Modeling Seafarers' Navigational Decision-Making for Autonomous Ships' Safety

Xue, J.

DOI
10.4233/uuid:f0268633-db06-4b40-b298-3b08db93f195

Publication date
2022

Document Version
Final published version

Citation (APA)
Xue, J. (2022). *Modeling Seafarers' Navigational Decision-Making for Autonomous Ships' Safety*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:f0268633-db06-4b40-b298-3b08db93f195

Modeling Seafarers' Navigational Decision-Making for Autonomous Ships' Safety

Jie XUE

Delft University of Technology

Modeling Seafarers' Navigational Decision-Making for Autonomous Ships' Safety

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen

chair of the Board for Doctorates

to be defended publicly on

Wednesday 19 January 2022 at 15:00 o' clock

by

Jie XUE

Bachelor of Engineering in Navigation Technology, Wuhan University of Technology, China

born in Linyi, Shandong, China

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof.dr.ir. P.H.A.J.M. van Gelder	Delft University of Technology, promotor
Dr. E. Papadimitriou	Delft University of Technology, copromotor

Independent members:

Prof.dr. R.R. Negenborn	Delft University of Technology
Prof.dr.ir. L.A. Tavasszy	Delft University of Technology
Dr. J. Szłapczyńska	Gdynia Maritime University, Poland
Dr. R. Szłapczyński	Gdańsk University of Technology, Poland

Other member:

Prof.dr.ir. G.L.L.M.E. Reniers	Delft University of Technology

The research was financially supported by Delft University of Technology (TU Delft) and China Scholarship Council (CSC) under grant 201706950088.

Published and distributed by Jie XUE
Cover illustration by Jie XUE
E-mail: J.Xue@tudelft.nl

ISBN: 978-94-6384-291-4

Printed in the Netherlands

"*We cannot solve our problems with the same thinking we used when we created them*."

—Albert Einstein (1879–1955 AD)

Dedicated to my parents
Yuban Wang and Huiyou Xue

Preface

Steve Jobs said that there is no reason not to follow your heart. On 28th November 2017, I traveled thousands of miles to the place where my heart eagers to reach, a kingdom of tulips, the Netherlands, and became a Ph.D. student of the Delft University of Technology, which completely changed the trajectory of my life. Inadvertently, four years passed by in a flash. When I am looking back on the fantastic four years' journey, countless feelings and emotions fulfill my heart, especially standing at the crossroads of graduation. A thousand words come together into one sentence: Many thanks to everyone who helped and encouraged me!

First and foremost, I would like to sincerely thank my promoter, Prof. Pieter van Gelder, for giving me an opportunity to become a Ph.D. student of TU Delft and a family member of our group. Your consistent patience, broad academic vision, and professional supervision led me to the door of autonomous ship research. Your rigorous academic attitude and unremitting hard-working spirit constantly inspired me to move forward. I enjoyed each meeting and discussion with you and the freedom I got in exploring various topics and sparks of thought. Your encouragement stimulated and made me to became an independent thinker and researcher. I am very grateful to you for everything you have done for me, both in my academia and life. I would also like to express my great appreciation for my co-promotor Dr. Eleonora Papadimitriou for your very careful, efficient, and rigorous supervision. You taught me to pay attention to scientific research details and learn critical thinking. You have strong logic and always could find unique opinions on scientific issues. You also encouraged me to balance academic and daily life. Without your guidance and help, I could not complete my Ph.D. program. Thanks again to you two for the devotion and generous help for me.

Additionally, I would like to thank all the nice colleagues in our warm Safety and Security Science Group. These wise and beautiful colleagues are Prof. Genserik Reniers, Prof. Job Groeneweg, Prof. Fuqiang Yang, Prof. Yunxiao Fa, Dr. Paul Swuste, Dr. Ming Yang, Dr. Simone Sillem, Dr. Irene Grossmann, Dr. Nima Khakzad, Dr. Peter Roelofsma, Dr. Frank Guldenmund, Dr. Yuling Li, Dr. Laobing Zhang, Dr. Yamin Huang, Dr. Dan Jiang, Dr. Pengfei Chen, Dr. Chao Chen, Dr. Wei Ge, Dr. Saba Chockalingam, Dr. Weitao Xiong, Dr. Yunfeng Yang, Dr. Chaoning Lin, Dick Hoeneveld, Karolien van Nunen, Eelco de Groot, Peter Blokland, Karolien van Nunen, Peter Schmitz, Eileen Focke-Bakker, Folkert van Delden, Xin Ren, Rongshui Qin, Xingyuan Zhou, Gexin Shen, Mengxia Li, Rongxin Song, Shuaiqi Yuan, Yuanyuan Xu, Sunyue Geng, Hao Sun, and Yiping Bai. I will always remember all your brilliant smiles and the good memories we experienced. Besides, thanks to the secretaries of our faculty and group, especially Astrid Pinzger-van Staalduinen and Nathalie van den Heuvel, who helped me a lot.

Furthermore, special thanks to Prof. Chaozhong Wu and Dr. Zhijun Chen, who were my promoters at the Wuhan University of Technology and guided, supported, and encouraged me

unconditionally. Moreover, I would also like to thank the colleagues and friends in Wuhan University of Technology: Prof. Xinping Yan, Prof. Ming Huang, Prof. Qing Wu, Prof. Shidong Fan, Prof. Bing Xiong, Prof. Di Zhang, Prof. Yuanqiao Wen, Prof. Nengchao Lv, Prof. Duanfeng Chu, Prof. Kezhong Liu, Prof. Xing Yang, Prof. Junmin Mou, Prof. Yanmin Xu, Prof. Qingxin Yan, Prof. Ming Zhong, Prof. Xiumin Chu, Prof. Zhaozheng Hu, Dr. Hui Zhang, Dr. Feng Ma, Dr. Jinfen Zhang, Dr. Yi He, Dr. Jialun Liu, Dr. Man Zhu, Dr. Chengpeng Wan, Dr. Lei Zhang, Dr. Hongxu Guan, Dr. Ping Wan, Dr. Lixin Yan, Dr. Shanshan Fu, Dr. Chen Chen, Dr. Chao Deng, Dr. Tengfei Wang, Dr. Shiqi Fan, Chang Zheng, Gan Zhang, Chao Gao, Xiaojing Qin, Rongrong Zhu, Gang Li, Yan Li, Xinyu Liu, Li Tang, Qiongzhen Shen and Tan Huafeng, Haoran Li, Yaqiu Li, Wenhui Chu, Haoran Wu, Lian Xie, Liqun Liu, Xin Li, Qi Zhang, Tingting Cheng, Ke Meng, Jinchong Zhang, Ze Li, Zejian Deng, Shuxian He, Yue Hu, Jian Cui, Zhu Meng, Man Yang, Zhicheng Duan, Shidong Liu, Ling Qin, Wei Xiong, Junru Yang, Chaoyong Zhang, Xiaozhao Cheng, Xinpeng Wang, and Zihao Huang. Many thanks for all your support and accompany.

In addition, I would also thank my dear friends in the Netherlands: Xiao Lin, Xukang Wei, Longjian Piao, Meng Meng, Kai Wang, Tian Gao, Yuchen Tang, Renfei Bu, Kailun Yang, Huanhuan Liu, Changrang Zhou, Wenting Ma, Shuhong Li, Kailan Wu, Ni Wang, Hongpeng Zhou, Qiyao Peng, Yingfu Xu, Xiaocong Lyu, Hao Wu, Xiao Deng, Tao Sun, Yu Chen, Yuxin Liu, Ailun Li, Fuqi Liang, Zhaolong Li, Tian Tao, Meng Wang, Xinyi Wang, Dadi Zhang, Dong Zhang, Yiquan Gao, and Jing Zheng, and the other kind friends not listed. Thanks to all of you for giving me lots of unforgettable and valuable memories of this land during the past four years. Besides, I would like to thank my friends met in the Association of Chinese Students and Scholars in the Netherlands. The lovable people are Ding Ding, Minkang Tan, Geng Chen, Tianrun Yang, Yingying Geng, Quanxing Wan, Yixiang Wang, Qizhi Ren, Jing Wang, Han Bao, Simin Chen, Xiaoyang Zhong, Ziying Huang, Zhao Yang, Yiyi Zhang, Haoyu Zhu, Jing Xu, Shubo Zhang, Runfei Liu, Zhongxuan Hou, Xiangjun Chen, Feiyang Ge, Yixin Zhang, Rui Feng, Boyu Chen, Beihong Yang, Yanzhuo Zhou, Jialun Liu, Qiaoyun Lu, Qi Shen, Haoyang Tan, Linghuan Sun, Shiya Xiao, Yujia Liu, Bingjing Lu, Lei Zhang, Yawei Du, and Shuoyuan Liang. Thank you for your selfless support and kind encouragement for me, and I will always remember the days we struggled together.

Moreover, I would like to express my gratitude to the committee members for my Ph.D. defense for reviewing my thesis and giving me very valuable and helpful comments. Besides, I would also like to express my gratitude to these experts in my research domain who helped me. They are Prof. Xinping Yan, Prof. Genserik Reniers, Prof. Hao Hu, Dr. Tsz Leung Yip, Dr. Jie Li, Dr. Ming Yang, Dr. Floris Goerlandt, Dr. Bing Wu, Dr. Valdez Banda Osiris, Dr. Changzhen Li, Dr. Lei Dai, Dr. Feng Xu, Dr. Zhipeng Zhang, Dr. Jakub Montewka, and Dr. Lei Du. I enjoyed the time we shared ideas, and the cooperation yielded many valuable papers.

Thanks to China Scholarship Council (CSC) for the financial support to my research and offers the opportunity for me to enjoy a wonderful journey in the Netherlands.

Last but not least, I would like to express my special gratitude to my dearest and beloved family, 感谢我的父母将我含辛茹苦地养育成人，在我生活、学习上最艰难的时刻给予的无私关怀与鼓励，使我能够安心、专注地完成学业，你们是我不断前进的最大动力，愿你们能够健康、长寿、快乐。此外，感谢所有亲属对我的关心、支持与帮助。

Jie XUE
Delft, November 2021

Contents

List of Figures

List of Tables

List of Symbols

a	the experimental data of each maneuvering decision-making factor are trisected into three levels, "a" means the "small" level
A	a feature/attribute
A^i	the i-th fuzzy attribute
$A=(a',b',c')$	the triangular fuzzy number corresponding to the linguistic term
$A(X)$	the crisp number
b	the experimental data of each maneuvering decision-making factor are trisected into three levels, "b" means "medium" level
c	the experimental data of each maneuvering decision-making factor are trisected into three levels, "c" means "large" level
C	the set of classes
D	the training/example data set
D_{C_k}	a fuzzy subset in D whose class is C_k
$\lvert D \rvert$	the total number of samples or the sum of the membership values in data set D
$\lvert D_{a_j^i} \rvert$	the sum of the membership values of the set of data $D_{a_j^i}$
$\lvert D_{C_k} \rvert$	the sum of the membership values of the set of data D_{C_k}
$\lvert D_j \rvert$	the number of samples in subset D_j
e_k	the element in the example set D
E	the number of error instances that occur after pruning

$E_{\max}$	the maximum value of the expected false positive rate
$E(A)$	the expected information of feature/attribute *A*, i.e., entropy
$E(A^i, D)$	the expected information for the attribute A^i in example data set *D*
$E(X_i)$	the entropy correlation of X_i
f	the actual observed error rate, $f = E/N$
$G(A^i, D)$	the fuzzy information gain for the attribute A^i in example data set *D*
$Gain(A)$	the information gain of feature/attribute *A*
$GainRatio(A)$	the Gain Ratio of feature/attribute *A*
H_{m}	the maximum value of grey entropy
$H(R_i)$	the grey relation entropy of X_i
$I(D)$	the average amount of information for *D*
$I(S_i)$	the amount of information on events S_i
l	the number of fuzzy attribute
N	the total number of instances under the pruned subtree
p_i	a prior probability of the event
P_i	the density value of the grey correlation coefficient distribution
P_k	if we randomly select a sample from *D*, and this sample belongs to class C_k, then we can get a prior probability P_k of the event
$p(S_i)$	the probability of occurrence of event S_i
q	the estimated error rate
R_i	the set of grey correlation coefficient for the sample data
$\lvert S_i \rvert$	the number of samples in data set *D* that belongs to S_i
$\lvert S_j^i \rvert$	the number of samples in subset D_j that belongs to S_i
$S_0^{'}$	the standard deviation of the reference series

$S_i^{'}$	the standard deviation of the comparative series
$Split(A)$	the split information of feature/attribute A
$u_{T_i}(x)$	the triangular membership function for each linguistic term T_i
X	a grey relation factor set (discrete series)
X_0	a reference series
X_i	the comparative series
$X^{'}$	the original data series
$X_0^{'}$	the processed reference series
$X_i^{'}$	the processed comparative series
α	the significance level
$\alpha^{'}$	the confidence level
β	the truth level threshold
β_i	the relative weights of the experts
γ_i	the grey relational grade
λ_k	the weight of each influencing factor
$\lambda_i(x_0(k), x_i(k))$	the relational grade between the reference series and comparative series
μ_{ij}	the corresponding membership degree
$\mu_A(x)$	the membership function for linguistic terms from the judgments of domain experts
$\xi_i(x_0(k), x_i(k))$	the correlation coefficient between the comparative series X_i and the reference series X_0 at point k
ρ	the resolution ratio
ω	the number of influencing factors
$\Delta_i(k)$	the absolute value of the difference between the reference series and each sub-series at each point

$\Delta_i(\max)$	the first-level maximum range
$\Delta_i(\min)$	the first-level minimum range
$\Delta_{\max}$	the second-level maximum range
$\Delta_{\min}$	the second-level minimum range

List of Acronyms and Abbreviations

AAWA	Advanced Autonomous Waterborne Applications Initiative
AI	Artificial Intelligence
AIS	Automatic Identification System
ANFIS	Adaptive Neuro-Fuzzy Inference System
ARPA	Automatic Radar Plotting Aids
CART	Classification And Regression Trees
CDT	Clear Decision Tree
CFP	Crisp Failure Possibility
CHAID	Chi-squared Automatic Interaction Detector
CLS	Concept Learning System
CNC	Computer Numerical Control
COA	Center Of Area
COG	Center Of Gravity
COLREGs	Convention on the International Regulations for Preventing and Collisions at Sea
CPC	Common Performance Condition
CREAM	Cognitive Reliability and Error Analysis Methods
CV	Cross-Validation
DNV	Det Norske Veritas
DOFs	Degrees Of Freedom
ECDIS	Electronic Chart Display and Information System
ENC	Electronic Nautical Charts

FAHP	Fuzzy Analytic Hierarchy Process
FDT	Fuzzy Decision Tree
FER	Fuzzy Evidential Reasoning
FLNG	Floating Liquefied Natural Gas
FN	False Negatives
FP	False Positives
FPSO	Floating, Production, Storage, and Offloading
GMDSS	Global Maritime Distress and Safety System
GPS	Global Positioning System
GRA	Grey Relational Analysis
HDMDR	Human-like Decision-making Maneuvering Decision Recognition
IBS	Integrated Bridge System
ICDM	IEEE International Conference on Data Mining
ICT	Information and Communication Technology
ID3	Iterative Dichotomiser 3
IMO	International Maritime Organization
IoT	Internet of Things
IT	Information Technology
k-NN	k-Nearest Neighbours
LNG	Liquefied Natural Gas
LR	Lloyd's Register
MARISA	MAritime RISk Assessment
MCDM	Multiple Criteria Decision Making
MCG	Maritime and Coast Guard
MOM	Mean Of Maximum
MRV	Monitoring, Reporting and Verification
MSA	Maritime Safety Administration

NT	Network Technology
OOW	Officer On Watch
OS	Own Ship
RICAS	Risk-Informed ship Collision Alert System
SEEMP	Ship Energy Efficiency Management Plan
SQ	Sub-Question
SVM	Support Vector Machine
TRO	Telegraph and Rudder Order
TN	True Negatives
TP	True Positives
VTS	Vessel Traffic Service
WEKA	Waikato Environment for Knowledge Analysis

Chapter 1 Introduction

This thesis focuses on the concept of human-like maneuvering for autonomous ships and studies the human-like decision-making method of autonomous ships. By establishing an independent machine learning method, the maneuvering decision-making mechanism in typical navigation scenarios is explored, and the processes of autonomous acquiring and learning seafarers' maneuvering decision-making characteristics for autonomous ships are studied. Eventually, the autonomous ship human-like decision-making models are constructed. We collected data on the full-task handling simulation platform for merchant ships named Navi-Trainer Professional 5000. Moreover, we used data mining, machine learning, statistical analysis, fuzzy theory, etc., comprehensively to conduct this research. In particular, we applied different decision tree algorithms to study the decision-making mechanisms of different maneuvering behaviors in the specific maritime transport scenario. Then we validate the trees with empirical data from a full-task handling simulation platform to find the optimal decision tree algorithm. Thus to realize the automatic acquisition and representation of a seafarer's decision-making and accurately identify the current maneuvering behavior. This research provides a new perspective and methodology for the development of autonomous ship technology in theory and practice and promotes the application and spreading of autonomous ships. In addition, it also gives theoretical guidance and feasibility bases for the simulation and realization of autonomous ship automatic maneuvering and berthing systems. In this chapter, the background of the thesis is introduced, followed by motivations, research questions, main contributions, and the outline of this thesis.

1.1 Background

Maritime shipping is the lifeblood of the global economy, transports approximately 90% of international merchandise trade (ICS, 2020). According to the statistics, currently, waterway transportation accounts for 95% of total crude oil transportation and 99% of total iron ore transportation, and there are over 50,000 merchant ships trading internationally (AGCS, 2018). Maritime shipping becomes an irreplaceable transportation method. Therefore, the safety of vessels is a critical issue in global maritime shipping.

At the same time, waterway 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 speed of ships is increasing. Coupled with the increase in the transportation of dangerous goods, the density of maritime traffic is increasing, and the navigation environment of ships is deteriorating, causing frequent maritime traffic accidents, which brings more attention to the risk of navigation (Akyuz and Celik, 2014; Goerlandt and Montewka, 2015). Maritime accidents frequently occur, for instance, there were 2712 reported shipping incidents or casualties in 2017 (AGCS, 2018), and hull collisions and damages caused by personnel errors account for more than 80% of maritime accidents (Hanzu-Pazara et al., 2008; Rothblum, 2000), and one of the important ways to solve ship accidents caused by human factors is to utilize autonomous maneuvering of ships. Additionally, the safety of the seafarers in extreme weather conditions in recent years has also become a problem that cannot be ignored (Wang et al., 2014a). Besides, the number of seafarers is declining recently, while the wages of seafarers are rising year by year, which has become the second largest expenditure item after the fuel costs of shipping (Lun et al., 2016).

In addition, with the development of marine technologies, information technologies, and "big data" intelligent applications, autonomous ship emergence is accelerating. Moreover, the world merchandise trade is moving in the direction of informatization and intelligence. Therefore, the study of autonomous merchant ships has become a "hot" topic internationally, as this would reduce the number and necessity for operators/seafarers onboard and increase maritime transport as a more environmental-friendly alternative to transport by trucks on land. The improvement of autonomous ships over the next 10 to 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, autonomous ships are listed as one of the 18 critical 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 autonomous ships. "Manufacturing in China 2025" views marine engineering equipment and high-tech shipbuilding as one of the top ten critical areas in which autonomous ships will be an essential 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 engine room 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 seafarer will be greatly reduced, which will lead to increased ship safety requirements.

Furthermore, although the current level of ship automation is relatively high, the regular operation of ships is always inseparable from human participation (Perera et al., 2015a).

Additionally, when an emergency occurs, it must be handled by the seafarers. Although the ship is maneuvered and supported by Automatic Identification System (AIS), Automatic Radar Plotting Aids (ARPA), Electronic Chart Display and Information System (ECDIS), Global Maritime Distress and Safety System (GMDSS), and ship autopilot system, etc., the ship bridge has not been to the level of unmanned. Autonomous ship technology has developed rapidly in recent years. However, there are still many problems that need to be solved and improved. Moreover, the existing research for automatically achieving the autonomous ships' maneuvering decision-making by acquiring the seafarers' operation characteristics is still scanty. In addition, it also lacks the appropriate theoretical methods to explore the problem of autonomous ship human-like maneuvering decision-making modeling.

Overall, as autonomous ships have outstanding advantages in improving operational efficiency, safety management, decision-making efficiency, and energy consumption management of ships, the research on autonomous ships has become an inevitable tendency for future ship development and gained the interest of many researchers in both academia and industry. Besides, it is necessary to strengthen the relevant theoretical and technical research. The study for the autonomous ship human-like maneuvering decision-making mechanisms is indispensable and meaningful.

1.2 Motivations

This thesis explores the decision mechanisms of different maneuvering behaviors to realize the automatic acquisition and representation of the seafarer's decision-making knowledge in the typical navigation scenario. In order to analyze and reproduce the seafarer's decisions in a typical navigation scenario for autonomous ship maneuvering and let the autonomous ship make decisions like a human, we mainly use data mining, machine learning, fuzzy and grey theories, etc., comprehensively to conduct this research. The main motivations of the research are as follows:

- Research on autonomous ship human-like decision-making mechanisms in the typical navigation scenario.

- Propose an integrated conceptual framework and methodology for modeling human-like decision-making of autonomous ships.

- Help policy-makers and stakeholders make optimal management decisions during the typical navigational scenario for autonomous ships in the coming future.

- Provide an easier understanding of all sorts of choice behaviors in contexts as diverse as traffic safety, marine management, and ocean governance, etc.

1.3 Research questions

The research objective of this thesis is to prioritize safety influencing factors of autonomous ships' maneuvering decisions and develop a series of ship maneuvering knowledge learning models to give autonomous ship the ability to make decisions like a human, i.e., to recognize and model seafarers' navigational decision-making characteristics and mechanisms for autonomous ships' safety in a typical scenario. In order to achieve the above motivations, the main research question addressed in the thesis is as follows:

How can the decision mechanisms of automatic acquisition and representation of the seafarer's *decision-making knowledge in a typical navigation scenario be recognized?*

To answer the main research question, the following research sub-questions (SQ) should be addressed:

Questions on research setup

a) Which data analysis method is more suitable and effective for the selection of the main influencing factors of seafarers' maneuvering decisions?

Although numerous studies in academia have been conducted upon influencing factors assessment based on the grey and fuzzy theories, they seldom consider the relative importance of different influencing factors and lack expertise. In addition, some studies only consider the same weight to determine the judgments from different experts or only use the standard fuzzy number functions to evaluate the linguistic terms given by experts. However, the standard fuzzy membership function sometimes cannot determine different linguistic terms from different domain experts reasonably. In some specific situations, it treats different indexes equally, specifically, the same linguistic term from different domain experts. The answer to the SQ is looking for the method that not only considers the priority of the selection of the main maritime traffic safety influencing factors for seafarers' maneuvering decisions but also has more suitable and applicable for the development of autonomous ships in a specific navigational scenario.

b) Which approaches can be used to automatically acquire and represent the decision-making mechanism of the experienced seafarers' maneuvering behavior in a typical navigation scenario?

The automatic acquisition and representation of ship maneuvering decision-making are essential for getting accurate and rapid ship maneuvering decisions and ensuring water traffic safety. Currently, knowledge acquisition and representation are mainly based on knowledge-based research methods, such as Support Vector Machine (SVM), neural network, statistical analysis, etc. However, there is no unified, comprehensive theoretical system, and there are shortcomings in the evaluation methods. For instance, the SVM method needs to compute 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. At the same time, a neural network model exhibits over-fitting and under-fitting phenomena. The knowledge is implied, not easily tested, and has poor flexibility. Therefore, to address the SQ, the approaches/algorithms that can reasonably complete the decision-making knowledge acquisition of the ship's automatic maneuvering in a specific waterway/port and have a high degree of application and promotion still need to be reviewed and further developed.

Questions on modeling

c) What are the advantages and disadvantages of the prioritizing model of safety influencing factors of autonomous ships' maneuvering decisions?

The seafarers' maneuvering decision-making is always influenced by multi-source information, for instance, the other ships in waterways/ports, the natural environmental factors, etc. (Kim et al., 2017). This requires ship maneuvering decision-making procedures expressed along with higher accuracy and effectiveness. However, due to the limited information acquiring and processing capacity, the Officer On Watch (OOW) cannot achieve the multi-attribute or multi-source information in a particular time and space concurrently. For instance, under high-intensity work pressure, the OOW cannot always ensure to make correct decisions timely when facing constant changing factors in different navigation scenarios; thus, maneuvering decisions cannot still be made accurately and quickly, which could lead to maritime traffic accidents. Therefore, the automatic acquisition and representation of maneuvering decision-making are necessary, and it is essential to identify, analyze, prioritize the main influencing factors for efficient selection of the corresponding maneuvering decisions for autonomous ships. The answer to this SQ is to establish the safety influencing factors' prioritizing model for autonomous ships' maneuvering decisions and evaluate the performance of the proposed model in respect of advantages and drawbacks.

d) How can the maneuvering decision-making processes of experienced seafarers under the typical navigation scenario be modeled?

The ship maneuvering process is a multi-functional integrated system integrating multiple automation systems. However, the improvement of the degree of automation of ships has a certain gap from the ships with automatic perception, subjective analysis, and autonomous decision-making. The answer to this SQ is to explore the decision-making mechanisms of different maneuvering operations in order to realize the automatic acquisition and representation of a seafarers' decision-making and develop a series of methods for ship maneuvering knowledge learning models to give autonomous ships the ability to make decisions like a human.

e) How to evaluate and maintain the proposed Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model to ensure its appropriate functioning throughout its entire life cycle?

Based on the actual experienced seafarers' operational data from the full-task handling simulation platform, this thesis uses the C4.5 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, i.e., an autonomous ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model. It is crucial to ensure and evaluate the performance of the model and have a clear way to update the proposed model to maintain it has wide applicability. The answer to this SQ addresses the way to evaluate and update the proposed model throughout its entire life cycle.
Questions on application

f) To what extent could the proposed models in this thesis be applied in reality?

The final objective of this study is to achieve the Degree Four: fully autonomous ship (IMO, 2021) in a specific scenario for a particular ship type and explore the possibilities for a wide application for the proposed model for various scenarios and types of ships. However, considering the reality at this stage, there are still many challenges on the way to achieving fully autonomous shipping, and lots of conditions need to be satisfied. The answer to this SQ is to illustrate the detailed applicable scenarios and objects for the proposed models, explain how they would be extended at this stage, and present the advantages and significance for the models to be applied in the real-life maritime domain.

1.4. Main contributions

In view of the shortcomings of the existing knowledge representation and acquisition methods, this thesis mainly focuses on the problem of modeling seafarers' navigational decision-making in the typical scenario for autonomous ships' safety. We propose the method to prioritize safety influencing factors of autonomous ships' maneuvering decisions and a series of ship maneuvering knowledge acquiring and learning models to give autonomous ships the ability to make decisions like a human. In particular, we propose a novel method that has not been previously tested, which combines decision trees with fuzzy theories to identify the seafarers' decision-making knowledge in a typical navigation scenario. Moreover, we develop the knowledge representation and acquisition method based on different decision tree algorithms, establish the maneuvering decision recognition model, and then verify the performance of the proposed method.

The decision mechanisms of different maneuvering behavioral patterns and the execution mechanisms of ship operating modes are two important steps in simulating task aggregation and multi-source information stimulation (Xiao et al., 2015; Zheng et al., 2012). In addition, while we know of no other studies on autonomous ship maneuvering decision-making based on the experimental data from a simulation platform, several efforts have been made to construct decision mechanism using the simulator in other research domains (Kennedy et al., 2010; Paschalidis et al., 2018). Therefore, it is essential and necessary to test the proposed method for ensuring accurate and rapid maneuvering decisions and maritime traffic safety.

This thesis provides theoretical guidance and a feasibility basis for research into seafarers' maneuvering decision-making and the realization of autonomous ships development and practical application of subsequent unmanned merchant ships as well as autonomous ship piloting and berthing systems. It is of significance to ensure transport safety, manage transport risks, reduce transport uncertainties, and prevent potential losses. The innovations of this research are as follows:

- In this thesis, the problem of autonomous ship maneuvering is considered a machine learning issue, which transforms the problem of autonomous ship maneuvering decision-making into a problem of establishing machine learning method to learn the seafarer's maneuvering decision-making characteristics and constructing multi-constraint decision model.

- "Autonomous" 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.

- This research considers the main influencing factors on the maneuvering rules in the typical navigation scenario. It solves the problem that the implementation of the experiment for an actual merchant ship in the real world is challenging due to the objective conditions of cost, feasibility, and other factors. It is unique and very valuable to obtain experimental data operated by an experienced senior seafarer on the full-task handling simulation platform. The number of experienced crew, especially captain and chief officers, is small, so it is challenging to organize large-scale multi-batch experiments in a certain time and space. Additionally, the cost of using the real 30,000-ton ship to carry out this kind of experiment is very high, and the feasibility of collecting the data of multiple voyages from this kind of ship is too slow. Moreover, it takes almost over ten years for the seafarer to grow from a third officer to a chief officer or a captain, so the data acquisition is rare and valuable.

- This research establishes an autonomous ship human-like decision-making model with the optimal decision tree algorithms based on a comparative study of different decision tree theories, which can accurately characterize the seafarer's perception, decision-making, and operation process. The impact of the main influencing factors during the seafarer's decision-making process is considered in our model, and it can reflect the seafarer's decision-making characteristics in a typical traffic scenario. The proposed model provides new insights and methods for the development of autonomous ship technology both in theory and in practice and promotes the application and promotion of autonomous ships.

1.5 Thesis outline

This thesis is divided into seven chapters and the schematic overview of these chapters and their relationships are demonstrated in Figure 1.1.

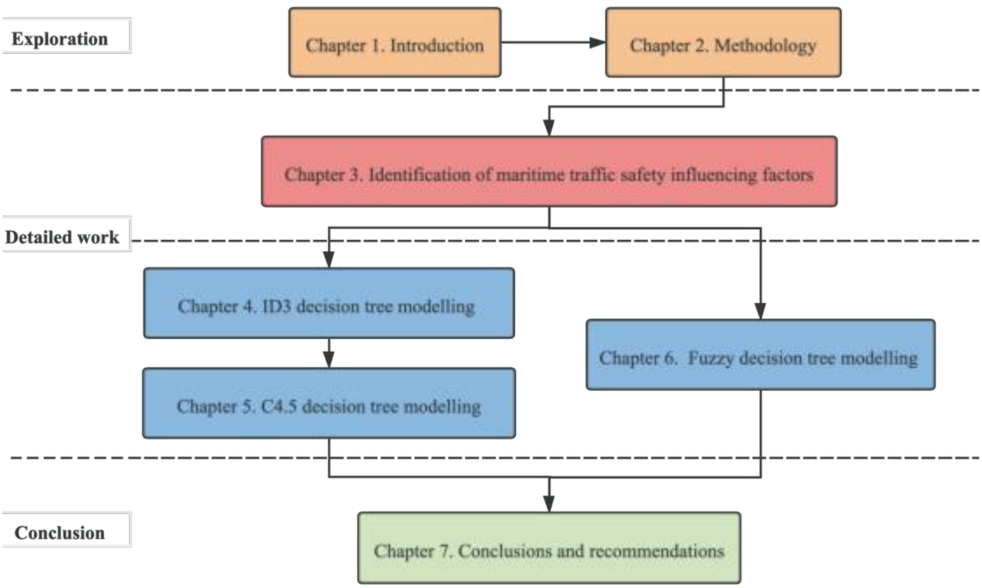


Figure 1.1 Outline of the thesis.

Chapter 1 introduces the background, motivations, research questions, and main contributions of this thesis.

Chapter 2 provides a comprehensive overview of the literature relevant to grey system theory, fuzzy theory, and decision tree algorithm and identifies the material contributing to our research. The overall background of knowledge acquisition and representation methods are briefly introduced, the advantages and disadvantages of these methods are discussed, and the related gaps are demonstrated.

Chapter 3 illustrates the multi-attribute decision-making method-based model for prioritizing maritime traffic safety influencing factors of autonomous ships' maneuvering decisions. The framework of the prioritizing model and the applying framework is presented for the real-world merchant ship.

Chapter 4 proposes an ID3 decision tree model for recognizing human-like decisions of autonomous ships in the specific ship maneuvering scenario for the first exploration and pre-study. This chapter gives a detailed description of the experimental scenario design and the preparations and primary conditions for our research setup. In addition, the standardization principle of ship maneuvering is introduced.

Chapter 5 develops a C4.5 decision tree model for human-like decision-making of autonomous ships to acquire the knowledge under multiple environmental constraints to give autonomous ships the ability to make decisions like a human: An autonomous ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model. The decision-making mechanism for the maneuvering behavior of Officer On Watch (OOW) is analyzed, and the OOW's decision-making knowledge is automatically acquired and represented.

Chapter 6 demonstrates a fuzzy decision tree model for human-like decision-making of autonomous ships, which could address problems of fuzziness and uncertainty, possess a high reasoning speed, and can accurately identify the experienced seafarers' operations. The model acquires the decision mechanisms and knowledge of various berthing operations and realizes the simulated reproduction of the seafarers' maneuvering behavior under the scenario of inbound ship analysis.

Chapter 7 summarizes the main findings and conclusions of this thesis, and the recommendations for future research are provided as well.

Chapter 2 Methodology and relevant models

In order to address the proposed research questions, this chapter presents a comprehensive overview of the literature relevant to the grey system theory, fuzzy theory, decision tree algorithm and identifies the materials contributing to our research. In Section 2.1, the overall background of knowledge acquisition and representation methods is briefly introduced, and the reason for the selection of these theories and the combination of the methods for implementation in this thesis are detailed. Then, the specific illustration for the grey system theory and the representative research is demonstrated in Section 2.2. In Section 2.3, the introduction about the fuzzy theory and relevant literature review is given, especially the research in various aspects of the maritime domain. Finally, the mechanism of the decision tree algorithm is detailed, and the advantages and disadvantages of different decision tree algorithms are discussed in Section 2.4.

