Application of neural network in abnormal AIS data identification

Yongming Wang
School of Navigation dept.
Dalian Maritime University
Dalian, China
E-mail: wangyongming@cttic.cn

Abstract—**Due to human tampering, equipment failure, channel congestion and other reasons, AIS data received by base station may have errors. These abnormal AIS data are not conducive to the identification and supervision of ship navigation intention, which greatly reduces the application value. Based on the analysis of the characteristics of the abnormal AIS data, through preprocessing and normalization of several adjacent AIS data, a model of the abnormal AIS data screening based on neural network is constructed, and the model is verified by the AIS data of the sea area near the Bohai Bay, Chengshantou Water Area, with an accuracy of 95.16%. At the same time, the influence of AIS data length and number of hidden layer nodes selected in the screening model on the accuracy rate is analyzed through experiments. The experimental results show that unreasonable data length and number of hidden layer nodes will reduce the accuracy rate of the screening model. When the data length is 4 and the number of hidden layer nodes is 6, the accuracy rate of the screening model reaches the highest.**

Keywords—AIS data, abnormal data, data screening, neural network

I. INTRODUCTION

The data of automatic identification system (AIS) contains rich ship traffic information, mainly including static information such as MMSI number, ship type, ship's on-going state and overall dimension, and dynamic information such as ship's speed, course and position (longitude and latitude)[1, 2] . It is not only the main equipment for the pilot to understand the navigation intention of other ships around the ship, but also the main technical means for the maritime regulatory authorities to track and supervise the ship. As AIS uses the connectionless user datagram protocol (UDP) for broadcasting, there is no corresponding fault-tolerant mechanism, neither packet grouping and assembly nor packet sorting is supported. After the message is sent, it is impossible to know whether it arrives completely. Any error in any position of the packet will lead to the failure of the entire packet[3,4]. Therefore, the data reliability in AIS message is poor, and there is the possibility of data loss and error[5]. Moreover, the static information of AIS message is manually input by the crew. Some crew members input wrong data to avoid supervision. The dynamic information of AIS message comes from gyroscope, GPS, velocimeter and other navigation equipment. The failure of these equipment will also lead to inaccurate AIS data. Generally speaking, the AIS data received by the base station in engineering practice is often incomplete or inaccurate[6]. These abnormal AIS data are not conducive to the identification and supervision of the ship's navigation intention, which greatly reduces the application value. It is necessary to screen the data in the AIS message and clear the abnormal AIS data.

II. RELATED WORK

AIS data reflects the movement of the ship in a period of time. The real movement track of the ship must be continuous, and there must be a certain correlation between several adjacent AIS data[7] , meeting certain rules.

There are many reasons for AIS data abnormal. In addition to human tampering and navigation equipment failure, channel congestion is also an important reason. In 2012, Ma Feng et al. Revealed the influence of AIS short message length and channel congestion on AIS packet error rate through hardware in the loop simulation, gave a prediction model of packet error rate, and put forward the problem of AIS abnormal data screening for the first time[8]. After that, according to the receiving range of AIS base station and the maneuvering characteristics of the ship, Lin-Zhi Sang et al. Divided the AIS abnormal data into three categories: position abnormality, ship speed abnormality and course abnormality, and made the judgment rules according to the maneuvering characteristics of the ship to identify the AIS abnormal data[9]. According to the transmission characteristics of the radio, the method calculates the latitude and longitude range of AIS message that the AIS base station can receive from the ship station. If the ship position in the AIS data is within the range, the data is normal; otherwise, the data is abnormal. Similarly, Wei Guirong et al. Calculated the average speed V_{sp} between two adjacent ship positions in the AIS data, and judged whether the AIS data was abnormal according to the relationship between the average speed and the speed threshold $V_{limitsp}$ [5] . This kind of rule-based screening method is simple in operation, high in efficiency and fast in processing. It has been widely used in AIS abnormal data screening. However, this kind of screening method also has some shortcomings. Because the threshold needs to be set for rule-based screening methods, whether the threshold is reasonable or not directly determines the accuracy of abnormal data screening. For different base stations, different ships, and even different waters, these thresholds are usually different, and the set thresholds are usually difficult to achieve better results in both the missing alarm rate and the false alarm rate.

