Extraction of Frequently Active Areas of Ships Based on Advanced Grid Density Peak Clustering

Xuanrui Xiong[1], Han Shen[1], Lanke Zhu[2,3], and Jianbo Zheng[2,3](✉)

[1] School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
xiongxr@cqupt.edu.cn, s210131206@stu.cqupt.edu.cn

[2] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
{1120200314,jianbo.zheng}@smbu.edu.cn

[3] Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China

Abstract. The cognition of the frequent activity areas of ships based on AIS data is of great significance in reducing port navigation risks and improving the efficiency of ships entering and leaving ports. Traditional extraction methods only consider spatial information and ignore the impact of temporal information on clustering results, resulting in inaccurate extraction of frequently active areas. We propose an advanced grid density peak clustering method (AGDPC) to extract frequently active areas, which can advanced select cluster centers and density thresholds to solve the problem that grid density peak clustering methods cannot advanced select cluster centers. The improved grid density peak clustering method is used to extract frequent ship motion regions under a single spatial-temporal granularity according to a given spatial-temporal granularity. Then, we fuse multiple ship frequent activity areas to obtain multi-temporal and spatial granularity ship frequent activity areas. Experimental results show that this method can extract frequent motion are-as more accurately than traditional methods, and better reflect the ship's navigation rules.

Keywords: Trajectory clustering · Grid density peak clustering · Frequent activity areas extraction · AIS data

1 Introduction

The mining and analysis of existing data is one of the important means to predict and evaluate the future situation of objects. With the development of machine learning and deep learning, data mining and analysis techniques are widely used in economics, edge computing [1–3], blockchain [4–7] and other fields [8–10]. The Automatic Identification System (AIS) is a type of ship navigational system that contains essential information

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 158–164, 2024.
https://doi.org/10.1007/978-3-031-65126-7_16

such as Maritime Mobile Service Identity (MMSI), vessel position, and speed. By analyzing and mining AIS data, extracting frequent activity areas of vessels can provide technical support for research in detecting abnormal vessel behavior, predicting port traffic flow, voyage planning [11], and recognizing maritime target intentions.

At present, clustering algorithms have been widely used in the research of object hotspots region extraction [12]. Wang et al. [13] developed a rapid clustering model of trajectories based on hierarchical modeling. Each ship state establishes its trajectory similarity model and performs recursive clustering of ship trajectories from top to bottom, avoiding the cumbersome calculation of existing ship clustering models, high time complexity, difficult parameter adjustment process, and other shortcomings. In [14], a spectral clustering algorithm is used to cluster the sub-trajectory segments to identify representative ship maneuvering behavior trajectories. Hartawan et al. [11] suggested that a typical motion model of ships in the area could be obtained based on AIS data by DBSCAN clustering of ship trajectory segments combined with track similarity measure and extraction of typical trajectories.

The above methods only focus on the spatial information of moving objects and ignore the time information, which will result in the identification of an area with no ships or only a few ships in a certain interval as an area where ships are frequently active. In this regard, this paper proposes an advanced grid density peak clustering method (AGDPC). This method can extract frequent ship motion areas at multiple time granularities while using spatial clustering and considering ship time information.

2 Advanced Grid Density Peak Clustering

Traditional grid density peak clustering requires manual determination of cluster centers, which can easily lead to inaccurate clustering results. To address this problem, this paper uses the boxplot method and the elbow method to automatically determine the cluster center and number of clusters. First, in order to solve the problem that the local density and relative distance of grid objects affect the selection of cluster centers due to different dimensions, this paper first uses the minimum and maximum normalization method to map the value range of grid objects to the local density and relative distance of grid objects. Perform normalization processing between 0 and 1, and preselect the cluster center set. Then, use the box plot method to calculate its upper and lower bounds and quartiles to obtain its box plot distribution. Grid objects with higher local density and relatively far distance is further filtered according to the box plot distribution as a cluster set. At this time, the cluster center candidate set may contain more cluster centers than the actual situation, causing the classification of clusters to be too detailed. Because there may be grid objects in the cluster set that have a high local density but a small relative distance, or a grid object that has a small local density but a large relative distance, it is necessary to further screen the cluster center candidate set. This paper uses the elbow method to filter the cluster set. By finding the inflection point of the cluster center, the candidate points before the inflection point are used as the cluster center, and the remaining candidate points are assigned to the same cluster as its nearest high-density neighboring grid object., complete the cluster analysis of ship AIS data and obtain the ship activity area.} Based on the above ideas, the core regions of the clusters can be identified. First, count the number of times that each mesh object in the cluster is the nearest higher-density mesh object to other mesh objects:

