공학석사 학위논문

기계학습 기법을 이용한 바지형 선박의 횡동요 RAO 예측

Prediction of RAO in Barge Ships
Using Machine Learning Method

지도교수 조 효 제

2020년 2월

한국해양대학교 해양과학기술전문대학원

해양과학기술융합학과

임 재 환

본 논문을 임재환의 공학석사 학위논문으로 인준함

위원장 이 성 욱 (인)

위 원 조 효 제 (인)

위 원 최 민 주 (인)

2019년 12월

한국해양대학교 해양과학기술전문대학원

Contents

List of Tables

List of Figures

Prediction of RAO in Barge Ships Using Machine Learning Method

Jae Hwan Lim

Department of Convergence Study on the Ocean Science and Technology,
Ocean Science and Technology School(OST),
Korea Maritime and Ocean University

Abstract

Recently, Artificial Intelligence(AI) technology has been applied in various industries due to increased interest in the fourth industrial revolution. However, research related to AI is relatively lacking in shipbuilding and maritime industries because of the characteristic of the industrial structure that information disclosure is limited. But shipbuilding and maritime industries are a complex industry in which various sectors are merged, and large amounts of data are generated. In addition, various forms of data are generated by the shipbuilding and marine industries in their respective processes, including design, construction, maintenance and operation. Therefore, it is believed that AI technology can be applied to optimizing design, efficiency of maintenance, and stability evaluation in operation to produce significant results. Many existing studies have mainly been done to improve efficiency in terms of production management and to optimize ship operation. In this paper, however, a machine learning prediction model was built for predicting the lateral homologous RAO from the design optimization perspective of ships. Data used in the paper were for barge types registered in advance, and RAO was generated for each vessel in an in-house code using the three-dimensional singularity distribution method. In addition, Python and Tensorflow was used to build the prediction model, and statistical techniques were used to measure the results of changing the hyper parameters of the prediction model to evaluate the accuracy of the results. Thus, unlike previous studies targeting specific targets, this paper has a difference in that it has targeted several vessels. Finally, the purpose of the study is to identify the approximate RAO for vessels where the drawing of ships does not exist, and to improve the efficiency of the modelling and analytical processes necessary to obtain RAO. Furthermore, it is thought that it will be possible to develop into a study that will help assess the stability of autonomous driving vessels in the future, given that the stability of vessels with specific dimensions can be identified.

KEY WORDS: Machine Learning; Barge Type Ship; Roll; Response Amplitude Operator(RAO); Deep Neural Network(DNN); Root Mean Square Error(RMSE) ; Standard Deviation(SD)

기계학습 기법을 이용한
바지형 선박의 횡동요 RAO 예측

임재환

한국해양대학교 해양과학기술전문대학원
해양과학기술융합학과

초 록

최근 4차 산업혁명에 대한 관심의 증가로 인공지능 기술이 다양한 산업에서 적용되고 있다. 그러나 조선 및 해양산업에서는 정보공개가 제한적이라는 산업구조 특성상 인공지능과 관련된 연구가 상대적으로 부족한 실정이다. 하지만 조선 및 해양산업은 다양한 분야가 융합된 복합적인 산업으로 많은 양의 데이터가 생성되는 산업 중 하나이다. 뿐만 아니라 조선 및 해양산업은 설계, 건조, 유지 보수, 운항 등 각각의 과정에서 갖가지 형태의 데이터가 발생한다. 따라서 인공지능 기술을 설계 최적화, 유지 보수의 효율성, 운항 시 안정성 평가 등에 적용하여 유의미한 결과를 도출할 수 있을 것이라 판단된다. 기존의 많은 연구들은 생산관리 측면의 효율을 개선하기 위한 연구와 선박의 운항의 최적화를 위한 연구가 주를 이루어왔다. 그러나 본 논문에서는 선박의 설계 최적화관점에서 횡동요 RAO를 예측을 위한 기계학습 예측모델을 구축하였다. 논문에 사용된 데이터는 선급에 등록된 바지

형 선박을 대상으로 하였으며, 각각의 선박에 대하여 3차원 특이점 분포법을 사용한 In-house code로 선박별로 RAO를 생성하였다. 그리고 예측모델을 구축함에 있어 Python의 Tensorflow를 사용하였으며, 결과의 정확도 평가를 위하여 예측모델의 하이퍼 파라미터를 변경한 결과들에 대해서 통계적인 기법들을 평가 지표로 사용하였다.

따라서 본 논문은 특정 대상을 타겟으로 하는 기존의 연구들과는 달리 여러 척의 선박을 대상으로 하였다는 점에서 차이가 존재한다. 최종적으로 연구의 목적은 선박의 도면이 존재하지 않는 선박에 대한 대략적인 RAO를 파악함과 동시에, RAO를 구하는데 있어 필요한 모델링 과정과 해석 과정에 대한 효율성 개선에 목적이 있다. 더 나아가 특정 제원을 가지는 선박의 적재상태에 따른 안정성을 파악할 수도 있다는 점에서 향후 자율운항선박의 항행 중 안정성 평가에 도움이 되는 연구로 발전할 수 있을 것이라 판단된다.

KEY WORDS: Machine Learning 기계학습, Barge Type Ship 바지형 선박, Roll 횡동요, Response Amplitude Operator(RAO) 응답진폭함수, Deep Neural Network(DNN) 심층신경망, Root Mean Square Error(RMSE) 평균 제곱근 오차, Standard Deviation(SD) 표준편차

Chapter 1 Introduction

1.1 Research Background

1.1.1 The Rolling of a Ship

The ships to be influenced by the external force in the marine environment have 6 degrees of freedom movement. Fig 1.1 is a picture of a ship's rolling movement. These rolling is a factor that determines the comfort, stability and working environment of a person aboard. In addition, a rolling is related to marine accidents such as the overturning of a ship, and accidents caused by the loss of a stability result in more physical and human damage than accidents caused by the ship's engine failure.

Fig. 1.1 Rolling of a ship

In order to prevent the physical and human damages, it is required to understand the characteristics of a ship in operation for its stability. The motion of a ship exposed to irregular external forces is determined by Wave energy spectrum and Response Amplitude Operator(RAO), unique response characteristics of a structure. Therefore, understanding the response by finding out the characteristics of RAO of a ship can be a way to prevent marine accidents.

The existing methods for obtaining RAO include experimental methods and computer analysis simulations. In case of obtaining RAO by an experiment, the difficulties are made due to various constraints such as experimental models, equipments and experimental environments etc. In order to obtain RAO by using computer simulations, the following three steps are required.

- 1st Step: The modeling process of configuration information of a ship. This is a step that generates the configuration information as a previous processing step for its analysis simulation.
- 2nd Step: The process of setting the conditions of a ship. This is a step in which its center of gravity and inertia radius etc. are entered in its configuration information in consideration of the loading state etc.
- 3rd Step: The process of analyzing the motion response in the frequency domain. As a final conclusion, in the 3rd Step, it is possible to get results such as Hydrostatic & Hydrodynamic Value and RAO by each external force direction etc.

However, the information on the configuration of a ship is not easily obtainable and there is a case where the medium and small ships do not have design drawings. And, in case a change in drawing occurs, an inefficiency will be a problem to be solved as repeating the above three steps is required. In addition, the commercial tools to be used for the simulations vary in proficiency depending on its user, which affects the reliability of its results.

1.1.2 Artificial Intelligence (AI)

As more people are interested in The Fourth Industrial Revolution, big data and machine learning have become hot issues in the overall industrial sector.

First of all, big data refer to a technology that analyzes structured and unstructured data by setting the structured data in text format or unstructured data such as picture, video, voice, and location beyond analysis of the existing database management tools.

The big data are used in many areas such as bio-industry, social networking, production, finance and telecommunications etc. For example, 'Google' predicts the spread of the flu through a big data-based system called 'Flu Trend', and one of korean companies, NC soft applies fraud detection algorithms using customer data analysis systems to monitor illegal activities in games.

Fig. 1.2 illustrates a concept of Artificial Intelligence(AI). A machine learning is a field of an AI, which means developing algorithms and technologies that enable computers to learn. If it is explained in full detail, this technique is used to identify rules from a certain amount of data and use them for classification, numerical prediction and grouping etc. Amazon, a world's leading online shopping mall, uses machine learning to create the optimal shopping environment for its customers. Also Tesla, an electric vehicle maker, is using machine learning techniques for fully automatic driving. Creating new values from a certain amount of data by using machine learning and artificial intelligence is a global phenomenon. In this way, regardless of a field, a research is made for an artificial intelligence technology as it is fused with various industries.

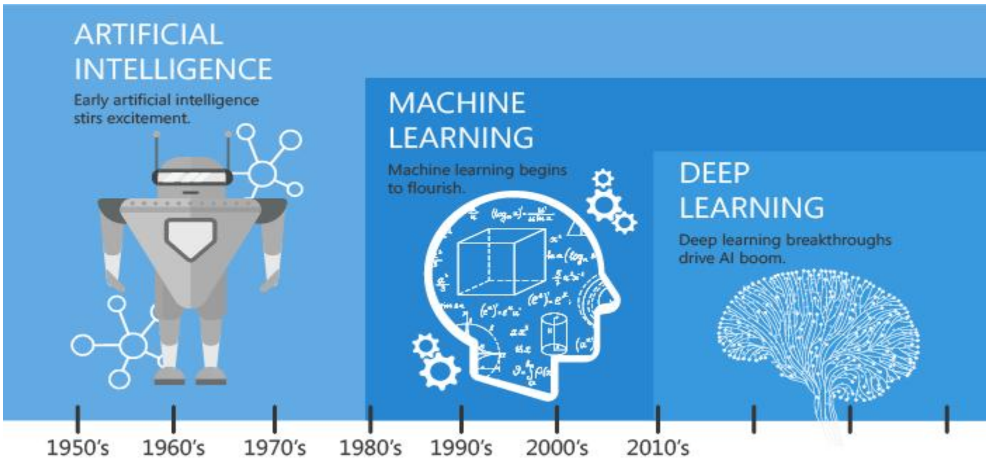

Fig. 1.2 Artificial Intelligence

However, as it is difficult to secure data because of the characteristics of shipbuilding and marine industries, a restrictive research is made. Under these limitations, in this research, it is judged that the existing restrictions will be reduced by developing an accurate prediction model for the motional characteristics of a ship as a machine learning technique is used.

1.2 The Trend of a Research

A variety of researches have been conducted as AI technology is applied to a shipbuilding and marine industry. As for a production management, Ham (2016) researched a prediction for the lead time of a supply in the consideration of the equipment's specifications and supply routes of fittings by using data mining techniques. And, Kim (2018) created and verified a predictive model for the lead time of a production by considering the features of blocks and piping materials in use of the data of a shipyard.

From the viewpoint of operating a ship, Park et, all (2004) assessed the stability of a disturbance occurring during the operation of a particular ship by using the 3D panel method. And, a research was made for a system to evaluate the optimal navigational paths by setting the kinematic of the body of a ship as variables.

Kang et al. (2012) developed a prediction system for the motion of a floating body by using an artificial neural network to predict, in real time, the response of a floating body caused by nonlinear waves.

And, Kim et al. (2018) conducted a research to predict the rolling by using variables of a 9600TEU container ship in operation. Also, Kim (2019) developed a predictive model for the fuel consumption of a ship based on the actual data of a ship in operation to make a model for giving support to make decisions on the abnormal states of equipments for a ship in navigation. In a similar research, Jeon (2019) created a model for predicting the fuel consumption of a ship by using a meta-model in which three machine learning models were combined. If the research trend which has progressed till now is reviewed, it can be classified into three fields. In addition, a variety of studies are being conducted in relation to various learning models. Anton(2016) conducted a study on a multi-output model with many neurons placed on the output layer, also Stathakis (2009) conducted a study on the number of hidden layers and the optimization of the number of neurons in the hidden layer. As such, research on optimization of models is on a trend.