In view of the shortcomings of the rule-based screening method, Chu Xiumin and others constructed a probability reasoning model to judge the correctness of AIS dynamic information from the perspective of probability. The method transformed the frequency distribution of ship speed, heading angle and position in AIS data into the evidence reliability between 0 and 1, and synthesized the rules of evidence reasoning to the AIS dynamic data of Tianxingzhou bridge water area of Wuhan is screened, and the recognition accuracy is close to the artificial level[10]. Nie Yang et al. Constructed a model of AIS data screening based on DS evidence theory to screen the AIS abnormal data of cargo ships in Baishazhou waters of the Yangtze River, and analyzed the temporal and spatial characteristics of the abnormal AIS data[11]. The above screening methods are based on the historical AIS data of a certain water area, and fully consider the behavior characteristics of ships in that water area. Because of this, it is often necessary to reconstruct the model for AIS data in other waters.

From the above analysis, it is not difficult to see that the key to identify anomalies is to establish a description model based on normal data, and then find the data inconsistent with the model and regard it as an exception. In the 1980s, great progress has been made in the research of artificial neural network. The relevant theories and methods provide new solutions for the establishment of complex models. Because it can fit nonlinear function, it has strong expression ability. Artificial neural network has been widely used in pattern recognition, image processing, intelligent control and other fields. Back propagation neural network(BP neural network) is a multilayer feedforward neural network trained according to the error back propagation algorithm, which is also the most widely used neural network. It has strong nonlinear mapping ability and flexible network structure. The number of intermediate layers and neurons in each layer of the network can be set arbitrarily according to the specific situation, and its performance varies with the difference of the structure. Therefore, in order to improve the applicability and reliability of AIS abnormal data screening model, this paper uses BP neural network model to solve the problem of AIS abnormal data screening, describes the correlation between several adjacent AIS data points through neural network, and then completes the AIS abnormal data screening.

III. MODEL BUILDING

The AIS dynamic data of the ship in a period of time can be expressed in Formula (1). Where x_i represents the longitude of the ship in the AIS message i, y_i represents the latitude of the ship in the AIS message i, v_i represents the speed of the ship in the AIS message i, c_i represents the course of the ship in the the AIS message i, and t_i represents the time when the base station receives the AIS message i.

$$\begin{bmatrix} longitude \\ latitude \\ SOG \\ COG \\ time \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \\ v_1 & v_2 & \cdots & v_n \\ c_1 & c_2 & \cdots & c_n \\ t_1 & t_2 & \cdots & t_n \end{bmatrix} \tag{1}$$

Because the transmission frequency of AIS message is related to the ship's own navigation speed, the frequency of data transmission of ships at different speed is different, and the frequency of the same ship at different speed is also different, so the time interval between AIS messages in Formula (1) is different. The time interval and speed between the two adjacent messages directly determine the displacement of the ship in this period. Therefore, it is impossible to directly use the dynamic data such as ship position and time in AIS data to reflect the movement of ships. In this paper, the displacement $\boldsymbol{R}$ and time difference Δt of the ship between two adjacent AIS data are used to represent the motion of the ship.

$$\mathbf{R}_i = \begin{bmatrix} x_i - x_{i-1} & y_i - y_{i-1} \end{bmatrix} \tag{2}$$

$$\Delta t_i = t_i - t_{i-1} \tag{3}$$

Where the displacement $\mathbf{R}_i$ is the vector. In order to reflect the magnitude and direction of the displacement vector at the same time, the module $|\mathbf{R}_i|$ of the displacement vector, the sine and cosine values of the angle between the displacement vector $\mathbf{R}_i$ and the x-axis, $cos <\mathbf{R}_i, \mathbf{i}>$ and $sin <\mathbf{R}_i, \mathbf{i}>$ are used to represent the displacement $\mathbf{R}_i$.

The average speed v_m of the ship between two adjacent AIS messages is used to express the ship's moving speed in this period of time.

