An Analysis Method of Ship Traffic Section Flow Based on Big Data

Liu Yi-fan1*

1 School of Navigation, Wuhan University of Technology, Wuhan 430036, China

*Corresponding author’s e-mail: liuyifan213@whut.edu.cn

Abstract. Traffic flow has important research value for traffic planning and traffic management. In waterway traffic, the cross-sectional flow of ships also has important application value in ship safety management, navigation rule making, bridge and wharf construction, and ship driver decision-making. In order to fully explore the characteristics of ship traffic flow in the big data environment, this paper studies the analysis method of ship cross-section flow based on big data. First of all, in order to ensure the accuracy of calculation and reduce the calculation time to improve calculation efficiency, based on AIS big data optimization data processing method, complete data storage strategy and construction of data management mechanism. Then, the ship section flow analysis algorithm is proposed and implemented. Finally, based on the ship section flow analysis algorithm, a ship section flow analysis and statistics system is constructed to complete the flow calculation and information statistics. The research results of this paper make the relevant analysis of ship traffic flow more convenient and faster, and have greater application value for waterway traffic planning and management.

1. Introduction

With the development of water transportation, ship traffic accidents occur frequently, causing unpredictable economic losses and damage to human life. Traffic accidents are often accompanied by management factors. For example, the density of ships in waters with heavy traffic flow is also higher, and the risk of accidents is usually greater. By understanding the traffic flow and carrying out corresponding traffic management or control, the occurrence of traffic accidents can be effectively reduced. Manual observation of ship cross-section flow is currently the most widely used flow observation method in our country, that is, the channel management department arranges for observers to select the most typical days of each month to observe the ship cross-section. However, the manual observation method has many shortcomings, such as large interference by human factors and limited sample size. Therefore, it is necessary to study a set of feasible monitoring methods for cross-section flow.

Qualitative and quantitative analysis methods are currently the most used methods to solve ship traffic flow forecasting. This article mainly uses quantitative analysis methods to achieve traffic flow forecasting. Tian Yanhua et al. [1] selected the traffic flow data of the Yangtze River Estuary for model training, and the results showed that the prediction model was effective in predicting traffic flow. Liu Jingxian et al. [2] used the nonlinear fitting ability of neural network to solve the small sample problem in the prediction process, and ensured the accuracy and stability of the prediction. Feng Hongxiang et al. [3] verified the feasibility of the prediction model in short-term prediction of ship flow. Wu Baoquan [4] used the statistical method of traffic flow on the road for reference, and proposed a method for monitoring the flow of ships in the channel section between the ship locks. By setting up two sets of
monitoring equipment in a certain segment, when the ship passed the two sets of monitoring equipment in order, then the cross-sectional flow of the channel can be measured. Hussein Dia [5] proposed that dynamic neural networks have superior prediction performance compared with static classifiers, and proved the feasibility of using object-oriented neural network methods to predict short-term traffic. Yisheng Lv et al [6] proposed a deep learning method based on SAE model to predict traffic flow, and successfully discovered potential traffic characteristics.

Based on the research of the above scholars, this paper proposes an analysis method of ship cross-sectional flow based on the AIS big data of the automatic ship identification system, which can quickly obtain the cross-sectional flow of the channel or route, so as to reduce the work intensity of the observer and grasp it in real time. The characteristics of traffic flow elements improve the accuracy of ship flow statistics, and further support waterway traffic risk analysis and traffic control to reduce ship traffic accidents and ensure water safety and environmental safety.

2. Data processing

2.1 AIS data preprocessing

2.1.1 AIS message analysis
In terms of logical function, it can be divided into 5 parts.
(1) Reading part of AIS message: read the original message of AIS.
(2) The check part of the AIS message: The AIS message is checked by the method of comparing the cyclic redundancy CRC (Cyclic Redundancy Check) with the message.
(3) The effective extraction part of the AIS message: extract the AIS message data after verification processing.
(4) Decoding part of AIS message data
(5) AIS analysis information storage part
   (√) The data after preliminary analysis is stored in the form of a notepad file
   (√) Then write Java code as needed, extract the data in notepad format, store it in excel, and finally import it into the database for storage and management.