The first is a research based on the viewpoint of a production management. A research has been conducted to reduce production lead time by integrating AI technology with the variables of the production process in use of the data obtained from actual shipyards. The second is a research on the operation of a ship. A research has been conducted to predict the fuel consumption rate of a ship. The third is a research on the movement of a ship. Similarly, a research has been conducted to predict the motion response of a ship in real time. In conclusion, a research has been made mainly on predictions such as lead times and fuel consumption rates. And, as for the movement of a floating body, a research based on a time series has mainly been made.

1.3 The Purpose of a Research

In this research, unlike the previous researches that targeted specific ships, the barge-type ships with various specifications were targeted. For the purpose of understanding the characteristics of the existing rolling, in order to cope with the restrictions of the analytical process for a ship, a research was made for a ship by integrating its motion characteristics with a machine learning. Therefore a research was made of RAO for ships with various specifications and RAO for a specific ship was predicted. The high-prediction model through this research will help to identify RAO for the rolling of small and medium ships without their design drawings. In addition, the modeling process, which was preceded in the analysis process, will be omitted. And, the problems resulting from its proficiency, which may occur in the use of an analytical simulation, will be solved. Furthermore, if data on many ships with various loads are gathered, it is expected that this research can be used as an early research for assessing the stability of a ship depending on its state, in the on-board system of a ship in navigation.

1.4 The Components of a Research

The paper consists of five chapters. Chapter 1 focuses on the background, purpose and necessity of a research. In chapter 2, a description is made of a theory on the general matters of a machine learning such as single layer perceptron, multi-layer perceptron and the learning process of a learning model in relation to an artificial neural network. The full-scale research process is mentioned, from chapter 3, and a description is made of the contents of the data gathering, the composition of an artificial neural network and the accuracy evaluation index. Chapter 4 has the contents in relation to the results of a learning as it includes the process for obtaining an optimal learning model by changing the number of data and the variables for hidden layers. Overall, the RAO data to be predicted by the optimal learning model can be compared with the existing RAO data. As chapter 5 contains a summary of the results of a learning in chapter 4, it ends with the contents for future tasks.

Chapter 2 Neural Network Algorithm

2.1 Machine Learning

A machine learning is a science of programming computers to learn from data. Fig. 2.1 shows a kind of machine learning. A machine learning is classified into a supervised learning, an unsupervised learning and a reinforcement learning according to yes or no of a label in the data. A supervised learning is divided into a classification and a regression depending on what is predicted as a result value. The classification is predefined and predicts one of several class labels that is likely to be output. A classification is classified into a binary classification that classifies as Yes or No, and a multi-class classification that classifies as three or more classes. Regression is a model that predicts a continuous number or floating point. The annual income forecast is also a regression model, given the level of education, age, and residence of a person. The difference between a regression being distinguished from a classification is continuity. There is continuity between the expected output values in the regression model.An unsupervised learning means a way computers produce their own results without human intervention, and its representative example is a clustering. A reinforcement learning means a learning to reinforce a behavior in the direction of the current behavior or in its opposite direction through a reward and a punishment. In other words, it means a technique of taking a measure or an action to maximize a reward among selectable actions through the recognition of a current state.

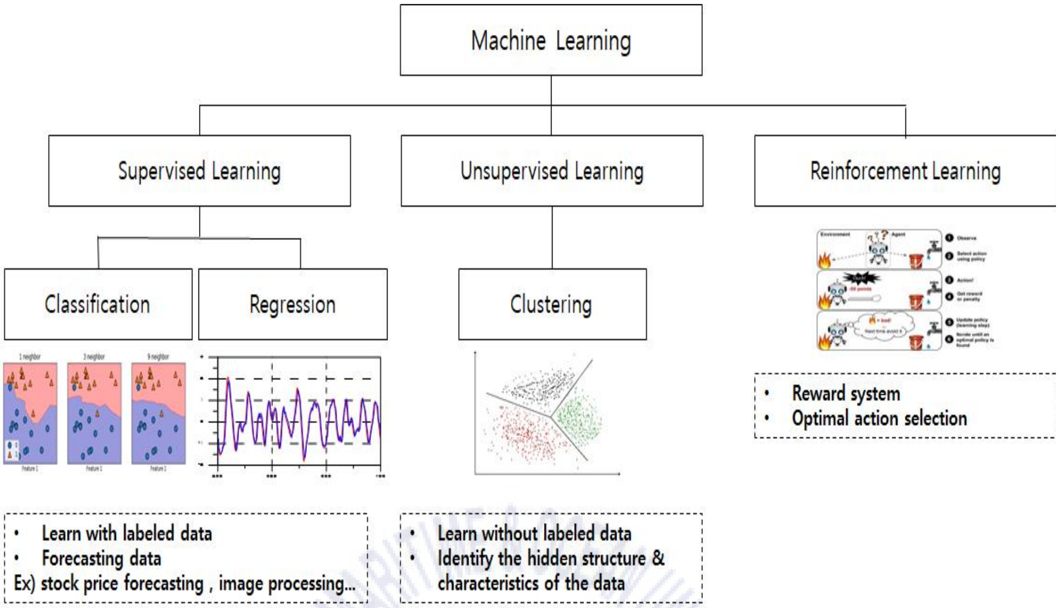

Fig. 2.1 Kind of machine learning

2.2 Artificial Neural Network

Fig. 2.2 is a Neuron and Artificial Neural Network(ANN) illustrating the structure. ANN is a mathematical modeling of a learning method of a brain. It is similar to a biological brain neuron. The neurons are connected by links, and each link has a numerical weight associated with them. So the weights represent each neuron input strength, i.e., importance, as a basic means for long-term memory in the ANN. Neural networks learn by repeatedly adjusting their weights. And, as they give and take information through a synapse that connects one neuron with another neuron in a brain, the information is transmitted and a learning is made. To put it simply an ANN refers to a model in which artificial neurons expressing a learning method of a brain mathematically have problem solving abilities by adjusting weights through a learning.

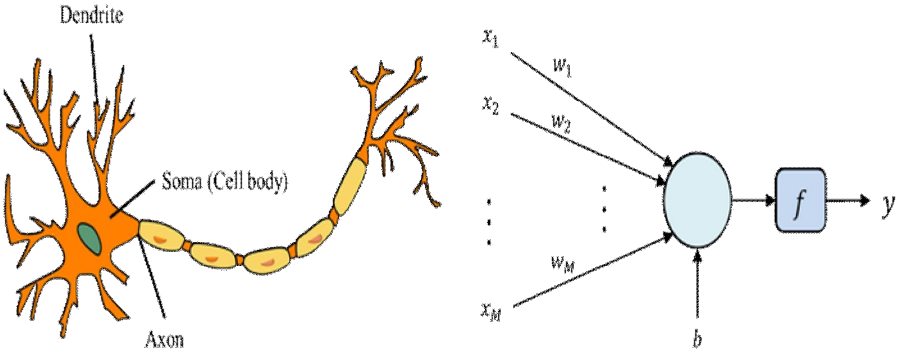

Fig. 2.2 Neuron structure(Left) & Artificial neural network structure(Right)

2.2.1 The Structure of an Artificial Neural Network

An artificial neural network is composed of an input layer, a hidden layer and an output layer. As many neurons as input variables exist in an input layer. In its hidden layer, there are neurons which are generated as the neurons of its input layer combined with the weights. In general, constructing an artificial neural network requires determining how many neurons will be used in the hidden layer and how the neurons will be connected in the neural network. That is, the structure of the neural network is chosen first, and which learning algorithms are used. Finally, neural network training takes place. In the end, its output layer contains neurons which are generated as the neurons in its hidden layer are combined with the weights. The number of neurons in an output layer depends on the dependent variables. And, the neurons which are present in the hidden layer and the output layer perform a function of adding an input value and a weighted value in the previous layer and an activation function in which the input values and the sum of weights of neurons are output in the form of signals.

2.2.2 Perceptron

Fig. 2.3 is a schematic diagram illustrating the working principles of perceptron. Its operation sequence is as follows.

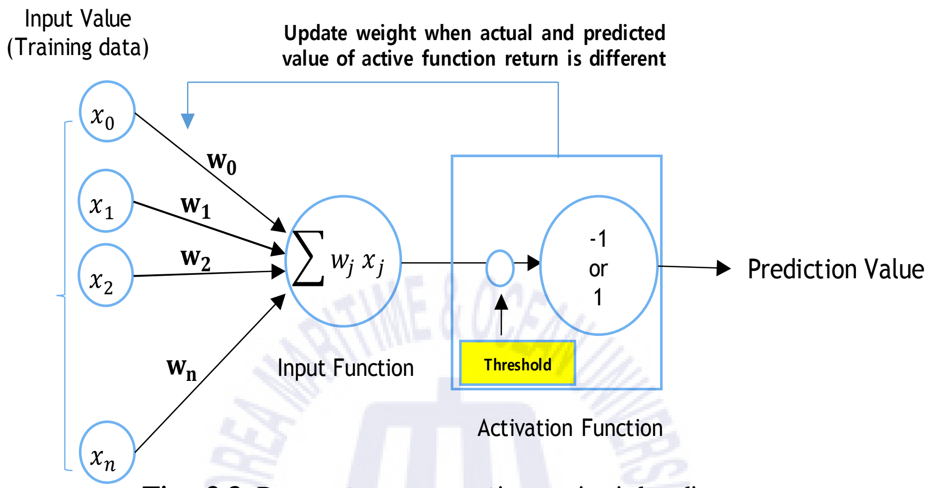

Fig. 2.3 Perceptron operating principle diagram

- Entering training data $(x_0, x_1, x_2 \cdots x_n)$
- Multiplying the weight $(w_0, w_1, w_2 \cdots w_n)$ with the input value
- Multiplying the input value with the weight, and pass it to the net input function.
- If the result value of a net input function is greater than the threshold value of an activation function, 1 is gotten and if less, −1 is gotten.
- After comparison with the actual result values, a weight should be updated in a way that minimizes the expected and actual values.

As shown in Fig. 2.3, a perceptron that consists of an input layer and an output layer is called a single-layer perceptron. The single-layer perceptron has a disadvantage that there is one activation function, so that there is a difficulty in learning about a nonlinear model.

2.2.3 Multi Layer Perceptron (MLP)

Fig. 2.3 is a structure diagram of Multi Layer Perceptron(MLP). MLP is a structure in which a hidden layer is placed between the input layer and the output layer to compensate for the shortcomings of the single layer perceptron. The complexity of a neural network is determined according to the number of hidden layers, and in general, an artificial neural network having a number of hidden layers is called a Deep Neural Network(DNN).

The operational principle of MLP is similar to that of single layer perceptron. And, it is as follows.

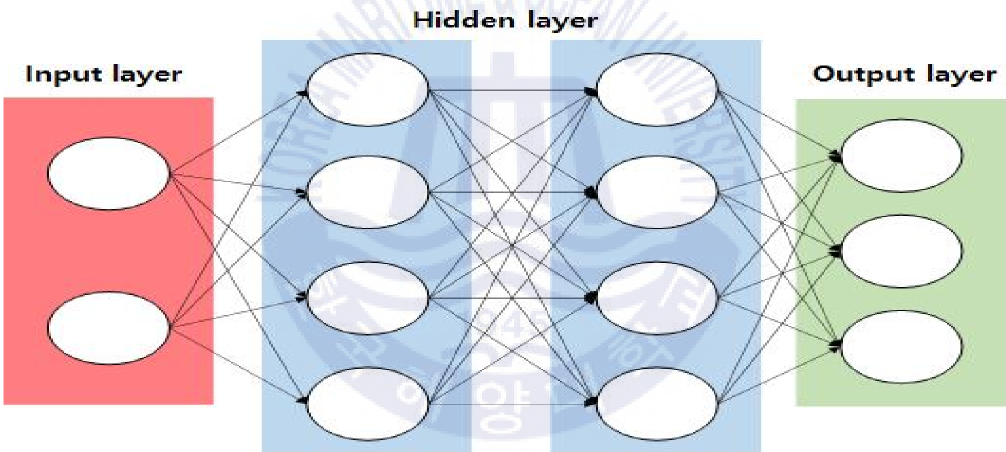

Fig. 2.4 Multi layer perceptron diagram

- Entering training data $(x_0, x_1, x_2 \cdots x_n)$
- Randomly setting a weight $(w_0, w_1, w_2 \cdots w_n)$ in each layer
- As for a set of training data, a net input function value should be calculated for each layer, and the output value by the activation function will be calculated.
- The learning continues until the end of the epoch and compares the actual value with the observation value

2.3 The Learning Process of an Artificial Neural Network

Fig. 2.5 shows the learning process of an artificial neural network. In general, before gathering data, an understanding of the characteristics of a problem to be solved should be the top priority.

