SPECIAL SECTION ON SMART CACHING, COMMUNICATIONS, COMPUTING AND CYBERSECURITY FOR INFORMATION-CENTRIC INTERNET OF THINGS

Received November 12, 2018, accepted November 21, 2018, date of publication November 27, 2018, date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883421

Mining Channel Water Depth Information From IoT-Based Big Automated Identification System Data for Safe Waterway Navigation

ZHENGWEI HE[1,2,3], FAN YANG[1], ZHONG LI[1], KEZHONG LIU[1,2,3], AND NAIXUE XIONG[4,5]

[1]School of Navigation, Wuhan University of Technology, Wuhan 430063, China
[2]Key Laboratory of Hubei Inland Shipping Technology, Wuhan 430063, China
[3]National Engineering Research Center for Water Transport Safety, Wuhan 430063, China
[4]College of Intelligence and Computing, Tianjin University, Tianjin, China, 300350
[5]Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK 74464, USA

Corresponding author: Fan Yang (yfan@whut.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 51479157, in part by the Fundamental Research Funds for the Central Universities of China under Grant WUT:2018ćó064GX, in part by the Fundamental Research Funds for the Central Universities of China under Grant WUT:2018-zy-127, and in part by the Fund of Hubei Key Laboratory of Inland Shipping Technology of China under Grant NHHY2017001.

ABSTRACT Internet of Things technology has been widely used in water traffic research. Many critical waterways in the world are becoming more crowded due to many factors in waterway environments, such as invisibility, variability, and uncertainty. Accurate water depth information is necessary to improve navigation safety. Water depth information of electronic charts cannot be updated in a timely way, while the actual water depth is unpredictable, and this factor threatens the safety of vessels in waterway environments. Based on the shore-based network, ship navigation data and other big data can be integrated to vessels navigation environments in real time. In this paper, we present a new scheme to quickly and accurately construct a vessel safety navigation depth reference map, which contains appropriate channel water depth information. This effective scheme is based on automated identification system (AIS) data and increases the travel safety through crowded waterways. AIS data include rich maritime traffic information. Both the static and the dynamic information about vessels through waterways can be extracted and processed from big real-time AIS data. Based on extensive actual experiments, we apply data mining techniques to extract the waterway depth information both draft-depth and vessel trajectories based on AIS data. The data are collected from vessels in both locations: 1) the Nantong port, in Jiangsu Province, China and 2) Meizhou Bay waterway, in Fujian Province, China. The Hermite interpolation scheme is used to patch the trajectories of vessels, and the BP neural network model is introduced to predict the maximum vessel draft. Clustering and data fusion methods are employed to construct a vessel safety navigation depth reference map according to the cluster area of vessel trajectories and draft information. The experimental results demonstrate that the vessel safety navigation depth reference map accurately reflects the current water depth profile of channels. This paper can provide accurate and timely channel water-depth information for the vessel navigation and the maritime supervision. The proposed scheme in this paper can also provide reference for trajectory data processing and mining.

INDEX TERMS Big AIS Data, IoT, BP neural network, clustering and data fusion, navigation safety, vessel trajectory.

I. INTRODUCTION

Waterway traffic is a critical part of the Internet of Things. A large number of shipboard sensors can provide real-time navigation information for ships to ensure safe navigation. Currently, ships can transmit vessel navigation information to shore-based networks in real time through Very High Frequency (VHF) radio stations and Automatic Identification System (AIS). If we make full use of the large amount of ship-based integrated navigation data, we can analyze the waterway traffic status online and provide real-time navigation safety guidance for safe navigation [1].

The shipping industry is the backbone of the global economy. In global trade, most goods are transported over water, and more than 90% of the goods are waterborne in the Chinese import and export trade. Waterway traffic is developing rapidly, making water traffic accidents more commonplace, with massive loss of life and property. In China, grounding accidents accounted for about 10% of all types of accidents on the Yangtze River. A large number of these accidents are related to insufficient water depth information.

Accurate water depth measurement and mapping is a requirement for safe waterway navigation. Siltation, sunken vessels, and changes in the underwater terrain, however make waterway depths variable and unpredictable. Channel and underwater topographic surveys are complex and expensive tasks, making it difficult to correct and update depth information in a timely way. Inaccurate charts and maps contribute to accidents such as grounding and collision, and threaten navigation safety.

Depth information for shipping areas is obtained from electronic charts and vessel logs. The water depth information contained in electronic charts is updated at a certain periodicity, closely related to the plan for channel measurement and updating electronic chart updates. Therefore, electronic charts do not provide current water depth information. Vessel logs can only measure the water depth information under the current vessel position, and are not suitable for pre planning routes that avoid in advance shallow areas, related to grounding accidents. At present, there is no effective and timely way to obtain water depth information for safe vessel navigation, this study addresses this problem using big Automatic Identification System (AIS) data and machine learning techniques.

AIS data contains rich vessel traffic information in real-time. This data is aggregated and sent through coastal base stations and contains a wealth of ship information. If water depth data for channels can be obtained and excavated through AIS data then water depth information can be extracted from vessel traffic data in real time. If a vessel has safely passed through a certain location of a channel, then vessels of the same type and draft should be able to pass safely through the same location and area. Among them, the ship water depth information by looking for the relevant strong captain, the breadth of the ship, draft information to predict access. Information about the safe navigation area of the ship is obtained from the trajectory of the ship through the method of convex hull. By establishing safe navigation zones for ships with different draft requirements, it is possible to determine the safe sailing water depth range. Water-depth information for a channel can be extracted from the historical vessel tracks found in AIS point data.

Trajectory data is a type of spatiotemporal data that records the movement of an object by positional sampling. According to the sparseness of positional sampling, trajectory data can be divided into two categories: sparse sampling and dense sampling. With the development of satellites, wireless networks, and positioning devices, trajectory data of a large number of moving objects is increasing rapidly, such as animal migration data, climate airflow data and personnel movement data [2], [3]. The study of trajectory data and the acquisition of unknown knowledge about motion of objects will become a future research hotspot.

In our research, we extracted channel water depth information from big AIS data, and then constructed Vessel Safe Navigation Depth Reference Map, which provides the channel water-depth information for different types of safe navigation. The vessel safe navigation depth reference map can provide timely reference information for safe vessel navigation and dynamic maritime supervision.

The rest of this paper is as follows, the second section reviews the literature on water depth and traffic flow prediction, and the third section introduces AIS data. The fourth section introduces the research methods in this paper and the proposed method for constructing a safe navigation, depth reference map. The fifth section presents prediction results for water depth and traffic flow in the Meizhou Bay and in the port of Nantong city, in China. The sixth chapter draws conclusions.

