Influence of environmental factors on human-like decision-making for intelligent ship

Xue, Jie; Chen, Zhijun; Papadimitriou, Eleonora; Wu, Chaozhong; Van Gelder, P. H.A.J.M.

DOI
10.1016/j.oceaneng.2019.05.042

Publication date
2019

Document Version
Final published version

Published in
Ocean Engineering

Ocean Engineering 186 (2019) 106060

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Influence of environmental factors on human-like decision-making for intelligent ship

Jie Xue [a,b,c,d,e], Zhijun Chen [a,b,d], Eleonora Papadimitriou [c], Chaozhong Wu [a,b,d,e,*], P.H.A.J. M. Van Gelder [c,**]

[a] Intelligent Transportation Systems Research Center(ITSC), Wuhan University of Technology, Wuhan, China
[b] National Engineering Research Center for Water Transport Safety(WTSC), Wuhan, China
[c] Faculty of Technology, Policy and Management, Safety and Security Science Group(S3G), Delft University of Technology, Delft, the Netherlands
[d] Engineering Research Center for Transportation Safety, Ministry of Education, Wuhan, China
[e] School of Energy and Power Engineering, Wuhan University of Technology, Wuhan, China

A R T I C L E I N F O

Keywords:
C4.5 algorithm
Ship maneuvering
Decision-making
Intelligent ship
Classification rule
Data mining

A B S T R A C T

To date, the increasing density of water traffic has caused the ship's navigation environment to deteriorate, resulting in frequent water traffic accidents. In addition, a majority of maritime accidents are caused by human factors, and one of the important ways to solve the ship accidents caused by human factors is to utilize intelligent maneuvering of ships. Based on the actual crews' operational data from full-task handling simulation platform, this study combines a 30,000-ton bulk carrier inbound navigation scenario and uses the decision tree method to propose a knowledge learning model under multiple environmental constraints to give intelligent ships the ability to make decisions like a human: An intelligent ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model. The decision-making mechanism for the maneuvering behavior of Officer On Watch (OOW) under the influence of the specific water traffic environment in the inbound scenario is analyzed, and the OOW's decision-making knowledge is automatically acquired and represented. The validation tests and the comparative analysis with the classic classification algorithms of k-Nearest Neighbours (k-NN) and Support Vector Machine (SVM) are performed to demonstrate the accuracy of the proposed HDMDR model. This paper provides a feasible basis for the human-like decision-making analysis of intelligent ships.

1. Introduction

Driven by economic globalization, the volume of trade between countries around the world continues to rise, and higher demands are placed on the transportation of goods. Due to its large volume and low cost, waterway transportation plays an increasingly important role in cargo transportation. It bears the main task of world cargo circulation and is the main means of trade transportation. Currently, waterway transportation accounts for 95% of total crude oil transportation and 99% of total iron ore transportation. It is an irreplaceable transportation method. However, with the increasing number of vessels and the increasingly busy routes, the environmental pollution related to waterway transportation, the high labor costs and the lack of safety have also received more attention (Lun et al., 2016). In recent years, the development of technologies, such as information, computers, communications, networks, new energy, artificial intelligence, application of the Internet of Things, big data, integrated bridge systems, and information physics systems, have greatly advanced the process of ship intelligence and made real green, safe, efficient and unmanned intelligence ships a possibility.

At the same time, water transportation is recognized as a high-risk industry. With the development of the domestic economy and world trade, transportation is becoming increasingly busy, the number of ships is increasing, ships are becoming larger and more specialized, and the

* Corresponding author. Intelligent Transport Systems Research Center (ITSC), Wuhan University of Technology, 1040 Heping Avenue, Wuhan, Hubei, 430063, China.
** Corresponding author. Faculty of Technology, Policy and Management, Safety and Security Science Group (S3G), Delft University of Technology, Jaffalaan 5, 2628 BX, Delft, the Netherlands.
E-mail addresses: J.Xue@tudelft.nl (J. Xue), chenzj556@whut.edu.cn (Z. Chen), E.Papadimitriou@tudelft.nl (E. Papadimitriou), wucz@whut.edu.cn (C. Wu), P.H.A.J.M.vanGelder@tudelft.nl (P.H.A.J.M. Van Gelder).

https://doi.org/10.1016/j.oceaneng.2019.05.042
Received 24 October 2018; Received in revised form 20 March 2019; Accepted 20 May 2019
Available online 24 June 2019

speed of ships is increasing. Coupled with the increase in the transportation of dangerous goods, the density of water traffic is increasing, and the navigation environment of ships is deteriorating, causing frequent water traffic accidents, which causes people to pay more attention to the risk of navigation (Akyuz and Celik, 2014; Goerlandt and Montewka, 2015). According to statistics (Hanzu-Pazara et al., 2008), in ship collision accidents, 89%–96% of accidents are caused directly or indirectly by human factors, and one of the important ways to solve ship accidents caused by human factors is to utilize intelligent maneuvering of ships. In addition, the safety of the crew in extreme weather conditions in recent years has also become a problem that cannot be ignored (Wang et al., 2014). Besides, the number of crews is declining recently, while the wages of crew are rising year by year, which has become the second largest expenditure item after the fuel costs of shipping (Lun et al., 2016). As intelligent ships have outstanding advantages in improving the safety management, energy consumption management, and operational efficiency of ships, therefore, the researches for intelligent ships have become an inevitable trend for future ship development, and gained the interest of many researchers in both academia and private sectors (Goerlandt and Montewka, 2015).

In addition, the natural environment is an important factor affecting the safety of waterborne traffic (Zhang et al., 2018). Among the natural environmental factors surrounding the ship, meteorological conditions, walrus conditions, topographical environments and water facilities will bring restrictions to the navigation of the ship. These factors affect the ship's navigation and the crew's decisions by affecting the ship's maneuverability, along with the skill and mentality of the shipper. The natural environmental factors that typically affect the safe environment of maritime traffic are weather conditions and ocean conditions, specifically, wind, current, and waves.

Intelligent ships use sensors, communications, Internet of Things, the Internet and other technical means to automatically sense and obtain information and data on the ship itself, the marine environment, logistics, ports, etc. Based on computer technology, automatic control technology, big data processing and analysis technology, it utilizes intelligent operation in ship navigation, management, maintenance, cargo transportation, etc. (Lazarowska, 2017), making ships safer, more environmentally friendly, more economical and more reliable. “Intelligent” here can be understood as “human-like thinking”. It can comprehensively consider the specific tasks and various information obtained and develop a series of optimal decisions that meet the safety requirements of the ship's navigation, economy, and environment. It takes a long transition period for an intelligent ship to fully realize unmanned maneuvering. Presently, although the current level of ship automation is relatively high, the normal operation of ships is always inseparable from human participation (Perera et al., 2015). Even in an unattended cabin, the crew must be handled when an emergency occurs. Although the ship is maneuvered by satellite navigation, electronic compass, electronic channel map and automatic rudder, the bridge has not been unmanned. Intelligent ship technology has developed rapidly in recent years, however, there are still many problems need to be solved. In addition, the existing research does not form a set of theoretical methods to solve the problem of autonomous learning of the intelligent ship for the maneuvering decision-making characteristics of Officer On Watch (OOW) and lacks the corresponding theoretical methods to solve the problem of intelligent ship human-like maneuvering decision-making modeling.

Researchers have proposed several different decision tree algorithms (see literature review) for both classification and decision-making problems based on different aspects and obtained good results. Based on the advantages of the C4.5 algorithm and the ability to analyze the characteristics of multifork trees, this paper uses the C4.5 algorithm to learn the OOW's maneuvering decision characteristics. We regard the intelligent ship human-like maneuvering decision-making problem as a machine learning problem based on the OOW's experience, the OOW's actual maneuvering data, and the environmental influencing factors, such as wind, wave, and current in specific water areas, and the problem is converted using the decision tree C4.5 method to learn the OOW's maneuvering decision-making characteristics, thus constructing a human-like decision-making model under multiple constraints.

In summary, this study focuses on the concept of human-like maneuvering for the intelligent ship and studies the human-like decision-making method of intelligent ships. By establishing autonomous learning method of maneuvering decision-making, the maneuvering decision-making rules of typical maneuvering style is explored, and the processes of autonomous learning OOWs' maneuvering decision-making characteristics for intelligent ships are studied, and the intelligent ship human-like decision-making model is constructed. This study provides a new perspective and methodology for the development of intelligent ship technology in theory and practice and promotes the application and spreading of intelligent ships. The main contributions of this study are as follows:

1) A novel intelligent ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model is proposed.
2) The standardization principle of environmental influencing factors and maneuvering decision-making factors is developed.
3) The decision-making mechanism of the OOW's maneuvering behavior is analyzed on the basis of the actual crews' operational data from full-task handling simulation platform, and the OOW's decision-making knowledge under the specific environmental influencing factors in the inbound scenario is automatically acquired and represented.
4) Considering the high cost of using the real 30,000-ton ship to carry out this kind of experiment, and the low feasibility of collecting the data of multiple voyages from the real-world ship, therefore, it is unique and very valuable to obtain the experimental data operated by an experienced OOW on the full-task handling simulation platform in a certain time and space.

The structure of this paper is organized as follows. Initially, Section 2 reviews the literature. Section 3 briefly presents the proposed decision-making model. The experimental processes are introduced in Section 4. Section 5 details the experimental results and the performance of our optimization methodology. The conclusions and future directions of research are addressed in Section 6.