2.1.2 AIS error message correction

2.1.2.1 Information error type
   ① Static information error
      IMO number is unavailable (zero), MMSI input error, ship name is missing, ship name input error, call sign is missing, ship type is not available, ship draught data is not available, captain data is missing (0), the distance from the bow to the antenna is missing, Missing ship width distance (zero), call sign format output error, ship length is less than ship width, the distance from the bow to the antenna compared with the ship length does not match the actual situation, the distance from the right chord to the antenna compared with the ship width does not match the actual situation, etc....
   ② Dynamic information error
      The AIS device is faulty or the AIS device has not been connected to other devices, such as the connection between the AIS device and GPS, etc., or the device connected to the AIS device is faulty, resulting in errors in the collected AIS information. On the other hand, the on-duty pilot or the responsible crew member has not modified the ship’s navigation status, for example, the ship has left the berth and is sailing normally, while the equipment is still at anchor. This can also cause errors in dynamic information.

2.1.2.2 Data error correction method
   ① Data cleaning
Data cleaning refers to removing the abnormal information part of the data. For example, in static data, the IMO number is not available (zero) or the field is empty. The MMSI input is incorrect. Missing ship name, ship name input error, call sign missing, ship type unavailable, ship draught data unavailable, ship length data missing (0), ship head to antenna distance missing, ship width distance missing (zero), call sign format output error. The length of the ship is less than the width of the ship, the distance from the bow to the antenna is greater than or equal to the length of the ship, the distance from the right chord to the antenna is greater than or equal to the width of the ship, etc. In the dynamic data, the ship’s longitude and latitude information is obviously inconsistent with the current GPS measured longitude and latitude data, that is, data where the longitude and latitude are not in the normal range, such as the longitude value exceeds 180°, and the latitude value exceeds 90° recording. The data record of the heading exceeding 360°, the data record of the course deviation between the ground course and the course calculated by continuous latitude and longitude. Data records with negative values for ground speed, data records with large deviations between ground speed and the speed calculated by continuous latitude and longitude.

② Data repair

In order to ensure the accuracy and completeness of the data, it is necessary to repair the missing parts of the AIS data during the process of collection and transmission. For example, if the trajectory data sent by the ship is interrupted and the position point jumps, it is necessary to clear the abnormal trajectory data point of the jump. After clearing the abnormal trajectory point, the original ship trajectory data point needs to be repaired.

2.2 Storage and management of trajectory data

2.2.1 Database design

Trajectory data includes MMSI code, ship name, call sign, ship length, ship width, ship type, ship latitude and longitude, course, speed, draft, time stamp, etc. In order to achieve data storage management and effective query, the data is imported into a table in the database.

2.2.1.1 Static information table design

This article takes the collected data of Fujian in January as an example. A static information table is created in the MySQL database to store the AIS data of the ship’s original trajectory, including MMSI, Shipname_EN, Shipname_CN, Callsign, IMO, Shiptype_EN, Shiptype_CN, Navi_status_EN, Navi_status_CN, Length_d, Width_d, Destination and other attribute fields. Among them, MMSI, Length_d, Width_d are of Double type, and no data length is assigned; Shipname_EN, Shipname_CN, Callsign, IMO, Shiptype_EN, Shiptype_CN, Navi_status_EN, Navi_status_CN, and Destination are of Varchar type, and the data length is 255. The specific table creation information is shown in the static information in Table 1:

| Column name         | type of data | length | Comment                                |
|---------------------|--------------|--------|----------------------------------------|
| MMSI                | Double       |        | Marine mobile communication service identification number |
| Shipname_EN         | Varchar      | 255    | Ship name (English)                     |
| Shipname_CN         | Varchar      | 255    | Ship name (Chinese)                     |
| Callsign            | Varchar      | 255    | Call sign                               |
| IMO                 | Varchar      | 255    | International Maritime Organization Number |
| Shiptype_EN         | Varchar      | 255    | Ship Type (English)                     |
| Shiptype_CN         | Varchar      | 255    | Ship Type (Chinese)                     |
| Navi_status_EN      | Varchar      | 255    | Navigation status (English)             |
| Navi_status_CN      | Varchar      | 255    | Navigation status (Chinese)             |
| Length_d            | Double       |        | Ship length                             |
| Width_d             | Double       |        | Ship width                              |
2.2.1.2 Dynamic information table design.

