World Maritime University

The Maritime Commons: Digital Repository of the World Maritime University

Maritime Safety & Environment Management Dissertations

Maritime Safety & Environment Management

8-27-2017

Research on improving navigation safety based on big data and cloud computing technology for Qiongzhou strait

Rui Wang

WORLD MARITIME UNIVERSITY

Dalian, China

RESEARCH ON IMPROVING NAVIGATION SAFETY BASED ON BIG DATA AND CLOUD COMPUTING TECHNOLOGY FOR QIONGZHOU STRAIT

By

WANG RUI

The People's Republic of China

A research paper submitted to the World Maritime University in partial Fulfillment of the requirements for the award of the degree of

MASTER OF SCIENCE

MARITIME SAFETY AND ENVIRONMENT MANAGEMENT

2017

DECLARATION

I certify that all the material in this research paper that is not my own work has been identified, and that no material is included for which a degree has previously been conferred on me.

The contents of this dissertation reflect my own personal views, and are not necessarily endorsed by the University.

Signature: Wang Rui

Date: June 29, 2017

Supervised by: Yang Tingting

ACKNOWLEDGEMENT

I am sincerely grateful to World Maritime University for offering me this infrequent opportunity to study for the master degree of Maritime Safety and Environment Management in Dalian Maritime University. Since the year of 2009, I graduated from DMU as a graduate of Navigation College, seven years has passed by. With seven years of working experiences and life experiences, the feeling of going back campus is full of pleasure.

Firstly, I would like to thank my loving wife and dear son, they are my spiritual pillar and power source. And I also want to thank my parents and parents in law, thank them always supporting me and encouraging me.

Secondly, I would like to pay my highest respects to my supervisor, Professor Yang Tingting, as an erudite scholar and outstanding professor, who gave me professional guidance and in-depth inspiration during the period of paper writing.

Thirdly, I would like to thank the Hainan MSA, who supplies me this opportunity going back school and carrying out in-deep academic research. Since the year of 2012, I have worked in Qiongzhou strait VTS center for five years, and VTS operator is a royal job, I would like to dedicate this paper to the Qiongzhou strait VTS center, which is the harvest of the working accumulation and study in the past five years.

Finally, I would like to give my best regards to all the professors and classmates of this programme. With profound theoretical foundation and rigorous academic attitude, professors of WMU have given me a completely new understanding of maritime safety and environment management, which will give me solid support in my future work.

ABSTRACT

Title of Dissertation: **Research on improving navigation safety based on big data and cloud computing technology for Qiongzhou strait**

Degree: **MSc**

It is said that the twenty-first century is the era of big data, and the application and thinking model of data in various industries will directly affect the development and innovation of the industry. This paper will establish the model of navigation risk identification and navigation safety assessment based on the thinking model of big data to improve the navigation safety. The main contributions of this paper is:

1. Establishing the model of navigation risk identification and navigation safety assessment based on big data and cloud computing technology, which is the core content of this paper.

2. Based on the model above, this paper utilizes the PSO algorithm to be the tool to carry out the task of identifying the navigation risk. The PSO algorithm has an obvious advantage, which is treating each vessel in the coverage as a target, and through comparing the fitness of the target with the DCPA and TCPA, therefore assess the navigation risk of the vessel.

3. The BP neutral networks algorithm is used to assess the navigation safety level via establishing the neutral networks and selecting the transfer and training functions, then training the sample and obtaining the system parameters.

Key Words: Navigation Safety and Risk, Big Data, Cloud Computing, Intelligence Algorithm, Model

TABLE OF CONTENTS

DECLARATION i

ACKNOWLEDGEMENT ii

ABSTRACT iii

TABLE OF CONTENTS iv

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ABBREVIATIONS x

Chapter 1: Introduction 1

1.1 Research background 1

1.2 Brief introduction of relevant research 2

1.3 Research objective 2

1.4 Summary of this paper 2

Chapter 2: Relevant concepts 4

2.1 What is big data 4

2.1.1 Definition of big data 4

2.1.2 Characteristics of big data 4

2.1.3 Technology system of big data 6

2.1.3.1 Collection of big data 7

2.1.3.2 Preparation of big data 7

2.1.3.3 Storage of big data 7

2.1.3.4 Analysis and mining of big data 8

2.1.3.5 Demonstration and visualization of big data 8

2.1.4 Application of big data 8

2.1.4.1 Business Intelligence(BI) 8

2.1.4.2 Big Data in transportation industry 9

2.2 What is cloud computing 10

2.2.1 Definition of cloud computing 10

2.2.2 Feature of cloud computing 11

2.2.3 Architecture layers of cloud computing 11

2.2.3.1 IAAS 12

2.2.3.2 PAAS 12

2.2.3.3 SAAS 12

2.3 Relationship between big data and cloud computing 13

Chapter 3: Navigation data and statistics of Qiongzhou strait 14

3.1 Brief introduction of Qiongzhou strait 14

3.1.1 Main channels in Qiongzhou strait 14

3.1.2 Traffic Separation Schemes(TSS) and Reporting System in Qiongzhou strait 15

3.1.3 Hydrological and meteorological information of Qiongzhou strait 16

3.1.4 Qiongzhou Strait VTS center 16

3.1.4.1 Staff structure of Qiongzhou Strait VTS center 17

3.1.4.2 Equipment allocation of Qiongzhou strait VTS center 18

3.2 Relevant statistic of Qiongzhou strait 18

Chapter 4: Model of navigation risk identification and navigation safety assessment based on big data and cloud computing for Qiongzhou strait 20

4.1 Data source 21

4.2 Data pretreatment 22

4.2.1 Data extraction 22

4.2.2 Data cleaning 23

4.2.3 Extract-Transform-Lord(ETL) of big data 24
4.2.4 Data integration 25
4.3 On-line transaction processing(OLTP) 25
4.4 Data warehouse 26
4.5 Data mining 27
4.6 Algorithm in ship risk identification model 27

Chapter 5: Navigational risk identification based on PSO algorithm and cloud computing 29
5.1 Algorithm principle 29
5.2 Basic formula of PSO algorithm 29
5.3 PSO algorithm process 31
5.4 Navigational risk identification in Qiongzhou strait based on PSO algorithm 32
5.4.1 Relative speed and relative course of Vessel-2 relative to Vessel-1 34
5.4.2 Calculation of DCPA and TCPA 35
5.4.3 Fitness evaluation 36
5.5 PSO algorithm based on cloud computing task scheduling 37

Chapter 6: Navigational safety assessment based on BP neural network algorithm and big data technology 38
6.1 Structure of Back Propagation(BP) neural network 38
6.2 BP learning algorithm 40
6.2.1 Forward Propagation computation output process 41
6.2.2 Back Propagation correct weights and threshold process 42
6.3 Navigational safety assessment based on BP neural networks algorithm 42
6.3.1 *Determine the dimension of network input and output vectors* 42
6.3.1.1 *Determine the dimension of network input vectors* 42
6.3.1.2 *Determine the dimension of network output vectors* 43

6.3.2 *Determine the transfer and training functions* 43

6.3.3 *Determine the number of hidden layer neurons* 44

Chapter 7: Conclusion 45

References 47

LIST OF TABLES

Table-3.1 Working data statistic of Qiongzhou strait VTS center(2012-2016) 19

Table-4.1 Example of AIS data 23

Table-5.1 Example of ships data 33

Table-5.2 Standard reference of DCPA and TCPA in Qiongzhou strait 36

LIST OF FIGURES

Figure-2.1	The four V`s of big data	5
Figure-2.2	Technology system of big data	6
Figure-2.3	Structure system of Business Intelligence	9
Figure-2.4	The diagrammatic sketch of big data in logistics	10
Figure-2.5	Architecture layers of cloud computing	13
Figure-3.1	Traffic Separation Schemes in Qiongzhou strait	15
Figure-3.2	Staff structure of Qiongzhou strait VTS center	17
Figure-3.3	Equipment allocation of Qiongzhou strait VTS center	18
Figure-4.1	Design drawing of navigation risk identification and navigation safety assessment for Qiongzhou strait	21
Figure-4.2	Data extraction principle	22
Figure-4.3	Process of data cleaning	24
Figure-4.4	Process of data ETL	25
Figure-4.5	Data warehouse design	26
Figure-4.6	Summarization flow of data mining	27
Figure-5.1	PSO algorithm process	31
Figure-5.2	A sketch map of the example above	33
Figure-6.1	Model of neuron	38
Figure-6.2	Structure of BP neural network	39

LIST OF ABBREVIATIONS

ADE	Application Development Environment
AIS	Automatic Identification System
BI	Business Intelligence
BP	Back Propagation
ETL	Extract-Transform-Load
GMDSS	Global Maritime Distress And Safety System
GPS	Global Positioning System
HDFS	Hadoop Distributed File System
IDC	International Data Corporation
IaaS	Infrastructure As A Service
MATLAB	Matrix Laboratory
MMSI	Maritime Mobile Service Identify
MSA	Maritime Safety Administration
N/M	Nautical Mile
NoSQL	Not Only Structured Query Language
OLTP	On-Line Transaction Processing
PaaS	Platform As A Service
PSO	Particle Swarm Optimization
RFID	Radio Frequency Identification
SaaS	Software As A Service
TB	True Bearing
TC	True Course
TSS	Traffic Separation Schemes
VHF	Very High Frequency
VTS	Vessel Traffic Service

Chapter 1 Introduction

1.1 Research background

Qiongzhou Strait, also known as Leizhou Strait, is located between Hainan Island and the Leizhou Peninsula, as one of the three major Straits in China, Qiongzhou Strait plays a key role in the development of South China Sea and the relevant Provinces. Qiongzhou Strait is the important economic and transportation hubs among Beibu Bay Economic Zone, Guangdong Province and Hainan Province; Hence, the navigational density in Qiongzhou Strait is high, and the situation of navigation safety in the Strait is severe.

Qiongzhou Strait is among the coverage of Qiongzhou Strait VTS Center. In general, the VTS operator utilizes information from Radar, AIS, GPS, VHF, VTS terminal and relevant regulations to supervise and manage the traffic in the Strait. However, this current supervision measure is dependent on the manpower to identify and assessment of the potential risk seriously, which is an undeniable factor influencing the navigation safety in Qiongzhou Strait.

The research contents of this paper is utilizing technical measures to enhance the service level of Maritime Safety Administration(MSA), in order to improve the navigation safety in Qiongzhou Strait.

Nowadays, the collection, storage, analysis and utilization of data attracts more and more public attention, and the application of big data and cloud computing technology becomes increasingly widespread. The digital era is liberating manpower from complex working gradually.

1.2 Brief introduction of relevant research

There are lots of relevant research on navigation safety in Qiongzhou Strait, ranging from universities, enterprises to government bodies. Nearly all the research mainly focus on the planning of navigation model, formulating regulations, ship design, safety management of company and training of crew. However, at the position of MSA, the VTS operators` ability of identifying and assess the potential navigation risk is still a difficult issue.

1.3 Research objective

The research objective of this paper is using the methodology of big data and cloud computing technology to find out a suitable model, using the maritime relevant data to identify navigation risk and assess navigation safety, and provide feedback on the identified risks to users, so then the users can analyze the risks and take suitable actions to improve the navigation safety level in Qiongzhou Strait.

This paper will introduce the big data and cloud computing technology in detail, and based on the technology, the data mining method of big data will be used to modeling the risk identification and safety assessment mechanism, which is the model of navigation risk identification and navigation safety assessment for Qiongzhou strait. So that, the main research method in this paper is modeling.

1.4 Summary of this paper

This paper contains totally seven chapters: the Chapter 2 will introduce several relevant concepts used in this paper, including the big data, data mining and the cloud computing; the Chapter 3 will introduce the approximate information of Qiongzhou strait, including the geographic information and marine data; Chapter 4 will introduce the structure and functions of the model in this paper; Chapter 5 and 6 will introduce the algorithms used in the model; and lastly the Chapter 7 is the conclusion.