II. RELATED WORK

Several approaches have been developed for improving navigation safety and enhancing the real-time data processing ability in electronic charts. Methods of determining safe nautical depths and effectively deploying e-Navigation are the most relevant to water-depth information mining and electronic reference map construction. Mcanally *et al.* [4] developed a nautical depth concept suitable for selected ports and waterways to reduce the need for the frequency and volume of dredging. However, this method uses a small amount of data and requires intensive coordinated efforts among national, regional, and local agencies. The nautical depth concept and data processing method is applicable to efforts towards mapping of fast changing channels in rivers and estuaries, essential to safe navigation. Weintrit introduced the e-Navigation concept and its definition, proposing the use of e-Navigation for water-depth safety requirements and safe navigation standards. E-Navigation can meet certain safe navigation challenges in a general way; however, the specific means for e-Navigation are inefficient and cannot address actual navigation problems such as prediction of real-time safe water depths and navigation conditions [5]. AIS data contains rich information, and thus a valuable resource for research. Big AIS data could be more effectively exploited to extract dynamic and static data from vessels to map safe navigation areas in a timely way.

Many different methods make use of AIS data. Among the AIS applications deploying data mining methods, the most widely used are analyses of AIS data to extract and model vessel motion trajectories and to establish ship-monitoring systems. Bomberger *et al.* [6] used AIS data and a neural network to extract and learn the pattern of regularities in the trajectories of vessels. This method supports the construction of historical vessel trajectories. Jalkanen *et al.* [7] proposed an algorithm to draw vessel trajectories based on an analysis of AIS data. This method is efficient and effective, but is not based on real-time AIS data with an off-line processing environment making it unsuitable for real time mapping. Focusing on analysis of vessel trajectories, Ulvila and Gaffney, Jr. *et al.* [8] used a nonlinear regression analysis algorithm to cluster AIS data and identify trajectory patterns, but this algorithm is complex and inefficient, and takes a lot of time to preprocess and cluster the AIS data. There are many studies on ship monitoring systems based on AIS data. Høye *et al.* [9] proposed a method for a global vessel supervision system based on AIS to boost the tracking ability for maritime navigation. This supervision system can boost the supervision ability and help the maritime sector to make decisions, but the system needs many sensors, and the data-mining algorithm is inefficient. A detailed case study of a Dutch and a Chinese supervision system based on AIS data analysis explores how lateral position, speed, heading and interval times for different types and sizes of vessels are treated in the two cases [10]. A supervision system can be used for safety assessment of ship traffic in a waterway, or for improving safety and efficiency, but the supervised area of the system is limited, and the installation and setting of the sensors is complex and cannot be monitored in a wide range and in a real time way.

The study reviewed that the method of e-navigation can solve the safety problem in ship navigation; however, the work process is inefficient and cannot make a reasonable response to the complex situation of maritime traffic. By mining the AIS data and building a maritime surveillance system, it is possible to make full use of the information of ship navigation. However, the lack of support of the corresponding electronic chart and other technologies cannot demonstrate the safety of navigation in a real time way. This paper combines these two methods and presents a novel method to extract the channel water-depth information from the real-time and historical trajectory data from vessels, based on big AIS data and analyze the characteristics of navigation trajectories about different types of vessels, and then generate the safe navigation, depth reference map for different vessels. By mining a large amount of AIS data, this new method can define quickly and accurately a safe navigation area for different types of ships.

III. AIS DATA

AIS was developed and deployed to improve navigation and ensure the safety of property and the lives of the crew [11]–[13]. An AIS has four goals, the first is to identify the vessel effectively, the second is to help maritime management identify unusual traffic patterns, the third simplifies vessel-vessel and vessel-shore information exchange, and the fourth goal is stopping vessel collision accidents. AIS is composed of onboard ship platform, onshore base station and communication link, and establishes a ship-shore, ship-ship information interaction system. The AIS system can be used to understand the ship traffic situation in the segment and take timely measures to prevent the danger of the ship and improve maritime affairs regulatory capacity and efficiency of the department.

AIS data contains static and dynamic data. The static data includes International Maritime Organization (IMO) number, vessel's name, vessel's Maritime Mobile Service Identity (MMSI) number, vessel length and vessel breadth, draft depth, type of vessel and other information. The dynamic data contains information such as longitude, latitude, navigation time, the speed, and heading. AIS data are constantly updated and sent to a base station. The dynamic location data enables the tracking and supervision of vessels in real time to reduce accidents.

There are many errors in AIS data. Taking the existing dynamic data for example, the waterway in Nantong China which are between the latitude and longitude from(longitude 120.725 °, latitude 32.049 °) to (longitude 120.835 °, latitude 31.981 °), the time of data is from 9:00 to 09:30 on February 6, 2017, among 170900 records 4953 records had negative latitude and longitude, while 103 MMSI numbers were not nine digits. There was only 2025 vessel data useful in 13330 dynamic data and the useful data only accounted for 15.3% for the dynamic data.

IV. METHODS

As shown in Figure 1, the method can be divided into two steps, but the data must be cleaned to remove abnormal and redundant data from the original AIS messages, seen at the top left of the flowchart. After preprocessing, the first step, illustrated on the left side, is AIS data mining and extraction of ship trajectories and prediction the draft information of ships. The second step, shown on the right, is data extraction and data fusion, linking ship trajectories and draft information.

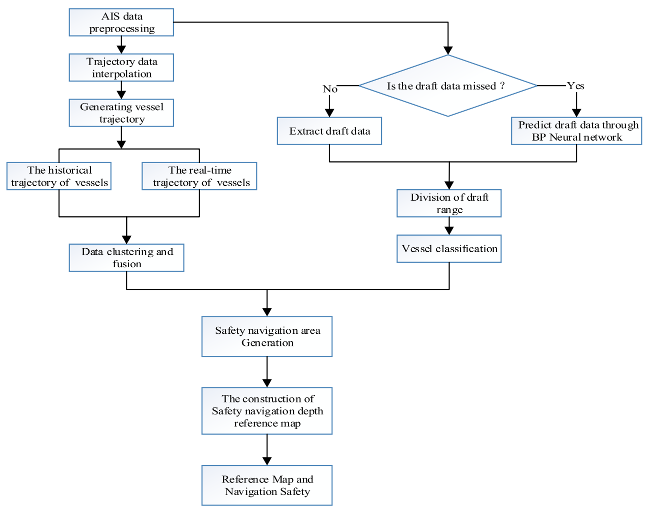


FIGURE 1. A New Method to Construct Real-time Vessel Safe navigation Depth Reference Map.

In the process illustrated in Figure 1, the Hermite method is used to interpolate vessel trajectories generated from AIS data, and data clustering generates the active area of the vessels, as detailed in part C of section IV. Real-time vessel trajectory data, extracted from vessels that recently passed through the channel, are the primary source of information used to create the reference map, as described in part C of section IV. Vessel trajectory and vessel draft data are fused to generate a safe navigation area for a vessel, based on draft data obtained from AIS data, discussed in part B of section IV. As shown in Figure 1, if the draft data is missing, the BP neural network is used to forecast and patch the missed draft data. A real-time, vessel safety navigation depth reference map is constructed from safe navigation areas of vessels, through computer graphics rendering; we discuss this process in part C of section IV. The impact of the ship's sailing depth on the accident of a ship has an important guiding role for the safe navigation of the ship, discussed in part D of section IV.