This chapter is based on the following papers:

Xue, J., van Gelder, P. H. A. J. M., Reniers, G., Papadimitriou, E., & Wu, C. (2019). Multi-attribute decision-making method for prioritizing maritime traffic safety influencing factors of autonomous ships' maneuvering decisions using grey and fuzzy theories. *Safety Science*, *120*, 323-340.

Xue, J., Chen, Z., Papadimitriou, E., Wu, C., & van Gelder, P. H. A. J. M. (2019). Influence of environmental factors on human-like decision-making for intelligent ship. *Ocean Engineering*, *186*, 106060.

2.1 State-of-the-art and models selection

The accuracy of maneuvering decisions of seafarers is directly related to the safety of maritime traffic. In the process of decision-making, seafarers' maneuvering operation is often influenced by multi-source information from the seafarer themselves, the ship, and the environment. Due to humans' limited information-processing capacity, the seafarer cannot consistently achieve knowledge acquisition and representation of the multi-source information so that maneuvering decisions can be carried out accurately and quickly, leading to maritime traffic accidents.

The decision mechanisms of different seafarers' maneuvering behavioral patterns and the execution mechanisms of ship operating modes are two important steps in simulating task aggregation and multi-source information stimulation (Wang and Yang, 2008; Wang and Yang, 2006; Xiao et al., 2015; Zheng et al., 2012). Therefore, the automatic acquisition and representation of maneuvering decision-making is essential in ensuring accurate and rapid maneuvering decisions and maritime traffic safety. The selection of different evaluation methods will affect the seafarers' decision-making efficiency and accuracy, thus affecting maritime traffic safety. At present, the methods of knowledge acquisition and representation are mainly based on knowledge-based research methods, such as fuzzy theory (Chen et al., 2000), Support Vector Machine (SVM) (Tsang et al., 2005), statistical analysis (Chen et al., 2015), rule extraction (Martınez et al., 2016; Moradi and Keyvanpour, 2015), neural networks (Liang et al., 2012), and 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 SVM method can solve optimization problems, but it needs to compute Hessian 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 maneuvering decision-making methods needs to be improved and further developed.

The grey system theory, proposed by Deng (Deng, 1982, 1989), is one of the most widely utilized pattern recognition methods. It is mainly utilized to analyze the proximity of the dynamic grey process development situation, determine the primary and secondary factors in the grey system, and control the main factors affecting the system (Huang et al., 2013). Grey system theory is suitable for multiple inputs and uncertain data. It can be utilized to resolve uncertainty problems, under partial information and discontinuous data effectively (Kumar et al., 2018). The grey relational analysis (GRA) is an effective algorithm for resolving uncertainty problems in the case of partial and discontinuous information (Deng, 1982). However, the traditional GRA has been largely criticized because it treats different indexes (influencing factors) equally and does not take the relative importance of different indexes into consideration. It does not fit with people's preferences for a specific index. Nevertheless, the fuzzy logic theory is a beneficial method for modeling processes which are too complicated for conventional quantitative analysis or information obtained from the process is qualitative, uncertain or inexact (Abbassi et al., 2017; Balin et al., 2018; Tseng and Cullinane, 2018; Zhou and Thai, 2016). Moreover, fuzzy numbers are more compatible with phrases and ambiguities; it is better to use them in real-world decision-making and reflect human thoughts (Hatefi and Tamošaitienė, 2018).

Decision tree is one of the most widely used classifiers in the machine learning research domain, mainly due to its straightforward model, its speed in classifying new patterns, and the ease with which the classification rules (which can also be graphically represented) can be understood. Specifically, the decision tree can handle both categorical and numerical data. It is good at processing non-numeric data, which can eliminate a significant amount of data preprocessing work when dealing with numerical data through algorithms, such as neural networks. Besides, the decision tree method is simple in structure and does not need much background knowledge in the process of learning. In addition, the decision tree model is more efficient and is more suitable for training sample sets with large amounts of data. Furthermore, 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 non-numeric data. Finally, the decision tree method has a higher classification accuracy. Therefore, the decision tree method is a crucial research direction in the field of machine learning.

Ship maneuvering 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 is 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 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 maneuvering.

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 and Wehenkel, 2003).

Therefore, in our thesis, we comprehensively utilize the grey system theory, fuzzy theory, and decision tree algorithm to address the research questions so as to obtain the objectives of this thesis. In the following sub-sections, we give an overview of these relevant contemporary studies in the literature relating to the above theories and algorithms and identify material that contributes to our research.

2.2 Grey system theory

Grey system theory is characterized by an uncertain system in which "partial information is known and some information is unknown". Through the research on some known information, the system can be accurately understood (Liu and Forrest, 2010). Specifically, as shown in the typical grey system in Figure 2.1, if white represents completely clear data/information, and black represents completely unknown data/information, grey represents other data/information that is known partially. If a system contains grey information, it can be called a grey system.

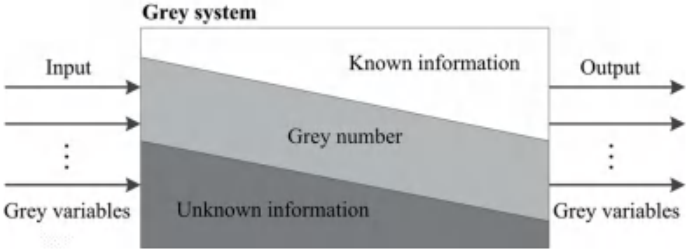


Figure 2.1 The concept of the grey system.

After more than twenty years of development, the grey system theory has penetrated many scientific research fields and has been confirmed and developed. It provides a new insight into to solve system problems in the case of poor information (Li, 1996). In order to analyze the system behavior of grey systems with uncertain information, the grey system theory develops a series of comprehensive analysis methods of grey systems, such as the Grey Relational Analysis (GRA) (Fu et al., 2017; Hao et al., 2017; Lee et al., 2018; Lilly Mercy et al., 2017; Rajesh et al., 2013).

Specifically, the GRA method is suitable for the data with uncertain, multiple inputs and discrete properties; it does provide techniques for determining an appropriate solution for real-world problems. Moreover, the GRA does not require too much sample size and does not require a typical distribution law during analysis. In addition, regardless of whether the system has adequate information, the GRA could capture the impact of the relationship between the main factor and influencing factors in the system (Deng, 1989; Shen and Du, 2005). As a systematic analysis technique, the GRA is a quantitative comparative analysis method, by calculating the correlation between the target value and the influencing factors, and the ranking of the relevance, the main factors affecting the target value are sought (Deng, 1982; Liu et al., 2010). The results are corresponding to the qualitative analysis results, so the method has wide practicality (Chen and Ting, 2002; Deng, 1989).

The GRA is applied to many research domains, for example, it was adapted to study the research output and growth of countries (Javed and Liu, 2017), and utilized to investigate the nonlinear multiple-dimensional model of the social economic activities' impact on the city air pollution (Li et al., 2017). In addition, Lu et al. (2010) applied a mathematical approach and GRA to analyze the traffic situation trends of China and investigate the potential solutions for enhancing road traffic safety. Wang et al. (2007) proposed a grey model-based smoothness predictions; the results showed that the model provides promising results and is useful for evaluating the riding quality of pavement performance. Zhou and Thai (2016) utilized GRA and grey theory to evaluate the failure modes and analyze the effect for tanker equipment failure prediction; the priority ranking results show that both fuzzy theory and grey theory are quite similar and the proposed fuzzy and grey Failure Mode and Effects Analysis (FMEA) method is more practical and flexible for risk evaluation with respect to tank shipping. Rajesh et al. (2013) introduced the optimization steps to investigate the effects of different operations in the Computer Numerical Control (CNC) machine by using the GRA with entropy. Hatefi and Tamošaitienė (2018) presented a novel improved GRA method to evaluate construction projects on the basis of the sustainable development criteria in social, economic, and environmental dimensions using experts' opinions.

2.3 Fuzzy theory

Fuzzy theory was introduced by Zadeh (1965) to solve uncertainty on decision-making by extending the traditional notation of sets. Fuzzy logic is a type of multi-valued logic. The truth values of variables are considered to be "fuzzy" may be any real number within the unit interval [0,1] (Novák et al., 2012). It is an effective method to design a system for decision-making, and it can be used to solve the problems related to conducting inaccurate and uncertain data (Balmat et al., 2011). Zadeh (1965) proposed the fuzzy sets in 1965, and it provides a useful mathematical tool for reliability analyses and to solve system vagueness and uncertainty on decision-making by extending the traditional notation of sets (Zadeh, 1983). A membership function specifies and assigns a value between 0 and 1 in the usual case for each element of discourse. The assigned value is called a membership degree and determines the extent to which a given element belongs to the fuzzy set. Besides, any fuzzy set can be uniquely determined by its membership (Wang et al., 2009; Zhou et al., 2018).

Fuzzy numbers are cases of fuzzy sets, and the most commonly used fuzzy numbers are trapezoidal and triangular fuzzy numbers (Hadi-Vencheh and Mokhtarian, 2011). In addition, the triangular fuzzy numbers have the advantages of promoting representation and processing imprecise information due to its computational simplicity (Pedrycz, 1994). In practical applications, fuzzy membership functions are utilized to convert the linguistic estimations into fuzzy numbers for quantitative evaluation. The triangular membership functions are shown in Figure 2.2, and respectively defined as follows:

$$\mu_A(x) = \begin{cases} 0, & x < a' \\ (x-a')/(b'-a'), & a' \le x \le b' \\ (c'-x)/(c'-a'), & b' \le x \le c' \\ 0, & x > c' \end{cases}. \tag{2.1}$$

In practical applications, linguistic estimations are converted into fuzzy numbers using fuzzy membership functions for quantitative evaluation. Assume $\tilde{a}=(a_1,a_2,a_3)$ and $\tilde{b}=(b_1,b_2,b_3)$ are two triangular fuzzy numbers, then the basic fuzzy arithmetic operations with these fuzzy numbers are defined as follows (Wang et al., 2009). Addition: $\tilde{a}+\tilde{b}=(a_1+b_1,a_2+b_2,a_3+b_3)$; Subtraction: $\tilde{a}-\tilde{b}=(a_1-b_1,a_2-b_2,a_3-b_3)$; Multiplication: $\tilde{a}\times\tilde{b}=(a_1\times b_1,a_2\times b_2,a_3\times b_3)$; Division: $\tilde{a}\div\tilde{b}=(a_1\div b_1,a_2\div b_2,a_3\div b_3)$.

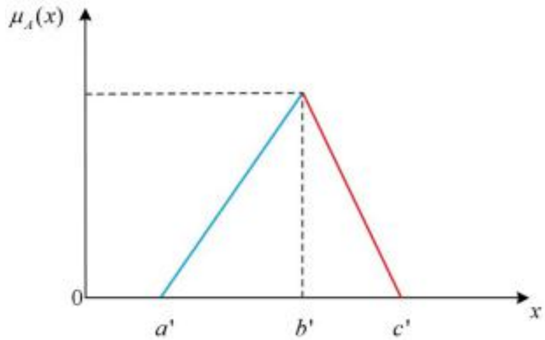


Figure 2.2 Triangular membership functions.

In the maritime domain, many studies using fuzzy theories have been implemented. For instance, in the aspect of shipping accident risk analysis and prevention, Senol and Sahin (2016) used the defuzzification process of fuzzy logic to transform the fuzzy numbers from Crisp Failure Possibility (CFP) to fault probability and proposed a dynamic real-time continuous fuzzy fault tree model for the analysis of ship collision and grounding. Balmat et al. (2011) applied a novel fuzzy technique to conduct a maritime risk assessment for the prevention of pollution on the open sea based on the decision-making system named Mritime RISk Assessment (MARISA). Yang and Wang (2015) developed an approach for analyzing engineering system risks based on a Fuzzy Evidential Reasoning (FER) method, and applied it to the safety modeling of an offshore engineering system, then performed the failure criticality analysis in a collision of a Floating, Production, Storage, and Offloading (FPSO) system with a shuttle tanker during tandem unloading operations. Celik et al. (2010) proposed a risk-based modeling algorithm on the basis of the fuzzy extended fault tree analysis to enhance the implementation process of the investigation for shipping accident; this approach allows accident stakeholders to clarify the technical failures that lead to the shipping accident. Yang et al. (2009) proposed a systematic framework to process the subjective maritime security assessment information based on the fuzzy evidential reasoning approaches. Goerlandt et al. (2015) developed a framework named: Risk-Informed ship Collision Alert System (RICAS), the result of the case-study for RICAS shows that it has an effective performance. Marken et al. (2015) used a fuzzy bow-tie analysis method to quantify the risk of delay for ships sailing in the northern sea route.

In addition, some fuzzy theory-based studies done for the reliability analysis for the human error and offshore operation issues for the shipping industry. Ung (2015) developed a novel fuzzy Cognitive Reliability and Error Analysis Methods (CREAM) methodology considering the weight of each Common Performance Condition (CPC), and validated the method using two axioms and demonstrated by the case of an oil tanker. Zhou et al. (2018) introduced a Bayesian network and fuzzy model for the quantitative analysis of human reliability of tanker shipping industry; the results show that the proposed model is up-and-coming and is in accordance with the CREAM approach. Similarly, Zhou et al. (2017) also proposed a quantitative CREAM method to estimate the human error probability in tanker operational safety using Fuzzy Analytic Hierarchy Process (FAHP) to establish a fuzzy congruous matrix. Abdussamie et al. (2018b) proposed an Adaptive Neuro-Fuzzy Inference System (ANFIS) algorithm to predict the ultimate strength reduction of locally corroded steel plates suffering from pitting corrosion for the marine structures. Abdussamie et al. (2018c) also developed a rule-based fuzzy logic model to calculate operational risk values of the transport barges and the offshore structure being loaded as well as the potential impacts on the safety of seafarers and environment. Rahman et al. (2019) proposed a robust logistics risk model based on the fuzzy and evidence theory to analyze criticality of the contributing factors for offshore oil and gas operations.

Moreover, the location selection problem is another aspect of concern in the academia. For instance, Wu et al. (2018) developed a fuzzy multiple attribute decision-making approach to select the location of an offshore wind farm in the busy waterway of the Eastern China Sea, the proposed method considered the maritime safety and economic feasibility of installation and determined an optimal site selection scheme for the wind farm. Guneri et al. (2009) conducted the shipyard location selection question based on the fuzzy analytical network process algorithm, which provided reference to the decision makers based on quantitative analysis.

Furthermore, many studies are explored by combining expert knowledge with fuzzy theories.

Such as, Abdussamie et al. (2018a) presented a rule-based fuzzy set approach to deal with the uncertainty of expert knowledge used for qualitative risk assessment for the hazardous scenarios of berthing operations of Liquefied Natural Gas (LNG) carrier and Floating LNG (FLNG) in open sea. Akyuz et al. (2016) integrated fuzzy rule-based expert system into fuzzy FMEA to identify potential failure and enhance maritime safety. Kose et al. (1995) presented an intelligent expert system for monitoring vessel safety by using the fuzzy logic inference engine. Perera et al. (2010) presented a fuzzy inference system for collision avoidance based on the expert knowledge and the International Maritime Organization Convention on the International Regulations for Preventing and Collisions at Sea (COLREGs) under critical situations.

Also, fuzzy theories are applied to the research area of ship maneuvering and performance evaluation of the management of shipping company. Bhattacharyya et al. (2011) illustrated a mathematical fuzzy autopilot algorithm for nonlinear maneuvering of surface ships and its performance has been found acceptable. Surendran and Kiran (2007) used the fuzzy logic control algorithm to reduce the roll motions of a ship by active fins; the algorithm proved to be versatile and can be utilized for irregular sea conditions. Wei et al. (2019) put forward a fuzzy algorithm to plan the variable values for hybrid boarding system to compensate the wave disturbance in roll direction as well as other disturbances. Chou and Liang (2010) dealt with an application for the performance evaluation of shipping company through the proposed fuzzy Multiple Criteria Decision Making (MCDM) model.

2.4 Decision tree algorithm

As we all know, 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.

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 tree 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 non-category 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.

Figure 2.3 shows an example for a typical decision tree model based a small training set shown in Table 1. From Figure 2.3, we can see that a decision node/attribute (e.g., Outlook, which represents the weather condition of a particular day) has two or more branches/values (e.g., Rainy, Overcast and Sunny, which represent several unique values of each attribute). Leaf node (e.g., Play) represents the class category or decision of each instance

Table 2.1 A small training set.

No.	**Attributes**				**Play /Class**
	Outlook	**Temperature**	**Humidity**	**Wind**	
1	Rainy	Mild	High	Strong	No
2	Sunny	Hot	Normal	Strong	Yes
3	Overcast	Mild	Normal	Strong	No
4	Overcast	cool	Normal	Weak	Yes
5	Rainy	Hot	Normal	Strong	No

Furthermore, the final decision can also be represented through the form of IF-THEN rule set shown as follows:

Rule 1: IF Outlook=Sunny THEN Play=Yes
Rule 2: IF Outlook=Overcast AND Wind=Strong THEN Play=No
Rule 3: IF Outlook=Overcast AND Wind=Weak THEN Play=Yes
Rule 4: IF Outlook=Rain THEN Play=No

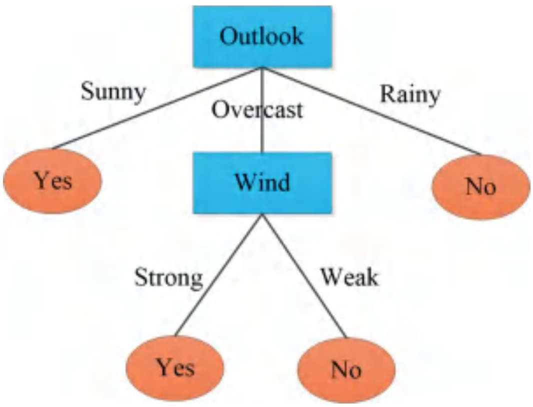


Figure 2.3 A decision tree generated using the small training set.

This example indicates whether a particular weather condition is suitable or not for some unspecified activity in a particular day. In this way, the attributes taken together provide a zeroth-order language for characterizing objects in the universe (Quinlan, 1986).

In the 1960s, decision tree algorithm was initially proposed by Hunt et al. (1966) to minimize the cost of classifying an object (Quinlan, 1986). The decision tree method is a key research direction in the field of machine learning. Muchoney et al. (2010) proposed the classification algorithms of decision tree, 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. 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; Quinlan, 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 algorithms, 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 or CHAID only supports building binary trees, while C4.5 allows two or more outcomes and supports binary or multi-fork trees (Wu et al., 2007). Several prior studies on the C4.5 decision tree could be found from the literature. A prior study (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 C4.5 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, the proposed algorithm could get excellent accuracy than the normal 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. Some more detailed technical information on the related algorithms will be provided in Chapters 4-6.

2.5 Conclusions

This chapter provides a comprehensive overview of the literature relevant to grey system theory, fuzzy theory, and decision tree algorithm that contribute to our research. The state-of-the-art research is highlighted, and the advantages and limitations of various methods relevant to our thesis are addressed. In this thesis, grey system theory is introduced to prioritize the influencing factors; the fuzzy theory is presented for more rational use of expert knowledge for judging the prioritization of the influencing factors and thus to fuzzify the experimental dataset into several language items for the process of constructing decision trees. In addition, as scholars have proposed several different decision tree algorithms for both classification and decision-making problems in various domains and obtained good results. Based on the advantages of the decision tree algorithms and the ability to analyze the characteristics of multi-fork trees, this thesis aims to utilize various decision tree algorithms to learn the seafarers' maneuvering decision characteristics. The autonomous ship human-like maneuvering decision-making problem is regarded as a machine learning problem based on the experts' knowledge, the seafarers' actual maneuvering data, and various influencing factors. The problem is converted using the decision tree algorithms to learn the seafarers' maneuvering decision-making characteristics, thus constructing a series of optimal human-like decision-making models under multiple constraints.

Chapter 3 Prioritizing safety influencing factors of autonomous ships' maneuvering decisions

Ship maneuvering decisions are influenced by several factors, and it is essential to prioritize the main influencing factors for efficient selection of the corresponding maneuvering decisions. Meanwhile, the autonomous ship maneuvering decision-making influencing factors constitute a typical grey system, which is suitable for research by grey relational analysis. Furthermore, linguistic assessment of factors is evaluated to obtain priorities numbers through the fuzzy approach. Therefore, this chapter mainly focuses on the concept of human-like maneuvering for autonomous ships. Based on experimental data of experienced seafarers and using a simulation platform under the scenario of the Shanghai Waigaoqiao wharf, an inference model utilizing grey and fuzzy theories is proposed. The proposed model is combined with expert linguistic terms to select the ship maneuvering decision-making main influencing factors from multi-source influencing factors (in overall and separated categories of natural environment, ship motion, force parameters, draft, and position), and to study the decision-making prioritization for maritime traffic safety for specific ship maneuvering scenarios. This method can prioritize the main factors which affect maneuvering decisions as well as guide an autonomous ship-assisted or automatic maneuvering evaluation system for the research of human-like maneuvering behavior. This chapter provides a new perspective on the identification of main ship maneuvering decision-making influencing factors in theory and in practice. It can be utilized for better decision-making concerning maritime traffic safety of autonomous ship maneuvering, which in turn makes shipping safer and promote the application and spreading of autonomous ships.

This chapter is organized as follows. First, Section 3.1 illustrates the background of this chapter, followed by the methodology and specific steps of our proposed model are described in Section 3.2. The experimental processes are introduced in Section 3.3 and Section 3.4 details the results of our experiment. Then, the discussions of the results are represented in Section 3.5. Finally, the conclusions are addressed in Section 3.6.

The content of this chapter is an edited version of the following published paper:

Xue, J., van Gelder, P. H. A. J. M., Reniers, G., Papadimitriou, E., & Wu, C. (2019). Multi-attribute decision-making method for prioritizing maritime traffic safety influencing factors of autonomous ships' maneuvering decisions using grey and fuzzy theories. *Safety Science*, *120*, 323-340.

3.1 Introduction

Maritime shipping industry transports approximately 90% of international merchandise trade (ICS, 2020). According to the statistics, there are over 50,000 merchant ships trading internationally (AGCS, 2018). Therefore, the safety of vessels is a critical issue in global seaborne transport. In addition, with the development of computer science and technology, especially the rapid development of technologies and theories such as The Internet of Things (IoT), Information Technology (IT), and Artificial Intelligence (AI), the world merchandise trade is moving in the direction of informatization and intelligence. Thereupon, the study of autonomous merchant ships has become a "hot" topic internationally, as this would reduce the need for operators/seafarers onboard, and increase maritime transport as a more environmental-friendly alternative to transport by trucks on land. Several large companies have started to test such vessels, for instance, the Advanced Autonomous Waterborne Applications Initiative (AAWA) project of Rolls-Royce Holdings plc (Rolls-Royce, 2018). In addition, for the shipping industry, advancements in Network Technology (NT), Information and Communication Technology (ICT), and IT create new opportunities for developing electrical systems such as ships autonomous navigation (Lee et al., 2009; Perera et al., 2015b), Integrated Bridge System (IBS), and decision support system (Pietrzykowski et al., 2017), and the level of shipping modernization has been rapidly improved (Pazouki et al., 2018). The development of autonomous ships has been technically feasible. In addition, the economy of the world is experiencing a period of slow-moving recovery; thus shipping industries are falling into the long-term overcapacity. Hence the world's major shipping companies have to shift their development planning to improve the operational efficiency and enhance the safety management of their merchant fleet, in order to reduce the seaborne transport costs and adapt to the market tendency. Moreover, the demands of ship owners and seafarers for safety and profitability of shipping are constantly increasing; it is also an essential influencing factor for the development of autonomous ships.

Furthermore, since the implementation of the international energy conservation and emission reduction rules and regulations promoted the development of autonomous ships, the EU's Monitoring, Reporting and Verification (MRV) regulations for greenhouse gas emissions of the shipping industry took effect on July 1, 2015, and began to monitor emissions according to MRV regulations on January 1, 2018. In addition, all ships larger than 5,000 gross tons and berthed in EU ports are required to meet MRV regulations. Moreover, the International Maritime Organization (IMO) has the program to start emissions monitoring under the Ship Energy Efficiency Management Plan (SEEMP) on January 1, 2019 (IMO, 2018). Besides, the number of seafarers in the world is declining recently, while the wages of seafarers are rising year by year, which has become the second largest expenditure item after the fuel costs of shipping (Lun et al., 2016). At the same time, maritime accidents frequently occur, for instance, there were 2712 reported shipping incidents/casualties in 2017 (AGCS, 2018), and hull collisions and damages caused by human errors account for more than 80% of marine accidents (Hanzu-Pazara et al., 2008; Rothblum, 2000). In addition, the safety of the seafarers in extreme sea conditions in recent years has also become a problem that cannot be ignored (Aziz et al., 2019; Baksh et al., 2018; Khan et al., 2018; Wang et al., 2014b).

In summary, as autonomous ships have outstanding advantages in improving operational efficiency, safety management, decision-making efficiency, and energy consumption management of ships, research for autonomous ships has become an inevitable tendency for future ship development, and gained the interest of many researchers in both academia and private sectors (Goerlandt and Montewka, 2015). Furthermore, although the control technology of ships has gradually begun to change from traditional electromechanical control (Gupta et al., 2018) to the trend of networking, digitization, and automation, the ship-handling process has become a multi-functional integrated system integrating multiple automation systems, which improves the safety, profitability and management efficiency of shipping. However, the improvement of the degree of automation of ships has a certain gap from the ships with automatic perception, subjective analysis, and autonomous decision-making.

The accuracy of ship maneuvering decisions is directly related to the safety of waterway transportation. The seafarers onboard vessels, especially the Officer On Watch (OOW), often perform duties in circumstances where technological, environmental factors, etc., emerge which may lead to the occurrence of human failures and marine accidents (Ugurlu et al., 2015). Likewise, in the process of autonomous ships human-like decision-making, the OOW maneuvering decision-making is also influenced by multi-source information, for instance, the other ships in waterways and ports, the natural environmental factors, etc. (Kim et al., 2017), this requires ship maneuvering decision-making procedures expressed along with higher effectiveness. However, due to the limited capacity of the information acquiring and processing, OOW cannot achieve the multi-attribute or multi-source information in a particular time and space concurrently (Xue et al., 2019). For instance, under high-intensity work pressure, the OOW cannot always ensure to make correct decisions timely when facing constant changing factors in different navigation scenarios, thus maneuvering decisions cannot still be made accurately and quickly, which could lead to maritime traffic accidents. Therefore, the automatic acquisition and representation of maneuvering decision-making are necessary for ensuring accurate maneuvering decisions and maritime traffic safety; moreover, it is essential to identify, analyze, and prioritize the main maritime traffic safety influencing factors for efficient selection from the multi-attribute or multi-source information for corresponding maneuvering decisions of autonomous ships.

Multi-attribute decision-making is widely used in economics, society, military, and engineering technology (Liu et al., 2015). Due to the uncertainty and complexity of decision problems, the problems of multi-attribute decision-making are always combined with uncertain and fuzzy matters, so fuzziness is an essential factor to be considered in practical decision-making of real-world (Jin and Liu, 2010). In addition, when conducting the problems with poor information, the characteristics of grey (the data/information that is known partially) are also shown within the decision problems. Therefore, the decision-making problems in the real world are often fuzzy and grey, which are called the grey fuzzy multiple attribute decision-making problems (Liu et al., 2015).

Although a variety of previous studies in academia have been conducted upon impact factors assessment based on the grey and fuzzy theories, they seldom take into consideration the relative importance of different influencing factors (they only consider different influencing factors in the same weight) and in the absence of expertise. Some studies only consider the same weight to determine the judgments from different experts. Some studies use the standard fuzzy number functions to evaluate the linguistic terms given by experts. However, the standard fuzzy membership function sometimes cannot determine different linguistic terms from different domain experts reasonably. In some specific situations, it equally treats different indexes, i.e., the same linguistic term from different domain experts.

In this chapter, the autonomous ship human maneuvering decision factors are modeled as a typical "grey system", and fuzzy numbers of the domain experts are utilized to optimize the proposed model. The maritime traffic safety influencing factors of autonomous ship maneuvering decision-making, such as force parameters, draft, environment, motion, and position, etc., are obtained using data from a simulation platform. After collecting the judgment knowledge from domain experts, the Delphi method was utilized for comprehensively determining the fuzzy numbers of different linguistic terms combined with varying weights of each domain expert. Finally, the novel improved GRA and fuzzy theories based model is proposed for analyzing the final weights and rankings of the influencing factors. With computer assistance, the algorithm/model proposed in this chapter permits an automatic conversion from the comparative series of maritime traffic safety influencing factors and the corresponding maneuvering decisions (the combination of ship telegraph and rudder order) reference series to autonomous ship maneuvering influencing factors analysis system.

3.2 The proposed prioritizing model

This chapter utilizes the grey and fuzzy theories combined with quantitative and qualitative analysis, and comprehensively evaluates the maritime traffic safety influencing factors of autonomous ship maneuvering decisions. On the one hand, it can deal with the problems of imprecision and uncertainty. On the other hand, giving various weights of different experts leads to a more rational use of expert knowledge for judging the prioritization of the influencing factors. Furthermore, the evaluation results of the specific criteria of different experts on each linguistic term will be more accurate and reasonable by comprehensively utilizing the fuzzy numbers. The specific method is introduced below.

Step 1 – Data preprocessing

There are differences in the dimension and magnitude of each factor in the ship's maneuvering decision system, in order to facilitate data processing, the original data need to be standardized, the dimension or the order of magnitude needs to be eliminated, and the data series need to be transformed into a comparative series due to the inconsistent dimension of various factors.

Assume X is a grey relation factor set (discrete series), $X_0 = \{x_0(k) \mid k = 1, 2, \cdots, m\}$ as a reference series, representing the ship maneuvering decisions, which is the combination of ship Telegraph and Rudder Order (TRO) in the research (see Figure 3.5). $X_i = \{x_i(k) \mid k = 1, 2, \cdots, m\}\ (i = 1, 2, \cdots, n)$ as comparative series, representing the influencing factors, such as wind, current, and waves. *k* stands for the number of the corresponding element (the specific value for the TRO and influencing factors Y1-Y33) in each row of the determinant (see Equation 3.7). Thus, the correlation mechanisms of the reference series and comparative series can be utilized to recognize the influential mechanism of four types of different factors (ship motion, natural environment, force parameters, and draft & position, shown in Table 3.3) for autonomous ship maneuvering.

In the analysis and calculation process of the GRA, there are three methods for the non-dimensionalization of the original data, namely, equalization, initialization, and standardization.

Equalization First, the average value of each series is calculated separately, and then the original data in the corresponding series is divided by the average value, that is, the new data column obtained by the mean transformation.

$$X_0' = \left\{ n x_0(k) / \sum_{k=1}^{m} x_0(k) \mid k = 1, 2, \cdots, m \right\}, \tag{3.1}$$

$$X_i^{'} = \left\{ nx_i(k) / \sum_{k=1}^{m} x_i(k) \mid k = 1,2,\cdots,m \right\} (i = 1,2,3,\cdots,n). \tag{3.2}$$

Initialization The data of the same series is divided by the subsequent original data to obtain new multiple series, which is an initial valued series.

$$X_0^{'} = \{ x_0(k) / x_0(1) \mid k = 1,2,\cdots,m \}, \tag{3.3}$$

$$X_i^{'} = \{ x_i(k) / x_i(1) \mid k = 1,2,\cdots,m \} \; (i = 1,2,3,...,n). \tag{3.4}$$

Standardization Firstly, the average value and standard deviation of each trait are respectively determined, and then the original data is subtracted from the average value and then divided by the standard deviation so that the new data column obtained is the standardized series.

$$X_0^{'} = \left\{ \left[x_0(k) \text{-} \frac{1}{m} \sum_{k=1}^{m} x_0(k) \right] / S_0^{'} \mid k = 1,2,\cdots,m \right\}, \tag{3.5}$$

$$X_i^{'} = \left\{ \left[x_i(k) \text{-} \frac{1}{m} \sum_{k=1}^{m} x_i(k) \right] / S_i^{'} \mid k = 1,2,\cdots,m \right\} (i = 1,2,3,...,n), \tag{3.6}$$

where X'_0 is a non-dimensionalized reference series; X'_i is a dimensionless comparative series; $S_0^{'}$ and $S_i^{'}$ are the standard deviation of the reference series and the comparative series, respectively.

The original data series can be described by

$$X^{'} = \begin{pmatrix} X_0^{'} \\ X_1^{'} \\ X_2^{'} \\ \vdots \\ X_\omega^{'} \end{pmatrix} = \begin{bmatrix} x_{01}^{'} & x_{02}^{'} & \cdots & x_{0m}^{'} \\ x_{11}^{'} & x_{12}^{'} & \cdots & x_{1m}^{'} \\ x_{21}^{'} & x_{22}^{'} & \cdots & x_{2m}^{'} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\omega 1}^{'} & x_{\omega 2}^{'} & \cdots & x_{\omega m}^{'} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{TRO} \\ \text{Influence Factor 1} \\ \text{Influence Factor 2} \\ \vdots \\ \text{Influence Factor } \omega \end{bmatrix}, \tag{3.7}$$

where ω is the number of influencing factors.

Step 2 – Range analyzing

First, calculate $\Delta_i(k)$, that is, the absolute value of the difference between the reference series and each sub-series at each point:

$$\Delta_i(k) = \left| x_0(k) - x_i(k) \right|, \tag{3.8}$$

among them, $k = 1,2,\cdots,m$, $i = 1,2,\cdots,n$.