This article takes the collected Fujian January data as an example. A dynamic information table is created in the MySQL database to store the AIS data of the ship’s original trajectory, mainly including MMSI, Draught, Heading, Course, Speed, Lat, Lon, ROT, Last_time, Unixtime, Lat_d, Lon_d and other attribute fields. Among them, MMSI, Heading, Course, Speed, ROT, Last_time, Unixtime, Lat_d, Lon_d are Double types, and the byte length is not allocated; Draught, Lat, Lon are Varchar types, The data length is 255. The specific table creation information is shown in Table 2 dynamic information:

| Column name | type of data | length | Comment                     |
|-------------|--------------|--------|-----------------------------|
| MMSI        | Double       |        | Marine mobile communication service identification number |
| Draught     | Varchar      | 255    | Draft                       |
| Heading     | Double       |        | Heading                     |
| Course      | Double       |        | Track                       |
| Speed       | Double       |        | Speed over ground           |
| Lat         | Varchar      | 255    | Latitude (varchar type)     |
| Lon         | Varchar      | 255    | Longitude (varchar type)    |
| ROT         | Double       |        | Heading angular velocity    |
| Last_time   | Timestamp    |        | World time                  |
| Unixtime    | Double       |        | Timestamp                   |
| Lat_d       | Double       |        | Latitude (double type)      |
| Lon_d       | Double       |        | Longitude (double type)     |

2.2.2 Track information storage

2.2.2.1 Database information

Import AIS data into the database according to the format of the static data table and the dynamic data table.

2.2.2.2 Database processing

① Group by MMSI and Lat_d field to find out MMSI, Unixtime, Lat_d, Lon_d

SQL statement: SELECT MMSI,Unixtime,Lat_d,Lon_d
FROM ‘Table Name’
GROUP BY MMSI,Lat_d;

② Sort the results in 1 in ascending order of the Unixtime field

SQL statement: SELECT*FROM(SELECT MMSI,Unixtime,Lat_d,Lon_d
FROM ‘TABLE NAME’
GROUP BY MMSI,Lat_d AS t
ORDER BY MMSI,Unixtime;

③ Remove part of the MMSI of ships that only sent one message

Code SQL statement: SELECT COUNT(*),MMSI
FROM ‘TABLE NAME’
GROUP BY MMSI
HAVING COUNT(*)>1;

④ Eliminate some missing ship information with longitude and latitude

SQL statement: DELETE FROM ‘TABLE NAME’
WHERE Lat_d IS NULL;
3. System realization

3.1 Realization of Section Analysis Algorithm

Java Topology Suite (JTS) is an API for processing data in Java. This article uses the GeometryFactory, Geometry and Coordinate classes in JTS, and uses the Intersects space operation and the LineString space data type to implement the flow algorithm.

First, extract the coordinate data of the same MMSI ship in chronological order, as shown in Figure 6, coordinates of ship \( MMSI_1 \) at time \( t_1 \): \( C_{t_1, MMSI_1} \), and coordinates of ship \( MMSI_2 \) at time \( t_2 \): \( C_{t_2, MMSI_2} \). Secondly, set the ship section and extract the coordinates A and B of the section. Finally, using the jts toolkit in Java, construct \( C_{t_1, MMSI_1} \), \( C_{t_2, MMSI_2} \), and coordinate A, coordinate B into line segments, and judge whether they intersect.

![Algorithm implementation](image)

3.2 Realization of Section Analysis System

3.2.1 System interface

3.2.1.1 Interface introduction

![Cross-section ship flow calculation and statistics interface](image)

As shown in Figure 2, the cross-sectional ship flow calculation and statistics interface has five parts, including the longitude and latitude input of the cross-section, the time selection of the flow calculation area, the selection of the flow information database table, the selection of the flow calculation area, and the output of the cross-sectional flow value.
Input the latitude and longitude of the section, select two points of the line segment where a vertical section of the calculation area intersects the water flow, obtain the latitude and longitude of the two points, and enter it into the latitude and longitude input box.

The time selection of the flow calculation area. In this paper, there are four areas including Wuhan, Nanjing, Fujian and Shandong. Each area has 12 months of data, so it is necessary to select the time of the calculation area.

The selection of the flow information database table, the information of ships passing through the section after the section flow calculation will be imported into the database table. Therefore, the flow information database table of a certain section in a certain month that needs to be calculated has been established in the database, so it is necessary to select the flow information database table to be imported in order to obtain and count the information of ships passing through the section.

Selection of flow calculation area, select the area that needs to calculate flow.

Output of cross-section flow rate value, the final flow rate value will be output to the output box.