2. Literature review

Data mining is a process that uses analytical tools to extract information and knowledge, including knowledge that is hidden, unknown, or incomplete but potentially useful, from a large amount of incomplete, noisy, fuzzy, and random data. Moreover, data mining determines the relationship between models and data and uses it to make predictions (Aguiar-Pulido et al., 2013; Sanil, 2001). The classification algorithm is a data analysis method belonging to predictive data mining. Its goal is to find models that accurately describe and distinguish data classes or concepts from important sample data sets, such that they can be grouped into a data category based on the entity's attribute values and other constraints. The current technologies and methods mainly include decision tree algorithms (Calistru et al., 2015; Xie et al., 2003), Bayesian classification and Bayesian networks (Baksh et al., 2018), neural networks (Kheradpisheh et al., 2018), genetic algorithms (Peng et al., 2015), rough sets (Zhang et al., 2012), etc.

In the 1960s, decision tree algorithm was initially proposed by Hunt et al. (1966) to minimize the cost of classifying an object (Quinlan, 1986). Decision trees can handle both categorical and numerical data and it is good at processing nonnumeric data, which can eliminate a significant amount of data preprocessing work when dealing with numerical data through algorithms, such as neural networks. In addition, the decision tree method is simple in structure and does not need much background knowledge in the process of learning. Second, the decision

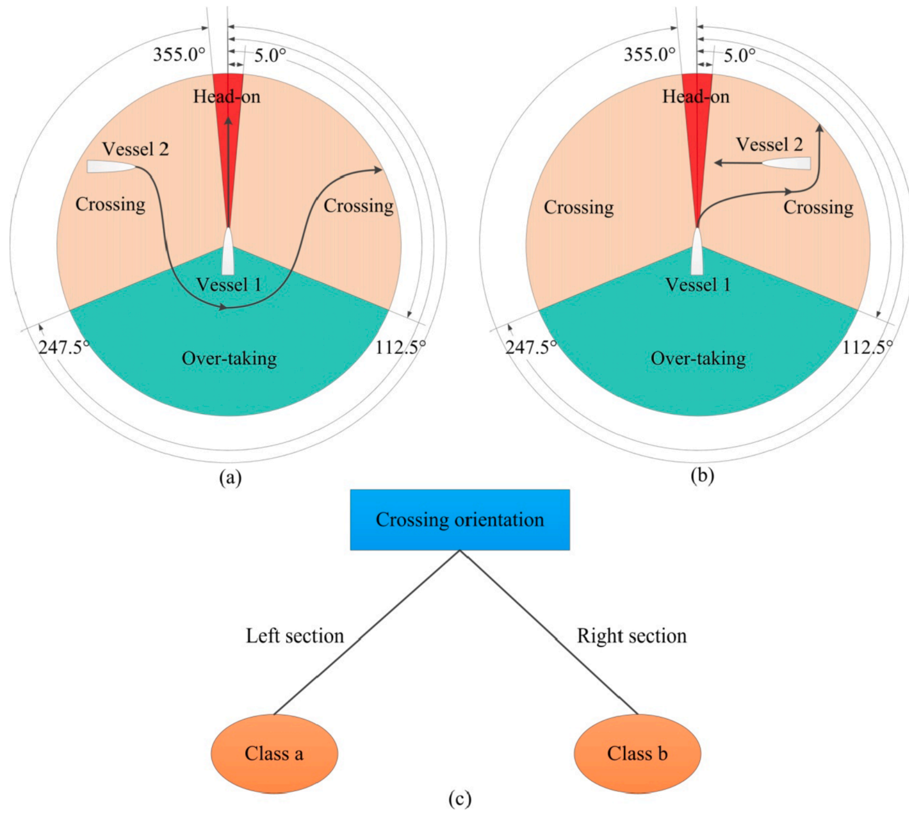

Fig. 1. The collision avoidance operation in the encounter scenario of different crossing situations and the decision tree generated from this case.

tree model is more efficient and is more suitable for training sample sets that have large amounts of data; third, the computational tree algorithm has a relatively small amount of computation; then, the decision tree method typically does not require knowledge outside the training data and is good at processing nonnumeric data; finally, the decision tree method has a higher classification accuracy. Therefore, the decision tree method is a key research direction in the field of machine learning. Muchoney et al. proposed the classification algorithms of decision tree (DT), artificial neural network (ANN), and maximum-likelihood to analyze the land cover classification problem in central United States, and the results show that the decision tree has the highest classification accuracy (Muchoney et al., 2010). Borak (1999) used decision trees to classify features from a large amount of data. The results show that the tree-based classifier can greatly reduce the dimensionality of the input data set without affecting the classification accuracy. Calistru et al. (2015) proposed a novel parallel decision tree algorithm, namely, PdsCART, to process a larger amount of data stream records and construct the tree efficiently. Saunier et al. (2011) used decision trees, the k-means algorithm, and the hierarchical agglomerative clustering method to identify patterns in the traffic event database and analyze the relationship between interaction attributes and collision.

Common decision tree algorithms are Concept Learning System (CLS) (Angluin, 1988; Hunt et al., 1966), Iterative Dichotomiser 3 (ID3) (Quinlan, 1979, 1986), C4.5 (Quinlan, 1993), C5.0 (Bujlow et al., 2012; Pandya and Pandya, 2015), Classification And Regression Trees (CART) (Calistru et al., 2015; Friedman et al., 1984), CHi-squared Automatic Interaction Detector (CHAID) (Kass, 1975; Rodriguez et al., 2016), etc. The internal variables of each subsample are highly consistent, and the corresponding variation/impurity falls between different subsamples as far as possible. All decision tree algorithms follow this criterion, and the data set is partitioned into subsets with different statistical approaches, such as Entropy (Lakkakula et al., 2014), Gain Ratio (Prasad and Naidu, 2013), Gini coefficient (Prasad et al., 2013; V et al., 2013), etc.

A series of follow-up decision tree programs, such as ID3, C4.5, and CART, etc. are all developed from CLS. Among them, the C4.5 algorithm developed based on ID3, is currently one of the most famous and popular decision tree algorithms (Lu et al., 2015), C4.5 is the most influential data mining algorithm identified by the IEEE International Conference on Data Mining (ICDM) in December 2006 (Wu et al., 2007). A comparative study of C4.5 and other learning algorithms shows that it can balance processing speed and error rate well (Lim et al., 2000). C4.5 can convert the decision tree into an equivalent production rule, solve the learning problem of continuous value data, classify multiple categories, increase the Boosting technology, and complete the processing of large databases more efficiently. The C4.5 algorithm also deals well with continuous and discrete values and attributes with missing attribute values (García-Laencina et al., 2015). The C4.5 algorithm solves the above problem well; however, the ID3 algorithm tends to favor more attributes and the data of discrete value attributes, but not the attributes with continuous values nor the samples with missing values, and is sensitive to noise (Hssina et al., 2014). C5.0 mainly adds support for Boosting, which also uses less memory. Compared with the C4.5 algorithm, it builds a smaller rule set; therefore, it is more accurate, but C5.0 is a commercial software, and the public cannot easily get the source code (Witten et al., 2016). CART uses the training set and the cross-validation set to continuously evaluate the performance of the decision tree to prune the decision tree, thus achieving a good balance between training error and test error. However, CART and CHAID only supports building binary trees, while C4.5 allows two or more outcomes and supports binary or multifork trees (Wu et al., 2007). Several prior studies on the C4.5 DT could be found from the literature. A prior study

Table 1
The data for the example.