A. SELECTION OF AIS DATA

Since there are large numbers of vessels in waterways and the time interval required for transmitting AIS data is small, so maritime and waterway management departments cannot easily process the large amount of AIS data. Thus, the data must be classified before processing. There are three methods to classify AIS data. The R language is used to select the data.

Select a specific area, such as a region with a longitude more than x_1 less than x_2 and a latitude more than y_1 less thany_2, bringing the data into compliance with the requirements using the data frame.

ais1<-ais[ais\$Longitude> x_1&ais\$Longitude< x_2 &ais\$Latitude> y_1& ais\$Latitude< y_2,]

AIS data can be filtered out of a data frame called ais1, which meet the requirements of the longitude and latitude.

When select a specific period t_1 to $t_2$of AIS data, using data frame to select the data.

ais2<-ais[ais\$UTC> t1&ais\$UTC> t2,]

B. PREDICTION OF DRAFT-DEPTH DATA

1) VESSEL MAXIMUM DRAFT DEPTH

Although the static data of AIS data includes the draft-depth information, it is not the actual draft-depth of vessel but an inherent attribute. In fact, although the actual draft-depth of the vessel is not necessarily related to the maximum draft-depth of the vessel, a range for the actual draft-depth of a vessel can be extracted by analyzing the maximum draft-depth of the vessel. For marine traffic, the actual draft-depth cannot exceed the maximum draft depth. If the actual draft-depth exceeds the maximum draft of the vessel, it may affect handling performance, and a vessel cannot effectively prevent accidents.

2) PREDICTION OF UNKNOWN MAXIMUM DRAFT DEPTH

Due to errors in the AIS system itself and data transmission problems, there will often be some information loss before receiving and analyzing AIS data. It is necessary to predict the maximum draft of the vessel based on other kinds of vessel information.

BP neural network model is a typical model in neural network technology. BP neural network applies network learning to ascertain the correspondence between input and output parameters. In a network learning process, the output of BP neural network will include error in the target output, and the network will feedback this error to the front layer, constantly adjusting the weights and thresholds between every node [14]–[18]. Because of these features, there is no need to build a mathematical relation between the input and output parameters when training the BP neural network. In general, the structure of a neural network contains three layers: input layer, hidden layer, output layer. A BP neural network is a model connected by a large number of nodes (neurons). Different individual neurons represent different excitation functions. Neurons have different weights, the network learns by training the neurons and changing weights. Figure 2 shows the structure of BP neural network we used.

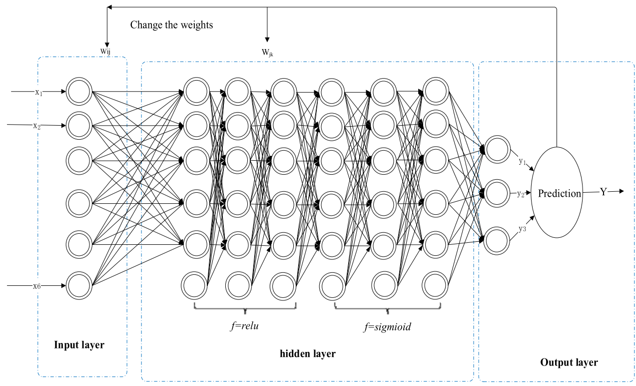


FIGURE 2. Structure of BP Neural Network.

In Figure 2, there are input layer, output layer and six hidden layers. The activation function of the first three is relu while the activation function of last three layers is sigmoid.

To predict the maximum draft depth, it is necessary to determine the parameters of the neural network. This design is based on four types of parameters to set the neural network.

Stochastic gradient descent algorithm can be used to find the optimal network parameters, making the network perform at the optimal level.

After training the AIS data, the range of parameters of the network can be expressed as;

$$\begin{aligned} f &\in \{\tanh(x), sigmoid(x), \max(x,0)\} \\ \lambda &\in \left\{10^{-4}, 10^{-2}\right\} \\ n &\in \{1, 15\} \\ N_l &\in \{1, 200\}^n \\ x^m_{a+h,b+h} &= (f_1 \cdot f_2 \cdot f_n)x^m \\ f_l &= f(w_l^T x_l + b_l) \end{aligned} \tag{1}$$

In (1), f is the activation function, λ is the learning rate, and n is the network layer number, N_l is the number of iterations, $x^m_{a+h,b+h}$ is the structure of the draft-depth prediction model.

According to vessel width, and the types of vessel based on the existing AIS data, it is necessary to see if there is correlation between these data. The duplicate records were deleted, and a total of static AIS data points for 2025 vessels were obtained. In this paper, static information for 1675 vessels was used for correlation analysis.

The BP neural network uses the parameters, vessel length, and vessel breadth to predict the draft-depth of the vessel; therefore, the vessel maximum draft-depth value was selected as the neural network output parameter.

In theory, only one hidden layer is required in a neural network when processing maps, but accuracy is low when using only one hidden layer. This can be corrected by increasing the number of hidden layers. However, many hidden layers will make the network structure complex, and data processing will consume a great deal of time. If the correlation between parameters is not particularly complex, a single hidden layer neural network is generally preferred.

The number of hidden neurons has an effect on the prediction of BP neural network, if the number of neurons is set too low, it may affect the effectiveness of training. If the number of neurons set too high, it will affect the training time.

The Pearson correlation coefficient is used to analyze the correlation of vessel width, vessel breadth, and the maximum draft depth. This is executed in two stages, first calculating the degree of linear correlation between the vessel width and maximum draft depth, and between the vessel breadth and maximum draft depth, and then between these results. (2) is the formula of the correlation coefficient, r is the coefficient, n is the number of samples, x_i and y_i are the values of samples [19]–[21].

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}} \tag{2}$$

Formula (2) indicates the correlation of x_i and y_i. When the positive correlation between x and y is stronger, the value of r is closer to 1. In the same time, if the negative correlation between x and y is stronger, the value of r is closer to −1.

C. CONSTRUCTING NAVIGATION AREA

1) VESSEL TRAJECTORY SEGMENTATION

Not all the vessel motion points are appropriate for constructing trajectories because vessel AIS information may be lost during the transmission of data from vessels to the base station. If information for vessel motion points is missing for a period, errors will occur when estimating ship trajectories. Therefore, we divide of the vessel trajectories into those with complete and those without complete information. The adjacent latitude and longitude information will be regarded as a different trajectory if the receiving time of the data exceeds T. The time threshold T cannot be set too small, if it is too small, the trajectory of a vessel may be divided into a large number of sub-tracks; short sub-tracks cannot reflect the overall movement of the vessel and navigation characteristics [22]. Therefore, we combine the longitude and latitude with the time of the data was received.

(a) Vessel latitude and longitude points are considered the same vessel's location information only if the MMSI number is also same with the vessel.

