



Modeling human-like decision-making for inbound smart ships based on fuzzy decision trees



Jie Xue^{a,b,c}, Chaozhong Wu^{a,b,*}, Zhiyun Chen^{a,b,*}, P.H.A.J.M. Van Gelder^c, Xinping Yan^a

^a Intelligent Transportation Systems Research Center(ITSC), Wuhan University of Technology, 1040 Heping Avenue, Wuhan, Hubei 430063, 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

ARTICLE INFO

Article history:

Received 24 April 2018

Revised 25 June 2018

Accepted 19 July 2018

Available online 30 July 2018

Keywords:

Fuzzy decision trees

Piloting decision

Smart ship

Classification rule

Data mining

Information entropy

ABSTRACT

With the further development of marine and information technologies, ship intelligence, green policies and automation will become mainstream with global cargo ships. Ship labor costs increase every year, so for the foreseeable future, the number of experienced crew members will be greatly reduced as smart ship emergence accelerates. At present, there is no mature research system for the human-like piloting of smart ships. In this paper, we use an improved decision tree, which could address problems of fuzziness and uncertainty. This will allow us to study the decision mechanisms of different piloting behaviors in order to realize the automatic acquisition and representation of the pilot's decision-making knowledge in inbound ship analysis as well as the simulated reproduction of the pilot's behavior. The simulation results show that the piloting decision recognition model, based on the fuzzy Iterative Dichotomiser 3 (ID3) decision tree, possesses a high reasoning speed and can accurately identify current piloting behavior. This provides theoretical guidance and a feasibility basis for research into human-like piloting behavior and the realization of automatic smart ship piloting systems.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

With the development of marine technologies, information technologies, and "big data" intelligent applications, smart ship emergence is accelerating. The improvement of smart ships over the next 10–20 years will be an important factor in determining the future direction of the shipping industry. According to Global Marine Technology Trends 2030, co-launched by Lloyd's Register (LR), the Aquinas TEEK group and the University of Southampton, smart ships are listed as one of the 18 key future marine technologies. German Industry 4.0, based on big data, is predicting technique-centric intelligent manufacturing. Through the integration of networks, entities, and "shore-sea integration" intelligent information service systems, it promotes the transformation of traditional manufacturing and the development of smart ships. "Manufacturing in China 2025" views marine engineering equipment and high-tech shipbuilding as one of top ten key areas, in which

smart ships will be an important part. Ship intelligence, green policies and automation will become the mainstream of global cargo ships. With the continued improvement of ship intelligence, the development of unmanned cabin maintenance, auxiliary piloting technology, fault self-diagnosis technology and other technologies will gradually reduce ship labor needs, potentially achieving unmanned operation of a ship. In the foreseeable future, the number of experienced crew members will be greatly reduced, which will lead to increased ship safety requirements. Therefore, it is necessary to strengthen the relevant theoretical and technical research. The study of manned piloting mechanisms and piloting behavioral decision-making is indispensable and meaningful.

The accuracy of piloting decisions is directly related to the safety of water traffic. In the process of decision-making, piloting behavior is often influenced by multi-source information, such as crew, the ship and the environment. Due to the limited information processing capacity, the pilot cannot concurrently achieve knowledge acquisition and representation of the multi-source information so that piloting decisions can be carried out accurately and quickly, which can lead to water traffic accidents. The decision mechanisms of different piloting behavioral patterns and the execution mechanisms of ship operating modes are two important steps in simulating task aggregation and multi-source information

* Corresponding authors at: Intelligent Transport Systems Research Center, Wuhan University of Technology, 1040 Heping Avenue, Wuhan, Hubei 430063, China.

E-mail addresses: J.Xue@tudelft.nl (J. Xue), wucz@whut.edu.cn (C. Wu), chenjz556@whut.edu.cn (Z. Chen), P.H.A.J.M.vanGelder@tudelft.nl (P.H.A.J.M. Van Gelder).

stimulation (Wang & Yang, 2006; Zheng, Zhang, Huang, & Li, 2012; Wang & Yang, 2008; Xiao, Ligteringen, Van Gulijk, & Ale, 2015). Therefore, the automatic acquisition and representation of piloting decision-making is essential in ensuring accurate and rapid piloting decisions and water traffic safety. The selection of different evaluation methods will affect the pilot's decision-making efficiency and accuracy, thus affecting water traffic safety. At present, the methods of knowledge acquisition and representation are mainly based on knowledge-based research methods, such as fuzzy theory (Chen, Liu, & Tsai, 2000), support vector machines (Tsang, Kwok, & Cheung, 2005), statistical analysis (Chen, Wu, Zhong, Lyu, & Huang, 2015), rule extraction (Martinez, Webb, Chen, & Zaidi, 2016; Moradi & Keyvanpour, 2015), neural networks (Liang, Zhiqiang, & Long, 2012) sparse representation (Chen et al., 2017), etc. However, there is no unified, comprehensive theoretical system, and there are shortcomings in the evaluation methods. For instance, the support vector machine method can solve optimization problems, but it needs to compute Hesse or the inverse matrix, and the time complexity is high. Meanwhile, the storage space and computation time requirements are large when the number of training samples is high. While a neural network can achieve knowledge acquisition and representation, the model exhibits over-fitting and under-fitting phenomena. The knowledge is implied, not easily tested and has poor flexibility. Any changes in the system must be re-learned, so learning convergence can be slow. Therefore, the research on ship piloting behaviors and decision-making methods needs to be improved and further developed.

Ship piloting behavior decision-making studies are a classification of the ship's operating behavior in accordance with certain rules. A decision tree is a classification method of data mining that can potentially find valuable information by classifying a large amount of data. It has the advantages of simple descriptions, fast classifications, and being suitable for large-scale data processing. It can learn from the sample, obtain classification rules, and classify the samples according to these rules. Decision tree methods can overcome the previously mentioned defects. They integrate knowledge representation and acquisition with a simple and intuitive form. This is convenient for expert testing and has higher reasoning efficiency. Therefore, it is feasible and reasonable to apply the decision tree classification method to the decision-making of ship piloting behavior. At present, the commonly used decision tree classification algorithms include the Iterative Dichotomiser 3 (ID3) algorithm, C4.5 algorithm, Classification and Regression Trees algorithm (CART) algorithm, etc. (Wang & Jiang, 2011). In these algorithms, the C4.5 algorithm is complex when continuously processing data, and its workload is large. The CART algorithm is a statistical analysis method appropriate for large samples but is not applicable when processing small sample sizes. The ID3 algorithm is the most influential decision tree generation algorithm. It chooses the attribute with the highest information gain as the test attribute of the current node. It divides the sample set based on the value of the test attribute, how many different values of the test attribute exist, the number of subset divisions, and then further divides the corresponding subset of the sample using a recursive method. However, the decision tree construction algorithms above are all based on the assumption that the attribute and classification values are clear, so these algorithms cannot address the uncertainties related to human thinking and behavior. Quinlan (1986) noted that while classification results of a decision tree are clear, it cannot address potential uncertainty during the classification process. When the attribute value has a slight change, mutations can inappropriately affect the classification results. The resulting decision tree generally is not robust, and inaccurate or missing data can prevent in the decision tree growing phase (Kantardzic, 2011). As a data mining method, the Fuzzy Decision Tree (FDT) is an extension of the classical decision tree. It integrates the advantages of

fuzzy theory and decision trees by combining the comprehensibility of decision trees and the comprehensive expressions of fuzzy technology. The FDT has strong decision-making abilities and can address the problems of ambiguity and uncertainty. Therefore, the decision tree is more robust, its comprehensibility is improved, and the expansion of the algorithm is enhanced (Janikow, 1998; Olaru & Wehenkel, 2003).

In view of this, in this paper, we collect data on the full-task handling simulation platform for large-scale ships named Navi-Trainer Professional 5000. We use the fuzzy ID3 decision tree to study the decision-making mechanisms of different piloting behaviors in order to realize the automatic acquisition and representation of a pilot's decision-making. This will overcome the shortcomings of phase separation between representation and acquisition. We use parameters α and β to control tree generation and carry out pre-pruning. We take the average of the optimal interval of the FDT, the significance level α , and truth level threshold β . This method can identify the current piloting behavior accurately and has high reasoning efficiency, which provides theoretical guidance and feasibility bases for the simulation and realization of smart ship automatic piloting systems.

The remaining contents of this paper are organized as follows. First, the piloting decision-making model and the optimization methodology are given. Then, the experimental processes, the method of classification interval division, and the standardization principle of piloting decision-making factors are introduced. Finally, the performance of the model and optimization methodology are shown in the next part, and we end with conclusions.

2. Piloting decision-making model

2.1. Grey relation entropy model

The grey relation analysis method is based on the degree of dissimilarity or similarity of target system to measure the correlation degree between factors or factors and system behaviors (Deng, 1989; Zhang, Guo, & Deng, 1996). The grey relation entropy analysis method is based on the grey relation analysis method. By using this method, it could avoid the loss when the local node correlation value controlling the tendency of the whole grey correlation in determining the grey correlation degree (Deng, 1990). Therefore, it can distinguish the impacts of major factors and secondary factors on the whole system more effectively.

2.1.1. Grey relation grade

Let X be grey relation factor set (discrete series), $X_0 = \{x_0(k)|k=1,2,\dots, m\}$ as reference columns and $X_i = \{x_i(k)|k=1,2,\dots, m\}(i=1,2,\dots, n)$ as comparison columns. Due to the inconsistent dimension of various factors, X_0 and X_i need to be standardized. Then we get the sequences X'_0 and X'_i , as shown in Eqs. (1) and (2):

$$X'_0 = \left\{ x_0(k) - \frac{1}{m} \sum_{k=1}^m x_0(k) / S_0 | k = 1, 2, \dots, m \right\} \quad (1)$$

$$X'_i = \left\{ x_i(k) - \frac{1}{m} \sum_{k=1}^m x_i(k) / S_i | k = 1, 2, \dots, m \right\} \quad (2)$$

which is a standardized matrix for evaluation problems consisting of n objects and m indicators.

Among them, S_i is the standard deviation of the sequence X'_i . the correlation coefficient of X'_i to X'_0 is:

$$\xi_i(x_0(k), x_i(k))$$

$$= \frac{\min_i(\min_k |x_0(k) - x_i(k)|) + \rho \max_i(\max_k |x_0(k) - x_i(k)|)}{|x_0(k) - x_i(k)| + \rho \max_i(\max_k |x_0(k) - x_i(k)|)} \quad (3)$$

among them, $\min_i(\min_k |x_0(k) - x_i(k)|)$ is the minimum difference of two levels, and $\max_i(\max_k |x_0(k) - x_i(k)|)$ is the maximum difference of two levels. ρ is a resolution ratio, in $(0, 1)$, if ρ is small, the greater the difference between the relationship coefficient, the stronger the ability to distinguish, and ρ usually takes a value of 0.5 (Wang, Wang, & Ai, 2014).

2.1.2. Grey relation entropy

Let X be grey relation factor set (discrete series), $X_0 = \{x_0(k) | k = 1, 2, \dots, m\}$ as reference columns and $X_i = \{x_i(k) | k = 1, 2, \dots, m\} (i = 1, 2, \dots, n)$ as comparison columns. $R_i = \{\xi_i(x_0(k), x_i(k)) | k = 1, 2, \dots, m\}$, so the grey correlation coefficient distribution map is called the density value of the distribution, as shown in Eq. (4):

$$P_i \triangleq \frac{\xi_i(x_0(i), x_i(i))}{\sum_{k=1}^m \xi_i(x_0(i), x_i(i))} \quad (4)$$

The grey relation entropy of X_i is expressed as:

$$H(R_i) \triangleq - \sum_{k=1}^m P_i \ln P_i \quad (5)$$

From the entropy law, we can see that when the grey entropy of sequence X_i is the largest, it means that the influence of X_i points on the reference column is equal, which indicates that the distance between X_i and the reference column is more balanced, so X_i is closer to the reference column geometry, and X_i is the strongest associated column. From the grey entropy theorem, the entropy correlation Eq. is:

$$E(X_i) \triangleq \frac{H(R_i)}{H_m} \quad (6)$$

Where H_m is the maximum value of grey entropy, $H_m = \ln n$, and n represents the maximum value of the difference information column consisting of n elements (Deng, 1989).

By entropy correlation criterion, the greater the entropy correlation degree of the comparison column, the stronger the correlation between the comparison column and the reference column. Therefore, using the above model, take the ship entry piloting decision (reflected on the control side, it is the ship rudder combination) as the reference sequence, with various influencing factors as the comparison sequence, and various influencing factors as the comparison sequence, then comprehensively judge the influence degree of each influence factors on the ship entry piloting decision, thus determine the order of each influence factors.