No.	Crossing orientation (attributes)	Class
1	Left	a
2	Right	b

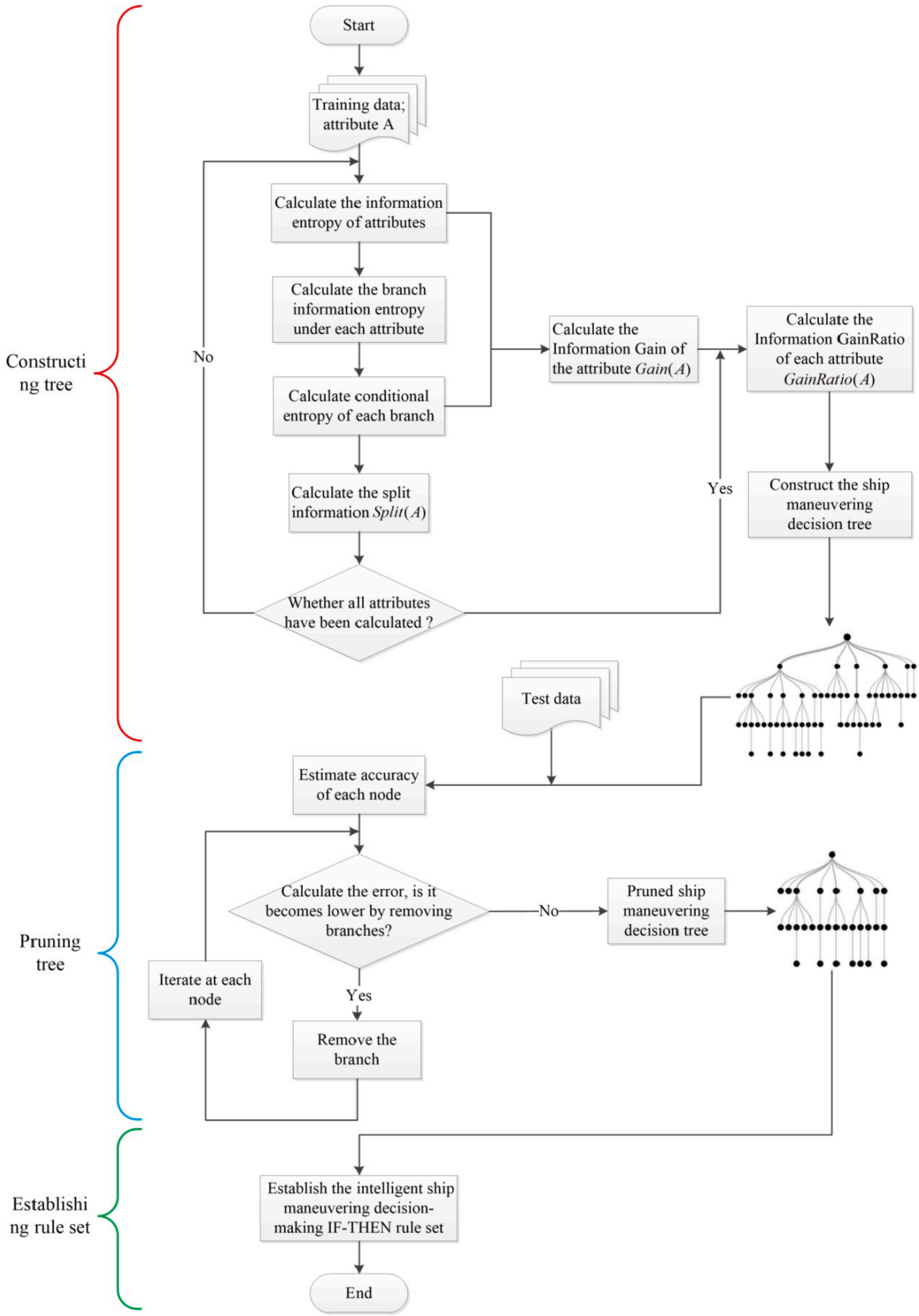

Fig. 2. The proposed HDMDR model logic schema.

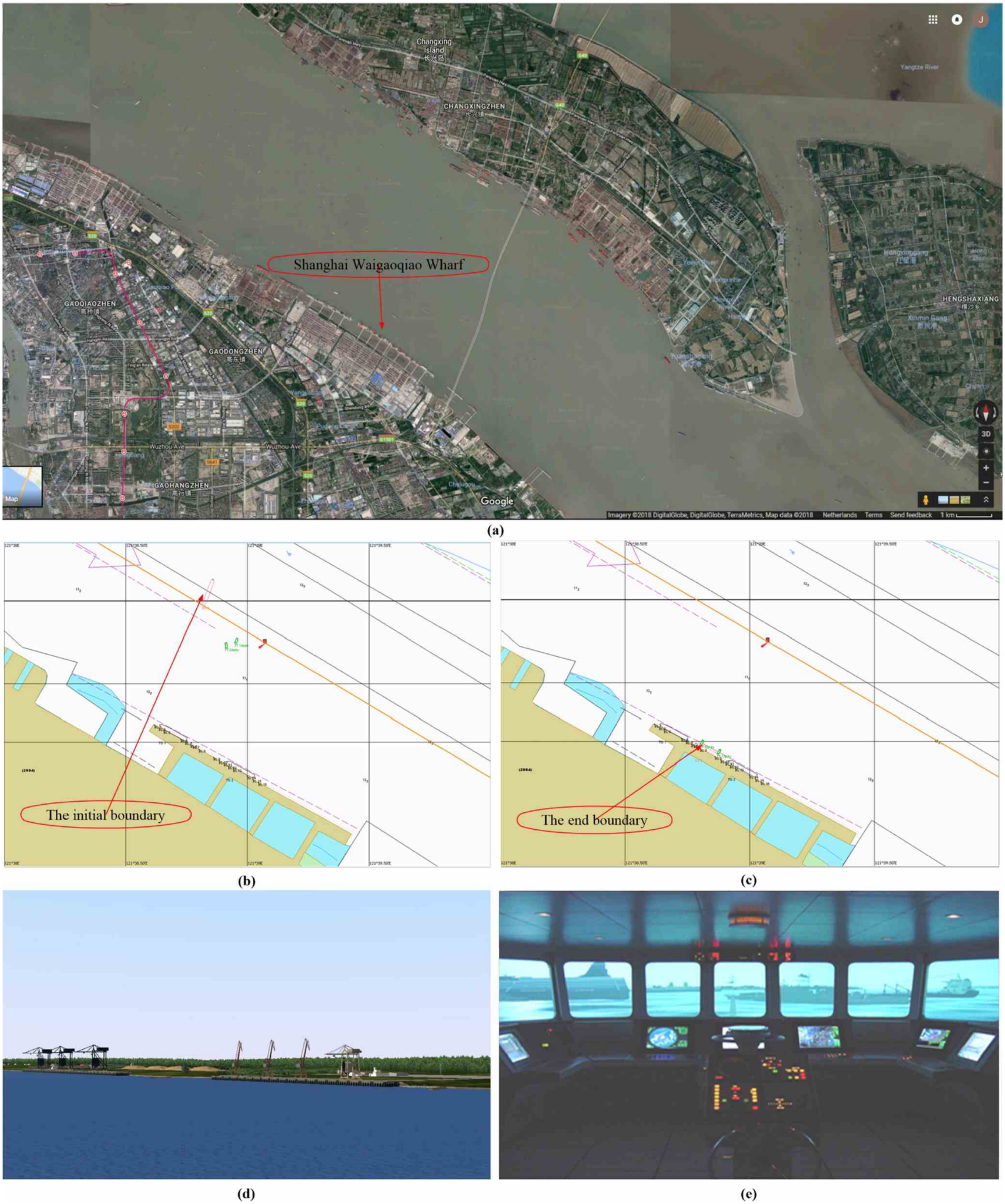

Fig. 3. The designed experimental scenario.

Table 2
Participants' information.

	Number of Participants	Age (years)		Piloting experience (years)	
		Mean	SD	Mean	SD
All	96	38.76	4.13	8.89	2.10
Captain	35	42.29	2.18	10.74	1.29
Chief Officer	61	36.74	3.59	7.82	1.69

(Provost and Domingos, 2003) found that a C4.5 introduction learner without pruning and without node "collapsing" (Quinlan, 1993) can achieve the best prediction accuracy. A novel VFC4.5 was proposed by Cherfi et al. (2018) to build decision trees through reducing the number of cut points by using the arithmetic mean and median, this algorithm could get excellent accuracy than a C4.5 algorithm. Reumers et al. (2013) used C4.5 decision tree-based model to infer activity types from Global Positioning System (GPS) traces, the results showed that the overfitting was minimal, in addition, the model enables researchers to infer activity types directly from activity start time and duration information obtained from GPS data. Dai and Ji (2014) proposed a parallel MapReduce algorithm to implement a typical C4.5 decision tree, the experimental results indicated that the algorithm exhibits both time efficiency and scalability.

3. Methodology

3.1. Decision tree

A decision tree is a mathematical method that generates decision trees or decision tree rules by inductive learning of training samples and then classifies new data using decision trees or decision rules. As a supervised case-based inductive learning algorithm, decision trees evolved from the artificial neural network method, which is a method to solve complex decision problems through tree-like logical thinking. It can infer the classification rules of the decision tree representation from a set of unordered and irregular cases. It typically forms a classifier and a prediction model, which can classify, predict and analyze the unknown data for knowledge discovery.

The decision tree consists of a root node, a series of internal nodes, and leaf nodes. Each node has only one root node and two or more leaf nodes, and the nodes are connected by branches (Yuan and Shaw, 1995). Each internal node of the decision tree corresponds to a collection of noncategory attributes, with each edge corresponding to each possible value of the attribute. The leaf nodes of the decision tree correspond to a category attribute value, and different leaf nodes can correspond to the same category attribute value. In addition to being represented in the form of a tree, a decision tree can also be represented as a set of production rules in the form of IF-THEN. Each root-to-leaf path in the decision tree corresponds to a rule. The condition of the rule is the rounding of all node attribute values on the path. The rule's conclusion is the category attribute of the leaf node on the path. Compared with decision trees, rules are more concise and easier for people to understand, use and modify, which form the basis of the expert system. Therefore, in practical applications, more rules are used.

The decision tree method consists of two main steps. The first step is to use the training sample set to build and generalize a decision tree and build a decision tree model. This process is actually a process of acquiring knowledge from the data and doing machine learning. It is usually divided into two phases: building and pruning. The second step is the process of classifying new data using a built-in decision tree.

The input of the decision tree learning algorithm is a set of training samples represented by attributes and attribute values, and the output is a decision tree (which can also be extended to other representations, such as rule sets). Decision tree generation typically uses a top-down recursive approach. The optimal attribute is selected as the node of the tree by some method, and the attribute values are compared on the node, and the branch from the node is judged according to the different attribute values that correspond to the training samples. The lower nodes and branches are repeatedly established in each branch subset, and the growth of the tree is stopped under certain conditions, and the conclusions are obtained at the leaf nodes of the decision tree to form a decision tree. The decision tree is generated by performing decision tree learning on the training samples. The decision tree can classify an unknown sample set according to the value of the attribute, which is the decision tree classification.

