

신경망기법을 활용한 선박 가치평가 모델 개발

김동균* · 최정석**†

* 목포해양대학교 항해학부 교수, ** 목포해양대학교 해상운송학부 교수

Development of Ship Valuation Model by Neural Network

Donggyun Kim* · Jung-Suk Choi**†

* Professor, Division of Navigation Science, Mokpo National Maritime University, Mokpo 58628, Korea

** Professor, Division of Maritime Transportation, Mokpo National Maritime University, Mokpo 58628, Korea

요 약 : 본 연구의 목적은 Neural Network Regression 모델을 활용하여 선박의 가치평가 모델을 개발하는 것이다. 가치평가의 대상은 중고 VLCC선이며, 선행연구를 통해 선박의 가치 변화를 유발하는 주요 요인들을 선별하여 변수를 설정하고, 2000년 1월부터 2020년 8월까지의 해당 데이터를 확보하였다. 변수의 안정성을 판단하기 위해 다중 공선성 검사를 수행하여 최종적으로 6개의 독립변수와 1개의 종속변수를 선정하고 연구 구조를 설계하였다. 이를 바탕으로 Linear Regression, Neural Network Regression, Random Forest Algorithm을 활용하여 총 9개의 시뮬레이션 모델을 설계하였다. 또한 각 모델간의 비교검증을 통해 평가결과의 정확성을 제고시켰다. 평가 결과, VLCC실제값과의 비교를 통해 2층으로 구성된 Hidden Layer의 Neural Network Regression 모델이 가장 정확도가 높은 것으로 나타났다. 본 연구의 시사점은 첫째, 기존 정형화된 평가기법에서 벗어나 기계학습기반 모델을 선박가치평가에 적용하였다는 점이다. 둘째, 해운시장 변화요인을 동태적 관점에서 분석하고 예측함으로써 연구결과의 객관성을 제고시켰다고 할 수 있다.

핵심용어 : 신경망, 랜덤포레스트, 선형회귀, 선박가치평가, VLCC 중고선가

Abstract : The purpose of this study is to develop the ship valuation model by utilizing the neural network model. The target of the valuation was secondhand VLCC. The variables were set as major factors inducing changes in the value of ship through prior research, and the corresponding data were collected on a monthly basis from January 2000 to August 2020. To determine the stability of subsequent variables, a multi-collinearity test was carried out and finally the research structure was designed by selecting six independent variables and one dependent variable. Based on this structure, a total of nine simulation models were designed using linear regression, neural network regression, and random forest algorithm. In addition, the accuracy of the evaluation results are improved through comparative verification between each model. As a result of the evaluation, it was found that the most accurate when the neural network regression model, which consist of a hidden layer composed of two layers, was simulated through comparison with actual VLCC values. The possible implications of this study first, creative research in terms of applying neural network model to ship valuation; this deviates from the existing formalized evaluation techniques. Second, the objectivity of research results was enhanced from a dynamic perspective by analyzing and predicting the factors of changes in the shipping market.

Key Words : Neural Network, Random Forest, Linear Regression, Ship Valuation, VLCC Secondhand Price

1. 서 론

선박의 가치평가는 해운 시장변화와 밀접한 관계가 있으며 평가가 필요한 시점의 시세를 산정하는 과정이 필요하므로 최적의 평가를 위해 선박의 상태와 같은 외적요인 뿐만 아니라, 체결된 운임계약 등 경제적 요인 등 다양한 시장요

인을 함께 고려해야 한다(Tvedt, 2003). 또한 선박의 담보가치가 경기여건 변화에 따라 일시적으로 하락할 경우 금융권은 추가 담보를 요구하거나 선대매각을 시도하며 해운기업들의 경쟁력 악화의 원인이 되고 있다. 그러나 해운산업은 호황과 불황이 반복되는 순환적 행태가 나타나므로 단기적인 여건변화 보다는 중장기적인 관점에서 보다 객관적인 방법을 통한 선박가치의 평가가 요구된다(Lee, 2017). 이러한 선박가치의 중요성에도 불구하고 국내외적으로 선박가치평가의 체계 개선 및 평가모델의 객관성과 신뢰성을 제고시키기

* First Author : dgkim@mmu.ac.kr, 061-240-7160

† Corresponding Author : jschoi@mmu.ac.kr, 061-240-7163

위한 연구는 부족한 실정이다. 따라서 선박가치의 객관적 평가를 위해 중장기적인 시장변화 요건과 불확실성 등을 평가 요소에 반영할 수 있는 분석방법을 활용한 선박가치평가 모델이 개발된다면 정확하고 투명한 선박가치 평가를 통해 단기 시장 위험성만으로 선박자산을 매각하는 행태를 방지하고, 해운기업의 경쟁력 강화를 견인할 것이다.

본 연구에서는 선박가치평가의 객관성 제고를 위한 대안으로 Neural Network Regression 방법을 이용한 선박가치평가 모델을 개발하고자 한다. 이를 위해 본 연구는 선박가치의 평가대상으로 중급 VLCC선을 선정하여 실증분석을 수행하였다.