Chapter 2 Relevant Concepts

2.1 What is big data

2.1.1 Definition of big data

Since the concept of “Big Data” appeared, there was no accepted and unified definition of the “Big Data” because the concept of big data is abstract. However, many famous institutions defined the big data in different aspects, below are some typical examples:

Gartner Group: Big data refers to the information assets which utilize new processing mode to have a stronger decision-making ability, deeper insight and process optimization capabilities, and have the characteristics of mass, diversity and high growth rate.

Mckinsey and Company: Big data refers to the data sets which cannot be collected, stored, managed or analyzed by traditional database tools within a given period.

IDC: Generally speaking, big data involves two or more than two forms of data set, which collects 100TB above data via high-speed and real-time data streams; or which collects small data, but the data volume is increasing by more than 60% each year.

Viktor Mayer-Schönberger: Big data is about applying math to huge quantities of data in order to infer probabilities(Mayer-Schönberger, 2013, p.12).

2.1.2 Characteristics of big data

There are four characteristics concerning the Big Data industry, which are volume, variety, velocity and value(Chen, Xu, 2015, p.3). As is shown in Figure-2.1 below.

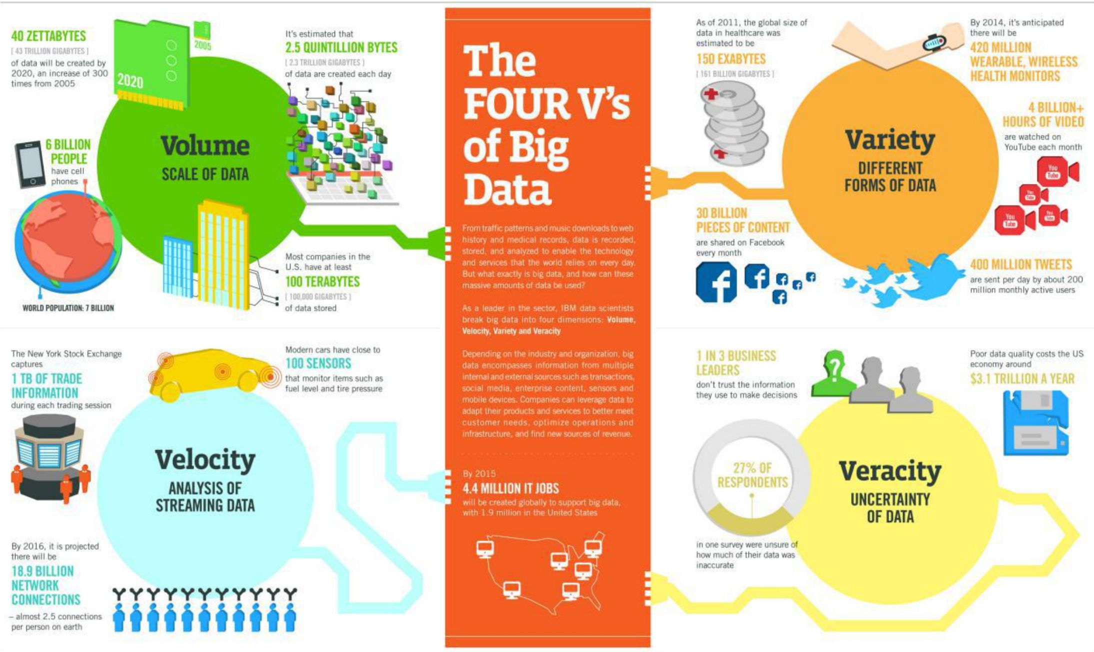

Figure-2.1: The four Vs of big data

Source: IBM Big Data & Analytics Hub. (2012). *Four V`s Of Big Data*. Retrieved May 2, 2017 from the World Wide Web: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

2.1.2.1 Volume

Huge volume is the basic characteristic of the big data. With the widespread application of internet technology and the sharp rise of Internet users, the production and share of data becomes convenient and easy, which produces an ever-increasing volume of data.

2.1.2.2 Variety

The data type includes not only traditional data which covered storage data, log data, record data, text data, location data, sensor data and so forth, but also the unprocessed, semi-structured and unstructured information.

2.1.2.3 Velocity

The frequency of data creation and updating is also an important characteristic of big data. For example, each day, the user updates the blog and browses stock information, all of the data need to be transmitted, which requires a higher speed for data processing.

2.1.2.4 Value

The volume of data presents sharp rise. However, the valuable information hidden behind the massive volume data does not show a corresponding increase. Instead collection of valuable information is gradually becoming more difficult.

Hence, the four characteristics reflect the big data not only has massive volume big data, meanwhile, the analysis of big data is more complex, which requires high speed and effectiveness.

2.1.3 technology system of big data

Big data is the technology which gains useful information from various types of data. A series of new technologies have been invented, which serve as the main method to the collection, preparation, storage, analysis and demonstration of big data. Figure-2.2 below shows the technology system of Big Data.

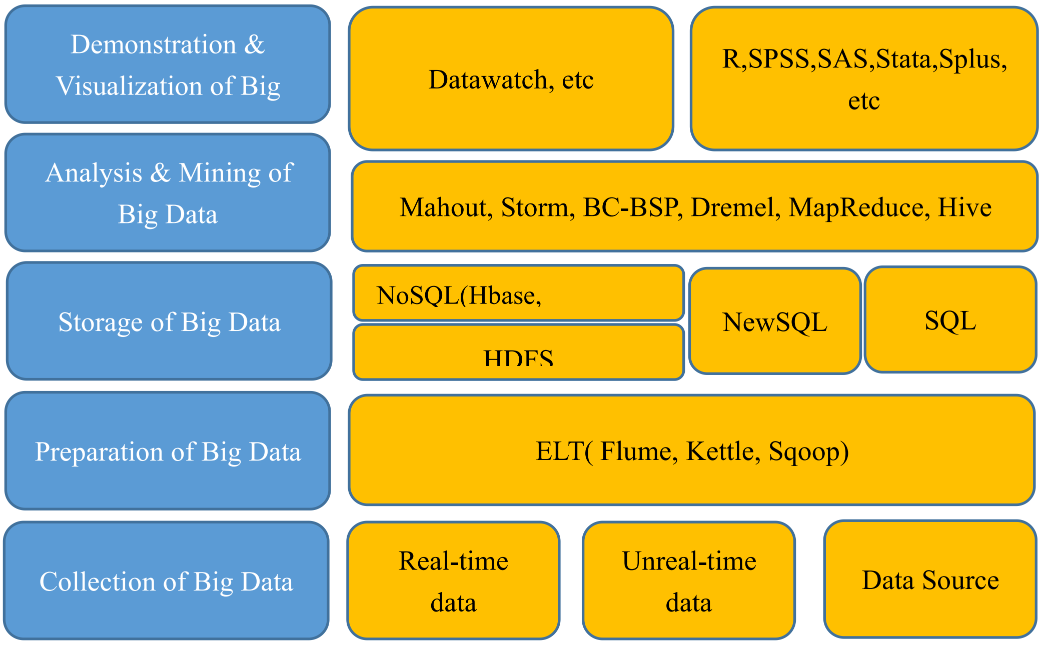

Figure-2.2 Technology system of big data

Source: Chen Gongmeng, Xu Chengzhong. (2015). *Introduction To Big Data*. Beijing: Tsinghua University Press.

2.1.3.1 Collection of big data

Collection of big data refers to the gain of multi-structure data, semi-structure data and unstructured data by means of RFID data, sensor data, interactive data and mobile internet data. The collection of big data is the basis of big data service system. In general, the collection of big data includes Intelligent Sensing Layer and Foundation Support Layer. Collecting method of big data mainly contains system log collection, Internet data collection, database collection and other data collection methods.

2.1.3.2 Preparation of big data

Preparation of big data is extracting, transforming and loading the data(ETL). The purpose of ETL is transforming the original complex data into simple structural data or conveniently processing structural data.

2.1.3.3 Storage of big data

Big data has strict challenge to the storage technology. The first is volume expansion, which is to ensure that the storage structure and file system can expand the storage capacity cheaply; the second is expansion of data format, which should meet the requirement of sorts of structured data. Nowadays, the main storage tool of big data includes HDFS and NoSQL.

2.1.3.4 Analysis and mining of big data

Analysis and mining of big data is the process of collection, collating, machining and analyzing data for particular information. Data analysis refers to the searching, and analysis of the prepared data via particularly analysis method or technique, in order to find out the causal relationship, internal link and regular pattern, which can supply reference resources for decision-making. Data mining refers to the process of extracting hidden and useful information from massive volume, random, obscure data.

2.1.3.5 Demonstration and visualization of big data

Visualization technology of big data can supply directly data representation, transform complex data and data relationship into simple graphical or intelligentized form which can be analyzed by the user. A visualization system has a basic element, which is the ability to deal with massive amount of multivariate time series data.

2.1.4 Application of big data

2.1.4.1 Business Intelligence(BI)

Business Intelligence is not a product or system. It is an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community with easy access to business data. Business Intelligence Roadmap Specifically addresses decision-support applications and databases(Moss, Atre, 2003, p.4).

There is another acceptable definition of BI: Business intelligence is the ability of the businesses utilizing modern information technology to improve the business decision, perfect the business procedure, increase the business performance and competitiveness when they collect, organize, manage and analyze the structured and unstructured data or information. Figure-2.3 below shows the structure system of BI.

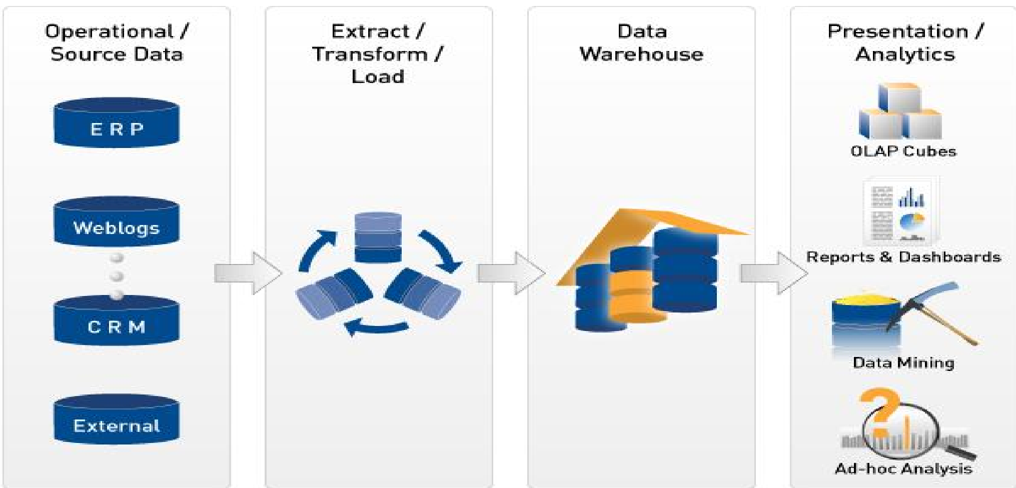

Figure-2.3 Structure system of Business Intelligence

Source: SQL Power Group. (2017). *What Is Business Intelligence*. Retrieved May 5, 2017 from the World Wide Web: http://www.sqlpower.ca.

From the structure system of BI we can see that the entire procedure of BI system requires four important technology sections, which contains Data source, ETL, Data

Warehouse and Front display. Business Intelligence can supply data analysis and decision analysis for particular bodies, and assist the enterprise in formulating strategy policy, which is an important application of big data.

2.1.4.2 Big data in transportation industry

Transportation industry is one of the important pillars of economic development. The complex and dynamic nature of transportation is the main problem in resolving the transportation problem. Big data technology can utilize the data of logistics companies, transportation companies and base-station real-time data to analyze and assess the traffic circumstance and plan the optimal route in order to relieve the transportation pressure and improve the transportation efficiency. That is the concept of intelligence transportation(Ahmed, 2017). Figure-2.4 shows the diagrammatic sketch of big data in logistics.