Then find the two-level maximum range and the two-level minimum range. First, calculate the first-level maximum range and the first-level minimum range:

$$\Delta_i(\max) = \max_k \Delta_i(k), \tag{3.9}$$

$$\Delta_i(\min) = \min_k \Delta_i(k). \tag{3.10}$$

Then calculate the second-level maximum range:

$$\Delta_{\max} = \max_i \max_k \Delta_i(k). \tag{3.11}$$

Similarly, the second-level minimum range is given by:

$$\Delta_{\min} = \min_i \min_k \Delta_i(k). \tag{3.12}$$

Step 3 – Relational coefficient calculating

The relational coefficient is used to measure the geometric difference between the comparative series and the reference series at each point. The relational coefficient of X_i to X_0 is:

$$\xi_i(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \rho \cdot \Delta_{\max}}{\Delta_i(k) + \rho \cdot \Delta_{\max}}, \tag{3.13}$$

where $\xi_i(x_0(k), x_i(k))$ represents the correlation coefficient between the comparative series X_i and the reference series X_0 at point k; 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 et al., 2014b); $k = 1, 2, \cdots, m$, $i = 1, 2, \cdots, n$.

Step 4 – Fuzzy membership functions of linguistic terms establishing

The traditional GRA does not fit with people's preference for a specific index. In order to overcome this shortcoming, this chapter considers the relative importance weights of the influencing factors, but it is difficult to be precisely determined. Moreover, in many situations, the information and experts' expertise are uncertain or vague. However, fuzzy sets provides a useful mathematical tool for directly working with the linguistic expression in reliability analyses (Lin and Wang, 1997; Page and Perry, 1994), and it is better to utilize fuzzy numbers in real-world decision-making to reflect human thoughts (Hatefi and Tamošaitienė, 2018). Therefore, we utilize fuzzy numbers of the domain experts to optimize our proposed model. The four domain experts are characterized as follows:

•**Expert No.1**: An experienced captain with more than 15 years of experience on the operation of board ships (classes of certificates: class A, ≥ 3000 gross tons, unlimited voyages).

•**Expert No.2**: A professor engaged in maritime research for more than ten years with particular reference to the ship operations.

•**Expert No.3**: A senior officer in charge of safety management of port operations of Yangtze River Three Gorges Navigation Authority.

•**Expert No.4**: A senior officer in charge of safety regulation of Shanghai Port from China Maritime Safety Administration.

The triangular fuzzy number, corresponding to linguistic terms, can be determined from domain expert knowledge based on the Delphi method (Ishikawa et al., 1993). Assuming that there are *n* experts, the *i*-th expert is assigned with the relative weight β_i (*i*= 1,. . . ,m), satisfying $\sum_{i=1}^{m}\beta_i = 1$ and β_i > 0 for *i* = 1,. . . ,m. And the fuzzy judgment linguistic term for the specific influencing factors is $x_i = (a_i, b_i, c_i)$, then according to the experts' judgment, the triangular fuzzy number $A = (a', b', c')$ corresponding to the fuzzy linguistic term of the variable can be summarized according to Equations (3.14) to (3.16).

$$a' = \sum_{i=1}^{n} \beta_i a_i , \tag{3.14}$$

$$b' = \sum_{i=1}^{n} \beta_i b_i , \tag{3.15}$$

$$c' = \sum_{i=1}^{n} \beta_i c_i . \tag{3.16}$$

This chapter defines the maritime traffic safety influencing factors of autonomous ship maneuvering using five linguistic terms, namely, Very Low (VL), Low (L), Medium (M), High (H), Very High (VH). Different from each linguistic term utilized in the same separation distance, for instance, the corresponding midpoint or the *b'* in triangular fuzzy number *A* of each linguistic term Very Low (VL), Low (L), Medium (M), High (H), Very High (VH) is 0, 0.25, 0.5, 0.75, 1, respectively (Wang et al., 2009; Wu et al., 2018). In this research, the triangular fuzzy number of different linguistic terms is determined by the domain expert knowledge, and the weight of each expert is taken into consideration, as shown in Table 3.1. Hence, the fuzzy membership function of each linguistic term can be represented more rationally because we take into account the different evaluation criteria of each expert for various linguistic terms comprehensively. Fuzzy membership degrees of quantitative indexes can be obtained from Figure 3.1. Experts are invited to define the triangular fuzzy number of each linguistic term based their judgment, then the triangular fuzzy numbers of different linguistic terms are calculated through Equations (3.14) to (3.16), and the results are shown in Table 3.1.

Table 3.1 Triangular fuzzy numbers of different linguistic terms.

Expert No.	**Weights(β_i)**	**Triangular fuzzy numbers of different linguistic terms**				
		Very Low (VL)	Low (L)	Medium (M)	High (H)	Very High (VH)
1	0.30	(0, 0, 0.25)	(0, 0.25, 0.50)	(0.25, 0.50, 0.75)	(0.50, 0.75, 1)	(0.75, 1, 1)
2	0.25	(0, 0, 0.20)	(0, 0.20, 0.40)	(0.20, 0.40, 0.60)	(0.40, 0.60,0.80)	(0.80, 1, 1)
3	0.20	(0, 0, 0.25)	(0.10, 0.30, 0.50)	(0.30, 0.50, 0.70)	(0.70, 0.90, 1)	(0.90, 1, 1)
4	0.25	(0, 0, 0.30)	(0.20, 0.40, 0.50)	(0.30, 0.50, 0.65)	(0.60, 0.70, 0.90)	(0.85, 1, 1)
Total	1	(0, 0, 0.25)	(0.07, 0.29, 0.48)	(0.26, 0.48, 0.68)	(0.54, 0.73, 0.93)	(0.82, 1, 1)

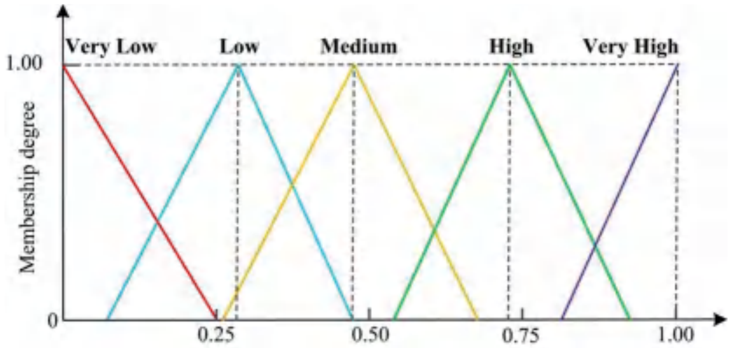


Figure 3.1 Triangular membership functions of different linguistic terms.

The specific process of utilizing fuzzy logic of this step is as follows:

(i) The maritime traffic safety influencing factors of autonomous ship maneuvering decisions are evaluated by the experts using the linguistic terms defined in Table 3.1;

(ii) The linguistic terms based on the judgments of domain expert are represented by the triangular fuzzy numbers, then the comprehensive evaluation fuzzy set of the weight of each influencing factor is established;

(iii) The relative weights β_i for each domain expert are taken into consideration. Specifically, the relative weights of experts are assigned based on their experience with the following relative weights: 0.30, 0.25, 0.20, and 0.25, respectively, then the optimized comprehensive evaluation fuzzy set is obtained;

(iv) The comprehensive evaluation weight of each influencing factor of autonomous ship maneuvering decisions is calculated.

Step 5 – Defuzzification

The linguistic terms from the judgments of domain experts need to be transformed into crisp values before further calculation. In other words, the fuzzy numbers should be converted into crisp numbers for priority ranking or comparison purpose, this process of transformation is called defuzzification. The defuzzification of fuzzy numbers is an important process, and it is the basis of applying the grey relational theory. Defuzzification can be conducted in many different ways, such as max criterion, center of gravity (COG), mean of maximum (MOM) methods, etc. (Akyuz et al., 2016; Balmat et al., 2011; Braae and Rutherford, 1978; Lee, 1990; Senol and Sahin, 2016).

The COG method, which also is known as center of area (COA), is the most extensively used technique developed by Sugeno (1999) as it is relatively accurate and takes the total output distribution into consideration (Patel and Mohan, 2002). Hence, the COG method can yield a better steady-state performance (Lee, 1990). This COG method can be used as a centroid defuzzification method to find the center of gravity point of the fuzzy set (Kumar et al., 2018).

The linguistic terms from the judgments of domain experts for maritime traffic safety influencing factors of autonomous ship maneuvering decisions can be defuzzified according to the fuzzy membership function; the crisp number can be calculated as follows:

$$A(X) = \frac{\int_X x\mu_A(x)dx}{\int_X \mu_A(x)dx}, \tag{3.17}$$

where $A(X)$ denotes the crisp value, x is the output variable, and $\mu_A(x)$ is the membership function for linguistic terms from the judgments of domain experts, as shown in Figure 3.1.

Specifically, the defuzzification of a triangular fuzzy number based the Equation (3.17) can be calculated as follows:

$$A(X) = \frac{\int_{a'}^{b'} x\frac{x-a'}{b'-a'}dx + \int_{b'}^{c'} x\frac{c'-x}{c'-b'}dx}{\int_{a'}^{b'} \frac{x-a'}{b'-a'}dx + \int_{b'}^{c'} \frac{c'-x}{c'-b'}dx} = \frac{1}{3}(a'+b'+c'). \tag{3.18}$$

Then, we can get a crisp number of different linguistic terms as shown in Table 3.2.

Table 3.2 The crisp number of different linguistic terms.

Name	**The triangular fuzzy number and crisp number of different linguistic terms**				
Linguistic term	Very Low (VL)	Low (L)	Medium (M)	High (H)	Very High (VH)
Fuzzy number	(0, 0, 0.25)	(0.07, 0.29, 0.48)	(0.26, 0.48, 0.68)	(0.54, 0.73, 0.93)	(0.82, 1, 1)
Crisp number	0.0833	0.2800	0.4733	0.7333	0.9400

Step 6 – Relational Grade Ranking

The traditional grey relational grade is calculated according to the Equation (3.19):

$$\gamma_i = \frac{1}{m}\sum_{k=1}^{m} \xi_i(x_0(k), x_i(k)), \tag{3.19}$$

where $k = 1, 2, \cdots, m$, $i = 1, 2, \cdots, n$.

Since the influence degree from each maritime traffic safety influencing factor of autonomous ship maneuvering decisions varies, assuming that the weight of each influencing factor is λ_k , then the relational grade between the reference series and comparative series can be obtained by the Equation (3.20):

$$\lambda_i(x_0(k), x_i(k)) = \frac{1}{m}\sum_{k=1}^{m} \lambda_k(\xi_i(x_0(k), x_i(k)), \tag{3.20}$$

where $\sum_{k=1}^{m} \lambda_k = 1$, λ_k can be determined by fuzzy sets based the domain expert knowledge.

When determining the relational grade, each sub-series of Y1-Y33 is compared to the reference series of TRO. Hence, the relationship between each sub-series and the reference series is sorted. Thereby, the main maritime traffic safety influencing factors of the autonomous ship maneuvering decisions in the specific navigational scenario are prioritized and identified.

The framework of our proposed model is shown graphically in Figure 3.3 that briefly illustrates the maritime traffic safety influencing factors of autonomous ship maneuvering decisions prioritizing procedure of the proposed GRA and fuzzy theories based methodology. The right-hand part of Figure 3.3 shows the steps of obtaining the weights for different influencing factors; the middle part presents the process of applying the traditional GRA theory, while the left-hand part provides the priority ranking and analyzing procedure of the maritime traffic safety influencing factors analysis system for autonomous ship maneuvering. And the logic framework for applying the proposed model is shown in Figure 3.2.

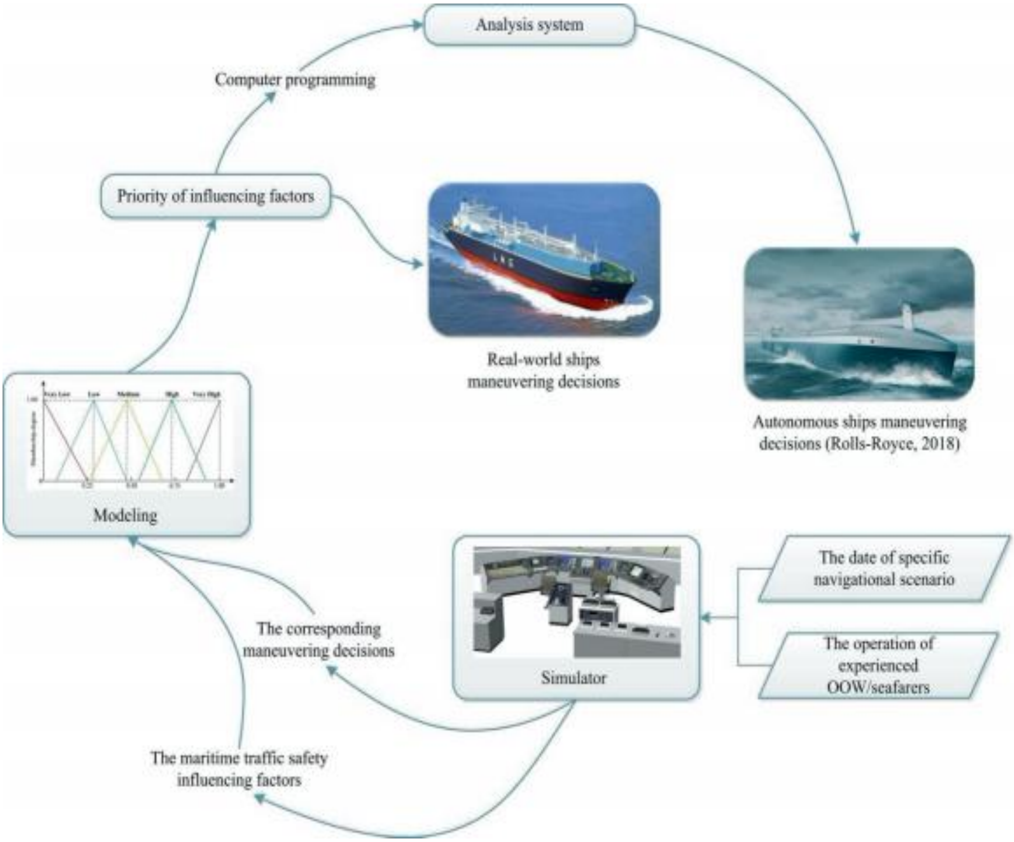


Figure 3.2 The logic framework for applying the proposed model.

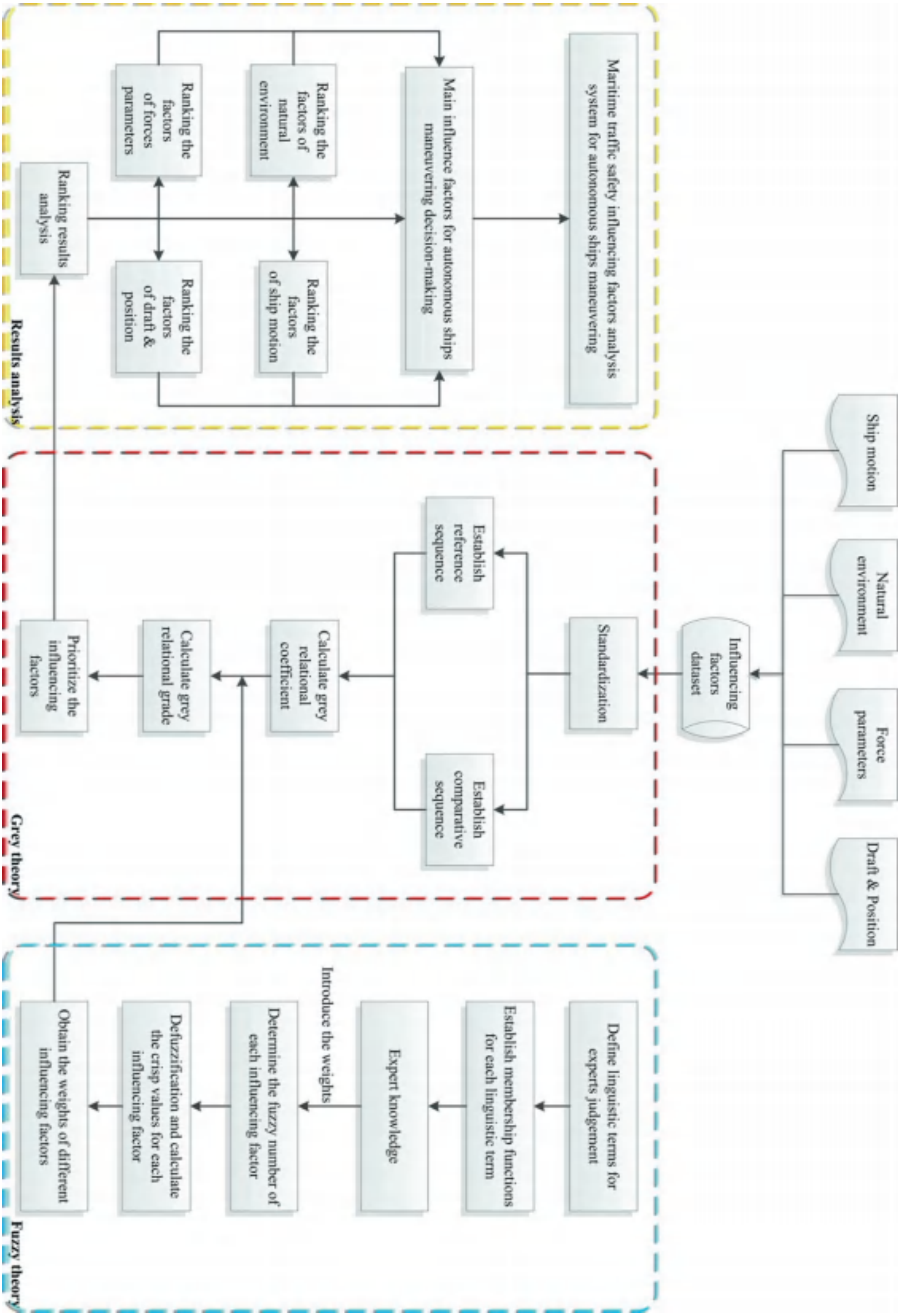


Figure 3.3 Framework of the proposed model using grey and fuzzy theory.

3.3 Experiments

3.3.1 Scenario design

Figure 3.4 The designed experimental scenario.

In our experiment, the Shanghai Waigaoqiao wharf was selected to be the simulator scenario when the ship was downstream of the berthing into the port. We use a 30,000-ton bulk carrier, i.e., Own Ship (OS) 1, as our experimental ship. 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 (Figure 3.4(b) shows the initial boundary) to the ship berths docked at the end of the cable (Figure 3.4(c) shows the end boundary) as a complete berthing process. The experimental scenario is shown in Figure 3.4. More details are presented in Chapter 4.

3.3.2 Data collection and processing

We collect the data from the full-task handling simulation platform (Navi-Trainer Professional 5000) from the Maneuvering Simulator Laboratory in Wuhan University of Technology Waterway Road Traffic Safety Control and Equipment Ministry of Education Engineering Research Center. The operational data from the exercises and assessment exams of unlimited navigational class seafarers are collected as our experimental data. In this experiment, there are 96 skilled maneuvering level captain/chief officer. More details are presented in Chapter 4. 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.

The multisource information of ship maneuvering traffic environment were collected. For instance, the location (longitude, latitude), environment (wind, current, etc.), control (rudder order, marine telegraph), ship movement (heading, roll rate, etc.), the ship's draft, tugs, mechanical contact force-related parameters, and other related parameters. The above factors, such as the environment, the control, location and the relevant parameters of the tug and other factors (see Table 3.3 and Table 3.4), were selected from the weakly related parameters. Table 3.4 lists some of the training samples.

Table 3.3 The category of influencing factors.

Influencing factors	Meaning	Units	Category
Y1	Current draft at ship bow	Meters	Draft
Y2	Current draft at ship stern	Meters	Draft
Y3	Under keel clearance aft	Meters	Draft
Y4	Under keel clearance fwd	Meters	Draft
Y5	Current direction	Degrees	Environment
Y6	Current speed	Knots	Environment
Y7	Relative current direction	Degrees	Environment
Y8	Relative wave direction	Degrees	Environment
Y9	Relative wind direction	Degrees	Environment
Y10	Relative wind speed	Knots	Environment
Y11	Water depth	Meters	Environment
Y12	Wave height	Meters	Environment
Y13	Lateral force	Tonne-force	Force Parameters
Y14	Longitudinal force	Tonne-force	Force Parameters
Y15	Summary force	Tonne-force	Force Parameters
Y16	Vertical force	Tonne-force	Force Parameters
Y17	Lateral force of mooring lines	Tonne-force	Force Parameters
Y18	Longitudinal force of mooring lines	Tonne-force	Force Parameters
Y19	Summary force of mooring lines	Tonne-force	Force Parameters
Y20	Vertical force of mooring lines	Tonne-force	Force Parameters
Y21	Heading	Degrees	Motion
Y22	Height above the water	Meters	Motion
Y23	Lateral speed	Knots	Motion
Y24	Longitudinal speed	Knots	Motion
Y25	Pitch angle	Degrees	Motion
Y26	Pitch rate	Degrees/min	Motion
Y27	Rate of turn	Degrees/min	Motion
Y28	Roll angle	Degrees	Motion
Y29	Roll rate	Degrees/min	Motion
Y30	Vertical speed	Knots	Motion

Y31	Yaw rate	Degrees/min	Motion
Y32	Latitude	Degrees	Position
Y33	Longitude	Degrees	Position

Table 3.4 Original data of the studied area (partially).

No.	**X**		**Y1 (Meters)**	**Y2 (Meters)**	**Y3 (Meters)**	**Y4 (Meters)**	**…**	**Y33 (Degrees)**
	Rudder order (Degrees)	**Telegraph order (%)**						
1	-1.0000	50.0000	10.1766	10.8138	4.2631	4.8818	**…**	121.6474
2	-1.0000	50.0000	10.1812	10.8184	4.2574	4.8783	**…**	121.6474
3	-1.0000	50.0000	10.1898	10.8270	4.2478	4.8706	**…**	121.6474
4	-1.0000	50.0000	10.2095	10.8468	4.2267	4.8523	**…**	121.6473
5	-1.0000	50.0000	10.2152	10.8526	4.2200	4.8474	**…**	121.6473
6	-1.0000	46.2955	10.1926	10.8300	4.2411	4.8714	**…**	121.6473
7	-1.0000	40.0000	10.1809	10.8183	4.2521	4.8837	**…**	121.6473
8	-1.0000	40.0000	10.1915	10.8290	4.2398	4.8748	**…**	121.6473
9	-1.0000	40.0000	10.2082	10.8457	4.2220	4.8591	**…**	121.6473
10	-1.0000	40.0000	10.2006	10.8381	4.2284	4.8678	**…**	121.6472
11	-3.3119	40.0000	10.1846	10.8221	4.2431	4.8849	**…**	121.6472
12	-11.2792	40.0000	10.1958	10.8334	4.2307	4.8747	**…**	121.6472
13	-11.9016	40.0000	10.2208	10.8584	4.2045	4.8507	**…**	121.6472
…	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**

According to the scenario shown in Figure 3.4, the principle of the rudder angle and the propeller speed are defined based on the data collected from the simulator and the navigation experience. Figure 3.5 and Table 3.5 show 64 possible maneuvering decisions based on various standardization principle of speed control (propeller state) and course control (rudder angle).

The OOW maneuvers the ship by operating different TROs to change ship's speed and direction. Figure 3.5 shows TROs of ship OS1 and the Table 3.5 shows the combining TROs; this control procedure is a multi-dynamic 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 TROs frequently in the inbound decision-making ship handing process; therefore, in this chapter, we do not consider "Stop/Standby/Finished with engines" and "Midships" regardless of the rudder angle and if the power output is 0. Table 3.5 shows the standardization principle for output maneuvering decision-making factor.

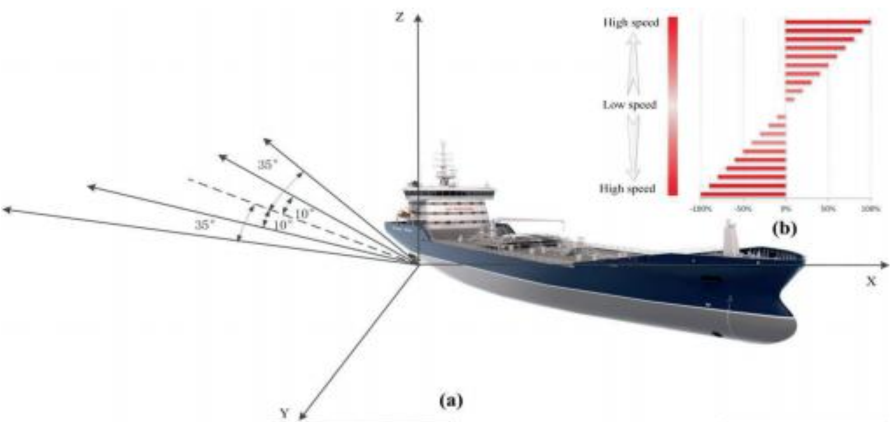


Figure 3.5 The telegraph and rudder orders of ship OS1.

Table 3.5 Ship maneuvering decision-making factors and standardization principle.

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
Maneuvering factors	**Decisions**		**Symbols**	**Decisions**		**Symbols**
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

3.4 Results

In our experiment, we select X and the related parameters Y1-Y33 to apply the proposed model, among them, X is the main factor and reference series, which consists of the 64 possible maneuvering decisions (the OOW's actual operation in the simulator, a different combination of TROs, see Table 3.5). Y1-Y33 are the influencing factors, and their values constitute the comparative series, such as the environment, ships, and other influencing factors. In addition, we collected a total of 20,534 samples as our data set.

3.4.1 Standardizing of the original data set

In this chapter, X presents the percentage of the number of each maneuvering decision of X1-X64 in a total number of the data set records. Limited to space, Table 3.6 lists only a part of multiple measured data. The data in Table 3.6 are standardized according to the principle of standardization of maneuvering decision-making influencing factors in Table 3.5 and the non-dimensionalization method of standardization (see Equations (3.5) and (3.6)).

Table 3.6 Dataset with the principle of standardization (partially).

No.	X			Y1	Y2	Y3	…	Y33
	Symbols	Proportion	Standardization					
1	X2	0.0300	-0.9848	0.5448	0.6840	-0.4284	…	0.7903
2	X2	0.0300	-0.9848	0.6719	0.7840	-0.4414	…	0.7782
3	X2	0.0300	-0.9848	0.9643	1.0135	-0.4704	…	0.7684
4	X2	0.0300	-0.9848	1.0498	1.0807	-0.4795	…	0.7555
5	X52	0.0196	-1.0784	0.7140	0.8186	-0.4506	…	0.7433
6	X52	0.0196	-1.0784	0.5404	0.6830	-0.4356	…	0.7320
7	X4	0.2955	1.4108	0.6975	0.8064	-0.4524	…	0.7214
8	X4	0.2955	1.4108	0.9452	1.0003	-0.4768	…	0.7100
9	X4	0.2955	1.4108	0.8325	0.9122	-0.4681	…	0.6986
10	X36	0.0098	-1.1667	0.5955	0.7270	-0.4479	…	0.6865
11	X35	0.0062	-1.1992	0.7622	0.8576	-0.4649	…	0.6744
12	X35	0.0062	-0.9848	0.5448	0.6840	-0.4284	…	0.7903
13	X35	0.0062	-0.9848	0.6719	0.7840	-0.4414	…	0.7782
…	…	…		…	…	…	…	…

3.4.2 Applying the proposed analysis model

According to the ranking criteria of the grey relational grade, the greater the grey relational grade of the comparative series, the greater the relevance of the comparative series to the reference series, the greater the degree of influence on the reference series, and the higher the ranking of the influencing factors. The GRA method is able to quantitatively describe the similarity and consistency degree between each comparative series and reference series and uses relational grades to complete the matching order of influencing factors. We use the original data matrix as defined by Equation (3.21).

$$X' = \begin{pmatrix} X_0' \\ X_1' \\ X_2' \\ \vdots \\ X_\omega' \end{pmatrix} = \begin{bmatrix} x_{01}' & x_{02}' & \cdots & x_{0m}' \\ x_{11}' & x_{12}' & \cdots & x_{1m}' \\ x_{21}' & x_{22}' & \cdots & x_{2m}' \\ \vdots & \vdots & \ddots & \vdots \\ x_{\omega 1}' & x_{\omega 2}' & \cdots & x_{\omega m}' \end{bmatrix} \Rightarrow \begin{bmatrix} \text{TRO} \\ \text{Influence Factor 1} \\ \text{Influence Factor 2} \\ \vdots \\ \text{Influence Factor } \omega \end{bmatrix} \Rightarrow \begin{bmatrix} \text{X} \\ \text{Y1} \\ \text{Y2} \\ \vdots \\ \text{Y33} \end{bmatrix}. \quad (3.21)$$

This way, we obtain the original data series. There is a case where the initial value is zero with respect to the influencing factors. Considering the value of the denominator should not be zero in a division operation; thus it is not suitable for the calculation based on Equation (3.4).

Moreover, the standardization method may genuinely reflect the relevance of the influencing factors to ship maneuvering decisions. Therefore, we use the standardization methods to explore the results of the interaction between ship maneuvering decisions and various influencing factors.

From Table A.1 and Equations (3.8) to (3.12), we can get the extreme values $\Delta_{max} = \mathbf{56.71438286}$, $\Delta_{min} = \mathbf{6.03501E-06}$, and we can calculate the grey relational coefficient based on the Equation (3.13), then calculating the traditional grey relational grade according to the Equation (3.19), the results are shown in Table A.2.

The convenient fuzzy numbers are defined for making pairwise comparisons shown in Table 3.1. Table A.3 shows the linguistic terms survey results from the four experts, and the crisp number and weights of different maneuvering influencing factors.

Then the defuzzification procedure is conducted based on Equation (3.18) and Table 3.2. The crisp number of different influencing factors are calculated with the relative weights β_i, then λ_k, the weights of different maneuvering influencing factors can be determined, the results are shown in Table A.3.

Finally, using Equations (3.19) and (3.20), and the results of grey relational coefficient from Table A.2, the priority ranking results of comparing grey algorithm with our proposed model are obtained and are shown in Table 3.7.

Table 3.7 Results of comparing grey method with our proposed model.

Influencing factors	**Grey method**				**Our proposed model**			
	Grey relational grade	**Rank No. 1**	**Category**	**Rank No. 2**	**Modeling grade**	**Rank No. 3**	**Category**	**Rank No. 4**
Y1	0.963331321	18	Draft	3	0.022296521	26	Draft	4
Y2	0.963022501	21	Draft	4	0.028357107	22	Draft	2
Y3	0.964702382	13	Draft	1	0.031169444	17	Draft	1
Y4	0.964360060	15	Draft	2	0.025634601	24	Draft	3
Y32	0.955548915	33	Position	6	0.016264792	30	Position	6
Y33	0.962805458	23	Position	5	0.018028349	28	Position	5
Y5	0.962321061	26	Environment	7	0.022824349	25	Environment	6
Y6	0.962607649	24	Environment	6	0.022279772	27	Environment	7
Y7	0.964744459	12	Environment	3	0.036003278	10	Environment	4
Y8	0.967877544	8	Environment	1	0.040086883	3	Environment	1
Y9	0.962919694	22	Environment	5	0.037689118	9	Environment	3
Y10	0.964861416	11	Environment	2	0.039961964	4	Environment	2
Y11	0.964247007	16	Environment	4	0.033350178	14	Environment	5
Y12	0.961966953	27	Environment	8	0.012658338	31	Environment	8
Y13	0.968696019	3	Forces	3	0.037915206	7	Forces	5
Y14	0.968659475	4	Forces	4	0.037913776	8	Forces	6
Y15	0.969245754	1	Forces	1	0.040143551	1	Forces	1
Y16	0.969236192	2	Forces	2	0.033081376	15	Forces	7
Y17	0.968609094	5	Forces	5	0.038352880	5	Forces	3
Y18	0.968266306	7	Forces	7	0.038339307	6	Forces	4
Y19	0.968451261	6	Forces	6	0.040110645	2	Forces	2
Y20	0.967668141	9	Forces	8	0.029048175	18	Forces	8
Y21	0.957594808	31	Motion	10	0.007249314	33	Motion	11

Y22	0.957995484	29	Motion	8	0.007667484	32	Motion	10
Y23	0.957976209	30	Motion	9	0.035314460	12	Motion	2
Y24	0.955638214	32	Motion	11	0.035228273	13	Motion	3
Y25	0.962322084	25	Motion	6	0.028887693	21	Motion	7
Y26	0.964491499	14	Motion	2	0.035554637	11	Motion	1
Y27	0.963209744	20	Motion	5	0.028914340	20	Motion	6
Y28	0.964126732	17	Motion	3	0.028941867	19	Motion	5
Y29	0.965110499	10	Motion	1	0.031182631	16	Motion	4
Y30	0.961761784	28	Motion	7	0.018008806	29	Motion	9
Y31	0.963209766	19	Motion	4	0.026155744	23	Motion	8

The rankings of ship maneuvering decision-making influencing factors are shown in Table 3.7, ranking result number 3. Furthermore, the result of grey method are sorted based on the ranking result number 1. Combining the results of Table 3.7 and Figure 3.6, the top ten influencing factors in Rank No. 1 and Rank No. 3 could be recognized, then the common seven influencing factors in the top ten most influential factors of both methods could be observed: Y15 (Summary force), Y19 (Summary force of mooring lines), Y8 (Relative wave direction), Y17 (Lateral force of mooring lines), Y18 (Longitudinal force of mooring lines), Y13 (Lateral force), Y14 (Longitudinal force), which should be given more attention when making decisions in ship maneuvering process. Furthermore, the result of top ten most influential factors sorted through our optimal model shows that: Y19 (Summary force of mooring lines) has risen four places to second place; Y8 (Relative wave direction) has risen five places to third place; Y10 (Relative wind speed) has risen seven places to fourth place; Y9 (Relative wind direction) has risen thirteen places to ninth place; Y7 (Relative current direction) has risen two places to tenth place. Y10, Y9, and Y7 became the new factors in the top ten of autonomous ship maneuvering decision process, which is corresponding to the judgment/operation of experienced seafarers in the real word shipping: when the seafarer (i.e., OOW) maneuvering the ship inbound the port, they need to pay more attention to the influencing factors of forces (e.g., forces of mooring lines and tugs), relative wave direction, relative wind direction, relative current direction, relative wind speed etc., so as to ensure the safety of ship and cargo. Therefore, the results indicate that our proposed model can identify the influencing factors of autonomous ship maneuvering decisions under real word maritime traffic safety context, and the priority ranking results are more reasonable than the original GRA method.