Fig. 1(c) shows an example of a typical binary decision tree based on the data shown in Table 1. From Fig. 1(c), we can see that a decision node/attribute (i.e., Crossing orientation, which represents the position of Vessel 2) has two branches/values (i.e., Right section and Left section, which represent the unique values for the specific attribute). Leaf node

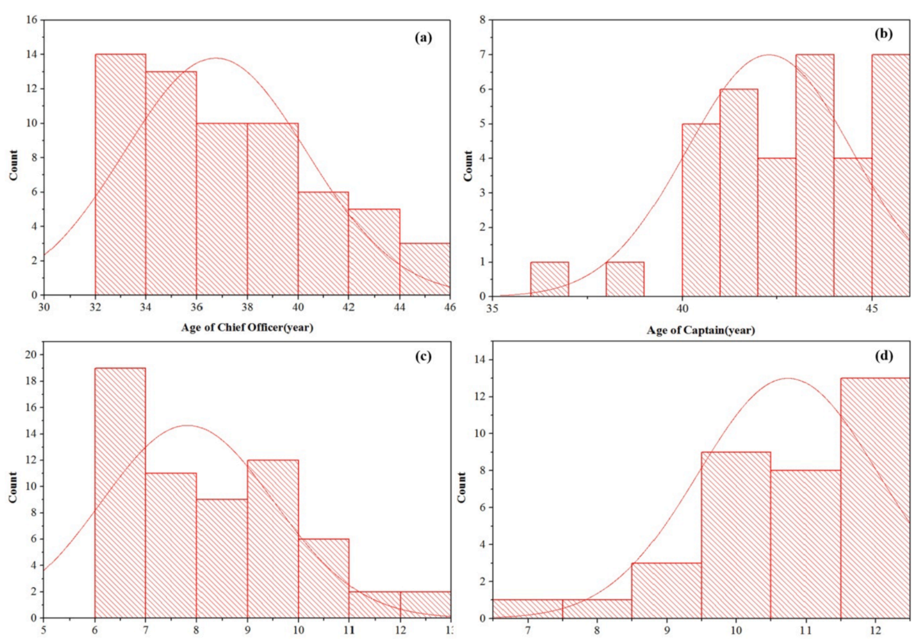

Fig. 4. The distribution of OOWs' age and their piloting experience.

Table 3
Training samples for evaluation of the studied area (partially).

No.	X		Y1	Y2	Y3	Y4	Y5	Y6
	Rudders Order	Telegraphs Order						
1	−1.0697	−30.0000	318.4000	1.1080	−173.7550	83.4139	−77.2000	7.7235
2	−1.7666	−30.0000	318.4000	1.1080	−173.3871	83.5533	−77.2000	7.7235
3	−2.0000	−30.0000	318.4000	1.1080	−173.2907	83.6046	−77.1954	7.7226
4	−2.0000	−30.0000	318.4000	1.1080	−172.9988	83.7506	−77.1000	6.7041
5	−2.0000	−30.0000	318.4000	1.1080	−172.8908	83.8092	−77.0954	6.7041
6	−2.0000	−30.0000	318.4000	1.1080	−172.6508	84.0000	−77.0000	6.7041
7	−2.0000	−30.0000	318.4000	1.1080	−172.6000	84.0000	−77.0000	6.7041
8	−2.0000	−30.0000	318.4000	1.1080	−172.2018	84.3491	−76.9313	6.6847
9	−2.0000	−30.0000	318.4000	1.1080	−172.1000	84.4000	−76.9000	6.6847
10	−2.0000	−30.0000	318.4000	1.1080	−171.7163	84.6837	−76.7582	6.6847
11	−2.0000	−30.0000	318.4000	1.1080	−171.6000	84.8000	−76.7000	6.6847
12	−2.0000	−30.0000	318.4000	1.1080	−171.3069	85.0931	−76.6334	6.6847
13	−2.0000	−30.0000	318.4000	1.1080	−171.2000	85.2000	−76.6000	6.6847
14	−2.0000	−30.0000	318.4000	1.1080	−170.9000	85.3000	−76.6000	6.6847
15	−2.0000	−30.0000	318.4000	1.1080	−170.9000	85.3000	−76.6000	6.6847
16	−2.0000	−30.0000	318.4000	1.1080	−170.3351	85.6474	−76.3763	6.6652
17	−2.0000	−30.0000	318.4000	1.1080	−170.1000	86.0000	−76.2000	6.6652
18	−2.0000	−35.9724	318.4000	1.0754	−169.9281	86.2437	−76.0806	6.6652
19	−2.0000	−40.0000	318.4000	1.0753	−169.8000	86.5000	−76.0000	6.6652
20	−2.0000	−45.6276	318.4000	1.0753	−169.4059	86.7626	−75.8687	6.6752
21	−2.0000	−50.0000	318.4000	1.0755	−169.2000	86.9000	−75.8000	6.6852
22	−2.0000	−50.0000	318.4000	1.1080	−168.9564	87.1718	−75.6761	6.7052
23	−2.0000	−50.0000	318.4000	1.1080	−168.5000	87.4000	−75.6000	6.7552
24	−2.0000	−50.0000	318.4000	1.1080	−168.6957	87.6273	−75.5497	6.8043
25	−2.0000	−50.0000	318.4000	1.1080	−169.3000	87.8000	−75.5000	7.1655
26	−2.0000	−62.0366	318.4000	1.1080	−168.0185	88.1405	−75.4000	7.2612
27	−2.0000	−70.0000	318.4000	1.0755	−167.7000	88.3000	−75.4000	7.3272
28	−1.9030	−70.0000	318.4000	1.1080	−167.5368	88.5632	−75.3000	7.6652
29	−1.8120	−70.0000	318.4000	1.1080	−167.3000	88.8000	−75.3000	7.6652
30	−1.7090	−70.0000	318.4000	1.1080	−166.6993	89.0801	−75.3000	7.6510
…	…	…	…	…	…	…	…	…

(i.e., Class, which represents the crossing situation) represents the class category or decision of each instance.

Furthermore, according to the COLREGs (International Regulations for Preventing Collisions at Sea) navigation rules which provide safe operation guidelines for maritime navigation. As shown in Fig. 1(a), if vessel 2 is at the left crossing section (Class a), the vessel 2 should turn right and vessel 1 should keep its course; if vessel 2 is at the right crossing section (Class b), the vessel 1 should turn right, and vessel 2 should keep its course, shown as Fig. 1(b). Therefore, the final decision can also be represented through the form of IF-THEN rule set shown as follows:

Table 4
Standardization principle of environmental influencing factors for inbound maneuvering decision-making (input).

Influencing factors	Meaning	Symbolic principle		
		Small (a)	Medium (b)	Large (c)
Y1	Current direction (degrees)	[313.9000, 315.5000)	[315.5000, 317.1000)	[317.1000, 318.7001)
Y2	Current speed (knots)	(1.0107, 1.0432)	[1.0432, 1.0756)	[1.0756, 1.1080]
Y3	Relative current direction (degrees)	[-60.0000, 0.0000) [0.0000, 60.0000)	[-120.0000, −60.0000) [60.0000, 120.0000)	(-180.0000, −120.0000) [120.0000, 180.0000]
Y4	Relative wave direction (degrees)	[-41.5000, 0.0000) [0.0000, 41.5000)	[-83.0000, −41.5000) [41.5000, 83.0000)	(-124.5000, −83.0000) [83.0000, 124.8000]
Y5	Relative wind direction (degrees)	[-59.0205, 0.0000) [0.0000, 59.0205)	[-118.0411, −59.0205) [59.0205, 118.0411)	(-179.7170, −118.0411) [118.0411, 179.8750)
Y6	Relative wind speed (knots)	(0.0228, 7.5154)	[7.5154, 14.7664)	[14.7664, 22.1793)

Rule 1: IF Crossing orientation = Left THEN Class = a (Vessel 1 keeps course and vessel 2 turns right)
Rule 2: IF Crossing orientation = Right THEN Class = b (Vessel 1 turns right and vessel 2 keeps course)

This example indicates a maritime problem of COLREGs situation: the decision tree generated from collision avoidance operation in the encounter scenario of different crossing situations In this way, the attributes taken together provide a zeroth-order language for characterizing objects in the universe (Quinlan, 1986).

3.2. The proposed HDMDR model

3.2.1. Theory information

Shannon (1948) proposed the information theory in 1948, and the amount of information on events could be calculated as follows:

$$I(S_i) = -p(S_i)\log_2 p(S_i) \tag{1}$$

where $p(S_i)$ is the probability of occurrence of event S_i.

Suppose that there are v mutually exclusive events $S_1, S_2, ..., S_v$, and only one of them happens. The average amount of information can be measured as follows:

$$I(S_1, S_2, ..., S_v) = -\sum_{i=1}^{v} p(S_i)\log_2 p(S_i) \tag{2}$$

When $p(S_i) = 0$, then $I(S_i) = -p(S_i)\log_2 p(S_i) = 0$.

3.2.2. Information entropy

Assume that D is the intelligent ship human-like decision-making training data set contains a set of m classes, $|D|$ stands for the total number of samples in data set D, and $|S_i|$ is the number of samples in data set D that belongs to class $S_i (i = 1, 2, ..., m)$. If we randomly select a sample from D, and this sample belongs to class S_i, then we can get a

Table 5
Maneuvering decision-making factors and standardization principle (output).

