improve the detection ability of offshore illegal use of the sea, and then solve the problem of difficulty and high cost of maritime natural resources supervision.

Acknowledgment

This work is supported by Marine Science and Technology Project of North China Sea Bureau, Ministry of Natural Resources(202108), Shandong Provincial Key Laboratory of Marine Ecology and Environment & Disaster Prevention and Mitigation(201803), National Natural Science Foundation of China (NSFC) (62201314), Natural Science Foundation of Shandong Province (No. ZR2020QF007), Research and exploration on the second classroom teaching reform guided by robot competition, Key technology tackling and industrialization demonstration projects of Qingdao(23-1-2-qdjh-18-gx) and Ministry of Education collaborative education project (220600643243903) .

References

[1] https://mp.weixin.qq.com/s?__biz=MzIzMDUxMDM0NQ==&mid=2247523927&idx=2&sn=a5f534512904d04830c48695b2ecc35f&chksm=e8b08431dfc70d27136a480b394977b3326b2cbf6d2f52d9ebff03a16b262f711c73ed268c81&scene=27

[2] Liu Yong. Research and implementation of spatial clustering algorithm based on

DBSCAN [D]. Yunnan University,2017. Xie Puchun. Ship based on convolution neural network [D]. Face detection research and application of donghua university,2022. The DOI: 10.27012 /, dc nki. Gdhuu. 2022.000699.
[3] Guo Naikun, Chen Mingjian, Chen Rui. A regard for the time characteristics o f ship trajectory DBSCAN clustering algorithm [J]. Journal of surveying and map ping engineering, 2021,30(3):5-58, DOI: 10.19349 / j.carol carroll nkiissn1006-7949.2021.03.009.
[4] Liu Y, PENG P F. Research on clustering method of ship trajectory based on improved DBSCAN [J]. Ship Electronic Engineering, 2012, 42(12):57-63.
[5] Dai Shao-sheng, Liu Xiao-bing, Lai Zhi-ying et al. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2022, Vol. 34(2):250-257.
[6] Zhang platinum. Based on the depth study of ship trajectory anomaly detectio n [D]. Dalian maritime university, 2021. The DOI: 10.26989 /, dc nki. Gdlhu. 2021.000213.
[7] Zhi-yuan zhang. Track prediction method based on the deep learning research [D]. Electronic research institute, China electronics technology group company 2021. DOI: 10.27728 /, dc nki. GDZKX. 2021.000061.
[8] Yu G. Urban road travel time prediction method based on deep learning of taxi spatio-temporal trajectory [D]. Southwest Jiaotong University, 2018.
[9] Lu Xiao. Clustering algorithm based on depth of learning research [D]. University of electronic science and technology, 2020. The DOI: 10.27005 /, dc nki.Gdzku. 2020.000269.
[10] Xie Puchun. Ship based on convolution neural network [D]. Face detection research and application of donghua university, 2022. The DOI: 10.27012 /, dc nki.Gdhuu. 2022.000699.
[11] Liu J,Qin H,Liu Z, et al. A Density-Based Spatial Clustering of Application with Noise Algorithm and its Empirical Research[C]//Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 3rd International Conference on Artificial Intelligence and Engineering Applications (AIEA 2022).2022:169-174.DOI:10.26914/c.cnkihy.2022.069447.
[12] Conte S I. The Sea of Stuff: a model to manage shared mutable data in a distributed environment[D]. University of St Andrews, 2019.
[13] Lin C Y, Dai G L, Chen Z, et al. Game of marine natural resources management: A strategy for determining rights registration[J]. Water, 2022, 15(1): 36.
[14] Wang S, Xu S. Marine Economic Development and Conservation[J]. Water, 2023, 15(12): 2253.
[15] Wang S, Tian W, Geng B, et al. Resource constraints and economic growth: Empirical analysis based on marine field[J]. Water, 2023, 15(4): 727.