2.2. Fuzzy decision tree model

A decision tree, also known as a tree model or tree structure model, is extensively applied in the field of data mining. Its principle is not complicated, as its basic idea is similar to variation analysis. Its basic purpose is to divide the total study sample into several relatively homogeneous sub-samples using some characteristic(s) (independent variable(s)). The internal variables of each sub-sample are highly consistent, and the corresponding variation/impurity falls between different sub-samples as far as possible. All decision tree algorithms follow this principle, with the difference being in the definition of variation/impurity, such as the use of P values, variance, entropy, Gini coefficient, etc. as a measurement index. According to the predicted dependent variable

type, the decision tree can be divided into two categories: classification tree and regression tree. A decision tree is a tree consisting of internal nodes and leaf nodes for classification and decision-making, where each internal node represents a test on an attribute, each branch represents a test output, and each leaf node represents a class or class distribution. The top node of the tree is the root node, and a path from the root node to the leaf node forms a classification rule. Decision trees are very intuitive classification representations and can be easily converted into classification rules.

The FDT is the expansion and perfection of a traditional decision tree, which extends decision tree learning to handle uncertainty. There is a lot of blurring in real life. Most knowledge is ambiguous and uncertain. Thus, experts usually use vague expertise to solve practical problems, and this transforms the traditional decision tree learning method.

To create a FDT, we first must select the classification attribute at each node. The fuzzy ID3 algorithm uses the concept of entropy. This concept is inversely proportional to the order degree of the data in the sample space. The more ordered the data, the smaller the entropy, and vice versa. If you select a classification attribute to classify the sample data at the node so that the entropy of the node decreases the most, then it is optimal to choose it as a classification attribute. The fuzzy ID3 algorithm defines the information gain (Umanol et al., 1994) to represent the reduction of this entropy, so the attribute with the largest information gain should be selected as the extended attribute of the node.

Set the domain as $D = \{e_1, e_2, \dots, e_n\}$ to represent the example set that summarizes the forecast rules. Each element $e_k (k = 1, 2, \dots, n)$ in the example set has ℓ fuzzy attributes: A^1, A^2, \dots, A^ℓ . The range of each attribute A^i is $\{a_1^i, a_2^i, \dots, a_m^i\} (i = 1, 2, \dots, \ell)$, the j -th example $e_j (j = 1, 2, \dots, m)$ around the value of the i -th attribute is represented by the corresponding membership degree μ_{ij} , which constitutes a fuzzy subset defined on the range $\{a_1^i, a_2^i, \dots, a_m^i\}$ of A^i , and the classification to be divided is $C = \{C_1, C_2, \dots, C_n\}$.

The information gain $G(A^i, D)$ for the attribute A^i is calculated as follows:

$$G(A^i, D) = I(D) - E(A^i, D) \quad (7)$$

where

$$I(D) = - \sum_{k=1}^n (P_k \cdot \log_2 P_k) \quad (8)$$

$$P_k = \frac{|D_{C_k}|}{|D|} \quad (9)$$

$$E(A^i, D) = \sum_{j=1}^m \mu_{ij} \quad (10)$$

$$P_{ij} = \frac{|D_{a_j^i}|}{\sum_{j=1}^m |D_{a_j^i}|} \quad (11)$$

$$\mu_{ij} = P_{ij} \cdot I(D_{a_j^i}) \quad (12)$$

Among them, let D_{C_k} to be a fuzzy subset in D whose class is C_k , $|D|$ the sum of the membership values of the set of data D , and $|D_{C_k}|$ the sum of the membership values of the set of data D_{C_k} ; then $|D_{a_j^i}|$ the sum of the membership values of the set of data $D_{a_j^i}$ and calculate the $\sum_{j=1}^m |D_{a_j^i}|$. After that, we obtain the fuzzy information gain of each attribute at each node calculated by $G(A^i, D)$ through Eqs. (7)–(12) and select the attribute with the largest information gain as the extended attribute of the node to realize the division of the example set.

The fuzzy ID3 algorithm needs to calculate the information gain of each decision attribute, and the attribute with the largest fuzzy information gain is selected as the decision attribute node of the given data set. We then set up the branch of the node by the value of each attribute.

The FDT algorithm consists of three steps:

- 1) Data preprocess: We need to fuzzify the data items of quantitative attributes and divide the quantitative attributes into several linguistic terms. In other words, they must be converted into character attribute values;
- 2) Establish decision tree: Using fuzzy entropy as the heuristic, we select the extended attribute from the root to the leaf, divide the example set, and establish the FDT;
- 3) Match: We predict unknown examples and use the fuzzy matching method to determine the category based on the FDT that has been generated.

2.2.1. Fuzzifying the training data

In a classification issue, the training data attributes are either categorical attributes or continuous numerical attributes. When the data is quantity type, it needs to be fuzzified, and the data set is fuzzified into several linguistic terms. That is, they must be converted into character type attribute values. This transformation process is a conceptual process of reducing decision information (Yuan & Shaw, 1995). There are two steps in the fuzzification process. The first step is to select an effective membership function, such as the triangular membership function, the trapezoidal membership function, or the Gaussian membership function (Chang, Fan, & Dzan, 2010; Fan, Chang, Lin, & Hsieh, 2011; Pulkkinen & Koivisto, 2008). The second step to find the center point of the fuzzy domain, but the number of central points (divided into several linguistic terms) need to pre-set by experience. Some studies have shown that the fuzzy effect of the Gaussian membership function is better. However, in practical applications, the triangular membership functions are used more often due to their simplicity (Wang, Zhai, & Lu, 2008; Fan et al., 2011; Wang, Liu, Pedrycz, & Zhang, 2015). Therefore, in this paper, we use the triangular membership function to fuzzify the quantitative database. The following is a triangular membership function definition (Yuan & Shaw, 1995).

Definition 1: For all examples, attribute that A has a quantitative attribute value x , expressed as $X = \{x(u), u \in U\}$. We want to cluster X to k linguistic terms $T_i, i = 1, 2, \dots, k$. And the triangular membership function equation for each linguistic term T_i is shown in Eqs. (13)–(15), the

$$u_{T_1}(x) = \begin{cases} 1, & x \leq m_1 \\ (m_2 - x)/(m_2 - m_1), & m_1 < x < m_2 \\ 0, & x \geq m_2 \end{cases} \quad (13)$$

$$u_{T_k}(x) = \begin{cases} 1, & x \geq m_k \\ (x - m_{k-1})/(m_k - m_{k-1}), & m_{k-1} < x < m_k \\ 0, & x \leq m_{k-1} \end{cases} \quad (14)$$

$$u_{T_i}(x) = \begin{cases} 0, & x \geq m_{i+1} \\ (m_{i+1} - x)/(m_{i+1} - m_i), & m_i < x < m_{i+1}, 1 < i < k \\ (x - m_{i-1})/(m_i - m_{i-1}), & m_{i-1} < x < m_i \\ 0, & x \leq m_{i-1} \end{cases} \quad (15)$$

Assume that the mentioned dataset D above, where each data has n values for each attribute and one classified class $C = \{C_S, C_M, C_L\}$, in addition, if the attribute has three center points, then the three alternative classified classes could be defined over a value range in fuzzy terms and expressed using the triangular membership function, as shown in Fig. 1.

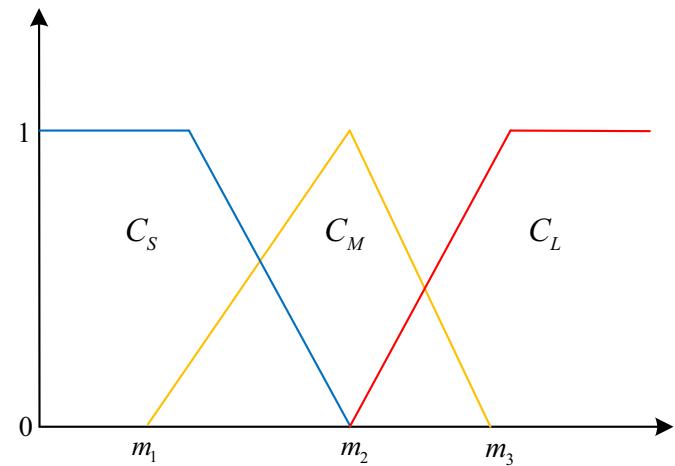


Fig. 1. Example of triangular membership function.

2.2.2. Pruning decision trees

The FDT algorithm is an improvement over the traditional Clear Decision Tree (CDT) algorithm. The fuzzy ID3 algorithm uses fuzzy entropy as the heuristic to select the extended attribute, establishes the FDT, converts each path from the root to the leaf into the rule, generates the fuzzy rule set, matches the example with the fuzzy matching method, and draws conclusions that are closer to human thought (Lior, 2014). The CDT pruning method can be used in the FDT algorithm with only a few modifications (Yuan & Shaw, 1995).

The CDT algorithm is not suitable for pre-pruning. Therefore, it focuses on post-pruning methods (Esposito, Malerba, Semeraro, & Kay, 1997; Quinlan, 2014). Unlike CDT, the FDT algorithm contains pre-pruning strategies. During the establishment of the FDT, the significance level α and truth level threshold β can be well controlled when constructing the FDT.

In addition, unlike the CDT algorithm, the training accuracy and test accuracy are not changed greatly when the rule of FDT is simplified and the lifting range is between 0.001 and 0.005. It has been proven that the decision tree produced by the FDT algorithm has strong prediction abilities, and regular set-pruning strategies cannot further improve its prediction accuracy. The main reasons are as follows (Sun & Wang, 2006):

- 1) The FDT algorithm establishes the decision tree after blurring the continuous data. The fuzziness expresses the correlation and dissimilarity of the attribute value better than the discretization of the CDT algorithm. It can also reduce the interference of the noise data to a certain extent.
- 2) In the construction process, the FDT algorithm uses α and β to control tree generation and carry out pre-pruning;
- 3) The matching strategy of the fuzzy rule set can reduce the noise in the test data so that the matching results are close to the best classification results.

In the process of the FDT generation, there is overlap between the examples covered by the same attribute value, which affects the selection of extended attributes. The introduction of significant level α can reduce the influence of this overlap and reduce the uncertainty of classification so that the entire generation process of the FDT is performed on a given significance level α . Meanwhile, in the FDT generation process, the parameter β is an important condition that is used to control the leaves' generation. The value of α and β directly affects the performance of the FDT. Often, as the value of α increases, the classification uncertainty in the process of building is reduced, but the excessive value of α will lose some sample information during the tree-building process. The higher

the value of β , the larger the decision tree, but the extension ability of the decision tree is reduced. If the value of β is too low, then the decision tree will be too small to summarize the feature set. In addition, in (Sun & Wang, 2006), the author, through the control variate method, noted that the vicinity of 0.7–0.8 usually obtains a better decision tree, and this conclusion has been verified by genetic algorithm; when α ranges from 0.3 to 0.45, good results are obtained.

The FDT algorithm can effectively control the pre-pruning due to its own parameters α and β , resulting in a small-shaped tree with strong forecasting ability. The values of parameters α and β are easily determined and can be derived empirically or experimentally. FDT's post-pruning method can also reduce the size of the tree and improve the prediction ability to a certain extent, but the effect is not obvious unless the parameters of the FDT algorithm are not selected suitably; then, the pruning effect will be obvious.

The FDT algorithm is superior to or equivalent to the CDT method with regular simplification in terms of efficiency tree size and prediction ability. Therefore, in this paper, we construct the FDT with a fuzzy ID3 algorithm and take the average of the optimal interval of the FDT, the significance level α , and truth level threshold β interval set $\alpha = 0.375$ and $\beta = 0.750$, to control the pre-pruning process of FDT.

3. Experiments

3.1. Scenario design

The data for this paper is compiled from the full-task handling simulation platform for large ships, Navi-Trainer Professional 5000, which conforms to the International Maritime Organization (IMO) STCW78/10 convention and the requirements of the China Maritime Safety Administration (MSA), the Det Norske Veritas (DNV), the British Maritime Safety Administration (MAS), the British Lloyd's Register, the British Maritime and Coast Guard (MCA), the Russian Ministry of Shipping (DMT) and other authoritative certifications. We collected the operational data of the exercises and assessment exams as our experimental data (unlimited navigational class crew, captain/chief officer). The simulator scenario was the Shanghai Waigaoqiao wharf, and the ship was downstream berthing into the port.

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. 2(f) shows the initial boundary) to the ship berths docked at the end of the cable (Fig. 2(g) shows the end boundary) as a complete berthing process. The experimental scenario is shown in Fig. 2.