Attributes	Speed control			Course control		
	Symbolic principle	Status	Symbol	Symbolic principle	Status	Symbol
Variety	$a_{i+1} - a_i \neq 0$	Changed	C1	$b_{1+1} - b_i \neq 0$	Changed	C2
	$a_{i+1} - a_i = 0$	Unchanged	U1	$b_{i+1} - b_i = 0$	Unchanged	U2
Direction	$a_i \geq 0$	Ahead	D1	$b_i \geq 0$	Starboard	D2
	$a_i < 0$	Astern	T1	$b_i < 0$	Port	T2
Maneuvering factors	Decisions		Symbols	Decisions		symbols
X(Dimensionless)	U1D1U2T2		X1	U1D1C2T2		X9
	U1T1U2T2		X2	U1T1C2T2		X10
	U1D1U2D2		X3	U1D1C2D2		X11
	U1T1U2D2		X4	U1T1C2D2		X12
	C1D1C2T2		X5	C1D1U2T2		X13
	C1T1C2T2		X6	C1T1U2T2		X14
	C1D1C2D2		X7	C1D1U2D2		X15
	C1T1C2D2		X8	C1T1U2D2		X16

prior probability of the event as follows:

$$p_i = |S_i|/|D| \tag{3}$$

The expected information (also referred to as entropy) needed to classify D into m classes is defined as:

$$I(|S_1|, |S_2|, \ldots, |S_m|) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{4}$$

Suppose a feature/attribute A has n distinct values, $\{a_1, a_2, \ldots, a_n\}$, feature/attribute A partitions D into n subsets, $\{D_1, D_2, \ldots, D_n\}$, $|D_j|$ is the number of samples in subset $D_j (j = 1, 2, \ldots, n)$, and $\left|S_j^i\right|$ stands for the number of samples in subset D_j that belongs to class S_i. Then, the expected information is defined as:

$$E(A) = \sum_{j=1}^{n} \frac{|D_j|}{|D|} I\left(\left|S_j^1\right|, \left|S_j^2\right|, \ldots, \left|S_j^i\right|\right) \tag{5}$$

Note that the smaller the entropy value is, the higher the purity of the subset partition, where m for a given subset D_j,

$$I\left(\left|S_j^1\right|, \left|S_j^2\right|, \ldots, \left|S_j^i\right|\right) = -\sum_{i=1}^{m} p_{ij} \log_2(p_{ij}) \tag{6}$$

3.2.3. Information Gain and Gain Ratio

The Information Gain of feature/attribute A is expressed as follows:

$$Gain(A) = I(|S_1|, |S_2|, \ldots, |S_m|) - E(A) \tag{7}$$

Table 6
Training set with the principle of standardization (partially).

No.	X	Y1			Y2			Y3			Y4			Y5			Y6		
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
2	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
3	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
4	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
5	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
6	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
7	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
8	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
9	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
10	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
11	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
12	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
13	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
14	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
15	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
16	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
17	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
18	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
19	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
20	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
21	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
22	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
23	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
24	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
25	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
26	X14	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
27	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
28	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
29	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
30	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
…	…				…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

The split information is defined as:

$$Split(A) = -\sum_{j=1}^{n}\frac{|D_j|}{|D|}\log_2\left(\frac{|D_j|}{|D|}\right) \tag{8}$$

where $Split(A)$ is the information generated by partitioning D based on the values of A; it indicates the outcome of the test rather than the class to which the sample belongs.

The Gain Ratio could be calculated by the following:

$$GainRatio(A) = \frac{Gain(A)}{Split(A)} \tag{9}$$

3.2.4. Constructing the C4.5 decision tree

C4.5 is an extension of ID3 and was presented by Quinlan (1993). ID3 selects the attribute with the largest Information Gain value as the node of the tree, as also shown by Xue et al. (2019). However, C4.5 introduces the concept of Information Gain Ratio and selects the attribute with the largest Information Gain Ratio. Moreover, each possible value is used as a branch of this node to recursively form a decision tree, In addition, C4.5 adds significant functions compared to ID3, such as rules generation, uncertainty processing functions and attribute discretization. C4.5 overcomes the shortcomings of the ID3 algorithm using Information Gain to select attributes when biasing the selection of more attributes and can build a decision tree with as simple a structure as possible while ensuring the accuracy of training set classification. Algorithm 1 depicts the procedures of the process of construction of the proposed maneuvering C4.5 decision tree of intelligent ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model.

Algorithm 1 Construct the proposed C4.5 decision tree of HDMDR model.

Input: The training dataset D of the maneuvering factor (X) and environmental factors (Y1~Y6 in our case; new factors can be upgraded here); attribute A.
Output: A proposed maneuvering C4.5 decision tree.

for every attribute A **do**
    Calculate the Information Gain Ratio for using A to splitting D;
end
if $GainRatio > threshold$ **then**
    return A degenerated tree with only one node
end
Construct a root node with the selected environmental factor;
for every subtree **do**
    Move all samples belonging in the subtree to a continuous memory area;
    Recursively call C4.5 to construct the subtrees, using the subset of training samples as its training set;
end

3.2.5. Pruning the decision tree

The initial construction of the C4.5 decision tree is often complicated by the inclusion of a large number of classification attributes and branches, and there are inevitably some errors, namely, noise. This noise gradually accumulates in the decision classification process, which will eventually cause the C4.5 decision tree to have a large deviation from the classification of the actual sample, and the accuracy is reduced, i.e., over-fitting. Thus, the C4.5 decision tree generated by the training set is very good for classifying the training set, but it may not be ideal to use it to classify the new data set that does not participate in the decision tree generation process. Therefore, the preliminary constructed C4.5 decision tree needs to be pruned, and the purpose of pruning is to optimize the C4.5 decision tree or simplify the generated rules. There are two kinds of decision tree pruning methods: prepruning and postpruning.

For the problem of over-fitting, this study uses postpruning methods to eliminate branching anomalies caused by noise data and isolated points. Quinlan (1993) proposed using pessimistic error pruning to compensate for optimistic bias in tree generation during pruning (Due to the decision tree is generated from the training data set, in most cases, the decision tree is consistent with the training data set. However, when the decision tree is used to classify data other than the training data, it is obvious that the error rate will be greatly increased).

The postpruning rule adopts the principle of minimum expected error rate, i.e., starting from the root node of the tree, and calculating the expected error rate that may occur for each branch node pruning/no pruning: If the node is clipped, resulting in a higher expected error rate, the subtree is retained. Otherwise, the subtree is clipped, and finally, the C4.5 decision tree with the smallest expected error rate is obtained.

In this paper, the upper limit of the quality confidence interval is used as the erroneous estimation under pessimistic conditions. Given a confidence level α (0.25 in the C4.5 algorithm), the total number of errors obeys the Bernoulli distribution; then, there is a probability equation:

$$P\left[\frac{|f-q|}{\sqrt{q(1-q)/N}} > \mu_{1-\alpha}\right] = \alpha \tag{10}$$

where N is the total number of instances under the pruned subtree, E is the number of error instances that occur after pruning, $f = E/N$ is the actual observed error rate, and q is the estimated error rate. Let $z = \mu_{1-\alpha}$, taking the upper bound of the confidence interval as the pessimistic error rate estimate of this node. Then, the formula for calculating the false positive rate of the node:

$$q = \frac{f + \frac{z^2}{2N} + z\sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4N^2}}}{1 + \frac{z^2}{N}} \tag{11}$$

Where $f = E/N$ is the actual observed error rate, and q is the estimated error rate.

Set the maximum value of the expected false positive rate to C. If the estimated false positive rate q after pruning is higher than C, the original subtree is retained. Otherwise, the subtree is cut and replaced with leaves. After the pruning, the inbound human-like decision-making tree is shown in Fig. 6. Fig. 2 is the basic process and framework for our proposed HDMDR model.

4. Experiments

4.1. Scenario design and data collection

In our experiment, the simulator scenario was the Shanghai Waigaoqiao wharf, and the ship was downstream of the berthing into the port. We use a 30,000-ton bulk carrier as our experimental ship OS1 (33089.0 t, 182.9 m long, 22.6 m wide). We define the process as when the ship's stern leaves the main channel near the port side of the boundary line in the electronic chart (Fig. 3(b) shows the initial boundary) to the ship berths docked at the end of the cable (Fig. 3(c) shows the end boundary) as a complete berthing process. The experimental scenario is shown in Fig. 3.

We collect the data from the full-task handling simulation platform (Navi-Trainer Professional 5000, which conforms to the IMO STCW78/10 convention and the Det Norske Veritas (DNV)) from the Maneuvering Simulator Laboratory in Wuhan University of Technology Waterway Road Traffic Safety Control and Equipment Ministry of Education Engineering Research Center.

We collect the operational data of the exercises and assessment exams as our experimental data (unlimited navigational class crew, 4 groups of 96 people, 32–45 years old, skilled maneuvering level, captain/chief officer). From Table 2, we can get the average age of the crew participating in this experiment is 38.76 years old and their average piloting experience is 8.89 years. The captains' average age and piloting age are both higher than those of chief officers'. From Fig. 4, we can get the distribution of OOWs' age and their piloting experience. The ship handing and environment, including inside and outside multisource information, were collected on the ship's berthing process, including the environment (wind, current, wave, etc.), control (rudder order, marine telegraph order - 2 factors). Table 3 lists some of the training samples.