Based on understanding it, the meaningful results can be obtained when the following steps are taken.

Fig. 2.5 Artificial neural network learning procedures

1) Data Gathering: It means a process of gathering training data and evaluation data to be used in a learning model.

2) Data Preprocessing : It means a preprocessing process of noise or biased data which can reduce an accuracy for the entire data set. It also includes the process of unifying the dimension by normalizing the input variables.

3) Neural Network: A choice should be made of what machine learning algorithm to be used for a problem to be solved.

4) Training: It is a process in which a learning is made for learning data.

5) Result Evaluation: It is a process of being evaluated for the result data based on learned model.

6) Parameter Change: It is a process of adjusting the changes to conditions set in the learning model for higher accuracy.

7) Prediction Result: It is a process of determining the accuracy of a learning model through the evaluation data finally.

2.4 The Key Concepts of an Artificial Neural Network

Table 2.1 lists the terms that are often used in artificial neural networks.

Table 2.1 Artificial neural network terms

Terms	Description
Training data	Data used in neural network learning
Test data	Data to evaluate learned neural network models
Loss	Differences between learning data and labels
Loss Function	Function for error measurement
Normalization	Pre-processing steps to reduce the impact from data dimensions
Optimizer	Parameters that determine how to learn by reducing the value of the loss function
Learning rate	Step in the gradient descent
Hyper parameter	Parameters that must be set by person in the learning model
Epoch	Number of times learning data has been ended during learning
Overfitting	State of optimized for training data (Therefore test data is less accurate)
Activation function	A function that converts the sum of the input signal into the output signal
Drop out	Disable for some neurons in learning (But use all neurons for evaluation)

Chapter 3 Research Procedure

3.1 Data Gathering

The research is made for the characteristics of RAO for the rolling of barge type ships. Fig. 3.1 represents the kind of ship’s classification. Among them data from Korean ship’ classification(KR), Japan ship’ classification(Class NK) and Denmark-Germany ship’ classification(DNV GL) were used. Data on length, breadth and draft of each ship were gathered.

Fig. 3.1 Kind of ship’s classification

The total number of data corresponded to 500 ships, and the ships with duplicated specifications were excluded from gathering data. Based on the data of a ship which were acquired, eight input variables to be used for a learning model were created. Here, the components related to a rolling were used to set the selection criteria for input variables. They are shown in Table

3.1. And, the radius of gyration of x-axis was set as 0.4 times of a line width, and the center of gravity of a ship was assumed to be located on the free water level.

Table 3.1 Input features

Property	Description
L	Length [m]
B	Breadth [m]
D	Draft [m]
V	Volume [m^3]
k_{44}	Radius of gyration of x-axis[m]
I_{44}	Mass moment of inertia of x-axis [kgf/m]
C_{44}	Restoring coefficient of x-axis [kgf]
GM_T	Transverse metacenter [m]

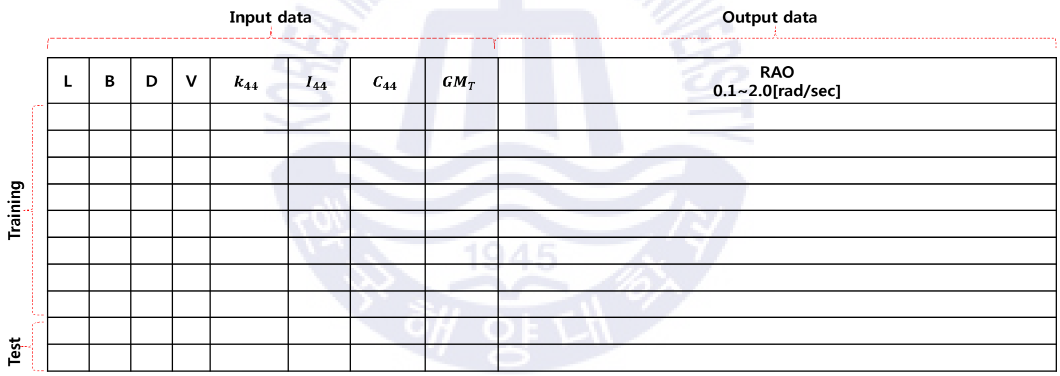

Fig. 3.2 Dataset concept

Figure 3.2 shows a schematic of a data set. In addition, 8 input variables were used as property values in In-house code based on 3D singularity distribution method, and were used to derive the RAO values. The code was developed by Jo in 1991. It was based on potential analysis program. And the analysis range of the RAO was 0.1 ~ 2.0[$rad/\sec$] considering the range of the external force environment of a real sea, a total of 20 original frequencies were set at 0.1[$rad/\sec$] intervals. Therefore, a data set consisting of 8 input variables and 20 output variables for 500 ships was created. Eight input variables played a role of a feature in the data set, and 0.1 to 2.0$[rad/\sec]$ with an interval of 0.1$[rad/\sec]$ played a role of a label.

3.2 Learning Model

A learning model was written in python language, and Tensorflow was used to make an artificial neural network. The number of input layer nodes in the learning model was set to 8 and it was equal to the number of input variables. The output layer was set to 20, which is the number of frequency domains of RAO. In this paper, the number of hidden layers and the number of neurons constituting the hidden layer was set as a variable. In addition, the normalization process was performed in the range of 0 ~ 1 for the data set to equalize the effect of input variables. Therefore the data distortion might not occur according to the dimension of the entire input variables. The ratio of training data and evaluation data was set to 8:2. Validation data was not used in this study because the number of total data is small and validation data may not represent the entire data characteristic. In the learning model, sigmoid function was used as activation function. The hyper parameters used in learning models of this paper are shown in Table 3.2.

Table 3.2 Hyper parameters

Parameter	Value
Input layer neurons	8
Hidden layer	Hidden layer : Variable Hidden layer neurons : Variable
Output layer neurons	20
Data ratio	Training data : 80 % Test data : 20 %
Learning rate	0.01
Epoch	10,000
Drop out	0.7
Batch size	80
Optimizer	Adam
Activation function	Sigmoid

3.3 Accuracy Evaluation Index

3.3.1 The Change of Random Numbers(Seed Number)

As securing the entire number of data was not satisfactory, it was judged that a large difference in numerical values for the accuracy of a learning model might be made. Therefore, to consider various training data and evaluation data, the random numbers inside a learning model were changed to calculate the accuracy for various training data and evaluation data. In this time 20 seed numbers from 0 to 19 were set to create 20 samples of evaluation data, and the accuracy of a sample was obtained for each random number. Therefore, one sample consisted of 400 learning data and 100 test data, and a total of 20 different samples were evaluated for accuracy.

3.3.2 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) was used as a general indicator to evaluate the regression model. It can be expressed as formula (1). And, y_o is RAO value obtained by simulation with In-house code and y_p means RAO value predicted by a learning model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_o - \hat{y_p})^2} \tag{1}$$

n : Data number

y_o : Measured value by In-house code

$\hat{y_p}$: Prediction value by learning model

3.3.3 Standard Deviation(SD)

In order to reflect the range of fluctuation of RMSE according to the evaluation data, Standard Deviation(SD), which represents a degree of scattering, was used. SD can be expressed as formula (2).

$$SD = \pm \sqrt{\frac{\sum (y_p - \overline{y_o})^2}{n-1}} \qquad (2)$$

n : Data number

y_p : Prediction value by learning model

$\overline{y_o}$: Mean value of Measured value by In-house code

3.3.4 Correlation Coefficient

To determine the relationship between two variables, it is required to check a joint probability distribution. Correlation coefficient(ρ) was used as an indicator to measure a direction and an intensity of a linear relationship. It can be expressed as formula (3).

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (3)$$

x_i : Variable x

y_i : Variable y

$\overline{x}$: Variable x' s mean value

$\overline{y}$: Variable y' s mean value

n : Data number

Chapter 4 Learning Result

4.1 Case Table

- The number of data: 100EA / 200EA / 300EA / 400EA / 500EA
- The number of hidden layers: 2nd floor / 3rd floor / 4th floor
- The number of neurons in the hidden layers: (256,256) / (200,200) / (100,100) / (115,95) / (18,18) / (14,14)

Table 4.1 shows the case according to the variables. The case is indicated in the form of 'DN[#data number]_L[#hidden layer number]_NN[(#the number of neurons in the first layer, #the number of neurons in the second laye r…)]'.

Table 4.1 Case table

Variable	Case
Data Number	DN100_L2_NN(256,256) DN200_L2_NN(256,256) DN300_L2_NN(256,256) DN400_L2_NN(256,256) DN500_L2_NN(256,256)
Hidden Layer Number	DN500_L2_NN(256,256) DN500_L3_NN(256,256,256) DN500_L4_NN(256,256,256,256)
Hidden Layer Neuron Number	DN500_L2_NN(256,256) DN500_L2_NN(200,200) DN500_L2_NN(115,95) DN500_L2_NN(100,100) DN500_L2_NN(18,18) DN500_L2_NN(14,14)

4.2 Change the Number of Data

The data used in this study consists of a total of 500 data. And the number of data was changed from 100 to 500 with an interval of 100. Fig. 4.1 shows the distribution of each data. Fig. 4.1 (a) shows the length and Fig. 4.1 (b) shows the width. Also, Fig. 4.1 (c) shows the distribution plot by draft. Because of depending on the distribution of data in the learning model, accuracy can be distorted. Therefore, in this research, the trend of distribution was set similarly for the accuracy of the result of a learning. Therefore, Fig. 4.1 shows that the trend of a distribution is similar even if the number of data changes.

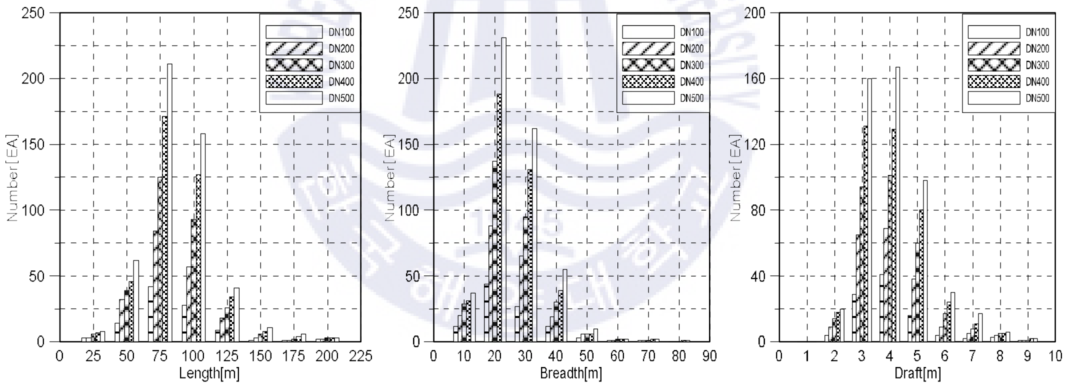

Fig. 4.1 Distribution plot by data number[(a)Length , (b)Breadth , (c)Draft]

Fig. 4.2 shows the results of RMSE and SD values according to the number of data. In Fig. 4.2 (a) shows that, as the number of data increases, RMSE decreases. It can be seen from Fig 4.2 (b) that the degree of variation in an accuracy is also reduced. To put it another way, this mean that as the number of data increase the RAO predictions become more accurate overall, and the difference between sets of data is reduced.