$$nt_j = \sum_{i=1}^{n} z\left(j - \underset{j:\rho_j > \rho_i}{arg\ min}\left(d_{ij}\right)\right) \tag{1}$$

In the formula, $z(x) = \begin{cases} 1, x = 0 \\ 0, other \end{cases}$. d_{ij} is expressed as the Euclidean distance between the grid object i and the grid object j. ρ is the local density of the mesh object. Since it is difficult for a point located on the boundary of a cluster to become the closest high-density mesh object to other mesh objects, when is 0, the mesh object is usually located on the boundary area, so the core area of the cluster can be defined as:

$$c_{core}^k = \left\{x_i | \rho_i > \max\left(\rho_j\right), x_i \in c^k, x_j \in c^k \& nt_j = 0\right\} \tag{2}$$

In the formula, c^k represents the clusters obtained by clustering, c_{core}^k represents the core region of the class cluster c^k, $\max(\rho_j)$ is the maximum function. In these cluster core areas, although their density is larger than their neighbors, from the overall data distribution, some areas have relatively few ships and should not be considered frequent activity areas. In order to obtain the frequent activity areas of ships that meet the actual situation, these core areas need to be further screened. In order to reduce human participation, this paper automatically selects the density threshold $d_{th} = \max_{ntj=0}(\rho_j)$ according to the distribution characteristics of grid density in various clusters, and selects the grids whose grid density exceeds the threshold in various clusters:

$$area_{fre} = \left\{x_i | \rho_i > d_{th}, x_i \in c_{core}^k\right\} \tag{3}$$

By merging adjacent high-density mesh objects, the ship frequent activity area can be obtained. However, the ship frequent activity area extracted by this method ignores the time information. In fact, the areas of ship activities are different at different times. In this paper, by fusing ship frequent areas with single spatio-temporal granularity on the time axis, more accurate ship frequent areas with multiple spatio-temporal granularities are obtained.

3 Experiment and Analysis

In order to validate the proposed method, this paper uses two common frequent activity region detection methods for comparison. The first is the classic grid clustering method Clustering In QUEst(CLIQUE) [9], which uses the number of data points in the grid as the grid density to extract areas with frequent ship activities; the second is the advanced grid density peak clustering method proposed in this paper. This experiment selected AIS data of ships in the sea area of 122°35′W–123°55′W, 48°06′W–48°30′N from January 1, 2019 to January 3, 2019, and the data comes from the open source website https://marinecadastre.gov/ais/.

3.1 Experimental Parameter Settings

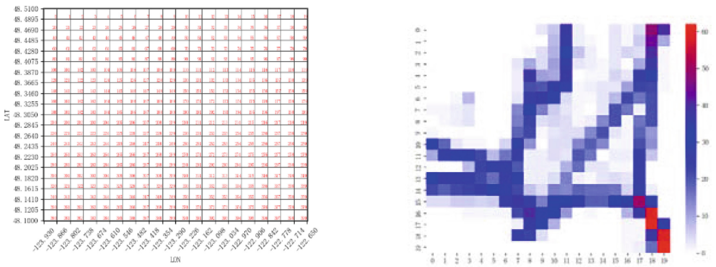
The parameter setting of the comparison method in the experiment is selected through manual tuning, as shown in Table 1.

Table 1. .

Parameter	Value
Meshing	20*20
Density threshold	200
Time granularity	1(day)

3.2 Results and Analysis

We first divide the experimental data into 20*20 grid areas. The number of ships in each grid area and the heat map are shown in Fig. 1. Subsequent experiments will be compared with Fig. 1.