To compare the results from the proposed method and the GRA method more intuitively and clearly, we settle different coordinate systems in the same specific figure to compare the trend of different graphics. The x-axis denotes the number of influencing factors, and the y-axis represents the grey relational grade get from grey method or the modeling grade get from our proposed method. The ranking results of comparing grey algorithm with our proposed model are visualized in Figure 3.6. Meanwhile, the priority ranking analysis for four types of influencing factors is shown in Figure 3.7.

As can be seen from Figure 3.6, the changing tendency of the curves for the GRA method and our proposed model are the same basically, however the fluctuation trend of the curve of our proposed model is more evident than the GRA method (the dispersion of the fluctuation for GRA is 0.0137; the dispersion of the fluctuation for the proposed model is 0.0329), which means that the sensitivity of the prediction result of each influencing factor of our proposed model is higher than GRA method. Meanwhile, the curve of the original GRA method is relatively flat (especially for the influencing factors with respect to force parameters of Y13-Y20), which also proves the drawbacks of the traditional GRA method: it treats different indexes (influencing factors) equally and takes no account of the relative importance of them. Moreover, it does not fit with people's preferences for a specific index.

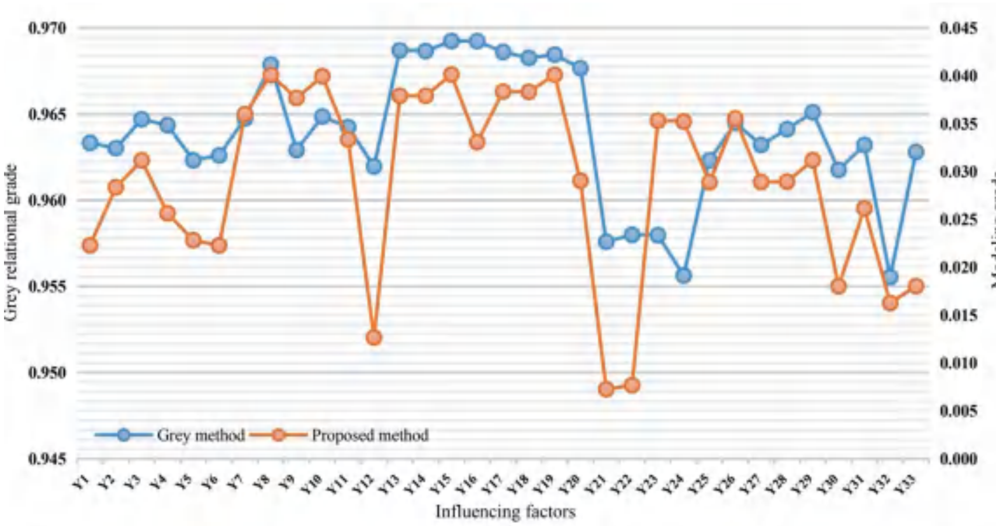


Figure 3.6 The results of comparing the grey method with our proposed model.

In addition, as shown in Figure 3.6, the comparing results of the histogram heights of the maritime traffic safety influencing factors Y9 (Relative wind direction), Y10 (Relative wind speed), Y23 (Lateral speed), and Y24 (Longitudinal speed) of our proposed method are obviously higher than the numbers in the GRA method, which indicates that OOW needs to take more attention about these factors when maneuvering the ship. In other words, when we design the program for the analysis system of the autonomous ship maneuvering decisions in the specific scenarios, we should assign a larger weight for these influencing factors than the original weight obtained from the grey method. Similarly, we should assign a smaller weight for the influencing factors Y12 (Wave height), Y21 (Heading), Y22 (Height above the water), Y30 (Vertical speed), and Y33 (Longitude) considering their histogram heights are obviously lower than the numbers in the GRA method.

It should be noted that, for the influencing factors of the same property, we may get different grey relational grades in different maritime traffic scenarios. For instance, in the specific experimental navigation scenario of Shanghai Waigaoqiao wharf, the ship's position of longitude did not change basically, and it's just a change in the position of latitude when it was berthing into the port, so the grey method gives us the different grey relational grades for the same property of longitude and latitude. However, when it is extended to the real general word maritime traffic scenarios or other domains, in common sense, the change of longitude and latitude always coincide. Thus the results are consistent with the proposed model. Therefore, the results displayed in Figure 3.6 are reasonable and meaningful, and the traditional GRA approach can sort the maneuvering influencing factors efficiently so that the OOW can get the main maritime traffic safety influencing factors intuitively through the correction and optimization of expert judgment knowledge and fuzzy theory. Then through the proposed model, the influencing factors which affect the ship maneuvering decisions are obtained, the proposed model could be applied in a more general and widespread maritime situation.

As shown in Figure 3.7, the diagrams of four categories of influencing factors are drawn independently (the histogram depicts the variation tendency of the proposed method and the scatter diagram in the form of a smooth curve represents the variation tendency of the GRA method). Overall, the changing tendency of each diagram for the GRA method and our proposed model are the same basically, but there are some details/differences which need to be described and explained.

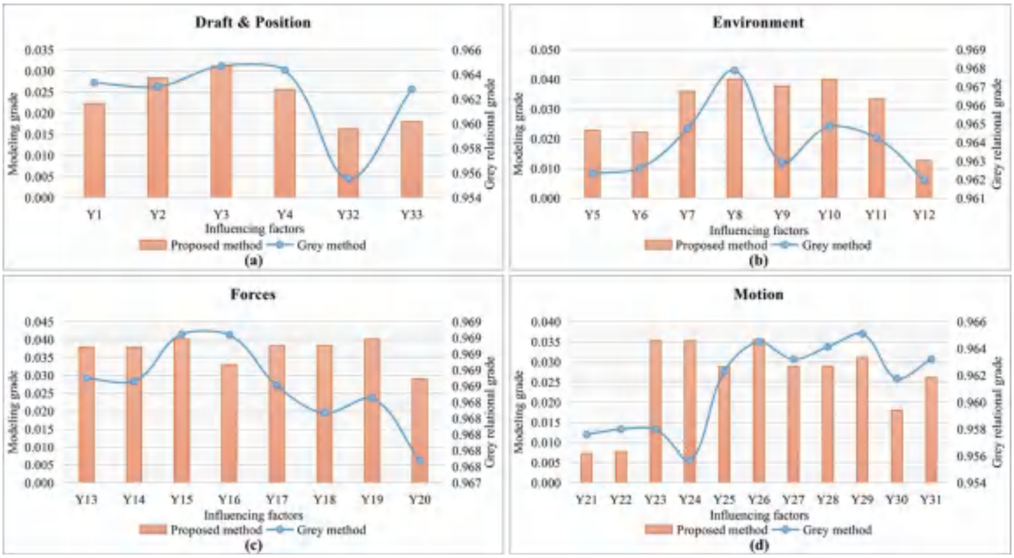


Figure 3.7 The ranking results analysis for four types of influencing factors.

Draft & Position: It can be seen from Figure 3.7(a), compared with the diagram of the grey method and the proposed method, the most influential factor within draft and position aspects is Y3 (Under keel clearance aft), it indicates that the OOW needs to take more attention about the under-keel clearance aft within the influencing factors of draft and position. Meanwhile, when we design the program for the analysis system of the autonomous ship maneuvering decisions in the specific scenarios considering maritime traffic safety, we should assign a larger weight for the keel clearance aft. Similarly, when it comes to the influencing factors longitude and latitude, the specific weight of Y32 (Latitude) has been increased, and the weight of Y33 (Longitude) has been reduced. As the above analysis, in the proposed method, the weight of latitude is higher, and the weight of longitude is lower than the original weight obtained via the grey method, that indicates the proposed model has a property of general flexibility for the analysis of the maritime traffic safety influencing factors for the ship maneuvering decisions.

Natural environment: As shown in Figure 3.7(b), Y8 (Relative wave direction) and Y10 (Relative wind speed) are the top two most influential factors in both the grey method and the proposed method, which indicates the OOW needs to focus on the relative wave direction and relative wind speed when it comes to the natural environment. In addition, the Y9 (Relative wind direction), Y10 (Relative wind speed), and Y11 (Water depth) have been increased in the results of proposed method. Among them, the increase of Y9 is greatest, which indicates that, in the scope of natural environment, according to the judgments of domain experts based on the fuzzy theory, the OOW should pay more attention to the relative wind direction when maneuvering the ship. Furthermore, it is similar to the program design for the analysis system, the heavyweight of relative wave direction and relative wind speed needs to be given. Moreover, the weight of influencing factor of relative wind direction needs to be increased.

Force parameters: According to Figure 3.7(c) and Figure 3.6, the ranking and grade of force parameters maintain a relatively stable trend in various influencing factors, meanwhile, all the force parameters keep a high ranking and grade in both two methods (all remain in the top 18, seen from Table 3.7). It indicates that all the force parameters play a crucial role in autonomous ship maneuvering decision-making in the specific scenario. Besides, it is also corresponding to the operation of experienced seafarers in the real world shipping, the force parameters are the crucial and direct influencing factors for the maneuvering of ships and maritime traffic safety. Furthermore, we can see that the most influential factor of force parameters is Y15 (Summary force); Y17 (Lateral force of mooring lines), Y18 (Lateral force of mooring lines), and Y19 (Lateral force of mooring lines) has been increased and occupy a heavyweight, and Y16 (Vertical force) has been decreased. Similarly, it is reasonable for the real word shipping, especially for the inbound scenario. For instance, when a ship inbound a port, the pilots always call the tugs for assistance, the tugs push (there is no vertical force in this procedure) or pull through the mooring lines then assist the ship in getting into the port, this has a great influence on the maneuvering of ships. For another example, when the ship is close to the berth, the ship usually uses the mooring winch to assist the berthing, so the forces from mooring lines are the main influencing factors for ship maneuvering and maritime traffic safety. Therefore, when the program design for the analysis system of the influencing factors of autonomous ship maneuvering decisions in the specific scenario, the force parameters should take into consideration and attach the heavyweights.

Ship motion: It is observed from Figure 3.7(d) that the most influential factor of ship motion is Y26 (Pitch rate); Y23 (Lateral speed) and Y24 (Longitudinal speed) have been increased, and Y30 (Vertical speed) has been decreased. In addition, the changing tendency of each influencing factor for the GRA method and our proposed model are the same basically, except Y 23 and Y24. The changes are reasonable and meaningful in the real-world shipping and traffic safety domain. When the ship is berthing to the port, the OOW/operator needs to pay attention to the lateral and longitudinal speed at all times, thus ensuring the safety of ships and cargo. For instance, if the ship has an obvious lateral speed, it would do damage for the berth and port; if the ship has a greater longitudinal speed, it will cause a collision with the ships before and after the berth. However, the vertical speed is usually not considered as the significant influencing factor of maritime safety when a ship is berthing into the port. Hence, when the OOW is maneuvering the ship, the lateral and longitudinal speed, as well as pitch rate, should be given more attention, as the same to the program design for the analysis system of the autonomous ship maneuvering decisions for evaluating maritime traffic safety influencing factors.

3.5 Discussion

Ship maneuvering decision-making is influenced by multi-source information, such as the information from the aspects of people, ships, environment, and it has an interaction with various influencing factors, and each factor plays a different role in the ship maneuvering decision-making process. At the same time, some factors interact with each other (e.g., when Y21 (Heading) of the ship changed, then Y8 (relative wave direction) changed correspondingly; when the position changed, i.e., Y32 (Latitude) and Y33 (Longitude) changed, then Y11 (Water depth) changed correspondingly) to form a grey system with clear and partially unclear information, thus constitute a typical "grey system". In this chapter, the maritime traffic safety influencing factors of autonomous ship maneuvering decision-making are identified and classified into four aspects: "Draft & Position", "Natural environment", "Force parameters", "Ship motion". Then the proposed grey and fuzzy algorithms are applied to prioritize these influencing factors using the linguistic terms of the judgments of domain experts; among these procedures, the relative importance of the linguistic terms of experts' judgments is also taken into consideration.

The results from the grey relational analysis showed that the values of grey relational grade for different influencing factors are relatively large (the minimum value is over 0.95), moreover, the values of grey relational grade between the reference series TRO and comparative series of different influencing factors are different, which indicates that the ship maneuvering decision-making is affected by different influencing factors and each influencing factor plays a specific role.

Furthermore, grey relational analysis combined with the fuzzy theory is a simple and practical method. The model elaborated in this innovative study is utilized to prioritize the influencing factors of autonomous ship maneuvering decision-making. The top ten most influential factors in the proposed method are Y15 (Summary force), Y19 (Summary force of mooring lines), Y8 (Relative wave direction), Y10 (Relative wind speed), Y17 (Lateral force of mooring lines), Y18 (Longitudinal force of mooring lines), Y13 (Lateral force), Y14 (Longitudinal force), Y9 (Relative wind direction), and Y7 (Relative current direction). In addition, among the four categories of influencing factors, the most influential factor within each aspect are Y3 (Under keel clearance aft), Y8 (Relative wave direction), Y15 (Summary force), and Y26 (Pitch rate), respectively. The results are corresponding to the judgment/operation of experienced seafarers in the real world shipping. Likewise, they are reasonable and meaningful in the specific navigational scenarios under maritime traffic safety domain.

Therefore, in the process of ship maneuvering decision-making, as well as the program design for the analysis system of the influencing factors of autonomous ship maneuvering decision-making in specific scenarios, the above ten factors should be taken as the main influencing factors. At the same time, the most influential factor in each category also needs to be paid particular attention, especially when the OOW/operators considering the impact of a certain type of influencing factors on ship maneuvering decision-making or the engineers design the maneuvering decisions programs for autonomous ships in specific maritime traffic scenarios. Furthermore, the degree of influence of various factors and the actual economic cost of ships operation should be further considered, thus to promote the development of autonomous merchant shipping, reduce transportation costs and improve transportation efficiency and maritime traffic safety.

Though the proposed grey and fuzzy model is a promising model, this study still has some shortcomings as follows, which should be solved in future research. In the specific experimental navigation scenario, as the above description and analysis for Figure 3.6 and Figure 3.7 in Section 3.4, our proposed model is rational and widely applicable to the analysis of the maritime traffic safety influencing factors for the ship maneuvering decisions. However, when in a specific navigational scenario, for instance, the influencing factors of longitude and latitude do not change correspondingly, there still has some shortcomings when adding the general expert knowledge using general common sense; in this case, the accuracy of our proposed model for analyzing these influencing factors is affected. Therefore, although the traditional grey theory has been largely criticized for the reason that it treats different indexes (influencing factors) equally and takes no account of the relative importance of them, and does not fit with people's preferences for a specific index, it still has the accuracy and sensitivity in specific experimental scenario for particular factors, so it is better to combine with the results from traditional grey method when we apply the proposed model. Hence, further research is needed to find out more influencing factors and navigational scenarios that can conduct a more comprehensive analysis of traffic safety influencing factors which affecting autonomous ship maneuvering decision-making.

3.6 Conclusions

With the development of modern science and technology, the improvement of autonomous ships has been technically feasible. However, autonomous ship maneuvering decisions are influenced by several influencing factors. The main purpose of our study is to select/prioritize the main influencing factors from all the decision-making influencing factors, thereby establishing the decision-making model efficiently for our subsequent autonomous ships human-like decision-making algorithm studies.

In this chapter, the standardization principle of ship maneuvering is introduced, and an innovative grey and fuzzy theories based inference model combined with the expert linguistic terms with different weights is proposed. This model can recognize the main decision-making factors of ship maneuvering from multi-source influencing factors, so as to study the decision-making prioritization for maritime traffic safety in specific ship maneuvering scenario accurately and efficiently, and it also can provide the theoretical basis for the decision-making of OOW and improve the maritime traffic safety as well as the program design for the analysis system of the influencing factors of autonomous ship maneuvering decisions in specific scenarios.

In this chapter, the overall influencing factors and four categories of influencing factors are analyzed and prioritized separately. The result provides guidance for the OOW's attention to different navigational information for ship maneuvering decision-making under specific maritime traffic scenarios. It not only emphasizes the main influencing factors in the overall attributes but also pays attention to the maritime traffic safety influencing factors and their dynamic change features in each category. The results of the proposed model are more related to real word shipping scenarios and are found to be satisfactory.

Furthermore, the fuzzy number functions are utilized to apply expert knowledge to the process of the main influencing factors selecting/prioritizing of autonomous ship maneuvering decisions, which realizes the identification of the main influencing factors. Moreover, through using the fuzzy theory with expert knowledge, the order of the ranking results of various influencing factors obtained from the traditional grey relational analysis is changed. The results show that the proposed model improves the ranking results of the influencing factors, it is more rational and applicable. Likewise, it provides guidance for autonomous ship maneuvering decisions. In addition, with computer assistance, the model proposed in this chapter permits an automatic conversion from the comparative series of maritime traffic safety influencing factors and the corresponding maneuvering decisions (the combination of ship telegraph and rudder order) reference series to autonomous ship maneuvering influencing factors analysis system. The proposed algorithm solves the computational problem of complex fuzzy systems under big data by computer programming (computing advantage), which is of great significance to the development of autonomous ship maneuvering decisions analysis system.

Overall, this chapter proposes a prioritizing model for the influencing factors of autonomous ship maneuvering decision-making using grey and fuzzy theories. Based on the actual operation data of the experienced seafarers collected from the simulator, a reference series is established by using the combination of ship telegraph and rudder orders which directly corresponding to the control of a ship. Additionally, we establish the comparative series for various influencing factors concerning motion and environment, which affect ship maneuvering decision-making. Moreover, combined with the expert knowledge, the proposed model is further optimized to ensure the rationality, accuracy, and generalizability of it, to select/prioritize the main maritime traffic safety influencing factors of the autonomous ship maneuvering decisions in the specific navigational scenario.

The results of this research provide theoretical and practical insights for prioritizing/evaluating the influencing factors in the autonomous ship maneuvering and maritime safety management for the shipping industry. The model can be further applied to the more general widespread way of the analysis system for autonomous ships human-like decision-making in specific scenarios. In further research, we will explore more about the optimization method for the selection/prioritization of influencing factors and use different data sets to compare the research findings. Moreover, we need to illustrate and combine the expert knowledge for various specific navigational scenarios when we apply our proposed model.

Chapter 4 Modeling for recognizing human-like decisions of autonomous ships

Chapter 3 introduces a prioritizing model using grey and fuzzy theories to recognize the main safety decision-making factors of ship maneuvering from multi-source influencing factors to study the decision-making prioritization for maritime traffic safety in a specific ship maneuvering scenario accurately and efficiently. In this chapter, a novel algorithm for modeling human decision-making of inbound merchant ships is proposed. This method can be used to realize the automatic acquisition and representation of the seafarer's decision-making knowledge in inbound merchant ships analysis. To verify the performance of the model, a case study based on this method is conducted in the Waigaoqiao Phase IV Port of Shanghai. The experimental results indicate that the maneuvering decision recognition model combined with the method of classification interval division, which is proposed in this chapter, can accurately and scientifically standardize the boundary of the interval of influencing factor data and identify current maneuvering behavior. The proposed methods and the evaluation results provide useful insights for effective safety management of the inbound merchant ships. In this chapter, Section 4.1 introduces the background of this chapter; the proposed model is developed in Section 4.2; then the experiments are conducted in Section 4.3; Section 4.4 illustrates the results of this chapter and ends with conclusions in Section 4.5.

Parts of this chapter have been published in the following papers:

Xue, J., Wu, C. Z., & van Gelder, P. (2019). A Novel Algorithm for Modeling Human Decision Making of Inbound Merchant Ships–A Case Study of the Shanghai Waigaoqiao Phase IV Port. In *Advances in Marine Navigation and Safety of Sea Transportation* (pp. 51-56). CRC Press Taylor & Francis Group.

Xue, J., Wu, C., Chen, Z., van Gelder, P. H. A. J. M., & Yan, X. (2019). Modeling human-like decision-making for inbound smart ships based on fuzzy decision trees. *Expert Systems with Applications*, *115*, 172-188.

4.1 Introduction

Currently, waterway transportation plays an increasingly important role in cargo transportation. It accounts for 95% of total crude oil transportation and 99% of total iron ore transportation. 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 addition, with the development of technologies, such as computers science, information and communications technologies (ICT), artificial intelligence (AI), internet of things (IoT), and information physics systems, have greatly advanced the process of ship intelligence and made unmanned autonomous ships a possibility. Autonomous 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 autonomous merchant ship for the maneuvering decision-making characteristics of crew.

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 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). Moreover, the intensity of seafarers' duty is very large, the OOW is on duty for eight hours a day in three shifts (chief officer: 04:00-08:00, 16:00-20:00; second officer: 00:00-04:00, 12:00-16:00; third officer: 08:00-12:00, 20: 00-00:00). However, people's energy is limited, especially in the ocean navigation environment, which may lead to visual fatigue, distraction, and other situations. At the same time, under high-intensity work pressure, the OOW cannot always ensure to make correct decisions timely when facing changing factors in the navigation environment. Besides, due to the water resistance and the limitation of the huge ship type, the ship has slower speed and poor flexibility in the water, and its slow pace may easily cause the OOW to relax his vigilance, leading to poor timeliness and promptness of the ship's maneuvering decision-making, and mistakes in the decision-making of OOW, coupled with the inertia of large vessels, are often prone to accidents and cause irreversible losses. According to statistics (Hanzu-Pazara et al., 2008), in ship collision accidents, 89% to 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 seafarer in extreme weather conditions in recent years has also become a problem that cannot be ignored (Wang et al., 2014a). Besides, the number of seafarers is declining recently, while the wages of seafarer are rising year by year, which has become the second largest expenditure item after the fuel costs of shipping (Lun et al., 2016). As autonomous ships have outstanding advantages in improving the safety management, energy consumption management, and operational efficiency of ships, therefore, the researches for autonomous 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).

Classification is a data mining (DM) technique used to predict or forecast the unknown information using the historical data. Many classification algorithms have been developed such as decision tree (Cohen et al., 2007), classification and regression tree (Friedman et al., 1984), Bayesian classification (Heckerman, 1998), neural networks (Rojas, 1996) and K-nearest neighbor classification (Denœux, 1995), etc. Among them, the decision tree has become more popular algorithm as it has several advantages over others. Common decision tree algorithms are Iterative Dichotomiser 3 (ID3), C4.5, C5.0, Classification And Regression Trees (CART), Chi-squared Automatic Interaction Detector (CHAID), etc. In these algorithms, the ID3 algorithm is the influential and wide used 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. The C4.5 algorithm is complex when continuously processing data, and its workload is large. C5.0 mainly adds support for Boosting, which uses less memory and 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 or CHAID only supports building binary trees, while ID3 or C4.5 allows two or more outcomes and supports binary or multi-fork trees (Wu et al., 2007). Therefore, this chapter utilizes the ID3 algorithm to learn the seafarer's maneuvering decision characteristics considering the advantages of it, thus to construct a human-like decision-making model under multiple constraints in a specific scenario.

In summary, this chapter focuses on the concept of human-like maneuvering for autonomous merchant ships and studies the human-like decision-making mechanism for autonomous merchant ships. We proposed an autonomous ship human-like decision-making recognition model. By establishing the autonomous learning method of maneuvering decision-making, the maneuvering decision-making rules of the typical maneuvering style in the specific scenario are explored, and the processes of autonomous learning OOW's maneuvering decision-making characteristics for autonomous ships are studied. This chapter provides a new perspective and methodology for the development of autonomous ship technology in theory and practice and promotes the application and spreading of autonomous merchant ships.

4.2 The proposed recognizing model

The machine learning technique for inducing a decision tree from data is called decision tree learning, or decision trees. In decision theory and decision analysis, a decision tree is a graph or model of decisions and their possible consequences. It is a method to solve complex decision problems through tree-like logical thinking and can be used to create a plan to reach a goal. Decision trees are constructed in order to help with making decisions. A decision tree is a special form of tree structure and a descriptive means for calculating conditional probabilities.

Decision tree learning is a common method used in data mining. Each internal node corresponds to a variable. A leaf node represents a possible value of target variable given the values of the variables represented by the path from the root node. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner. The recursion is completed when splitting is either non-feasible, or a singular classification can be applied to each element of the derived subset. Figure 4.1 shows the flowchart for tree-based classification.

Decision tree-based classification is also one of the most widely used classification methods in the field of data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. Each internal node corresponds to one of the input variables. Each leaf node represents a value of the target variable given the values of the input variables represented by the path from the root node to the leaf node. Some of the key advantages of using decision trees are the ease of use and overall efficiency. A decision tree can 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, and the rule set can be derived that are easy to interpret.

The ID3 algorithm is a typical decision tree learning algorithm. It uses the information gain as the attribute selection criterion to determine the appropriate attributes to be used when generating each node. In this way, the attribute with the maximum information gain can be selected as the test attribute of the current node, so that the information required for classification using the training sample subset obtained by the attribute is minimized.

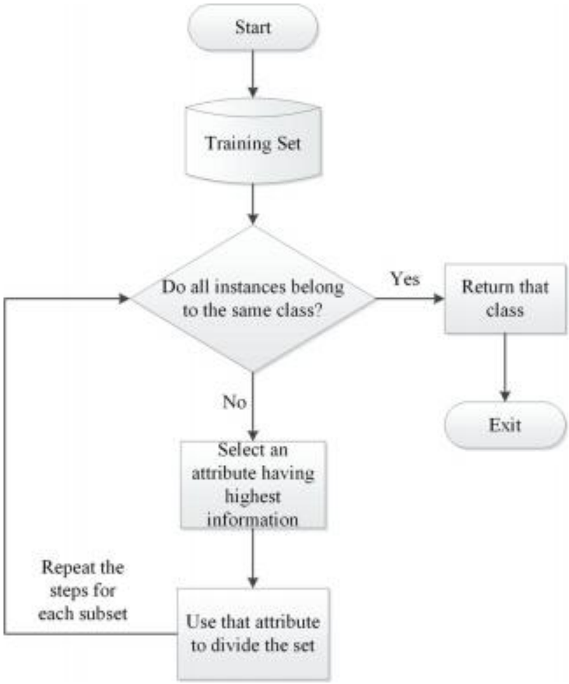


Figure 4.1 The flowchart for tree-based classification.

Information theory

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

$$I(S_i) = -p(S_i)\log_2 p(S_i)\,, \tag{4.1}$$

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:

$$I(S_1, S_2, ..., S_v) = -\sum_{i=1}^{v} p(S_i)\log_2 p(S_i)\,, \tag{4.2}$$

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

Information entropy

Assume that D is the autonomous ship human-like decision-making training data set containing 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 prior probability of the event as follows:

$$p_i = |S_i| / |D| . \tag{4.3}$$

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

$$I(|S_1|, |S_2|, ..., |S_m|) = -\sum_{i=1}^{m} p_i \log_2(p_i) . \tag{4.4}$$

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

$$E(A) = \sum_{j=1}^{n} \frac{|D_j|}{|D|} I(|S_j^1|, |S_j^2|, ..., |S_j^i|) . \tag{4.5}$$

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

$$I(|S_j^1|, |S_j^2|, ..., |S_j^i|) = -\sum_{i=1}^{m} p_{ij} \log_2(p_{ij}) . \tag{4.6}$$

Information gain

The information gain of feature/attribute A is expressed as follows:

$$Gain(A) = I(|S_1|, |S_2|, ..., |S_m|) - E(A) . \tag{4.7}$$

A good rule of thumb would seem to be to choose that attribute to branch on which gains the most information. ID3 examines all candidate attributes and chooses A to maximize $Gain(A)$, then forms the tree and then uses the same process recursively to build decision trees for the residual subsets.

4.3 Experiments

4.3.1 Scenario identification

The data for this thesis 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 MSA, the British

Lloyd's Register, the British Maritime and Coast Guard (MCG), the Russian Ministry of Shipping and other authoritative certifications. We collected the operational data of the exercises and assessment exams as our experimental data (unlimited navigational class seafarer, captain/chief officer). The simulator scene 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 (Figure 4.2(f) shows the initial boundary) to the ship berths docked at the end of the cable (Figure 4.2(g) shows the end boundary) as a complete berthing process. The experimental scene is shown in Figure 4.2.

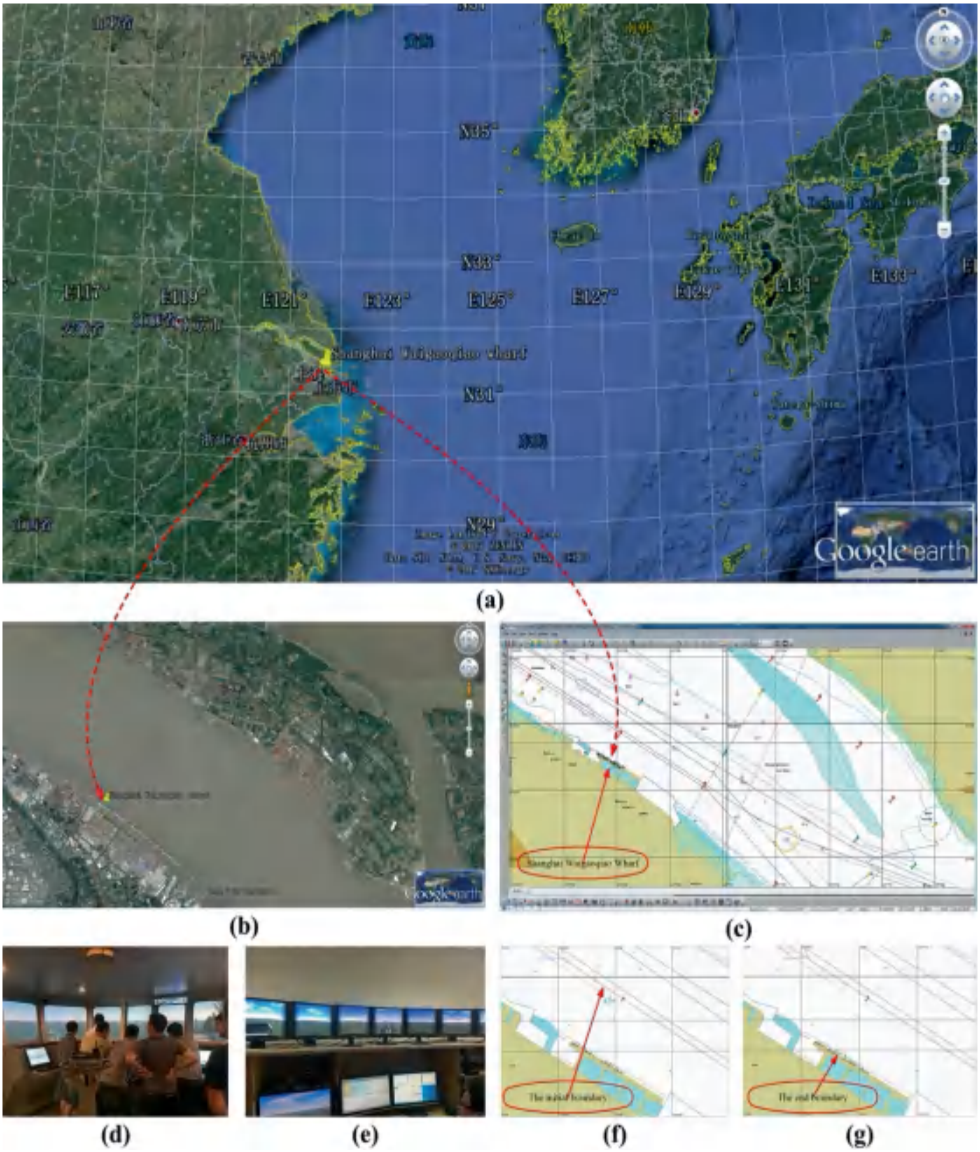
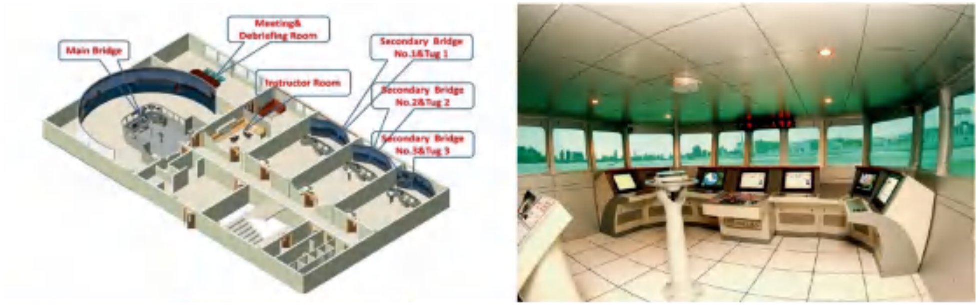


Figure 4.2 The experimental scene.

4.3.2 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 maneuvering behavior and environment, including inside and outside the multi-source information.



(a) **(b)**

Figure 4.3 The Navi-Trainer Professional 5000 system.

Figure 4.3(a) shows the spatial arrangement for The Navi-Trainer Professional 5000 system, it is mainly composed by the Main Bridge, Instructor room, and three Secondary Bridge systems. Figure 4.3(b) is the panorama of the Main Bridge. The experimental scheme of this chapter is shown in Table 4.1.

Table 4.1 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 ships.
Seafarer	Unlimited navigational class A chief officer or captain, 4 groups of 96 people, 32-45 years old, skilled maneuvering level.
Ship	30,000 tons bulk carrier (experimental simulation ship OS1, see Figure 4.4(a)). 33089.0t, 182.9 meters long, 22.6 meters 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 Figure 4.2(d), (e) and Figure 4.3(a), (b).