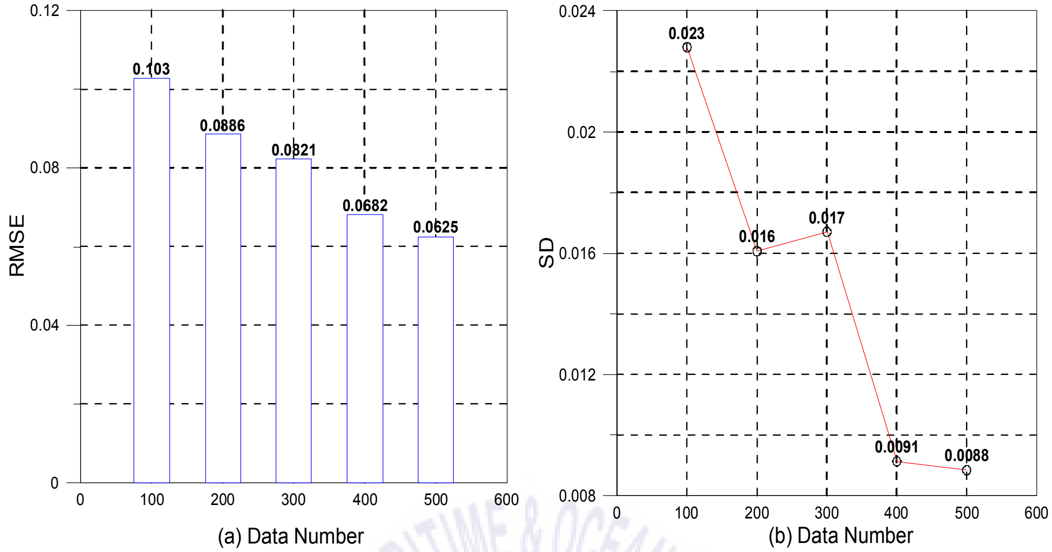

Fig. 4.2 (a) RMSE & (b) SD by data number

As mentioned in section 3.3.1, in this study, 20 test samples were used. That is to say, The results of Fig. 4.2 show the average value of 20 test samples. And we used the average value of RMSE for 20 data sets to ensure the reliability of the learning results. Therefore, the RMSE of a particular data set has a bigger value than the average, also, RMSE of some data sets may have a smaller value than the average value. This is because the configuration of the test data is different. That's why we used a statistic figure called standard Deviation(SD). The larger the number of data, the smaller the variation in the RMSE at each test sample. Ultimately, it is understood that more accurate predictions were made when there was the largest amount of data.

However, the disadvantage exists that average value is representative value, making it difficult to consider each test data. This being so, considerations are required for each test sample. So, in Fig 4.3, RMSE is shown for each data set of DN500_L2_NN(256,256).

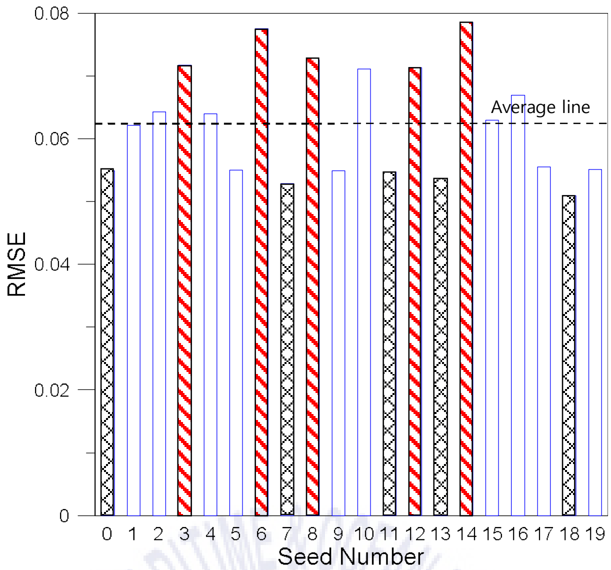

Fig. 4.3 DN500_L2_NN(256,256)'s RMSE

Among 20 data sets, the data sets whose RMSE corresponded to the higher 25% (Seed number 3, Seed number 6, Seed number 8, Seed number 12, and Seed number 14) were displayed as slanted bar graphs and the data sets whose RMSE corresponded to the lower 25% (Seed number 0, Seed number 7, Seed number 11, Seed number 13 and Seed number 18) were displayed as cross-bar graphs.

Table 4.2 RMSE of seed number

Seed Number	RMSE	Mean
#3	0.0717	
#6	0.0775	
#8	0.0725	0.0743
#12	0.0712	
#14	0.0785	
#0	0.0548	
#7	0.0528	
#11	0.0547	0.0534
#13	0.0536	
#18	0.0509	

4.2.1 The Analysis of RAO Data by Each Ship-Seed Number 6&14

In section 4.1, for seed numbers 3, 6, 8, 12, and 14, RMSE which was larger than the average was observed. As i mentioned, RMSE of one seed number is an average value of RMSE of 100 ships. So it is necessary to analyze each ship for the reason why RMSE are large for seed numbers of 6 and 14. Therefore, in Fig. 4.4, a horizontal axis is represented by a ship for the evaluation data and a vertical axis is represented by RMSE of the corresponding ship. And Fig. 4.4 (a) means RMSE of a ship with seed number 6 and Fig 4.4 (b) means RMSE of a ship with seed number 14. In Fig. 4.4, for the entire ships, a prediction is made with a low accuracy, and high RMSE tends to be observed for some specific ships. Thus, ‘X’ symbol for a particular ship on a graph means a ship with a relatively large RMSE which is observed, ‘O’ symbol for a particular ship on a graph means a ship with a small RMSE which is observed. For each Seed number, three ships marked with ‘X’ and three ships marked with ‘O’ were set and used for analyzing RAO.

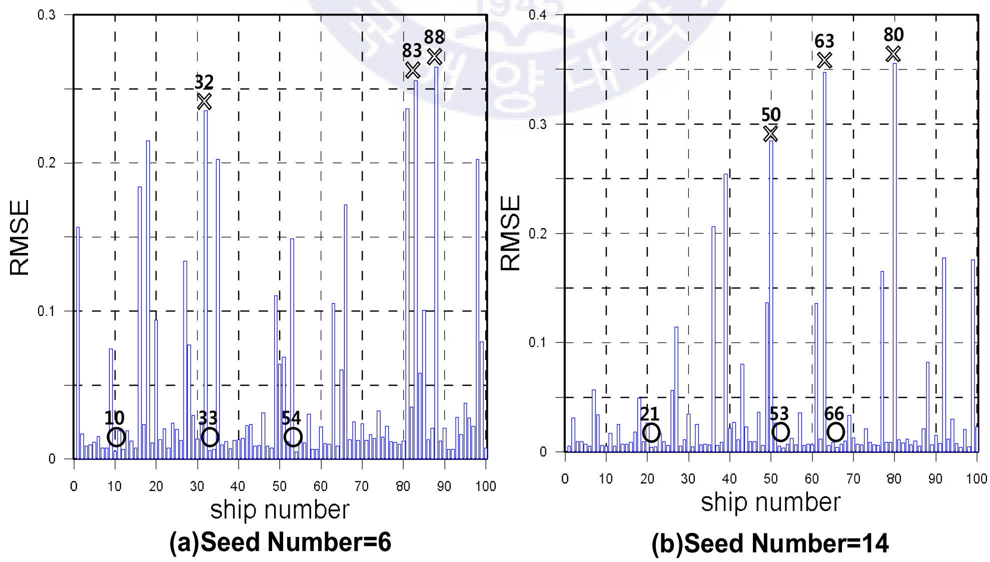

Fig. 4.4 (a) RMSE at seed number 6 & (b) seed number 14

In Fig. 4.5, RAO is shown for comparing the measured value of a ship with the prediction value by learning model in relation to #10, #33 and #54 ships corresponding to seed number 6. We know that slight difference was made between the measured value and the prediction value in a specific section, but it could be confirmed that the entire size or the location of a resonance point was accurately predicted.

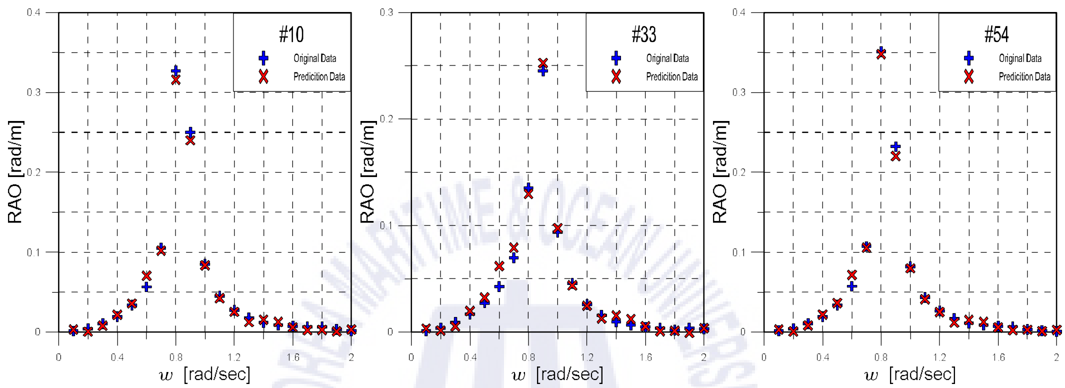

Fig. 4.5 RAO comparison(seed number 6 : #10, #33, #54)

In Fig. 4.6, RAO is shown for comparing the measured value of a ship with the prediction value by learning model in relation to #21, #53 and #66 ships corresponding to seed number 14. Fig. 4.6 shows that there is no difference between measured value and prediction value throughout the RAO.

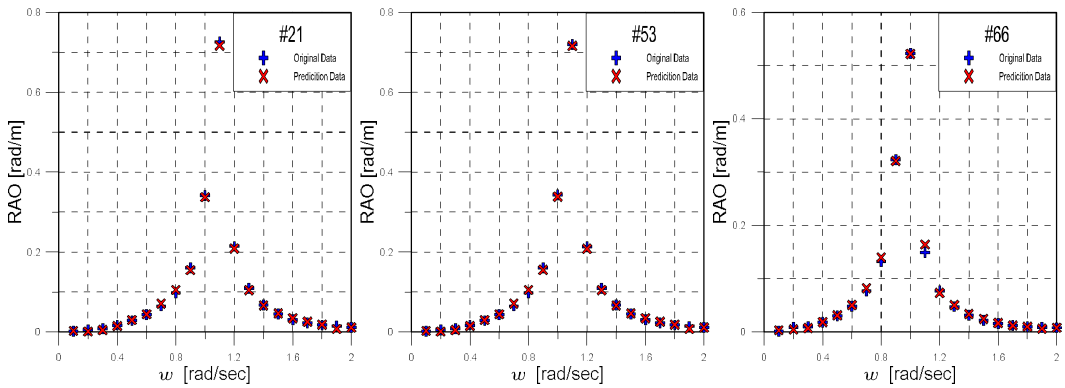

Fig. 4.6 RAO comparison(seed number 14 : #21, #53, #66)

In Fig. 4.7, the accuracies of low and high frequencies are highly accurate except for the range near a resonance point. However, an error occurs due to the difference in values at a resonance point and the position of a resonance point. In addition, The dimensions of the longitudinal axis confirm that the RAO value is greater than 1 and that these numerical differences have resulted in low accuracy.