It should be noted that, in our case, the OOW is the captain or chief officer, although, in the real situation, the captain is not on duty. The captain will go to the bridge only in special circumstances, and if necessary, the captain may take over the duty of the OOW to maneuver the ship, but it is an assessment and evaluation scenario in our experiment; therefore, the captain also acts as the OOW. In addition, we regard the tugboat as a power plant system of target ship OS1 to facilitate the ship's overall situation of a simplified analysis and we consider the tugs and the ship OS1 as a whole dynamic model. Under the premise of this hypothesis, the ship OS1 completes the inbound operation through the combination of rudder orders and telegraph orders, according to the actual navigational situation of its force and movement.

4.2. Standardization principle setting

Maneuvering decision-making processes are often influenced by multisource information, such as human, ship and environmental factors. These influencing factors act together to determine the next action strategy of the ship's OOW.

For a particular person-ship unit, the overall reliability is constant for a certain period of time or during a trip; therefore, the person and ship factors have less influence on maneuvering decisions. With the operation of the ship, the OOW's waterway and the environment will change with time and space, and the changing waterway and environmental factors will have a greater impact on maneuvering decisions. In this research, we mainly focus on the environmental influencing factors and study their effect on the decision-making of the OOW. Based on the strategy and the current maneuvering environment, the experienced OOW can quickly and accurately make maneuver decisions, thus laying the foundation for the study of human-like maneuvering behavior for the application to intelligent ships. We select six environmental influencing factors as the input of our proposed HDMDR model to study the decision-making mechanisms for different maneuvering behaviors.

In order to let the maneuvering decision-making knowledge to be automatically obtained and expressed along with higher decision-making knowledge effectiveness. It is typically necessary to divide the number of linguistic terms by experience (Yuan and Shaw, 1995). In this paper, experimental data of each maneuvering decision-making factor are trisected into three levels, namely, small (a), medium (b), and large (c), see Table 4, to objectively describe the characteristics of each influencing factor, and make it easier to describe how each factor influences final maneuvering decisions. We select six environmental influencing factors as the input of our proposed model to study the decision-making mechanisms for different maneuvering behaviors: Current direction, current speed, relative current direction, relative wave direction, relative wind direction, relative wind speed (In other cases, the other new factors can also be upgraded according to Algorithm 1 in section 3.2.4 using specific standardization principle).

The OOW maneuvers the ship by operating different telegraph and rudder orders to change ship's speed and direction and to complete the ship's control. Table 5 shows the combining telegraph and rudder orders (speed and course control respectively); this control is a multidynamic process. Moreover, it should be noted that, in combination with the actual situation of the experimental scenario, unlike the ship sailing on the open sea, the OOW needs to call the rudder and telegraph orders frequently in the inbound decision-making ship handing process; therefore, in this paper, we do not consider "Midships" and "Stop engine," regardless of the rudder angle and if the power output is 0. Table 5 shows the standardization principle for output maneuvering decision-making factors.

5. Results and discussion

5.1. Standardizing of training set

The data in Table 3 are standardized according to the principle of standardization of maneuvering decision influencing factors in Tables 4 and 5; the results are shown in Table 6.

5.2. Constructing and pruning the decision tree

The C4.5 algorithm can be divided into two phases. First, a certain attribute is selected according to the criterion of maximum Information Gain to divide the training set, and the recursive call is performed until all the examples in each division belong to the same class; then, the established tree is pruned, i.e., the branch established above the noise data is cut. In the decision tree analysis, approximately 80% of the data is randomly selected as the training set, and the remaining 20% is used as the test set. Then, through Eqs. (3)–(11), we could obtain the decision tree structure, as shown in Fig. 6, partitioned into 3 parts, Part I, II and III. The number and proportion of different decisions are shown in Fig. 5.

5.3. Establishing maneuvering decision classification rules

The result of our proposed HDMDR model is a set of classification rules in the form of IF-THEN. Each path from the root node to the leaf node constitutes a rule. The characteristics of the internal nodes of the path correspond to the conditions of the rule, and the classification of the leaf nodes corresponds to the conclusion of the rule. As a result, we can easily extract the human-like decision-making knowledge using the decision tree and rule set. The optimized maneuvering decision

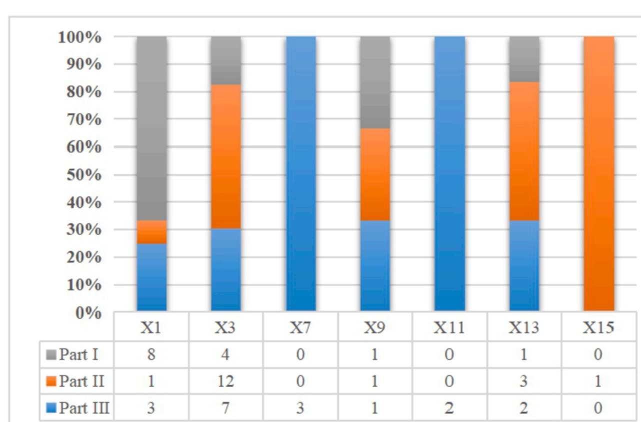

Fig. 5. The number and proportion of different decisions.

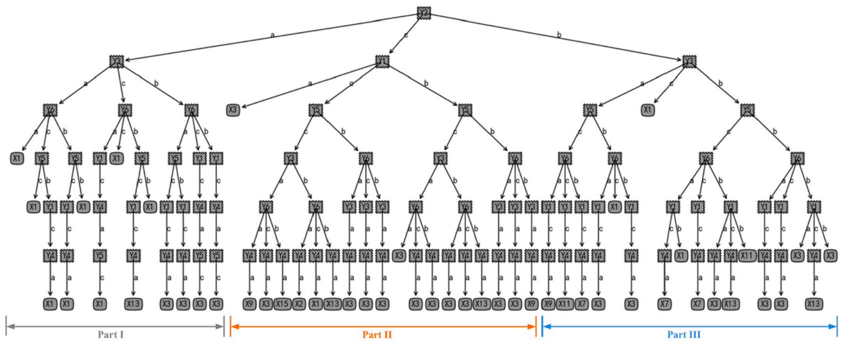

Fig. 6. The decision tree structure.

recognition rule set is shown in Table 7.

From Figs. 5 and 6, and Table 7, we can draw the following conclusions:

1) It can be seen from the generated decision tree that Y2 as the root node, i.e., current speed, is the most informative attribute of all samples. In other words, in the environment of the simulation experimental scenario, the current speed in the environmental influencing factors has the most significant impact on the OOW's maneuvering decision-making, followed by the relative current direction and current direction.
2) The ordering of environmental factors provides the OOW with a set of variables for decision-making reference, which has certain guiding significance for the formulation of maneuvering decisions.
3) Through the analysis of the rule set, this designed scenario outputs a number of standardized maneuvering decision operations: X1 (12/50) and X3 (23/50), X1 (U1D1U2T2) maneuvering decision knowledge can be interpreted and conceptualized into linguistic term or operation order: {Keep the propeller-ahead and keep the current rudder angle-port}, the same, X3 (U1D1U2D2): {Keep the propeller-ahead and keep the current rudder angle-starboard}, which is consistent with actual ship maneuvering experience.
4) The specific factors and decision rule sets in a specific scenario obtained by the HDMDR model proposed in this paper can be used as an important reference for the intelligent ship human-like decision-making and can also be used to create a knowledge base of expert systems. It has a high reference value and practical value for the development of intelligent ship's maneuvering algorithm.

Table 7
Maneuvering decision classification rule set.