3.2. Data collection and processing

Information was collected on the ship's berthing process, including the environment (wind, flow, tide and 15 other factors), location (longitude, latitude - 2 factors), control (rudder order, marine telegraph - 2 factors), the target ship in the channel (Ship types, speed, quantities and other factors), ship movement (ship heading, steering rate and 11 other factors), the ship's draft (ship's bow, the bow and other factors), tugs, the main collection control (power, rudder order - 7 factors), mechanical contact force-related parameters (4), cable force-related parameters (4), ship movements (bow, 11 ratio factors, etc.), and other related parameters. These above factors, such as the ship's own movement, the environment, the control, location and the relevant parameters of the tug and other factors, were extracted from fixed factors and the weakly related parameters. We record the piloting behavior and environment, including inside and outside the multi-source information.

Fig. 3(a) shows the spatial arrangement for The Navi-Trainer Professional 5000 system, it mainly composed by the Main Bridge, Instructor room, and three Secondary Bridge systems. Fig. 3(b) is the panorama of the Main Bridge. The experimental scheme of this paper is shown in Table 1.

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. From Fig. 5, we can get the distribution of crew age and their piloting experience. The captains' average age and piloting age are both higher than those of chief officers'. It should be noted that, in this article, we consider the tugboat itself as a power plant system of ship OS1 in order to facilitate the ship's overall situation of a simplified analysis.

Fuzzifying the data set as a linguistic term is essentially a conceptualization of reduced decision information. It is usually necessary to divide the number of linguistic terms by experience (Yuan & Shaw, 1995). For instance, the temperature can be conceptualized into three linguistic terms: Cool, Mild, and Hot. In this paper, experimental data of each piloting decision-making factor are tri-sected into three levels: Small (a), Medium (b), and Large (c), to objectively describe the characteristics of each influencing factor, facilitate the construction of a hierarchical and fuzzy decision tree model, and make it easier to describe and mine in detail how each factor influences final driving decisions. Among them, for directional vector influencing factors, such as direction, speed, turning rate, etc., if there is a situation in which the directions are different (There are positive and negative signs in the original data), and the data is asymmetrical, then the extreme value with a large absolute value is selected as the endpoint of the equalization to perform equalization processing. Moreover, the actual physical meaning of each influencing factor, such as Wave Height, Height above the Water and other influencing factors should be fully considered. Although they are vectors and have positive and negative values, they are still directly divided equally and are no longer considered using the above absolute value to get the endpoint. When the data are preprocessed, all the influence factors whose internal data are all zero or unchanged are removed, and the remaining influence factors are sorted in descending order. And the 30 sets of data with large saltation at both ends of the descending order data set are removed, and the extreme values at both ends of the processed data set are selected as the equalization endpoint values to determine the intermediate split points, in order to describe the various characteristics of each factor more objectively. Table 3 shows the processed data set partitions.

Furthermore, when the optimized middle split point is determined, the extreme value of the sorted original data is selected as the extreme value of both ends when the saltation is inconspicuous; when it is judged that the saltatorial extremum is beyond the normal range extension according to the real physical meaning of the influence factors, then the corresponding row of the influence factor data at the moment of saltatorial occurrence is deleted, and then the processed endpoint extreme value is selected. Meanwhile, in order to contain all the data in each influencing factor, the selected rules do not completely follow the principle of rounding. And after four digits are retained after the decimal point, the open interval and closed interval of the split points are determined flexibly according to the trade-off situation. For instance, when the number 12.364512 is the left boundary of the interval, it is retained as 12.3645, and the open interval is selected; when it is the right boundary of the interval, then it is retained as 12.3646 and the open interval is selected; and if the boundary is an integer, it is selected as the boundary value and the closed interval is selected. The partition of the optimized data set segmentation interval is shown in Table 4.

The method of classification interval division that is proposed in this paper fully considers the distribution of data sets and the

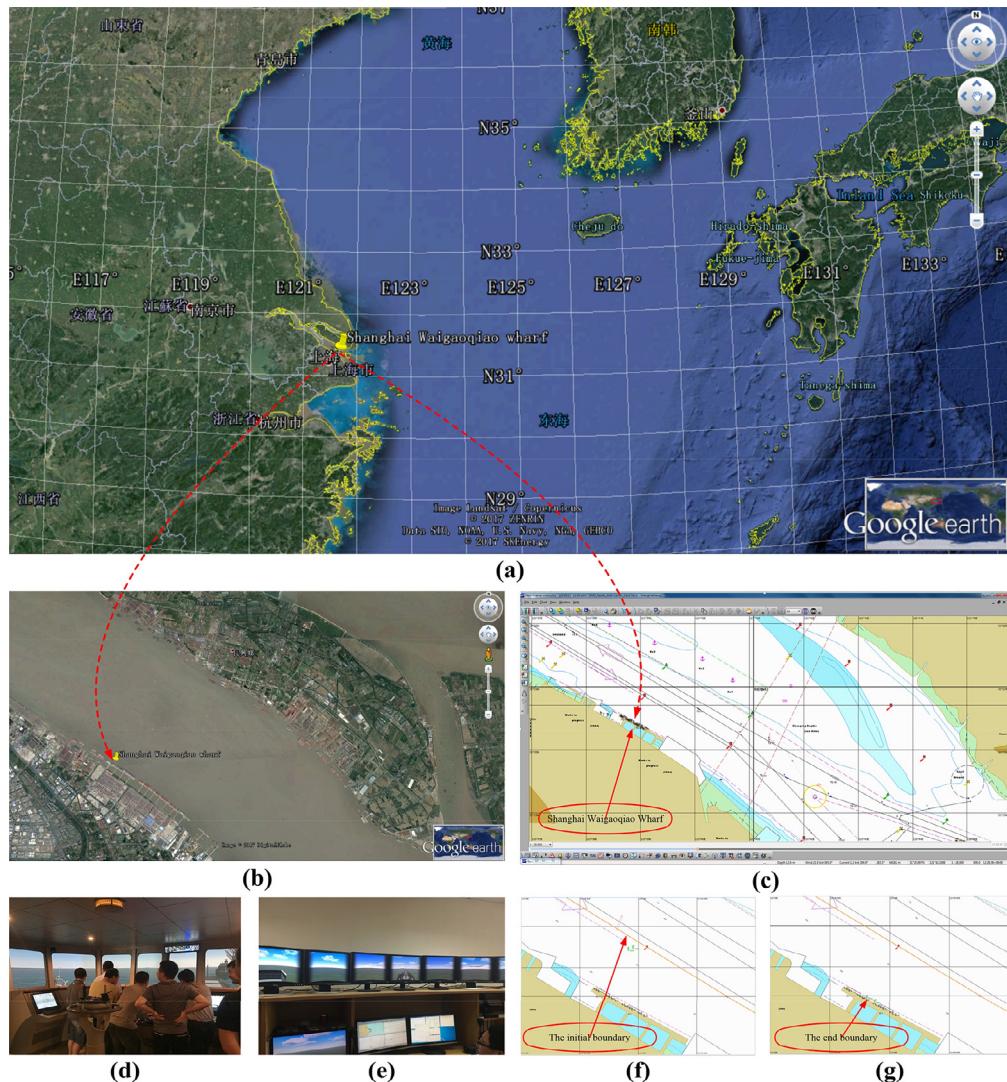


Fig. 2. The experimental scenario.

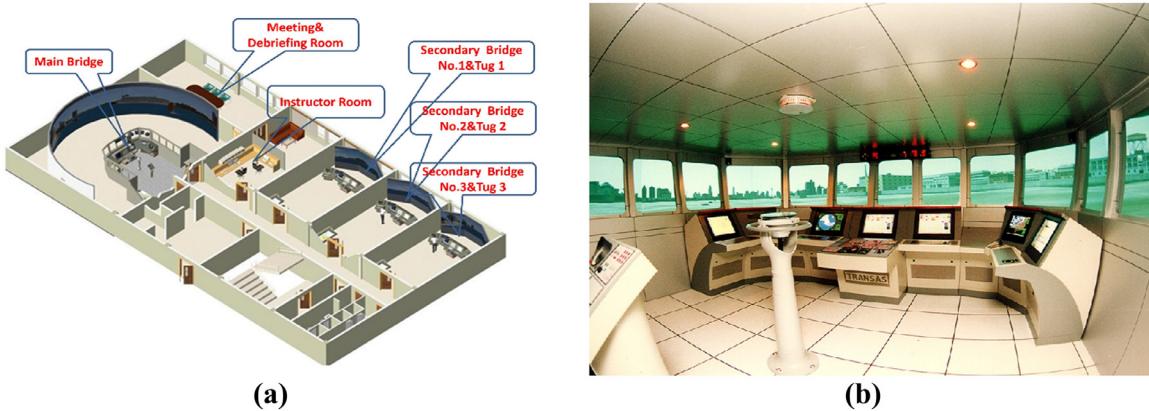


Fig. 3. The Navi-Trainer Professional 5000 system.

endpoint extreme value within a reasonable range, so that all influencing factor data can be standardized accurately and scientifically, and maximally respect and restore the pilot's actual operation and decision-making for the inbound ship under typical scenarios on the simulator. The Performance of the optimized data set partitions and standardization principle is shown in Table 5.

According to the simulation scenario shown in Figs. 2 and Fig. 6, the size of the rudder angle and the propeller speed are defined according to the navigation experience and the situation of data collection from the emulator. When the output power $\geq 50\%$, it is defined as the propeller rapid rotation state, the value range is $[-100\%, -50\%] \cup [50\%, 100\%]$. When the output power $< 50\%$, it

Table 1
Experimental program.

Name	Contents
Time	8: 00–11: 00 and 14: 00–17: 00 on May 17 to June 2.
Place	Wuhan University of Technology Waterway Road Traffic Safety Control and Equipment Ministry of Education Engineering Research Center, Maneuvering Simulator Laboratory for large ships.
Crew	Unlimited navigational class A chief officer or captain, 4 groups of 96 people, 32–45 years old, skilled piloting level.
Ship	30,000 tons bulk carrier (experimental simulation ship OS1, see Fig. 4(a)). 33,089.0t, 182.9 m long, 22.6 m wide.
Scenes	1) Ship downstream berthing into the Shanghai Waigaoqiao Phase IV Port. 2) Sailing in narrow water. 3) Poor visibility. 4) Two tugs help berthing.
Equipment	Navi-Trainer Professional 5000 and 40 Desktop NT-Expert V3.35 system for full-task handling simulation platform for large ships. See Fig. 2(d), (e) and Fig. 3(a), (b).

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

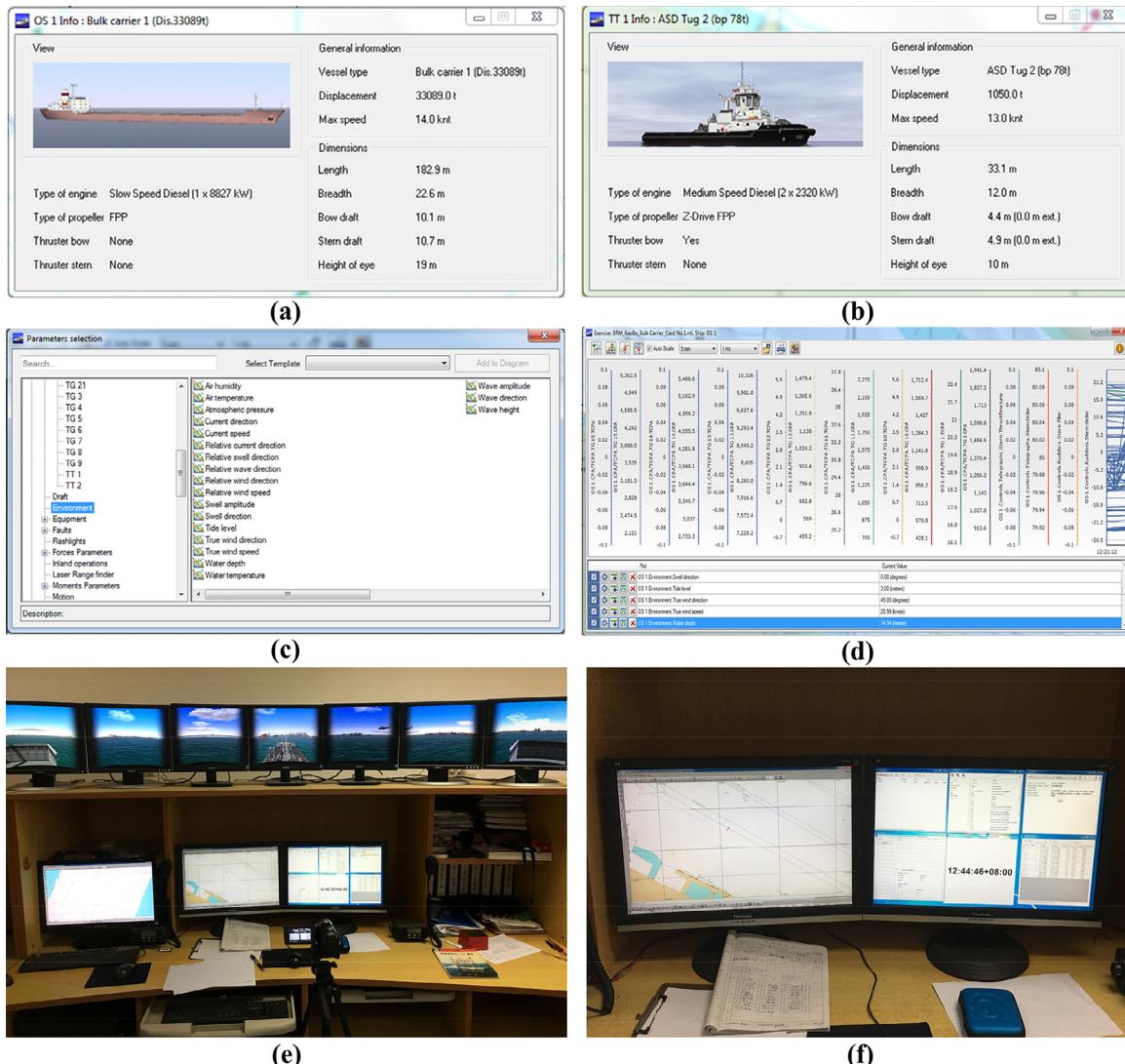


Fig. 4. Data collection and processing.