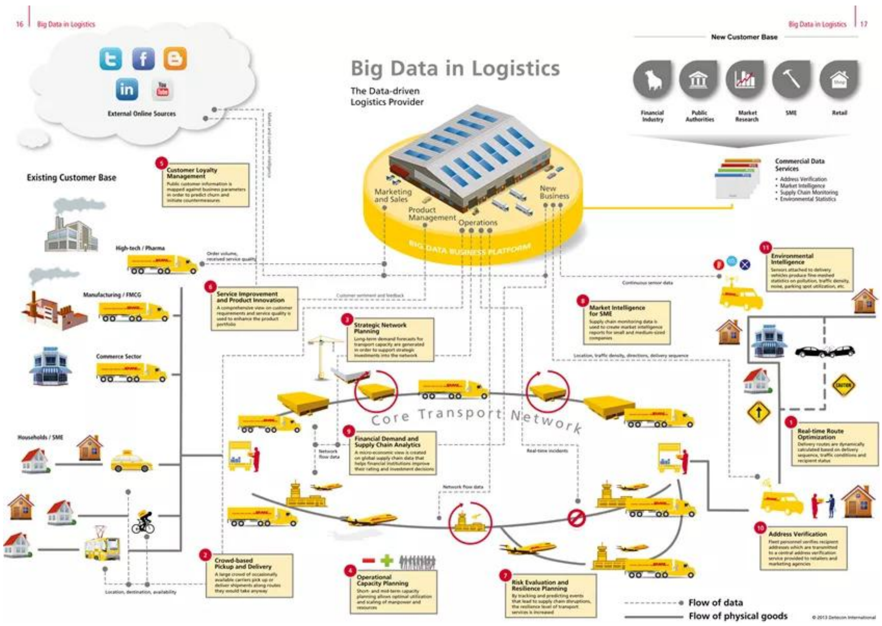

Figure-2.4 Big data in logistics

Source: Kevin Jessop. (November 14, 2016). *What Is the Impact of Big Data in the Transportation & Supply Chain Industries*. Retrieved May 10, 2017 from the World Wide Web: http://cerasis.com/?s=What+Is+the+Impact+of+Big+Data+in+the+Transportation+%26+Supply+Chain+Industries

2.2 What is cloud computing

2.2.1 Definition of cloud computing

The National Institute of Standards and Technology of U.S. defined the Cloud Computing as "*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources(e.g., networks, servers, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*." (Mell, Grance, 2011)

Cloud computing includes two aspects. One is the applications delivered as services on the Internet; the other is the hardware and systems software in the data centers that provide such services(AVM, 2010, p.52). Cloud computing is not a term for an independent technology, but rather a general term for all the technologies needed to implement a cloud computing model, which includes Distributed Computing Technology, Virtualization, Network Technique, Server Technique, Data Center Technique and Cloud Computing Platform Technique. Broadly speaking, Cloud Computing technology almost involves the vast majority of current information technology(Fan, Liu& Huo, 2016, p.124).

2.2.2 Features of cloud computing

According to the definition of Cloud Computing from the National Institute of Standards and Technology of the U.S., the Cloud Computing should have the following feature(Rosenblum, 2012):

- Elasticity of resource pooling
- Service on demand
- Universality and automatic feature

- Virtualization

2.2.3 Architecture layers of cloud computing

The Cloud service provider mainly supplies three kind of service: Infrastructure as a Service(IaaS), Software as a Service(SaaS) and Platform as a Service(PaaS). Figure-5 shows the resources managed at each layer(Fan, Liu& Huo, 2016, p.127).

2.2.3.1 IaaS

IaaS is located at the bottom of cloud computing service, which supplies information service based on the server, storage device, network device. In general, the information service contains basic calculation and storage.

2.2.3.2 PaaS

PaaS is located in the middle level of cloud computing service. It provides Application Development Environment(ADE) to the end user, including Application Programming Interface and Operation Platform.

2.2.3.3 SaaS

SaaS is located on the top of cloud computing service, ant it is the commonest Cloud Computing service. The user uses software on the Internet via standard Web browser, and the service provider is in charge of management and maintenance of the software.

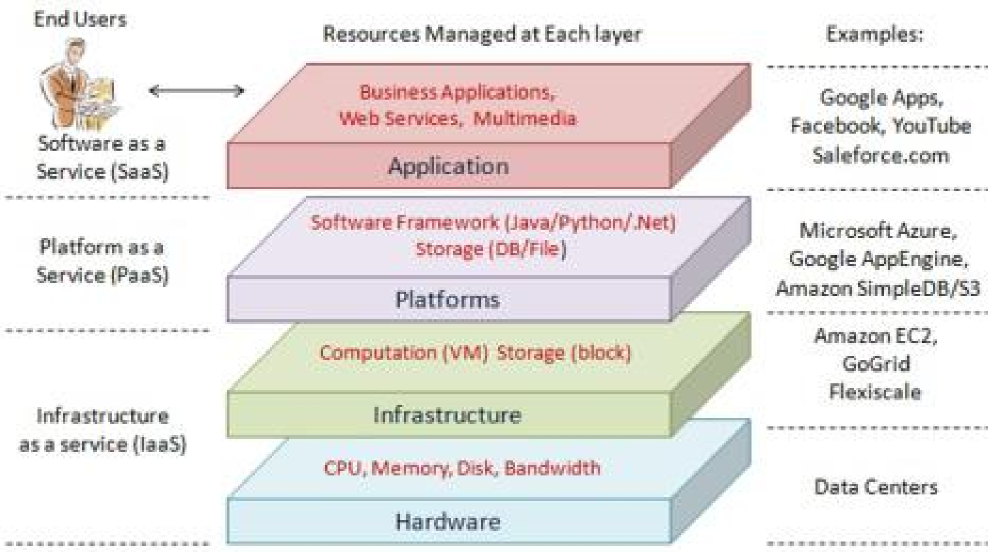

Figure-2.5 Architecture layers of cloud computing

Source: Krešimir Popović, Zeljko Hocenski. (2010). Cloud Computing Security Issues and Challenges. *Proceedings of the 33rd International Convention*. Opatija, Croatia, MIPRO.

2.3 Relationship between big data and cloud computing

Essentially, the big data and cloud computing is interdependent. Cloud computing focuses on calculation, and it is the technological foundation of big data. On the contrary, big data focuses on the calculated object, which is the application of cloud computing(Ferkoun, 2014).

However, the concept of big data and cloud computing is not distinct. The big data requires the ability of computing to deal with data, which contains a calculation feature. And the data storage device is the important component in the infrastructure of cloud computing. Therefore, the big data and cloud computing are inextricable.

Chapter 3 Navigation data and statistics of Qiongzhou strait

3.1 Brief introduction of Qiongzhou strait

Qiongzhou Strait is part of the South China Sea, which is located between the Leizhou Peninsula and Hainan Island, connecting the Beibu Gulf and serving as the important sea channel for the Southeast coast of China to enter into Beibu Gulf.

The eastern coverage of Qiongzhou Strait is the sea area within the baseline with Mulantou Lighthouse at the center and 22 nautical miles in radius. There are four channels in the eastern coverage and the hydrological condition is complex. The Western sea area has a the straight line between Dengloujiao and Lingaojiao as boundary. In the middle of Qiongzhou Strait, there is Traffic Separation Schemes(TSS), which is the busiest sea lane between Haian Port(Guangdong Provence) and Haikou Port(Hainan Province).

3.1.1 Main channels in Qiongzhou strait

In the eastern coverage of Qiongzhou Strait, there are totally four channels: Wailuo Channel, North Channel, Middle Channel and South Channel. The depth of water in Wailuo Channel is below 5.0 meters, therefore it is only suitable for the sailing of small ships and fishing boat. The North Channel can be used by Chinese (including Hongkong and Taiwan) vessels whose maximum draft is below 7.0 meters. The Middle Channel is the only deep-water fairway of Qiongzhou Strait, which can be used by all foreign vessels and some of Chinese vessels whose maximum draft is over 7.0 meters. The South Channel is located at the southernmost of Qiongzhou Strait, linking the Strait and South China Sea, which is the main sea lane between north and south of Hainan Island.

3.1.2 Traffic Separation Schemes(TSS) and Reporting System in Qiongzhou strait

The TSS in Qiongzhou Strait(Figure-3.1) is composed of eight separation zones, which regularizes the east-west and north-south vessel flow. The True Course(TC) of east-west vessel flow is 081^0 and 261^0， and that of north-south vessel flow is 170^0 and 350^0(MSA, 2009).

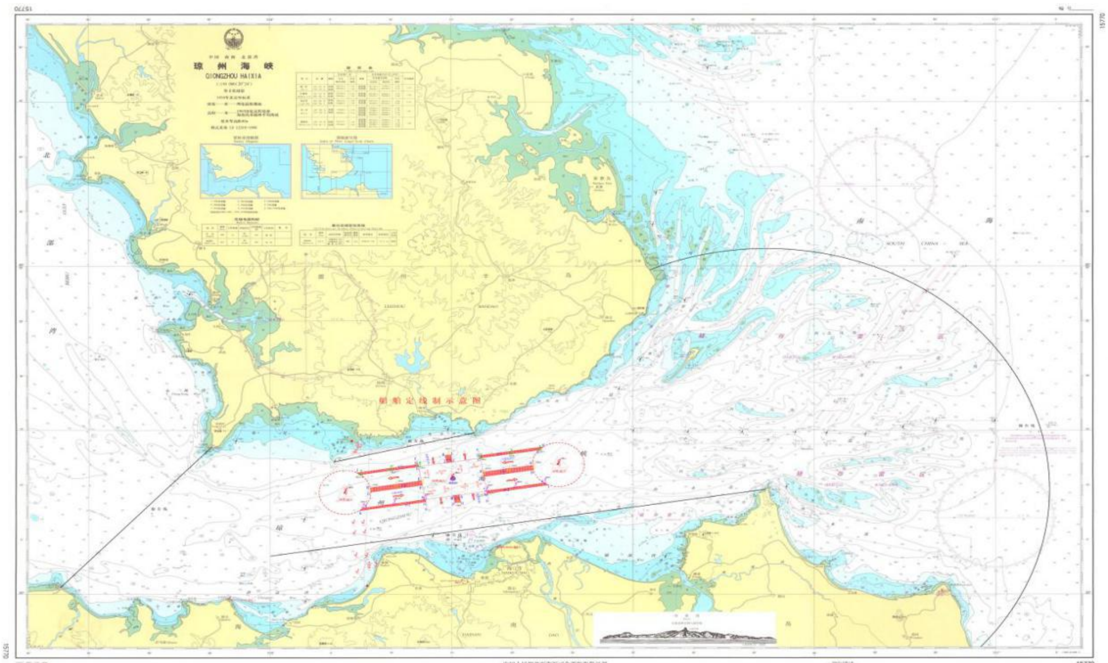

Figure-3.1 Traffic Separation Schemes in Qiongzhou strait
Source: Maritime Safety Administration. (2017). *Chart15770: Qiongzhou Strait*. Shanghai: Donghai Navigation Safety Administration(DNSA) MOT.

According to the Guidelines of Qiongzhou Strait VTS, ships sailing into or leaving Qiongzhou Strait should report to the VTS center. There are four reporting lines in the Strait:(MSA, 2009)

North Reporting Line is the line between Sandun Light-buoy(20^{0}12`06``N, 110$^{0}$05`24``E) and Paiweijiao Light Beacon(20^{0}14`48``N, 110$^{0}$16`54``E);

South Reporting Line is the line between Mulantou Lighthouse(20^{0}09`36``N, 110$^{0}$41`04``E) and the point(20^{0}03`30N, 110$^{0}$00`00``E);

West Reporting Line is the line between Lingaojiao Lighthouse(20^{0}00`38``N, 109$^{0}$42`42``E) and Dengloujiao Lighthouse(20^{0}13`28``N, 109$^{0}$55`07``E);

East Reporting Line is the line taking the Mulantou Lighthouse(20^{0}09`36``N, 110$^{0}$41`04``E) as the center, the curve of the arc drown from north to east of the 22 nautical miles as radius, and the curve connecting with the coastline.