연구의 구성은 다음과 같다. 먼저 선박가치에 영향을 미치는 다양한 요소들을 기존 문헌연구 등을 통해 도출해내고 이를 바탕으로 가치평가 모델을 설계하였다. 변수들 가운데는 다중공선성 검증을 통해 데이터의 안정성을 검사하고 최종적으로 도출된 시계열 데이터들을 대상으로 평가 모델을 구축하였다. 다음으로 Microsoft Azure Studio의 Neural Network Regression 모델을 활용해 모델의 구조를 식별하고 가치평가를 위한 최적의 모델을 산출하여 선박가치평가를 수행하였다. 또한 Linear Regression Model과 Random Forest를 이용하여 선박가치를 평가하는 교차비교 분석을 통해 두 분석 기법간의 정확성을 검증하였다. 이를 위해 RMSE(평균제곱근오차)를 통해 실제 선박가치와의 오차를 비교 검토하였다. 마지막으로 본 연구의 결론과 시사점을 도출하였다.

2. 선행연구

전 세계적으로 선박가치평가 체계의 개선 및 평가모델의 객관성 및 신뢰성을 제고시키기 위한 연구는 부족한 실정이며, 일부 특정 선박을 대상으로 몬테카를로 시뮬레이션 모델을 적용한 선박가치평가 연구가 수행되었으나, 평가요인으로 용선료, 이자율, 해체가격만을 반영했다는 점에서 연구결과의 객관성과 신뢰성을 담보하는데 한계점을 가지고 있다(Choi et al., 2015). 현재까지 연구되고 있는 대다수의 선박가치평가의 방법으로는 경제적 타당성을 기반으로 선박의 수익을 합리적으로 추정하기 위한 DCF(Discounted Cash Flow)기반 연구들이 대부분이다(Tvedt, 2006; Adland and Koekebakker, 2007; French and Gabrielli, 2005). 이들 연구에서는 선박가치를 평가하는 기법으로 수익추정 접근방식의 중요도를 주장하면서, 수익추정에 투입되는 변수의 불확실성을 제거하고 합리적 추정을 위한 평가기법의 개선을 강조하고 있다. 그러나 DCF 수익추정방식은 선박의 재무적 가치만을 평가함으로써 해운시장의 변화를 제대로 반영할 수 없다. 이를 입증하며, ROA(Real Option Analysis) 모델을 활용한 선박투자가치

평가 연구가 진행되었으나, 설정변수와 데이터의 제약으로 인해 정확도가 높지 않은 연구의 한계점이 나타났다(Kim, 2009).

반면 Neural Network모델은 FTSE100지수의 옵션 가격 결정에 대해 ANN(Artificial Neural Network)과 Black-Sholes option모형을 통해 비교 분석한 연구결과 ANN의 예측력이 우수하다는 것이 입증되었다(Bennell and Sutcliffe, 2003). 이 밖에도 시계열 평가모델인 ARIMA 기법과 Neural Network 기법을 사용하여 유럽국가의 경기동향을 평가한 논문에서도 Neural Network모델의 정확도가 우수한 것으로 나타났다. 또한 Neural Network모델과 ARIMA를 활용하여 수에즈운하의 통항량을 예측한 후 두 결과를 비교 분석한 연구에서도 Neural Network모델의 우수성이 입증되었으며, Neural Network모델이 예측과 경기분류 연구분야에 많이 사용되고 있으며 기존의 Linear Regression Model과는 다른 관점의 통계기법으로 최근 많이 사용되고 있다고 설명했다(Mostafa, 2006). 특히 Neural Network모델은 시장 변화에 영향을 미치는 요인들이 수요요인, 공급요인, 경제요인, 금융요인 등으로 매우 다양한 해운시장의 변동성을 예측하는데 높은 정확도를 보이고 있다. Neural Network모델을 활용하여 중단기 탱커 운임시장을 예측한 연구에서는 ARIMA 모형과 같은 시계열 예측 모형에 비해 Neural Network모델의 예측 정확도가 높은 것으로 입증되었다(Li and Parsons, 2006).

Neural Network모델을 활용하여 벌크 선박의 가치와의 상관성을 연구한 사례에서는, 선박 가치에 영향을 미치는 6가지 요인에 대해 각각의 요인별 중요도를 분석하여 3가지 요인을 선정한 후 Linear Regression Model과 가치추정치 결과 비교하여 Neural Network모델의 정확도를 입증하였다. 그러나 동 연구는 히든 레이어를 한 개층으로 구성한 단순 Neural Network모델이며, 가치 추정을 위해 3개의 요인만을 사용하였다는 한계점을 내포하고 있다(Lim et al., 2019).

Table 1. Comparison with preceding studies

Type	Existing Study	This Study
Target	1. Tanker Freight Forecasting 2. Drybulk Price Forecasting	Tanker Price Forecasting
Method	1. ANN vs ARIMA 2. ANN vs Linear Regression	Linear Regression vs Neural Network Regression (ANN) vs Random Forest
Limitation	1. Single Hidden Layer 2. 1 Hidden Layer, 3 Independent variable	2 Hidden Layer 6 Independent variable
Researcher	1. Jun and Parsons (2006) 2. Lim et al. (2019)	

따라서 본 연구에서는 Table 1과 같이 선박가치의 정확한 평가를 위해 해운시장의 변화를 반영하는 변수와 데이터 기반의 Neural Network모형을 이용하고, 선박 가치에 영향을 미치는 요인들을 바탕으로 선박의 가치를 측정하는 모델을 개발함으로써 기존 연구들이 갖는 한계점을 극복하고자 한다.