Table 3

The processed data set partitions and standardization principle of piloting decision-making factors (input).

Influence factors	Meaning	Symbolic principle		
		Small (a)	Medium (b)	Large (c)
Y1	Current draft at ship bow(meters)	[9.9751, 10.16570)	[10.1657, 10.3562)	[10.3562, 10.5468)
Y2	Current draft at ship stern(meters)	[10.5908, 10.8096)	[10.8096, 11.0283)	[11.0283, 11.2470)
Y3	Under keel clearance aft(meters)	[1.3714, 3.3516)	[3.3516, 5.3318)	[5.3318, 7.3120)
Y4	Under keel clearance fwd(meters)	[3.3407, 4.8641)	[4.861, 6.3874)	[6.3874, 7.9108)
Y5	Current direction(degrees)	[313.9000, 315.5000)	[315.5000, 317.1000)	[317.1000, 318.7000)
Y6	Current speed(knots)	[1.0108, 1.0432)	[1.0432, 1.0756)	[1.0756, 1.1080)
Y7	Relative current direction(degrees)	[−60.0000, 0.0000)	[−120.0000, −60.0000)	[−180.0000, −120.0000)
		[0.0000, 60.0000)	[60.0000, 120.0000)	[120.0000, 180.0000)
Y8	Relative wave direction(degrees)	[−41.5000, 0.0000)	[−83.0000, −41.5000)	[−124.5000, −83.0000)
		[0.0000, 41.5000)	[41.5000, 83.0000)	[83.0000, 124.5000)
Y9	Relative wind direction(degrees)	[−59.0205, 0.0000)	[−118.0411, −59.0205)	[−177.0616, −118.0411)
		[0.0000, 59.0205)	[59.0205, 118.0411)	[118.0411, 177.0616)
Y10	Relative wind speed(knots)	[0.2644, 7.5154)	[7.5154, 14.7664)	[14.7664, 22.0174)
Y11	Water depth(meters)	[13.0034, 14.0023)	[14.0023, 15.00113)	[15.00113, 16.0000)
Y12	Wave height(meters)	[−0.4155, −0.1234)	[−0.1234, 0.1686)	[0.1686, 0.4607)
Y13	Forces Parameters. Lateral force(tonne-force)	[−37.5423, 0.0000)	[−75.0846, −37.5423)	[−112.6269, −75.0846)
		[0.0000, 37.542302)	—	—
Y14	Forces Parameters. Longitudinal force(tonne-force)	[−142.0715, 0.0000)	[−284.1429, −142.0715)	[−426.2144, −284.1429)
		[0.0000, 142.0715)	—	—
Y15	Forces Parameters. Summary force(tonne-force)	[0.0000, 160.2039)	[160.2039, 320.4078)	[320.4078, 480.6117)
Y16	Forces Parameters .Vertical force(tonne-force)	[−11.6871, 0.0000)	[−23.3742, −11.6871)	[−35.0612, −23.3742)
		[0.0000, 11.6871)	[11.6871, 23.3742)	[23.3742, 35.0612)
Y17	Mooring lines. Lateral force(tonne-force)	[−162.9374, 0.0000)	—	—
		[0.0000, 162.9374)	[162.9374, 325.8748)	[325.8748, 488.8122)
Y18	Mooring lines. Longitudinal force(tonne-force)	[−44.9968, 0.0000)	[−89.9937, −44.9968)	[−134.9905, −89.9937)
		[0.0000, 44.9968)	[44.9968, 89.9937)	[89.9937, 134.9905)
Y19	Mooring lines. Summary force(tonne-force)	[0.0000, 167.8068)	[167.8068, 335.6137)	[335.6137, 503.4205)
Y20	Mooring lines. Vertical force(tonne-force)	[−33.1281, 0.0000)	[−66.2562, −33.1281)	[−99.3843, −66.2562)
Y21	Heading(degrees)	[100.5245, 178.1916)	[178.1916, 255.8587)	[255.8587, 333.5258)
Y22	Height above the water(meters)	[1.4207, 2.0598)	[2.0598, 2.6989)	[2.6989, 3.3381)
Y23	Lateral speed(knots)	[−0.7564, 0.0000)	[−1.5129, −0.7564)	—
		[0.0000, 0.7564)	[0.7564, 1.5129)	[1.5129, 2.2693)
Y24	Longitudinal speed(knots)	[−2.6398, 0.0000)	—	—
		[0.0000, 2.6398)	[2.6398, 5.2797)	[5.2797, 7.9195)
Y25	Pitch angle(degrees)	[0.1575, 0.1887)	[0.1887, 0.2200)	[0.2200, 0.2512)
Y26	Pitch rate(degrees/min)	[−0.9887, 0.0000)	[−1.9774, −0.9887)	[−2.9660, −1.9774)
		[0.0000, 0.9887)	[0.9887, 1.9774)	[1.9774, 2.9660)
Y27	Rate of turn(degrees/min)	[−13.3700, 0.0000)	[−26.7401, −13.3700)	[−40.1101, −26.7401)
		[0.0000, 13.3700)	[13.3700, 26.7401)	—
Y28	Roll angle(degrees)	[−1.4621, 0.0000)	—	—
		[0.0000, 1.4621)	[1.4621, 2.9242)	[2.9242, 4.3864)
Y29	Roll rate(degrees/min)	[−13.9547, 0.0000)	[−27.9093, −13.9547)	[−41.8640, −27.9093)
		[0.0000, 13.9547)	[13.9547, 27.9093)	[27.9093, 41.8640)
Y30	Vertical speed(knots)	[−0.2744, 0.0000)	[−0.5488, −0.2744)	[−0.8232, −0.5488)
		[0.0000, 0.2744)	[0.2744, 0.5488)	[0.5488, 0.8232)
Y31	Yaw rate(degrees/min)	[−13.3700, 0.0000)	[−26.7401, −13.3700)	[−40.1101, −26.7401)
		[0.0000, 13.3700)	[13.3700, 26.7401)	—
Y32	Latitude(degrees)	[31.3410, 31.3447)	[31.3447, 31.3483)	[31.3483, 31.3520)
Y33	Longitude(degrees)	[121.6420, 121.6444)	[121.6444, 121.6467)	[121.6467, 121.6490)

is defined as the propeller slow rotation state, the value range is $(-50\%, 0) \cup (0, 50\%)$. When the rudder angle value belongs to the interval $(-10, 0) \cup (0, 10)$, it is defined as the small steering angle. When the value of the rudder angle belongs to the interval $[-35, -10] \cup [10, 35]$, it is defined as the large steering angle. See Fig. 6 and Table 6 (showing 64 possible piloting decisions). In addition, this article does not consider "Midships" and "Stop engine," regardless of the rudder angle and if the power output is 0.

Captains manoeuvre ships by operating different telegraph and rudder orders, so as to change ship's speed and direction, and to complete the ship's control. Combining telegraph and rudder orders, this control is a multi-dynamic process. Fig. 6(c), (d) reveals the changing rule of telegraph order and rudder order within the time in a typical situation that a ship sails from the initial boundary to the end boundary designed in scenario design part. See Fig. 2(c), (f), (g).

4. Results and discussions

Piloting decision-making is stimulated and influenced by multi-source information, such as human, ships, environment, as well as real-time requirements. This requires piloting decision-making knowledge to be automatically obtained and expressed along with higher decision-making knowledge effectiveness.

4.1. Determining piloting decision main influence factor

Piloting decision-making processes are often influenced by multi-source information such as human, ship and environmental factors. These factors are collectively referred to as piloting decision-making factors. They act together to determine the next action strategy of the ship's pilot. According to this strategy and the current piloting environment, the pilot can quickly and accurately develop piloting decisions and thus lay the foundation

Table 4

The optimized data set partitions and standardization principle of piloting decision-making factors (input).

Influence factors	Meaning	Symbolic principle		
		Small (a)	Medium (b)	Large (c)
Y1	Current draft at ship bow(meters)	(9.0160, 10.16570)	[10.1657, 10.3562)	[10.3562, 10.8347)
Y2	Current draft at ship stern(meters)	(9.6172, 10.8096)	[10.8096, 11.0283)	[11.0283, 14.3239)
Y3	Under keel clearance aft(meters)	(0.7235, 3.3516)	[3.3516, 5.3318)	[5.3318, 7.9108)
Y4	Under keel clearance fwd(meters)	(2.7321, 4.8641)	[4.861, 6.3874)	[6.3874, 8.9840)
Y5	Current direction(degrees)	[313.9000, 315.5000)	[315.5000, 317.1000)	[317.1000, 318.7001)
Y6	Current speed(knots)	(1.0107, 1.0432)	[1.0432, 1.0756)	[1.0756, 1.1080]
Y7	Relative current direction(degrees)	[−60.0000, 0.0000)	[−120.0000, −60.0000)	(−180.0000, −120.0000)
Y8	Relative wave direction(degrees)	[0.0000, 60.0000)	[60.0000, 120.0000)	[120.0000, 180.0000)
Y9	Relative wind direction(degrees)	[−59.0205, 0.0000)	[−118.0411, −59.0205)	(−179.7170, −118.0411)
Y10	Relative wind speed(knots)	(0.0228, 7.5154)	[7.5154, 14.7664)	[14.7664, 22.1793)
Y11	Water depth(meters)	(12.5530, 14.0023)	[14.0023, 15.00113)	[15.00113, 16.00000)
Y12	Wave height(meters)	[−0.7300, −0.1234)	[−0.1234, 0.1686)	[0.1686, 0.6900]
Y13	Forces Parameters. Lateral force(tonne-force)	[−37.5423, 0.0000)	[−75.0846, −37.5423)	(−1025.9804, −75.0846)
Y14	Forces Parameters. Longitudinal force(tonne-force)	[−142.0715, 0.0000)	[−284.1429, −142.0715)	(−24,592.3230, −284.1429)
Y15	Forces Parameters. Summary force(tonne-force)	[0.0000, 160.2039)	[160.2039, 320.4078)	[320.4078, 25,144.6958)
Y16	Forces Parameters .Vertical force(tonne-force)	[−11.6871, 0.0000)	[−23.3742, −11.6871)	(−4547.0121, −23.3742)
Y17	Mooring lines. Lateral force(tonne-force)	(−162.9374, 0.0000)	—	—
Y18	Mooring lines. Longitudinal force(tonne-force)	[0.0000, 162.9374)	[162.9374, 325.8748)	[325.8748, 953.2618)
Y19	Mooring lines. Summary force(tonne-force)	[0.0000, 167.8068)	[167.8068, 335.6137)	[335.6137, 983.4908)
Y20	Mooring lines. Vertical force(tonne-force)	[−33.1281, 0.0000)	[−66.2562, −33.1281)	[−246.0158, −66.2562)
Y21	Heading(degrees)	[100.2441, 178.1916)	[178.1916, 255.8587)	[255.8587, 339.7877)
Y22	Height above the water(meters)	(1.1746, 2.0598)	[2.0598, 2.6989)	[2.6989, 4.0339)
Y23	Lateral speed(knots)	[−0.7564, 0.0000)	[−1.5129, −0.7564)	(−1.2774, −1.5129)
Y24	Longitudinal speed(knots)	[0.0000, 0.7564)	[0.7564, 1.5129)	[1.5129, 4.0375)
Y25	Pitch angle(degrees)	[−0.9631, 0.1887)	[0.1887, 0.2200)	[0.2200, 1.4851)
Y26	Pitch rate(degrees/min)	(−92.3133, 0.0000)	[−1.9774, −0.9887)	[−2.9660, 37.1897)
Y27	Rate of turn(degrees/min)	[0.0000, 0.9887)	[0.9887, 1.9774)	[1.9774, 2.9660)
Y28	Roll angle(degrees)	[−13.3700, 0.0000)	[−26.7401, −13.3700)	(−216.4195, −26.7401)
Y29	Roll rate(degrees/min)	[0.0000, 13.3700)	[13.3700, 26.7401)	[26.7401, 82.5781)
Y30	Vertical speed(knots)	(−7.1696, 0.0000)	—	—
Y31	Yaw rate(degrees/min)	[0.0000, 1.4621)	[1.4621, 2.9242)	[2.9242, 12.1096)
Y32	Latitude(degrees)	(−13.9547, 0.0000)	[−27.9093, −13.9547)	(−219.3305, −27.9093)
Y33	Longitude(degrees)	[0.0000, 13.9547)	[13.9547, 27.9093)	[27.9093, 220.8617)

for the research of human-like piloting behavior. For a particular person-ship unit, the overall reliability is constant for a certain period of time or during a trip, so the person and ship factors have less influence on piloting decisions. With the operation of the ship, the pilot's waterway and the environment will change with time and space, and the changing waterway and environmental factors will have a greater impact on piloting decisions. Therefore, this paper uses the grey relation entropy analysis to focus on the piloting decision-making factors from the waterway and environment, where the factors are squeezed and sorted. The piloting

of decision-making factors related to the information are shown in [Tables 6](#) and [Table 7](#).