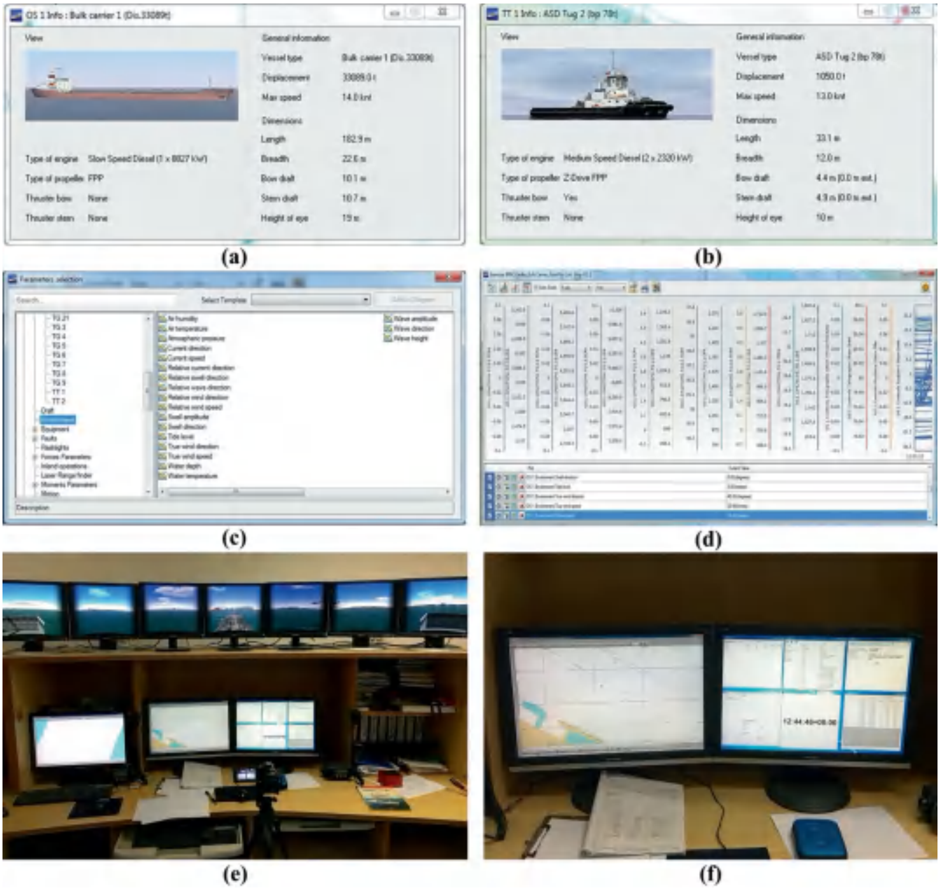


Figure 4.4 Data collection and processing.

From Table 4.2, we can get the average age of the seafarer participating in this experiment is 38.76 years old, and their average maneuvering experience is 8.89 years. From Figure 4.5, we can get the distribution of seafarers' age and their maneuvering experience. The captains' average age and maneuvering age are both higher than those of chief officers'. Note that for the definition of "ideal OOW", there is no precedent in the industry for the study of autonomous maneuvering decision characteristics by using the maneuvering data of experienced seafarers. Thus, there is no reference standard for the definition of "ideal OOW". However, in this chapter, in the context of security as a prerequisite, we consider that ship maneuvering experience (most directly reflected in driving age) is the core reference standard of an "ideal OOW", and it provides guarantees for high-quality experimental data. In addition, the chief officers or captains have a high social status and sense of responsibility. Most of them are trained professionals so both experimental and professional qualities of experimental personnel are guaranteed. Additionally, it also should be noted that, in this thesis, 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. Moreover, for ships, there are six different motion components are utilized to determine the orientation and position in six degrees of freedom (DOFs). In our case, there is a fixed dynamic ship model in the full-task handling simulation platform (Navi-Trainer Professional 5000) for the target ship OS 1. It is the same with the actual situation in the natural world port with changing environment surroundings when the OOW maneuvers the ship to berth on the ship bridge using the simulator, and there are six DOFs for the fixed ship model in that case. In our thesis, the complicated dynamic conditions are simplified to learn the procedures that the experienced OOW maneuver ships by operating different telegraph and rudder orders to change ship's speed and direction, and to complete the ship's control in the designed specific navigational scenario (no matter the changing environmental conditions and the dynamic situations for the ship, the OOW always needs to maneuver the ship through the combination of telegraph and rudder orders).

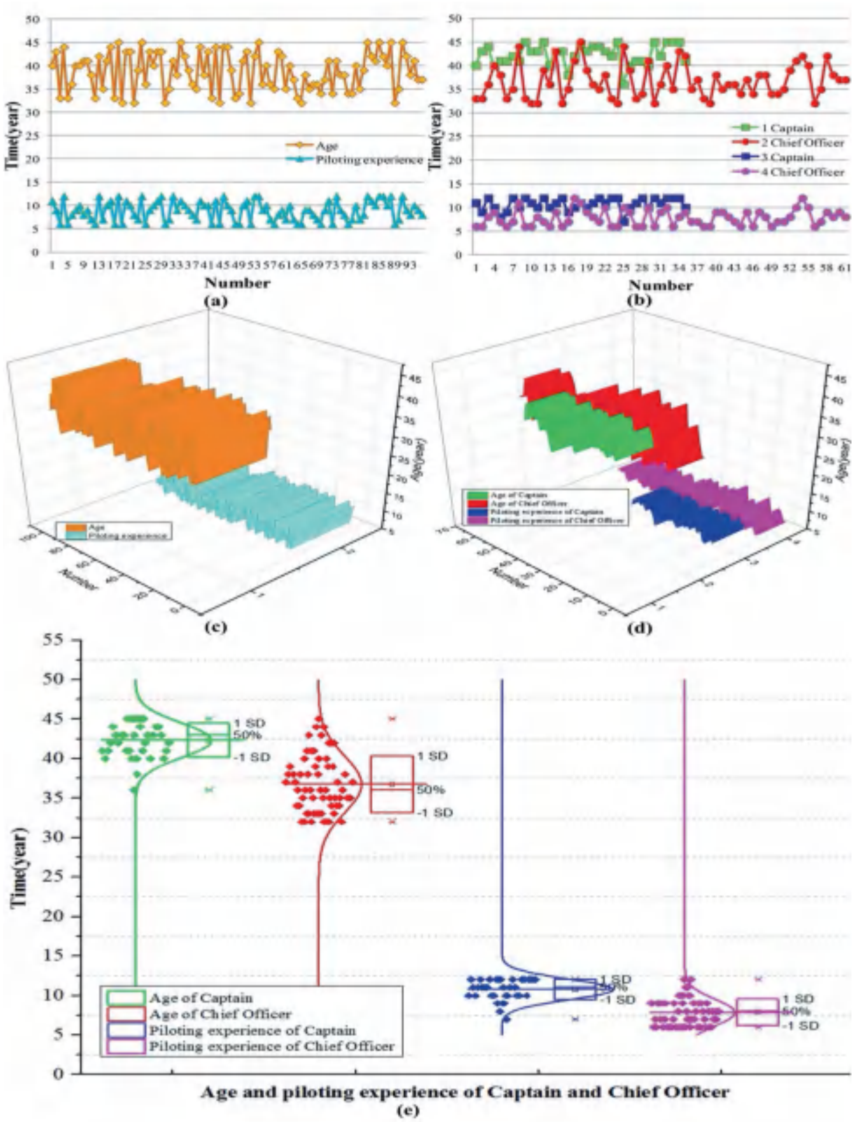


Figure 4.5 Analysis of participants' information.

Table 4.2 Participants' information.

	Number of Participants	Age (years)		Maneuvering 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

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 chapter, experimental data of each maneuvering decision-making factor are trisected into three levels, namely, small (a), medium (b), and large (c).

Table 4.3 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]	[-120.0000, -60.0000]	(-180.0000, -120.0000)
		[0.0000, 60.0000]	[60.0000, 120.0000]	[120.0000, 180.0000]
Y4	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, 124.8000]
Y5	Relative wind direction(degrees)	[-59.0205, 0.0000]	[-118.0411, -59.0205]	(-179.7170, -118.0411)
		[0.0000, 59.0205]	[59.0205, 118.0411]	[118.0411, 179.8750]
Y6	Relative wind speed(knots)	(0.0228, 7.5154)	[7.5154, 14.7664]	[14.7664, 22.1793]

Table 4.4 Training samples for the studied area (partially).

No.	X		Y1	Y2	Y3	Y4	Y5	Y6
	Rudder order (Degrees)	Telegraph order (%)						
1	-1.070	-30.000	318.400	1.108	-173.755	83.414	-77.200	7.724
2	-1.767	-30.000	318.400	1.108	-173.387	83.553	-77.200	7.724
3	-2.000	-30.000	318.400	1.108	-173.291	83.605	-77.195	7.723
4	-2.000	-30.000	318.400	1.108	-172.999	83.751	-77.100	6.704
5	-2.000	-30.000	318.400	1.108	-172.891	83.809	-77.095	6.704
6	-2.000	-30.000	318.400	1.108	-172.651	84.000	-77.000	6.704
7	-2.000	-30.000	318.400	1.108	-172.600	84.000	-77.000	6.704
8	-2.000	-30.000	318.400	1.108	-172.202	84.349	-76.931	6.685
9	-2.000	-30.000	318.400	1.108	-172.100	84.400	-76.900	6.685
10	-2.000	-30.000	318.400	1.108	-171.716	84.684	-76.758	6.685
11	-2.000	-30.000	318.400	1.108	-171.600	84.800	-76.700	6.685
12	-2.000	-30.000	318.400	1.108	-171.307	85.093	-76.633	6.685
13	-2.000	-30.000	318.400	1.108	-171.200	85.200	-76.600	6.685

14	-2.000	-30.000	318.400	1.108	-170.900	85.300	-76.600	6.685
15	-2.000	-30.000	318.400	1.108	-170.900	85.300	-76.600	6.685
…	…	…	…	…	…	…	…	…

Moreover, 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 (Y1), current speed (Y2), relative current direction (Y3), relative wave direction (Y4), relative wind direction (Y5), relative wind speed (Y6). Note that the factors Y1-Y6 here are different from the influencing factors, which have the same subscript number in Chapters 3 and 6. Table 4.4 lists some of the training samples.

Table 4.5 The symbol of telegraph orders (speed control) and rudder orders (course control).

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
Direction	$a_i > 0$	Ahead	D1	$b_i > 0$	Starboard	D2
	$a_i = 0$	Stop/Standby/Finished with engines	M1	$b_i = 0$	Midships	M2
	$a_i < 0$	Astern	T1	$b_i < 0$	Port	T2

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, the symbol are shown in Table 4.5. The standardization principle of the input selected influence factors for inbound maneuvering decision-making can be seen from Table 4.3. Moreover, Table 4.6 shows the combining telegraph and rudder orders and the standardization principle for output maneuvering decision-making factors. For instance, X1 (U1D1U2T2) indicates that the maneuvering decision-making is: {Keep the propeller forward and keep the current rudder angle-port rudder}.

Table 4.6 Ship maneuvering decision-making factors and standardization principle (output).

Maneuvering factors	**Decisions**	**Symbols**	**Decisions**	**Symbols**
X(Dimensionless)	U1D1U2T2	X1	C1M1C2T2	X19
	U1T1U2T2	X2	C1M1C2D2	X20
	U1D1U2D2	X3	U1M1C2T2	X21
	U1T1U2D2	X4	U1M1C2D2	X22
	C1D1C2T2	X5	C1M1U2T2	X23
	C1T1C2T2	X6	C1M1U2D2	X24
	C1D1C2D2	X7	U1D1U2M2	X25
	C1T1C2D2	X8	U1T1U2M2	X26
	U1D1C2T2	X9	C1D1C2M2	X27
	U1T1C2T2	X10	C1T1C2M2	X28
	U1D1C2D2	X11	U1D1C2M2	X29
	U1T1C2D2	X12	U1T1C2M2	X30
	C1D1U2T2	X13	C1D1U2M2	X31

C1T1U2T2	X14	C1T1U2M2	X32
C1D1U2D2	X15	U1M1U2M2	X33
C1T1U2D2	X16	U1M1C2M2	X34
U1M1U2T2	X17	C1M1U2M2	X35
U1M1U2D2	X18	C1M1C2M2	X36

4.4 Results

4.4.1 Standardizing of training set

The data in Table 4.4 are standardized according to the principle of standardization of maneuvering decision influence factors in Table 4.3, Table 4.5 and Table 4.6; then we get the training set as shown in Table 4.7.

Table 4.7 Training set for constructing the decision tree (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
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

4.4.2 Constructing the decision tree

In the ID3 decision tree algorithm 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 the proposed model in Section 4.2, we could obtain the decision tree structure, as shown in Figure 4.6. Then we can get the decision-making rule set based on the decision tree structure in the form of IF-THEN. Each path from the root node to the leaf node constitutes a rule. For instance, we can get the rule from the left side of the tree structure: IF Y2=a AND Y3=a AND Y6=a THEN X=X1. 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.

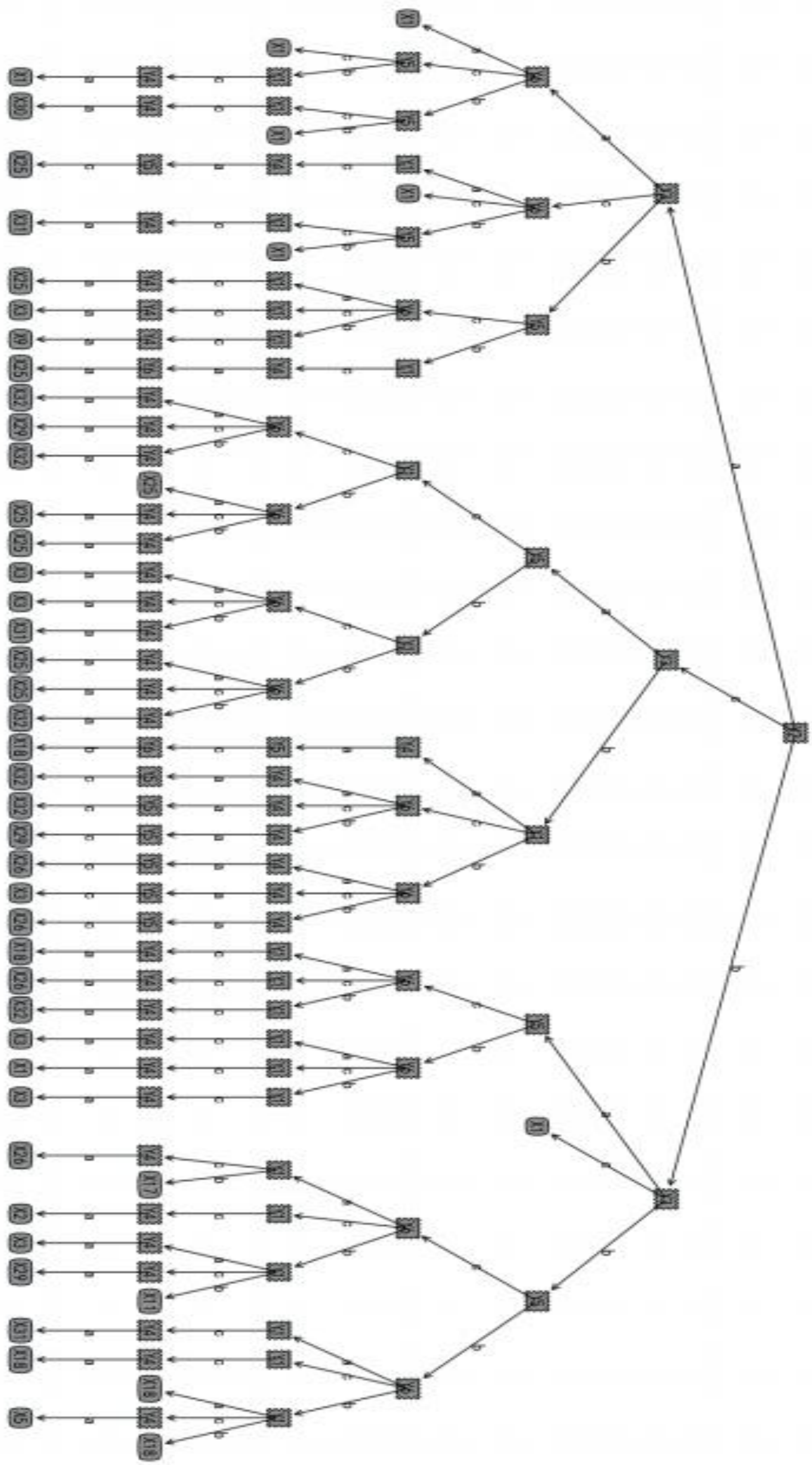


Figure 4.6 The experimental decision tree structure.

4.5 Conclusions

With the continuous development of large-scale, high-speed and professional ships, and the increasing construction of modern intelligent deep-water ports, the safety of inbound merchant ships receives more and more attention. In addition, with the development of modern science and technology, the improvement of autonomous ships has been technically feasible. The purpose of this chapter is to recognize the automatic acquisition and representation of the seafarer's decision-making knowledge and to provide a basis and reference for the development of decision-making algorithms for autonomous ships. In this chapter, to achieve the above research objective, some preparations and primary conditions for our research are conducted and set up, e.g., the experimental scenario, the research design, the data collection and processing, etc. In addition, the standardization principle of ship maneuvering is introduced, and the ID3 decision tree model for learning human-like decision-making mechanisms of autonomous ships is proposed for the first exploration and pre-study.

This chapter provides a new perspective for the development of autonomous ships and promotes the application and promotion of autonomous ships in a specific scenario. In the follow-up study, we will pay attention to the detailed standardization principles of various influencing factors and maneuvering decision-making factors and the application of our proposed model, both in the aspects of the detailed application framework and decision tree algorithm.

Chapter 5 Modeling human-like decision-making for autonomous ships using navigational decision tree

Chapter 4 presents the primary experimental preparations for our research. Moreover, the standardization principle of ship maneuvering is introduced, and the ID3 decision tree model for learning human-like decision-making mechanisms of autonomous ships is proposed. Based on the actual seafarers' operational data from the full-task handling simulation platform, this chapter combines a 30,000-ton bulk carrier inbound navigation scenario and uses the C4.5 decision tree method to propose a knowledge learning model under multiple environmental constraints to give autonomous ships the ability to make decisions like a human: An autonomous 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 study provides a feasible basis for the human-like decision-making analysis of autonomous ships.

The structure of this chapter is organized as follows. Initially, the background and the main contributions of this chapter are introduced in Section 5.1. Section 5.2 briefly presents the proposed decision-making model. The experimental processes are introduced in Section 5.3. Section 5.4 details the experimental results and the performance of our optimization methodology. The conclusions and future directions of research are addressed in Section 5.5.

The main content of this chapter is based on the following published paper:

Xue, J., Chen, Z., Papadimitriou, E., Wu, C., & van Gelder, P. H. A. J. M. (2019). Influence of environmental factors on human-like decision-making for intelligent ship. *Ocean Engineering*, *186*, 106060.

5.1 Introduction

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 directly or indirectly 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. 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, sea states, topographical environments and water facilities will bring restrictions to the navigation of the ship. These factors affect the ship's navigation and the seafarer'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.

Autonomous ships use sensors, communications, Internet of Things, the Internet and other technical ways 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 autonomous 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., 2015a). Even under a good navigation situation, the seafarer must be handled when an emergency occurs. Although the ship is maneuvered by satellite navigation, electronic compass, Electronic Nautical Charts (ENC), and autopilot system, the bridge has not been unmanned. Autonomous 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 autonomous ship for the maneuvering decision-making characteristics of Officer On Watch (OOW) and lacks the corresponding theoretical methods to solve the problem of autonomous ship human-like maneuvering decision-making modeling.

Researchers have proposed several different decision tree algorithms 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 chapter uses the C4.5 algorithm to learn the OOW's maneuvering decision characteristics. We regard the autonomous 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.

Overall, this study focuses on the concept of human-like maneuvering for the autonomous ship and studies the human-like decision-making method of autonomous 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 autonomous ships are studied, and the autonomous ship human-like decision-making model is constructed. The main contributions of this chapter are as follows:

1) A novel autonomous 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 OOW' 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.

5.2 Methodology

5.2.1 A navigational decision tree

The input of the decision tree learning algorithm is a set of training samples represented by combinations of attributes, attribute values and decisions registered for them, 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.

Figure 5.1(c) shows an example of a typical binary decision tree based on the data shown in Table 5.1. From Figure 5.1(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 (i.e., Class, which represents the crossing situation) represents the class category or decision of each instance.

Table 5.1 The data for the example.

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

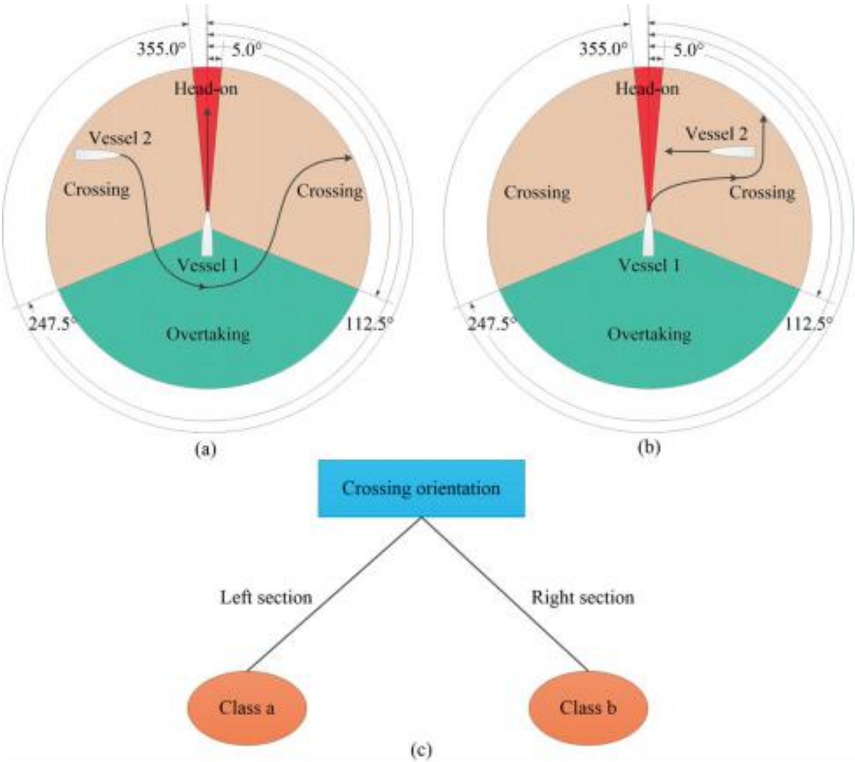


Figure 5.1 The collision avoidance operation in the encounter scenario of different crossing situations and the decision tree generated from this case.

Furthermore, the COLREGs (Convention on International Regulations for Preventing Collisions at Sea) navigation rules provide safe operation guidelines for maritime navigation. As shown in Figure 5.1(a), if Vessel 2 is at the left crossing section (Class a), Vessel 2 should turn right and Vessel 1 should keep its course; if Vessel 2 is at the right crossing section (Class b), Vessel 1 should turn right, and Vessel 2 should keep its course, shown as Figure 5.1(b). Therefore, the final decision can also be represented through the form of IF-THEN rule set shown as follows:

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.

This way, the attributes taken together provide a zeroth-order language for characterizing objects in the universe (Quinlan, 1986). Note that the COLREGs should be implemented explicitly and followed by every Offer On Watch (OOW) when navigating at sea. In this case, as shown in Figures 5.1 (a) and 5.1(b), it strictly followed the COLREGs of crossing situation: when two power-driven vessels are crossing so as to involve risk of collision, the vessel which has the other on her own starboard side shall keep out of the way and shall, if the circumstances of the case admit, avoid crossing ahead of the other vessel. This is only a simple example to let the readers easily understand the decision tree's basic logic and structure and how the final decision can be represented through the form of the IF-THEN rule set. However, in this thesis, no other ships sail in the designed experimental inbound scenario (from the initial boundary and end boundary shown in Figure 3.4), so some situations, such as overtaking, head-on, as well crossing situation, are not applicable in our case.

Generally, 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.

5.2.2 The proposed HDMDR model

Information gain ratio

From Chapter 4, we could get the information gain of feature/attribute A based on the Equations (4.1) to (4.7).

The split information is defined as:

$$Split(A) = -\sum_{j=1}^{n} \frac{|D_j|}{|D|} \log_2 \left(\frac{|D_j|}{|D|}\right), \tag{5.1}$$

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:

$$GainRatio(A) = \frac{Gain(A)}{Split(A)}. \tag{5.2}$$

Constructing the C4.5 decision tree

C4.5 is an extension of ID3 and was presented by J.R. Quinlan (Quinlan, 1993). ID3 selects the attribute with the largest information gain value as the node of the tree. 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 autonomous ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model.

Algorithm 5.1 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

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 (Because 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 chapter, the upper limit of the 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:

$$P\left[\frac{|f-q|}{\sqrt{q(1-q)/N}} > \mu_{1-\alpha'}\right] = \alpha', \tag{5.3}$$

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 limit of the confidence interval as the pessimistic error rate estimate of this node. Then, the equation for calculating the false positive rate of the node:

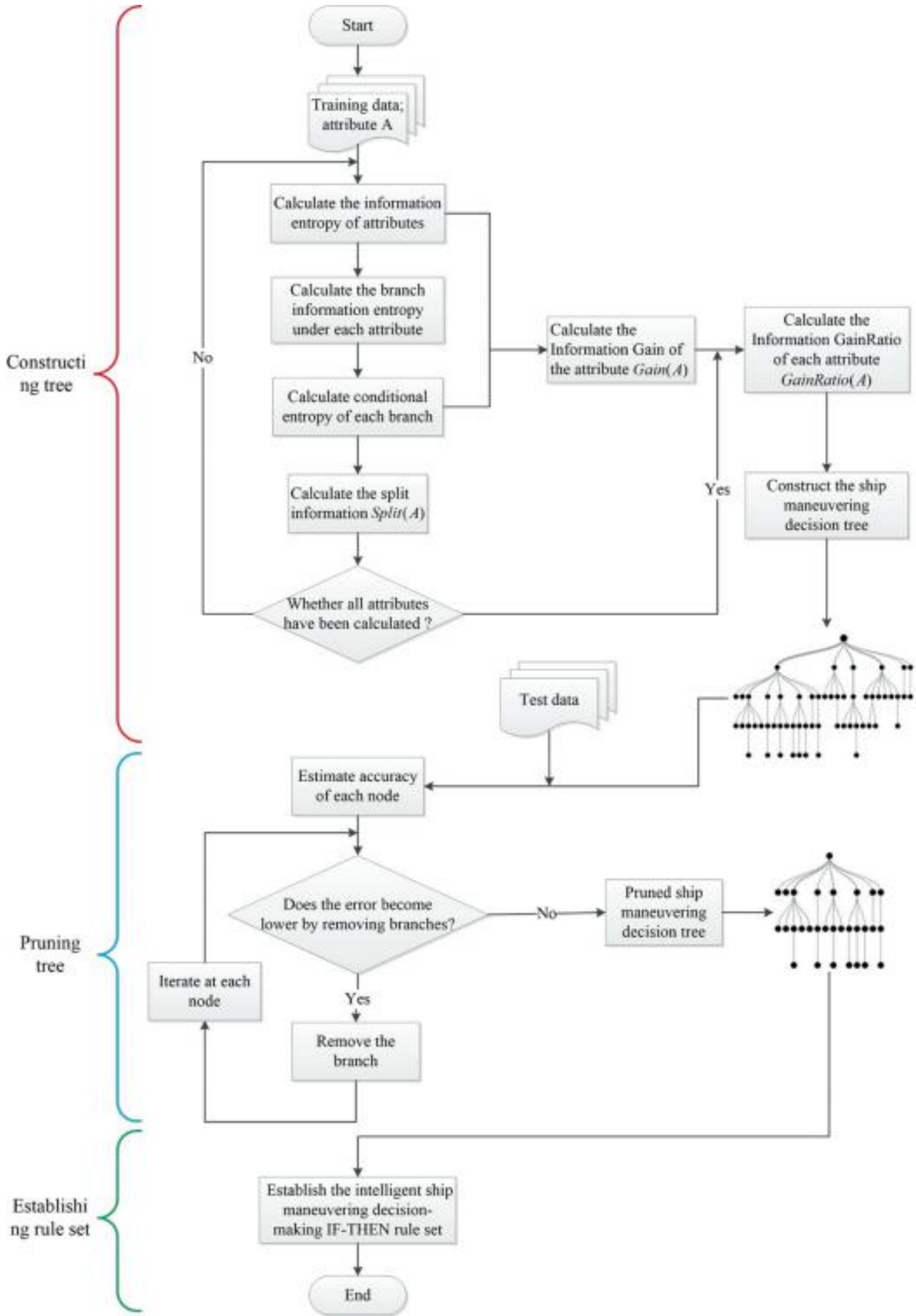


Figure 5.2 Framework of the proposed HDMDR model.

$$q=\frac{f+\frac{z^2}{2N}+z\sqrt{\frac{f}{N}-\frac{f^2}{N}+\frac{z^2}{4N^2}}}{1+\frac{z^2}{N}}, \tag{5.4}$$

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 $E_{\max}$. If the estimated false positive rate q after pruning is higher than $E_{\max}$, 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 Figure 5.5. Figure 5.2 is the basic process and framework for our proposed HDMDR model.

5.3 Experiments

5.3.1 Scenario design and data collection

The scenario design was the same as the aforementioned scenario in Chapters 3 and 4, i.e., the ship was downstream berthing into the Shanghai Waigaoqiao wharf port. Additionally, the data collection and basic processing procedures are the same as well.

We collect the operational data of the exercises and assessment exams as our experimental data (unlimited navigational class seafarer, 4 groups of 96 people, 32-45 years old, skilled maneuvering level, captain/chief officer). From Figure 5.3, we can get the distribution of OOWs' age and their maneuvering 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 5.2 lists some of the training samples. Note that the factors Y1-Y6 here are the same as the influencing factors in Chapter 4.

Table 5.2 Training samples for evaluation of the studied area (partially).

No.	X		Y1	Y2	Y3	Y4	Y5	Y6
	Rudder order (Degrees)	Telegraph order (%)						
1	-2.0000	-30.0000	318.4000	1.1080	-170.3351	85.6474	-76.3763	6.6652
2	-2.0000	-30.0000	318.4000	1.1080	-170.1000	86.0000	-76.2000	6.6652
3	-2.0000	-35.9724	318.4000	1.0754	-169.9281	86.2437	-76.0806	6.6652
4	-2.0000	-40.0000	318.4000	1.0753	-169.8000	86.5000	-76.0000	6.6652
5	-2.0000	-45.6276	318.4000	1.0753	-169.4059	86.7626	-75.8687	6.6752
6	-2.0000	-50.0000	318.4000	1.0755	-169.2000	86.9000	-75.8000	6.6852
7	-2.0000	-50.0000	318.4000	1.1080	-168.9564	87.1718	-75.6761	6.7052
8	-2.0000	-50.0000	318.4000	1.1080	-168.5000	87.4000	-75.6000	6.7552
9	-2.0000	-50.0000	318.4000	1.1080	-168.6957	87.6273	-75.5497	6.8043
10	-2.0000	-50.0000	318.4000	1.1080	-169.3000	87.8000	-75.5000	7.1655
11	-2.0000	-62.0366	318.4000	1.1080	-168.0185	88.1405	-75.4000	7.2612
12	-2.0000	-70.0000	318.4000	1.0755	-167.7000	88.3000	-75.4000	7.3272
13	-1.9030	-70.0000	318.4000	1.1080	-167.5368	88.5632	-75.3000	7.6652

14	-1.8120	-70.0000	318.4000	1.1080	-167.3000	88.8000	-75.3000	7.6652
15	-1.7090	-70.0000	318.4000	1.1080	-166.6993	89.0801	-75.3000	7.6510
…	…	…	…	…	…	…	…	…

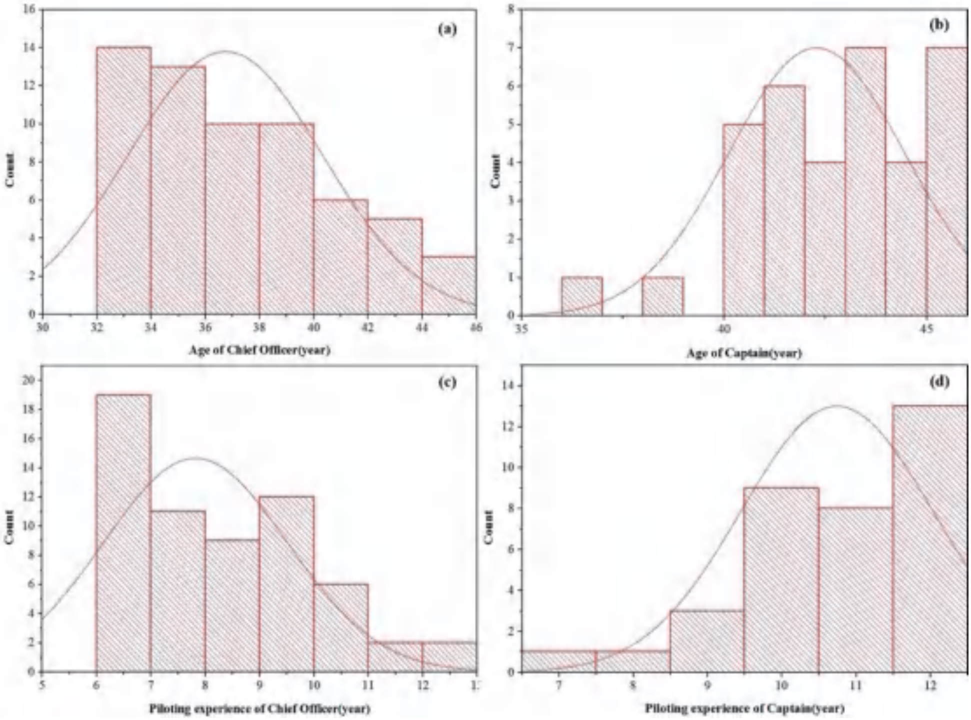


Figure 5.3 The distribution of OOWs' age and their maneuvering experience.