3. 모델 설계

선박가치평가를 위해 사용된 Neural Network Regression Model은 명령을 내리는 과정에서 복잡한 연산 등의 신호를 받아들이는 Input layer와 신호를 전송하는 Output layer의 역할을 수행하는 각각의 노드들이 연결되어 동태적으로 계산을 처리하는 모델이다(McNelis, 2005).

본 연구에서는 Neural Network기법의 일반적인 과정에 따라 총 4단계로 연구를 진행하였다. 첫 번째, 선박가치에 영향을 미치는 변수를 설정하고 유의한 데이터를 확보한다. 둘째, Linear Regression, Neural Network Regression, Random Forest Algorithm을 활용한 모델을 설계한다. 셋째, 각 모델의 시뮬레이션을 수행하여 선박가치를 산출한다. 넷째, 산출된 결과에 대한 정확도를 비교 검증한다.

3.1 변수 설정

선박의 가치에 영향을 미치는 주요 변수들을 식별하여 검증과정을 통해 최종적으로 종속변수인 VLCC secondhand Price 변화에 영향을 미치는 변수들을 선정하였다.

선박가치평가는 선박을 소유한 선사의 재무, 영업, 시장상황 등을 고려하여 선박의 상업적인 가치를 평가하고 금융기관이 이를 담보로 대출금 및 이자율 조정에 활용하는 것을 목적으로 한다. 따라서 선박의 가치에 영향을 미치는 변수를 선정하기 위해서는 선박의 수요를 책임지는 글로벌 경기 여건과 선박의 공급을 담당하는 조선시장, 그리고 운임시장의 변화 등을 종합적으로 검토할 필요가 있다. 이를 위해 본 연구에서는 선행연구에 대한 고찰을 통해 다양한 관점에서 선박가치에 영향을 미치는 요인들을 도출하였다.

먼저, 선박의 가치 변화에 따라 직접적인 영향을 받게 되는 선주, 금융기관, 중개업체 등의 종사자를 대상으로 선박의 가치결정 요인에 대한 중요도를 AHP분석을 통해 도출한 연구결과에서는 1. 선박공급량, 2. 화물수요량, 3 운임, 4. 이자율, 5.유사선박 가격, 6. 선박해체 가격이 중요한 것으로 나타났다(Table 2; Lee, 2017). 또한 Secondhand Price에 영향을 미치는 요인을 Demand factor과 Supply factor, 그리고 Ship condition factor로 구분하여 연구한 결과 변수의 중요도 순위가 Age, TC rate, Newbuilding price, Orderbook/Fleet ratio, Demolition price로 나타났다(Lim et al., 2019). 이 밖에도 선박

의 가치 변화와 운임시장 간의 영향관계를 분석한 연구(Beenstock, 1985)와 신조선 가격과 수요시장간의 관계를 분석한 연구(Tsolakis et al., 2003) 등에서는 모두 선박가치에 수요시장이 중요한 영향을 미친다고 설명했다. 또한 공급측면에서도 선박의 교체수요가 증가하기 위해서는 선박해체 요인이 중요하며, 선박해체 가격과 선박중고선 가격은 밀접한 관계를 형성하고 있다는 연구결과도 나타났다(Thalassinos and Politis, 2014). 이와 같은 문헌 연구를 통해 중복되는 변수를 제외한 8가지 변수를 대상으로 연구 모델을 설계하였다.

Table 2. Description of Research Factor

Researcher	Factor	Description
	Demand	Cargo Demand
Lee (2017)	Supply	Supply, Similar Price, Scrap Value
Lim et al. (2019)		Orderbook/Fleet Ratio, Newbuilding Price
Beenstock (1985)		
Tsolakis et al. (2003)	Economy	TC Rate, Labor Rate
Thalassinos and Politis (2014)	Ship Condition	Age

식별된 변수들을 대상으로 데이터의 안정성을 확인하고자 다중 공선성 검증을 수행한다. 다중 공선성이란 변수 간 상관관계가 높아 종속변수를 설명하는 데 부정적인 영향을 미치는 것을 의미한다. 변수들의 다중 공선성을 확인하기 위해서는 분산팽창요인(VIF, Variance Inflation Factor)지표를 평가한다. VIF는 상관성으로 인해서 발생하는 분산의 증가를 나타내는 계수로서 10이 넘는다면 일반적으로 다중공선성의 문제가 발생한 것으로 확인된다(Lim et al., 2019). 이를 통해 Table 3과 같이 VIF가 10 이상이 되어 데이터의 안정성이 확보되지 않은 Cargo Demand와 Age변수를 본 연구 모델에서 제외하여 최종적으로 6가지 변수를 적용하여 선박가치를 평가하는 모델을 설계하였다.

제외된 변수 가운데 Age의 경우 선박 중고시세에 영향을 줄 수 있는 요인이지만, 선박의 노후화는 운임 수입 감소를 유발하며 두 요인간의 높은 상관관계가 발생하면서 다중공선성 문제가 나타난 것으로 분석된다.