(a)Meshing and grid number (b) heat map of ship distribution in grid

Fig. 1. The attributes of the research area (a) Meshing and grid number, (b) heat map of ship distribution

The extraction of frequent ship activity areas at a single spatio-temporal granularity refers to extracting frequent ship activity areas in different time periods within the same research area, given the time granularity and spatial granularity. We divide the time range into several uniform equal parts, and divide the space range into m*m grids. An improved grid density peak clustering algorithm is used to automatically select the cluster center and extract its frequent activity areas at a single spatio-temporal granularity. Figure 2 shows the frequent activity areas of ships extracted by CLIQUE. It extracts 6 frequently active regions. However, compared with Fig. 1, it can be seen that the ship density in some of the six frequent activity areas is very low, such as grid 107 in area 1, area 3, and

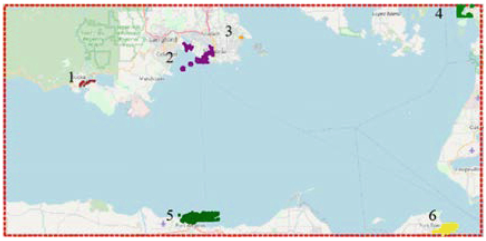


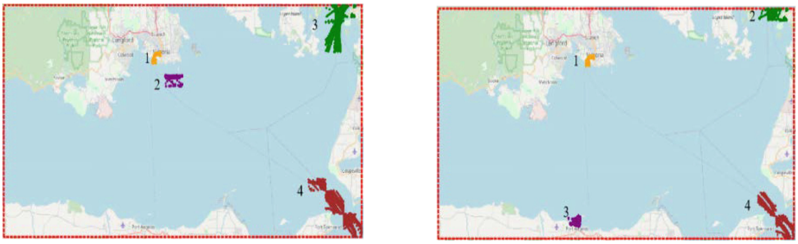
Fig. 2. Map visualization of frequent activity areas by using CLIQUE method

area 2, but it is identified as a frequent activity area. However, some areas among the screened-out areas have very high ship densities, such as grid 18, grid 38, grid 399, grid 379 and grid 358, which represent areas with significantly higher ship density than other areas., but was identified as an infrequently active area.} Fig. 3 shows the frequently active regions extracted by grid density peak clustering. Compared with Fig. 2, the density of frequent active areas extracted in Fig. 3 is in the forefront, which shows the effectiveness of the advanced selection threshold method proposed in this paper, which can correctly screen out high-density grids.

Fig. 3. Map visualization of frequent activity areas by using AGDPC method

Figure 2 and Fig. 3 only consider the spatial information of the frequent activity area extraction method, and can only obtain the frequent activity area in the entire large time period, but cannot obtain the frequent activity area in different time periods. In order to extract frequent activity areas more accurately, this paper firstly extracts the frequent activity areas of ships with single spatio-temporal granularity in different time periods in the same area under the given time granularity and spatial granularity. On this basis, on the time axis, if there is an intersection between frequent ship activity areas in adjacent time periods, the frequent activity areas in adjacent time periods are merged. Otherwise, the fusion of the next time period is performed until the time span is traversed.

The frequently active regions extracted at a single spatio-temporal granularity using the grid density peak clustering method are shown in Fig. 4. And the frequent movement area of ships with multiple spatial and temporal granularities is shown in Fig. 5. It can be seen that the frequent ship activity areas under multiple spatial-temporal granularity tend to be consistent on different dates, and the propagation activity area of a single



(a) Areas with frequent ship movements on January 1

(b) Areas with frequent ship movements on January 2

Fig. 4. Visualization of the map of the frequent movement area of ships with a single spatiotemporal granularity

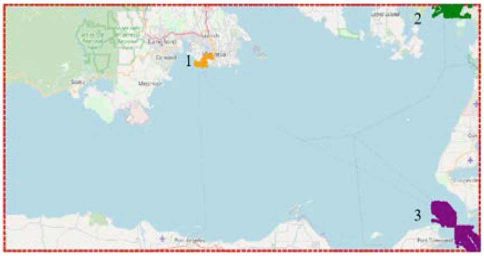


Fig. 5. Visualization of the map of the frequent movement area of ships with a multi-temporal and spatial granularity based on two-day data from January 1st to 2nd

spatial-temporal granularity shows great differences on different dates, which proves the effectiveness of the method proposed in this paper.