5.3.2 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 autonomous 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 study, experimental data of each maneuvering decision-making factor are trisected into three levels, namely, small (a), medium (b), and large (c), see Table 4.3 in Chapter 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 case, the other new factors can also be upgraded according to Algorithm 4.1 in Section 4.2.2 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 5.3 shows the combining telegraph and rudder orders (speed and course control respectively); this control is a multi-dynamic process. Moreover, it should be noted that, 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 in the actual situation of the experimental scenario. Therefore, in this chapter, we do not consider "Midships" and "Stop/Standby/Finished with engines", regardless of the rudder angle and if the power output is 0. Table 5.3 shows the standardization principle for output maneuvering decision-making factors.

Table 5.3 Maneuvering decision-making factors and the proposed 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
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

5.4 Results and discussion

5.4.1 Standardizing of training set

The data in Table 5.2 are standardized according to the principle of standardization of maneuvering decision influencing factors in Table 4.3 and Table 5.3. Then the standardized training set is obtained, as shown in Table 5.4.

Table 5.4 The standardized training set (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	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
2	X2	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
3	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
4	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
5	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
6	X14	0	0	1	0	1	0	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	X14	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
12	X14	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
13	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
14	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
15	X10	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0
…	…				…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

5.4.2 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 Equations (4.3) to (4.7) and (5.1) to (5.4), we could obtain the decision tree structure, as shown in Figure 5.5, partitioned into 3 parts, Part I, II and III. The number and proportion of different decisions are shown in Figure 5.4.

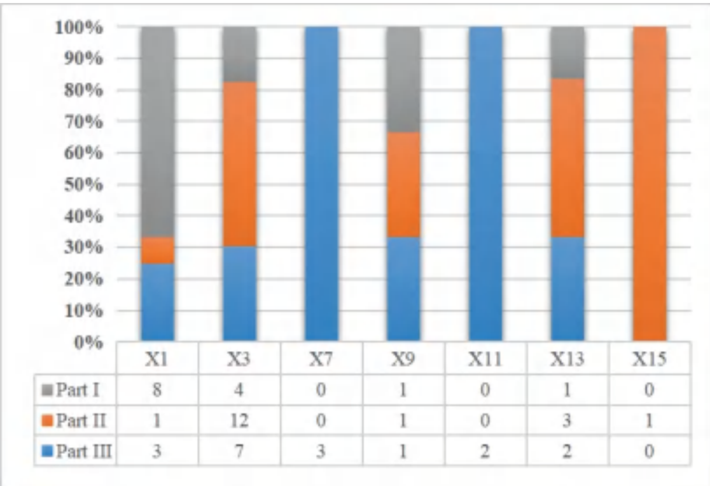


Figure 5.4 The number and proportion of different decisions.

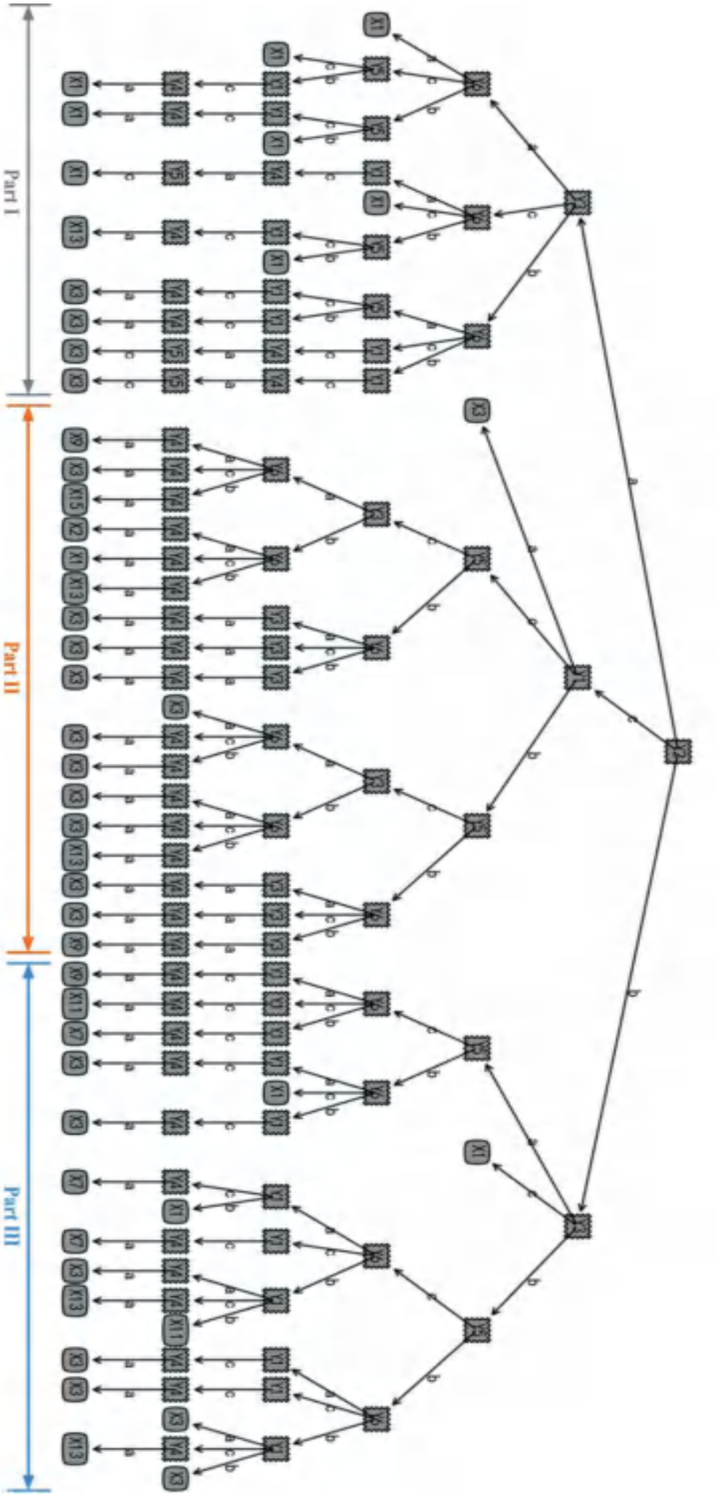


Figure 5.5 The proposed navigational decision tree structure.

5.4.3 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 recognition rule set is shown in Table 5.5.

Table 5.5 The established ship 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*

5.4.4 Performances assessment

Applying rules for classification

We use the maneuvering decision-making model proposed in this chapter to identify the decision-making data to be identified in Table 5.6. 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 5.6, 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. For the classified dataset, the performance could be measured by using a confusion matrix/contingency, the accuracy of the proposed module (ACC) could be calculated as:

$$ACC = \frac{TN + TP}{TN + TP + FN + FP}, \tag{5.5}$$

where TN is true negatives (correct negative assignments), TP is true positives (correct positive assignments), FN is false negatives (incorrect negative assignments), and FP is false positives (incorrect positive assignments). The classification average accuracy of our proposed HDMDR model using C4.5 decision trees based on the test data set can reach about 80.58%.

Table 5.6 Maneuvering decision data to be identified and its standardization.

Maneuvering decision aata to be identified

No.	**X (Actual maneuvering decision)**		**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6**
	Rudder order	**Telegraph 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**		
	Rudder order	**Telegraph 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

Comparative analysis

To further validate the effectiveness of the HDMDR model, in this chapter, we compare the performance of the proposed C4.5 decision tree algorithm with two classic classification algorithms: k-NN and 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 Equation (4.12), to measure the proposed C4.5 algorithm. In addition, in this chapter, 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 5.7 and Figure 5.6. According to the classification accuracy results, the proposed method can achieve the highest accuracy among these three algorithms.

Table 5.7 The performance of different classifier algorithms.

Classifier algorithms	**Accuracy (performance is measured in %)**	
k-NN	**Fold**	**Accuracy**
	1	75.36
	2	73.79
	3	73.81
	4	74.62
	5	72.87
	6	76.86
	7	72.37
	8	75.72
	9	74.89
	10	75.66
Average	-	74.60
SVM		**Accuracy**
	1	70.26
	2	72.62
	3	75.43
	4	73.62
	5	77.79
	6	72.83
	7	70.29
	8	71.63
	9	74.13
	10	73.09
Average	-	73.17
Proposed method		**Accuracy**
	1	80.33
	2	79.88
	3	83.42
	4	76.59
	5	79.16
	6	83.76
	7	79.78

	8	82.56
	9	81.86
	10	78.43
Average	-	80.58

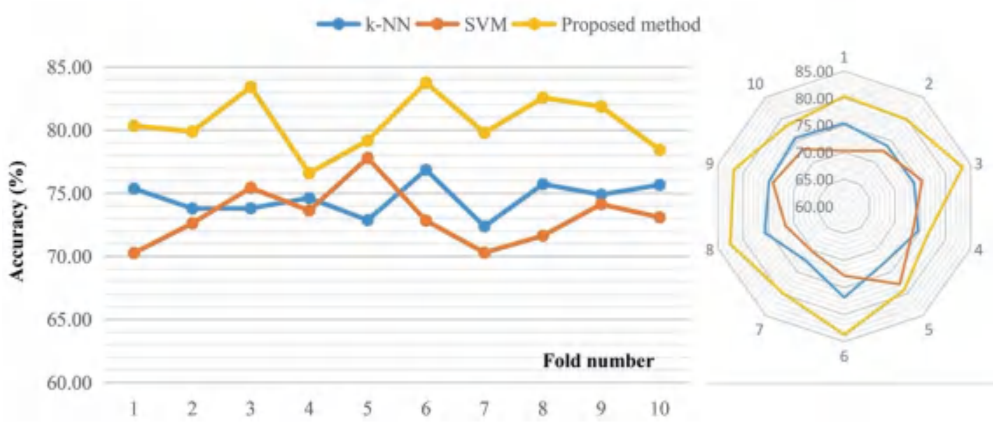


Figure 5.6 The accuracy of different classifier algorithms.

5.5 Conclusions

In this chapter, an autonomous 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 autonomous 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 chapter provides a new perspective and methodology for the development of autonomous ship maneuvering decision-making technology in theory and practice, promotes the application and spreading of autonomous ships under specific scenarios, and is conducive to the development of water transportation in the direction of safety, sustainability and economy. In the subsequent research, some advanced machine learning methods and various inbound and outbound scenarios with more complicated navigational situations will be explored combined with COLREGs rules or predicted trajectories based on the historical AIS data (in the broader region from offshore to the port). Additionally, the remote human-machine interaction assistant systems in combination with the information from VTS could also be employed in the future. In addition, in Chapter 6, we will study the classification of the influencing factors, the fuzzy processing of data sets, etc., to further optimize our proposed algorithm.

Chapter 6 Modeling human-like decision-making for autonomous ships based on fuzzy decision tree

In this chapter, 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 maneuvering behaviors in order to realize the automatic acquisition and representation of the seafarer's decision-making knowledge in inbound ship analysis as well as the simulated reproduction of the seafarer's behavior. The simulation results show that the maneuvering decision recognition model, based on the fuzzy Iterative Dichotomiser 3 (ID3) decision tree, possesses a high reasoning speed and can accurately identify current maneuvering behavior. This provides theoretical guidance and a feasibility basis for research into human-like maneuvering behavior and the realization of automatic autonomous ship maneuvering systems.

This chapter is organized as follows. First, Section 6.1 introduces the motivation and background of this chapter. Second, the maneuvering decision-making model and the optimization methodology are given in Section 6.2. Then, the experimental processes, the method of classification interval division, and the standardization principle of maneuvering decision-making factors are introduced in Section 6.3. Finally, the performance of the model and optimization methodology are shown in section 6.4, and we end with conclusions in Section 6.5.

The content of this chapter is an edited version of the following published paper:

Xue, J., Wu, C., Chen, Z., van Gelder, P. H. A. J. M., & Yan, X. (2019). Modeling human-like decision-making for inbound smart ships based on fuzzy decision trees. *Expert Systems with Applications*, *115*, 172-188.

6.1 Introduction

The further development of marine and information technologies promotes intelligence, green policies and automation become mainstream for global cargo ships. Ship labor costs increase every year, so for the foreseeable future, the number of experienced seafarers will be greatly reduced as autonomous ship emergence accelerates. At present, there is no mature research system for the human-like maneuvering of autonomous ships. Ship maneuvering 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 is 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 in the introduction part of Chapter 2. 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 maneuvering.

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 and Jiang, 2011). In these algorithms, the C4.5 algorithm is very 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 and Wehenkel, 2003).

In view of this, in this chapter, 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 maneuvering behaviors in order to realize the automatic acquisition and representation of a seafarer'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 maneuvering behavior accurately and has high reasoning efficiency, which provides theoretical guidance and feasibility bases for the simulation and realization of autonomous ship automatic maneuvering systems.

6.2 Maneuvering decision-making model

6.2.1 Grey relation entropy model

The grey relational 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 et al., 1996). The grey relation entropy analysis method is based on the grey relational 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.

Grey relational grade

Let X be grey relation factor set (discrete series), $X_0 = \{x_0(k) \mid k = 1,2,\cdots,m\}$ as reference columns and $X_i = \{x_i(k) \mid k = 1,2,\cdots,m\}\ (i = 1,2,\cdots,n)$ 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 Equations (6.1) and (6.2):

$$X_0^{'} = \left\{ \left[x_0(k) \text{-} \frac{1}{m} \sum_{k=1}^{m} x_0(k) \right] / S_0^{'} \mid k = 1,2,\cdots,m \right\}, \tag{6.1}$$

$$X_i^{'} = \left\{ \left[x_i(k) \text{-} \frac{1}{m} \sum_{k=1}^{m} x_i(k) \right] / S_i^{'} \mid k = 1,2,\cdots,m \right\} (i = 1,2,3,...,n), \tag{6.2}$$

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:

$$\xi_i(x_0(k), x_i(k)) = \frac{\min\limits_i(\min\limits_k |x_0(k) - x_i(k)|) + \rho \max\limits_i(\max\limits_k |x_0(k) - x_i(k)|)}{|x_0(k) - x_i(k)| + \rho \max\limits_i(\max\limits_k |x_0(k) - x_i(k)|)}, \tag{6.3}$$

among them, $\min\limits_i(\min\limits_k |x_0(k) - x_i(k)|)$ is the minimum difference of two levels, and $\max\limits_i(\max\limits_k |x_0(k) - x_i(k)|)$ 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 et al., 2014b).

Grey relation entropy

Let X be grey relation factor set (discrete series), $X_0 = \{x_0(k) \mid k = 1,2,\cdots,m\}$ as reference columns and $X_i = \{x_i(k) \mid k = 1,2,\cdots,m\}\ (i = 1,2,\cdots,n)$ as comparison columns.

$R_i = \{\xi_i(x_0(k), x_i(k)) \mid k = 1, 2, \cdots, m\}$, so the grey correlation coefficient distribution map is called the density value of the distribution, as shown in Equation (6.4):

$$P_i \stackrel{\Delta}{=} \frac{\xi_i(x_0(i), x_i(i))}{\sum_{k=1}^{m} \xi_i(x_0(i), x_i(i))}. \tag{6.4}$$

The grey relation entropy of X_i is expressed as:

$$H(R_i) \stackrel{\Delta}{=} -\sum_{k=1}^{m} P_i \ln P_i . \tag{6.5}$$

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 equation is:

$$E(X_i) \stackrel{\Delta}{=} \frac{H(R_i)}{H_{\mathrm{m}}}. \tag{6.6}$$

Where H_{m} is the maximum value of grey entropy, $H_{\mathrm{m}} = \ln n$, 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 maneuvering decision (reflected on the control side, it is the ship rudder combination) as the reference sequence, and various influencing factors as the comparison sequence, then comprehensively judge the influence degree of each influence factor on the ship entry maneuvering decision, thus determining the order of each influence factors.

6.2.2 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 deceases the most, then it is optimal to choose it as a classification attribute. The fuzzy ID3 algorithm defines the information gain to represent the reduction of this entropy (Umanol et al., 1994), so the attribute with the largest information gain should be selected as the extended attribute of the node.

Set the domain as $D=\{e_1,e_2\cdots,e_n\}$ to represent the example set that summarizes the forecast rules. Each element $e_k(k=1,2,\cdots,n)$ in the example set has l fuzzy attributes: $A^1,A^2,\cdots,A^l$. The range of each attribute A^i is $\{a_1^i,a_2^i,\cdots,a_m^i\}(i=1,2,\cdots,l)$, the *j*-th example $e_j(j=1,2,\cdots,m)$ 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 $\{a_1^i,a_2^i,\cdots,a_m^i\}$ of A^i, and the classification to be divided is $C=\{C_1,C_2,\cdots,C_n\}$.

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

$$G(A^i,D)=I(D)-E(A^i,D), \tag{6.7}$$

where

$$I(D)=-\sum_{k=1}^{n}(P_k\cdot\log_2 P_k), \tag{6.8}$$

$$P_k=\frac{|D_{C_k}|}{|D|}, \tag{6.9}$$

$$E(A^i,D)=\sum_{j=1}^{m}\mu_{ij}, \tag{6.10}$$

$$P_{ij}=\frac{\left|D_{a_j^i}\right|}{\sum_{j=1}^{m}\left|D_{a_j^i}\right|}, \tag{6.11}$$

$$\mu_{ij}=P_{ij}\cdot I(D_{a_j^i}), \tag{6.12}$$

among them, let D_{C_k} to be a fuzzy subset in D whose class is C_k, $|D|$ is the sum of the membership values of the set of data D, and $|D_{C_k}|$ is the sum of the membership values of the set of data D_{C_k}, $\left|D_{a_j^i}\right|$ is the sum of the membership values of the set of data $D_{a_j^i}$ and calculate the $\sum_{j=1}^{m}\left|D_{a_j^i}\right|$. After that, we obtain the fuzzy information gain of each attribute at each node calculated by $G(A^i,D)$ through Equations (6.7) to (6.12) 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.

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 and 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 et al., 2010; Fan et al., 2011; Pulkkinen and Koivisto, 2008). The second step is 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 (Fan et al., 2011; Wang et al., 2015; Wang et al., 2008). Therefore, in this chapter, we use the triangular membership function to fuzzify the quantitative database. The following is a triangular membership function definition (Yuan and Shaw, 1995).

Definition 1: For all examples, attribute that A has a quantitative attribute value x, expressed as $X_A = \{x(u), u \in U\}$, We want to cluster X_A to k linguistic terms $T_i, i = 1, 2, \cdots, k$. And the triangular membership function equation for each linguistic term T_i is shown in Equations (5.13) to (5.15), the

$$u_{T_1}(x) = \begin{cases} 1, x \le m_1 \\ (m_2 - x)/(m_2 - m_1), m_1 < x < m_2, \\ 0, x \ge m_2 \end{cases} \tag{6.13}$$

$$u_{T_k}(x) = \begin{cases} 1, x \ge m_k \\ (x - m_{k-1})/(m_k - m_{k-1}), m_{k-1} < x < m_k, \\ 0, x \le m_{k-1} \end{cases} \tag{6.14}$$

$$u_{T_i}(x) = \begin{cases} 0, x \geq m_{i+1} \\ (m_{i+1} - x)/(m_{i+1} - m_i), m_i < x < m_{i+1} \\ (x - m_{i-1})/(m_i - m_{i-1}), m_{i-1} < x < m_i \\ 0, x \leq m_{i-1} \end{cases}, 1 < i < k. \tag{6.15}$$

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 Figure 6.1.

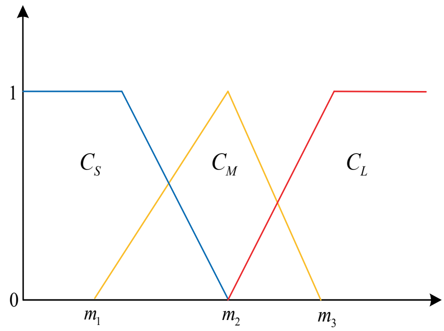


Figure 6.1 Example of triangular membership function.

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 (Maimon and Rokach, 2014). The CDT pruning method can be used in the FDT algorithm with only a few modifications (Yuan and Shaw, 1995).

The CDT algorithm is not suitable for pre-pruning. Therefore, it focuses on post-pruning methods (Esposito et al., 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 to 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 and 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 and Wang, 2006), the author, through the control variate method, noted that the vicinity of 0.7 to 0.8 usually obtains a better decision tree, and this conclusion has been verified by genetic algorithm; when α ranges from 0.30 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 chapter, 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 α =0.375 and β = 0.750, to control the pre-pruning process of FDT.

6.3 Experiments

The scenario design was the aforementioned scenario in Chapters 3-5, i.e., the ship was downstream berthing into the Shanghai Waigaoqiao wharf port. Additionally, the data collection and basic processing procedures are the same as well. Note that the factors Y1-Y33 here are the same as the influencing factors in Chapter 3 but different from the influencing factors, which have the same subscript number in Chapters 4 and 5.

Moreover, in this chapter, the optimized data set partitions and standardization principle for the input maneuvering decision-making factors are proposed. 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 and Shaw, 1995). For instance, the temperature can be conceptualized into three linguistic terms: Cool, Mild, and Hot. In this chapter, experimental data of each maneuvering decision-making factor are trisected 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 6.1 shows the processed data set partitions.

Table 6.1 The processed data set partitions and standardization principle of maneuvering 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)

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 6.2.

Table 6.2 The optimized data set partitions and standardization principle of maneuvering 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)
		[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, 124.8000]
Y9	Relative wind direction(degrees)	[-59.0205, 0.0000)	[-118.0411, -59.0205)	(-179.7170, -118.0411)
		[0.0000, 59.0205)	[59.0205, 118.0411)	[118.0411, 179.8750)
Y10	Relative wind	(0.0228, 7.5154)	[7.5154, 14.7664)	[14.7664,

	speed(knots)			22.1793)
Y11	Water depth(meters)	(12.5530, 14.0023)	[14.0023, 15.00113)	[15.00113, 16.0000]
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)
		[0.0000, 37.542302)	[37.5423, 75.0846)	[75.0846, 2607.3083)
Y14	Forces Parameters. Longitudinal force(tonne-force)	[-142.0715, 0.0000)	[-284.1429, -142.0715)	(-24592.3230, -284.1429)
		[0.0000, 142.0715)	—	—
Y15	Forces Parameters. Summary force(tonne-force)	[0.0000, 160.2039)	[160.2039, 320.4078)	[320.4078, 25144.6958)
Y16	Forces Parameters .Vertical force(tonne-force)	[-11.6871, 0.0000)	[-23.3742, -11.6871)	(-4547.0121, -23.3742)
		[0.0000, 11.6871)	[11.6871, 23.3742)	[23.3742, 2024.0176)
Y17	Mooring lines. Lateral force(tonne-force)	(-162.9374, 0.0000)	—	—
		[0.0000, 162.9374)	[162.9374, 325.8748)	[325.8748, 953.2618)
Y18	Mooring lines. Longitudinal force(tonne-force)	[-44.9968, 0.0000)	[-89.9937, -44.9968)	(-286.8233, -89.9937)
		[0.0000, 44.9968)	[44.9968, 89.9937)	[89.9937, 333.4608)
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)
		[0.0000, 0.7564)	[0.7564, 1.5129)	[1.5129, 4.0375)
Y24	Longitudinal speed(knots)	[-2.6398, 0.0000)	—	(-1.7988, -5.2797)
		[0.0000, 2.6398)	[2.6398, 5.2797)	[5.2797, 8.6558)
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)
		[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)	(-216.4195, -26.7401)
		[0.0000, 13.3700)	[13.3700, 26.7401)	[26.7401, 82.5781)
Y28	Roll angle(degrees)	(-7.1696, 0.0000)	—	—
		[0.0000, 1.4621)	[1.4621, 2.9242)	[2.9242, 12.1096)
Y29	Roll rate(degrees/min)	[-13.9547,	[-27.9093, -	(-219.3305, -

		0.0000)	13.9547)	27.9093)
		[0.0000, 13.9547)	[13.9547, 27.9093)	[27.9093, 220.8617)
Y30	Vertical speed(knots)	[-0.2744, 0.0000)	[-0.5488, -0.2744)	(-1.1612, -0.5488)
		[0.0000, 0.2744)	[0.2744, 0.5488)	[0.5488, 2.5488)
Y31	Yaw rate(degrees/min)	[-13.3700, 0.0000)	[-26.7401, -13.3700)	(-216.4195, -26.7401)
		[0.0000, 13.3700)	[13.3700, 26.7401)	[26.7401, 82.5781)
Y32	Latitude(degrees)	(31.1022, 31.3447)	[31.3447, 31.3483)	[31.3483, 31.3521)
Y33	Longitude(degrees)	(121.3144, 121.6444)	[121.6444, 121.6467)	[121.6467, 121.6494)

The method of classification interval division that is proposed in this chapter fully considers the distribution of data sets and the 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 OOW'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 6.3.

Table 6.3 Performance of the optimized data set partitions and standardization principle.

Segmentation of intervals	Total number	Selected number	Accuracy (%)	Total time (s)
Table 6.1	677650.00	675670.00	99.71	7.65
Table 6.2	677650.00	677650.00	100.00	7.73

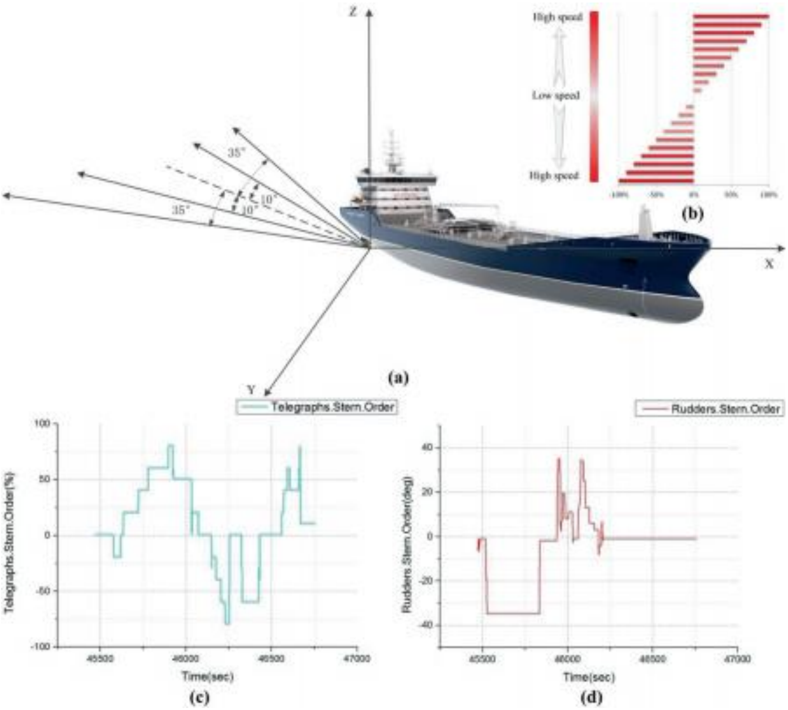


Figure 6.2 Analysis of the experimental ship.

According to the simulation scene shown in Figure 4.2 of Chapter 4 and Figure 6.2, 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 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 Figure 6.2 and Table 3.5 of Chapter 3 (showing 64 possible maneuvering factors, i.e., maneuvering decisions, are regarded as the output of our proposed model). In addition, this thesis does not consider "Midships" and "Stop/Standby/Finished with engines", regardless of the rudder angle and if the power output is 0.

Captains maneuver 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. Figures 6.2(c) and (d) reveal 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 Figures 4.2 (c), (f), and (g) of Chapter 4.

6.4 Results and discussions

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

6.4.1 Determining maneuvering decision main influence factor

Maneuvering decision-making processes are often influenced by multi-source information such as human, ship and environmental factors. These factors are collectively referred to as maneuvering decision-making factors. They act together to determine the next action strategy of the OOW. According to this strategy and the current maneuvering environment, the OOW can quickly and accurately develop maneuvering decisions and thus lay the foundation for the research of human-like maneuvering 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 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. Therefore, this chapter uses the grey relation entropy analysis to focus on the maneuvering decision-making factors from the waterway and environment, where the factors are squeezed and sorted. The maneuvering of decision-making factors related to the information are shown in Table 3.5 of Chapter 3 and Table 6.4.

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 uses entropy correlation degrees to complete the matching order of influencing factors. In this chapter, we select X0 (X0 presents the percentage of the number of each maneuvering 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 6.4 lists only a part of multiple measured data.

Table 6.4 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	…

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 6.4; the results are shown in Table 6.5 and Table 6.6.

Table 6.5 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	…

Table 6.6 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	—	—	—

According to Table 6.6, the influence factors are sorted according to the influence degree: Y8> Y22> Y4> Y31> Y27> Y16> Y33> Y11> Y3> Y13> Y14> Y15> Y12> Y30> Y5> Y18> Y6> Y26> Y20> Y10> Y17> Y19> Y23> Y25> Y29> Y24> Y28> Y32> Y2> Y1> Y9> Y21> Y7. For simplicity, this chapter selects the first six factors to study the decision-making mechanisms for different maneuvering behaviors. Table 6.7 lists some of the training samples.

Table 6.7 Training samples (partially).

No.	**X**		**Y4**	**Y8**	**Y16**	**Y22**	**Y27**	**Y31**
	Rudder order (Degrees)	**Telegraph 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
…	…	…	…	…	…	…	…	…

The data in Table 6.7 are standardized according to the principle of standardization of maneuvering decision influence factors in Table 6.2 and Table 3.5 of Chapter 3.

Table 6.8 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
…	…				…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

In accordance with the optimized standardization principle of influence factors for maneuvering decisions in Table 6.2, the attributes of the six factors selected in Table 6.7 are fuzzified, the number of center points k is 3, and the set of center points is $M = \{m_i,\ i = 1,\ 2,\ \cdots,\ k\}$. By simple division method, the linguistic term Small is calculated using Equation (5.13), the linguistic term Medium is obtained by Equation (6.15), and the linguistic term Large is obtained by Equation (6.14). 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 6.9. The training set with fuzzy representation is shown in Table 6.10.

Table 6.9 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

Table 6.10 Training set with fuzzy representation (partially).

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

6.4.2 Inducing maneuvering fuzzy decision tree

The maneuvering decision classification tree is constructed by using the fuzzy ID3 classification algorithms and fuzzy membership standard training samples in Table 6.10. The fuzzy ID3 classification algorithm is summarized as follows. First, select the maneuvering 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 maneuvering decision classification tree constructed by the fuzzy ID3 classification algorithm. And the classification rules are graphically represented by the decision tree structure in Figure 6.3. The algorithm scheme is as follows:

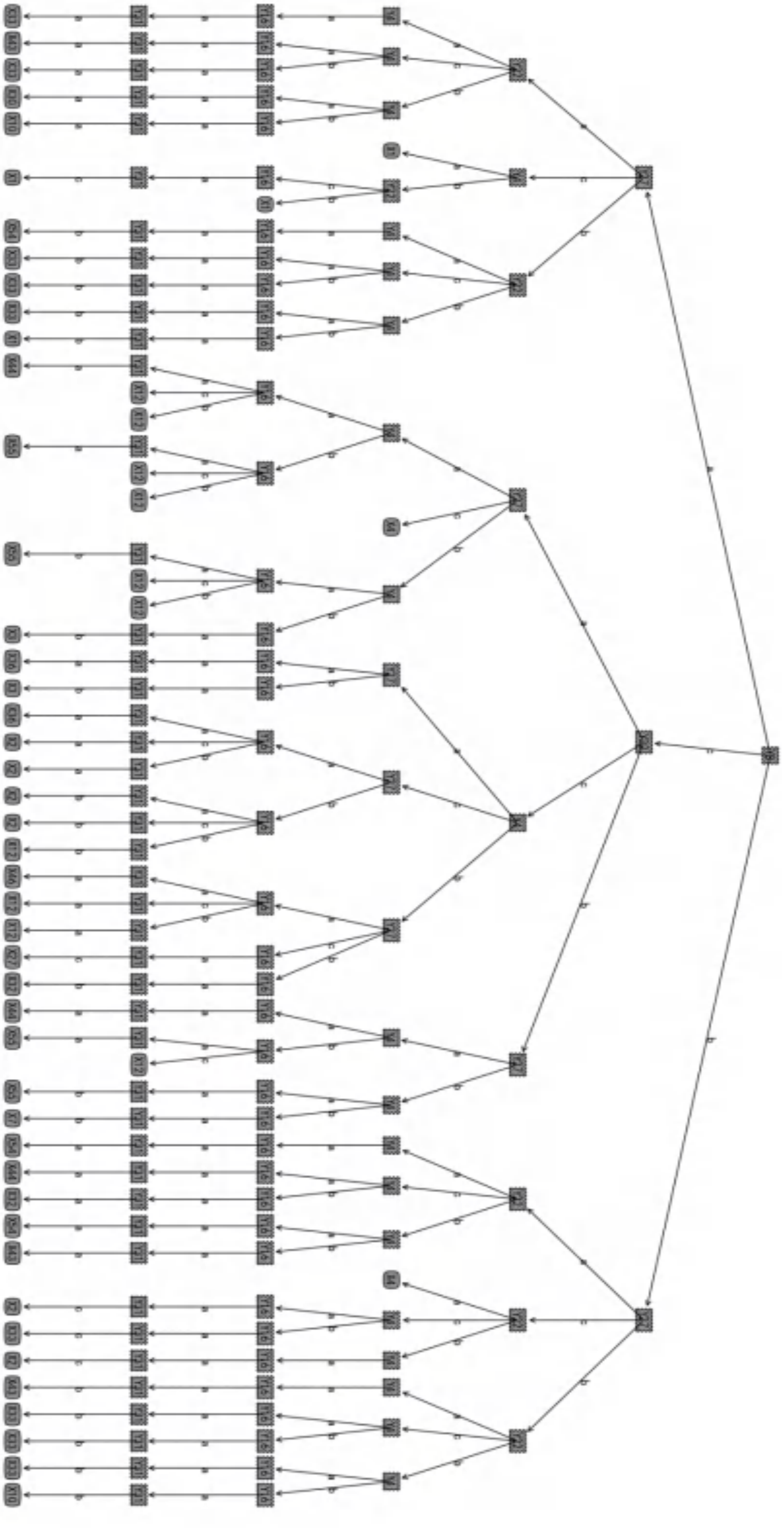


Figure 6.3 Decision tree structure for maneuvering decision classification rules.