According to the sorting criteria of the grey relational sequence, the greater the degree of entropy correlation of the comparison column, the greater the relevance of the comparison column to the reference column, the greater the degree of influence on the reference column, and the higher the ranking of the influencing factors. The grey entropy analysis method uses information entropy to quantitatively describe the similarity and consistency degree between each comparison column and reference column and

Table 5

Performance of the optimized data set partitions and standardization principle.

Segmentation of intervals	Total number	Selected number	Accuracy (%)	Total time (s)
Table 3	677,650.00	675,670.00	99.71	7.65
Table 4	677,650.00	677,650.00	100.00	7.73

Table 6
Piloting 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	$b_{i+1} - b_i \neq 0$	Changed	C2
	$a_{i+1} - a_i = 0$	Unchanged	U1	$b_{i+1} - b_i = 0$	Unchanged	U2
Value	$[-100\%, -50\%] \cup [50\%, 100\%]$	Fast	F1	$[-35, -10] \cup [10, 35]$	Large	L2
	$(-50\%, 0) \cup (0, 50\%)$	Slow	S1	$(-10, 0) \cup (0, 10)$	Small	S2
Direction	$a_i > 0$	Ahead	D1	$b_i > 0$	Starboard	D2
	$a_i < 0$	Astern	T1	$b_i < 0$	Port	T2
Influence factors	Decisions		Symbols	Decisions		symbol
X(Dimensionless)	U1F1D1U2L2T2		X1	U1F1D1C2L2T2		X33
	U1F1D1U2S2T2		X2	U1F1D1C2S2T2		X34
	U1S1D1U2L2T2		X3	U1S1D1C2L2T2		X35
	U1S1D1U2S2T2		X4	U1S1D1C2S2T2		X36
	U1F1T1U2L2T2		X5	U1F1T1C2L2T2		X37
	U1F1T1U2S2T2		X6	U1F1T1C2S2T2		X38
	U1S1T1U2L2T2		X7	U1S1T1C2L2T2		X39
	U1S1T1U2S2T2		X8	U1S1T1C2S2T2		X40
	U1F1D1U2L2D2		X9	U1F1D1C2L2D2		X41
	U1F1D1U2S2D2		X10	U1F1D1C2S2D2		X42
	U1S1D1U2L2D2		X11	U1S1D1C2L2D2		X43
	U1S1D1U2S2D2		X12	U1S1D1C2S2D2		X44
	U1F1T1U2L2D2		X13	U1F1T1C2L2D2		X45
	U1F1T1U2S2D2		X14	U1F1T1C2S2D2		X46
	U1S1T1U2L2D2		X15	U1S1T1C2L2D2		X47
	U1S1T1U2S2D2		X16	U1S1T1C2S2D2		X48
	C1F1D1C2L2T2		X17	C1F1D1U2L2T2		X49
	C1F1D1C2S2T2		X18	C1F1D1U2S2T2		X50
	C1S1D1C2L2T2		X19	C1S1D1U2L2T2		X51
	C1S1D1C2S2T2		X20	C1S1D1U2S2T2		X52
	C1F1T1C2L2T2		X21	C1F1T1U2L2T2		X53
	C1F1T1C2S2T2		X22	C1F1T1U2S2T2		X54
	C1S1T1C2L2T2		X23	C1S1T1U2L2T2		X55
	C1S1T1C2S2T2		X24	C1S1T1U2S2T2		X56
	C1F1D1C2L2D2		X25	C1F1D1U2L2D2		X57
	C1F1D1C2S2D2		X26	C1F1D1U2S2D2		X58
	U1S1D1C2L2D2		X27	C1S1D1U2L2D2		X59
	C1S1D1C2S2D2		X28	C1S1D1U2S2D2		X60
	C1F1T1C2L2D2		X29	C1F1T1U2L2D2		X61
	C1F1T1C2S2D2		X30	C1F1T1U2S2D2		X62
	C1D1T1C2L2D2		X31	C1S1T1U2L2D2		X63
	C1D1T1C2S2D2		X32	C1S1T1U2S2D2		X64

uses entropy correlation degrees to complete the matching order of influencing factors. In this paper, we select X0 (X0 presents the percentage of the number of each piloting decision of X1-X64 in a total number of the data set records) as the reference column and Y1 – Y33 as the comparison column. Limited to space, Table 7 lists only a part of multiple measured data.

According to the grey relation entropy principle, the grey correlation coefficient, and the entropy correlation degree of each comparison column is obtained by quantitative calculation of the data in Table 7; the results are shown in Tables 8 and Table 9.

According to Table 9, the influence factors are sorted according to the influence degree: $Y_8 > Y_{22} > Y_4 > Y_{31} > Y_{27} > Y_{16} > Y_{33} > Y_{11} > Y_3 > Y_{13} > Y_{14} > Y_{15} > Y_{12} > Y_{30} > Y_5 > Y_{18} > Y_6 > Y_{26} > Y_{20} > Y_{10} > Y_{17} > Y_{19} > Y_{23} > Y_{25} > Y_{29} > Y_{24} > Y_{28} > Y_{32} > Y_2 > Y_1 > Y_9 > Y_{21} > Y_7$. For simplicity, this paper selects the first six factors to study the decision-making mechanisms for different piloting behaviors. Table 10 lists some of the training samples.

The data in Table 10 are standardized according to the principle of standardization of piloting decision influence factors in Tables 4 and Table 6.

In accordance with the optimized standardization principle of influence factors for piloting decisions in Table 4, the attributes of the six factors selected in Table 10 are fuzzified, the number of center points k is 3, and the set of center points is $M = \{m_i, i=1,$

$\dots, k\}$. By simple division method, the linguistic term Small is calculated using Eq. (13), the linguistic term Medium is obtained by Eq. (15), and the linguistic term Large is obtained by Eq. (14). Here we opt for K-means clustering algorithm combines with the algorithm proposed in data collection and processing part, then generate the center points of impact factors as shown in Table 12. The training set with fuzzy representation is shown in Table 13.

4.2. Inducing piloting fuzzy decision tree

The piloting decision classification tree is constructed by using the fuzzy ID3 classification algorithms and fuzzy membership standard training samples in Table 13. The fuzzy ID3 classification algorithm is summarized as follows. First, select the piloting decision-making main influence factors with the maximum fuzzy information gain to generate decision tree nodes and establish a branch by the different values of the nodes. Second, take the instance subset of the branch and use this method to establish the nodes and branches of the decision tree until the instances in a subset belong to the same classification. Finally, the piloting decision classification tree constructed by the fuzzy ID3 classification algorithm. And the classification rules are graphically represented by the decision tree structure in Fig. 7. The algorithm scheme is as follows:

Table 7
Sample data set for evaluation of the studied area (partially).

Influence factors	Sample set						
	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	...
X0 (Dimensionless)	0.0384	0.0728	0.0384	0.2728	0.0454	0.0128	...
Y1 (meters)	10.1110	10.1516	10.1538	10.1560	10.1296	10.1355	...
Y2 (meters)	10.7500	10.7948	10.7961	10.7974	10.7643	10.7614	...
Y3 (meters)	4.2833	4.2363	4.2345	4.2327	4.2611	4.2589	...
Y4 (meters)	4.7445	4.7017	4.6990	4.6963	4.7179	4.7068	...
Y5 (degrees)	314.9000	314.9000	314.9000	314.9000	314.9000	314.9000	...
Y6 (knots)	1.0497	1.0497	1.0497	1.0497	1.0497	1.0497	...
Y7 (degrees)	28.0777	122.9068	-12.7913	-148.4894	-179.4163	-179.4000	...
Y8 (degrees)	-8.9422	-8.9000	-8.9000	-8.9000	-8.8513	-8.8000	...
Y9 (degrees)	69.8398	70.0158	70.0536	70.0913	70.5728	71.0000	...
Y10 (knots)	1.9827	1.9827	1.9827	1.9827	1.9733	1.9633	...
Y11 (meters)	14.9400	14.9400	14.9400	14.9400	14.9351	14.9300	...
Y12 (meters)	0.0401	-0.0016	-0.0054	-0.0091	0.0562	-0.0900	...
Y13 (tonne-force)	2.3463	2.3463	2.3463	2.3463	2.3463	2.3463	...
Y14 (tonne-force)	0.2803	0.2803	0.2803	0.2803	0.2803	0.2803	...
Y15 (tonne-force)	2.3630	2.3630	2.3630	2.3630	2.3630	2.3630	...
Y16 (tonne-force)	-0.0092	-0.0092	-0.0092	-0.0092	-0.0092	-0.0092	...
Y17 (tonne-force)	0.0067	0.0067	0.0067	0.0067	0.0067	0.0067	...
Y18 (tonne-force)	-0.0213	-0.0213	-0.0213	-0.0213	-0.0213	-0.0213	...
Y19 (tonne-force)	0.0786	0.0786	0.0786	0.0786	0.0786	0.0786	...
Y20 (tonne-force)	-0.0754	-0.0754	-0.0754	-0.0754	-0.0754	-0.0754	...
Y21 (degrees)	233.9447	233.9200	233.9183	233.8878	233.8264	233.7855	...
Y22 (meters)	2.9483	2.9222	2.9225	2.9399	2.9623	2.9477	...
Y23 (knots)	1.0905	1.0908	1.0910	1.0939	0.9871	1.0972	...
Y24 (knots)	5.6254	5.6152	5.6152	5.6149	5.4510	5.5839	...
Y25 (degrees)	0.2091	0.2100	0.2097	0.2020	0.2193	0.2035	...
Y26 (degrees/min)	-0.3208	-0.3458	-0.3713	-0.7965	0.5707	-0.4372	...
Y27 (degrees/min)	-2.1088	-2.1241	-2.1421	-2.4411	-1.9075	-1.8242	...
Y28 (degrees)	0.0188	0.0233	0.0236	0.0272	0.0322	0.0331	...
Y29 (degrees/min)	0.3844	0.3141	0.3100	0.2430	0.0469	-0.0326	...
Y30 (knots)	-0.0395	-0.0112	-0.0076	0.0528	-0.0161	-0.0336	...
Y31 (degrees/min)	-2.1088	-2.1241	-2.1421	-2.4411	-1.9075	-1.8242	...
Y32 (degrees)	31.3495	31.3494	31.3494	31.3494	31.3494	31.3494	...
Y33 (degrees)	121.6494	121.6493	121.6493	121.6493	121.6493	121.6492	...

Table 8
Grey correlation coefficient for the sample data (partially).