3.1.3 Hydrological and meteorological information of Qiongzhou strait

The Qiongzhou Strait is located between subtropic zone and tropic zone, and the climate in the Strait is oceanic tropic monsoon climate. The ocean current in Strait is strong: when the southwest monsoon blows in summer each year, the ocean current flows from west to east and the rate is high, with the highest rate being 5~6 nautical miles; and in other seasons the ocean current flows east to west, with a law velocity(COIN, 2017).

The tide in Qiongzhou Strait is irregular lunar tide; the average tide range is about 1.0 meter, and the biggest tide range can achieve 7.8 meters(COIN, 2017).

3.1.4 Qiongzhou strait VTS Center

Qiongzhou Strait VTS Center was built in 1996 and the test run began in 1997, until the year of 2002, when was formally operated. From then on, Qiongzhou Strait started implementing Reporting System and Routing System. And the Qiongzhou Strait VTS Center is in charge of traffic management and information service in the strait.

3.1.4.1 Staff structure of Qiongzhou strait VTS Center(Figure-3.2)

The VTS Center has five teams, each team has one chief operator and three minor operators. The five operator teams are under the lead of VTS Directors, who report their duties to the Assigned Executive of Haikou MSA, and the Assigned Director of Hainan MSA receives working report from the Assigned Executive of Haikou MSA.

The working schedule is divided each day into three periods: morning period is 0800-1300; afternoon period is 1300-2200; and the evening period is 2200-0800. All the three operator teams are needed each day, and every five day is a cycle.

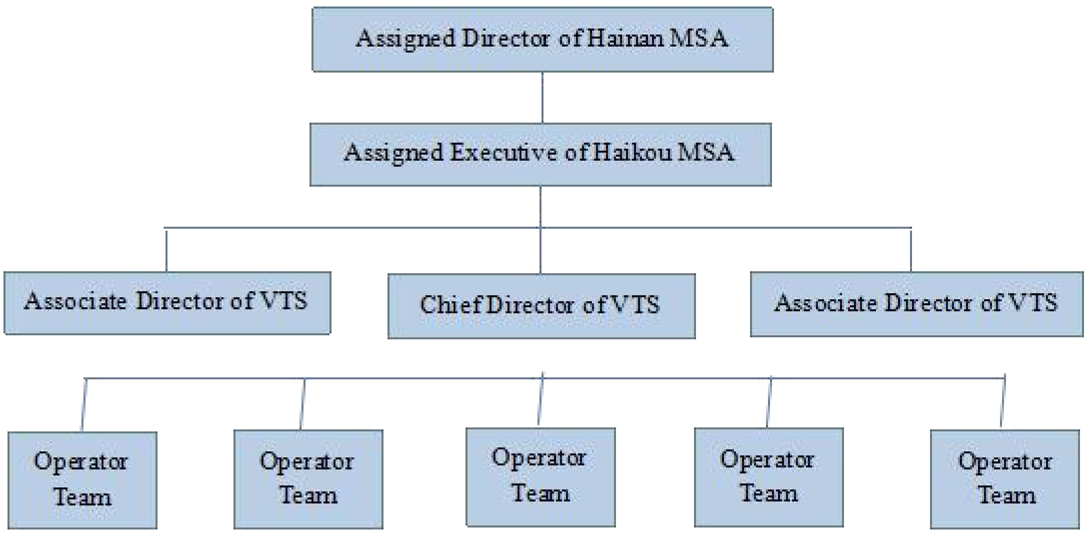

Figure-3.2 Staff structure of Qiongzhou strait VTS Center

Source: Qiongzhou Strait VTS Center. (2002). *VTS Operator Manual Guideline*. Unpublished internal working documents, Hainan MSA.

3.1.4.2 Equipment allocation of Qiongzhou strait VTS Center(Figure-3.3)

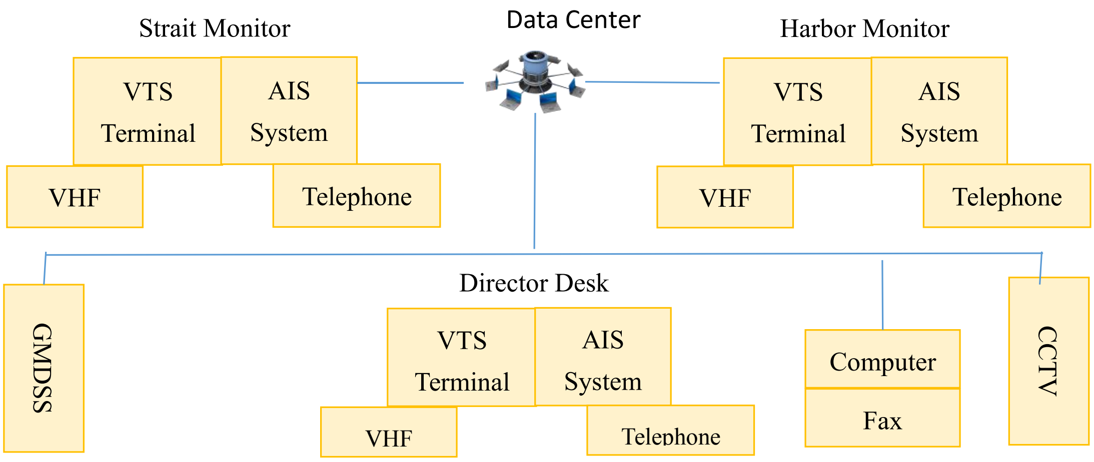

Figure-3.3 Equipment allocation of Qiongzhou strait VTS Center
Source: Qiongzhou Strait VTS Center. (2002). *VTS Operator Manual Guideline*. Unpublished internal working documents, Hainan MSA.

Qiongzhou Strait VTS Center allocated lots of equipment to assist the traffic management and information service: VTS system and AIS system can supply the dynamic circumstance of the Strait; VHF is the means of communication between VTS Center and ships; GMDSS is the method of receiving distress alerting, search and rescue communication and broadcast of safety information; CCTV can supply on-scene supervision of the harbor.

3.2 Relevant statistics of Qiongzhou strait

Table-3.1 is the working data statistics of Qiongzhou strait VTS center, from the statistics it can be seen that:

$$Information\text{-}Service_{(2012-2016)} = Broadcast_{(2012-2016)} + \mathrm{Remind}_{(2012-2016)} + \mathrm{Re}quest_{(2012-2016)}$$

$$= 258700(\mathrm{pieces})$$

$$Total - flow - of - vessels_{(2012-2016)} = 104876 + 106934 + 108527 + 109928 + 111691$$

$$=541956(\text{ships})$$

$$\frac{Information-Service_{(2012-2016)}}{Total-flow-of-vessels_{(2012-2016)}}=\frac{258700\,pieces}{541956\,ships}=0.477({pieces}/{ship})$$

Based on above calculation, it can be seen that each vessel passing Qiongzhou strait received 0.477 pieces information service within the period from 2012 to 2016. So that the working pressure of VTS operators is obvious; and on the other hand, it reflected that the current working model of Qiongzhou strait VTS center has many areas deserved to improve in order to enhance the working efficiency.

Table-3.1 Working statistics of Qiongzhou strait VTS Center(2012-2016)

Items \ Year		2012	2013	2014	2015	2016
Total Flow of Vessels(ships)		104876	106934	108527	109928	111691
Reports Received(pieces)		140986	144897	149737	153532	159231
Ships Followed(ships)		188945	190876	193876	196903	200837
Particular Monitoring(ships)		60154	61309	62874	63238	60254
Information Service (pieces)	Broadcast	7345	7963	8976	7588	13045
	Remind	23984	23795	24578	25784	22573
	At Request	18596	18340	18974	19359	17800
Early Warning Information (pieces)	Tropical Storm	124	130	139	121	131
	Gale Warning	408	419	434	422	432
	Dense Fog	571	548	579	564	557

Source: Haikou Maritime Safety Administration. (2016). *Working Statistics of Qiongzhou Strait VTS Center(2012-2016)*. Unpublished Internal Statistics. Haikou MSA.

Chapter 4 Model of navigation risk identification and navigation safety assessment based on big data and cloud computing for Qiongzhou strait

Qiongzhou Strait VTS operators are mainly responsible for supervising the navigation safety via supplying safety information service based on risks identified. And the major ship risks in Qiongzhou are mainly embodied in head-on, crossing and overtaking situation. However, the existing situation in Qiongzhou Strait is VTS operators utilizing VTS system and relevant equipment to identify and assess the ship risk, which is inefficient and imprecise.

This paper aims to model the navigation risk identification and navigation safety assessment based on big data technology. The model can divide the ship risk into different levels, and the VTS operator can send corresponding safety information or provide traffic control according to the risk level. Figure-4.1 is the design drawing of navigation risk identification and navigation safety assessment for Qiongzhou strait.

The model has two basic functions. The first is ship risk identification in real time ship flow based on On-Line Transaction Processing(OLTP), and the risk identification and assessment criteria in the algorithm are mainly Distance to the Closest Point of Approach(DCPA) and Time to the Closest Point of Approach(TCPA); the second function of the model is navigation safety assessment is based on statistics and analysis of amount of historical data, which can identify the risk in special area and circumstance. The two functions are independent and interrelated, and the link between them is the iterative processing of the model.

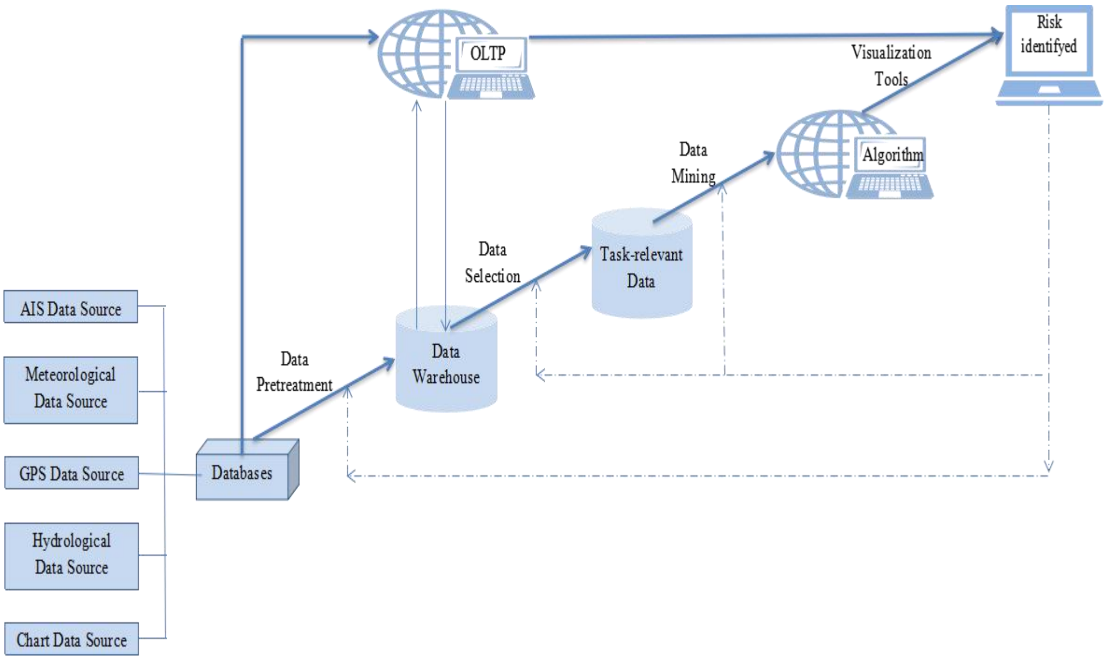

Figure-4.1 Design drawing of navigation risk identification and navigation safety assessment for Qiongzhou strait

Source: The Author

4.1 Data source

Data source is an important foundation of model, which is a device or raw medium providing required data for the model, and all the information that builds the database connection is stored in the data source. Modeling process is not only technology, but also it is the process that solves systematic process problem. Therefore, selecting suitable data source is very important. In this paper, the following data are required by the ship risk identification model requires below data:

AIS data, which includes: Ship Name, Call Sign, Draft, MMSI, Dangerous Cargo, Ship`s Location(Latitude, Longitude), Speed, Course, Last Port of Call and Destination.