1) Create a root node. 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 the root node. Then, calculate the fuzzy information gain $G(A^i, D)$ generated by each attribute A^i at the root node.

3) If the confidence level of a certain class in the node is greater than β (the truth level threshold used in this chapter 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. Conduct 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.

6.4.3 Establishing maneuvering decision classification rules

For the resulting maneuvering 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 maneuvering decision classification 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 maneuvering 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 maneuvering decision recognition rule base, as shown in Table 6.11.

Table 6.11 Maneuvering decision classification rules.

No.	Maneuvering 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=a 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*

The following conclusions can be drawn from Figure 6.3 and Table 6.11:

1) In the maneuvering decision-making process, maneuvering behavior is often stimulated and influenced by seafarers, ships, waterways, environment and other factors. These factors together lead the seafarer to gradually form the next moment action (action, strategy or tactics) in their mind. According to this long-term strategy and the current maneuvering environment, the seafarer can quickly and accurately develop maneuvering decisions and prepare to establish the seafarer's comprehensive cognitive sequence activity execution mechanism.

2) Maneuvering decision-making also depends on the seafarer's personality type. For instance, in the same navigational environment, when the variable speed change conditions are in a critical state, conservative seafarers will not risk changing speed or direction, even if this would lead to great traffic delay. Impulsive seafarers are more likely to risk shifting or changing direction.

3) Fuzzy ID3 decision tree sets features as main factors that determine whether the seafarer will take the main maneuvering force or take rudder operation during navigation, which is consistent with the actual navigation experience. These rules and the current skilled seafarer's background knowledge are also consistent. These factors can be an important reference for maneuvering 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%, close to or even exceeding the effect of an empirical seafarer judgment (under the premise of ensuring navigation safety, assuming duty hour is an 8-hour shift time, when the OOW could maintain efficient and correct maneuvering decisions more than 80% of the duty time, we consider that is an empirical judgment), which well proves the validity of the fuzzy ID3 decision tree algorithm in navigational maneuvering behavior data mining.

6.4.4 Comparative analysis

To validate the effectiveness of the fuzzy decision tree algorithm we proposed in this chapter, 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.

Table 6.12 The information of dataset used in our experiments.

Database	Number of Instances	Number of Features	Number of Classes
Our database	677655	18	64

To test the performance of the proposed Fuzzy ID3 algorithm, SVM, and Naïve Bayes (NB) are compared. And we use classification accuracy to measure the proposed Fuzzy ID3 algorithm. Assume the dataset $D=\{X_1, X_2, ..., X_n\}$, each data record is represented as $X_i=\{x_1, x_2, ..., x_n\}$, D also contains a set of classes $C=\{C_1, C_2, ..., C_n\}$, then we can get the classification performance symbol and equations or meanings (as the same as Chapter 5), as shown in Table 6.13.

Table 6.13 Classification performance symbols and their equations or meanings.

Symbol	Equation/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	$ACC = \dfrac{TN+TP}{TN+TP+FN+FP}$

For the dataset, described in Table 6.12, a ten-fold cross-validation (10-CV) is conducted, the performance of different classifier algorithms is shown in Table 6.14.

Table 6.14 The performance of different classifier algorithms with 10-fold cross-validation.

Classifier algorithms	**Accuracy (performance is measured in %)**	
SVM	**Fold**	**Accuracy**
	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**
	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
Proposed method		**Accuracy**
	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

The classification accuracy using different classifiers on our dataset is shown in Table 6.14. According to the classification accuracy results, the proposed method can achieve the highest accuracy among these three algorithms. And the proposed method can obtain the best average classification accuracy of 86.40%, followed by NB at 84.15%, and the SVM at 83.09%. Therefore, we can conclude that the proposed method outperforms the compared methods.

6.5 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 chapter, 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 maneuvering decision recognition model, and apply it to research on the decision-making mechanism of the different maneuvering 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 maneuvering decision recognition model, which has high reasoning efficiency (the results of this chapter show the proposed fuzzy ID 3 model has better performance (average accuracy is 86.40%) than the proposed C4.5 model in Chapter 5 (average accuracy is 80.58%)).

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 recognize the key factors which affect maneuvering decisions, accurately identify the current maneuvering behavior and provide guidance for an autonomous ship-assisted or automatic maneuvering system for the research of human-like maneuvering 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 seafarer 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 seafarer on the full-task handling simulation platform. However, there are still some problems with the models and experiments described in this chapter, 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 chapter, we do not take into account the impact of other vessels on the waterway. Based on the definition of the inbound scene proposed in this chapter, 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 maneuvering decisions.

2) In this chapter, considering the relevant output information of the maneuvering 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 maneuvering 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 maneuvering behavior decision-making theory and system for autonomous ship maneuvering behavior decision-making research to provide theoretical guidance and a feasibility basis for the development of autonomous ships.

Chapter 7 Conclusions and future research

This thesis focuses on the problem of modeling seafarers' navigational decision-making in typical scenario for autonomous ships' safety. We propose a method to prioritize safety influencing factors of autonomous ships' maneuvering decisions and a series of ship maneuvering knowledge learning models to give autonomous ship the ability to make decisions like a human.

In this chapter, the main research findings and the answers to the research questions are summarized in Section 7.1. Subsequently, the recommendations for future research and the limitations of this research are described in Section 7.2.

7.1 Answer to research questions

In this section, the main findings are summarized to answer the research questions in Section 1.3 of Chapter 1.

Main research question

How can the decision mechanisms of automatic acquisition and representation of the seafarer’s decision-making knowledge in a typical navigation scenario be recognized?

To answer the main research question, a comprehensive overview of the relevant literature related to the utilized methodology is conducted, and the materials contributing to our research are identified in Chapter 2. Then the grey system theory is introduced, in order to prioritize the influencing factors, and the fuzzy theory is presented for more rational use of expert knowledge to judge the prioritization of the influencing factors (Chapter 3) and to fuzzify the experimental dataset into several language items for the process of constructing decision trees (Chapter 6). Chapter 4 proposes an ID3 decision tree model for recognizing human-like decisions of autonomous ships in the specific ship maneuvering scenario for the first exploration and pre-study. An autonomous ship Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model is proposed in Chapter 5 to acquire knowledge under multiple environmental constraints. Another improvement takes place in Chapter 6, where a fuzzy decision tree model is developed, combining the comprehensibility of decision trees and the comprehensive expression ability of fuzzy technology, which has strong decision analysis ability and can address the problem of ambiguity and uncertainty.

More specifically, the six research sub-questions (SQ) relevant to the main research question addressed in Chapter 1 are answered as follows.

Questions on state-of-the-art and methodology

a) Which data analysis method is more suitable and effective for the selection of the main influencing factors of seafarers’ maneuvering decisions?

- As presented in Chapter 2, the grey system theory is one of the most widely utilized pattern recognition methods, it is mainly utilized to analyze the proximity of the dynamic grey process development situation, determine the primary and secondary factors in the grey system, and control the main factors affecting the system. Additionally, as a systematic analysis technique, the Grey Relational Analysis (GRA) is a quantitative comparative analysis method. By calculating the correlation between the target value and the influencing factors, and the ranking of their relevance, the main factors affecting the target value are sought. The GRA method is suitable for the data with uncertain, multiple inputs and discrete properties; it does provide techniques for determining an appropriate solution for real-world problems. Moreover, the GRA does not require too much sample size and does not require a typical distribution law during analysis. In addition, regardless of whether the system has adequate information, the GRA could capture the impact of the relationship between the main factor and influencing factors in the system. The results correspond to the qualitative analysis results, and the method has proved practicality. The grey relational analysis is an effective algorithm for resolving uncertainty problems in the case of partial and discontinuous information. However, the traditional GRA has mainly been criticized because it treats different indexes (influencing factors) equally and does not consider the relative importance of different indexes. It does not fit with people's preferences for a specific index. Nevertheless, the fuzzy logic theory is a beneficial method for modeling processes that are too complicated for conventional quantitative analysis or information obtained from the process is qualitative, uncertain, or inexact. Moreover, fuzzy numbers are more compatible with phrases and ambiguities; it is better to use them in real-world decision-making and reflect human thoughts. Furthermore, many studies are explored by combining expert knowledge with fuzzy theories, and it is widely used in the maritime research domain (Chapter 2).

Therefore, this thesis proposes a prioritizing model for the influencing factors of autonomous ship maneuvering decision-making using grey and fuzzy theories in Chapter 3. Based on the actual operation data of the experienced seafarers collected from the simulator, a reference series is established by using the combination of ship telegraph and rudder orders, which directly correspond to a ship's control. Likewise, we establish the comparative series for various influencing factors of ship motion and natural and traffic environment that affect ship maneuvering decision-making. Moreover, combined with the expert knowledge, the proposed model is further optimized to ensure its rationality, accuracy, and generalizability, to select/prioritize the main maritime traffic safety influencing factors of the autonomous ship maneuvering decisions in the specific navigational scenario.

b) Which approaches can be used to automatically acquire and represent the decision-making knowledge of the experienced seafarers' maneuvering behavior in a typical navigation scenario?

- The basic logic for the ship maneuvering decision-making mechanism is the classification of the ship's maneuvering and operating behavior according to specific 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 is 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 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 maneuvering. Therefore, in Chapters 4, we proposed an ID3 decision tree model for recognizing human-like decisions of autonomous ships in the designed ship maneuvering scenario for the first exploration and pre-study. Then Chapter 5 develops the HDMDR model using the C4.5 decision tree algorithm to acquire knowledge under multiple environmental constraints to give autonomous ships the ability to make decisions like a human. The OOW's decision-making knowledge is automatically acquired and represented.

  However, the decision tree construction algorithms aforementioned in Chapter 2 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. 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 is generally not robust, and inaccurate or missing data can prevent the growth of the decision tree. As a data mining method, the Fuzzy Decision Tree (FDT) extends 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. In Chapter 6, the fuzzy ID3 decision algorithm is applied to autonomous ship decision-making for the first time based on empirical simulator maneuvering data. The algorithm can reasonably complete the simulation and decision-making knowledge acquisition of the ship's automatic driving in experimental waters and has a high degree of application and promotion. The improvement of the algorithm's training set is of great significance for the development of autonomous merchant ships. Moreover, it is unique and very valuable to obtain experimental data operated by an experienced senior crew on the full-task handling simulation platform in a certain time and space.

Questions on modeling

c) What are the advantages and disadvantages of the prioritizing model of safety influencing factors of autonomous ships' maneuvering decisions?

- The proposed prioritizing model in Chapter 3 has the following four advantages: (i) By applying the expert knowledge to the process of prioritizing autonomous ship maneuvering decisions influencing factors, furthermore, by establishing fuzzy linguistic terms sets and the corresponding fuzzy numbers, the basis is provided for qualitative evaluation of the influencing factors of the autonomous ship maneuvering decision-making. (ii) Through the procedure of defuzzification, the fuzzy numbers are transformed into crisp numbers for priority ranking and comparison purpose. Therefore, the analysis of maritime traffic safety influencing factors for autonomous ship maneuvering decision-making can be conducted, thereby improving accuracy and rationality as well as expanding the application scope of the proposed model. (iii) both the weight of each expert and the weight of each influencing factor in the whole grey system are introduced, allowing to rank and compare the order of various influencing factors more reasonably and more accurately. Hence, the importance degree of each influencing factor and the preference of decision makers are comprehensively considered according to the actual situation. (iv) The simulator used in this research can simulate various actual navigational scenarios in different ports all over the world, combining with the actual operation data of experienced seafarers, thus, it can provide meaningful guidance for the selection/prioritization of the maritime traffic safety influencing factors of the autonomous ship maneuvering decisions and promote the development of autonomous ships.

  Although the proposed grey and fuzzy model is a promising model, this study still has some shortcomings as follows: in the specific experimental navigation scenario, as the above description and analysis in Section 3.4 of Chapter 3, the proposed model is rational and widely applicable to the analysis of the maritime traffic safety influencing factors for the ship maneuvering decisions. However, when in a specific navigational scenario, for instance, the influencing factors of longitude and latitude do not change correspondingly, there are still some shortcomings when adding the general expert knowledge using general common sense; in this case, the accuracy of our proposed model for analyzing these influencing factors is affected. Therefore, although the traditional grey theory has mainly been criticized because it treats different indexes (influencing factors) equally and takes no account of their relative importance, it does not fit with people's preferences for a specific index. It still has the accuracy and sensitivity in the specific experimental scenario for certain factors, so it is better to combine with the results from the traditional grey method when we apply the proposed model.

d) How can the maneuvering decision-making processes of experienced seafarers under the typical navigation scenario be modeled?

- ➢ This thesis collects the empirical data on the full-task handling simulation platform for large-scale ships named Navi-Trainer Professional 5000. A series of decision tree algorithms are utilized to study the decision-making mechanisms of different maneuvering behaviors in order to realize the automatic acquisition and representation of seafarers' decision-making in Chapters 4-6. In addition, a framework for the complete navigational decision tree method, containing the procedures of constructing the decision tree, pruning the decision tree, establishing manoeuvring decision classification rules, is designed. This thesis conducts preliminary exploratory research on the influencing factors on human-like decision-making theory for autonomous ships. It eliminates the limitations of objective conditions of cost, feasibility, and other factors for the traditional real-world merchant ship experiment and uses the operations from the experienced seafarers based on the simulator to conduct modeling research.

e) How to evaluate and maintain the proposed Human-like Decision-making Maneuvering Decision Recognition (HDMDR) model to ensure its appropriate functioning throughout its entire life cycle?

- ➢ The proposed HDMDR model in Chapter 5 is based on the C4.5 decision tree algorithm, which is one of the most famous and popular decision tree algorithms and the most influential data mining algorithm. In addition, in pruning the decision tree, we also utilized the post pruning method to eliminate branching anomalies caused by noise data and isolated points, thus overcoming the over-fitting problem in this model during its entire life cycle.

  Furthermore, we developed the optimized standardization principle of influencing factors to maintain the functioning of the model. The method of classifying the interval division that is proposed in this thesis fully considers the distribution of data sets and the endpoint extreme value within a reasonable range, such that all influencing factors data can be standardized accurately and scientifically and fully respect and restore the OOW's actual operation and decision-making for the inbound ship under typical scenarios on the simulator. Specifically, 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 influencing factors, then the corresponding row of the influencing factor data at the moment of saltatorial occurrence is deleted, and then the processed endpoint extreme value is selected. Meanwhile, to contain all the data in each influencing factor, the selected rules do not completely follow the rounding principle. When four digits are retained after the decimal point, the split points' open and closed intervals are determined according to the trade-off situation.

  Moreover, in the process of machine learning training of the model, approximately 80% of the data is randomly selected as the training set, and the remaining 20% is used as the test set. We utilized the test set to assess the accuracy and feasibility of the HDMDR model. In addition, we conducted a ten-fold cross-validation (10-CV) experiment using two classic classification algorithms: k-Nearest Neighbours (k-NN) and Support Vector Machine (SVM) to demonstrate further and evaluate the accuracy of the proposed method. Therefore, in the specific navigation scenario, our proposed HDMDR model has high accuracy and applicability. Overall, this can ensure the accuracy and maintain the functioning of our proposed model.

Questions on application

f) To what extent could the proposed models in this thesis be applied in reality?

- This thesis is exploration research on the influencing factors on human-like decision-making theory for the merchant ship in the real shipping industry (a 30,000-ton bulk carrier) under a specific navigation scenario. The simulation scenario is consistent with the actual port (the Shanghai Waigaoqiao wharf) environment according to the historical data at the same time, which has high value and application significance. Besides, the simulator can simulate various actual navigational scenarios using different types of merchant ships in different ports worldwide, combining with the actual operation data of the experienced seafarers (note that the corresponding various scenarios design with different target merchant ships' characteristics in specific experiments are needed when to apply the models more widely). Therefore, the model has a high commercial and promotional value. The stakeholders could design and develop the berthing system for autonomous ships based on the seafarer's operation rules obtained in the typical navigation scenario according to the empirical data from the full-task handling simulation platform. Additionally, the managers of ports and maritime authorities could also apply the optimal maneuvering decision-making rules (decision-making mechanisms) to enhance maritime traffic safety management. It is of great significance for the stakeholders to address the safety management and realize the intelligent decision-making of the merchant ships properly.

7.2 Recommendations for future research

In this section, the limitations of our research and several directions and challenging issues for future research are indicated.

a) Recommendation on improving and applying HDMDR model in practice

- The HDMDR model presented in Chapter 5 is a suitable initiative for processing existing environmental data, and it considers the analysis of current factors of maritime ship navigation in its current operational mode. The proposed model is established based on the experimental data collected from the designed navigational scenario (data-driven method). Most of these factors are relevant to consider in developing new intelligent agents supporting the navigation of autonomous ships. However, the factors covered by the model represent just a part of the components needed for the development of the agents.

  Therefore, it is necessary to collect more environmental factors which may affect the autonomous operation of ships to further improve and optimize the model (the training dataset *D* of the maneuvering factor X and environmental factors Y1~Y6 in Chapter 5;

new factors can be upgraded in the *Algorithm 5.1* using specific standardization principle; attribute *A*). When the proposed model is utilized in practice, depending on the specific environmental influencing factors in different scenarios, the factors should be selected according to the particular situation. Apart from setting the principle of standardization of environmental influencing factors to adjust the actual case, the proposed model could be upgraded and utilized to different operational scenarios.

In addition, further study should constantly optimize the practicality of the decision tree algorithm through continuous learning and training to improve the accuracy of the algorithm and promote the application of the proposed model in different navigation scenarios. Thus, the model will have broader applicability and recognition accuracy.

Moreover, achieving Degree Four: fully autonomous ship (IMO, 2021) is the final/future objective of this study. However, there are still have many challenges on the way to achieving fully autonomous shipping. In a specific scenario, the assumed ship can finally be the ship of Degree Four need to satisfy many conditions, such as the quality and quantity of data collection meet the requirements (from experienced OOW), the proposed model gets very high accuracy after a large amount of training, the port has complete intelligent infrastructure construction (could achieve port-autonomous ship interaction in real-time), and specific laws/regulations for autonomous ships have been formulated (some require amendments to current provisions for traditional human-operated ships, while others require the reconstruction of regulations), etc. Additionally, even though there are strong demands for autonomy shipping, human still plays an essential role from the first three degrees of autonomy based on the IMO's classification (First, at this stage, a pilot is mandatory in most ports worldwide. Port pilotage is a symbol of national sovereignty. Most countries implement compulsory pilotage to ensure the safety of ports, ships, and facilities. Furthermore, according to the suggestions from regulatory scoping for the use of maritime autonomous surface ships, it seems difficult to determine the most appropriate way at this stage, and some actions such as developing unified interpretation/regulations should be avoided to prevent creating confusion and contradiction for the traditional human-operated ships and management of stakeholders (IMO, 2021). That means, at the policy/regulations level, the development of autonomous ships in port scenarios will require humans' participation for a long time in the future. In addition, some sudden collision dangers still need the navigational assistance of experienced OOW, etc.).

b) Recommendation on improving the segmentation of intervals

- In Chapter 6, we have not elaborated on this classification interval division method. Besides, we mainly focus on guaranteeing that all the influencing factors data can be standardized to maximally respect and restore the seafarers' actual operation and decision-making for the typical inbound berthing scenarios. Furthermore, an experiment to show the accuracy difference between these two partitions is presented (the performance is shown in Table 6.3). A future study could conduct detailed research about the method that we proposed or use a standard approach for the segmentation of intervals of maneuvering decision-making attributes according to the actual navigation situation, which is more suitable for the real-world ship-handing orders (specifically the combined rudder orders and telegraph orders). Therefore, a comparative analysis can be conducted about this method, mainly expounding the data pre-processing process, accuracy and time-consuming analysis, and mathematical statistics mechanism.

c) Recommendation on studying the impact and characteristics of human behavior

- This thesis conducts exploratory research on the influencing factors on human-like decision-making theory for the autonomous ship. It addresses the cost limitations for the traditional real-world merchant ship experiment and uses the operation from the experienced seafarers based on the simulator to carry out modeling research. The proposed model is of considerable significance for the development of intelligent merchant ships. Moreover, it is unique and useful to obtain experimental data operated by an experienced senior crew on the full-task handling simulation platform at a specific time and space. Nevertheless, the waterborne transport and maneuver system will consist of autonomous ships and human-operated ships simultaneously for a very long period. In other words, the situation of a human-ship-environment interaction system with different autonomous levels will exist in various shipping scenarios for a long time. Therefore, a follow-up study could conduct a detailed research about human behavior and explore their performance and characteristics in specific navigational scenarios.

d) Recommendation on the berthing module that supports autonomous ship maneuvering decision-making in real-time

- In our case, the proposed models in Chapters 4-6 can reasonably complete the simulation and decision-making knowledge acquisition of the ship's automatic maneuvering process and has a high degree of application and promotion (the simulator, Navi-Trainer Professional 5000, which conforms to the IMO STCW78/10 convention and the Det Norske Veritas, has simulation scenarios for most ports worldwide). Moreover, as shown in Chapter 5, the upper limit of the quality confidence interval is used as the erroneous estimation under pessimistic conditions to maintain the accuracy and efficiency of our proposed model. The improvement of the algorithm is of great significance for the development of autonomous merchant ships.

  However, the decision tree cannot represent any feedback loop to inform the effect of the input action on a component and adapt its performance. It would be interesting to optimize our model for the autonomous operation of ships, decision classification rules and constantly optimize the practicality of our algorithm through continuous learning and training. Alternatively, reinforcement learning, deep learning, and transfer learning methods can be used to learn the human-like decision-making mechanism for autonomous merchant ships, which suits different navigational scenarios and proposes a model that supports autonomous ship maneuvering decision-making takes action in real-time.

e) Recommendation on considering the function of tugs during the berthing maneuvering

- In the thesis, we gave the hypothetical premise for our proposed model that the tugs are as a power plant system of target Own Ship (OS) 1 to facilitate the ship's overall situation of 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 by combining rudder orders and telegraph orders, according to the actual navigational situation of its force and movement.

  In our case, as the designed experimental scenario in Chapter 3, 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 to the ship berths docked at the end of the cable as a complete berthing process. The actual situation is that our data is collected in the initial boundary while the tug starts to work at the end of the whole inbound process of the ship OS1. This needs to be regarded as two different situations and modeling in different ways. In addition, taking the operations of ship OS1 and two tugs as the output of our proposed model into consideration at the same time is too complicated to establish the model.

Therefore, based on the idea of simplification, this thesis considers the tugs and the ship OS1 as a whole dynamic model (the tug is considered to be a power plant system of ship OS1, in the future intelligent port construction, unmanned is also a significant trend, the tug would also be unmanned or would be uniformly deployed by autonomous ships, as part of the overall dynamics model of the autonomous ships. Therefore, this assumption has a specific practical significance).

However, generally, in the berthing scenario, the tugs are always used in this kind of maneuver, and they significantly influence the decision model of the navigator. OOW/pilot uses propulsion and rudder very specifically (and indifferent way than maneuvering alone) in the presence of tugs, and engine orders are correlated highly with tugs orders. Moreover, the type and power of tugs play an essential role in maneuvering tactics. Future research could collect the relevant data from the tugs and add the data to the model analysis to further optimize our model and promote the application of our proposed models in different navigational scenarios.

Appendix

Table A.1 The extreme values of our data set.

Influencing factors	Standardization	
	$\Delta_i(\max)$	$\Delta_i(\min)$
Y1	10.75723437	0.000149400
Y2	9.286000215	2.97525E-05
Y3	6.670632875	0.000162331
Y4	4.939213846	0.000240429
Y5	2.677718534	0.001937135
Y6	2.607298241	0.002782460
Y7	4.896570329	4.70016E-05
Y8	6.238392243	0.000341300
Y9	5.742657263	0.000149654
Y10	2.699055325	4.80284E-05
Y11	6.230599999	0.000794324
Y12	8.023167652	0.000179697
Y13	45.23686934	0.001040272
Y14	37.19534450	0.010007617
Y15	36.16702220	0.006453297
Y16	**56.71438286**	0.005779491
Y17	26.88140323	0.001041084
Y18	26.76096695	0.029507153
Y19	25.52296248	0.005543666
Y20	31.57740192	0.041646945
Y21	6.406334561	3.47088E-05
Y22	4.576141174	0.000154554
Y23	4.212766847	0.000149660
Y24	5.285008067	0.000186862
Y25	13.21063113	7.98433E-05
Y26	24.45508796	0.001488166
Y27	6.267063219	0.000109524
Y28	10.38202823	9.73156E-05
Y29	12.12034909	6.66299E-05
Y30	8.602456594	0.000166826
Y31	6.267064612	0.000108035
Y32	3.862857951	**6.03501E-06**
Y33	4.661142861	2.04946E-05

Table A.2 The grey relational coefficient (partially).

Influencing factors	**Grey Relational Coefficient (Standardization)**								**Grey Grade**
	No. 1	**No. 2**	**No. 3**	**No. 4**	**No. 5**	**No. 6**	**…**	**No. 20534**	
Y1	0.948821333	0.944800559	0.935685389	0.933054446	0.940548179	0.945995493	…	0.936525683	0.963331321
Y2	0.944422670	0.941285170	0.934169836	0.932105721	0.937298072	0.941518452	…	0.930200250	0.963022501
Y3	0.980756009	0.981198262	0.982182514	0.982491117	0.978341825	0.977832987	…	0.985724706	0.964702382
Y4	0.984269772	0.984635896	0.985511601	0.985748630	0.981409613	0.980825559	…	0.926098511	0.964360060
Y5	0.958719718	0.958719718	0.958719718	0.958719718	0.955695626	0.955695626	…	0.915302540	0.962321061
Y6	0.940306134	0.940306134	0.940306134	0.940306134	0.937396914	0.937396914	…	0.915797346	0.962607649
Y7	0.977801220	0.977891825	0.977931508	0.977993022	0.974925396	0.974957017	…	0.941043104	0.964744459
Y8	0.971251135	0.971392017	0.971608734	0.971735174	0.974998194	0.975095592	…	0.973872465	0.967877544
…	…	…	…	…	…	…	…		…
Y28	0.966142965	0.966234013	0.966354440	0.966462437	0.963518894	0.963701914	…	0.941754820	0.964126732
Y29	0.966915878	0.966277561	0.966579003	0.966698608	0.963888008	0.964242401	…	0.951127489	0.965110499
Y30	0.970562344	0.972582036	0.973802135	0.958892072	0.950538172	0.968438806	…	0.952762326	0.961761784
Y31	0.993807833	0.993415906	0.992347753	0.991987132	0.988237248	0.987176821	…	0.968141097	0.963209766
Y32	0.902638993	0.902687368	0.902726071	0.902784132	0.900150231	0.900198339	…	0.919241512	0.955548915
Y33	0.941088142	0.941467051	0.941775139	0.942178328	0.939634966	0.939989087	…	0.977156185	0.962805458

Table A.3 The linguistic terms from the experts and the crisp number and weights of different maneuvering influencing factors.

Influencing factors	**Expert No. 1**		**Expert No. 2**		**Expert No. 3**		**Expert No. 4**		**Crisp numbers (with β_i)**	**Weights (λ_k)**
	Linguistic terms	**Crisp numbers**	**Linguistic terms**	**Crisp numbers**	**Linguistic terms**	**Crisp numbers**	**Linguistic terms**	**Crisp numbers**		
Y1	M	0.4733	M	0.4733	H	0.7333	M	0.4733	0.5253	0.0231
Y2	H	0.7333	M	0.4733	H	0.7333	H	0.7333	0.6683	0.0294
Y3	H	0.7333	H	0.7333	H	0.7333	H	0.7333	0.7333	0.0323
Y4	H	0.7333	M	0.4733	H	0.7333	M	0.4733	0.6033	0.0266
Y5	M	0.4733	M	0.4733	M	0.4733	H	0.7333	0.5383	0.0237
Y6	M	0.4733	M	0.4733	H	0.7333	M	0.4733	0.5253	0.0231
Y7	VH	0.9400	H	0.7333	H	0.7333	VH	0.9400	0.8470	0.0373
Y8	VH	0.9400	VH	0.9400	VH	0.9400	VH	0.9400	0.9400	0.0414
Y9	VH	0.9400	VH	0.9400	VH	0.9400	H	0.7333	0.8883	0.0391
Y10	VH	0.9400	VH	0.9400	VH	0.9400	VH	0.9400	0.9400	0.0414
Y11	H	0.7333	VH	0.9400	H	0.7333	H	0.7333	0.7850	0.0346
Y12	M	0.4733	L	0.2800	VL	0.0833	L	0.2800	0.2987	0.0132
Y13	VH	0.9400	VH	0.9400	VH	0.9400	H	0.7333	0.8883	0.0391
Y14	VH	0.9400	VH	0.9400	VH	0.9400	H	0.7333	0.8883	0.0391
Y15	VH	0.9400	VH	0.9400	VH	0.9400	VH	0.9400	0.9400	0.0414
Y16	H	0.7333	H	0.7333	VH	0.9400	H	0.7333	0.7746	0.0341
Y17	VH	0.9400	VH	0.9400	H	0.7333	VH	0.9400	0.8987	0.0396
Y18	VH	0.9400	VH	0.9400	H	0.7333	VH	0.9400	0.8987	0.0396
Y19	VH	0.9400	VH	0.9400	VH	0.9400	VH	0.9400	0.9400	0.0414
Y20	H	0.7333	H	0.7333	M	0.4733	H	0.7333	0.6813	0.0300
Y21	VL	0.0833	L	0.2800	L	0.2800	VL	0.0833	0.1718	0.0076
Y22	L	0.2800	VL	0.0833	L	0.2800	VL	0.0833	0.1817	0.0080
Y23	H	0.7333	VH	0.9400	H	0.7333	VH	0.9400	0.8367	0.0369
Y24	H	0.7333	VH	0.9400	H	0.7333	VH	0.9400	0.8367	0.0369
Y25	H	0.7333	H	0.7333	M	0.4733	H	0.7333	0.6813	0.0300
Y26	H	0.7333	VH	0.9400	H	0.7333	VH	0.9400	0.8367	0.0369
Y27	H	0.7333	H	0.7333	M	0.4733	H	0.7333	0.6813	0.0300
Y28	H	0.7333	H	0.7333	M	0.4733	H	0.7333	0.6813	0.0300
Y29	H	0.7333	H	0.7333	H	0.7333	H	0.7333	0.7333	0.0323
Y30	M	0.4733	M	0.4733	M	0.4733	L	0.2800	0.4250	0.0187
Y31	H	0.7333	M	0.4733	M	0.4733	H	0.7333	0.6163	0.0272
Y32	M	0.4733	L	0.2800	L	0.2800	M	0.4733	0.3863	0.0170
Y33	M	0.4733	L	0.2800	M	0.4733	M	0.4733	0.4250	0.0187
Weights (β_i)	-	0.30	-	0.25	-	0.20	-	0.25	-	**Sum=1**

References

Abbassi, R., Khan, F., Khakzad, N., Veitch, B., Ehlers, S., 2017. Risk analysis of offshore transportation accident in arctic waters. Transactions of the Royal Institution of Naval Architects Part A: International Journal of Maritime Engineering 159, 213-223.

Abdussamie, N., Daboos, M., Elferjani, I., Chai, S.H., Alaktiwi, A., 2018a. Risk assessment of LNG and FLNG vessels during manoeuvring in open sea. Journal of Ocean Engineering and Science 3, 56-66.

Abdussamie, N., Ojeda, R., Daboos, M., 2018b. ANFIS method for ultimate strength prediction of unstiffened plates with pitting corrosion. Ships and Offshore Structures 13, 540-550.

Abdussamie, N., Zaghwan, A., Daboos, M., Elferjani, I., Mehanna, A., Su, W., 2018c. Operational risk assessment of offshore transport barges. Ocean Engineering 156, 333-346.

AGCS, 2018. An annual review of trends and developments in shipping losses and safety. Allianz Global Corporate and Specialty SE. Retrieved from: https://www.agcs.allianz.com/assets/PDFs/Reports/AGCS_Safety_Shipping_Review_2018.pdf. (Accessed October 17, 2019).

Aguiar-Pulido, V., A Seoane, J., Gestal, M., Dorado, J., 2013. Exploring patterns of epigenetic information with data mining techniques. Current Pharmaceutical Design 19, 779-789.

Akyuz, E., Akgun, I., Celik, M., 2016. A fuzzy failure mode and effects approach to analyse concentrated inspection campaigns on board ships. Maritime Policy & Management 43, 887-908.

Akyuz, E., Celik, M., 2014. A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. Safety Science 68, 169-179.

Angluin, D., 1988. Queries and concept learning. Machine Learning 2, 319-342.

Aziz, A., Ahmed, S., Khan, F., Stack, C., Lind, A., 2019. Operational risk assessment model for marine vessels. Reliability Engineering & System Safety 185, 348-361.

Baksh, A.-A., Abbassi, R., Garaniya, V., Khan, F., 2018. Marine transportation risk assessment using Bayesian Network: Application to Arctic waters. Ocean Engineering 159, 422-436.

Balin, A., Demirel, H., Celik, E., Alarcin, F., 2018. A fuzzy dematel model proposal for the cause and effect of the fault occurring in the auxiliary systems of the ships' main engine. International Journal of Maritime Engineering 160, A141-A153.

Balmat, J.-F., Lafont, F., Maifret, R., Pessel, N., 2011. A decision-making system to maritime risk assessment. Ocean Engineering 38, 171-176.

Bhattacharyya, S., Rajesh, G., Gupta, D.K., 2011. Fuzzy autopilot for ship maneuvering. International Shipbuilding Progress 58, 191-218.

Borak, J.S., 1999. Feature selection and land cover classification of a MODIS-like data set for a semiarid environment. International Journal of Remote Sensing 20, 919-938.

Braae, M., Rutherford, D., 1978. Fuzzy relations in a control setting. Kybernetes 7, 185-188.

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.

Celik, M., Lavasani, S.M., Wang, J., 2010. A risk-based modelling approach to enhance shipping accident investigation. Safety Science 48, 18-27.