다음으로 안정성 검증 과정을 거쳐 최종 선정된 변수 6가지(VLCC Freight, VLCC Supply, VLCC Scrap Value, VLCC Newbuilding Price, VLCC Orderbook/Supply Ration, Labor Rate)를 대상으로 종속변수인 VLCC Secondhand Price를 평가하기 위해 관련 데이터를 확보하였다. 데이터는 영국의 해운시장 리서치 전문기관인 Clarkson Research에서 매월 발간하는 데

Table 3. Description of Research Factor

Factor	Non Standardization Factor		Standardization Factor	t	P-Value	Multicollinearity	
	B	Standard Error	Beta			Tolerance	VIF
Freight	4.084E-05	.000	.137	3.480	.001	.622	1.609
Supply	-3.073E-07	.000	-.026	-.738	.462	.785	1.274
Scrap Value	-.160	.150	-.053	-1.067	.288	.387	2.582
NB Price	.997	.097	.775	10.244	.000	.1697	5.978
Orderbook/fleet Ratio	.705	.066	.697	10.677	.000	.225	4.452
LIBOR	-5.498	.959	-.498	-5.731	.000	.127	7.892
Cargo Demand	.161	.081	.212	2.002	.048	.085	11.731
Age	2.623	.236	.913	11.121	.000	.142	10.033

Table 4. Variable and Data

Type	Contents
Dependent Variable	VLCC Second-Hand Price
Independent Variable	1. VLCC Freight
	2. VLCC Supply
	3. VLCC Scrap Value
	4. VLCC Newbuilding Price
	5. VLCC Orderbook/Supply ratio
	6. LIBOR Interest Rate
Data to Use	Jan 2000 ~ Aug 2020 Monthly Data 248 Data
Resource	Clarkson Research Intelligence

이터를 기반으로 Table 4과 같이 2000년 1월~2020년 8월까지의 248개월간 데이터를 확보하여 분석을 수행하였다.

Fig. 1~6은 학습에 사용될 6가지의 독립변수이며 각각 VLCC Freight, VLCC Supply, VLCC Demolition Price, VLCC Newbuilding Price, VLCC Orderbook/Supply Ration, Libor Rate의 월별 변화를 나타낸다. Fig. 2의 VLCC Supply의 경우, 변화의 정도가 매우 심한 것을 알 수 있다.

Fig. 3의 경우, Scrap Value가 2008년 9월 크게 하락 후 2011년 5월까지 서서히 증가하는 경향을 보였다.

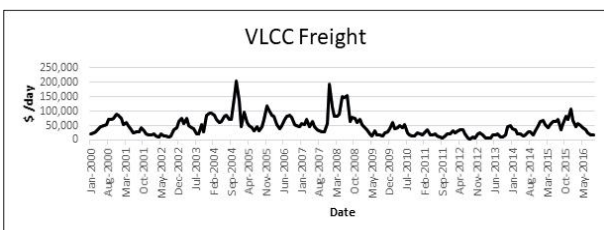


Fig. 1. Monthly Change of VLCC Freight.

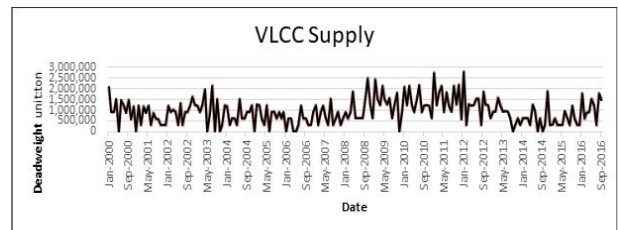


Fig. 2. Monthly Change of VLCC Supply.

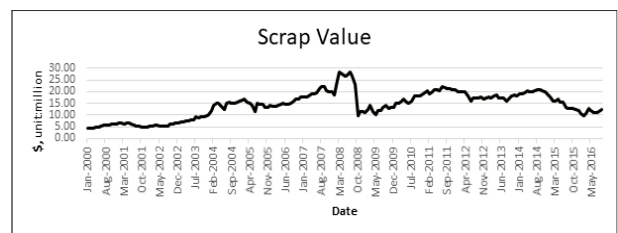


Fig. 3. Monthly Change of Scrap Value.

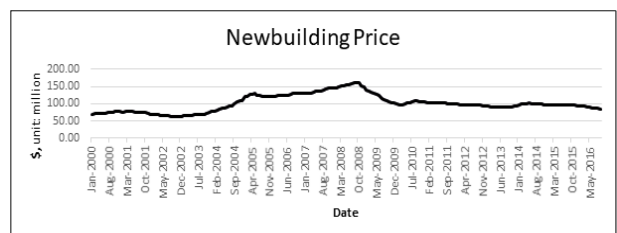


Fig. 4. Monthly Change of New building Price.

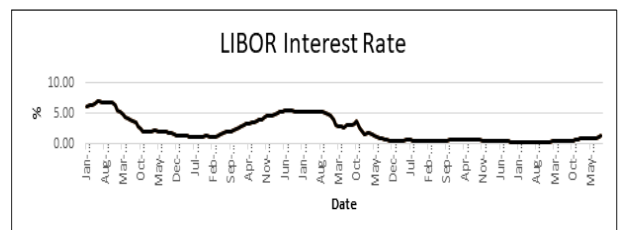


Fig. 5. Monthly Change of LIBOR Interest Rate.

신경망기법을 활용한 선박 가치평가 모델 개발

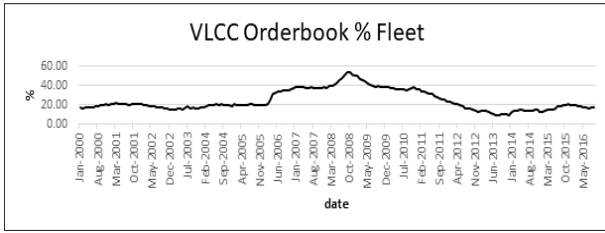


Fig. 6. Monthly Change of VLCC Orderbook.