GPS data, which includes Time data, Location and Voyage Plan of ships.

Chart data, which includes: Geographic Information, Depth at Chart Datum, Shallow Water information and Channel information.

Meteorological data, which includes Wind information(force, direction), Fog information.

Hydrological data, which includes Ocean Current information(direction, speed), Tide information.

4.2 Data pretreatment

4.2.1 Data extraction

The core of data extraction is the process of extracting the data from its data sources(DS) according to a rule. The principle of data extraction is listed as follows: given the DS, determining a mapping P from DS to the database, the mapping P extract data from DS and assemble the data into data set R in a certain format. The computing process achieving the mapping process can be seen as data extraction.

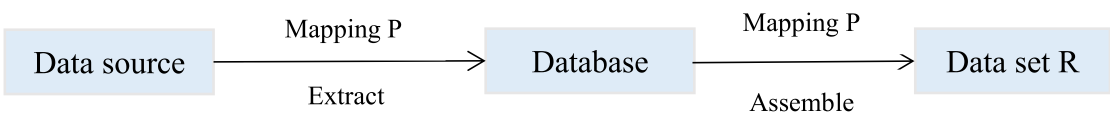

Figure-4.2 Data extraction principle

Source: Fan Chongjun, Liu Cheng& Huo Liangan. (2016). *Analysis And Application Of Big Data*. Shanghai: Lixinapn Press.

The ship risk identification model in Qiongzhou strait uses the data from AIS, GPS, chart and so on, so that the data extraction can be achieved via certain mapping. The Table-4.1 below is an example of AIS data:

Table-4.1 Example of AIS data

MMSI	Name	Location	Speed	Course	Draft	Destination	Call sign
***001	ABC	***	***	***	***	Beihai	***
***002	DEF	***	***	***	***	Fangcheng	***
***003	HIJ	***	***	***	***	Haiphong	***
......	......	......	......	......	......	......	......

Source: The Author

If the model searches the ship data whose destination is Beihai, then the extracting mapping code is:

Select MMSI, Name, Location, Speed, Course, Draft, Destination, Call sign

From AIS data

Where Destination= "Beihai"

The code has two mapping processes: the first is searching ships whose destination is Beihai in AIS data according to "*Where*" statement; the second is mapping the row where the destination is Beihai with the field selected by the "*Select"* statement and outputting the data searched.

4.2.2 Data cleaning

Data extraction is the first step in ETL, however, only the extraction cannot meet the requirement of users. That is because there are some errors in the extracted data. Therefore further data processing is needed for the extracted data, which is data cleaning. Data cleaning refers to detecting and removing errors and inconsistencies from data in order to improve the quality of data, and the quality of data including completeness, accuracy, concision, applicability, believability and accessibility(Dasu, Johnson, 2003,p.99).

According to certain rules and strategies, data cleaning process is mainly conducted through detecting, matching, statistics and merging method, and using specialized techniques or tools, transforming the polluted data into high-quality data meeting the users` requirement and outputting the data. Figure-4.3 below shows the process of data cleaning.

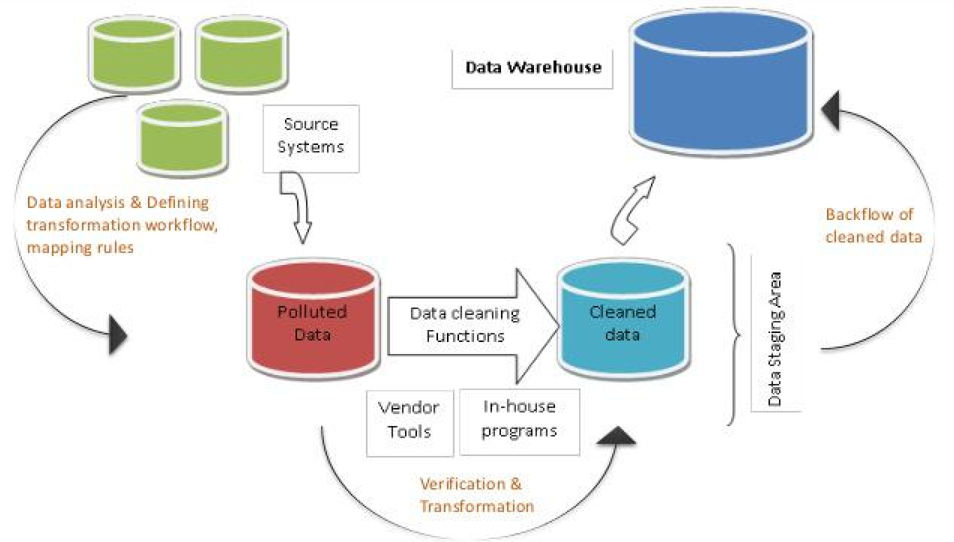

Figure-4.3 Process of data cleaning

Source: Erhard Rahm. (2000). *Data Cleaning: Problems and Current Approaches*. Germany, University of Leipzig: Author.

4.2.3 Extract-Transform-Lord(ETL) of big data

ETL is the process of data extraction, transform and load, which is an important part of building database. The process of ETL can extract data from the distributed and heterogeneous data sources to the staging database, and then clean, transform and integrate the data, and finally load data into the data warehouse, which is the basis of data mining. Figure-4.4 shows the process of ELT.

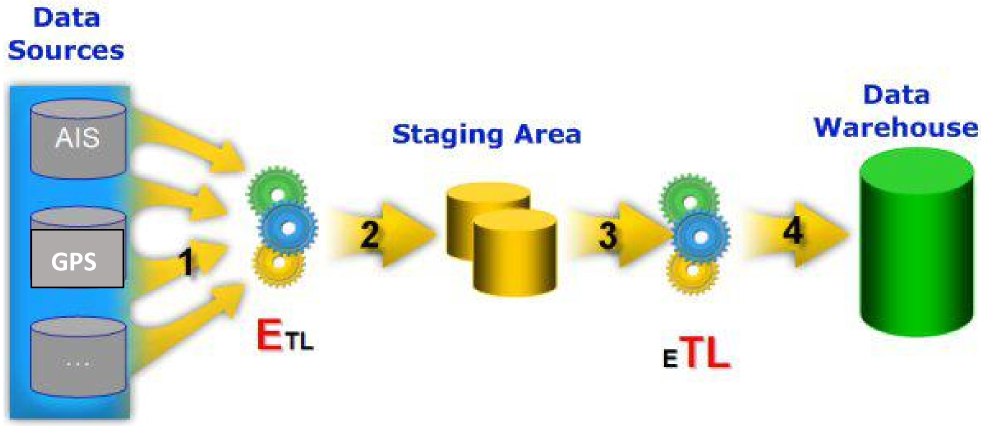

Figure-4.4 Process of data ETL

Source: Extract-Transform-Load (ETL) Technologies– Part 1. (December 31, 2012). *DB. Best Technologies*. Retrieved May 15, 2017 from the World Wide Web: https://www.dbbest.com/blog/extract-transform-load-etl-technologies-part-1/

4.2.4 Data integration

The ship risk identification model utilizes multiple databases, and these databases are formulated into different data sets. However, the ship risk identification requires cooperation between these data sets, in order to improve the accuracy of algorithm identifying the ship risk. Therefore, data integration is an important section of data pretreatment.(Doan, Halevy& Ives, 2012, p.17)

Data integration has many models, and data warehouse is one of them. This paper uses the data warehouse as the data integration in ship risk identification model, the detailed introduction of data warehouse in the following section.

4.3 On-Line Transaction Processing(OLTP)

On-line transaction processing usually means executing a very large number of small transactions very quickly(Garmany, Walker& Clark, 2005, p.139). One of the purposes of ship risk identification model in Qiongzhou Strait is utilizing dynamic information in Qiongzhou Strait identifying the potential navigational risk for vessels, which need the model can assess and response risk in real time. In this aspect, OLTP can implement part of the functions of the model.

4.4 Data warehouse

Data Warehouse is a kind of technology that aggregates structured data from several data sources in order to analyze and calculate according to the requirement of the higher algorithm. Data Warehouse is constructed by integrated, subject oriented, nonvolatile and time variant data from multiple heterogeneous data sources.(Inmon, Strauss& Neushloss, 2008, p.7) Data Warehouse can be utilized in two aspect: one is supporting decision-making based on data analysis and statistics; the other is integrating several multiple heterogeneous data sources and data reorganization. Figure-4.5 below is the Data Warehouse design chart.

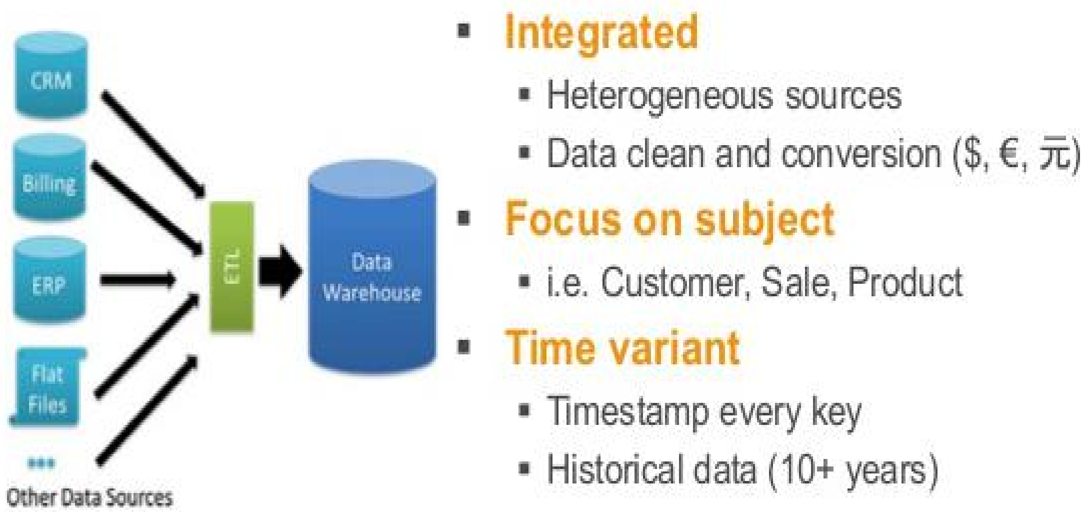

Figure-4.5 Data warehouse design

Source: Ivo Andreev. (October 13, 2014). *Data Warehouse Design and Best Practices*. Retrieved May 18, 2017 from the World Wide Web: https://www.slideshare.net/ivoandreev/data-warehouse-design-and-best-practices?qid=a6a215c1-9282-402d-98fd-98a350735ef1&v=&b=&from_search=1.

4.5 Data mining

The most commonly accepted definition of "data mining" is the discovery of "models" for data(Leskovec, Rajaraman& Ullman, 2014, p.1).

Data mining has several important directions in modeling, like Statistical Modeling, Machine Learning, Computational Approaches to Modeling, Summarization, Feature Extraction and so on.

In the ship risk identification model of this paper, the Data mining is seen as data summarization, which extracts and abstracts data to the algorithm of the model. Figure-4.6 shows the summarization flow of data mining.