(b) If the receive time of data are more than 20 minutes, the trajectory will be divided into two different sub-tracks.

(c) If there are less than two points in a sub-track, it will be considered as useless information and be deleted.

After getting the data of vessel trajectory segmentation, it is found that due to the lack of data in the course of data transmission, the data of vessel segmentations have discontinuity, so the data needs to be processed, the missing point must be interpolated and repair the vessel trajectories.

2) VESSEL TRAJECTORY INTERPOLATION

Since there are missing points in the trajectory of a vessel, the longitude and latitude data must be imputed by trajectory interpolation, making the vessel trajectory more complete.

Hermite interpolation is used to complete the missing points [23], [24]. Hermite interpolation uses values and derivative values of the unknown function to construct an interpolation polynomial. Formula (3) shows the polynomial.

$$a \le x_0 < x_1 \ldots < x_n \le b \tag{3}$$

In (3), a and b are the adjacent time points, and x_j is the point to be interpolated:

$$y_j = f(x_j), \quad m_j = f'(x_j), \ (j = 0, 1, \ldots n) \tag{4}$$

In (4), y_j is the longitude and latitude of the points, $f(x_i)$ is the constructor function for the points to be interpolated, m_j is the derivative of $f(x_i)$ [25]. Formula (5) is the formula for Hermite interpolation.

$$H_{2n+1}(x) = \sum_{j=0}^{n}[y_j\alpha_j(x) + m_j\beta_j(x)]. \tag{5}$$

Hermite interpolation can determine a polynomial of no more than $2n + 1$, and a_{2n+1} in (5) is the polynomial coefficient:

Formula (6)-(8) are the other coefficients in (5).

$$\alpha_j(x) = (1 - 2(x - x_j)\sum_{k=0,k=j}^{n}\frac{1}{x_j - x})l_j^2(x) \tag{6}$$

$$\beta_j(x) = (x_j - x)l_j^2(x) \tag{7}$$

$$l_j(x) = \frac{(x-x_0)\ldots(x-x_{j-1})(x-x_{j+1})\ldots(x-x_n)}{(x_j-x_0)\ldots(x_j-x_{j-1})(x_j-x_{j+1})\ldots(x_j-x_n)} \tag{8}$$

x_j is the point to be interpolated, y_j is the longitude and latitude of the points, and $\alpha_j(x)$, $\beta_j(x)$, and $l_j(x)$ are coefficients used in (5).

After interpolating missing latitude and longitude points, all the points will be connected to temporal periods. Taking the AIS data of the specific vessel whose MMSI is 413769173 for example. Table 1 shows the AIS data fragments of the vessel to be interpolated,

TABLE 1. AIS Data Fragment of a Vessel whose MMSI is 413769173.

Longitude	Latitude	SOG	COG	UTC
114.30	30.58	6.0	33.8	2017/4/23 1:08
114.31	30.59	6.0	31.6	2017/4/23 1:10
114.31	30.59	6.1	34.0	2017/4/23 1:11
114.31	30.59	6.1	32.0	2017/4/23 1:12
114.31	30.59	6.2	33.0	2017/4/23 1:14
114.34	30.624	6.2	48.0	2017/4/23 1:37
114.34	30.62	6.2	50.7	2017/4/23 1:38
114.34	30.62	6.3	50.2	2017/4/23 1:39
114.35	30.62	6.3	50.5	2017/4/23 1:40
114.35	30.63	6.3	50.2	2017/4/23 1:41
114.35	30.63	6.3	52.0	2017/4/23 1:42
114.35	30.63	6.4	52.8	2017/4/23 1:44
114.35	30.63	6.5	51.5	2017/4/23 1:45
114.36	30.63	6.5	53.3	2017/4/23 1:47
114.36	30.63	6.6	50.8	2017/4/23 1:49
114.36	30.64	6.7	54.7	2017/4/23 1:51
114.36	30.64	6.7	55.0	2017/4/23 1:51
114.37	30.64	6.8	56.9	2017/4/23 1:54
114.37	30.64	6.7	58.9	2017/4/23 1:56
114.38	30.64	6.7	61.2	2017/4/23 1:58
114.38	30.65	6.6	63.8	2017/4/23 2:03
114.39	30.65	6.7	63.8	2017/4/23 2:04
114.39	30.65	6.7	66.2	2017/4/23 2:05
114.39	30.65	6.7	66.5	2017/4/23 2:06

3) SAFE NAVIGATION AREA

In order to obtain a safe draft-depth navigational area, the convex hull method is applied to connect the outer edge points to obtain the safe navigation areas for different types of vessels [26]. The convex hull is a convex polygon that joins the outermost points, which contain all the trajectory points. Convex hull is a concept in graphics. If there is a certain type of point set on a plane, the convex hull is a convex polygon formed by connecting the outermost points, which can contain all the points in the point set. Vessel trajectories are all surrounded by a convex hull. Using the convex hull to connect a point on the outer edge of the region, different types of vessel navigation areas can be constructed. The method of constructing vessel water depth using a convex hull is as:

(a) For the same type of vessel, the vessel has different trajectory points $p_1, p_2, \ldots, p_n$.Therefore, the aggregate trajectory P of this type of vessel can be obtained by connecting the trajectory points, and (9) produces the trajectory of a vessel as:

$$P = p_1 \cup p_2 \ldots \cup p_n \tag{9}$$

(b) To calculate the aggregate trajectories of all the vessels of all types in an area, a threshold method is adopted to set the threshold u. If the distance d between the two vessel trajectories is $d > u$, then the historical trajectories between the two trajectories are removed, so that a part of the empty space in the trajectory of all the vessels is generated. This part of a waterway may be isolated island or shoal. A gift-wrapping method is used to construct a convex hull for the aggregated trajectories of all vessels in an area, a point in each trajectory is selected randomly [27]. These points are taken as initial points while the outermost points from all the vessel trajectories are selected in a counterclockwise direction. When the algorithm returns to the initial point, the set of collected points outlines a convex hull for the aggregated trajectories of all the vessels in an area, as illustrated in Figure 3.

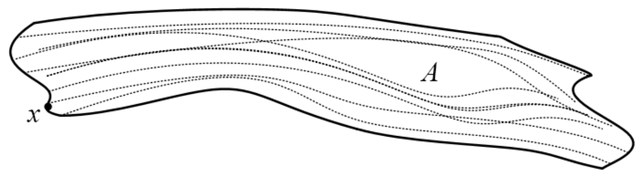


FIGURE 3. Construction of a Convex Hull.

Figure 3 shows a convex hull construction diagram, where x is the initial point selected by the gift wrapping method, and area A is a blank area in the trajectories of all the vessels in the area obtained through our threshold method.

(c) The safe navigation area R of a vessel is expressed by (9), where Convex () is the convex hull function. The hull area of this vessel R, is obtained by convex hull method as

$$R = Convex(P) \tag{10}$$

Formula (10) shows the method of convex hull, the safe navigation area of vessels will be constructed by this method.