Fig. 7은 학습에 사용될 label로 VLCC Secondhand Price의 월별 변화를 나타낸다.

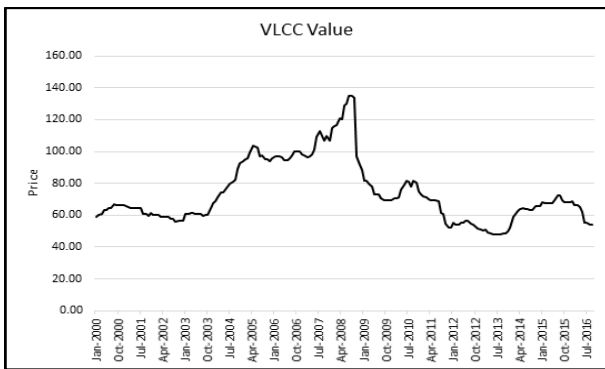


Fig. 7. Monthly Change of VLCC Secondhand Price.

3.2 모델 생성

VLCC Secondhand Price를 추정하기 위해 Linear Regression, Neural Network Regression, Random Forest를 사용하였다. 학습에는 전체 데이터 중 80%를 모델 학습에 사용하였으며 나머지 20%를 테스트용으로 사용하였다. 따라서 전체 248개 데이터 가운데 198개는 학습용으로 사용되고, 50개는 테스트용으로 활용되었다.

Table 5. Generated Models for estimating VLCC second-hand price

Algorithm	Number of Generated Models	Remarks
Linear Regression	1	Each feature's weights can be computed-easy to understand.
Neural Network Regression	4	Two kinds of the number of hidden layer's node. Two kinds of the number of hidden layers
Random Forest	4	Two kinds of decision trees. Two kinds of maximum depth of the trees.

Table 5는 본 연구에 사용된 알고리즘과 이를 활용해 생성된 모델의 수를 의미한다. Linear Regression의 경우 한 종류의 모델만을 생성하였다. Neural Network Regression과 Random Forest의 경우, 서로 다른 두 가지의 파라미터를 설정한 후 이들을 조합하여 각각 4개의 모델을 생성하였다. 먼저 Linear Regression에 대해 설명한다.

1) Linear Regression Model

Ordinary Least Squares(최소제곱법)을 활용하여 실제 해와 근사적으로 구하려는 오차의 제곱의 합이 최소가 되도록 하여 가중치를 결정한다. 이 OLS는 Linear Regression Model뿐만 아니라 Non Linear Regression Model에도 사용될 수 있다.

데이터에 너무 과적합될 경우 새로운 데이터를 정확하게 예측할 수 없다. 이 때 과적합의 경향을 낮추기 위한 방법이 Regularization이다.

머신러닝을 활용해 모델을 학습시킬 때 제공된 데이터에 맞게 과도하게 학습되지 않도록 비용함수에 어떤 항을 추가하여 준다. 본 논문에서는 L2 Regularization(또는 Ridge)라고 불리는 값을 비용함수에 추가하였다. L2는 제곱을 취하므로 실제 값과 오차가 클 경우, 더욱 큰 L2 Loss가 발생된다.

2) Neural Network Regression

Fig. 8은 Neural Network Regression 모델의 구성을 나타낸다. Input layer에는 Bias를 포함해 VLCC Freight, VLCC Supply, Scrap Value, Newbuilding price, LIBOR Interest rate, VLCC Orderbook을 입력 변수로 설정하였다. 히든 레이어1에서는 입력 레이어에서 받은 값과 가중치를 곱해 각각의 합을 더한 후 식(1)의 sigmoid 활성화 함수의 입력으로 전달한다.

$$\phi(z) = \frac{1}{1 + e^{-z}} \quad (1)$$

1층의 i번째 노드와 l+1층의 j번째 노드를 연결하는 가중치 $w_{j,i}^{(l)}$ 이며 1층의 i번째 노드는 $a_i^{(l)}$ 이다.

예를 들어 $z_1^{(2)} = w_{1,1}^{(1)} \cdot a_1^{(1)} + w_{1,2}^{(1)} \cdot a_2^{(1)}$ 에서 $z_1^{(2)}$ 은 은닉층의 입력 함수이며, $a_1^{(1)}$ 은 입력층의 첫 번째 노드를 의미한다. $z_1^{(2)}$ 은 활성화 함수의 입력값으로 사용된다. 히든 레이어 2도 히든 레이어 1과 같은 방식으로 히든 레이어 1에서 받은 값과 가중치를 곱해 Sigmoid 활성화 함수를 통해 출력 레이어로 전달한다. 출력 레이어는 히든 레이어2에서 전달받은 값과 실제 VLCC Second-hand Price를 비교한다.

이 때 Neural Network에서 평균 제곱 오차(MSE)를 비용함수로 사용하였다. 비용함수는 식(2)와 같다.

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (y - \hat{y})^2 \quad (2)$$

y 는 중고선박의 가치를 의미하며 \hat{y} 는 모델이 예측한 중고선의 가치를 의미한다.