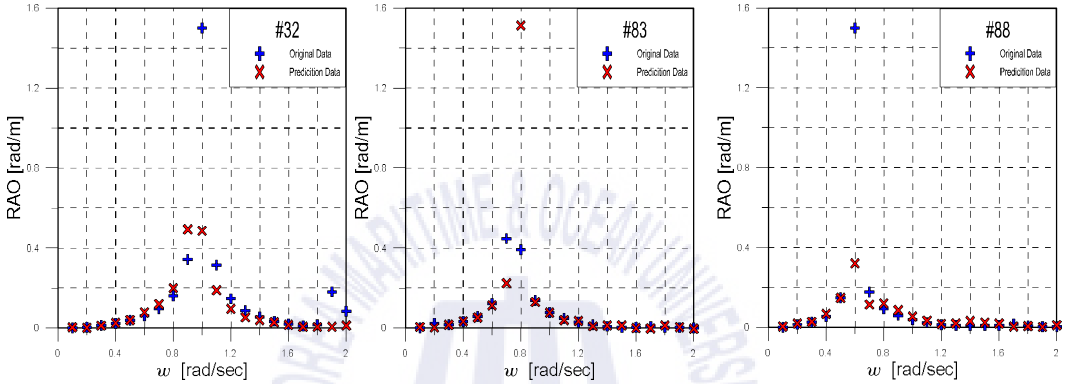

Fig. 4.7 RAO comparison(seed number 6 : #32, #83, #88)

In Fig. 4.8 shows that a difference between original data and prediction data. And numerical difference is made in the entire frequency domain. Moreover, the difference by the dimensions of the longitudinal axis is also judged to be a factor that reduces accuracy .

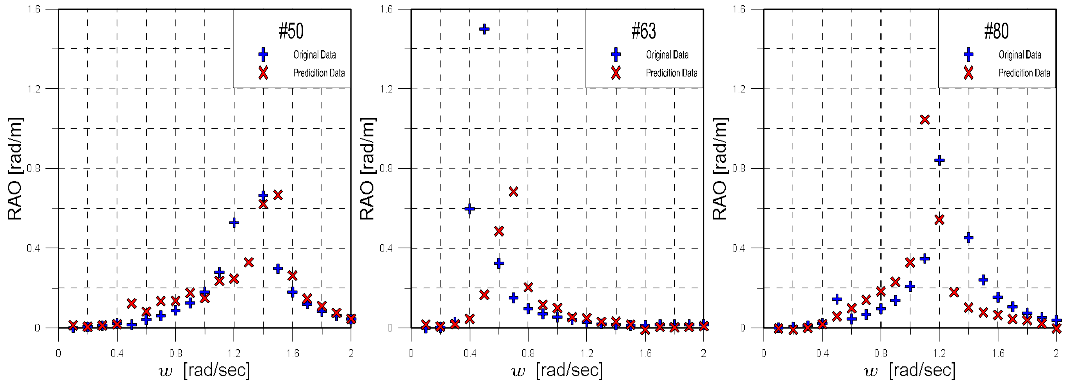

Fig. 4.8 RAO comparison(seed number 14 : #50, #63, #80)

4.2.2 The Analysis of RAO Data by Each Ship-Seed Number 7&18

Fig. 4.9 (a) and (b) is a graph of RMSE along with 100 ships at a seed number 7 and 18. Small RMSE (high accuracy) than the average was observed for seed number 0, 7, 11, 13 and 18 in section 4.2. Especially, the RMSE is 0.528 when the seed number is 7, and 0.509 when seed number is 18. As for seed number 7 and 18, an analysis was made of RMSE for each ship in the same way as section 4.2.1. In case of a seed number of 7, vessels(#1, #33, #92) were selected as the case of high accuracy, and vessels(#10, #38, #59) were selected as the case of low accuracy. And in case of a seed number of 18, vessels(#9, #56, #69) were selected as the case of high accuracy, and vessels(#11, #14, #83) were selected as the case of low accuracy.

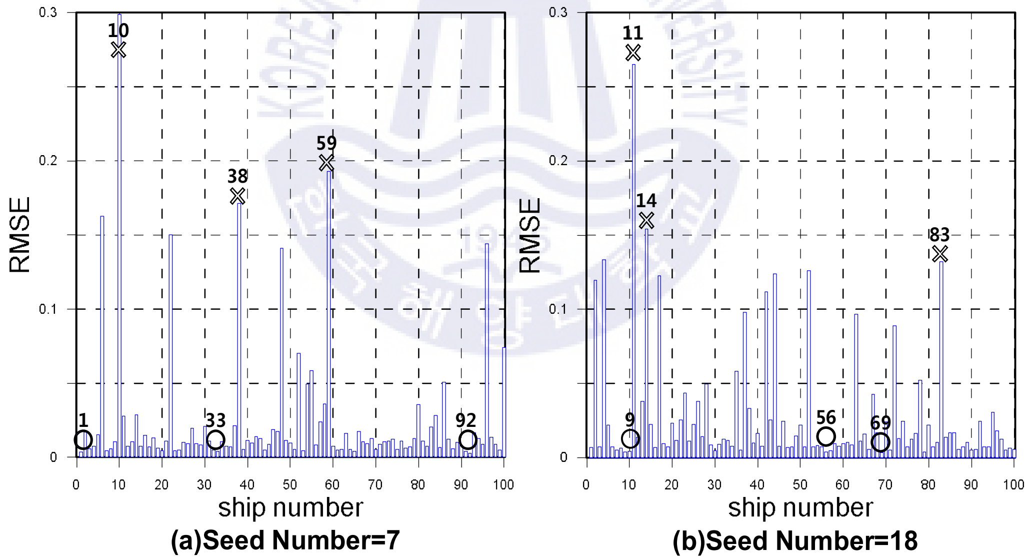

Fig. 4.9 (a) RMSE at seed number 7 & (b) seed number 18

In Fig. 4.10 represents that RAO is shown for comparing the measured value of a ship with the prediction value by learning model in relation to #1, #33 and #92 ships corresponding to seed number 7. As shown in Fig. 4.10 it appears that there is no difference between the original data and prediction data in the overall area of the analysis as shown in section 4.2.1. As mentioned earlier, commonly high-accuracy vessels have low-level RAO results.

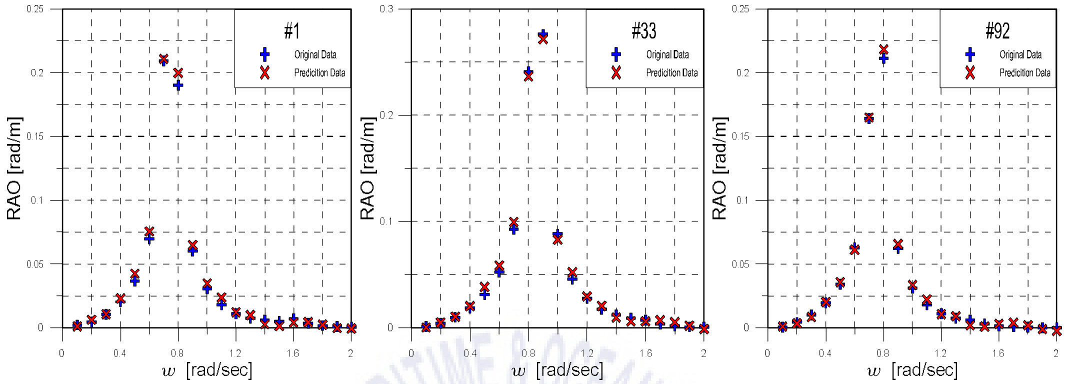

Fig. 4.10 RAO comparison(seed number 7 : #1, #33, #92)

Fig. 4.11 represents that RAO is shown for comparing the original data with the prediction data in relation to #9, #56 and #69 ships corresponding to seed number 18. Similar to section 4.2.1, the ships with high accuracy show a slight difference between a location of a resonance point and the value for a range around a resonance point. Furthermore the RAO figure is below 0.3 may have affected the high accuracy.

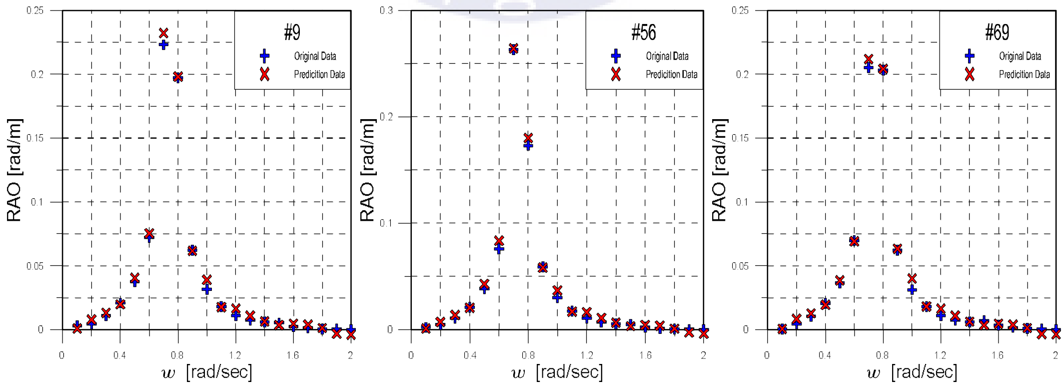

Fig. 4.11 RAO comparison(seed number 18 : #9, #56, #69)

Fig. 4.12 that RAO is shown for comparing the original data of a ship with the prediction data in relation to #10, #38 and #59 ships corresponding to seed number 7.

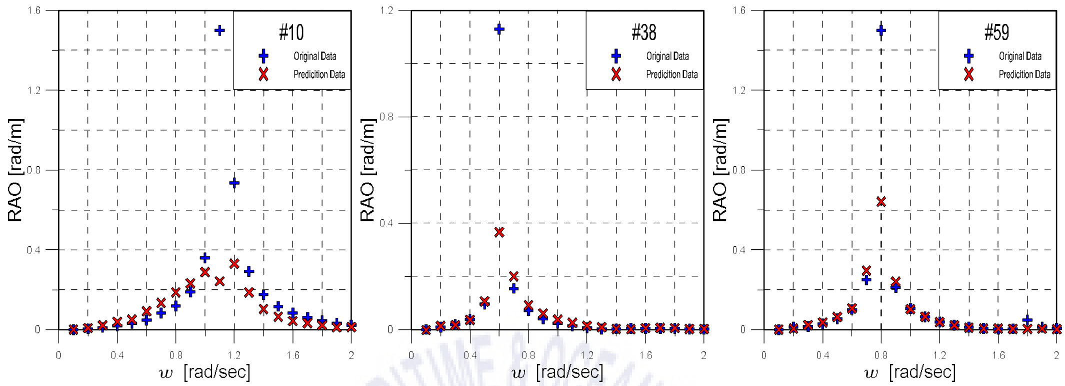

Fig. 4.12 RAO comparison(seed number 7 : #10, #38, #59)

Especially #10 ship in Fig. 4.12, it is observed that an error occurs due to the difference between the position of a resonance point and the value for a range around a resonance point. In detail original data of #10 ship has a high-level dimension and prediction data does not accurately predict the magnitude of the resonance point. Thus it makes a biggest RMSE when the seed number 7.

On the other hand, in case of #38 ship and #59 ship, although the location of a resonance point is the same, it is confirmed that the error to be caused due to the difference of a magnitude in a resonance point has the greatest influence. Although the accuracy of the vessels(#10, #38, #59) were low, identifying the location of the resonance point is the most important aspect of the roll motion characteristics, it was judged that the better results could be obtained through further supplementation.

In Fig. 4.13, RAO is shown for comparing the original data of a ship with the prediction data in relation to #11, #14 and #83 ships corresponding to seed number 18.