$$v_{m-i} = \frac{v_{i-1} + v_i}{2} \tag{4}$$

The cosine value $cos(c_i)$ and sine value $sin(c_i)$ of the ship's course are used to express the ship's course in this period of time. Then n pieces of AIS dynamic data of the ship in a continuous period of time can be expressed as :

$$\begin{bmatrix} |\mathbf{R}_i| \\ \cos\langle \mathbf{R}_i, \mathbf{i} \rangle \\ \sin\langle \mathbf{R}_i, \mathbf{i} \rangle \\ \frac{v_{i-1}+v_i}{2} \\ \cos\langle c_i \rangle \\ \sin\langle c_i \rangle \\ \Delta t_i \end{bmatrix} = \begin{bmatrix} |\mathbf{R}_2| & |\mathbf{R}_3| & \cdots & |\mathbf{R}_n| \\ \cos\langle \mathbf{R}_2, \mathbf{i} \rangle & \cos\langle \mathbf{R}_3, \mathbf{i} \rangle & \cdots & \cos\langle \mathbf{R}_n, \mathbf{i} \rangle \\ \sin\langle \mathbf{R}_2, \mathbf{i} \rangle & \sin\langle \mathbf{R}_3, \mathbf{i} \rangle & \cdots & \sin\langle \mathbf{R}_n, \mathbf{i} \rangle \\ \frac{v_{2-1}+v_1}{2} & \frac{v_{3-1}+v_3}{2} & \cdots & \frac{v_{n-1}+v_n}{2} \\ \cos\langle c_2 \rangle & \cos\langle c_3 \rangle & \cdots & \cos\langle c_n \rangle \\ \sin\langle c_2 \rangle & \sin\langle c_3 \rangle & \cdots & \sin\langle c_n \rangle \\ \Delta t_2 & \Delta t_3 & \cdots & \Delta t_n \end{bmatrix} \tag{5}$$

When all AIS dynamic data are normal, each element in the matrix shown in Formula (5) satisfies the ship's motion law, for example, the distance $\mathbf{R}_i$ is equal to the product of average speed $\frac{v_{i-1}+v_i}{2}$ and time interval Δt_i. If one or more data are abnormal during this period, the elements in the matrix shown in Formula (5) must not conform to the ship motion law in one aspect. Therefore, this paper constructs the neural network model shown in Figure 1 to identify the abnormal AIS data, and takes the AIS data in Formula (5) as the input of the network model, and the AIS dynamic data discrimination result as the output. 1 indicates that this group of AIS data is normal; 0 indicates that this group of AIS data is abnormal.

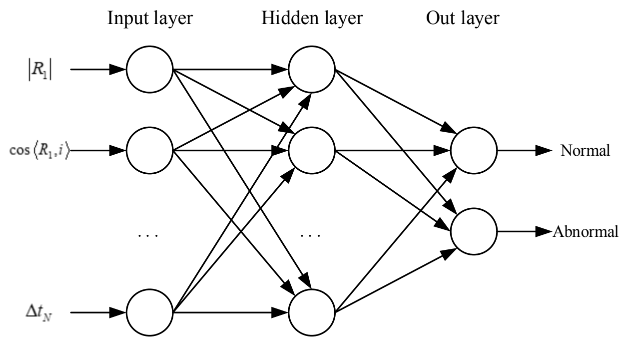


Figure 1: The model of AIS data screening based on Neural Network

IV. Reliability Verification of Abnormal AIS Data Screening Model

A. Experimental data

2056 pieces of AIS dynamic data near Chengshantou Water Area are selected as the experimental data, the navigation speed of the ship is about 7 knots, and the maximum navigation speed is 9 knots. After manual screening and statistics, it is found that the collected AIS dynamic data are normal data, and there is no abnormal AIS data. The training data set of network model is selected by data_size in turn, and 2056 - data_size + 1 groups of the normal AIS experimental data are obtained.

Because the abnormal AIS dynamic data can be generally divided into three types: position anomaly, velocity anomaly and heading anomaly[9], and a large number of abnormal AIS dynamic data is difficult to obtain. In this paper, the abnormal AIS dynamic data is obtained by adding random noise to the normal AIS data set. For the abnormal AIS training data, in addition to adding distance noise, this paper also adds speed noise and heading noise. After the above processing, a total of 2056 - data + 1 groups of abnormal AIS experimental data were obtained, and a total of (2056 - data + 1) × 2 groups of normal AIS experimental data and abnormal AIS experimental data were obtained.

B. Experimental results

In order to evaluate the effect of abnormal AIS data screening, this paper uses the accuracy rate, false alarm rate and missing alarm rate as the evaluation criteria, and the calculation Formula (6)-(8) is shown in.

$$R = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{6}$$

$$FPR = \frac{FP}{TN+FP} \times 100\% \tag{7}$$

$$FNR = \frac{FN}{TP+FN} \times 100\% \tag{8}$$

Among them, R represents the accuracy of AIS dynamic data screening; FPR represents the false alarm rate; FNR represents the missing alarm rate; TP (True Positive) represents the test sample is abnormal and the test result is abnormal; TN (True Negative) represents the test sample is normal and the test result is normal; FP (False Positive) represents the test sample is normal and the test result is abnormal; FN (False Negative) indicates that the test sample is abnormal data and the test result is normal.