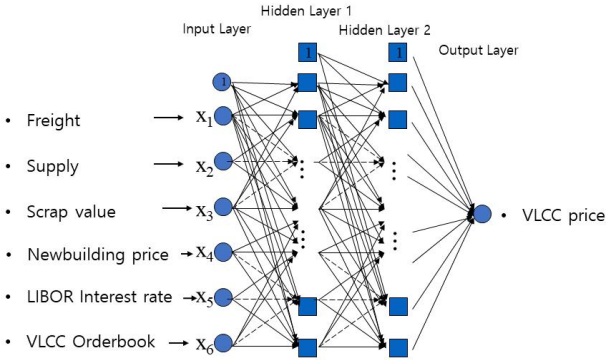


Fig. 8. Neural Network Regression Model for VLCC second-hand price.

학습된 값과 실제 VLCC Secondhand Price가 같아지도록 역전파방식을 통해 가중치를 수정한다. 과적합을 막기 위해 조기 종료 기능을 설정하였다. 학습 시 5회 이상 성능이 좋아지지 않을 경우, 학습이 종료된다.

또한 사용된 각 데이터 특징들의 값의 범위 차이가 크다고 판단하여 모든 데이터를 [0, 1] 범위로 만드는 최소최대 정규화를 적용하였다. 이는 학습이 안정적이고 빠르게 실행되기 위함이다.

추가로 Neural Network Regression의 경우, 하이퍼 파라미터에 따라 4개의 모델을 생성하였다. 하이퍼 파라미터를 찾는 것은 어려운 작업이지만 경험적으로 히든 레이어의 노드의 수를 탐색하기 위해서는 다음과 같은 방법으로 시작하는 것을 제안하고 있다(Jeff, 2008; Roger et al., 2015). 히든 레이어의 노드 수는 1) 입력 레이어의 노드 수와 출력 레이어의 노드 수 사이의 값 또는 2) 입력 레이어의 노드 수와 출력 레이어의 노드수의 총합의 2/3의 값 또는 3) 입력 레이어의 노드 수의 2배보다 작은 값을 추천한다. 본 연구에서 첫 번째 방법을 사용하였으며 히든 레이어의 노드의 수를 입력 노드 수와 같도록 설정하였다.

히든 레이어 층을 각각 1층과 2층으로 구성되도록 설계하였다. Table 6과 같이 히든 레이어의 노드 수와 히든 레이어

층수의 조합으로 총 4개의 Neural Network Regression Model을 생성하였다.

Table 6. Initial settings for Neural Network Regression Models

Model	1L6N	1L12N	2L6N	2L12N
Number of Nodes	6	12	6	12
Number of hidden layers	1		2	
Early termination	MedianStoppingPolicy (evaluation_interval=1, delay_evaluation=5)			
Activation function	Sigmoid			
Loss function	Root Mean Square Error			

예를 들어 Table 6의 모델 1L6N은 1층의 히든 레이어로 구성되었으며 이 때 이 1층의 히든 레이어의 노드 수는 6개를 의미한다. 같은 방법으로 2L6N은 2층의 히든 레이어로 구성되었으며 각각의 히든 레이어는 6개의 노드로 구성되어 있다.

3) Random Forest

Random Forest는 분류와 회귀 등에 적용 가능한 앙상블 학습 방법이다. 훈련과정에서 다수의 결정 트리가 생성되며 Linear Regression 분석을 출력한다. 가장 큰 특징은 랜덤성에 의해 트리들이 조금씩 다른 특성을 갖는다는 것이다. 중요한 매개변수로는 Forest의 크기, 즉 트리의 개수와 최대 허용 깊이이다. 본 연구에서는 트리의 개수를 8, 16개의 두 가지 경우와, 최대 깊이가 32, 64 두 가지 경우의 조합으로 모델을 설계하였다. Resampling으로는 bagging을 적용하였으며 Table 7과 같이 설정하였다.

Table 7. Initial Settings for Random Forest Models

Model	RF-1	RF-2	RF-3	RF-4
Number of decision trees	8	16	8	16
Maximum depth of the decision trees	32		64	
Number of random splits per nodes	128			
Minimum number of samples per leaf node	1			
Re-sampling method	Bagging			

신경망기법을 활용한 선박 가치평가 모델 개발

Fig. 9는 16개의 결정 트리와 최대 깊이 32의 깊이를 갖는 RF-2 모델의 예이다.

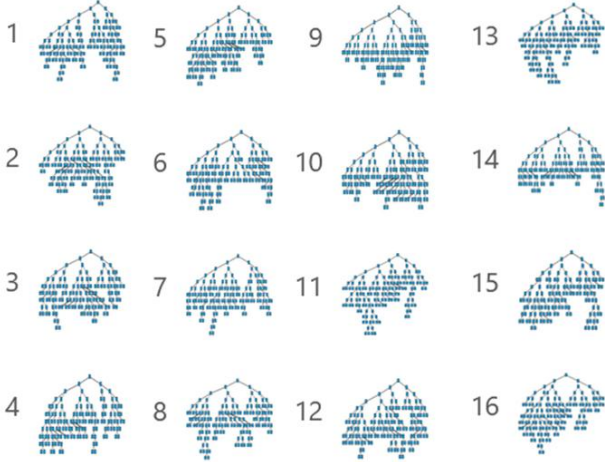


Fig. 9. Instance of RF-2 with 16 decision trees and 32 maximum depth.