4 Conclusion

In this paper, we propose a method for extracting frequent ship moving areas based on grid density peak clustering, which solves the problem that grid density peak clustering methods need to manually select cluster centers. To learn more fine-grained spatial-temporal information, we consider frequently active regions of both spatial and temporal information. We fuse the frequently active regions with single spatiotemporal granularity on the timeline to obtain frequent active regions with multiple spatiotemporal granularities, which makes the extracted frequent active regions more accurate. In simulation experiments, we evaluate the effectiveness of the proposed ship frequent activity area extraction method and compare it experimentally with other methods. The results show that our method can more accurately and effectively extract the areas with frequent ship activities.

References

1. Wang, X., Ning, Z., Guo, L., Guo, S., Gao, X., Wang, G.: Mean-field learning for edge computing in mobile blockchain networks. IEEE T Mob. Comput. **22**(10), 5978–5994 (2022)

2. Ning, Z., et al.: Blockchain-enabled intelligent transportation systems: a distributed crowdsensing framework. IEEE T Mob. Comput. **21**(12), 4201–4217 (2021)
3. Ning, Z., et al.: Intelligent resource allocation in mobile blockchain for privacy and security transactions: a deep reinforcement learning based approach. Sci. China Inf. Sci. **64**(6), 162303 (2021)
4. Ning, Z., et al.: Dynamic computation offloading and server deployment for UAV-enabled multi-access edge computing. IEEE T Mob. Comput. **22**(5) (2021)
5. Ning, Z., Chen, H., Ngai, E.C., Wang, X., Guo, L., Liu, J.: Lightweight imitation learning for real-time cooperative service migration. IEEE T Mob. Comput. (2023)
6. Ferreira, M.D., Campbell, J., Purney, E., Soares, A., Matwin, S.: Assessing compression algorithms to improve the efficiency of clustering analysis on AIS vessel trajectories. Int. J. Geogr. Inf. Sci. **37**(3), 660–683 (2023)
7. Egala, B.S., Pradhan, A.K., Badarla, V., Mohanty, S.P.: Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control. IEEE Internet Things **8**(14), 11717–11731 (2021)
8. Xiong, L., Xiong, X., Zhang, F., Chen, H.: Unsupervised Deep Embedding Clustering for AIS Trajectory. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022, pp. 2283–2286 (2022)
9. Deebak, B.D., Al-Turjman, F.: Smart mutual authentication protocol for cloud based medical healthcare systems using internet of medical things. IEEE J. Sel. Area Commun. **39**(2), 346–360 (2020)
10. Ali, A., et al.: Security, privacy, and reliability in digital healthcare systems using blockchain. Electronics **10**(16), 2034 (2021)
11. Bureva, V., Popov, S., Traneva, V., Tranev, S.: Generalized net model of cluster analysis using CLIQUE: clustering in quest. Int. J. Bioautom. **23**(2), 131 (2019)
12. Wang, S., Li, Y., Xing, H.: A novel method for ship trajectory prediction in complex scenarios based on spatio-temporal features extraction of AIS data. Ocean Eng. **281**, 114846 (2023)
13. Ganesh, E.N., Rajendran, V., Ravikumar, D., Kumar, P.S., Revathy, G., Harivardhan, P.: Detection and route estimation of ship vessels using linear filtering and ARMA model from AIS data. Int. J. Oceans Oceanogr. **15**(1), 1–10 (2021)
14. Hartawan, I.P.N., Widyantara, I.M.O., Karyawati, A.A.I.N.E., Er, N.I., Artana, K.B., Sastra, N.P.: AIS data pre-processing for trajectory clustering data preparation. In: 2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), Bali, Indonesia, 2021, pp. 1–5. (2021)