Chang, P.-C., Fan, C.-Y., Dzan, W.-Y., 2010. A CBR-based fuzzy decision tree approach for database classification. Expert Systems with Applications 37, 214-225.

Chen, C.N., Ting, S.C., 2002. A study using the grey system theory to evaluate the importance of various service quality factors. International Journal of Quality & Reliability Management 19, 838-861.

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, 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., Ran, B., 2017. Vehicle Behavior Learning via Sparse Reconstruction with l_2-l_p Minimization and Trajectory Similarity. IEEE Transactions on Intelligent Transportation Systems 18, 236-247.

Cherfi, A., Nouira, K., Ferchichi, A., 2018. Very Fast C4.5 Decision Tree Algorithm. Applied Artificial Intelligence 32, 119-137.

Chou, T.-Y., Liang, G.-S., 2010. Application of a fuzzy multi-criteria decision-making model for shipping company performance evaluation. Maritime Policy & Management 28, 375-392.

Cohen, S., Rokach, L., Maimon, O., 2007. Decision-tree instance-space decomposition with grouped gain-ratio. Information Sciences 177, 3592-3612.

Dai, W., Ji, W., 2014. A MapReduce Implementation of C4.5 Decision Tree Algorithm. International Journal of Database Theory and Application 7, 49-60.

Deng, J., 1982. Control problems of grey systems. Systems & Control Letters 1, 288-294.

Deng, J., 1989. Introduction to grey system theory. The Journal of grey system 1, 1-24.

Deng, J.L., 1990. Grey system theory course. Huazhong University of Science and Technology Press: Wuhan, China, 55-56.

Denœux, T., 1995. A k-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory. IEEE Transactions on Systems Man & Cybernetics 25, 804-813.

Esposito, F., Malerba, D., Semeraro, G., Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE Transactions on Pattern Analysis and MachineIntelligence 19, 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, 632-644.

Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. Belmont, CA: Wadsworth & Brooks.

Fu, B., Liu, W., Jin, Y., Yang, B., 2017. Evaluation on Crew's Information Processing Capability Based on Grey Relational Analysis.

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. Computers in Biology and Medicine 59, 125-133.

Goerlandt, F., Montewka, J., 2015. Maritime transportation risk analysis: Review and analysis in light of some foundational issues. Reliability Engineering & System

Safety 138, 115-134.

Goerlandt, F., Montewka, J., Kuzmin, V., Kujala, P., 2015. A risk-informed ship collision alert system: Framework and application. Safety Science 77, 182-204.

Guneri, A.F., Cengiz, M., Seker, S., 2009. A fuzzy ANP approach to shipyard location selection. Expert Systems with Applications 36, 7992-7999.

Gupta, D.K., Vasudev, K.L., Bhattacharyya, S.K., 2018. Genetic algorithm optimization based nonlinear ship maneuvering control. Applied Ocean Research 74, 142-153.

Hadi-Vencheh, A., Mokhtarian, M., 2011. A new fuzzy MCDM approach based on centroid of fuzzy numbers. Expert Systems with Applications 38, 5226-5230.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software:an update. Acm Sigkdd Explorations Newsletter 11, 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. Journal of Maritime Research 5, 3-18.

Hao, Y., Hu, S., Chen, X., 2017. Grey relational analysis method for human error identification based on the MAIB accident reports. Journal of Safety & Environment.

Hatefi, S.M., Tamošaitienė, J., 2018. Construction Projects Assessment Based on the Sustainable Development Criteria by an Integrated Fuzzy AHP and Improved GRA Model. Sustainability 10, 991.

Heckerman, D., 1998. A tutorial on learning with Bayesian networks, Learning in graphical models. Springer, pp. 301-354.

Hssina, B., Merbouha, A., Ezzikouri, H., Erritali, M., 2014. A comparative study of decision tree ID3 and C4. 5. International Journal of Advanced Computer Science and Applications 4.

Huang, L., Hou, J., Liu, Y., Guo, Y., 2013. Grey Entropy Relation Algorithm of Choosing the Optimum Diagnostic Nodes in Analogue Circuits. Chinese Journal of Electronics 22, 615-620.

Hunt, E.B., Marin, J., Stone, P.J., 1966. Experiments in induction. Oxford, England: Academic Press.

ICS, 2020. The International Chamber of Shipping Annual Review. Retrieved from: https://www.ics-shipping.org/wp-content/uploads/2020/11/annual-review-2020-final-compressed.pdf. (Accessed May 20, 2021).

IMO, 2018. Energy Efficiency Measures. Retrieved from: http://www.imo.org/en/ourwork/environment/pollutionprevention/airpollution/pa ges/technical-and-operational-measures.aspx. (Accessed January 3, 2019).

IMO, 2021. Outcome of the Regulatory Scoping Exercise for the use of Maritime Autonomous Surface Ships. Retrieved from: https://wwwcdn.imo.org/localresources/en/MediaCentre/PressBriefings/Documents/MSC.1-Circ.1638.pdf. (Accessed December 20, 2021).

Ishikawa, A., Amagasa, M., Shiga, T., Tomizawa, G., Tatsuta, R., Mieno, H., 1993. The max-min Delphi method and fuzzy Delphi method via fuzzy integration. Fuzzy Sets and Systems 55, 241-253.

Janikow, C.Z., 1998. Fuzzy decision trees: issues and methods. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society 28, 1-14.

Javed, S.A., Liu, S., 2017. Predicting the research output/growth of selected countries: application of Even GM (1, 1) and NDGM models. Scientometrics 115, 395-413.

Jin, F., Liu, P., 2010. The multi-attribute group decision making method based on the interval grey linguistic variables. African Journal of Business Management 4, 3708-3715.

Kantardzic, M., 2011. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.

Kass, G.V., 1975. Significance testing in automatic interaction detection (AID). Applied Statistics, 178-189.

Kennedy, Q., Taylor, J.L., Reade, G., Yesavage, J.A., 2010. Age and expertise effects in aviation decision making and flight control in a flight simulator. Aviation, Space, and Environmental Medicine 81, 489-497.

Khan, B., Khan, F., Veitch, B., Yang, M., 2018. An operational risk analysis tool to analyze marine transportation in Arctic waters. Reliability Engineering & System Safety 169, 485-502.

Kheradpisheh, S.R., Ganjtabesh, M., Thorpe, S.J., Masquelier, T., 2018. STDP-based spiking deep convolutional neural networks for object recognition. Neural Networks 99, 56-67.

Kim, S.-W., Kim, J.-H., Seo, M.-G., Choi, J.-W., Lee, Y.-B., Han, S.-K., 2017. Assessment of Ship Operating Performance by Using Full Scale Measurement. International Society of Offshore and Polar Engineers.

Kose, E., Gosine, R.G., Dunwoody, A.B., Calisal, S.M., 1995. An expert-system for monitoring dynamic stability of small craft. IEEE Journal of Oceanic Engineering 20, 13-22.

Kumar, A.M., Rajakarunakaran, S., Pitchipoo, P., Vimalesan, R., 2018. Fuzzy based risk prioritisation in an auto LPG dispensing station. Safety Science 101, 231-247.

Lakkakula, N.P., Naidu, M.M., Reddy, K.K., 2014. An Entropy Based Elegant Decision Tree Classifier to Predict Precipitation. 11-19.

Lazarowska, A., 2017. A new deterministic approach in a decision support system for ship's trajectory planning. Expert Systems with Applications 71, 469-478.

Lee, C.-C., 1990. Fuzzy logic in control systems: fuzzy logic controller. I. IEEE Transactions on Systems, Man, and Cybernetics 20, 404-418.

Lee, G., Surendran, S., Kim, S.-H., 2009. Algorithms to control the moving ship during harbour entry. Applied Mathematical Modelling 33, 2474-2490.

Lee, P.T.-W., Lin, C.-W., Shin, S.-H., 2018. Financial Performance Evaluation of Shipping Companies Using Entropy and Grey Relation Analysis, Multi-Criteria Decision Making in Maritime Studies and Logistics. Springer, Switzerland, pp. 219-247.

Li, X., 1996. Research on the Computation Model of Grey Interconnet Degree. Systems Engineering.

Li, X., Zheng, W., Yin, L., Yin, Z., Song, L., Tian, X., 2017. Influence of Social-economic Activities on Air Pollutants in Beijing, China. Open Geosciences 9.

Liang, J., Zhiqiang, S., Long, C., 2012. Dynamic neural network-based integrated learning algorithm for driver behavior. Journal of Transportation Systems Engineering and Information Technology 12, 34-40.

Lilly Mercy, J., Prakash, S., Krishnamoorthy, A., Ramesh, S., Alex Anand, D., 2017. Multi response optimisation of mechanical properties in self-healing glass fiber reinforced plastic using grey relational analysis. Measurement 110, 344-355.

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. Machine Learning 40, 203-228.

Lin, C.-T., Wang, M.-J.J., 1997. Hybrid fault tree analysis using fuzzy sets. Reliability Engineering & System Safety 58, 205-213.

Liu, P., Chu, Y., Li, Y., 2015. The multi-attribute group decision-making method based on the interval grey uncertain linguistic generalized hybrid averaging operator. Neural Computing and Applications 26, 1395-1405.

Liu, S., Forrest, J., Yang, Y., 2010. Advances in Grey Systems Research. Springer Berlin Heidelberg.

Liu, S., Forrest, J.Y.L., 2010. Grey systems: theory and applications. Springer.

Lu, M., Wevers, K., Wang, J., 2010. Implementation Road Map for In-Vehicle Safety Systems in China. Transportation Research Record Journal of the Transportation Research Board 2148, 116-123.

Lu, Z., Wu, X., Bongard, J.C., 2015. Active Learning through Adaptive Heterogeneous Ensembling. IEEE Transactions on Knowledge and Data Engineering 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.

Maimon, O.Z., Rokach, L., 2014. Data mining with decision trees: theory and applications. World scientific.

Marken, V.B., Ehlers, S., Khan, F., 2015. Delay risk analysis of ship sailing the northern sea route. Ship Technology Research 62, 26-35.

Martınez, A.M., Webb, G.I., Chen, S., Zaidi, N.A., 2016. Scalable learning of Bayesian network classifiers. Journal of Machine Learning Research 17, 1-35.

Moradi, M., Keyvanpour, M.R., 2015. An analytical review of XML association rules mining. Artificial Intelligence Review 43, 277-300.

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. International Journal of Remote Sensing 21, 1115-1138.

Novák, V., Perfilieva, I., Mockor, J., 2012. Mathematical principles of fuzzy logic. Springer Science & Business Media.

Olaru, C., Wehenkel, L., 2003. A complete fuzzy decision tree technique. Fuzzy Sets and Systems 138, 221-254.

Page, L.B., Perry, J.E., 1994. Standard deviation as an alternative to fuzziness in fault tree models. IEEE Transactions on Reliability 43, 402-407.

Pandya, R., Pandya, J., 2015. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. International Journal of Computer Applications 117, 18-21.

Paschalidis, E., Choudhury, C.F., Hess, S., 2018. Modelling the effects of stress on gap-acceptance decisions combining data from driving simulator and physiological sensors. Transportation Research Part F: Traffic Psychology and Behaviour 59, 418-435.

Patel, A.V., Mohan, B., 2002. Some numerical aspects of center of area defuzzification method. Fuzzy sets and systems 132, 401-409.

Pazouki, K., Forbes, N., Norman, R.A., Woodward, M.D., 2018. Investigation on the impact of human-automation interaction in maritime operations. Ocean Engineering 153, 297-304.

Pedrycz, W., 1994. Why triangular membership functions? Fuzzy sets and systems 64, 21-30.

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. Transportation Research Record Journal of the Transportation Research Board 2503, 29-38.

Perera, L.P., Carvalho, J.P., Guedes Soares, C., 2010. Fuzzy logic based decision making system for collision avoidance of ocean navigation under critical collision conditions. Journal of Marine Science and Technology 16, 84-99.

Perera, L.P., Ferrari, V., Santos, F.P., Hinostroza, M.A., Guedes Soares, C., 2015a. Experimental Evaluations on Ship Autonomous Navigation and Collision Avoidance by Intelligent Guidance. IEEE Journal of Oceanic Engineering 40, 374-387.

Perera, L.P., Ferrari, V., Santos, F.P., Hinostroza, M.A., Soares, C.G., 2015b. Experimental evaluations on ship autonomous navigation and collision avoidance by intelligent guidance. IEEE Journal of Oceanic Engineering 40, 374-387.

Pietrzykowski, Z., Wołejsza, P., Borkowski, P., 2017. Decision support in collision situations at sea. The Journal of Navigation 70, 447-464.

Prasad, N., Kumar, P., Mm, N., 2013. An Approach to Prediction of Precipitation Using Gini Index in SLIQ Decision Tree. IEEE, pp. 56-60.

Prasad, N., Naidu, M.M., 2013. Gain Ratio as Attribute Selection Measure in Elegant Decision Tree to Predict Precipitation. 141-150.

Proske, D., Van Gelder, P., 2006. Prediction of complex systems using grey models. Engineering Conferences International on Geohazards.

Provost, F., Domingos, P., 2003. Tree induction for probability-based ranking. Machine Learning 52, 199-215.

Pulkkinen, P., Koivisto, H., 2008. Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. International Journal of Approximate Reasoning 48, 526-543.

Quinlan, J., 1979. Interactive Dichotomizer, ID3. Eds. Morgan Kauffmann, Springer-Verlag.

Quinlan, J.R., 1986. Induction of decision trees. Machine Learning 1, 81-106.

Quinlan, J.R., 1993. C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco.

Quinlan, J.R., 2014. C4. 5: programs for machine learning. Elsevier.

Rahman, M.S., Khan, F., Shaikh, A., Ahmed, S., Imtiaz, S., 2019. Development of risk model for marine logistics support to offshore oil and gas operations in remote and harsh environments. Ocean Engineering, 125-134.

Rajesh, S., Rajakarunakaran, S., Sudhkarapandian, R., 2013. Optimization of the red mud–aluminum composite in the turning process by the Grey relational analysis with entropy. Journal of Composite Materials 48, 2097-2105.

Reumers, S., Liu, F., Janssens, D., Cools, M., Wets, G., 2013. Semantic Annotation of Global Positioning System Traces. Transportation Research Record: Journal of the Transportation Research 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. Journal of Infection 72, 143-151.

Rojas, R., 1996. Neural Networks-A Systematic Introduction Springer-Verlag. New York.

Rolls-Royce, 2018. The next steps of remote and autonomous ships. Rolls-Royce Holdings plc. Retrieved from: https://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf. (Accessed April 16, 2020).

Rothblum, A.M., 2000. Human error and marine safety, National Safety Council Congress and Expo, Orlando, p. 7.

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: Journal of the Transportation Research Board 2237, 41-50.

Senol, Y.E., Sahin, B., 2016. A novel Real-Time Continuous Fuzzy Fault Tree Analysis (RC-FFTA) model for dynamic environment. Ocean Engineering 127, 70-81.

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379-423.

Shen, D.-H., Du, J.-C., 2005. Application of Gray Relational Analysis to Evaluate HMA with Reclaimed Building Materials. Journal of Materials in Civil Engineering 17, 400-406.

Sugeno, M., 1999. Fuzzy modelling and control. CRC Press, Florida, USA.

Sun, J., Wang, X., 2006. A Comparative Analysis of Rule Simplification and Pruning Fuzzy Decision Trees. Computer Engineering 32, 210-211.

Surendran, S., Kiran, V., 2007. Control of ship roll motion by active fins using fuzzy logic. Ships and Offshore Structures 2, 11-20.

Tsang, I.W., Kwok, J.T., Cheung, P.-M., Cristianini, N., 2005. Core vector machines: Fast SVM training on very large data sets. Journal of Machine Learning Research 6.

Tseng, P.H., Cullinane, K., 2018. Key criteria influencing the choice of Arctic shipping: a fuzzy analytic hierarchy process model. Maritime Policy & Management 45, 422-438.

Ugurlu, O., Yildirim, U., Basar, E., 2015. Analysis of grounding accidents caused by human error. Journal of Marine Science and Technology-Taiwan 23, 748-760.

Umanol, M., Okamoto, H., Hatono, I., Tamura, H., Kawachi, F., Umedzu, S., Kinoshita, J., 1994. Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems. IEEE, pp. 2113-2118.

Ung, S.-T., 2015. A weighted CREAM model for maritime human reliability analysis. Safety Science 72, 144-152.

V, N.P.L., Patro, K.R., Naidu, M.M., 2013. A Gini Index Based Elegant Decision Tree Classifier to Predict Precipitation. IEEE, pp. 46-54.

Wang, J., Li, M., Liu, Y., Zhang, H., Zou, W., Cheng, L., 2014a. Safety assessment of shipping routes in the South China Sea based on the fuzzy analytic hierarchy process. Safety Science 62, 46-57.

Wang, K.C.P., Li, Q., Hall, K.D., Elliott, R.P., 2007. Experimentation with Gray Theory for Pavement Smoothness Prediction. Biulleten Eksperimentalnoĭ Biologii I Meditsiny 1990, 3-13.

Wang, X., Liu, X., Pedrycz, W., Zhang, L., 2015. Fuzzy rule based decision trees. Pattern Recognition 48, 50-59.

Wang, X., Yang, X., 2008. Study on decision mechanism of driving behavior based on decision tree. Journal of System Simulation 20, 415-420.

Wang, X.-W., Jiang, Y.-M., 2011. Analysis and improvement of ID3 decision tree algorithm. Computer Engineering and Design 9.

Wang, X.-Z., Zhai, J.-H., Lu, S.-X., 2008. Induction of multiple fuzzy decision trees based on rough set technique. Information Sciences 178, 3188-3202.

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, 2691-2637.

Wang, Y.-M., Chin, K.-S., Poon, G.K.K., Yang, J.-B., 2009. Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean. Expert Systems with Applications 36, 1195-1207.

Wang, Z., Wang, Q., Ai, T., 2014b. 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.

Wei, Y., Wang, A., Han, H., 2019. Ocean wave active compensation analysis of inverse kinematics for hybrid boarding system based on fuzzy algorithm. Ocean Engineering 182, 577-583.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Wu, B., Yip, T.L., Xie, L., Wang, Y., 2018. A fuzzy-MADM based approach for site selection of offshore wind farm in busy waterways in China. Ocean Engineering 168, 121-132.

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. Knowledge and Information Systems 14, 1-37.

Xiao, F., Ligteringen, H., van Gulijk, C., Ale, B., 2015. Comparison study on AIS data of ship traffic behavior. Ocean Engineering 95, 84-93.

Xie, C., Lu, J., Parkany, E., 2003. Work travel mode choice modeling with data mining: decision trees and neural networks. Transportation Research Record: Journal of the Transportation Research 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 Systems with Applications 115, 172-188.

Yang, Z., Wang, J., 2015. Use of fuzzy risk assessment in FMEA of offshore engineering systems. Ocean Engineering 95, 195-204.

Yang, Z.L., Wang, J., Bonsall, S., Fang, Q.G., 2009. Use of Fuzzy Evidential Reasoning in Maritime Security Assessment. Risk analysis 29, 95-120.

Yuan, Y., Shaw, M.J., 1995. Induction of fuzzy decision trees. Fuzzy sets and systems 69, 125-139.

Zadeh, L.A., 1965. Fuzzy sets. Information and control 8, 338-353.

Zadeh, L.A., 1983. The role of fuzzy logic in the management of uncertainty in expert systems. Fuzzy sets and systems 11, 199-227.

Zhang, L.-Y., Peng, Z.-R., Sun, D., Liu, X., 2012. Rule-Based Forecasting of Traffic Flow for Large-Scale Road Networks. Transportation Research Record: Journal of the Transportation Research Board 2279, 3-11.

Zhang, Q., Guo, X., Deng, J., 1996. Grey relation entropy method of grey relation analysis. Systems Engineering Theory & Practice 8, 7-11.

Zhang, X., Zhang, J., Fan, C., Meng, J., Wang, J., Wan, Y., 2018. Analysis of Dynamic Characteristics of the 21st Century Maritime Silk Road. Journal of Ocean University of China 17, 487-497.

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, Applied Mechanics and Materials. Trans Tech Publ, pp. 256-260.

Zhou, Q., Wong, Y.D., Loh, H.S., Yuen, K.F., 2018. A fuzzy and Bayesian network CREAM model for human reliability analysis – The case of tanker shipping. Safety Science 105, 149-157.

Zhou, Q., Wong, Y.D., Xu, H., Van Thai, V., Loh, H.S., Yuen, K.F., 2017. An enhanced CREAM with stakeholder-graded protocols for tanker shipping safety application. Safety Science 95, 140-147.

Zhou, Q.J., Thai, V.V., 2016. Fuzzy and grey theories in failure mode and effect analysis for tanker equipment failure prediction. Safety Science 83, 74-79.

Summary

Maritime shipping is essential to the global economy, while waterway transportation is recognized as a high-risk industry. Additionally, maritime accidents are frequently caused by human errors, and with the rapid improvement of science and technology, the improvement of autonomous ships has been technically feasible, which attracts the wide attention of researchers in academia and industry. However, the knowledge acquisition and representation methods are mainly based on knowledge-based research methods, while the existing research for automatically achieving the autonomous ships' maneuvering decision-making by acquiring the seafarers' operation characteristics is still scanty. In addition, it also lacks the appropriate theoretical methods to explore the problem of autonomous ship human-like maneuvering decision-making modeling. Therefore, the research on ship maneuvering decision-making methods still needs to be improved and further developed.

This thesis focuses on the problem of modeling seafarers' navigational decision-making in a typical scenario for autonomous ships' safety. We propose the method to prioritize safety influencing factors of autonomous ships' maneuvering decisions and a series of ship maneuvering knowledge learning models to give the autonomous ship the ability to make decisions like a human. The autonomous ship human-like maneuvering decision-making problem has been considered as a machine learning problem, and we translate the problem into learning the maneuvering decision characteristics of the officer on watch (OOW) using various decision tree algorithms. By constructing autonomous ship human-like decision-making maneuvering decision recognition models under multiple constraints in the specific scenarios, the decision-making mechanism of the OOW's maneuvering behavior under specific water traffic safety influencing factors in the inbound scenario is analyzed, and the OOW's decision-making knowledge is automatically acquired and represented.

Maritime traffic safety influencing factors prioritizing model of autonomous ships'

maneuvering decisions

Ship maneuvering decisions are influenced by several factors. Meanwhile, the autonomous ship maneuvering decision-making influencing factors constitute a typical grey system suitable for research by grey relational analysis. Furthermore, in the fuzzy approach, linguistic assessment of factors is evaluated to obtain priorities numbers. Therefore, this thesis proposes a maritime traffic safety influencing factors prioritizing model, which utilized the grey and fuzzy theories and combined with expert linguistic terms, thus to select the ship maneuvering decision-making main influencing factors from multi-source influencing factors (in overall and separated categories of natural environment, ship motion, force parameters, draft, and position), and to study the decision-making prioritization for maritime traffic safety for specific ship maneuvering scenarios. This proposed model can prioritize the main factors that affect maneuvering decisions and guide an autonomous ship-assisted or automatic maneuvering evaluation system for the research of human-like maneuvering behavior. This study provides a new perspective on identifying primary ship maneuvering decision-making influencing factors in theory and practice. It can be utilized for better decision-making concerning maritime traffic safety of autonomous ship maneuvering, making shipping safer, and promoting the application and spreading of autonomous ships.

Navigational decision tree model of human-like decision-making for autonomous ships' maneuvering

Based on the advantages of the Iterative Dichotomiser 3 (ID3) and C4.5 algorithms and the ability to analyze the characteristics of multifork trees, a framework for the complete navigational decision tree method, containing the procedures of constructing the decision tree, pruning the decision tree, establishing maneuvering decision classification rules, is designed. Additionally, the novel navigational decision tree algorithms are utilized to recognize the OOW's maneuvering decision characteristics. In this thesis, the autonomous ship human-like maneuvering decision-making problem is regarded as a machine learning problem based on the OOW's experience, the OOW's actual maneuvering data, and the environment influencing factors, such as wind, wave, and current in specific water areas, and the problem is converted using the decision tree methods to learn the OOW's maneuvering decision-making characteristics, thus constructing a human-like decision-making model under multiple constraints. The proposed model could be applied to realize the automatic acquisition and representation of the OOW's decision-making knowledge in inbound merchant ships analysis. Moreover, to verify the model's performance, the case study based on this method is conducted in the Waigaoqiao Phase IV Port of Shanghai. 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 model.

Fuzzy decision tree model of human-like decision-making for autonomous ships'

maneuvering

The classification results of a decision tree are clear and easy to understanding; while it cannot address potential uncertainty during the classification process, the resulting decision tree is generally not robust. However, the Fuzzy Decision Tree (FDT) integrates the advantages of fuzzy theory and decision trees by combining the comprehensibility of decision trees and the comprehensive expressions of fuzzy technology, which could address problems of fuzziness and uncertainty. The FDT has strong decision-making abilities and can address the problems of ambiguity and uncertainty. Because of this, in this thesis, we collect data on the full-task handling simulation platform for large-scale ships and use an improved fuzzy ID3 decision tree to explore the decision-making mechanisms of different maneuvering behaviors to realize the automatic acquisition and representation of a seafarer's decision-making. The significance level α and truth level threshold β are utilized to control tree generation and carry out pre-pruning. The simulation results combined with the method of classification interval division indicate that the proposed method can identify the seafarers' maneuvering operations and characteristics accurately and has high reasoning efficiency.

Insights into autonomous berthing system design and port safety management

With the continuous development of large-scale, high-speed, and professional ships and the increasing construction of modern intelligent deep-water ports, inbound merchant ships' safety is increasingly important. Thus, it is of great significance for the stakeholders to adequately address the safety management and realize the intelligent decision-making of the inbound merchant ships. In this thesis, based on the actual seafarers' operational data from the full-task handling simulation platform, this study combines the navigation scenario of a 30,000-ton bulk carrier ship inbound port to propose a series of knowledge acquisition models under multiple constraints to give autonomous ships the ability to make decisions like a human. The experimental results indicate that the maneuvering decision recognition models can accurately and scientifically standardize the boundary of the interval of influencing factor data and identify current maneuvering behavior. The proposed methods and the evaluation results provide valuable insights for effective safety management of the inbound merchant ships for the stakeholders and managers of the port. In addition, the thesis also provides theoretical guidance and a feasibility basis for research into human-like piloting behavior and the realization of autonomous ship piloting and berthing systems.

In summary, this thesis focuses on the concept of human-like maneuvering for autonomous merchant ships and studies the human-like decision-making mechanism for autonomous merchant ships. We propose the autonomous ship human-like decision-making recognition models. Specifically, by establishing the autonomous learning method of maneuvering decision-making, the maneuvering decision-making rules of the *typical maneuvering style in the specific scenario are explored, and the processes of autonomous learning seafarer's maneuvering decision-making characteristics for autonomous ships are analyzed. This study provides a new perspective and methodology for developing autonomous ship maneuvering decision-making technology in theory and practice, promotes the application and spreading of autonomous merchant ships, and is conducive to the development of water transportation in the direction of safety, sustainability, and economy.*

Curriculum vitae

Faculty of Technology, Policy and Management
Delft University of Technology
2628CD, Delft, The Netherlands

Tel: +31-621596821
Fax: +31-15-2782502
E-mail: J.Xue@tudelft.nl

Personal information

- Name: Jie Xue (Chinese: 薛杰)
- M/F: Male
- Nationality: P.R. China
- Birth: April 5th, 1989

Education

09/2009-06/2013 B.Eng. in Navigation Technology
Wuhan University of Technology

03/2011-06/2013 B.B.A. in Business Administration
Wuhan University

09/2014-10/2017 Major in Traffic Information Engineering and Control
(successive postgraduate and doctoral program)
Wuhan University of Technology

11/2017- present PhD candidate in Safety and Security Science Group
Delft University of Technology

Research interests

Modeling and Optimization, Machine Learning, Decision Making, Intelligent Transportation Systems, Maritime Safety, Autonomous Ships, Accident Analysis

Publications

- **Journal papers**

[1] **Jie Xue**, Tsz Leung Yip, Bing Wu, Chaozhong Wu, P.H.A.J.M. van Gelder. 2021. A novel fuzzy Bayesian network-based MADM model for offshore wind turbine selection in busy waterways: An application to a case in China. *Renewable Energy*. 172, 897-917.

[2] **Jie Xue**, Eleonora Papadimitriou, Reniers Genserik, Chaozhong Wu, Dan Jiang, P.H.A.J.M. van Gelder. 2021. A comprehensive statistical investigation framework for characteristics and causes analysis of ship accidents: A case study in the fluctuating backwater area of Three Gorges Reservoir region. *Ocean Engineering*. 229, 108981.

[3] **Jie Xue**, Chaozhong Wu, Zhijun Chen, P.H.A.J.M. van Gelder, Xinping Yan. 2019. Modeling human-like decision-making for inbound smart ships based on fuzzy decision trees. *Expert Systems with Applications*. 115, 172-188.

[4] **Jie Xue**, P.H.A.J.M. van Gelder, Reniers Genserik, Eleonora Papadimitriou, Chaozhong Wu. 2019. Multi-attribute decision-making method for prioritizing maritime traffic safety influencing factors of autonomous ships' maneuvering decisions using grey and fuzzy theories. *Safety Science*. 120, 323-340.

[5] **Jie Xue**, Zhijun Chen, Eleonora Papadimitriou, Chaozhong Wu, P.H.A.J.M. van Gelder. 2019. Influence of environmental factors on human-like decision-making for intelligent ship. *Ocean Engineering*. 186, 106060.

[6] **Jie Xue***, Genserik Reniers, Jie Li, Ming Yang, Chaozhong Wu, P.H.A.J.M. van Gelder. 2021. A bibliometric and visualized overview for the evolution of process safety and environmental protection. *International Journal of Environmental Research and Public Health*. 18(11), 5985.

[7] **Jie Xue**, Eleonora Papadimitriou, Tsz Leung Yip, Zhijun Chen, Chaozhong Wu, P.H.A.J.M. van Gelder. 2021. Machine learning for maritime safety of autonomous ships: A bibliometric analysis and future research trends. *Reliability Engineering & System Safety.* (Submitted)

[8] Junfeng Jiang, Qiushi Chen, **Jie Xue***, Haobo Wang, Zhijun Chen. 2020. A novel method about the representation and discrimination of traffic state. *Sensors*. 20(18), 5039.

[9] Changzhen Li, Junyi Yu, **Jie Xue***, Wei Chen, Shoufeng Wang, Kun Yang. 2021. Maritime broadband communication: Wireless channel measurement and characteristic analysis for offshore waters. *Journal of Marine Science and Engineering*, 9(7), 783.

[10] Zhijun Chen, **Jie Xue**, Chaozhong Wu, Lingqiao Qin, Liqun Liu, Xiaozhao Cheng. 2018. Classification of vessel motion pattern based on Automatic Identification System. *Ocean Engineering*. 161, 69-76.

[11] Dan Jiang, Bing Wu, Zhiyou Cheng, **Jie Xue**, P.H.A.J.M. van Gelder. 2021. Towards a probabilistic model for estimation of grounding accidents in fluctuating backwater zone of the Three Gorges Reservoir. *Reliability Engineering & System Safety*. 205, 107239.

[12] Xinpeng Wang, Chaozhong Wu, **Jie Xue**, Zhijun Chen. 2020. A method of personalized driving decision for smart car based on deep reinforcement learning. *Information*. 11(6), 295.

- **Conference papers (presentations)**

[1] **Jie Xue**, Eleonora Papadimitriou, Chaozhong Wu, P.H.A.J.M. van Gelder. Statistical analysis of the characteristics of ship accidents for Chongqing maritime safety administration district. *Forum on Integrated and Sustainable Transportation Systems (FISTS)*. Delft, Netherlands, 2020, November 3-5.

[2] **Jie Xue**, P.H.A.J.M. van Gelder, Eleonora Papadimitriou, Zhijun Chen, Chaozhong Wu. Grey relational analysis of environmental influencing factors of autonomous ships' maneuvering decision-making. *The 5th International Conference on Transportation Information and Safety (ICTIS)*. Liverpool, UK, 2019, July 14-17.

[3] **Jie Xue**, P.H.A.J.M. van Gelder. Modeling pilot's navigational decision-making in typical scenarios for autonomous ship safety. *The 6th Technology, Management & Policy (TMP) Graduate Consortium*. Washington, D.C., USA, 2019, June 16-18.

[4] **Jie Xue**, P.H.A.J.M. van Gelder, Chaozhong Wu, Zhijun Chen. Influence of environmental factors on intelligent ship decision-making: rules learned from inward scenario case study. *Transportation Research Board 98th Annual Meeting*. Washington, D.C., USA, 2019, January 13-17.

[5] **Jie Xue**, Chaozhong Wu, Zhijun Chen. Research on decision-making factors of ship's driving behavior based on grey relation entropy analysis method. *The 18th COTA International Conference of Transportation Professionals (CICTP)*. Beijing, China, 2018, July 5-8.

[6] **Jie Xue**, Chaozhong Wu, Zhijun Chen. A novel interpolation algorithm for ship trajectory restoration. *Transportation Research Board 97th Annual Meeting*. Washington, D.C., USA, 2018, January 07-11.

[7] **Jie Xue**, Chaozhong Wu, Zhijun Chen, Xiaoxuan Chen. A novel estimation algorithm for interpolating ship motion. *The 4th International Conference on Transportation Information and Safety (ICTIS)*. Alberta, Canada, 2017, August 8-10.

[8] **Jie Xue**, Chaozhong Wu, Zhijun Chen. Ship AIS data mining and processing method in bridge waters of inland river. *The 17th COTA International Conference of Transportation Professionals (CICTP)*. Shanghai, China, 2017, July 07-09.

- **Chapter**

[1] **Jie Xue**, Chaozhong Wu, P.H.A.J.M. van Gelder. 2019. A novel algorithm for modeling human decision making of inbound merchant ships-A case study of the

Shanghai Waigaoqiao phase IV port, in *Advances in Marine Navigation and Safety of Sea Transportation*. ed. Adam Weintrit, Tomasz Neumann. CRC Press Taylor & Francis Group.