1) Length of AIS data segment data_size

In AIS data screening, the selected AIS data segment length, that is, the value of data_size, will produce a larger image for the accuracy of screening. In order to find out the influence of data segment length on abnormal AIS data, this paper compares the recognition accuracy of different data segment lengths.

The neural network model used to identify the abnormal AIS dynamic data consists of three layers: input layer, hidden layer and output layer. The number of input layer nodes is 7 × (data_size -1), and the input amount of each node is as shown in Formula (1); the number of output layer nodes is 2, which respectively represents the normal and abnormal AIS data in this section; the number of hidden layer nodes is calculated by Formula (9).

$$N_h = \left[\sqrt{N_i + N_o} + a\right] \tag{9}$$

Where, N_h is the number of hidden layer nodes; N_i is the number of input layer nodes; N_o is the number of output layer nodes; a is a constant in [0 10], in this paper, a = 2. During the experiment, the number of input layer nodes and hidden layer nodes are shown in table I.

Table I: The Number of Neural Network Nodes Under Different DATA_SIZE

data_size	2	3	4	5	6	7
N_i	7	14	21	28	35	42
N_h	5	6	7	7	8	9

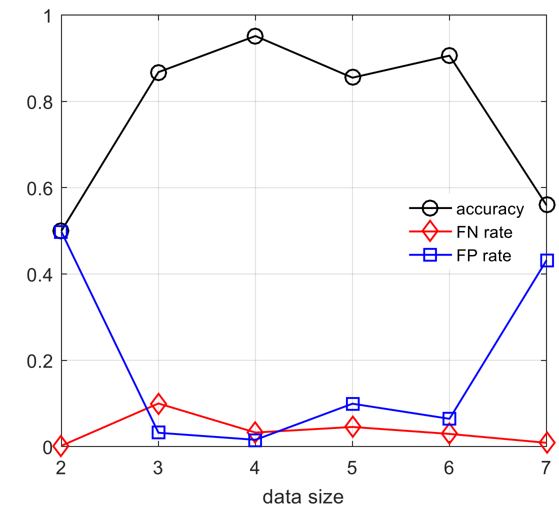


Figure 2: The accuracy rate(accuracy), false alarm rate (FP) and missing alarm rate (FN) of AIS data screening under different data_size

The experimental data of (2056 - data_size + 1) × 2 groups were divided into two parts. 1000 groups were randomly selected as training data and the rest as test data. During neural network training, the algorithm in this paper is based on the deep learning toolbox of MATLAB platform, in which epochs = 3, batchsize = 100. After the training, the neural network model is used to screen the abnormal data in the test sample. Under different data_size, the screening accuracy, false alarm rate and missing alarm rate are shown in Figure 2 .

For accuracy, when data_size is less than 4, the accuracy of data screening increases gradually with the increase of data_size; when data_size is 4, that is 4 consecutive AIS data are selected for screening, the accuracy of AIS dynamic data screening is the highest, reaching 95.16%; when data_size is greater than 4, the accuracy of data screening decreases gradually with the increase of data_size. This shows that in the AIS dynamic data screening, the length of data is not the longer the better. For the data selected in this paper, it is reasonable to select 4 consecutive AIS data for identification and judgment.

The trend of false alarm rate (FP) is just the opposite of accuracy rate with the change of data_size. When data_size is less than 4, the false alarm rate of data screening decreases gradually with the increase of data_size. When data_size is 4, the false alarm rate of data screening is the lowest, reaching 1.6%. When data_size is larger than 4, the false alarm rate of data screening increases gradually with the increase of data_size.

As for the missing alarm rate (FN), with the increase of data_size, the missing alarm rate decreases gradually. When data_size is 4, the missing alarm rate is 3.26%.

Generally speaking, for AIS data screening model based on neural network, the length of data segment is not the longer the better. When the data_size is 4, the accuracy rate, false alarm rate and missing alarm rate are all better.