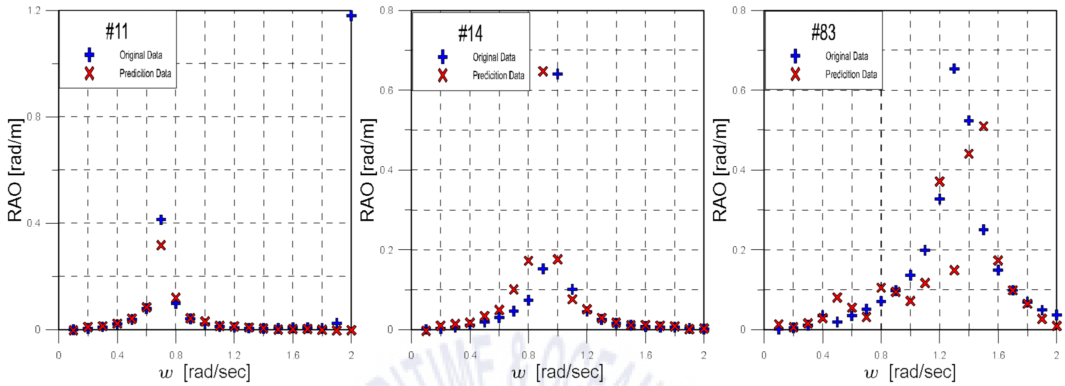

Fig. 4.13 RAO comparison(seed number 18 : #11, #14, #83)

In case of #11 ship, when the original data is 2[rad/sec], it is confirmed that an abnormal value has occurred. Rather, the figure of a prediction value is considered to be a valid result. Therefore the error in #11 ship is considered to be caused from the noise in a data set. In this respect, learning result enables feedback on incorrect simulated data.

However, in case of #14 ship, A high RMSE was derived in that the location of the resonance point was not accurately predicted. Then #83 ship, Not only did we not accurately predict the resonance point, but we can also see that the predictive accuracy of the figures is poor across the whole area. In detail, #14 has the same vertical axis as #83, but in the case of #14, the difference between the original and the predicted values was observed in 0.9 [rad/sec] and 1.0[rad/sec]. In the case of #83, he trend is difficult to identify, but only 1.3[rad/sec] has made a big difference, which finally results in lower RMSE than #14.

4.3 Change the Number of Hidden Layers

A hidden layer is a layer that receives an input value from an input layer, calculates the sum of weights and applies the sum of weights to the activation function to deliver the corresponding values to the output layer. Generally, one-layer or two-layer neural network is used for a hidden layer. But, in some cases, many hidden layers are needed according to the purpose and complexity a neural network. Usually, as the number of hidden layers increases, the complexity of a neural network tends to increase. As a result, the accuracy can decrease. However, in section 4.2, to optimize a learning model, as the number of hidden layers was changed to 2 layers, 3 layers and 4 layers RMSE and SD were calculated and the corresponding values were compared.

In Fig. 4.14, the number of hidden layers is shown in the horizontal axis and RMSE and SD are shown the vertical axis. As for the number of hidden layers, it can be confirmed that RMSE and SD increase slightly when 2 layers are changed into 3 layers. Above all, it can be confirmed that RMSE and SD increase non-linearly in case the number of hidden layer is 4 layers.

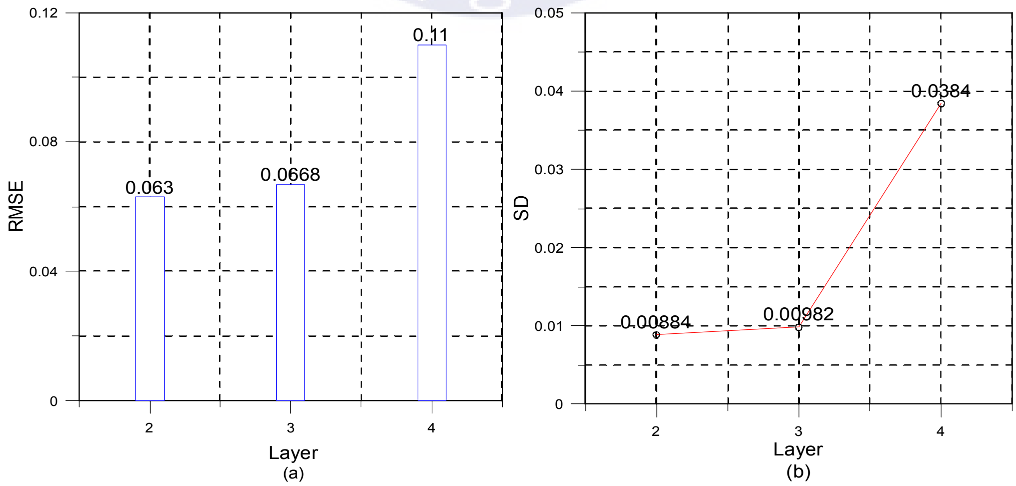

Fig. 4.14 (a) Change in the number of hidden layers RMSE & (b) SD

4.3.1 DN500_L4_NN(256,256,256,256) Learning Result Analysis

Fig. 4.15 is a graph showing a horizontal axis as seed number and a vertical axis as RMSE for DN500_L4_NN(256,256,256,256). Among the 20 test samples, 12 test samples have a higher RMSE than average. It mean, the RAO was not estimated to be reasonably high through the RMSE. Second, the high SD values mean low confidence in the predict results. And, 12 cases RMSE above the average were marked with red slants bar.

Especially when the seed number 8 has the biggest RMSE. Above all, look at Fig. 4.15, and you can see that the difference between RMSE for each seed number is obvious. The reasons for this can be found in Fig. 4.14. The SD value between 2 layers and 3 is about 11%. But the variation between 2 layers and 4 layers is about 430%. Namely if there are four layers, the accuracy of the learning model is low and the SD is large by seed number. So learning models with 4 layers mean that they are not good. However, it is necessary to consider the reasons for having high RMSE and SD in the same way as before.

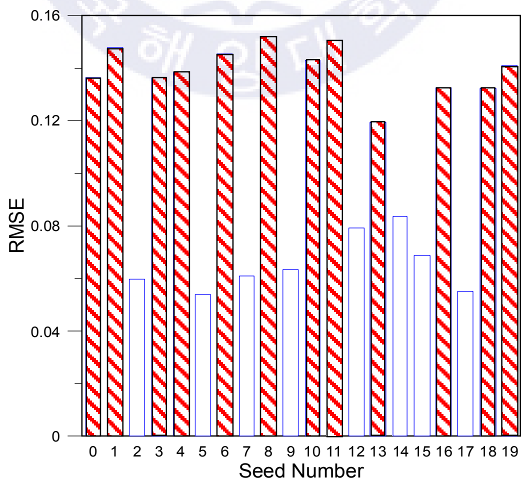

Fig. 4.15 DN500_L4_NN(256,256,256,256)’s RMSE

Therefore, it was intended to analyze the seed number 8 with the highest RMSE in Fig. 4.16. In the same way as in the previous chapter, the three vessels with high RMSE are marked with 'X' symbol.

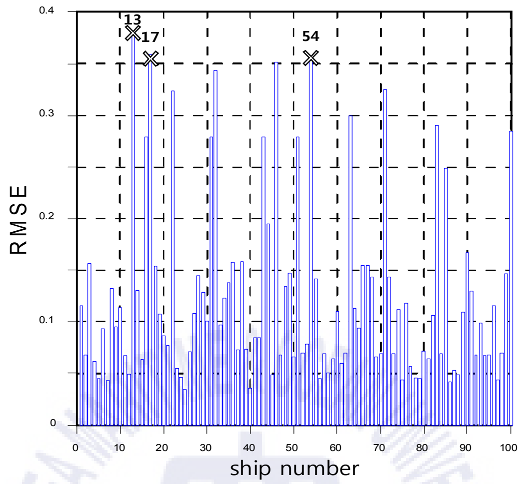

Fig. 4.16 RMSE at seed number 8

Fig. 4.17 shows a graph of RAO in which the measured value are compared with the prediction value for #13, #17 and #54 ships marked with an 'X' in Fig. 4.15.

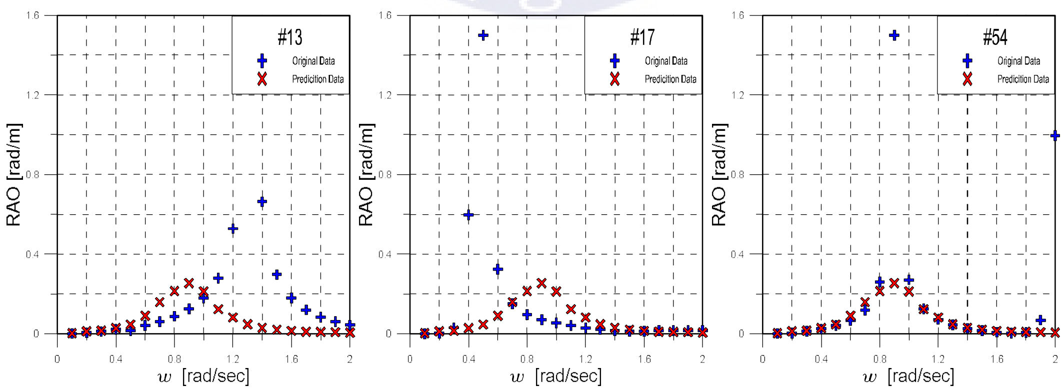

Fig. 4.17 RAO comparison(seed number 18 : #13, #17, #54)

And, if you look at the prediction value for three ships in Fig 4.16, although the ships have different specifications, they have the same RAO. Also, the same phenomenon was observed for the remaining 97 ships.

As for 12 seed numbers marked with a red slant, it was commonly observed that the same RAO was predicted for 100 ships. It is judged that the above phenomenon occurred because the complexity of a system increased as the number of hidden layers increased. The above complexity increased because a learning was not enough as a search was made for not the global minima in the entire loss function but the local minima in a specific part of a function.

4.4 Change of the Number of Neurons in a Hidden Layer

In section 4.1 and 4.2, we conclude that the efficiency is the best when the number of data is 500 and the number of hidden layers is 2. While the number of neurons input&output layer is fixed, the number of neurons in a hidden layer depends on the experience of a user, which makes a decision difficult. So, in section 4.3, a comparison was made of RMSE of learning models by changing the number of neurons constituting a hidden layer in the case of DN500_L2. In making a case, we used a backward approach to learn and test neural networks by reducing the number of neurons in the hidden layer step by step. Table 4.3 shows the neural network structure for each case. Case 1, 2 and 4 set the number of random neurons for a learning model. The number of neurons in Case 3 was selected with reference to the previous researches (D.Stathakis, 2009). The number of neurons in Case 5 as well as Case 6 was selected with reference to the empirical laws of the previous researches (Kim, 2017).

Table 4.3 Changes in the Number of Hidden Layers of Neurons

Case	Name	Description
1	DN500_L2_NN(256,256)	First floor neuron : Random selection(256) Second floor neuron : Random selection(256)
2	DN500_L2_NN(200,200)	First floor neuron : Random selection(200) Second floor neuron : Random selection(200)
3	DN500_L2_NN(115,95)	First floor neuron : $\sqrt{(m+2)DN}+2\sqrt{DN/(m+2)}$ Second floor neuron : $m\sqrt{DN/(m+2)}$
4	DN500_L2_NN(100,100)	First floor neuron : Random selection(100) Second floor neuron : Random selection(100)
5	DN500_L2_NN(18,18)	First floor neuron : $2(n+m)/3$ Second floor neuron : $2(n+m)/3$
6	DN500_L2_NN(14,14)	First floor neuron : $(n+m)/2$ Second floor neuron : $(n+m)/2$

※ n : $Input\,Layer\,Node\,Nmber$, m : $Output\,Layer\,Node\,Nmber$, DN : $Data\,Nmber$

If Fig. 4.18 (a) is reviewed, it can be confirmed that the trend of RMSE increases as the number of neurons increases. In Case 3, RMSE decreased slightly but in Cases 5 and 6, it increased dramatically. In contrast, in Fig 4.18 (b), it can be confirmed that the smaller the number of neurons will be, the smaller SD value will be. In other words, in Case of 1, 2, 3 and 4, even though average RMSE is small, according to the evaluation data, it is found that the fluctuation range of RMSE is big. On the contrary, Case 5, 6 have a rather large RMSE, however, the fluctuation range of RMSE is small.