No.	Maneuvering decision classification rule set
1	*IF Y2=a AND Y3=a AND Y6=a THEN X=X1*
2	*IF Y2=a AND Y3=a AND Y6=b AND Y5=b THEN X=X1*
3	*IF Y2=a AND Y3=a AND Y6=b AND Y5=c AND Y1=c AND Y4=a THEN X=X1*
4	*IF Y2=a AND Y3=a AND Y6=c AND Y5=c THEN X=X1*
5	*IF Y2=a AND Y3=a AND Y6=c AND Y5=b AND Y1=c AND Y4=a THEN X=X1*
6	*IF Y2=a AND Y3=b AND Y6=a AND Y5=b/c AND Y1=c AND Y4=a THEN X=X3*
7	*IF Y2=a AND Y3=b AND Y6=b/c AND Y1=c AND Y4=a AND Y5=c THEN X=X3*
8	*IF Y2=a AND Y3=c AND Y6=a AND Y1=c AND Y4=a AND Y5=c THEN X=X1*
9	*IF Y2=a AND Y3=c AND Y6=b AND Y5=b THEN X=X1*
10	*IF Y2=a AND Y3=c AND Y6=b AND Y5=c AND Y1=c AND Y4=a THEN X=X13*
11	*IF Y2=a AND Y3=c AND Y6=c THEN X=X1*
12	*IF Y2=b AND Y3=a AND Y5=b AND Y6=a AND Y1=c AND Y4=a THEN X=X9*
13	*IF Y2=b AND Y3=a AND Y5=b AND Y6=b AND Y1=c AND Y4=a THEN X=X3*
14	*IF Y2=b AND Y3=a AND Y5=b AND Y6=c THEN X=X1*
15	*IF Y2=b AND Y3=a AND Y5=c AND Y6=a AND Y1=c AND Y4=a THEN X=X9*
16	*IF Y2=b AND Y3=a AND Y5=c AND Y6=b AND Y1=c AND Y4=a THEN X=X7*
17	*IF Y2=b AND Y3=a AND Y5=c AND Y6=c AND Y1=c AND Y4=a THEN X=X11*
18	*IF Y2=b AND Y3=b AND Y5=b AND Y6=a AND Y1=c AND Y4=a THEN X=X3*
19	*IF Y2=b AND Y3=b AND Y5=b AND Y6=b AND Y1=a/b THEN X=X3*
20	*IF Y2=b AND Y3=b AND Y5=b/c AND Y6=b AND Y1=c AND Y4=a THEN X=X13*
21	*IF Y2=b AND Y3=b AND Y5=b AND Y6=c AND Y1=c AND Y4=a THEN X=X3*
22	*IF Y2=b AND Y3=b AND Y5=c AND Y6=a AND Y1=b THEN X=X1*
23	*IF Y2=b AND Y3=b AND Y5=c AND Y6=a/c AND Y1=c AND Y4=a THEN X=X7*
24	*IF Y2=b AND Y3=b AND Y5=c AND Y6=b AND Y1=a AND Y4=a THEN X=X3*
25	*IF Y2=b AND Y3=b AND Y5=c AND Y6=b AND Y1=b THEN X=X11*
26	*IF Y2=b AND Y3=c THEN X=X1*
27	*IF Y2=c AND Y1=a THEN X=X3*
28	*IF Y2=c AND Y1=b AND Y5=b AND Y6=a/c AND Y3=a AND Y4=a THEN X=X3*
29	*IF Y2=c AND Y1=b AND Y5=b AND Y6=b AND Y3=a AND Y4=a THEN X=X9*
30	*IF Y2=c AND Y1=b AND Y5=c AND Y3=a AND Y6=a THEN X=X3*
31	*IF Y2=c AND Y1=b AND Y5=c AND Y3=a AND Y6=b/c AND Y4=a THEN X=X3*
32	*IF Y2=c AND Y1=b AND Y5=c AND Y3=b AND Y6=a AND Y4=a THEN X=X3*
33	*IF Y2=c AND Y1=b/c AND Y5=c AND Y3=b AND Y6=b AND Y4=a THEN X=X13*
34	*IF Y2=c AND Y1=b AND Y5=c AND Y3=b AND Y6=c AND Y4=a THEN X=X3*
35	*IF Y2=c AND Y1=c AND Y5=b AND Y3=a AND Y4=a THEN X=X3*
36	*IF Y2=c AND Y1=c AND Y5=c AND Y3=a AND Y6=a AND Y4=a THEN X=X9*
37	*IF Y2=c AND Y1=c AND Y5=c AND Y3=a AND Y6=b AND Y4=a THEN X=X15*
38	*IF Y2=c AND Y1=c AND Y5=c AND Y3=a AND Y6=c AND Y4=a THEN X=X3*
39	*IF Y2=c AND Y1=c AND Y5=c AND Y3=b AND Y6=a AND Y4=a THEN X=X2*
40	*IF Y2=c AND Y1=c AND Y5=c AND Y3=b AND Y6=c AND Y4=a THEN X=X1*

5.4. Performances assessment

5.4.1. Applying rules for classification

We use the maneuvering decision-making model proposed in this paper to identify the decision-making data to be identified in Table 8. We compare the recognition results with the actual ship maneuvering decisions and use the accuracy of the recognition to verify the validity of the model. The standardized maneuvering decision-making data are identified in Table 8, using classification rules 33, 29, and 37, and the recognition result is X13, X13, X13, X9, X9, and X15. This result is consistent with actual maneuvering decisions and demonstrates high reasoning efficiency.

The test data set was evaluated and validated using the generated decision tree model. There were 135531 samples participating in the test, accounting for 20% of the overall data set. To assess the accuracy of the HDMDR model, the data in the test data set is used for prediction, and the degree of agreement between the test results and the actual situation is compared. The accuracy of the proposed module (ACC) could be calculated as:

$$ACC = \frac{TN + TP}{TN + TP + FN + FP} \tag{12}$$

where TN is true negatives, TP is true positives, FN is false negatives, and FP is true positives.

The classification accuracy of our proposed HDMDR model using C4.5 decision trees based on the test data set can reach more than 81.6%.

Table 8
Maneuvering decision data to be identified and its standardization.

Maneuvering Decision Data to Be Identified

No.	X (Actual Maneuvering Decision)		Y1	Y2	Y3	Y4	Y5	Y6
	Rudders Order	Telegraphs Order						
1	−35.0000	16.3207	315.3000	1.0802	−65.1521	−1.2042	121.7873	9.5846
2	−35.0000	18.9076	315.3000	1.0802	−62.0192	−0.9383	120.5850	9.5745
3	−35.0000	20.0000	315.3000	1.0802	−60.9662	−0.8662	119.8690	9.5714
4	−10.0000	20.0000	315.3000	1.0802	−59.8343	−0.6030	117.7910	9.5586
5	−5.0000	20.0000	316.5000	1.0802	−59.7830	−0.4652	116.2045	9.5551
6	10.0000	5.3301	317.2000	1.0802	−59.5314	0.0373	118.1805	9.5551

Standardized Maneuvering Decision Data to Be Identified

No.	X		Y1			Y2			Y3			Y4			Y5			Y6		
	Rudders Order	Telegraphs Order	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	−35.0000	16.3207	0	1	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0
2	−35.0000	18.9076	0	1	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0
3	−35.0000	20.0000	0	1	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0
4	−10.0000	20.0000	0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0
5	−5.0000	20.0000	0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	0	1	0
6	10.0000	5.3301	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	0	1	0

5.4.2. *Comparative analysis*

To further validate the effectiveness of the HDMDR model, in this paper, we compare the performance of the proposed C4.5 decision tree algorithm with two classic classification algorithms: k-Nearest Neighbours (k-NN) and Support Vector Machine (SVM). In our case, we use the Radial Basis Function (RBF) to conduct the SVM and $k = 1$ in the K-NN. Besides, we use classification accuracy, shown as Eq. (12), to measure the proposed C4.5 algorithm. In addition, in this paper, the code for the basic versions of k-NN and SVM classifiers is adopted from the Waikato Environment for Knowledge Analysis (WEKA), which is open source data mining software (Hall et al., 2009). WEKA is a comprehensive software that implements many state-of-the-art machine learning and data mining algorithms.

We conduct a ten-fold cross-validation (10-CV) experiment using the data from training set. 10-CV breaks data into ten sets equally, then trains the classifier on nine data sets and uses it to test the remaining one data set. Repeating ten times like this, and finally taking an average accuracy, thus to compare the performance of the proposed C4.5 decision tree algorithm with k-NN and SVM. The performance of different classifier algorithms on our data set is shown in Table 9 and Fig. 7. According to the classification accuracy results, the proposed method can achieve the highest accuracy among these three algorithms.

6. Conclusions

With the development of the economy, the continuous advancement of technology, and the continuous increase of labor costs, it has become an urgent trend to realize intelligent maneuvering of ships. The purpose of this research is to recognize the automatic acquisition and representation of the OOW's decision-making knowledge and to provide a basis and reference for the development of decision-making algorithms for intelligent ships.

In this paper, a intelligent ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model and a novel standardization principle of maneuvering decision-making factors are proposed for the learning of human-like decision-making mechanisms of intelligent ships. By establishing an autonomous learning method of maneuvering decision-making, the processes of autonomous learning OOWs' maneuvering decision-making characteristics are studied. In addition, it is unique and very valuable to obtain experimental data operated by an experienced OOW on the full-task handling simulation platform in a certain time and space. To validate the performance and effectiveness of our proposed model, the assessment of applying rules for classification and the comparative analysis with the k-NN and SVM are compared. According to the results, the classification accuracy of our proposed HDMDR model can reach more than 81.6%. In addition, the proposed method is superior to the representative classification algorithms.

This study provides a new perspective and methodology for the development of intelligent ship maneuvering decision-making technology in theory and practice, promotes the application and spreading of intelligent ships under specific scenarios, and is conducive to the development of water transportation in the direction of safety, sustainability and economy.

Nevertheless, the HDMDR model still has some shortcomings which need to be improved in further research.

1) The proposed method/model is a data-driven method. We need more data to further train the machine learning model and improve recognition accuracy. Besides, the feedback loop to inform the effect of this model still need to be optimized.
2) The application scenarios of the proposed model still need to be enriched, and more environmental influencing factors which may affect piloting decisions need to be added as well considering the specific situation, thus to make the proposed model more widely applicable.
3) The standardization principle of maneuvering decision-making attributes need to be further detailed according to the actual navigation situation, and more suitable for the real-world ship-handing

Table 9
The performance of different classifier algorithms with 10-fold cross-validation.