2) Number of neural network nodes N_h

In addition to the length of AIS data segment, the structure of neural network itself will also have a greater impact on the screening results. In order to find out the influence of the number of nodes in different hidden layers N_h on the result of AIS data screening, we use different number of nodes to build an abnormal AIS data screening model and carry out AIS data screening experiments. In this paper, the screening model with the number of hidden layer nodes N_h between 2 and 7 is verified by experiments. In the process of verification, the data_size is taken as 4.

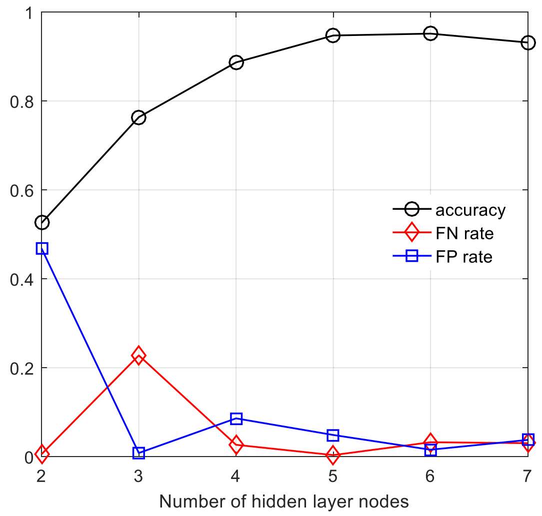


Figure 3: The accuracy rate, false alarm rate and missing alarm rate of AIS data screening with different number of hidden layer nodes

Make data_size = 4, divide the experimental data of (2056 – data_size + 1) × 2 groups into two parts, randomly select 1000 groups as training data and the rest as test data. The algorithm of this paper is based on the deep learning toolbox of MATLAB platform, in which epochs = 3, batchsize = 100. After the training, the neural network model is used to screen the abnormal data in the test sample. The accuracy, false alarm rate (FP) and missing alarm rate (FN) of AIS data screening under different N_h conditions are shown in Figure 3. When the number of hidden layer nodes is 6, the accuracy rate is 95.16%, the false alarm rate is 1.58%, and the missing alarm rate is 3.26%.

3) Analysis of experimental results

The following conclusions can be drawn from the analysis and discussion of the above two experiments:

(1)The model of AIS data identification constructed in this paper can effectively identify the abnormal dynamic data. The accuracy rate of identification is 95.16%, the false alarm rate is as low as 1.58%, and the missing alarm rate is as low as 3.26%.

(2)In the abnormal AIS data screening model, the value of data_size has a great influence on the accuracy of screening, and the larger or smaller will reduce the accuracy of screening. The main reason is that the smaller data_size can not accurately describe the characteristics of ship's accelerating, decelerating, turning to the left, turning to the right, direct navigation and other navigation behaviors; the larger data_size covers the characteristics of ship's multiple navigation behaviors due to its long duration, and the relationship between adjacent AIS message data is complex and the correlation is not clear, so the screening model will have many normal AIS messages data is regarded as abnormal data, which is also the main reason why the false alarm rate increases sharply with the increase of data_size.

(3)The number of hidden layer nodes N_h will also affect the accuracy of the screening model. When the number of nodes in the hidden layer is small, the neural network model appears under fitting, unable to describe the behavior characteristics of the ship, so the accuracy rate of screening is low; when the number of nodes is large, the neural network model appears over fitting, although the model has a high screening accuracy rate for the training sample data, the accuracy rate of other data is low.

Generally speaking, for the neural network model constructed in this paper, 4 adjacent AIS data are selected, and 6 hidden layer nodes are set, which can achieve a high accuracy rate in identifying the abnormal AIS data.

V. Example Analysis

In order to verify the application scope of AIS data screening model, this paper uses the above training results to screen AIS data of ships sailing in Bohai Bay waters.

A. AIS data preprocessing

The preprocessing of AIS data in the database is as follows:

(1) Sort all AIS data with speed over ground (SOG) greater than 3 knots according to MMSI number.

(2) The AIS data with the same MMSI number are sorted according to the increasing time to get the *Track* of each

ship.

(3)Calculate the time difference between two adjacent AIS data in the ship *Track*. According to the performance standard of AIS data, the time difference between two adjacent AIS data will not exceed 6 minite[9] . Therefore, if the time difference Δt between two adjacent AIS data is greater than 360 second, then the ship *Track* is divided into two *sub-Tracks*.