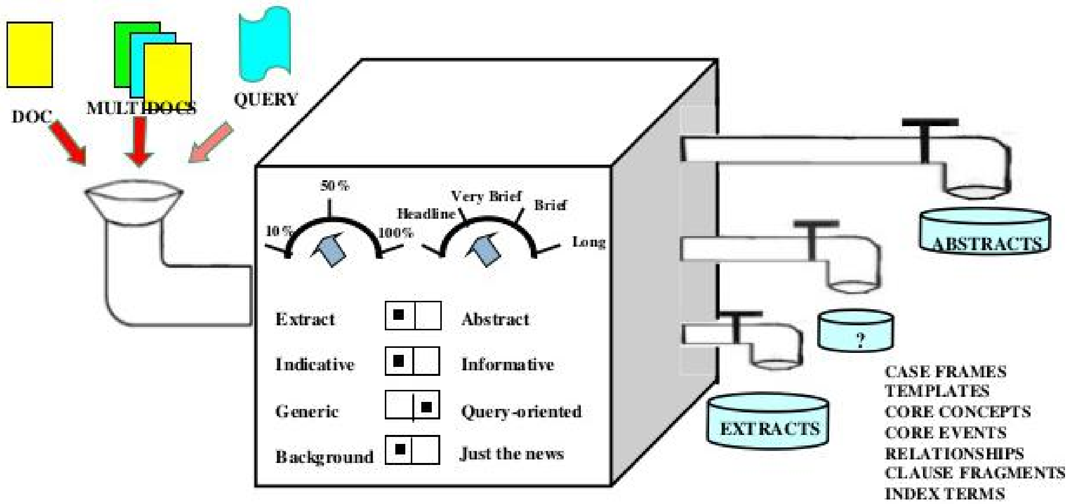

Figure-4.6 Summarization flow of data mining

Source: Chennai. (2015). *Text Data Mining*. Retrieved May 19, 2017 from the World Wide Web: https://www.slideshare.net/

4.6 Algorithm in ship risk identification model

As mentioned earlier, the ship risk identification model has two functions, one is identifying and analyzing the navigational risk based on real time data; the other is assessing the navigational safety based on the amount of historical data. Therefore, there are different algorithms for these two functions.

In this paper, Particle Swarm Optimization(PSO) algorithm is selected to identify and analyze the navigational risk, and BP Neural Network Algorithm is used to assess the navigational safety of Qiongzhou Strait. Chapter 5 and 6 supply detailed introduction of the two algorithms.

Chapter 5 Navigational risk identification based on PSO algorithm and cloud computing

5.1 Algorithm principle

The Particle Swarm Optimization(PSO) algorithm was proposed by Kennedy and Eberhart in 1995, which sees each individual in the swarm as a particle(point) of no mass and size in the search space. The particles has a certain speed(direction, velocity) in the search space, and dynamically adjust the speed according to the motion experience of itself and others. That means each particle continuously up-date the direction and velocity aims to form the positive feedback mechanism of group optimization. In conclusion, the PSO algorithm is the process of searching for the optimal solution of group problem according to the fitness of each particle in the group environment(Kennedy, Eberhart, 1995).

5.2 Basic formula of PSO algorithm

There are *m* individuals of a certain swarm in a *D* dimensional search space, each with three features: the current location, speed and the optimal location. The swarm has certain topology structure, and the individual can interact with another individual of the swarm based on the topology structure. When the algorithm iterates, each individual can update the status according to the information of itself and other particles in the swarm. And the basic formula is like(Liu, 2010, p.34):

Index: $1 \le i \le m$

$1 \le d \le D$

$\vec{x}_i = (x_{i1}, x_{i2}, ..., x_{iD})$

Particle: $\vec{v}_i = (v_{i1}, v_{i2}, ..., v_{iD})$

$\vec{p}_i = (p_{i1}, p_{i2}, ..., p_{iD})$

Particle Swarm: $\vec{p}_g = (p_{g1}, p_{g2}, ..., p_{gD})$

Formula: $v_{id}^{*} = v_{id} + C_1\xi(p_{id} - x_{id}) + C_2\eta(p_{gd} - x_{id})$ ①

$x_{id}^{*} = x_{id} + v_{id}^{*}$ ②

m: Swarm size

x_i: Particle`s dimension in search space

$\vec{v}_i$: Speed vector of individual

$\vec{p}_i$: Location vector of individual

C_1, C_2: Learning Factor, which makes the particle have the learning ability, thus the particle can approach the optimal location in the swarm. In general, the learning factor is 2.

ξ, η: Random numbers between 0-1.

$\vec{p}_g$: Global optimal solution

v_{id}^* : Updated speed of particle

x_{id}^* : Particle`s updated dimension in search space

5.3 PSO algorithm process

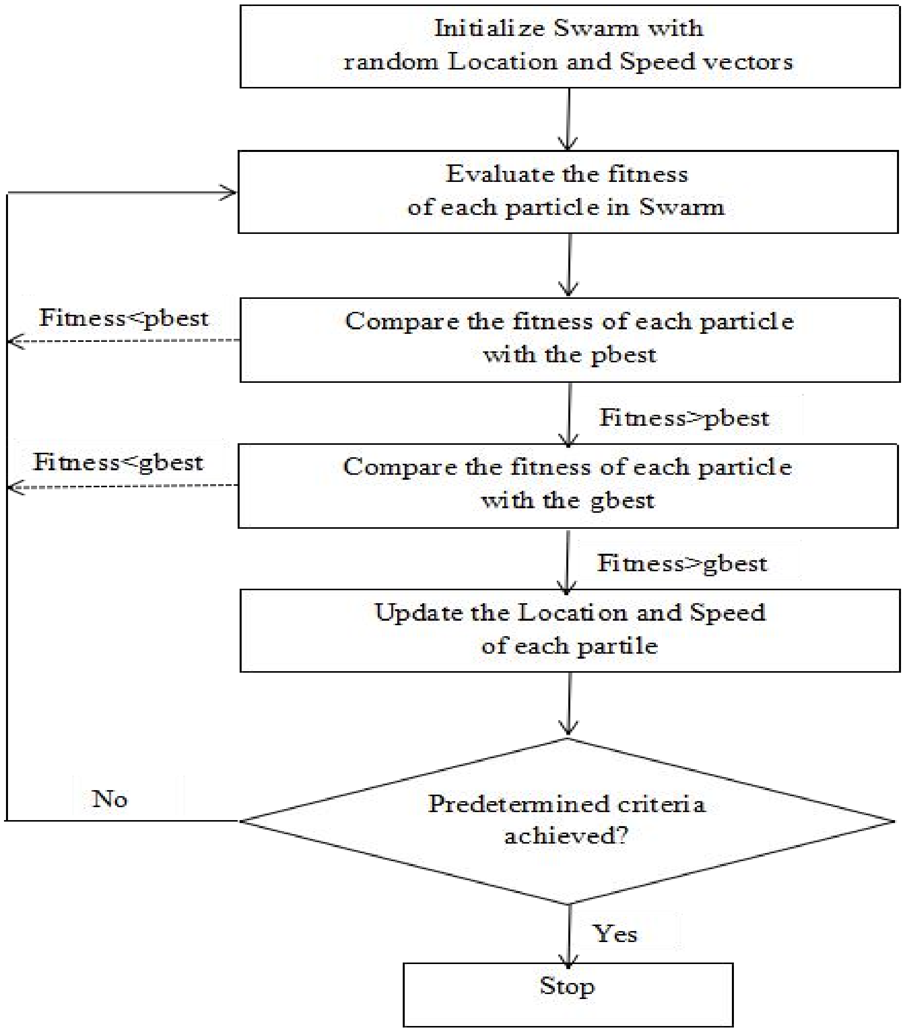

Figure-5.1 PSO algorithm process

Source: Alireza Abdollahi. (2013). A Comprehensive Survey: Applications Of Multi-objective Particle Swarm Optimization(MOPSO) Algorithm. *Transactions on Combinatorics, 13*, 2-1.

Step 1. Initialize swarm with random location and speed. The particle`s location of pbest is the current location, and calculating the pbest. And the gbest is the optimal pbest of the swarm;

Step 2. Evaluate the fitness of each particle;

Step 3. Compare each particle's fitness with the pbest: if the fitness is greater than pbest, adjust the pbest;

Step 4. Compare each particle's fitness with the gbest: if the fitness is greater than gbest, adjust the gbest;

Step 5. Update the speed and location of each particle according to formula ① ②;

Step 6. Start again from Step 2 until the predetermined criteria is satisfied.

5.4 Navigational risk identification in Qiongzhou strait based on PSO algorithm

The key point of PSO algorithm is to determine the pbest and gbest, and then evaluate the fitness of each individual with pbest and gbest and adjust the status of the individual until the swarm achieves optimal solution.

In the navigation risk identification model based on PSO algorithm, the Qiongzhou Strait coverage can be seen as the swarm, and the ship sailing within the Strait can be seen as the particle in the swarm. Based on PSO algorithm process, firstly the fitness of vessels and pbest should be determined. In this paper, the DCPA and TCPA of each vessel are evaluated as the fitness, and the pbest and the gbest are determined by the parameter of the model, which is further determined by the marine environmental circumstance, like the wind, current, navigational density and so forth.

Below is an example:

Figure-5.1 Example ships Data

	Location	Course	Speed
Vessel-1	φ 20^{0}14\`N $\lambda$ 110$^0$38\`E	090^0	9.0knots
Vessel-2	φ 20^{0}18\`N $\lambda$ 110$^0$42\`E	238^0	13.5knots

Source: The author

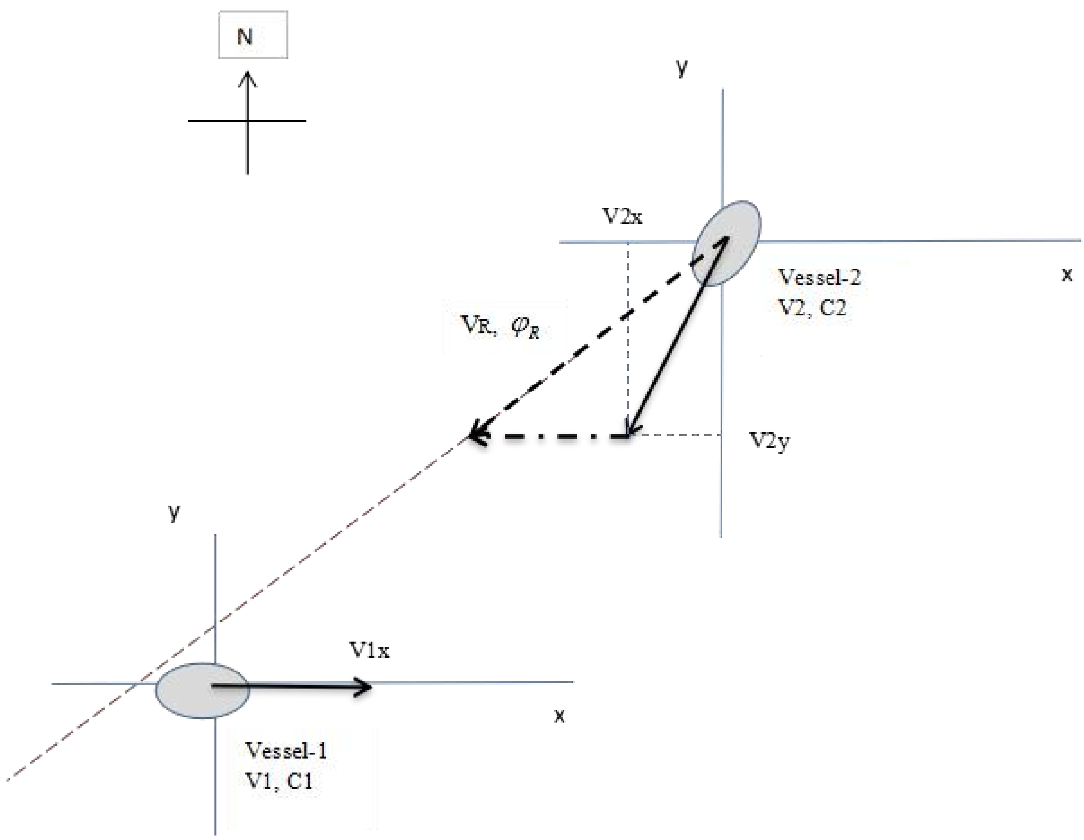

Figure-5.2 A sketch map of the example above
Source: Author

V_1, C_1, Speed and Course of Vessel-1

V_2, C_2, Speed and Course of Vessel-2

V_R, φ_R, Relative Speed and Course of Vessel-2 relative to Vessel-1

5.4.1 Relative speed and relative course of Vessel-2 relative to Vessel-1

$$\begin{cases} V_{x1} = V_1 * \sin C_1 \\ V_{y1} = V_1 * \cos C_1 \end{cases}$$

$$\begin{cases} V_{x2} = V_2 * \sin C_2 \\ V_{y2} = V_2 * \cos C_2 \end{cases}$$

$$\begin{cases} V_{xR} = V_{x2} - V_{x1} \\ V_{yR} = V_{y2} - V_{y1} \end{cases} \Rightarrow \quad V_R = \sqrt{V_{XR}{}^2 + V_{YR}{}^2}$$

$$\varphi_R = \arctan \frac{V_{xR}}{V_{YR}} + \alpha$$

α is the coefficient, which is:

$$\alpha = \begin{cases} 000^0, if: V_{xR} \geq 0, V_{yR} \geq 0 \\ 180^0, if: V_{xR} < 0, V_{yR} < 0 \\ 180^0, if: V_{xR} > 0, V_{yR} < 0 \\ 360^0, if: V_{xR} < 0, V_{yR} > 0 \end{cases} \text{(Tao, 2001, p.59)}$$