### 3.2.1.2 The realization of the interface

The interface consists of a window and four components. The components include a label component, a text box component, a drop-down box component, and a button component, respectively, as shown in Figure for the latitude and longitude of the interrupt surface A, the number input box, the selection box and the Wuhan button. Use Java language and swing classes, use MyEclipse to compile software, and build a visual graphical interface.

### 3.2.2 Realization of traffic statistics

#### 3.2.2.1 MySQL database and Java connection

Based on MyEclipse software, using Java language to access MySQL database using JDBC (Java DataBase Connectivity, Java) connection.

Start the MySQL server  
MyEclipse import jar package  
Use the MySQL server client SQLyog to establish a database  
Load the data access driver  
Configure the database link address  
Configure user name and password  
Establish connection  
Execute SQL language  
Processing result set  
Close the database

#### 3.2.2.2 Java extracts data from MySQL

The trajectory data of a ship is associated with the MMSI code. Each time the trajectory information sent by the ship must have an MMSI code, which is used to distinguish the trajectory information of different ships in order to generate the trajectory of each ship. Therefore, this article uses MMSI code to access other track information of this ship. Part of the track information is shown in Table 3:

| MMSI    | Unixtime  | Lat_d     | Lon_d     | Shipname\_CN | Shiertype\_CN | Length_d | Width_d |
|---------|-----------|-----------|-----------|---------------|---------------|----------|---------|
| 200023823 | 1546496522 | 26.61769  | 119.845582 | NULL          | Fishing boat  | 31       | 6       |
| 200023823 | 1546496696 | 26.6188   | 119.842112 | NULL          | Fishing boat  | 31       | 6       |
| 200023823 | 1546496757 | 26.621167 | 119.839095 | NULL          | Fishing boat  | 31       | 6       |
| 200023823 | 1546497507 | 26.631592 | 119.819435 | NULL          | Fishing boat  | 31       | 6       |
| 89789785  | 1547182814 | 26.62077  | 119.834027 | NULL          | NULL         | 40       | 7       |
| 89789785  | 1547185374 | 26.585217 | 119.821852 | NULL          | NULL         | 40       | 7       |
| 100153451 | 1546304260 | 26.586967 | 119.789642 | NULL          | Fishing boat  | 33       | 6       |
| 100153451 | 1546304749 | 26.556982 | 119.801277 | NULL          | Fishing boat  | 33       | 6       |

First of all, the MMSI is processed in the MySQL database, and the processing results are shown in Table 4.
In MyEclipse, execute the SQL statement 
\[ \text{SELECT * FROM 'MMSI' '码表名';} \]
At the same time, due to the large amount of actual data, it is impossible to determine the length of the data queried from the MySQL database after the SQL statement is executed by MyEclipse, so you need to define the list container first, and then store the data found from the MySQL database in the list container. Then, define a one-dimensional array, the length of the array is the length of the list container, and store the data of the list container in the one-dimensional array in order. Finally, use the MMSI code in MyEclipse to access other track information in the MySQL database, and store the track information in multiple arrays. The executed SQL statements are as follows: 
\[ \text{SELECT track information FROM 'Table Name' WHERE MMSI=MMSI[i]} \]
Where MMSI[] is an array of MMSI codes. The results of the interview are shown in Table 5.

Table 5. MMSI code to access other track information data

| MMSI[0](200023823) | MMSI[1](89789785) | MMSI[2](100153451) |
|------------------|------------------|------------------|
| ARR1[][Unixtime]  | ARR1[][Unixtime]  | ARR1[][Unixtime]  |
| ARR2[][Lat_d]     | ARR2[][Lat_d]     | ARR2[][Lat_d]     |
| ARR3[][Lon_d]     | ARR3[][Lon_d]     | ARR3[][Lon_d]     |
| ARR4[][Shipname_EN]| ARR4[][Shipname_EN]| ARR4[][Shipname_EN]| |
| ARR5[][Shipname_CN]| ARR5[][Shipname_CN]| ARR5[][Shipname_CN]| |
| ARR6[][Shiptype_EN]| ARR6[][Shiptype_EN]| ARR6[][Shiptype_EN]| |
| ARR7[][Shiptype_CN]| ARR7[][Shiptype_CN]| ARR7[][Shiptype_CN]| |
| ARR8[][Length_d]  | ARR8[][Length_d]  | ARR8[][Length_d]  |
| ARR9[][Width_d]   | ARR9[][Width_d]   | ARR9[][Width_d]   |

For example, through MMSI[0] that is, MMSI=200023823, store the Unixtime, Lat_d, Lon_d, Shipname_EN, Shipname_CN, Shiptype_EN, Shiptype_CN, Length_d, Width_d information of this ship into different arrays respectively. Traverse all the MMSI codes in the MMSI[] array in turn to obtain all the track information of each ship.