Therefore, in case 6, even the SD, the fluctuation range of RMSE, is small, since the absolute RMSE is big, so it is regarded that reliability is low. When we put RMSE and SD in the right compromise point, it is regarded that Case 4 could become into the optimal model.

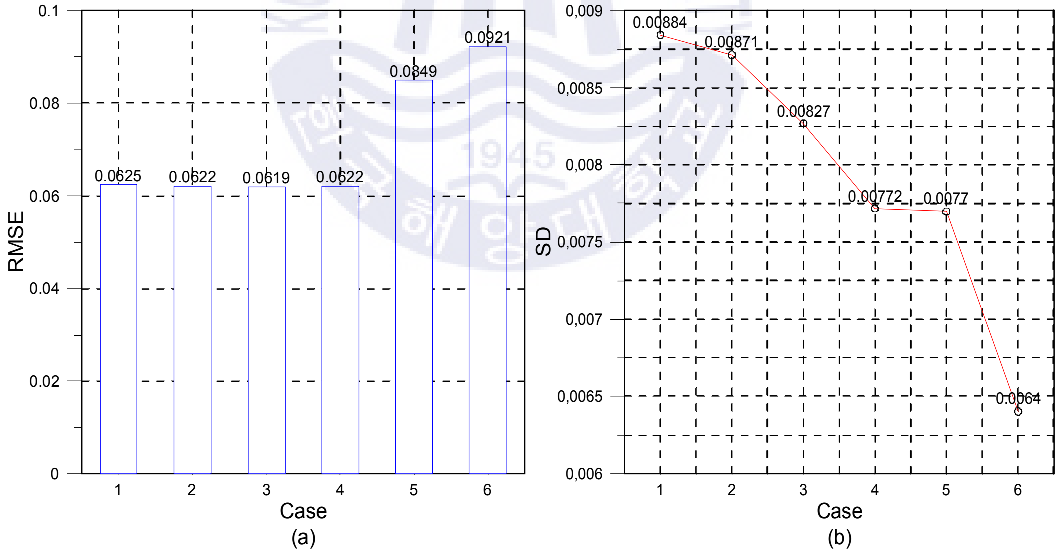

Fig. 4.18 (a) Change number of neurons in the hidden layer RMSE & (b) SD

4.5 Case 4: DN500_L2_NN(100,100) Analysis

Fig. 4.19 (a) is a graph showing RMSE for each evaluation data of DN500_L2_NN(100,100)case. And, when the seed number is 14, RMSE becomes higher than other evaluation data. Therefore, Fig. 4.19 (b) is a graph showing RMSE for each ship of seed number 14. So, in Table 4.4, the zones are divided based on the specific RMSE.

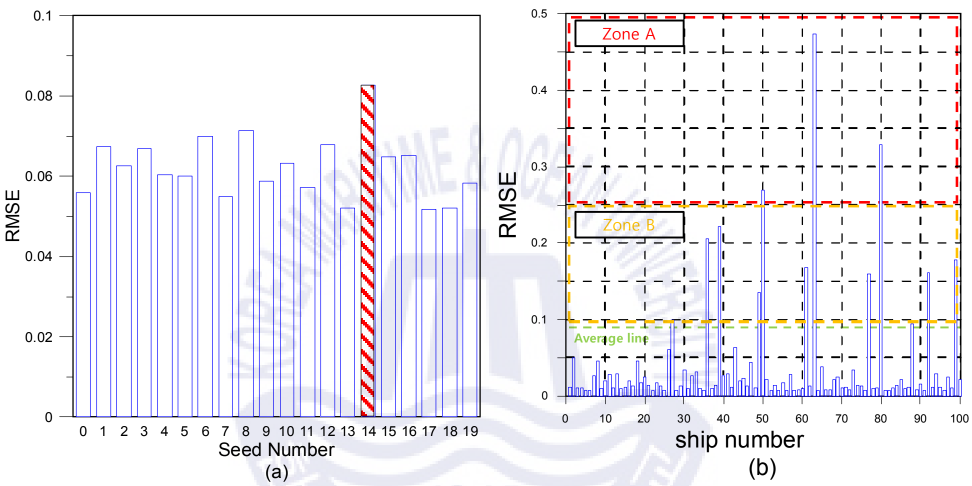

Fig. 4.19 (a) DN500_L2_NN(100,100)' s RMSE & (b) seed number 14 RMSE

Table 4.4 Classification of vessels according to RMSE range

Zone(RMSE Range)	Ship number	RMSE
Zone A (0.25 ≤ RMSE ≤ 0.5)	#50	0.269
	#63	0.473
	#80	0.328
Zone B (0.08226 ≤ RMSE < 0.25)	#27	0.099
	#36	0.205
	#39	0.221
	#49	0.135
	#61	0.167
	#77	0.159
	#80	0.328
	#88	0.094
	#92	0.161
	#99	0.178

Fig. 4.20 is a graph comparing the RAO of #50, #63, and #80 ships in Zone A. Commonly, it has the overall configuration of RAO, however, there were significant differences from the measured value over the resonance point position, magnitude, low frequency region, and high frequency region. Due to this difference, it has been found to have high RMSE.

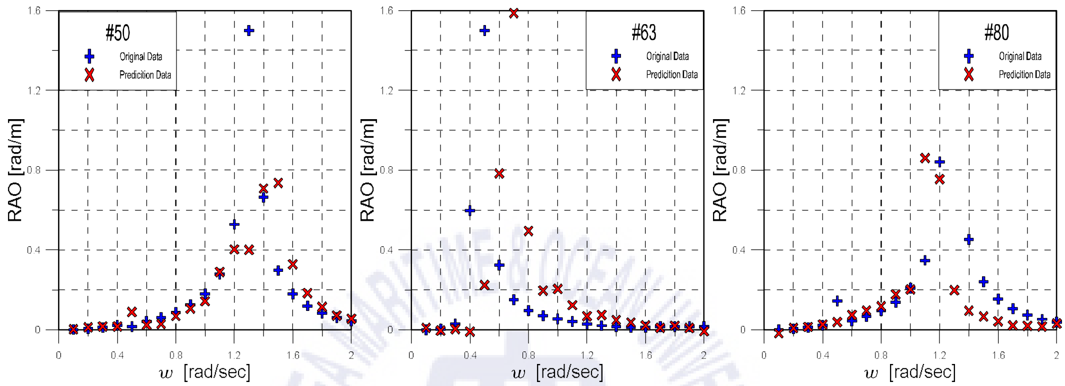

Fig. 4.20 RAO comparison(Zone A : #50, #63, #80)

Fig. 4.21 is a graph comparing RAOs of #27, #77, and #92 ships in Zone B. Commonly, the configuration of RAO was similar in the entire domain of a frequency and the resonance point position could be predictable, however due to the size difference in the resonance point, it is confirmed that this is a type has a RMSE.

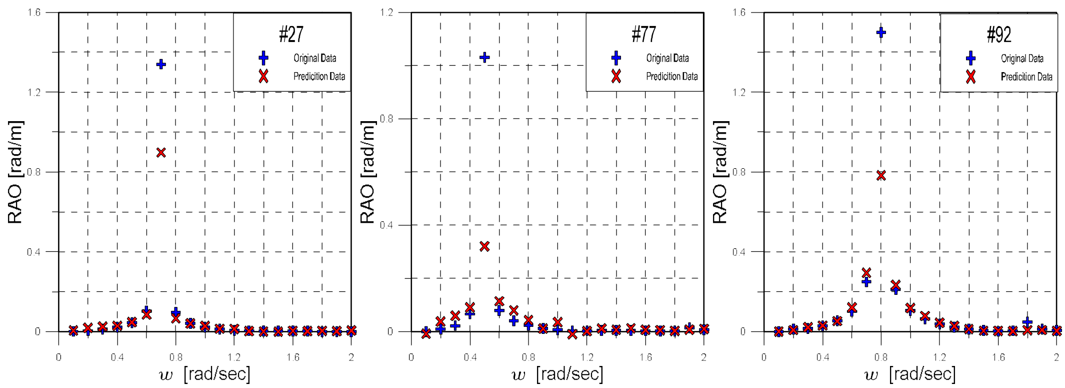

Fig. 4.21 RAO comparison(Zone B : #27, #77, #92)

Fig. 4.22 is a graph comparing the RAO of #36, #39, #49, #61, #88, #99 ships in Zone B. It can be confirmed that, as for six ships, the configurations of RAO are similar in the entire frequency domain but a difference is made around a resonance point.

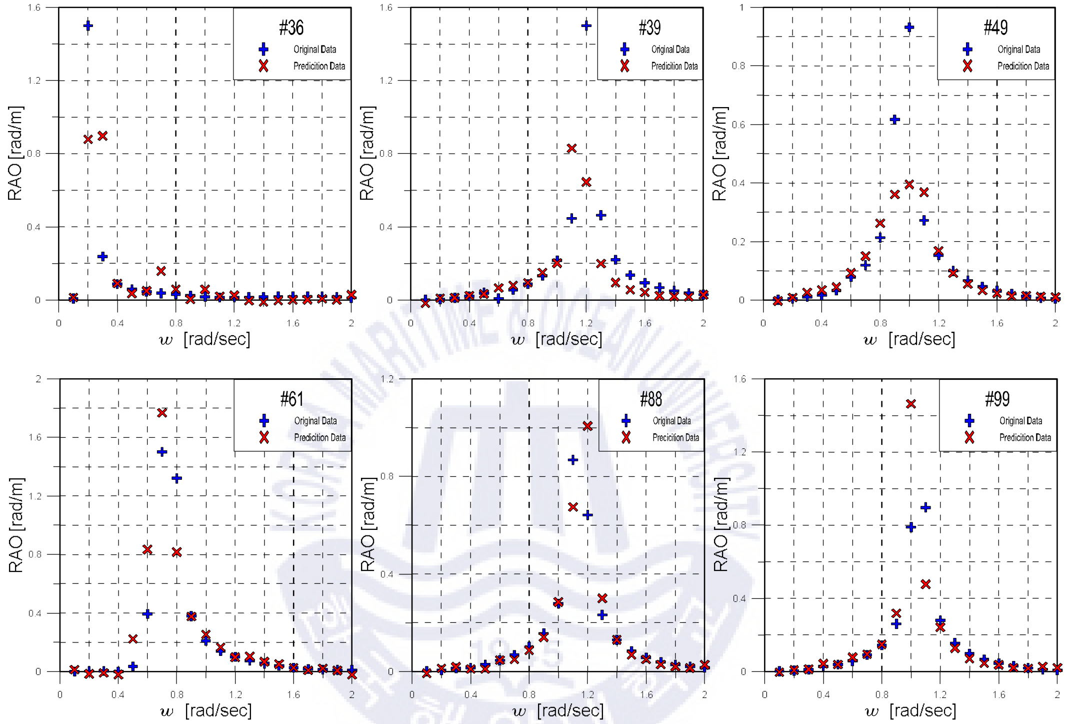

Fig. 4.22 RAO 비교(Zone B : #36, #39, #49, #61, #88, #99)

Moreover, it is necessary to understand the characteristics of the distribution of a training data and the distribution of an test data so that a grasp may be made of the reason why RMSE is large. Therefore, in Table 4.5, the correlation coefficients of the remaining input variables ($L, D, V, I_{44}, C_{44}, GM_T$) of the ship breadth were obtained to identify the characteristics of the data set.