Classifier Algorithms	Accuracy (performance measures in %)										
	F 1	F2	F3	F4	F5	F 6	F7	F 8	F 9	F 10	Average
k-NN	75.36	73.79	73.81	74.62	72.87	76.86	72.37	75.72	74.89	75.66	74.60
SVM	70.26	72.62	75.43	73.62	77.79	72.83	70.29	71.63	74.13	73.09	73.17
Proposed method	80.33	79.88	83.42	76.59	79.16	83.76	79.78	82.56	81.86	78.43	80.58

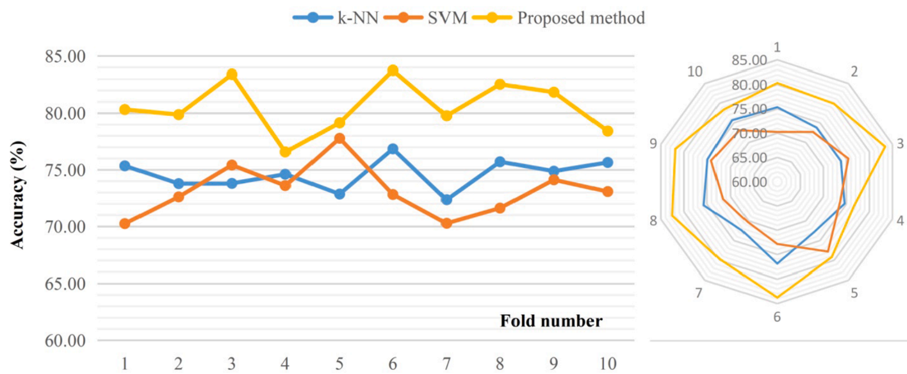

Fig. 7. The accuracy of different classifier algorithms.

orders (specifically the combined rudder orders and telegraph orders), thus further increasing the applicability of the model.

In the subsequent research, we will study the classification of the influencing factors, the fuzzy processing of data sets, the detailed connection of our model with human behavior and their performance in specific navigational scenarios, the application of multi-navigation scenarios, etc. Besides, we will further collect the relevant data from the tugs and add the data to the model analysis, to further optimize our proposed algorithm.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (51775396, 61703319, U1764262), the Major Project of Technological Innovation of Hubei Province (2016AAA007, 2017CFA008), and the China Scholarship Council.

References

Aguiar-Pulido, V., A Seoane, J., Gestal, M., Dorado, J., 2013. Exploring patterns of epigenetic information with data mining techniques. Curr. Pharmaceut. Des. 19, 779–789.
Akyuz, E., Celik, M., 2014. A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. Saf. Sci. 68, 169–179.
Angluin, D., 1988. Queries and concept learning. Mach. Learn. 2, 319–342.
Baksh, A.-A., Abbassi, R., Garaniya, V., Khan, F., 2018. Marine transportation risk assessment using Bayesian Network: application to Arctic waters. Ocean Eng. 159, 422–436.
Borak, J.S., 1999. Feature selection and land cover classification of a MODIS-like data set for a semiarid environment. Int. J. Remote Sens. 20, 919–938.
Bujlow, T., Riaz, T., Pedersen, J.M., 2012. A Method for Classification of Network Traffic Based on C5. 0 Machine Learning Algorithm. IEEE, pp. 237–241.
Calistru, I.T., Cotofrei, P., Stoffel, K., 2015. A Parallel Approach for Decision Trees Learning from Big Data Streams. Springer, pp. 3–15.
Cherfi, A., Nouira, K., Ferchichi, A., 2018. Very fast C4.5 decision tree algorithm. Appl. Artif. Intell. 32, 119–137.
Dai, W., Ji, W., 2014. A MapReduce implementation of C4.5 decision tree algorithm. Int. J. Database Theory Appl. 7, 49–60.
Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth & Brooks, Belmont, CA.
García-Laencina, P.J., Abreu, P.H., Abreu, M.H., Afonoso, N., 2015. Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. Comput. Biol. Med. 59, 125–133.
Goerlandt, F., Montewka, J., 2015. Maritime transportation risk analysis: review and analysis in light of some foundational issues. Reliab. Eng. Syst. Saf. 138, 115–134.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software:an update. Acm Sigkdd Explor. Newsl. 11 (1), 10–18.
Hanzu-Pazara, R., Barsan, E., Arsenie, P., Chiotoroiu, L., Raicu, G., 2008. Reducing of maritime accidents caused by human factors using simulators in training process. J. Marit. Res. 5 (1), 3–18.
Hssina, B., Merbouha, A., Ezzikouri, H., Erritali, M., 2014. A comparative study of decision tree ID3 and C4. 5. Int. J. Adv. Comput. Sci. Appl. 4.
Hunt, E.B., Marin, J., Stone, P.J., 1966. Experiments in Induction. Academic Press, Oxford, England.
Kass, G.V., 1975. Significance testing in automatic interaction detection (AID). Appl. Stat. 178–189.
Kheradpisheh, S.R., Ganjtabesh, M., Thorpe, S.J., Masquelier, T., 2018. STDP-based spiking deep convolutional neural networks for object recognition. Neural Network. 99, 56–67.
Lakkakula, N.P., Naidu, M.M., Reddy, K.K., 2014. An Entropy Based Elegant Decision Tree Classifier to Predict Precipitation, pp. 11–19.
Lazarowska, A., 2017. A new deterministic approach in a decision support system for ship's trajectory planning. Expert Syst. Appl. 71, 469–478.
Lim, T.S., Loh, W.Y., Shih, Y.S., 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Mach. Learn. 40, 203–228.
Lu, Z., Wu, X., Bongard, J.C., 2015. Active learning through adaptive heterogeneous ensembling. IEEE Trans. Knowl. Data Eng. 27, 368–381.
Lun, Y.V., Lai, K.-h., Wong, C.W., Cheng, T., 2016. Introduction to Green Shipping Practices, Green Shipping Management. Springer, pp. 3–15.
Muchoney, D., Borak, J., Chi, H., Friedl, M., Gopal, S., Hodges, J., Morrow, N., Strahler, A., 2010. Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America. Int. J. Remote Sens. 21, 1115–1138.
Pandya, R., Pandya, J., 2015. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. Int. J. Comput. Appl. 117, 18–21.
Peng, Z.R., Wang, Z., He, H.D., Lu, F., Lu, Q.C., 2015. A hybrid model for prediction of carbon monoxide and fine particulate matter concentrations near Road intersection. Transport. Res. Rec. J. Transport. Res. Board 2503, 29–38.
Perera, L.P., Ferrari, V., Santos, F.P., Hinostroza, M.A., Guedes Soares, C., 2015. Experimental evaluations on ship autonomous navigation and collision avoidance by intelligent guidance. IEEE J. Ocean. Eng. 40, 374–387.
Prasad, N., Kumar, P., Mm, N., 2013. An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree, pp. 56–60.
Prasad, N., Naidu, M.M., 2013. Gain Ratio as Attribute Selection Measure in Elegant Decision Tree to Predict Precipitation, pp. 141–150.
Provost, F., Domingos, P., 2003. Tree induction for probability-based ranking. Mach. Learn. 52, 199–215.
Quinlan, J., 1979. In: Interactive Dichotomizer, ID3. Morgan Kauffmann. Springer-Verlag.
Quinlan, J.R., 1986. Induction of decision trees. Mach. Learn. 1, 81–106.
Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco.
Reumers, S., Liu, F., Janssens, D., Cools, M., Wets, G., 2013. Semantic annotation of global positioning system traces. Transport. Res. Rec.: J. Transport. Res. Board 2383, 35–43.
Rodriguez, A.H., Aviles-Jurado, F.X., Diaz, E., Schuetz, P., Trefler, S.I., Sole-Violan, J., Cordero, L., Vidaur, L., Estella, A., Pozo Laderas, J.C., Socias, L., Vergara, J.C., Zaragoza, R., Bonastre, J., Guerrero, J.E., Suberviola, B., Cilloniz, C., Restrepo, M.I., Martin-Loeches, I., Group, S.G.W., 2016. Procalcitonin (PCT) levels for ruling-out bacterial coinfection in ICU patients with influenza: a CHAID decision-tree analysis. J. Infect. 72, 143–151.
Sanil, A.P., 2001. Principles of Data Mining. MIT Press.
Saunier, N., Mourji, N., Agard, B., 2011. Mining microscopic data of vehicle conflicts and collisions to investigate collision factors. Transportation research record. J. Transport. Res. Board 2237, 41–50.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423.
V, N.P.L., Patro, K.R., Naidu, M.M., 2013. A Gini Index Based Elegant Decision Tree Classifier to Predict Precipitation, pp. 46–54.
Wang, J., Li, M., Liu, Y., Zhang, H., Zou, W., Cheng, L., 2014. Safety assessment of shipping routes in the South China Sea based on the fuzzy analytic hierarchy process. Saf. Sci. 62, 46–57.
Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2007. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14, 1–37.
Xie, C., Lu, J., Parkany, E., 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. Transport. Res. Rec.: J. Transport. Res. Board 50–61.

Xue, J., Wu, C., Chen, Z., Van Gelder, P., Yan, X., 2019. Modeling human-like decision-making for inbound smart ships based on fuzzy decision trees. Expert Syst. Appl. 115, 172–188.
Yuan, Y., Shaw, M.J., 1995. Induction of fuzzy decision trees. Fuzzy Sets Syst. 69, 125–139.
Zhang, L.-Y., Peng, Z.-R., Sun, D., Liu, X., 2012. Rule-based forecasting of traffic flow for large-scale Road networks. Transport. Res. Rec.: J. Transport. Res. Board 2279, 3–11.
Zhang, X., Zhang, J., Fan, C., Meng, J., Wang, J., Wan, Y., 2018. Analysis of dynamic characteristics of the 21st century maritime silk Road. J. Ocean Univ. China 17, 487–497.