Impact factors	Grey correlation coefficient R						
	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	...
Y1	0.948821	0.944801	0.935685	0.933054	0.940548	0.945995	...
Y2	0.944423	0.941285	0.934170	0.932106	0.937298	0.941518	...
Y3	0.980756	0.981198	0.982183	0.982491	0.978342	0.977833	...
Y4	0.984270	0.984636	0.985512	0.985749	0.981410	0.980826	...
Y5	0.958720	0.958720	0.958720	0.958720	0.955696	0.955696	...
Y6	0.940306	0.940306	0.940306	0.940306	0.937397	0.937397	...
Y7	0.977801	0.977892	0.977932	0.977993	0.974925	0.974957	...
Y8	0.971251	0.971392	0.971609	0.971735	0.974998	0.975096	...
Y9	0.841280	0.841107	0.840897	0.840735	0.838168	0.838051	...
Y10	0.998258	0.998170	0.998126	0.998081	0.998665	0.998729	...
Y11	0.983231	0.983231	0.983231	0.983231	0.980051	0.980051	...
Y12	0.944236	0.944220	0.994049	0.977954	0.934740	0.944329	...
Y13	0.965905	0.965905	0.965905	0.965905	0.962835	0.962835	...
Y14	0.964395	0.964395	0.964395	0.964395	0.961335	0.961335	...
Y15	0.968953	0.968953	0.968953	0.968953	0.965865	0.965865	...
Y16	0.966102	0.966102	0.966102	0.966102	0.963031	0.963031	...
Y17	0.969357	0.969357	0.969357	0.969357	0.966265	0.966265	...
Y18	0.967034	0.967034	0.967034	0.967034	0.963957	0.963957	...
Y19	0.969569	0.969569	0.969569	0.969569	0.966476	0.966476	...
Y20	0.963393	0.963393	0.963393	0.963393	0.960339	0.960339	...
Y21	0.909890	0.910029	0.910146	0.910297	0.907702	0.907830	...
Y22	0.954179	0.955253	0.956160	0.956446	0.951070	0.950563	...
Y23	0.907782	0.907409	0.908501	0.908866	0.906425	0.906430	...
Y24	0.950644	0.950756	0.950573	0.950536	0.947538	0.947568	...
Y25	0.938662	0.938578	0.938564	0.938505	0.935553	0.935517	...
Y26	0.965855	0.965857	0.965923	0.965895	0.962822	0.962899	...
Y27	0.993808	0.993416	0.992348	0.991987	0.988237	0.987177	...
Y28	0.966143	0.966234	0.966354	0.966462	0.963519	0.963702	...
Y29	0.966916	0.966278	0.966579	0.966699	0.963888	0.964242	...
Y30	0.970562	0.972582	0.973802	0.958892	0.950538	0.968439	...
Y31	0.993808	0.993416	0.992348	0.991987	0.988237	0.987177	...
Y32	0.902639	0.902687	0.902726	0.902784	0.900150	0.900198	...
Y33	0.941088	0.941467	0.941775	0.942178	0.939635	0.939989	...

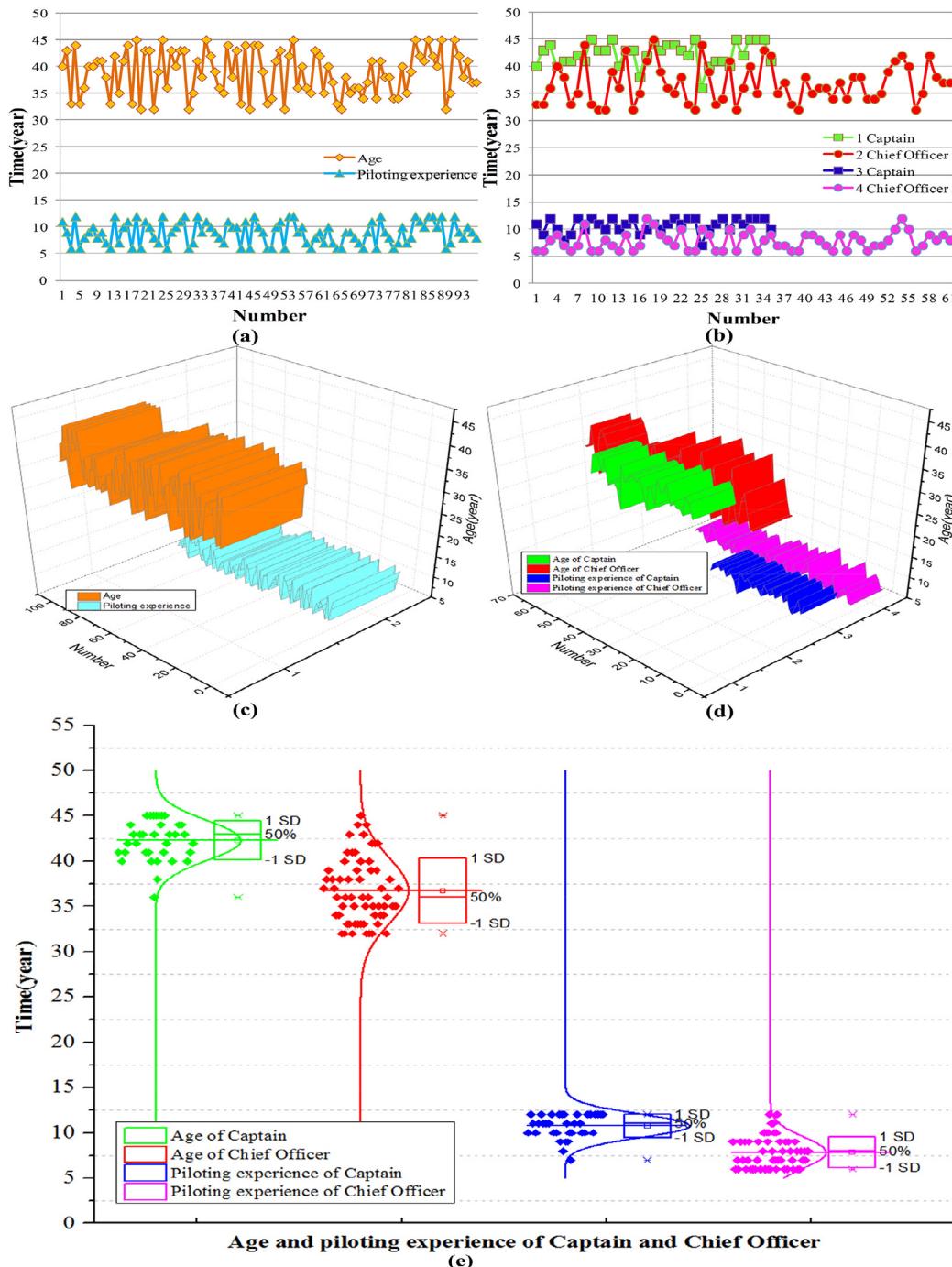


Fig. 5. Analysis of participants' information.

- 1) Create a root node D. Set the fuzzy training as the root node of the input and the other nodes of the tree as a fuzzy subset of the training set.
- 2) Classify the example set. Select the attribute with the highest fuzzy information gain as the extended attribute of the node. Each attribute A^i performs a fuzzy segmentation on D. Then, calculate the fuzzy information gain $G(A^i, D)$ generated by each attribute A^i at the node D.
- 3) If the confidence level of a certain class in the node is greater than β (the truth level threshold used in this paper is 0.750), then the leaf is generated.
- 4) If all the attributes on a node have been used, then the leaf is generated.
- 5) Otherwise, select the unused attribute with the highest fuzzy information gain as the extension attribute. If the fuzzy information gain is less than the given value, the leaf is generated. Blur the current node by the extended attribute value to generate its sub-node. Then, repeat the above process until the whole decision tree is established.

4.3. Establishing piloting decision classification rules

For the resulting piloting decision tree, the path from the root node to each leaf node of the decision tree corresponds to the combination of a set of attribute tests. The decision tree represents these conjunctive separations. With the piloting decision classifi-

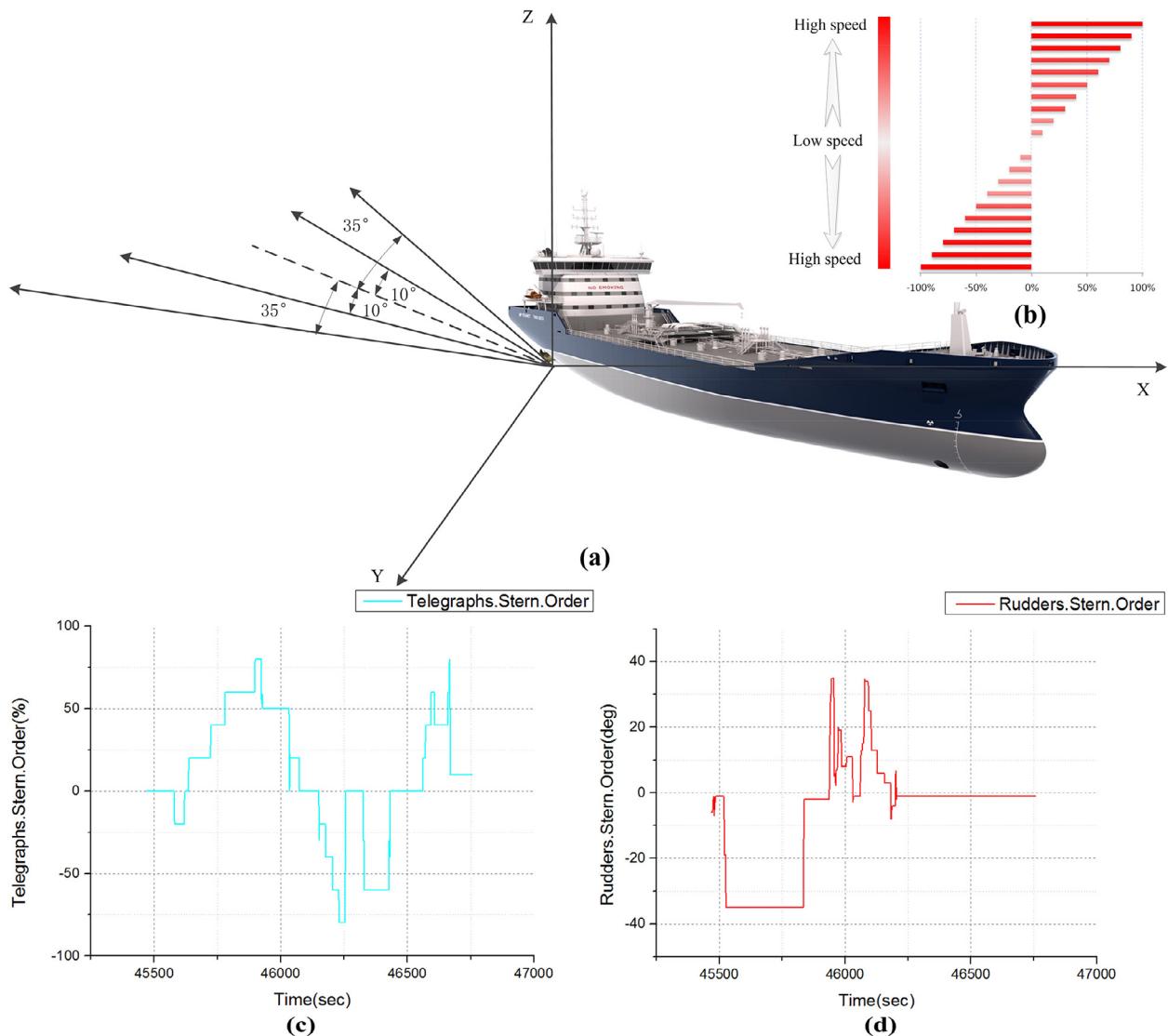
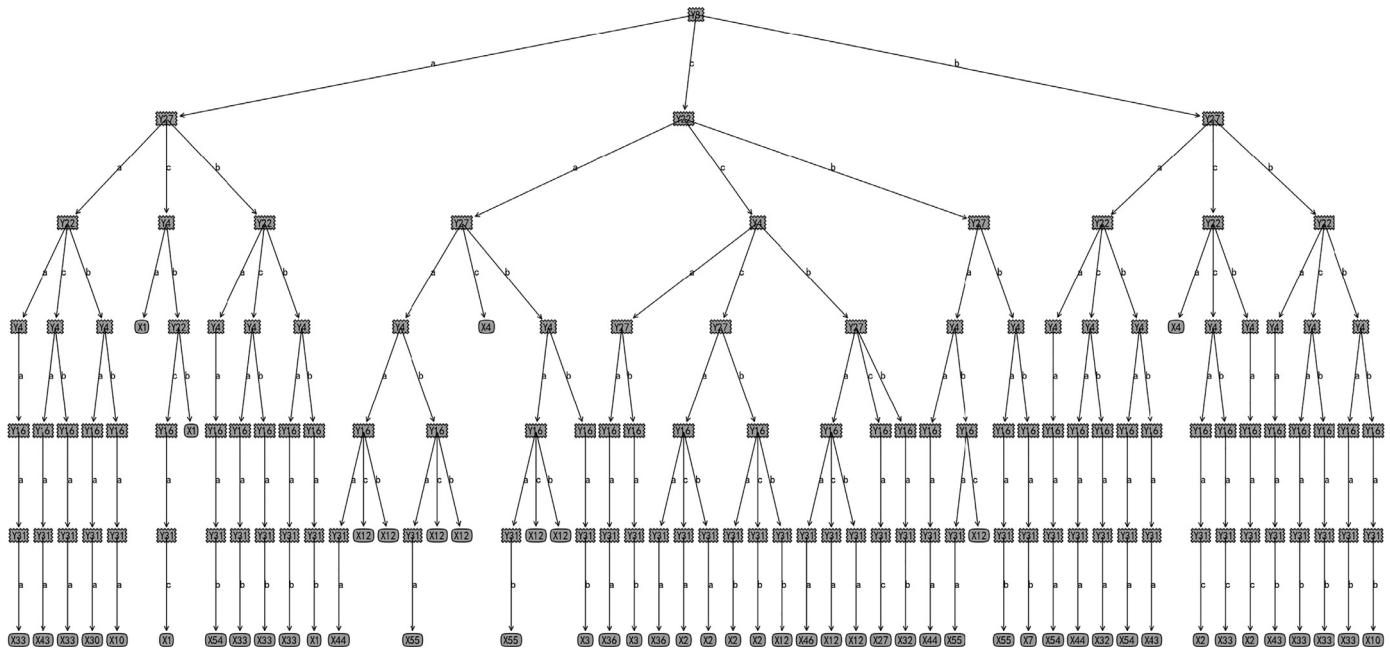


Fig. 6. Analysis of the experimental ship.