4) SAFE NAVIGATION DEPTH REFERENCE MAP

According to the maximum draft-depth of the vessel, the points will be divided into different categories, the trajectory points of the same type of vessel are connected.

$$G_1 = F(R_A); \quad G_1 = F(R_A); \; G_3 = F(R_C); \; G_4 = F(R_D) \tag{11}$$

Formula (11) shows the construction of draft-depth map for each type of vessel, G_i is the map of each type of vessel, $F(*)$ is the function to construct the draft-depth map.

$$G = G_1 \cup G_2 \ldots \cup G_i \ldots \cup G_n \quad (i = 1, 2, 3 \ldots) \tag{12}$$

Formula (12) shows the method of creating a comprehensive draft-depth map by adding each map of the specific vessel, G is the draft map, G_i is the map of each type of vessel.

$$S = G \cup F \tag{13}$$

Formula (13) is to combine the different types of draft-depth maps with waterway electronic chart, S is the safe navigation draft-depth reference map, F is the electronic chart.

D. REFERENCE MAP AND NAVIGATION SAFETY

1) SAFETY ANALYSIS

Accidents caused by insufficient actual navigational water depth include vessels hitting rocks, stranded vessels, and other accidents. Among the accidents, vessel hitting rocks and stranded vessels are the main types of accidents related to insufficient water depth, caused by channel siltation. The stranded vessels disrupt cargo transport but also are a threat to the environment.

The safe navigation depth is composed of the full draft of the vessel and the Under Keel Clearance (UKC). Formula (14) indicates safe navigation of the vessel. D is the vessel safe navigation depth, ΔZ is the vessel UKC. Formula (15) represents the composition of the vessel surplus,

$$D = T + \Delta Z \tag{14}$$

$$\Delta Z = Z_0 + Z_1 + Z_2 + Z_3 \tag{15}$$

where Z_0 is the amount of navigation sinking, related to navigation speed, and channel water depth. Z_1 is the smallest UKC of water depth, related to submarine obstacles, Z_2 is the wave UKC, related to wave factors, and Z_3 is the increased stern draft due to an uneven load. As can be seen from (13), adequate water depth and draft are critical factors for the safe navigation of the ship.

In order to analyze the influence of the water depth when the vessel is stranded, the accidents are analyzed by selecting the appropriate parameters of AIS data. Table 2 shows the sample of data in the dataset from Nantong waterway, we choose 6 vessels with different tonnage as sample data. DWT is the dead weight tonnage of vessel, COG is the course of the vessel, SOG is the speed of the vessel.

TABLE 2. Sample of data.

DWT	COG	SOG	BREADTH	LENGTH	DRAFT
5	113.4	0.78	33.0	229	12.8
10	262.1	0.81	43.0	246	14.8
15	146.6	0.79	50.0	274	17.1
20	109.4	0.76	55.0	321	18.6
25	171.1	0.82	60.0	333	19.8
30	137.1	0.81	60.0	334	22.6

Based on the statistics and analysis of 252 vessel accidents from 2002 to 2015, it can be concluded that when the vessel is stranded, the actual navigational depth of the vessels described in Table 2 do not satisfy the draft requirements of the vessel as, the actual draft is related to the navigation speed.

An analysis of stranded vessel accidents shows that draft conditions are different for varying vessel lengths and breadths, and the data of COG and SOG is the same. For the vessels with the same length and breadth, the drafts are close to the actual depth, and the data of COG and SOG is the same for all vessels of the same type.

Since the water depth in the electronic chart is updated slowly, the safe navigation area for different types of vessels can be separated in the safe navigation depth reference map to meet varying draft-depth requirements for different types of vessels. The depth reference map for safe navigation can guide vessels. Vessels with the greatest draft-depth demands sail near the deep-water areas; while vessels with smaller draft-depth sailing closer to shallow water areas. This will not only ensure the safe navigation and avoid the stranded vessel accidents, but will also guide the speed of the vessels, thus reducing occurrences of waterway congestion to improve the effective utilization of waterways with safer navigation practices.

2) VESSEL RISK ASSESSMENT UNDER DIFFERENT WATER DEPTHS

More attention must focus on accidents related to water depth conditions that have a greatest impact such as strandings or collisions with rocks. When the actual water depth surrounding a vessel is too small to meet the water draft requirements for safe navigation, the resistance of the vessel will increase. At this time, the propulsion efficiency of the vessel is reduced, which may mean a vessel goes out of control. In addition, when the water depth is small, navigating shallow waterways may compromise the stability of a vessel, and could possibly cause damage to the vessel, rupture the hull, or have other consequences that may adversely affect navigation. In serious cases, the safety of the crew may even be endangered.

The classification of risk is based on the Low as Reasonably Practical (ALARP) principle. The risk level of a vessel under different water depth conditions were established as in [28], and shown in Table 3.

TABLE 3. Vessel risk classification.

Vessel risk level	Meaning
Low Risk	vessels can sail in this area, but should pay attention to the regulations of the waterway depth
Medium Risk	vessels can decide according to their own situation whether to sail under this conditions, but should try to avoid long-term navigation
High Risk	vessels should completely avoid sailing under these conditions

3) DETERMINATION OF VESSEL RISK RATING STANDARDS UNDER DIFFERENT WATER DEPTHS

Establishing $4 * 4$ dimensional fuzzy matrices for water depth conditions

$$\begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ D(t) = & \begin{bmatrix} d_{11}(t) & d_{12}(t) & d_{13}(t) & d_{14}(t) \\ d_{21}(t) & d_{22}(t) & d_{23}(t) & d_{24}(t) \\ d_{31}(t) & d_{32}(t) & d_{33}(t) & d_{34}(t) \\ d_{41}(t) & d_{42}(t) & d_{43}(t) & d_{44}(t) \end{bmatrix} \\ & = \begin{bmatrix} r_1(t) & r_2(t) & r_3(t) & r_4(t) \end{bmatrix} \end{array} \tag{16}$$

In (16), $d_{ij}(t)$ (i = 1, 2, 3, 4; j = 1, 2, 3, 4) indicates the degree of the j-level risk when the water depth condition is i.

The probability of different water depth conditions is $A = \left[a_1\ a_2\ a_3\ a_4 \right]$, a_i represents the probability of different water depth conditions. The fuzzy subset of the risk rating for bulk carriers is

$$B = \left[\max_{1\le k\le 4} \{A \circ a_1(t_k)\} \quad \max_{1\le k\le 4} \{A \circ a_2(t_k)\} \quad \max_{1\le k\le 4} \{A \circ a_3(t_k)\} \quad \max_{1\le k\le 4} \{A \circ a_4(t_k)\} \right] = \left[x_1 \quad x_2 \quad x_3 \quad x_4 \right] \tag{17}$$

According to the fuzzy matrix method and related experts questionnaire survey, the determination of Vessel Risk Rating was established in the Table 5 (D in the Table 5 represents water depth).