So that,

$V_R = 21.66$ knots

$\varphi_R = 250^0 43^`$

5.4.2 Calculation of DCPA and TCPA

$$TB = C_1 + Q$$

$$\cos S = \sin\varphi_1 * \sin\varphi_2 + \cos\varphi_1 * \cos\varphi * \cos D\lambda$$

$$\tan C_0 = \frac{\sin D\lambda}{\cos\varphi_1 * \tan\varphi_2 - \sin\varphi_1 * \cos D\lambda}$$

$$DCPA = S * \sin(\varphi_R - TB - \pi)$$

$$TCPA = \frac{S * \cos(\varphi_R - TB - \pi)}{V_R}$$ (Tao, 2001, p.59)

TB, True Bearing of Vessel-2 relative to Vessel-1;

Q, Relative bearing of Vessel-2 relative to Vessel-1;

C_0, Initial Great Circle Course from Vessel-1 to Vessel-2;

S, Great Circle Distance between Vessel-1 and Vessel-2.

So that,

$S = 5.66$ n.mile

$C_0 = 086^0 37$\`, and TB=$C_0$=$086^0 37$\`

Finally, the DCPA and TCPA can be calculated:

$$DCPA = S * \sin(\varphi_R - TB - \pi) \approx 1.57 \text{ n.mile}$$

$$TCPA = \frac{S * \cos(\varphi_R - TB - \pi)}{V_R} \approx 15 \text{ min}$$

5.4.3 Fitness evaluation

So far there is no standard reference for the evaluation of DCPA and TCPA due to the marine environmental difference among regions and sea areas. However, the customary practices which are suitable in specialized areas are recommended for reference. In Qiongzhou Strait, the navigational circumstance is complex, and there are lots of crossing points along the channel. Thus, the pbest and gbest of navigational risk identification model in Qiongzhou Strait should be formulated cautiously. Table-5.1 below shows the standard reference of DCPA and TCPA in Qiongzhou Strait.

Table-5.2 Standard reference of DCPA and TCPA in Qiongzhou strait

DCPA / TCPA	DCPA≤5 Cables	5 Cable<DCPA<1 N/M	DCPA≥1 N/M
TCPA≤5 min	Collision risk	Assess collision situation	Assess collision situation
TCPA >5 min	Assess collision situation	Assess collision situation	Clear

Source: Qiongzhou Strait VTS Center. (2002). *VTS Operator Manual Guideline*. Unpublished internal working documents, Hainan MSA.

In the above example, the fitness is :

DCPA=1.57 N/M> 1 N/M

TCPA=15 min> 5 min

So that the navigational situation of the example is safety and clear, and the model will provide feedback on the assessment and thus continuously assess other vessels in the Strait.

5.5 PSO algorithm based on cloud computing task scheduling

Cloud computing utilizes virtualization technology producing different types of virtual machines, which can carry out users` task, and improving the resource utilization of Hardware Notes(Wang, Zhang, 2013).

In the navigational risk identification in Qiongzhou Strait, the cloud computing center can distribute virtual machines to the assessment of each particle according to the preset parameter, instead of each virtual machine performs matching calculation, which can decrease the calculation quantity dramatically. Furthermore that can improve the performance of PSO algorithm in identifying the navigational risk in Qiongzhou Strait.

Chapter 6 Navigational safety assessment based on BP neural network algorithm and big data technology

6.1 Structure of Back Propagation(BP) neural network

In the 1980s, the book of D.E. Rumelhart and J.L.McClelland, which is entitled *Parallel Distributed Processing: Explorations in The Microstructure of Cognition*, put forward the back propagation algorithm used to train multilayer neural network. The topology of BP neural network comprises input layer, hidden layer and output layer(Rumelhart, McClelland, 1986, Vol. 1).

BP neural networks` basic unit of work is BP neurons, whose basic principle is mimicking the weighting, activation and sum function of biological neuron. Figure-6.1 is the model of neuron.

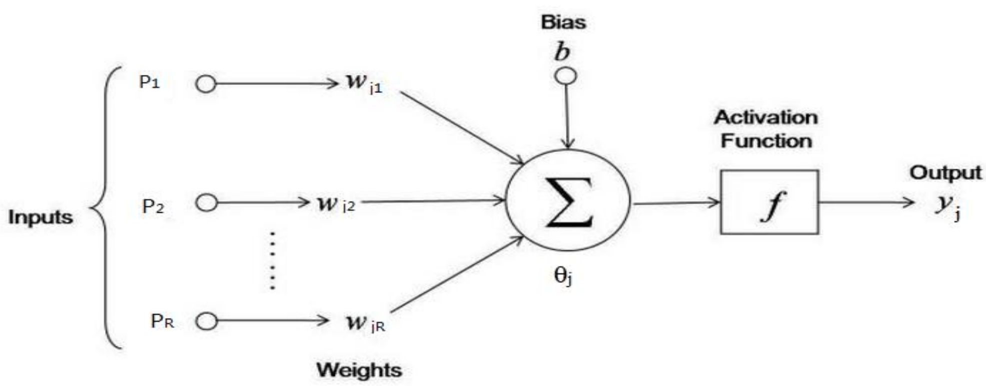

Figure-6.1 Model of Neuron
Source: Martin T. Hagan, Howard B. Demuth& Mark H. Beale. (2002). *Neural Network Design*. Beijing: China Machine Press.

The output of Neuron is: $y_j = f(\sum_{i=1}^{R} \omega_{ji} \cdot P_i - \theta_j)$, and

$P_1, P_2, ..., P_i ... P_R$, the input to the neuron 1, 2, ...,i...R;

$\omega_{j1}, \omega_{j2}, ..., \omega_{ji} ... \omega_{jR}$, the connection weights between neuron 1, 2, ...,i...R and the neuron j;

θ_j , the threshold;

$f(.)$, the transfer function;

y_j , the output of neuron j.

Based on the model of neuron, the structure of Back Propagation Neural Network can be designed as in figure-6.2 below.

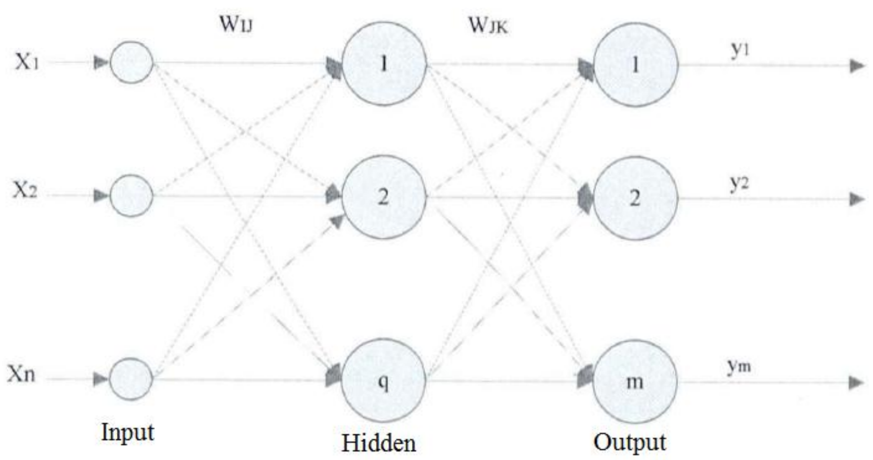

Figure-6.2 Structure of BP neural network

Source: Sun Zhonghua. (2012). *Research on PSC Targeting Model Based on Intelligent Optimization Algorithms*. Unpublished doctor`s thesis, Dalian Maritime University, Dalian, China.

Actually, BP neural network has one or more Hidden layers, and the neuron in Hidden layer can select and apply different transfer functions, like Sigmoid Function, which can actively learn the linear and nonlinear relationship between input and output(Rojas, 1996, pp.151-154). However, the linear transfer function in output layer is to broaden network output. Besides, the transfer functions in BP neural network are Differentiable Monotone Increasing Functions:

6.2 BP learning algorithm(Shen, Li& Xuan, 2008)

BP neural networks learning algorithm corrects the weights and threshold of neurons along the negative gradient direction of the performance function. The formula is :

$x_{k+1} = x_k + \alpha_k \cdot g_k$, and,

x_k, Current Weights and Threshold Matrices;

g_k, Current Performance Function Gradient;

α_k, Learning Rate.

To assume that the BP neural networks in this paper have three layers: there are ***n*** neurons in input layer, ***q*** neurons in hidden layer, and ***m*** neurons in output layer:

ω_{ij}, the network weight between the input node and the hidden node;

V_{jk}, the network weight between the hidden node and the output node;

θ_j, the threshold of the hidden layer;

γ_k , the threshold of the output layer;

(P,T) is the set containing ***N*** samples, P is arbitrary one of the ***N*** samples;

x_i is the input;

Sample input is $P=(P_1,P_2,...P_n)$;

Desired output is $T=(t_1,t_2,...t_n)$, and,

$i=(1,2,...n)$, $j=(1,2,...q)$, $k=(1,2,...m)$, then the BP neural networks algorithm is:

6.2.1 Forward Propagation computation output process:

① Initializing the network weights ω_{ij}, V_{jk}; and the threshold θ_j, γ_k;

② Inputting sample training set (P,T);

③ Computing the output of neuron ***j*** in hidden layer: $O_j = f(\sum_{i=1}^{n}\omega_{ij}x_i-\theta_j)$;

④ Computing the output of neuron ***k*** in output layer: $O_k = f(\sum_{j=1}^{q}\omega_{jk}O_j-\theta_k)$.

6.2.2 Back Propagation correct Weights and Threshold process:

① Computing the sample error

Sample Error $\boldsymbol{p}$: $E^p = \sum_{k=1}^{m} \frac{1}{2}(t_k^p - O_k^p)^2$; Sample Error: $E_T = \sum_{p=1}^{N} E^p$;

② Generalization error of output layer: $d_k = O_k(t_k - O_k)(1 - O_k)$;

③ Generalization error of hidden layer: $e_j = O_j(t_k - O_j)[\sum_{k=1}^{m} d_k \cdot V_{jk}]$;

④ Correcting the weight and threshold:

$$V_{jk}(N+1) = V_{jk}(N) + \alpha \cdot d_k \cdot O_j$$

$\gamma_k(N+1) = \gamma_k(N) + \alpha \cdot d_k$, therein, $0 < \alpha < 1$

$$\omega_{ij}(N+1) = \omega_{ij}(N) + \alpha \cdot e_j \cdot P_i$$

$\theta_j(N+1) = \theta_j(N) + \beta \cdot e_j$, therein, $0 < \beta < 1$

⑤ Return to ③ and ④ of Forward Propagation process and recalculate until the sample error E_T is less than the set value or achieving the maximum training times, then the training process ends.