Table 9
Grey relation entropy and entropy correlation of each comparative column.

Impact factors	Grey relation entropy $H(R)$	Entropy correlation $E(Y)$	Impact factors	Grey relation entropy $H(R)$	Entropy correlation $E(Y)$
Y1	9.9293545	0.9999514	Y18	9.9294521	0.9999612
Y2	9.9293573	0.9999517	Y19	9.9294144	0.9999574
Y3	9.9294876	0.9999648	Y20	9.9294297	0.9999590
Y4	9.9295372	0.9999698	Y21	9.9293154	0.9999474
Y5	9.9294551	0.9999615	Y22	9.9295775	0.9999738
Y6	9.9294348	0.9999595	Y23	9.9294079	0.9999568
Y7	9.9293085	0.9999467	Y24	9.9293957	0.9999555
Y8	9.9295981	0.9999759	Y25	9.9294056	0.9999565
Y9	9.9293199	0.9999479	Y26	9.9294307	0.9999591
Y10	9.9294223	0.9999582	Y27	9.9295344	0.9999695
Y11	9.9295045	0.9999665	Y28	9.9293906	0.9999550
Y12	9.9294607	0.9999621	Y29	9.9294052	0.9999565
Y13	9.9294823	0.9999642	Y30	9.9294555	0.9999615
Y14	9.9294797	0.9999640	Y31	9.9295355	0.9999696
Y15	9.9294764	0.9999636	Y32	9.9293865	0.9999546
Y16	9.9295335	0.9999694	Y33	9.9295294	0.9999690
Y17	9.9294209	0.9999581	—	—	—

**Fig. 7.** Decision tree structure for piloting decision classification rules.**Table 10**
Training samples (partially).

No.	X			Y4	Y8	Y16	Y22	Y27	Y31
		Rudders order	Telegraphs order						
1	-35.0000	0.0000		4.7793	-1.6463	-0.0092	2.7666	-10.1748	-4.9876
2	-35.0000	0.0000		4.7833	-1.5230	-0.0092	2.8756	-10.0944	-4.6709
3	-35.0000	0.0000		4.8618	-1.3379	-0.0092	2.8889	-10.4877	-4.8759
4	-35.0000	-4.3207		4.9425	-1.2042	-0.0092	3.0291	-10.3155	-4.7104
5	-35.0000	-17.9076		5.0001	-0.9383	-0.0092	3.0278	-10.4462	-4.5943
6	-35.0000	-20.0000		4.9737	-0.8662	-0.0092	2.9371	-10.3930	-4.4932
...

Table 11
Training set with the principle of standardization (partially).

No.	X	Y4			Y8			Y16			Y22			Y27			Y31		
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	X3	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0
2	X3	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0
3	X3	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0
4	X55	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0
5	X55	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0
6	X55	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0
...

Table 12
The center points of selected impact factors.

Impact factors	m_1	m_2	m_3
Y4	4.0241	5.3161	6.6080
Y8	31.2000	62.4000	93.6000
Y16	-0.0521	-0.0274	-0.0027
Y22	2.0750	2.4000	2.7250
Y27	-2.1606	22.4190	46.9986
Y31	-4.4106	17.9190	40.2486

From the decision tree, we can easily extract the decision-making knowledge described by the decision tree and can use the "IF-THEN" form to extract the rules. Each piloting decision can be obtained along the path from the root node to the leaf node of the decision tree. A collection of attributes and their values encountered along the

given path constitutes a prerequisite for the rule (IF part). The leaf node gives the predicted value of the classification, forming the conclusion part of the rule (THEN part). Finally, all rules are merged to form the piloting decision recognition rule base, as shown in Table 14.

The following conclusions can be drawn from Fig. 7 and Table 14:

- In the piloting decision-making process, piloting behavior is often stimulated and influenced by pilots, ships, waterways, environment and other factors. These factors together lead the pilot to gradually form the next moment action (action, strategy or tactics) in their mind. According to this long-term strategy and the current piloting environment, the pilot can quickly and accurately develop piloting decisions and prepare to establish

Table 13

Training set with fuzzy representation (partially).

No.	X	Y4			Y8			Y16			Y22			Y27			Y31		
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	X3	0.629	0.371	0.000	0.000	0.019	0.981	0.720	0.280	0.000	0.992	0.008	0.000	0.918	0.082	0.000	0.975	0.025	0.000
2	X3	0.625	0.375	0.000	0.000	0.016	0.984	0.706	0.294	0.000	0.997	0.003	0.000	0.926	0.074	0.000	0.963	0.037	0.000
3	X3	0.601	0.399	0.000	0.000	0.013	0.987	0.710	0.290	0.000	0.999	0.001	0.000	0.938	0.062	0.000	0.953	0.047	0.000
4	X55	0.593	0.407	0.000	0.000	0.009	0.991	0.726	0.274	0.000	0.997	0.003	0.000	0.949	0.051	0.000	0.943	0.057	0.000
5	X55	0.597	0.403	0.000	0.000	0.005	0.995	0.710	0.290	0.000	0.998	0.002	0.000	0.970	0.030	0.000	0.926	0.074	0.000
6	X55	0.618	0.382	0.000	0.000	0.001	0.999	0.691	0.309	0.000	1.000	0.000	0.000	0.985	0.015	0.000	0.926	0.074	0.000
...	

Table 14

Piloting decision classification rules.

No.	Piloting decision classification rules
1	IF Y8 = a AND Y27 = a AND Y22 = a AND Y4 = a AND Y16 = a AND Y31 = a THEN X = X33
2	IF Y8 = a AND Y27 = a AND Y22 = c AND Y4 = a AND Y16 = a AND Y31 = a THEN X = X43
3	IF Y8 = a AND Y27 = a AND Y22 = c AND Y4 = b AND Y16 = a AND Y31 = a THEN X = X33
4	IF Y8 = a AND Y27 = a AND Y22 = b AND Y4 = a AND Y16 = a AND Y31 = a THEN X = X30
5	IF Y8 = a AND Y27 = a AND Y22 = b AND Y4 = b AND Y16 = a AND Y31 = a THEN X = X10
6	IF Y8 = a AND Y27 = c AND Y4 = a THEN X = X1
7	IF Y8 = a AND Y27 = c AND Y4 = b AND Y22 = c AND Y16 = a AND Y31 = c THEN X = X1
8	IF Y8 = a AND Y27 = c AND Y4 = b AND Y22 = b THEN X = X1
9	IF Y8 = a AND Y27 = b AND Y22 = a AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X54
10	IF Y8 = a AND Y27 = b AND Y22 = c AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X33
11	IF Y8 = a AND Y27 = b AND Y22 = c AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X33
12	IF Y8 = a AND Y27 = b AND Y22 = b AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X33
13	IF Y8 = a AND Y27 = b AND Y22 = b AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X1
14	IF Y8 = c AND Y22 = a AND Y27 = a AND Y4 = a AND Y16 = a AND Y31 = a THEN X = X44
15	IF Y8 = c AND Y22 = a AND Y27 = a AND Y4 = a AND Y16 = c THEN X = X12
16	IF Y8 = c AND Y22 = a AND Y27 = a AND Y4 = b AND Y16 = b THEN X = X12
17	IF Y8 = c AND Y22 = a AND Y27 = a AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X55
18	IF Y8 = c AND Y22 = a AND Y27 = a AND Y4 = b AND Y16 = c THEN X = X12
19	IF Y8 = c AND Y22 = a AND Y27 = a AND Y4 = b AND Y16 = b THEN X = X12
20	IF Y8 = c AND Y22 = a AND Y27 = c THEN X = X4
21	IF Y8 = c AND Y22 = a AND Y27 = b AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X55
22	IF Y8 = c AND Y22 = a AND Y27 = b AND Y4 = a AND Y16 = c THEN X = X12
23	IF Y8 = c AND Y22 = a AND Y27 = b AND Y4 = a AND Y16 = b THEN X = X12
24	IF Y8 = c AND Y22 = a AND Y27 = b AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X3
25	IF Y8 = c AND Y22 = c AND Y4 = a AND Y27 = a AND Y16 = a AND Y31 = a THEN X = X36
26	IF Y8 = c AND Y22 = c AND Y4 = a AND Y27 = b AND Y16 = a AND Y31 = b THEN X = X3
27	IF Y8 = c AND Y22 = c AND Y4 = c AND Y27 = a AND Y16 = a AND Y31 = a THEN X = X36
28	IF Y8 = c AND Y22 = c AND Y4 = c AND Y27 = a AND Y16 = c AND Y31 = a THEN X = X2
29	IF Y8 = c AND Y22 = c AND Y4 = c AND Y27 = a AND Y16 = b AND Y31 = a THEN X = X2
30	IF Y8 = c AND Y22 = c AND Y4 = c AND Y27 = b AND Y16 = a AND Y31 = b THEN X = X2
31	IF Y8 = c AND Y22 = c AND Y4 = c AND Y27 = b AND Y16 = c AND Y31 = b THEN X = X2
32	IF Y8 = c AND Y22 = c AND Y4 = c AND Y27 = b AND Y16 = b AND Y31 = b THEN X = X12
33	IF Y8 = c AND Y22 = c AND Y4 = b AND Y27 = a AND Y16 = a AND Y31 = a THEN X = X46
34	IF Y8 = c AND Y22 = c AND Y4 = b AND Y27 = a AND Y16 = c AND Y31 = a THEN X = X12
35	IF Y8 = c AND Y22 = c AND Y4 = b AND Y27 = a AND Y16 = b AND Y31 = a THEN X = X12
36	IF Y8 = c AND Y22 = c AND Y4 = b AND Y27 = c AND Y16 = a AND Y31 = c THEN X = X27
37	IF Y8 = c AND Y22 = c AND Y4 = b AND Y27 = b AND Y16 = a AND Y31 = b THEN X = X32
38	IF Y8 = c AND Y22 = b AND Y27 = a AND Y4 = a AND Y16 = a AND Y31 = a THEN X = X44
39	IF Y8 = c AND Y22 = b AND Y27 = a AND Y4 = b AND Y16 = a AND Y31 = a THEN X = X55
40	IF Y8 = c AND Y22 = b AND Y27 = a AND Y4 = b AND Y16 = c THEN X = X12
41	IF Y8 = c AND Y22 = b AND Y27 = b AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X55
42	IF Y8 = c AND Y22 = b AND Y27 = b AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X7
43	IF Y8 = b AND Y27 = a AND Y22 = a AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X54
44	IF Y8 = b AND Y27 = a AND Y22 = c AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X44
45	IF Y8 = b AND Y27 = a AND Y22 = c AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X32
46	IF Y8 = b AND Y27 = a AND Y22 = b AND Y4 = a AND Y16 = a AND Y31 = a THEN X = X54
47	IF Y8 = b AND Y27 = a AND Y22 = b AND Y4 = b AND Y16 = a AND Y31 = a THEN X = X43
48	IF Y8 = b AND Y27 = c AND Y22 = a THEN X = X4
49	IF Y8 = b AND Y27 = c AND Y22 = c AND Y4 = a AND Y16 = a AND Y31 = c THEN X = X2
50	IF Y8 = b AND Y27 = c AND Y22 = c AND Y4 = b AND Y16 = a AND Y31 = c THEN X = X33
51	IF Y8 = b AND Y27 = c AND Y22 = b AND Y4 = a AND Y16 = a AND Y31 = c THEN X = X2
52	IF Y8 = b AND Y27 = b AND Y22 = a AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X43
53	IF Y8 = b AND Y27 = b AND Y22 = c AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X33
54	IF Y8 = b AND Y27 = b AND Y22 = c AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X33
55	IF Y8 = b AND Y27 = b AND Y22 = b AND Y4 = a AND Y16 = a AND Y31 = b THEN X = X33
56	IF Y8 = b AND Y27 = b AND Y22 = b AND Y4 = b AND Y16 = a AND Y31 = b THEN X = X10

Table 15

The information of dataset used in our experiments.