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. STUDY AREAS AND DATA

In this paper, we selected two waterways which are between the latitude and longitude from(longitude 120.725 °, latitude 32.049 °) to (longitude 120.835 °, latitude 31.981 °) and (longitude 118.926 °, latitude 25.236 °) to (longitude 119.003 °, latitude 25.199 °). The data for Nantong Port, Jiangsu Province and Meizhou Bay in Fujian Province were collected from 9:00 to 11:00 on February 6, 2017. The Nantong waterway is on the Yangtze River estuary, while the Meizhou Bay waterway is in Fujian Province. The traffic situation of Nantong is a complex inland waterway with many kinds of vessels passing through the area. In contrast, the Meizhou Bay is next to the East China Sea and it has been a critical port in China with convenient traffic. The number of data in the experiment is up to 100,000,000 and the big data processing environment is introduced in the process of the case study.

B. EXPERIMENT RESULTS

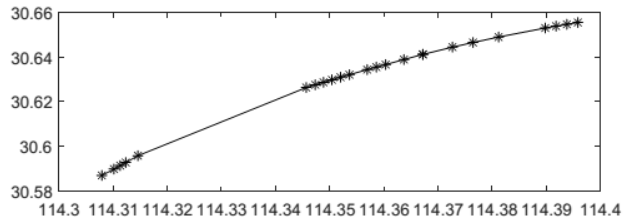
BP neural network was used for the prediction of the draft-depth data. Table 4 shows the results of correlation of vessel width, vessel breadth and the draft-depth for the input parameters of BP neural network.

TABLE 4. Correlation of vessel width, vessel breadth and the draft depth.

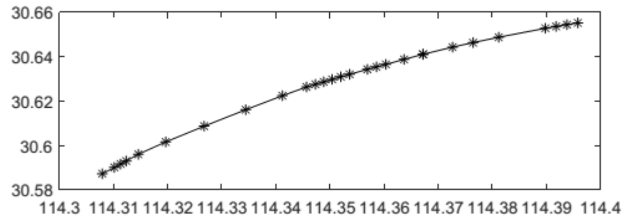
		vessel width	vessel breadth	draft depth
vessel width	Pearson correlation	1	0.905	0.689
	significance level		0.000	0.000
	N	1675	1675	1675
vessel breadth	Pearson correlation	0.905	1	0.685
	Significance level	0.000		0.000
	N	1675	1675	1675
draft Depth	Pearson correlation	0.689	0.685	1
	significance level	0.000	0.000	
	N	1675	1675	1675

Table 4 indicates that there is a strong correlation between the vessel length, vessel breadth and the depth of the draft. Thus vessel length and vessel breadth can be used as the input parameters of the BP neural network.

Figure 4 shows the results of Hermite interpolation for interpolating the missing AIS data.



(a)

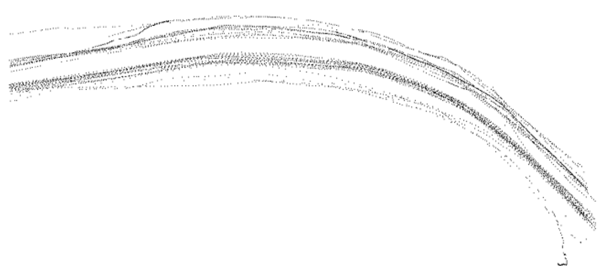


(b)

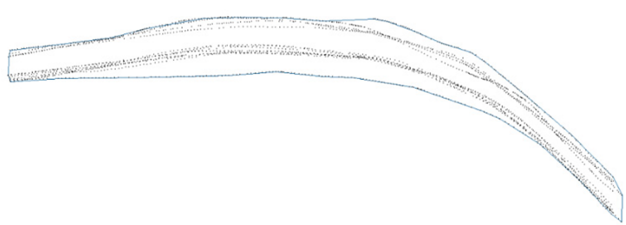
FIGURE 4. Results of Interpolation.

In Figure 4 (a), the trajectory connected by the primitive AIS data that have not been interpolated, Figure 4 (b) shows that after Hermite interpolation, the vessel trajectory is smooth.

Figure 5 shows the construction results of safe navigation area. Figure 5 (a) shows a vessel trajectory which



(a)



(b)

FIGURE 5. Results of Safe Navigation Area.

is constructed by connecting the vessel movement points. Figure 5 (b) shows the safe navigation area produced by the convex hull method.

It can be seen from Figure 5 that the vessel's navigation trajectory is generated by superimposing different locations of the same type of vessel and can reflect the traffic characteristics of ships in a specific area of the waterway. By using the convex hull method to process the vessel trajectories, a closed area can be obtained, in which the ship sailing is more frequent and can meet the draft requirements of this type of vessel, so the area is the safe navigation area for that vessel.

Before the construction of safe navigation reference map, the vessels were divided into four types: type A includes vessels with a maximum draft between 0.1-10.8 meters; and type B were vessels with a maximum draft between 10.8-14.6 meters. Type C were vessels with a maximum draft between 14.6-20 meters; and type D included vessels with a maximum draft of more than 20 meters. Figure 6 shows the reference map we constructed in Nantong waterway.

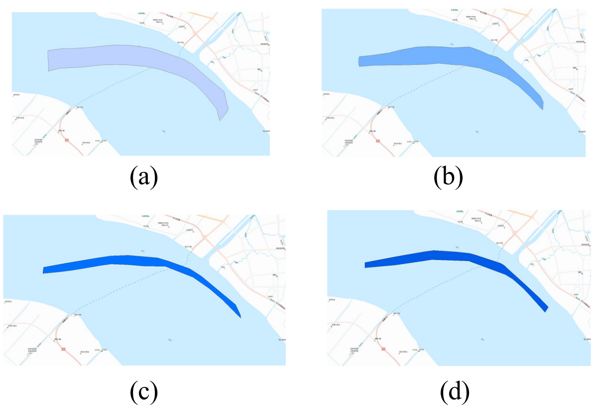


FIGURE 6. **Results of Depth Map in Nantong waterway.**

Figure 6 (a) is the map of type A vessel; Figure 6 (b) is the map of type B vessel; Figure 6 (c) is the map of type C vessel; Figure 6 (d) is the map of type D vessel. In the figures, the different blue areas represent the safe navigation areas of each type of ships, where the darker the blue is, the higher the requirements of the ship for draft, the more concentrated the navigation areas is in the center of the waterway. In order to compare the Nantong Port and Meizhou Bay navigation area, the same safe navigation area map was established for the Meizhou Bay navigation waterway in Figure 7.

