Shipping Density Assessment Based on Trajectory Big Data

Zeyuan Dai\(^1\)*, Lihua Zhang\(^1\), Shuaidong Jia\(^4\), and Hengxiu Pang\(^2\)

\(^1\)Department of Military Oceanography and Hydrography & Cartography, Dalian Naval Academy, 116018 Dalian, Liaoning, China

\(^2\)North Theater Command Navy Support Team of Battlefield Environment of PLA, 266003 Qingdao, Shandong, China

* Corresponding author: carl-day@foxmail.com

Abstract. Due to the current shortage of maritime shipping density assessment methods, a method of shipping density assessment based on trajectory big data is proposed. Firstly, in order to obtain the trajectories following the actual situation, the AIS data is pre-processed. Secondly, as the minimum calculation unit, the grid is constructed, and the shipping density assesses evaluation index is established by comprehensively considering among the travel time, information transmission and the number of vessel crossing grids. Finally, with the proposed evaluation index, the shipping density is assessed quantitatively. The experimental results show that the proposed method can characterize the maritime shipping density and provide a new solution for maritime traffic situational awareness.

1. Introduction

With the continuous advancement of China’s “One Belt, One Road” initiative, the shipping industry has received more attention than before. How to assess the density of marine vessels and quantitatively extract the hotspots of maritime traffic has become an important part of the current maritime intelligent service. As a system that reports the position and identity of vessel automatically, the Automatic Identification System (AIS) can not only provide help for vessel’s collision avoidance but also contain trajectory information, reflecting the density of vessels at sea [1-3]. In order to get the situation quickly and timely, many scholars have used AIS data to do related research.

At present, the use of AIS data that extract shipping density mainly draws on the idea of land transportation, that is, the number of vessels per unit length (or area) in a certain channel (or open water). Some scholars use the selected cross-sectional line from the channel, counting the vessels’ distribution [4-6]. Similarly, some other scholars evaluate the shipping density by counting the number of trajectories in the traffic separation schemes [7]. In addition to the methods above, some scholars think that maritime traffic is different from land transportation, for there is no certain concept of “road,” and attempt to characterize the shipping density by using grids [8-10]. Some of them count the quotient between the number of AIS points and the area of the grid to reflect the shipping density [11]. However, vessels’ navigation is a dynamic process, and the density calculated by existing methods at a certain time cannot reflect the flow characteristics of maritime traffic, even conclude that inconsistent with the actual situation.

In order to solve the problems mentioned above, this paper proposes a shipping density assessment method based on big trajectory data. By cleaning, filtering and deleting invalid AIS data, the reliable trajectories is obtained; as the minimum calculation unit, the grids are constructed; taking dynamic and static properties into account, vessels’ travel time, message transmission count and crossing count is...
analyzed to obtain the shipping density assessment index; with this index, shipping density is quantitatively evaluated.

2. The method of shipping density assessment

2.1. AIS data pre-processing

The quality of the trajectories is not only a key to the comprehensiveness of analysis but also an essential factor in avoiding mistaken results [12]. During the voyage, bad weather, rough sea, channel bandwidth, and some other factors may affect the propagation of AIS data and will limit the usage of it greatly. In the absence of a systemic pre-processing of AIS data, the studies will be limited when using AIS data mining. Therefore, before using AIS data, it must be pre-processed to get the trajectories that match the real situation.

2.1.1 Cleaning the redundant data

There are 27 different types of AIS messages be broadcast. According to the International Convention for the Safety of Life at Sea (SOLAS), all vessels of 300 gross tonnes and above displacement engaged on international voyages, cargo ships of 500 gross tonnes and above displacement not engaged in international voyages and all passenger ships regardless of size are required to fit an AIS transceiver [13].

For the shipping density assessment requirement of this paper, the position information (messages 1&18) and the static ship information (message 5&19) are chosen, the rest of data is considered as the redundant data and discarded.

2.1.2 Deleting the unreliable data

The AIS data follows the recommendation M.1371–4 of International Telecommunication Union (ITU) [14], the reliability of AIS data means the majority amount of data respect to this recommendation. By using it, trusted AIS data could be extracted from the original data, and unreliable data can also be deleted. More than that, AIS data contains many coast towers, lighthouses, buoys and some useless objects for shipping density assessment, this paper also considers them as unreliable data. As mentioned above, this paper mainly uses position information and ship information in AIS data. Therefore, the integrity of AIS data can be checked from the information of the maritime mobile service identity (MMSI), speed of ground (SOG), course of ground (COG), heading of ground (HDG), navigation status, and the length, width, type of vessels.