Database	Number of instances	Number of features	Number of classes
Our database	677,655	18	64

Table 16

Classification performance symbols and their equations or meanings.

Symbol	Formula/meaning
TN	x_i not predicted to be in C_i and is not actually in it
TP	x_i predicted to be in C_i and is actually in it
FN	x_i not predicted to be in C_i but is actually in it
FP	x_i predicted to be in C_i but is not actually in it
Accuracy	$ACC = \frac{TN+TP}{TN+TP+FN+FP}$

the pilot's comprehensive cognitive sequence activity execution mechanism.

- 2) Piloting decision-making also depends on the pilot's personality type. For instance, in the same navigational environment, when the variable speed change conditions are in a critical state, conservative pilots will not risk changing speed or direction, even if this would lead to great traffic delay. Impulsive pilots are more likely to risk shifting or changing direction.
- 3) Fuzzy ID3 decision tree sets features as main factors determine whether the pilot will take the main piloting force or take rudder operation during navigation, which is consistent with the actual navigation experience. These rules and the current skilled pilot's background knowledge are also consistent. These factors can be an important reference for piloting behavior selection and can also be used to create a knowledge base of an expert system. The results contain high reference value and practical value.
- 4) The classification accuracy of data mining using fuzzy ID3 decision trees can reach more than 86.40% (will be discussed later), close to or even exceeding the effect of an empirical pilot judgment, which well proves the validity of the fuzzy ID3 decision tree algorithm in navigational piloting behavior data mining.

4.4. Comparative analysis

To validate the effectiveness of the fuzzy decision tree algorithm we proposed in this paper, the experiments are conducted with the experimental environment: an Intel (R) Core (TM) i7-5600U 2 Duo Processor 2.6 GHz processor (4 MB Cache), 12 GB of RAM, Windows10, and Python 2.7.14.

In implementing the algorithm, the experimental samples are divided into two categories; we randomly select 80% of samples as a training sample set, and the remaining 20% of samples are used as a test sample set.

To test the performance of the proposed Fuzzy ID3 methods, Support Vector Machine (SVM), and Naïve Bayes (NB) are compared. And we use classification accuracy to measure the proposed Fuzzy ID3 algorithm. Assume the dataset $D=\{X_1, X_2, \dots, X_n\}$, each data record is represented as $X_i=\{X_{i1}, X_{i2}, \dots, X_{in}\}$, D also contains a set of classes $C=\{C_1, C_2, \dots, C_n\}$, then we can get the classification performance symbol and equations or meanings, as shown in Table 15.

For the dataset, described in Table 15, a ten-fold cross-validation (10-CV) is conducted, the performance of different classifier algorithms is shown in Table 17.

The classification accuracy using different classifiers on our dataset is shown in Table 17. According to the classification accuracy results, the proposed method can achieve the highest accuracy among these three algorithms. And the proposed method can ob-

Table 17

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

Classifier algorithms	Performance measures in %	
	Fold	Accuracy
SVM	1	80.57
	2	79.26
	3	81.29
	4	81.89
	5	87.33
	6	79.72
	7	83.33
	8	86.56
	9	83.79
	10	87.11
Average	-	83.09
	NB	Accuracy
NB	1	86.91
	2	82.55
	3	85.75
	4	81.21
	5	80.54
	6	84.23
	7	91.46
	8	82.57
	9	79.75
	10	86.54
Average	-	84.15
	Our method	Accuracy
Our method	1	86.33
	2	85.32
	3	87.23
	4	86.75
	5	85.44
	6	89.12
	7	84.77
	8	86.86
	9	85.46
	10	86.76
Average	-	86.40

tain the best average classification accuracy of 86.40%, followed by NB at 84.15%, and the SVM at 83.09%. Therefore, we can know that the proposed method outperforms the compared methods.

5. Conclusions

Based on the experimental data of the full-task handling simulation platform and in view of the shortcomings of the existing knowledge representation and acquisition methods, in this paper, we use decision trees to integrate the advantages of knowledge representation and acquisition. We combine a decision tree with a fuzzy theory to address the potential uncertainty in the process of classification. We then put forward the knowledge representation and acquisition method based on the fuzzy ID3 decision tree, establish the piloting decision recognition model, and apply it to research on the decision-making mechanism of the different piloting behavior of an inbound ship to verify the performance of the method. We achieve the following conclusions from the simulation results:

- 1) The proposed method uses the fuzzy ID3 decision tree to express the piloting decision recognition model, which has high reasoning efficiency.
- 2) The method integrates the advantages of fuzzy theory and decision trees, combining the comprehensibility of decision trees and the comprehensive expression ability of fuzzy technology. It has strong decision analysis ability and can address the problem of ambiguity and uncertainty. It improves the decision tree's robustness, comprehensibility, and efficiency.

3) This method can mine the key factors which affect piloting decisions, accurately identify the current piloting behavior and provide guidance for a smart ship-assisted or automatic piloting system for the research of human-like piloting behavior.

Furthermore, considering the cost and feasibility of using real ships, computer simulations and simulator experiments are more commonly used. In addition, the number of captain, officers and experienced crew is small, so it is difficult to organize large-scale multi-batch experiments in a certain time and space. With the opportunity of Wuhan University of Technology's training assessment, it is valuable and unique to obtain experimental data operated by an experienced senior crew on the full-task handling simulation platform. However, there are still some problems with the models and experiments described in this paper, which need to be improved in subsequent studies:

- 1) According to the actual situation of the Shanghai Waigaoqiao Phase IV Port and the actual simulation data of an inbound ship, in this paper, we do not take into account the impact of other vessels on the waterway. Based on the definition of the inbound scenario proposed in this paper, there are no other ships interfering with ship OS1 into the port, so there is no need to consider this situation. However, follow-up studies should consider outward ship piloting decisions.
- 2) In this paper, considering the relevant output information of the piloting decision influence factors, only the third level of information is considered. This includes the forward and reverse rotation of the propeller, the propeller speed condition, the rudder angle direction, the rudder angle size and the corresponding maintenance and change conditions. Although we made a detailed division of the information and its guiding significance on the ship into the port piloting decision-making process, the ship rudder is a multi-dynamic factor, so follow-up research needs to do further scientific division and consideration.

In future research, the above problems will be further studied and explored. We hope to improve the ship piloting behavior decision-making theory and system for smart ship piloting behavior decision-making research to provide theoretical guidance and a feasibility basis for the development of smart ships.

Acknowledgments

This study is partially supported by the National Natural Science Foundation of China ([51775396](#), [61703319](#)), the Major Project of Technological Innovation of Hubei Province ([2016AAA007](#), [2017CFA008](#)), and the China Scholarship Council ([201706950088](#)). Grateful acknowledgment is made to Dr. Hongxu Guan, Dr. Chao Chen, Dr. Cunlong Fan, and Dr. Chao Deng who provided us with considerable help and thoughtful comments. In addition, the constructive comments of anonymous reviewers are greatly appreciated.

References

- Chang, P. C., Fan, C. Y., & Duan, W. Y. (2010). A CBR-based fuzzy decision tree approach for database classification. *Expert Systems with Applications*, *37*(1), 214–225.
- Chen, W. H., Liu, C. W., & Tsai, M. S. (2000). On-line fault diagnosis of distribution substations using hybrid cause-effect network and fuzzy rule-based method. *IEEE Transactions on Power Delivery*, *15*(2), 710–717.
- Chen, Z., Wu, C., Zhong, M., Lyu, N., & Huang, Z. (2015). Identification of common features of vehicle motion under drowsy/distracted driving: A case study in Wuhan, China. *Accident Analysis & Prevention*, *81*, 251–259.
- Chen, Z. J., Wu, C. Z., Zhang, Y. S., Huang, Z., Jiang, J. F., & Lyu, N. C. (2017). Vehicle behavior learning via sparse reconstruction with l2-lp minimization and trajectory similarity. *IEEE Transactions on Intelligent Transportation Systems*, *18*(2), 236–247.
- Deng, J. L. (1990). *Grey system theory course* (pp. 55–62). Huazhong University of Science and Technology Publishing House.
- Deng, J. L. (1989). Introduction to grey system theory. *The Journal of grey system*, *1*(1), 1–24.
- Esposito, F., Malerba, D., Semeraro, G., & Kay, J. (1997). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(5), 476–491.
- Fan, C. Y., Chang, P. C., Lin, J. J., & Hsieh, J. C. (2011). A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Applied Soft Computing*, *11*(1), 632–644.
- Janikow, C. Z. (1998). Fuzzy decision trees: Issues and methods. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, *28*(1), 1–14.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms* (2nd ed.). John Wiley & Sons, Inc.
- Liang, J., Zhiqiang, S. H. A., & Long, C. H. E. N. (2012). Dynamic neural network-based integrated learning algorithm for driver behavior. *Journal of Transportation Systems Engineering and Information Technology*, *12*(2), 34–40.
- Lior, R. (2014). *Data mining with decision trees: Theory and applications*: 81. World scientific.
- Martinez, A. M., Webb, G. I., Chen, S., & Zaidi, N. A. (2016). Scalable learning of Bayesian network classifiers. *Journal of Machine Learning Research*, *17*(44), 1–35.
- Moradi, M., & Keyvanpour, M. R. (2015). An analytical review of XML association rules mining. *Artificial Intelligence Review*, *43*(2), 277–300.
- Olaru, C., & Wehenkel, L. (2003). A complete fuzzy decision tree technique. *Fuzzy Sets and Systems*, *138*(2), 221–254.
- Pulkkinen, P., & Koivisto, H. (2008). Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. *International Journal of Approximate Reasoning*, *48*(2), 526–543.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106.
- Quinlan, J. R. (2014). *C4. 5: Programs for machine learning*. Elsevier.
- Sun, J., & Wang, X. (2006). A comparative analysis of rule simplification and pruning fuzzy decision trees. *Computer Engineering*, *32*(12), 210–211.
- Tsang, I. W., Kwok, J. T., & Cheung, P. M. (2005). Core vector machines: Fast SVM training on very large data sets. *Journal of Machine Learning Research*, *6*(Apr), 363–392.
- Umanol, M., Okamoto, H., Hatono, I., Tamura, H. I. R. O. Y. U. K. I., Kawachi, F., & Umedzu, S. (1994, June). Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems. In *Fuzzy systems, 1994. IEEE world congress on computational intelligence., Proceedings of the Third IEEE conference on* (pp. 2113–2118). IEEE.
- Wang, X., Liu, X., Pedrycz, W., & Zhang, L. (2015). Fuzzy rule based decision trees. *Pattern Recognition*, *48*(1), 50–59.
- Wang, X. W., & Jiang, Y. M. (2011). Analysis and improvement of ID3 decision tree algorithm. *Computer Engineering and Design*, *32*(9), 3069–3072.
- Wang, X. Y., & Yang, X. Y. (2008). Study on decision mechanism of driving behavior based on decision tree. *Journal of System Simulation*, *20*(2), 415–420.
- Wang, X. Y., & Yang, X. Y. (2006). Simulation model of driving behavior based on the cubic spline nonparametric fitting regression. *Journal of System Simulation*, *18*(9), 2691–2697.
- Wang, X. Z., Zhai, J. H., & Lu, S. X. (2008). Induction of multiple fuzzy decision trees based on rough set technique. *Information Sciences*, *178*(16), 3188–3202.
- Wang, Z., Wang, Q., & Ai, T. (2014). Comparative study on effects of binders and curing ages on properties of cement emulsified asphalt mixture using gray correlation entropy analysis. *Construction and Building Materials*, *54*, 615–622.
- Xiao, F., Ligteringen, H., Van Gulijk, C., & Ale, B. (2015). Comparison study on AIS data of ship traffic behavior. *Ocean Engineering*, *95*, 84–93.
- Yuan, Y., & Shaw, M. J. (1995). Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, *69*(2), 125–139.
- Zhang, Q. S., Guo, X. J., & Deng, J. L. (1996). Grey relation entropy method of grey relation analysis. *Systems Engineering Theory & Practice*, *8*(8), 7–11.
- Zheng, Z., Zhang, Z. J., Huang, Q. Z., & Li, H. (2012). Research on the integrated cognitive topological structure of driver in the multi-resource information fusion. In *Applied mechanics and materials: 198* (pp. 256–260). Trans Tech Publications.