(4)Select the *sub-Track* with the number of ship position points greater than 10. Because the abnormal AIS data screening model based on neural network needs to be based on a section of AIS data for data screening, the *sub-Track* which is too short can not reflect the overall motion of the ship. In the experiment, AIS data corresponding to *sub-Track* of ship with more than 10 ship position points are selected.

After the above preprocessing, 37956 pieces of AIS data were obtained.

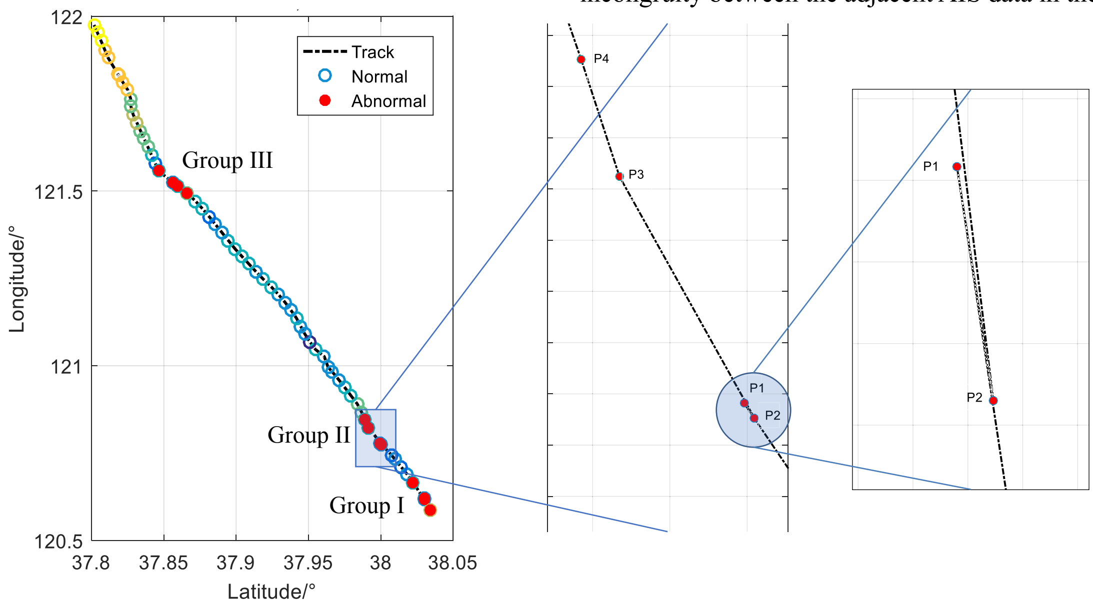
B. Screening results of abnormal AIS data

The AIS data in the preprocessing results are screened by the screening model. Each data_size pieces of data is taken as a group. In this experiment, the data_size is taken as 4, and the number of hidden layer nodes N_h is taken as 6, that is, the abnormal AIS data screening model is used to screen 37956 / 4 = 9489 groups of data. Among the screening results, 9155 were normal AIS data, 334 were abnormal AIS data, accounting for 3.6%. Through manual comparison, among the abnormal AIS data in the screening result, there are 27 groups of AIS data that are actually normal, as shown in table II.

TABLE II: SCREENING RESULT STATISTICS

Abnormal AIS data of screening results	Actual state		Proportion
334 groups	Normal AIS data	27	8.1%
	Abnormal AIS data	307	91.9%

The abnormal AIS data in the screening results can be divided into the following two categories: one is the AIS data group abnormality caused by the entanglement of ship's trajectory, and the other is the abnormality caused by the incongruity between the adjacent AIS data in the AIS data.



(a)*Track-A* (b)Part 1 (c)Part 2

Figure 4: Screening results of abnormal AIS data

1) Trajectory entanglement

According to the time sequence, the ship trajectory obtained by connecting the ship positions in AIS data may be entangled, as shown in Figure 4. In the figure, the red solid dot represents the abnormal AIS data, the hollow dot represents the normal AIS data, and the dotted line represents the *Track-A* of the ship. In the figure, the group II abnormal AIS data of *Track-A* is enlarged and displayed. As can be seen from the figure, the ship's trajectory is entangled at points P_1 and P_2. Table III shows the AIS data of ship *Track-A*, including time(TIME), course(COURSE), speed over ground(SOG), longitude(LON) and latitude(LAT). Where, the TIME column indicates the time interval between AIS data *i* and AIS data *i*-1 within this group of AIS data. It can be seen from the table that the ship's heading is stable at about 100 °, under normal circumstances, the longitude LON in AIS data will gradually increase, and the latitude LAT will gradually decrease. However, comparing the longitude and latitude of Group II in table III, it is found that *Track-A* decreases in longitude and increases in latitude at point (120.7753, 38.0005).