The Coefficient_Train in Table 4.5 shows the value of a correlation coefficient for the ship's width and the remaining input variables in relation to the learning data. And, the ranking for each correlation coefficient is showed in Rank. Similarly, the Coefficient_Test shows the value of a correlation coefficient for the test data. And, the ranking is showed in Table 4.5. For example, the volume(V) according to the breadth(B) of the training data shows a correlation of 0.839, but in the test data, it is 0.737, which is lower than the training data. It means that training data has a strong correlation between breadth and volume, test data show a relatively weak correlation between breadth and volume. Therefore, Table 4.5 shows that the accuracy is low due to the difference between the trend of a training data and the trend of an test data.

Table 4.5 Number and rank of correlation between training and test data

Parameter	Coefficient_Train	Rank	Coefficient_Tests	Rank
L	0.740	3	0.660	5
D	0.719	4	0.609	6
V	0.839	2	0.737	4
I_{44}	0.658	6	0.759	3
C_{44}	0.715	5	0.874	1
GM_T	0.879	1	0.811	2

Fig. 4.23~Fig. 4.25 show the positions of the ships in Zone A and Zone B on the distribution. Fig. 4.23 ~ Fig. 4.25 show a distribution in which a horizontal axis is set to the breadth of a ship and a vertical axis is set to a length (Fig. 4.23), a draft (Fig. 4.24) and a transverse meta center (Fig. 4.25). In general, ships have a certain range of dimensions such as a breadth and a draft, depending on their lengths. This means that the specifications of a ship have a certain correlation with each other, and the correlation can be confirmed in the distribution, too.

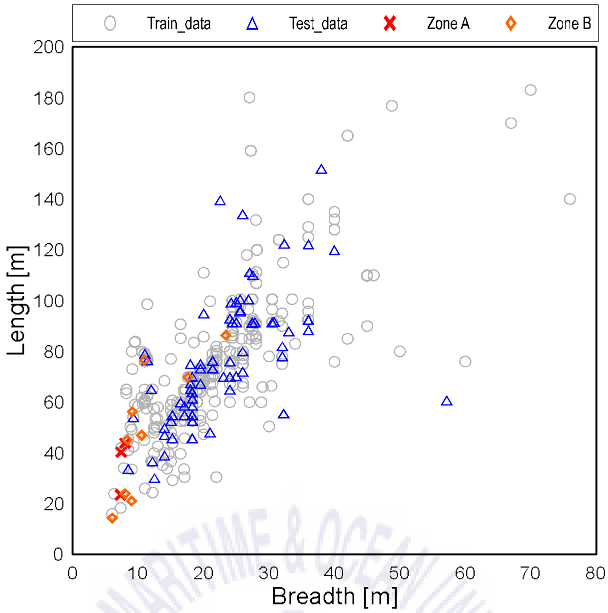

Fig. 4.23 Scatter plot(B-L)

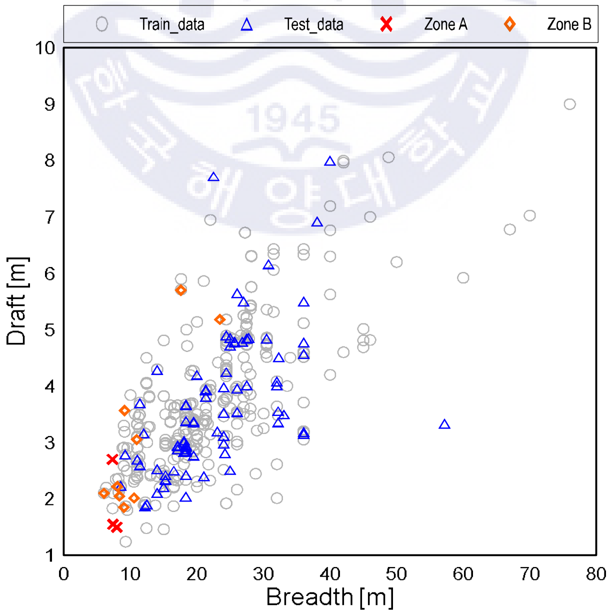

Fig. 4.24 Scatter plot(B-D)

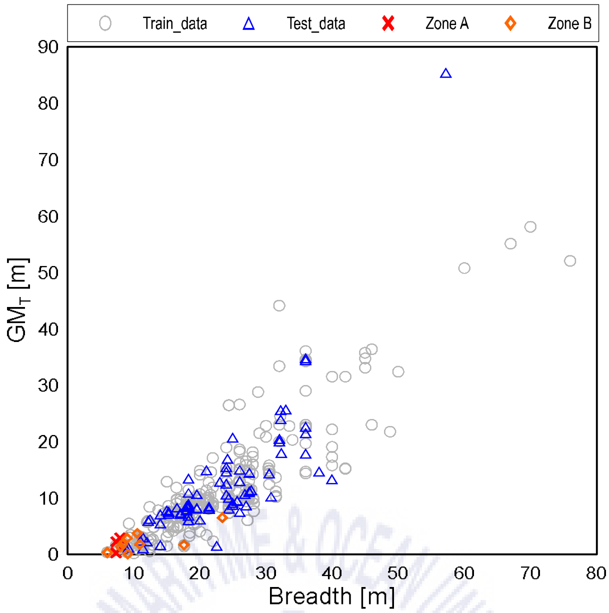

Fig. 4.25 Scatter plot(B-GM$_T$)

In the distributions of Fig. 4.22 ~ Fig. 4.24, it can be seen that data are concentrated in a certain section, and a ship with high RMSE is located at the low concentration of the data. In other words, a ship with less learning amount would have less accuracy.

However, it is confirmed that the overall trend of the RAO is able to be predicted even for ships belonging to Zone A and Zone B with big RMSE. Certain ships are found to be highly accurate, even though they are located at a point where the density of data is low. However, the previous results show that although the data are located in a low density area, ships with high accuracy are usually small in the vertical axis dimension.

To sum up, ranking of the correlation coefficient of the transverse meta center in table 4.5 and the location of vessels with low accuracy in Fig. 4.25, it is fact that the transverse meta center is a major factor.

Chapter 5 Conclusion

In this paper, a research was made to predict roll RAO of a barge-type ship by a using machine learning. The specifications of 500 ships among the barge-type ships which were registered of shipping classification were used to create input variables ($L, B, D, V, k_{44}, I_{44}, C_{44}, GM_T$). And In-house codes were used to create RAO data for each vessel. The code based on the three-dimensional singularity distribution method was used to simulate 500 ships and obtain values for roll RAO. Finally, RAO ranged from 0.1 to 2.0$[rad/\sec]$ according to the main specifications of the barge-type ship was created as a data set.

Then, deep neural network model was created using Tensorflow of python, and DNN technique with more than 2 hidden layers was used. The results of a learning were derived by changing the number of data, the number of hidden layers and the number of neurons of hidden layers. RMSE, SD, correlation coefficient and scatter plot were used to estimate the accuracy as the indicators. And, And the optimal model was chosen from a variety of variables thorough the consideration between RMSE and SD together. Finally, the accuracy of optimal model was analyzed to investigate the lack and improvement of the learning model.

The conclusions drawn from this research are as follows.

(1) It can be confirmed that the accuracy increased and the fluctuation range decreased, as the number of data increased.

(2) Since there was clearly a difference between the high accuracy data and the low test data among 20 test data, an accuracy analysis should be performed on some evaluation data.

(3) An accuracy was low in the particular ships among 100 ships.

(4) When comparing comparison is made of RAO of each ship for all the ships, they can be categorized into three types: (a) Ships with high accuracy that do not differ from the actual values, (b) Ships with reduced accuracy due to size differences at the resonance point, (c) Ships with low accuracy due to poor prediction of the location and size of the resonance point

(5) When the number of hidden layers was changed, the accuracy and fluctuation range were not significantly different in the 2 and 3 layers, but in case of the 4 layer, the accuracy decreased and the fluctuation range was increased.

(6) When the number of hidden layers was 4 layers, there are the cases in which the RAO of 100 test data was predicted as the same, which is considered to be influenced by local minima as a complexity increases.

(7) The model should be evaluated with more neurons between 100 and 18.

(8) It is judged that Case 4 was the best model, if the effects on an accuracy and a fluctuation range were considered.

(9) In order to improve the accuracy of Case 4, an analysis was made according to test data (seed number 14) with low accuracy. It is considered that the accuracy for test data of seed number 14 was low because the distribution characteristics of the learning data and the evaluation data were different based on the correlation coefficients.

(10) Based on a degree of scattering to be drawn for a ship with a low accuracy, it was found that the data with low accuracy were located in the domain of data with low density.

(11) When the trends of an accuracy and a fluctuation range by each number of data are judged, it is found out that more data will be needed in the future because the accuracy of a learning model increases and its fluctuation range decreases after sufficient data are obtained.

(12) As changing the number of neurons produced completely different results, it is necessary to make an optimal model based on combinations of various variables (learning rates, activation functions, etc).

(13) 100 neurons were rapidly changed into 18 neurons in a hidden layer. Thus, it is be required that an additional verification is made of a result value between 18 neurons and 100 neurons.

(14) The difficulties were caused in evaluating a reliability and a reliability interval as an accuracy did not follow a normal distribution. If the evaluation of a reliability interval is made for a variable which does not follow a normal distribution, it will help to grasp the accuracy of a prediction model.

Reference

Jo H.J, 1991. "渓流式浮遊海洋構造物の応答特性に及ぼす多方向波の影響に関する研究", 東京大学大學院工学系研究科船舶海洋工学専攻博士課程.

Mahfouz, A.B., 2004. Identification of the nonlinear ship rolling motion equation using the measured response at sea. Ocean Engineering, 31(17-18), 2139-2156. https://doi.org/10.1016/j.oceaneng.2004.06.001

Park, G.I, Choi, K,Y, Lee, J.H, Kim, M.S, 2004. Introduction of Optimum Navigation Route Assessment System based on Weather Forecasting and Seakeeping Prediction. Journal of Korean Navigation and Port Research, 28(10), 833-841.

Lee, J.H, Park, G.I, Choi, K,Y, Kim, M.S, Bang, C.S, 2005. Basic Research of Optimum Routing Assessment System for Safe and Efficient Voyage. Journal of the Society of Naval Architects of Korea, 42(1), 57-63.

Stathakis, D, 2009. How many hidden layers and nodes. International Journal of Remote Sensing, 30(8), 2133-2147. https://doi.org/10.1080/01431160802549278

Kang, M.S, Kim, Y.W, Jung, H, 2012. Floating body motion prediction system using artificial neural network. Proceedings of The Korean Association of Ocean Science and Technology Societies. 1041-1050.

Ham, D.G., 2016. A study of Data-Mining Methodology in Offshore Plant's Outfittings Procurement Management. Master thesis, Korea Maritime and Ocean University, Busan, Korea.

Anton, B, Brian, B, Reinaldo S, 2016. Multi-Output Artificial Neural Network for Storm Surge Prediction in North Carolina. ArXiv 2016. Vol 1609.07378

Kim, G.H., 2017. Artificial Neural Network for Short-Term Load Forecasting Algorithm Considering Temperature Characteristics. Master thesis, Korea Maritime and Ocean University, Busan, Korea.

Kim, J.H., 2018a. A study on Machine Learning for Prediction of the Shipbuilding Lead Time. Master thesis, Korea Maritime and Ocean University, Busan, Korea

Kim, Y.R., 2018b. Prediction of Ship Roll Motion using Machine Learning-based Surrogate Model. Journal of Korean Navigation and Port Research 42(6), 395-405.

Jeon, M.Y. 2019. Development of Predictive Model of Ship's Fuel Consumption using Machine Learning Method, Master thesis, Pusan National University, Busan, Korea.

Kim, Y.R, 2019. Development of a Fuel Consumption Prediction Model Based on Machine Learning Using Ship' s Data. Master thesis, Korea Maritime and Ocean University, Busan, Korea