Figure 7 shows the reference map we construct in Meizhou Bay waterway. Figure 7 (a) is the map of type A vessel; Figure 7 (b) is the map of type B vessel; Figure 7 (c) is the map of type C vessel; Figure 7 (d) is the map of type D vessel. In the safe navigation area of Meizhou Bay, the bigger the draft, the safe navigation area is concentrated in the center of the waterway, to meet the draft of vessels and ensure the safety of navigation. By superposing the safe navigation area map and the electronic chart of the two waterways, a safe

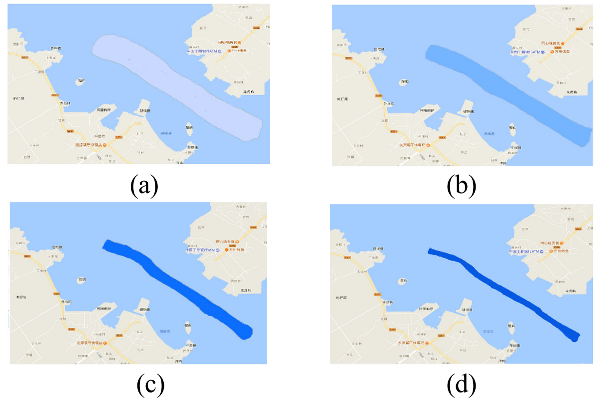


FIGURE 7. **Results of Depth Map in Meizhou Bay waterway.**

navigation reference map of the ship is generated as shown in Figure 8.

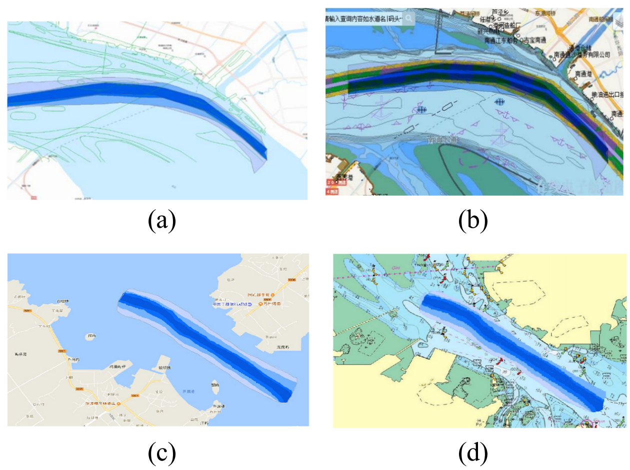


FIGURE 8. **Results of Reference Map.**

Figure 8 (a) is the Nantong waterway map combined with bathymetric chart, the green line is depth contour. Figure 8 (c) is the Meizhou Bay waterway reference map. The navigation areas are substantially in line with the water depth in the waterway.

Figure 8 (b) and (d) are the maps combined with electronic chart. The safe navigation, depth reference map covers 86% of the waterway part of the electronic chart, and the overlap area of the safe navigation reference map of the deep-water area and the deep-water part of the channel map is 80 %. The reference map demonstrates that vessels with smaller drafts are beyond the scope of the waterway, which indicates that the depth of the waters outside the waterway can meet the requirements of these vessels. In the deep- water area of the reference map, the distribution of the deep-water area is not average, indicating that the depth of the deep-water area in the map may not satisfy the demands of vessels with large drafts.

The experimental results show that the safe navigation, depth reference map is consistent with the waterway map provided by Changjiang Waterway Bureau. To some extent, our reference map reflects the actual water-depth of the waterway. This map could boost the ability to supervise traffic within less time.

As discussed in part D of section IV, the water depth affects the possibility of ship accidents, and the Determination of Vessel Risk Rating is obtained by fuzzy matrix method in Table 5.

TABLE 5. Determination of vessel risk rating

Vessel risk rating / Type of vessel	Low Risk	Medium Risk	High Risk
4000t-5000t bulk carrier	D>65m	25<D≤65m	D≤25m
Large ro-ro passenger boat	D>55m	22<D≤55m	D≤22m
Below 4000t bulk carriers	D>45m	18<D≤45m	D≤18m
general ro-ro passenger boat	D>35m	16<D≤35m	D≤16m

It can be seen from the table that the vessel with the highest requirement for water depth is 4000t-5000t bulk carrier and the lowest requirement for water depth is a general ro-ro passenger boat type. As the water depth decreases, the vessel risk rating tends to increase. When the water depth is below 25m, all types of ships are at or about to be in high-risk navigational conditions. Effective use of safe depth maps, in ship navigation, the rational design and planning of ship navigation routes, can avert and mitigate ship accidents, reduces the potential risks.

We can see similarities and differences in the reference maps when comparing the Nantong Port case and the Meizhou Bay cases. In Figure 6 and Figure 7, for ships with larger drafts, the safe navigation area is concentrated in the middle of the waterway. The water depth of this area is large, which can meet the draft requirements of large ships and conducive to the safe navigation of ships. For ships with smaller drafts, the safe sailing area is located more on outer margins than that of a large ship, a wider area can meet the draft requirements of smaller ships.

Table 5 shows that larger draft vessels have safer navigable conditions and will experience medium or high risk when the depth of the watercraft cannot reach safe depth. For ships with smaller draft, its safe sailing depth is medium, and its water depth threshold at medium or higher risk is smaller and larger than that of a large ship.

The proposed method based on ship motion trajectory aggregation is simpler than the other methods. From the experimental results, it can be seen that the safe navigation area of the ship is consistent with the actual ship navigation area. Through the calculation of the ship's risk, it can be seen that the ship's safe navigation reference map can help to avoid the propagation risk. The safe navigation map can control the navigation area of the ship so that the actual water depth of each type of ship sailing satisfies the requirements of this type of ship draft demand, thereby enhancing the safety of navigation. In addition, since AIS data is obtained in real time, the construction method of safe navigation reference map is also in a timely way.

VI. CONCLUSION AND FUTURE WORK

For safe navigation of vessel, the safe navigation reference map is constructed from big AIS data. Firstly, AIS dynamic data and static data are preprocessing respectively. Secondly, for some AIS data, there are many missed data during the transmission period. The BP neural network is used to predict maximum draft data according to vessel length and breadth. In this step, the correlation of vessel width, vessel breadth and the draft-depth is analyzed by the method of Pearson correlation coefficient. Thirdly, the method of Hermite interpolation is introduced to complete the data between two location points of vessel. The safety navigation depth-reference map is constructed according to different types of vessel. At last, the latest version of the electronic chart data is used to verify the accuracy of the safe navigation depth-reference map. The results show that, the safe navigation depth-reference map can more accurately reflect the water depth of waterway in time, and can provide full safety navigation information for vessel navigation. The method of this study can provide reference for data processing and mining trajectory data.

However, the constructing method of the safe navigation depth-reference map will consume much time for the transmission and processing of AIS data, which may mean the system cannot construct the map in a real-time state. Future works might combine the constructing method of map with big data processing platform in a distributed environment to enhance the processing ability, to improve the accuracy and timeliness of vessel navigation, and make maritime sector supervision of the depth conditions of waterways timelier.

REFERENCES

[1] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "Big data analysis-based secure cluster management for optimized control plane in software-defined networks," *Trans. Netw. Service Manag*, vol. 15, no. 1, pp. 27–38, Mar. 2018.