TABLE III:ABNORMAL AIS DATA GROUPING OF *TRACK-A*

Goup of Abnormal	TIME/s	COG/°	SOG/knot	LON/°	LAT/°
Group I	----	98	13.2	120.618	38.0302
	39	99.6	13.2	120.6209	38.0299
	310	102.5	13.5	120.5876	38.0345
	290	102.4	13.3	120.6662	38.0224
Group II	-----	106.6	13.2	120.7782	37.9998
	291	106	13.2	120.7753	38.0005
	299	99.4	13.3	120.8224	37.9917
	301	98	13.3	120.8453	37.9893

There are two main reasons for the entanglement of ship's trajectory: one is the large positioning error of GPS at the ship's end, which causes the ship's position in the AIS data to deviate from the real ship's position seriously; the other is the errors in the transmission or parsing of AIS data, which results in the large errors in TIME of the AIS data.

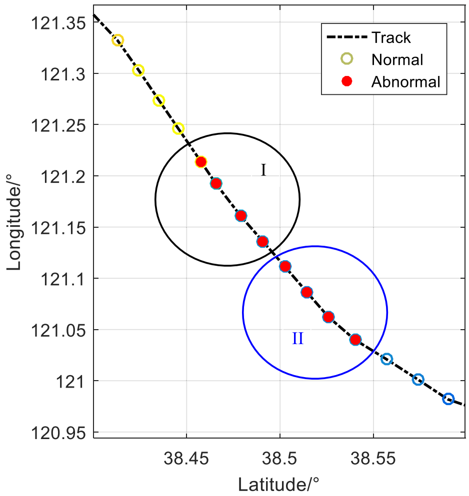


Figure 5: Screening results of abnormal AIS data of *Track-B*

2) Inconsistency between adjacent AIS data

Different from the abnormal AIS data of track entanglement, it is impossible to find the inconsistency between the adjacent AIS data simply according to the track of the ship. For example, in the *Track-B* track in Figure 5, the ship position points are evenly distributed and the lines are smooth. The reason why this algorithm regards it as abnormal AIS data is that the time interval between AIS data in group I and group II data in *Track-B* is almost the same, and the track point spacing is the same, but the speed (18.5 knot) of the first point in group I and the direction (318 °) of the fourth point in group II are obviously different from other three data, See table IV for abnormal track data.

TABLE IV: ABNORMAL AIS DATA OF *TRACK-B*

Goup of Abnormal	TIME/s	COG/°	SOG/knot	LON/°	LAT/°
Group I	----	295.4	18.5	121.2139	38.4577
	299	295.4	16.7	121.1925	38.4656
	301	301.4	16.4	121.1606	38.479
	300	301.4	16.3	121.1358	38.4909
Group II	----	301.3	16.2	121.1112	38.5026
	301	301.4	16.1	121.0867	38.5143
	300	301.4	16.2	121.0622	38.5259
	300	318.8	16.3	121.0399	38.5402

From the screening results of *Track-B*, we can see that the algorithm in this paper can identify the inconsistency between adjacent AIS data, even if the degree of the inconsistency is small. It should be noted that in some engineering applications, the slight inconsistency between the adjacent AIS data may not affect the normal use of the data, and it is not necessary to treat it as an abnormal AIS data.

C. Result analysis

Figure 6 shows the distribution of normal AIS data and abnormal AIS data near the Bohai Bay. It can be seen from the figure that the more intensive the area is, the more likely the number of abnormal AIS data will appear, and the more abnormal AIS data near the port will obviously increase.

From the statistical results, it can be seen that the probability of abnormal AIS data is high when the ship speed is low. This does not mean that AIS data is more prone to errors in the process of transmission and analysis when the ship speed is low. The main reason is that the algorithm in this paper has a large error rate when screening AIS data of ships with low speed. Because the abnormal AIS data screening model based on neural network describes the movement characteristics of the ship in a period of time, when the ship speed is low, due to the influence of random error, the correlation between the data points is small, the movement characteristics are not obvious, resulting in the error of the screening model. In addition, when the ship speed is low, the frequency of AIS sending data is low, which will also lead to the error of screening model.