With the AIS data collected on August 12th, 2017, the statistical results of unreliable data are shown in Table 1. Using COG as an example, there are 0.49% of data had values that smaller or larger than the range of 0 degrees and 360 degrees approximately. Basing on this result, the amount of data that does not meet the credible evaluation is deleted.

| Field          | Invalid range                  | Number of messages | Proportion |
|----------------|--------------------------------|--------------------|------------|
| MMSI           | $(-\infty, 2000000000) \cup (9000000000, +\infty)$ | 89769              | 0.89%      |
| SOG(kn)        | $(-\infty, 0) \cup (102.3, +\infty)$            | 12179              | 0.12%      |
| COG(°)         | $(-\infty, 0) \cup [360, +\infty)$              | 49779              | 0.49%      |
| HDG(°)         | $(-\infty, 0) \cup [360, +\infty)$              | 33440              | 0.33%      |
| navigation status | $[9, +\infty)$                | 122696             | 1.22%      |
| length(m)      | $(-\infty, 0) \cup (10220, +\infty)$           | 168195             | 1.67%      |
| width(m)       | $(-\infty, 0) \cup (1260, +\infty)$            | 412049             | 4.09%      |
| type           | $(-\infty, 20) \cup [90, +\infty)$              | 297241             | 2.95%      |
2.1.3 Filtering the data by using spatiotemporal characteristics

As mentioned in section 1, AIS data collect the vessels’ information and location information. Although the spatiotemporal characteristics of AIS data are implicit in the location information, it does not characterize the vessels’ trajectories directly. Without loss of the generality, in a short time between two AIS location point, the navigation state of the vessel can be approximated as a uniform linear motion, and any two adjacent points in the timeline can be connected as a segment of this vessel. Furthermore, for the vessels’ segment collection \( L = \{l_1, l_2, \ldots, l_n\} \), it is necessary to filter according to its spatiotemporal characteristics and obtain the trajectories that conform to the actual situation.

During the voyage of a vessel, the state of navigation in a short period is stable, and the host speed should not exceed the maximum value of the design speed. Although the speed of ground may be greater than the maximum host speed due to the influence of the ocean environment such as following the wind, current, and wave, there is generally no significant difference. Conversely, if a vessel has an AIS broadcast error during the voyage, it may cause the calculation of the speed between adjacent data points to be unreasonable. In order to avoid such data anomalies, the speed threshold \( V_{\text{max}} \) between two points needs to be set. As a criterion, the expression of \( V_{\text{th}} \) is as follows:

\[
V_{\text{th}} = V_{(p_{i-1}, p_i)} = \frac{\mathcal{D}(p_{i-1}, p_i)}{\Delta T} = \frac{\text{Dis}(p_{i-1}, p_i)}{T_{p_{i-1}} - T_{p_i}}
\]

Where \( V_{\text{th}} \) is the speed between the adjacent position \( i \) and \( i-1 \), \( \Delta T \) is the time interval between these two positions and \( \text{Dis}(p_{i-1}, p_i) \) means the function that calculates the true distance from longitude and latitude. With the same meaning, during the navigation of vessels, the broadcast interval of AIS messages is strictly following the standards shown in Table 2. Although the signal may lose due to various reasons, there is generally no significant difference. Basing that, the time interval threshold \( T_{\text{max}} \) between the adjacent points is set as the criterion for judgment. If the average speed or time interval is larger than the threshold \( V_{\text{max}} \) or \( T_{\text{max}} \), segment \( l \) is considered as a mistake and be deleted from the segment collection \( L \).

| Navigation speed (kn) | Broadcast interval (s) |
|-----------------------|-----------------------|
| In the anchorage       | 180                   |
| 0-14                  | 12                    |
| 0-14 with course changing | 4                     |
| 14-23                 | 6                     |
| 14-23 with course changing | 2                     |
| larger than 23        | 3                     |
| larger than 23 with course changing | 2               |

With all these works above, the pre-processing of the AIS track data can be completed according to the spatial and temporal distribution characteristics of the vessel position.

2.2. The grid-based method of quantitatively assess shipping density

2.2.1 Grids construction and assessment index expression

When sailing at sea, the largest difference between land transportation and maritime traffic is the former one does not need to fully comply with the restrictions of the traffic network, as long as the security can be fully guaranteed, it is navigable for wherever water area. In order to characterize and evaluate the density of vessels in different areas, the minimum calculation unit for shipping density
assessment can be provided by constructing grids. The size of grids reflects shipping density information under different precisions.