실험 환경은 다음과 같다. Windows 10 Home, AMD 3700X, RAM: 64GB, RTX 2070 그래픽카드, Microsoft Azure Studio에서 실험을 진행하였다.

4. 실험 결과

본 연구에서는 VLCC Secondhand Price의 가치를 측정하기 위해 Linear Regression, Neural Network Regression, Random Forest 알고리즘을 사용하여 실험을 진행하였다. 총 9개의 모델은 2016년 10월 이후의 VLCC Secondhand Price를 예측하도록 하였다. 또한, 실제값과 예측값의 차이인 Error를 비교하였다.

4.1 각 모델이 예측한 VLCC Second-hand Price

Fig. 10은 9가지 모델과 실제값을 나타낸 그래프이다. 모든 모델은 실제값의 추이를 예측하는 경향을 보여주었다. 그러나 실제 중고선박의 가치의 변화에 대해 민감하게 반응하는 경향을 보였다. 이는 Fig. 7에 사용된 훈련 데이터의 경우 큰 폭으로 증가 감소하는 경향을 보이는데 이러한 경향을 학습하여 테스트 데이터에 적용한 결과로 분석된다.

결과를 이해하기 쉽도록 Fig. 11과 같이 RMSE가 가장 작은 2L6N와 RMSE가 가장 큰 1L6N 그리고 LR만을 따로 나타내었다. LR은 실제 값보다 조금 더 높게 예측하는 경향을 보였다. Neural Network Regression 모델은 LR이 예측한 값보다 낮게 예측하는 경향을 보였다.

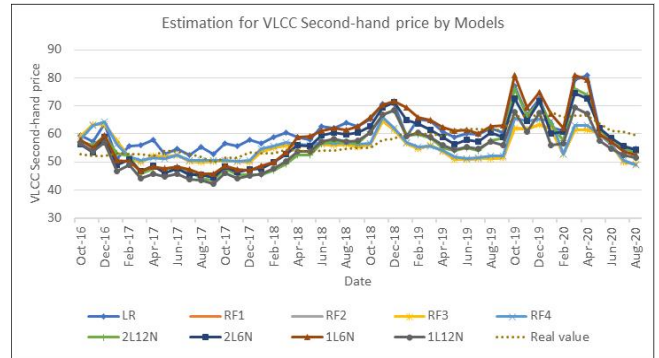


Fig. 10. Estimation for VLCC Secondhand price by Models.

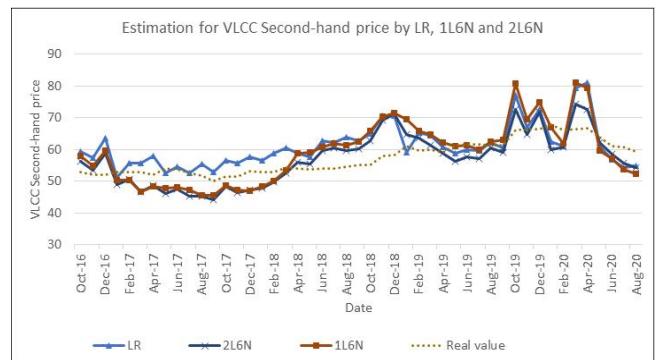


Fig. 11. Estimation for VLCC Secondhand price by LR, 1L6N and 2L6N.

Table 8은 Linear Regression Model의 학습 결과에 따른 입력 변수의 가중치를 의미한다. BIAS가 가장 큰 가중치를 가졌다. BIAS가 높다는 의미는 예측값과 실제값과의 차이를 의미한다. 이는 Fig. 10, 11과 같이 각각의 모델들이 예측한 값은 실제 값과 비슷한 경향을 보였으나 실제 값의 변화 정도에 비해 모델들의 예측값이 크게 변화하는 경향을 보였기 때문이다. 이는 학습에 사용된 데이터의 변화 정도가 테스트에 사용된 데이터의 변화 정도에 비해 더 컸기 때문인 것으로 판단된다. 두 번째로 큰 가중치는 LIBOR Interest Rate로 런던의 주요 은행 사이에서 단기자금을 조달하는 이자율을 의미한다. 이자율 변화에 따라 자금조달 금리가 결정되어 중고선박 구입을 위한 수요변화 및 중고선 가치에도 영향을 미친다고 판단된다. 세 번째로 큰 가중치는 Newbuilding Price로 VLCC선 신조선박의 가격을 의미한다. 신조선박의 가격이 상승하면 중고선 가치에도 정(+)의 영향을 미친다고 판단된다. 네 번째로 큰 가중치는 VLCC Order book으로 VLCC선이 조선소에서 건조되고 있는 발주잔량을 의미한다. 건조중인 선박량에 따라 향후 공급과잉 또는 부족을 유발하여 중고선 가격변화에도 영향을 미친다고 판단된다. 다섯 번째로 큰 가중치는 Scrap Price로 고철 가격을 의미한다. 단순히 고

철의 가치가 높아질 경우, 중고선박의 가치도 높아진다고 판단된다. 여섯 번째로 큰 가중치는 VLCC Freight로 VLCC선의 운임을 의미하며, 운임시장의 변화에 따라 중고선의 수요가 결정되므로 중고선 가격에도 영향을 미친다. 가장 작은 가중치는 VLCC Supply이며 이는 선복량을 의미하여 유일하게 -값을 가진다. 선복량이 증가하면 중고선의 가치는 떨어질 것이라는 상식에는 맞으나 그 영향이 다른 변수에 비해 가장 작은 결과를 나타냈다.