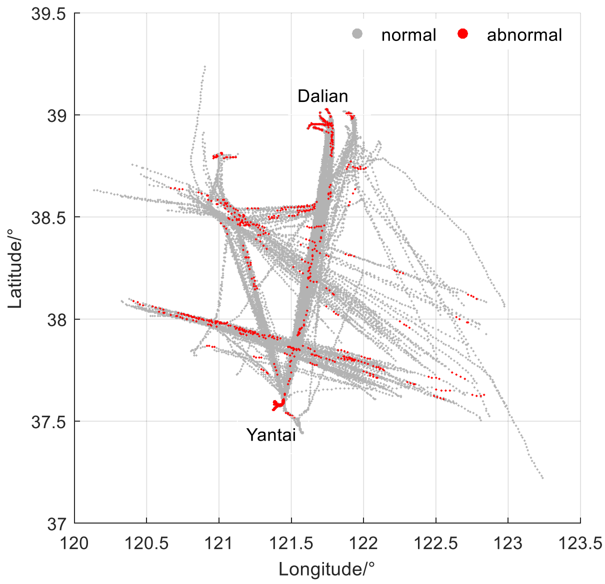


Figure 6: Distribution of abnormal AIS data (Bohai Sea)

VI. CONCLUSION

In the abnormal AIS data screening model based on neural network, the length of AIS data segment *data_size* and the number of hidden layer nodes N_h will have a greater impact on the accuracy. In order to find out the influence of these two factors on the screening effect, this paper conducts experiments on different *data_size* and N_h models. The experimental results show that:

(1) If the *data_size* is too large or too small, the accuracy of the screening model will be reduced.

(2) If N_h is too small, the screening model will appear under fitting; if N_h is too large, the screening model will appear over fitting, both of which will reduce the accuracy of the screening model.

(3) In the screening model, when *data_size* = 4, N_h = 6, the accuracy is the highest.

(4) The screening model can effectively identify the track entanglement and the inconsistency between adjacent

AIS data.

(5) The screening model has a large error for AIS data with low ship speed.

REFERENCES

[1] Giuliana P, Michele V, Karna B. Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction[J]. Entropy, 2013, 15(12): 2218–2245.

[2] Talavera A, Aguasca R, Galván B, et al. Application of Dempster–Shafer theory for the quantification and propagation of the uncertainty caused by the use of AIS data[J]. Reliability Engineering & System Safety, 2013, 111(1): 95–105.

[3] Ma Feng,Chu Xiu-min,Yan Xin-ping. Short message characteristics of AIS base station [J]. Journal of Traffic and Transportation Engineering, 2010, 12(6): 111‑118.

[4] Zhu Fei-xiang, Zhang Ying-jun, Gao Zong-jiang, et al. Research on ship behavior based on Data Mining [J]. Navigation of China, China Institute of Navigation, 2012(02): 50‑54.

[5] Wei Gui-rong,Yang Chun. The method of AIS data error detection [J]. Navigation of China, 2016, 39(4): 11‑14.

[6] Liu Xing-long, Chu Xiu-min, Ma Feng, et al. Research on the distance between the virtual navigation mark base stations in the mountainous waterway [J]. Journal of Harbin Engineering University, 2016(03): 90–95.

[7] Zhen Rong, Jin Yong-xing, Hu Qin-you, et al. Prediction of ship navigation behavior based on AIS information and BP neural network [J]. Navigation of China, 2017(2): 6-10.

[8] Ma Feng, Yan Xin-ping, Chu Xiu-min, et al. Correlation between signal failure and field strength of ship automatic identification system [J]. Journal of Dalian Maritime University, 2011, 37(3): 111–114.

[9] SANG L, WALL A, MAO Z, et al. A novel method for restoring the trajectory of the inland waterway ship by using AIS data[J]. Ocean Engineering, Elsevier, 2015, 110(1): 183–194.

[10] Liu Xing-long, Chu Xiu-min, Ma Feng, et al. AIS ynamic information screening method for abnormal messages [J]. Journal of Traffic and Transportation Engineering, 2016, 16(5): 142–150.

[11] Nie Yang, Chu Xiu-min, Liu Xing-long, et al. Comparison of analysis methods for validity of inland river AIS data [J]. Navigation of China, China Institute of Navigation, 2016(02): 59‑62.