Based on the grids, it is necessary to provide an evaluation index for the vessels’ trajectories in the grids to quantitatively evaluate the shipping density value within the grids. In order to take the maritime navigation trend into account, compared with the method of simple statistics the number of AIS location points, this method combines the “dynamic” information of the vessels’ navigation with the “static” information of the vessels’ location. Integrating the vessels’ travel time, the number of messages sending, the shipping density evaluation index ρ is set, which can be described as:

\[ \rho_i = \frac{T_i}{M_i} \]  

(2)

Where \( \rho_i \) is the density of the \( i \)th grid, \( T_i \) is the crossing time, and \( M_i \) is the messages sending count of vessels in this grid.

The density index \( \rho \) mentioned above can reflect the dynamic characteristics of vessels during navigation. However, in addition to the navigation statue, the maritime behavior at sea includes static or incomplete static characteristics such as anchoring, exploration, and fishing. When the vessels stay in a certain area or at a low speed, the frequency of AIS data broadcasting decrease, which means the crossing time for several grids are longer than other vessels, and \( \rho \) increases, but this does not mean the density of vessels in this area is rising. In order to solve this problem, the number of vessels crossing the grid \( N \) has been introduced, and the original shipping density assessment index \( \rho \) is expanded to:

\[ \rho_i = \frac{T_i \times N_i}{M_i} \]  

(3)

This index fully reflects the dynamic and static characteristics of shipping density during navigation and can realize the shipping density characterization in a long period.

2.2.2 Shipping density assessment based on grids

After pre-processing the AIS data as described in Section 2.1, the vessels’ segment collection \( L \) can be obtained, where \( L = \{ l_1, l_2, l_3, \ldots, l_n \} \). Any segment \( l_i \) in this collection must belong to one or several grids.

According to the quantitative evaluation index of shipping density proposed in Section 2.2.1, it is necessary to calculate the value as the basis for calculating and visualizing the shipping density.

Taking segment \( l_i \) as an example, assume that the starting point is \( p_s \) and the ending point is \( p_e \). From the perspective of spatial relationship, the relationship between the segment \( l_i \) and the grid \( g_s \) can be divided into \( p_s \), and \( p_e \) of the segment \( l_i \) is in the same grid, or the different grids.

1. If \( p_s \) and \( p_e \) are in the same grid \( g_s \), the sailing time of this segment is part of the vessels’ travel time in this grid; the number of messages is sent (in the AIS data, it is represented as the number of location point) is also part of the number of messages sending in this grid. Also, the vessel has not left this grid and does not change the number of vessels crossing this grid, which can be described as:

\[ \rho_i(T, N, M) = \begin{cases} T = T + \Delta T \\
N = N \\
M = M + 2 \end{cases} \]  

(4)

2. If \( p_s \) and \( p_e \) are in two grids \( g_s \) and \( g_e \), respectively, the density is calculated according to the different locations they are and the crossing grids which the segment traverse. From \( g_s \) and \( g_e \), the AIS information is sent only at \( g_s \) and \( g_e \), and the number of information transmissions is increased in these two grids. For the number of vessel crossing and the traveling time, each grid from \( g_s \) and \( g_e \) is correspondingly increased. In Section 2.1.3, the motion of the vessel in any segment is approximated
as uniform linear motion. Therefore, the ratio of time that vessel passing through the grid \( g_i \) to the total time of the segment can be converted into the intersection length between the segment and the grid and the total length. 

Basing all these above, for the segment \( l_i \), the shipping density assessment of the corresponding grid can be described as follows:

\[
\rho(T, N, M) = \begin{cases} 
T = T + \Delta T \\
N = N + 1 \ (i = j) \text{or} (i = k) \\
M = M + 1 
\end{cases} \quad (5)
\]

\[
\rho(T, N, M) = \begin{cases} 
T = T + \Delta T \\
N = N + 1 \ (i \neq j) \text{and} (i \neq k) \\
M = M 
\end{cases} \quad (6)
\]

3. Experiments and Analysis

In order to verify the effectiveness of the proposed method, the AIS data from August 1\(^{st}\) to September 30\(^{th}\), 2017 was selected as the experimental dataset for experiments. These experimental data were collected from the official data service provider in China. The distribution of AIS data on August 12\(^{th}\) is shown in Figure 1.