[2] L. Guo *et al.*, "A secure mechanism for big data collection in large scale Internet of Vehicle," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 601–610, Feb. 2017.

[3] J. Xu, K. Ota, and M. Dong, "Real-time awareness scheduling for multimedia big data oriented in-memory computing," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3464–3473, Oct. 2018.

[4] W. H. Mcanally *et al.*, "Nautical depth for U.S. navigable waterways: A review," *J. Waterway Port Coastal Ocean Eng.*, vol. 142, no. 2, pp. 1–13, Sep. 2015.

[5] B. Allotta *et al.*, "A new AUV navigation system exploiting unscented Kalman filter," *Ocean Eng.*, vol. 113, pp. 121–132, Feb. 2016.

[6] N. A. Bomberger, B. J. Rhodes, M. Seibert, and A. M. Waxman, "Associative learning of vessel motion patterns for maritime situation awareness," in *Proc. 9th Int. Conf. Inf. Fusion*, Jul. 2006, pp. 1–8.

[7] J. P. Jalkanen, A. Brink, J. Kalli, and H. Pettersson, "A modelling system for the exhaust emissions of marine traffic and its application in the Baltic Sea area," *Atmos. Chem. Phys.*, vol. 9, no. 23, pp. 9209–9223, 2009.
[8] J. W. Ulvila and J. E. Gaffney, Jr., "A decision analysis method for evaluating computer intrusion detection systems," *Decis. Anal.*, vol. 1, no. 1, pp. 35–50, Mar. 2004.
[9] G. K. Høye, T. Eriksen, B. J. Meland, and B. T. Narheim, "Space-based AIS for global maritime traffic monitoring," *Acta Astronaut.*, vol. 62, nos. 2–3, pp. 240–245, Feb. 2018.
[10] F. Xiao, H. Ligteringen, C. van Gulijk, and B. Ale, "Comparison study on AIS data of ship traffic behavior," *Ocean Eng.*, vol. 95, pp. 84–93, Feb. 2015.
[11] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Trans. Intell. Transport. Syst.*, vol. 19, no. 5, pp. 1559–1582, May 2018.
[12] H. Lang, S. Wu, and Y. Xu, "Ship classification in SAR images improved by AIS knowledge transfer," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 439–443, Mar. 2018.
[13] R. Pelich, N. Longépé, G. Mercier, G. Hajduch, and R. Garello, "AIS-based evaluation of target detectors and SAR sensors characteristics for maritime surveillance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3892–3901, Aug. 2015.
[14] K. Esfandiari, F. Abdollahi, and H. A. Talebi, "Adaptive control of uncertain nonaffine nonlinear systems with input saturation using neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2311–2322, Oct. 2015.
[15] R. Zhang, Z.-B. Xu, G.-B. Huang, and D. Wang, "Global convergence of online BP training with dynamic learning rate," *IEEE Trans. Neural Netw. Syst.*, vol. 23, no. 2, pp. 330–341, Feb. 2012.
[16] Y. Zhang, B. Chen, Y. Zhao, and G. Pan, "Wind speed prediction of ipso-bp neural network based on lorenz disturbance," *IEEE Access*, vol. 6, pp. 53168–53179, 2018.
[17] T. Longfei, X. Zhihong, and B. Venkatesh, "Contactor modeling technology based on an artificial neural network," *IEEE Trans. Magn.*, vol. 54, no. 2, Feb. 2018, Art. no. 4900108.
[18] F. Zhang and Q. Zhou, "Ensemble detection model for profile injection attacks in collaborative recommender systems based on BP neural network," *IET Inf. Secur.*, vol. 9, no. 1, pp. 24–31, 2015.
[19] D.-S. Lee, C.-S. Chang, and H.-N. Chang, "Analyses of the clustering coefficient and the pearson degree correlation coefficient of chung's duplication model," *IEEE Trans. Netw. Sci. Eng.*, vol. 3, no. 3, pp. 117–131, Jul./Sep. 2016.
[20] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Trans. Audio Speech Lang Process.*, vol. 16, no. 4, pp. 757–765, May 2008.
[21] J. Wang and N. Zheng, "A novel fractal image compression scheme with block classification and sorting based on pearson's correlation coefficient," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3690–3702, Sep. 2013.
[22] K. K. Delibasis and A. Kechriniotis, "A new formula for bivariate hermite interpolation on variable step grids and its application to image interpolation," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2892–2904, Jul. 2014.
[23] K. K. Delibasis and A. Kechriniotis, "A new formula for bivariate Hermite interpolation on variable step grids and its application to image interpolation," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2892–2904, July 2014.
[24] X. Li, F. Yang, D. Wu, Z. Zhou, and X. Zeng, "MOS table models for fast and accurate simulation of analog and mixed-signal circuits using efficient oscillation-diminishing interpolations," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 34, no. 9, pp. 1481–1494, Sep. 2015.
[25] C.-C. Tseng and S.-L. Lee, "Design of fractional delay filter using hermite interpolation method," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 7, pp. 1458–1471, Jul. 2012.
[26] J. Zou, T. Jin, W. Li, J. Lee, and S. Chang, "A hermite interpolation model to accelerate the calculation of the horizontal electric field of a lightning channel along a transmission line," *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 1, pp. 124–131, Feb. 2013.
[27] S. Lu, Y. Wang, and Y. Wu, "Novel high-precision simulation technology for high-dynamics signal simulators based on piecewise hermite cubic interpolation," *IEEE Aerosp. Electron.*, vol. 54, no. 5, pp. 2304–2317, Oct. 2018.
[28] R. E. Melchers, "On the ALARP approach to risk management," *Rel. Eng. Syst. Saf.*, vol. 71, no. 2, pp. 201–208, Jan. 2001.

ZHENGWEI HE received the Ph.D. degree from Wuhan University, China. He is currently an Associate Professor with the School of Navigation, Wuhan University of Technology, Wuhan, China. His research interests include traffic big data processing and mining, maritime information systems, and traffic environment simulation. His focus is on artificial intelligence application technology, deep learning, and smart navigation.

FAN YANG is currently pursuing the M.Sc. degree with the School of Navigation, Wuhan University of Technology, Wuhan, China. His interests have focused on traffic big data mining, machine learning, and artificial intelligence.

ZHONG LI received the M.E. degree from the School of Energy and Power Engineering, Wuhan University of Technology, Wuhan, China. His main research interests include ship data mining, ship navigation information visualization, and maritime traffic information modeling and simulation.

KEZHONG LIU is currently a Professor with the School of Navigation, Wuhan University of Technology, Wuhan, China. His research interests include pervasive computing and wireless sensor networks, traffic information system, and traffic safety technology.

NAIXUE XIONG is currently an Associate Professor with the School of Computer Science, Northeastern University, Boston, MA, USA. His research interests include computer science, big data processing and mining, and Internet of Things technology. His focus is on deep learning and artificial intelligence application technology.