6.3 Navigational safety assessment based on BP neural networks algorithm

6.3.1 Determine the dimension of network input and output vectors

6.3.1.1 Determine the dimension of network input vectors

In the Model of Navigational Safety Assessment, the assessment condition attributes include SHIP`S NAME, DRAFT, MMSI, DANGEROUS DARGO, LOCATION, SPEED, COURSE, LAST PORT OF CALL, DESTINATION, VOYAGE PLAN, CHART DATA, METEOROLOGICAL DATA and HYDROLOGICAL DATA. As the network input, all the data above should be transferred into numerical index in order to be utilized by the algorithm.

6.3.1.2 Determine the dimension of network output vectors

The network output vector of the Model in this paper is one-dimensional, which is the assessment value of the navigational safety.

6.3.2 Determine the transfer and training functions

In the Model of Navigational Safety Assessment in Qiongzhou Strait, according to the data type of input and output, the transfer function in the algorithm is Logsig function, and the training function is Traingdm function.

Logsig transfer function takes the input and squashes the output into the range 0 to 1, which is commonly used in multilayer networks trained by the back propagation algorithm(Dorofki, 2012, p.39).

Batch Gradient Descent with Momentum(Traingdm) can provide faster convergence, which is steepest descent with momentum. Both local gradient and the error surface of recent trends can be reflected by momentum. Just as the working mechanism of a low-pass filter, momentum will not response to the minor error. If the momentum is deducted, the network may stuck by a shallow local minimum. And just in contrast, when the momentum is counted, suck kind minimum could be ignored by the network.(Demuth, Beale, 2004, p.511).

6.3.3 Determine the number of Hidden Layer Neurons

There are several weights in the hidden layer neuron, and each of them is a parameter to improve the mapping ability of the network. If the number of hidden layer neurons is less, then the ability of the network to obtain sample information is poor, so that which cannot reflect the inherent law of the sample; however, if the number of hidden layer neurons is more, the irregular content will be remembered, and that will lead to the training time will be increased. In general, the number of neurons depends on the amount, redundancy, regularity and complexity of the sample. In this paper, the number of hidden layer neurons would be calculated based on the following empirical formula:

$j = \sqrt{m+n} + (1 \sim 9)$ (Sun, 2012, p.66)

Therein, j, the number of hidden layer neurons

m, the number of input layer notes

n, the number of output layer notes

The MATLAB software neural network toolbox can supply specialized functions to train sample. In this paper, the model of navigational assessment for Qiongzhou strait has already be introduced in detail. And the process of the algorithm is also explained in detail too. Limited to the time and space, the training process will not introduced in this paper.

Chapter 7 Conclusion

As an important support of the global economy, the significance of shipping industry is indubitable. However, the issue of safety in shipping industry has always been a topic of public discussion. From international organizations, national departments to academic institutions, the relevant research about safety has never stopped. In China, as the administration in charge of safety at sea, the MSA treated the safety with even more seriousness.

The author of this paper has worked in Qiongzhou strait VTS Center for several years. Based on the practical work experience and the summary of the insufficiency in the current working model of Qiongzhou Strait VTS, this paper established the navigation safety assessment model base on big data and cloud computing technology for Qiongzhou strait, which aims to improve the efficiency of risk identification of VTS operators, ultimately improve the safety level in Qiongzhou strait.

The modeling process is based on the data mining method in big data technology, including the selection of data sources, data pretreatment, task-relevant data selection and data mining, and lastly utilizing intelligence algorithm to assess the navigation safety and identify the navigation risk, which are the two functions of the model.

There are two algorithms used in the model, one is Particle Swarm Optimization(PSO) algorithm, and the other is BP Neural Network. PSO algorithm treats the each vessel in Qiongzhou strait as particle with location and vector speed, via calculating the fitness(DCPA, TCPA) of each vessel and assessing the fitness with pbest and gbest, so that identify the navigation risk in Qiongzhou strait; BP neural network is another different algorithm, which estimates the assessment model, and utilizes specialized software and functions to train the sample from data warehouse, aims to obtain an assessment mechanism in order to assess the navigation safety in Qiongzhou strait.

In conclusion, the modeling process in this paper tend to be more theoretical exploration. Although at the level of theory, the model may improve the safety level, however, the algorithms for the risk identification and safety assessment need further optimization based on practical application in the routine work of VTS center. The model of navigation risk identification and navigation safety assessment based on big data and cloud computing technology for Qiongzhou strait has great research meaning, and it has great research space and significant practical application value.

Reference

Alireza Abdollahi. (2013). A Comprehensive Survey: Applications Of Multi-objective Particle Swarm Optimization(MOPSO) Algorithm. *Transactions on Combinatorics, 13*, 2-1.

Anhai Doan, Alon Halevy& Zachary Ives. (2012). *Principles of Data Integration*. Waltham, USA: Morgan Kaufmann Press.

Association for Computing Machinery. (2010). A view of cloud computing. New York, USA: *Communication of the ACM*, ISSN: 0001-0782, EISSN: 1557-7317.

Chen Gongmeng, Xu Chengzhong. (2015). *Introduction To Big Data*. Beijing: Tsinghua University Press.

Chennai. (2015). *Text Data Mining*. Retrieved May 19, 2017 from the World Wide Web: https://www.slideshare.net/

China Oceanic Information Network. (2017). *Hydrological Information*. Retrieved May 11, 2017 from the World Wide Web:
http://www.coi.gov.cn/data/jichu/

China Oceanic Information Network. (2017). *Meteorological Information*. Retrieved May 12, 2017 from the World Wide Web:
http://www.coi.gov.cn/data/jichu/

Erhard Rahm. (2000). *Data Cleaning: Problems and Current Approaches*. Germany, University of Leipzig: Author.

Extract-Transform-Load (ETL) Technologies– Part 1. (December 31, 2012). *DB. Best Technologies*. Retrieved May 15, 2017 from the World Wide Web:
https://www.dbbest.com/blog/extract-transform-load-etl-technologies-part-1/

Fan Chongjun, Liu Cheng& Huo Liangan. (2016). *Analysis And Application Of Big Data*. Shanghai: Lixinapn Press.

Fazil Ahmed. (2017). *Big Data, What`s The Benefit For Transportation*. Retrieved May 20, 2017 from the World Wide Web:
http://www.wsp-pb.com/en/WSP-UK/Who-we-are/Newsroom/features/Big-Data-whats-the-benefit-for-transportation-/.

Howard Demuth, Mark Beale. (2004). *Neural Network Toolbox: For Use With MATLAB®*. The MathWorks, Inc: User`s Guide Version 4.

IBM Big Data & Analytics Hub. (2012). *Four V`s Of Big Data*. Retrieved May 2, 2017 from the World Wide Web:
http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Ivo Andreev. (October 13, 2014). *Data Warehouse Design and Best Practices*. Retrieved May 18, 2017 from the World Wide Web:
https://www.slideshare.net/ivoandreev/data-warehouse-design-and-best-practices?qid=a6a215c1-9282-402d-98fd-98a350735ef1&v=&b=&from_search=1.

Jack Rosenblum. (2012). *Key Features Of Cloud Computing*. Retrieved May 12, 2017 from the World Wide Web:
https://cloudtweaks.com/2012/09/key-features-of-cloud-computing/.

James Kennedy, Russell Eberhart. (1995). Particle Swarm Optimization. *Proceeding of IEEE International Conference on Neural Networks, IV*. Piscataway, NJ: IEEE Service Center, 1995.

John Garmany, Jeff Walker& Terry Clark. (2005). *Logical Database Design Principles*. London, UK: Taylor& Francis Group, CRC Press.

Jure Leskovec, Anand Rajaraman& Jeffrey David Ullman. (2014). *Mining Of Massive Datasets*. University of Cambridge, UK: Cambridge University Press.

Kevin Jessop. (November 14, 2016). *What Is the Impact of Big Data in the Transportation & Supply Chain Industries*. Retrieved May 10, 2017 from the World Wide Web:
http://cerasis.com/?s=What+Is+the+Impact+of+Big+Data+in+the+Transportation+%26+Supply+Chain+Industries

Krešimir Popović, Zeljko Hocenski. (2010). Cloud Computing Security Issues and Challenges. *Proceedings of the 33rd International Convention*. Opatija, Croatia, MIPRO.

Larissa T. Moss, Shaku Atre. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle For Decision-support Applications*. Indiana, USA: RR Donnelley Crawfordsville Press.

Liu Bo. (2010). *Particle Swarm Optimization and Application*. Beijing: Publishing House Of Electronics Industry.

Maamar Ferkoun. (2014).*Cloud Computing And Big Data: An Ideal Combination*. Retrieved May 16, 2017 from the World Wide Web: https://www.ibm.com/blogs/cloud-computing/2014/02/cloud-computing-and-big-data-an-ideal-combination/.

Maritime Safety Administration. (2017). *Chart15770: Qiongzhou Strait*. Shanghai: Donghai Navigation Safety Administration(DNSA) MOT.

Martin T. Hagan, Howard B. Demuth& Mark H. Beale. (2002). *Neural Network Design*. Beijing: China Machine Press.

Mohammad Dorofki. (2012). *Comparison Of Artificial Neural Network Transfer Functions Abilities To Simulate Extreme Runoff Data*. Proceeding in International Conference on Environment, Energy and Biotechnology 2012. Singapore: IACSIT Press.

Peter Mell, Timothy Grance. (2011). *The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology*. Department of Commerce, U.S.A.: Special Publication 800-145.

Qiongzhou Strait VTS Center. (2002). *VTS Operator Manual Guideline*. Unpublished internal working documents, Hainan MSA.

R. Rojas. (1996). *Neural Networks: A Systematic Introduction*. Berlin, Heidelberg, Germany: Springer Press.

Rumelhart, J.L.McClelland. (1986). *Parallel Distributed Processing: Explorations in The Microstructure of Cognition. Vol. 1*. UK: Cambridge, MA: MIT Press.

Shen Chunming, Li Xinghua& Xuan Dayang. (2008). Application of BP Neural Networks in Safety Assessment. *Industrial Safety and Environmental Protection*, 08, 15-11.

SQL Power Group. (2017). *What Is Business Intelligence*. Retrieved May 5, 2017 from the World Wide Web: http://www.sqlpower.ca.

Sun Zhonghua. (2012). *Research on PSC Targeting Model Based on Intelligent Optimization Algorithms*. Unpublished doctor`s thesis, Dalian Maritime University, Dalian, China.

Tamraparni Dasu, Theodore Johnson. (2003). *Exploratory Data Mining and Data Cleaning*. Hoboken, New Jersey: John Wiley & Sons, Inc. Press.

Tao Yuan. (2001). *Application Research of ship automatic collision avoidance decision system based on AIS*. Unpublished master`s thesis, Shanghai Maritime University, Shanghai, China.

Traffic Separation Schemes And Reporting System In Qiongzhou Strait 2009, MSA, People's Republic of China, (2009).

Viktor Mayer-Schönberger， Kenneth Cukier. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. New York： Houghton Mifflin Harcourt Publishing Company.

Wang Bo, Zhang Xiaolei. (2013). Task scheduling algorithm based on Particle Swarm Optimization Genetic Algorithms in cloud computing environment. *Computer Engineering and Applications*, 13, 51-6.

W.H. Inmon, Derek Strauss& Genia Neushloss. (2008). *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. Burlington, USA: Elsevier Inc, Morgan Kaufmann Press.