![AIS point](image)

**Fig. 1.** The distribution of AIS data on August 12\(^{th}\).

Using the method mentioned in Section 2.1, the AIS data used in experiments need to be pre-processed. The dataset is 614,996,243 location points, and after pre-processing, the number of points is 542,672,685. To verify the impact of pre-processing, Figure 2 and Figure 3 show the trajectories before and after pre-process within the shores of China and its adjacent waters in August 12\(^{nd}\), 2017. Compared with Figure 2, which was overwhelmed by noisy lines that crossing continents, it is obvious that the unreliable location points have been deleted in Figure 3 and the quality of trajectories is improved.
Based on the AIS data that had been pre-processed, grids at resolution 3 minute by 3 minute had been constructed. The basic parameters of the PC used for the experiments are shown in Table 3, and the visualization of shipping density map is as shown in Figure 4.

Table 3. Hardware configuration for the experiments.

| Item                  | Type and parameters       |
|-----------------------|---------------------------|
| Processor             | Intel Core i5-6400 2.70GHz|
| RAM                   | 8GB                       |
| Hard Disk             | 1TB                       |
| Graphics               | NVIDIA GTX 1050 Ti        |
| Operating System      | Windows7 X64              |
Fig. 4. The shipping density map in August & September, 2017

To further verify the effectiveness of the proposed method and experimental results, a world shipping map generalized by the Central Intelligence Agency (CIA) of America is selected and shown in Figure 5 [15], and the level of routes on the map indicate the significance of route, not fully volume the maritime traffic.

Fig. 5. World shipping map generalized by CIA

Compared with the world shipping map, it can be seen that the shipping density assessed by the proposed method is slightly different from that from the proposed method, but the difference is not significant. This may due to the period of experimental data has fewer crossing vessels. In the meanwhile, the main hot areas can be quantitatively extracted by the proposed method.

4. Conclusions
In order to overcome the limitation of the existing shipping density assessment method that cannot take dynamic elements into account, a shipping density assessment method based on AIS trajectories big data is proposed. By pre-processing the AIS data, actual trajectories that reliable are extracted from the original dataset. Then, constructing the grids and the index of shipping density is set considering the static and dynamic properties. Finally, using the pre-processed AIS data and assessment index, the shipping density is calculated and visualized. The experimental results show that the proposed method can get the shipping density of from the trajectories big data with reliability.

Additionally, this proposed method only uses a two-month dataset and only uses AIS data. How to conduct this method for a longer period and make use of BeiDou satellite navigation data will need to be further studied.

Acknowledgments
This study was supported by the National Natural Science Foundation of China (41871369; 41471380; 41601498; 41774014)

References
[1] F. Xiao, H. Lijteringen, C. Gulijk, B. Ale, Ocean. Eng. 95, 84-93 (2015).
[2] L. Cheng, Z.J. Yan, Y.J. Xiao, Y.M. Chen, F.L. Zhang, M.C. Li, Sci. China. Tech. Sci. 61, 1-10 (2018).
[3] W.J. Yan, R. Wen, A.N. Zhang, D.Z Yang, IEEE Int. Conf. Big. Data. 3798-3806 (2016).
[4] B. Kundakçi, S. Nas, Pol. Marit. Res. 25(4), 49-58 (2018).
[5] M. Mustaffa, M. Abas, S. Ahmad, N.A. Aini, W.F. Abbas, S.A. Abdullah, N.I. Razak, M.Y. Darus, J. EMS. 4(4), 333-341 (2016).
[6] Y.C. Altan, E.N. Otay, J. Navigation. 70(06), 1367-1382 (2017).
[7] Z. Pietrzykowski, P. Wolejsza, J. Magaj, TransNav. 9(1), 121-126 (2015)
[8] P.R. Lei, T.H Tsai, W.C. Peng, IEICE APNOMS. (2016).
[9] F.X. Zhu, IEEE Int. Conf. Emerg. Manage. Manage. Sci. 772-775 (2011).
[10] Y.T. Wen, C.H. Lai, P.R. Lei, W.C. Peng, IEEE Int. Conf. Mobile. Data. Manage. 353-356(2014).
[11] C.B. Tang, Z.P. Shao, J.S. Wu, Q.R. Tang, J. Guangzhou. Marit. Coll. 20(03), 7-10 (2012).
[12] L.B Zhao, G.Y Shi, J.X Yang, J. Navigation. 71(05), 1210-1230 (2018).
[13] International Maritime Organization (IMO). International Convention for the Safety of Life at Sea(SOLAS). (2003).
[14] International Telecommunications Union (ITU). ITU-R M.1371-4. (2010).
[15] Central Intelligence Agency (CIA). Map of the World Oceans. (2015).