3.2.3 Realization of flow calculation
Take the ship with MMSI=200023823 in Table 5 as an example, the array ARR2[i] storing latitude data and the array ARR3[i] storing longitude data respectively store latitude and longitude data in chronological order. Now take the adjacent latitude and longitude data in turn, as shown in Figure 3.

![Fig.3 Near latitude and longitude extraction](image-url)

Fig.3 Near latitude and longitude extraction
Use the Geometry class and GeometryFactory class in the JTS toolkit to construct the extracted longitude and latitude data into line segments, and then perform the Intersects space operation with the line segment formed by the longitude and latitude of the selected section. If they intersect, add 1 to traverse all ships in turn. Perform the above operations on the latitude and longitude of each ship in turn, and finally get the ship flow.
3.2.4 Implementation of Traffic Information Classification

The graphical interface for cross-section flow calculation can only obtain the number of times the ship has passed through the cross-section, and cannot obtain the detailed information of the ships that have passed through the cross-section. For this reason, the code is optimized. If the latitude and longitude data of a ship with MMSI=MMSI[i] is judged as intersecting by the algorithm, then MMSI=MMSI[i] ship’s Unixtime, Lat_d, Lon_d, Shipname_EN, Shipname_CN, Shiptype_EN, Shiptype_CN, Length_d, Width_d are imported into the database. Take Fujian’s January trajectory information and section AB as an example, as shown in Figures 4 and 5.

After calculation, the ship has passed through the section 380 times, and the information of the ship through the section is shown in Table 6.

| MMSI   | Shipname_CN | Shiptype_CN | Length_d | Width_d | Number of ship passing sections |
|--------|-------------|-------------|-----------|---------|---------------------------------|
| 101600921 | NULL       | Fishing boat | 22        | 6       | 2                               |
| 123698578 | NULL   | NULL         | 23        | 4       | 27                              |
| 170000000 | NULL | Fishing boat | 26        | 6       | 1                               |
| 200008668 | NULL   | Work boat    | 65        | 8       | 2                               |
| 200011984 | NULL | Passenger ship | 35       | 6       | 7                               |
| total: 89 |           |              |           |         | total: 380                      |

(1) Classification by ship type

(2) Classified by captain

![Fig.4 Sectional interception](image1)

![Fig.5 Flow calculation result](image2)

![Fig.6 Traffic information classified by ship type](image3)

![Fig.7 Traffic information classified by captain](image4)
4. Conclusion
There are more and more researches on cross-section ship flow statistics based on AIS big data, and they are becoming more and more perfect, but the implementation process is more complicated. This article proposes a relatively simple implementation process and more convenient algorithm, and uses the Java language to use JTS tools The package and MySQL database realize the flow statistics; the graphical interface for flow calculation and statistics is designed, and the flow calculation area and calculation period can be selected in the interface; and the statistics of the ships through the section are classified. Compared with the previous traffic statistics method, it is easy to operate and more practical. The system can conveniently count ship traffic flow information. It is of great significance for maritime administration agencies to formulate ship navigation management rules, bridge and wharf construction to avoid waters with large flows, and ship drivers to understand the flow of navigable waterways.

References
[1] Tian Yanhua, Chen Jinbiao. Research on Ship Traffic Flow Forecast Based on BP Neural Network[J]. Ship and Ocean Engineering, 2010, 39(01): 122-125.
[2] Liu Jingxian, Liu Zhendong, Zhou Feng. Prediction model of ship traffic volume based on generalized regression neural network[J]. China Navigation, 2011, 34(02): 74-77+85.
[3] Feng Hongxiang, Xiao Yingjie, Kong Fancun. Prediction model of ship traffic flow based on support vector machine[J]. China Navigation, 2011, 34(04): 62-66.
[4] Wu Baoquan. Detection technology of ship flow in the channel section between ship locks[J]. China Water Transport, 2010(05): 40-41.
[5] Dia H. An object-oriented neural network approach to short-term traffic forecasting[J]. European Journal of Operational Research, 2001, 131(2): 253-261.
[6] Lv Y, Duan Y, Kang W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2015, 16(2): 865-873.