Table 8. Variable and Value for Linear Regression

Feature	Weight
Bias	11.0362
LIBOR Interest Rate	2.62126
Newbuilding Price	0.387375
VLCC Orderbook	0.373167
Scrap Price	0.293133
VLCC Freight	0.00015588
VLCC Supply	-0.00000280495

4.2 각 모델의 정확도 비교

Fig. 12는 각 모델의 오차를 의미한다. 가장 낮은 오차를 기록한 모델은 2L6N로서 두 개의 히든 레이어 층과 6개의 히든 레이어 노드를 가지는 Neural Network Regression 모델이 기록하였다. 뒤이어 2L12N이 두 번째로 낮은 오차를 기록하였다. 같은 수의 히든 레이어 층수라도 히든 레이어의 노드가 증가할 경우, 오차가 증가하는 것을 확인하였다. RF-2와 RF-4 모델이 세 번째로 낮은 오차를 기록하였다. 가장 오차가 큰 모델은 1L6N이 기록하였다. 이는 다른 Neural Network Regression 모델에 비해 층수가 얇고 히든 레이어의 노드 수가 부족하기 때문인 것으로 보인다. 1L6N에서 노드 수만 2배 증가한 1L12N의 경우 오차가 낮아지는 것을 확인하였다.

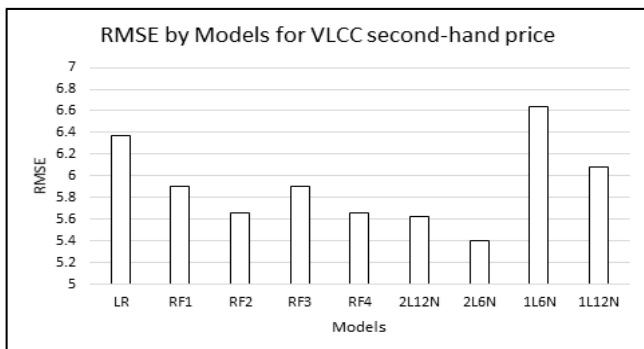


Fig. 12. RMSE by Models.

5. 결론

본 연구는 세 가지 알고리즘을 활용하여 선박가치를 평가하고 예측하기 위해 수행되었다. 선형 위주의 예측 모델의 한계점을 보완하기 위해 Neural Network Regression과 Random Forest 알고리즘을 활용하여 9개의 모델을 생성하였다. 이를 활용해 선박가치 평가결과의 정확도를 제고하는 모델을 제시하였다. 실증분석을 위한 가치평가 대상은 중고 VLCC선으로 선행연구와 다중공선성 검증을 통해 확인된 6가지의 독립변수(VLCC Freight, VLCC Supply, VLCC Demolition Price, VLCC Newbuilding Price, VLCC Orderbook/Supply Ration, Labor Rate)를 활용하여 가치평가 및 예측을 수행하였다. 제안된 모델의 정확도를 검증하기 위해서 각 모델이 예측한 값과 실제 선박가치를 비교하였으며, 예측결과에 대해 RMSE로 오차율을 확인하였다.

분석 결과, 2L6N의 6개의 히든 레이어 노드를 가지는 2개의 히든 레이어를 가지는 모델의 성능이 가장 좋게 나타났다. 이때 RMSE는 5.404767를 기록하였다. 그러나 Neural Network Regression의 하이퍼 파라미터를 찾는 것은 실험을 통해서만 알 수 있었다.

테스트에 사용된 실제 VLCC Secondhand Price의 변화폭이 작음에도(증가 또는 감소) 모델은 값의 변화보다 더 큰 변화 양상을 보였다. 이는 학습에 사용된 2000년부터 2016년 10월까지의 실제 데이터 VLCC Secondhand Price에서 큰 폭으로 증가 또는 감소한 부분이 존재하였으며 이러한 경향도 Neural Network Regression이 학습하였기 때문으로 판단된다.

본 연구의 시사점은 첫째, 기존 정형화된 평가기법에서 벗어나 기계학습기반 모델을 선박가치평가에 적용한 측면에서 독창적인 연구라고 할 수 있다.

둘째, 해운시장 변화요인을 동태적 관점에서 분석하고 예측함으로써 연구결과의 객관성을 제고시켰다고 할 수 있다. 반면 실증분석 대상이 VLCC선 한 선종에 한정되었다는 점에서 선박가치평가 모델의 일반화를 판단하기에는 한계점을 내포하고 있다. 또한 평가에 사용된 데이터의 샘플 수를 248개 이상을 사용했다면 보다 정교한 학습모델을 개발할 수 있을 것이다. 마지막으로 MS 애저 머신러닝 프로그램의 실험 환경 제약으로 인해 선형회귀 모델에 사용한 L2 Regularization을 뉴럴네트워크 모델에도 동일하게 적용하지 못한 점은 실험결과 비교의 정확도를 높이지 못하는 한계점을 가지고 있다.

따라서 향후 추가적인 선종과 데이터를 확보하여 보다 많은 사례와 새로운 실험환경과 연구방법을 통해 선박가치평가 추가 연구가 진행될 필요성이 있다.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5A8042768).

References

- [1] Adland, R. and S. Koekebakker(2007), Ship Valuation Using Cross-Sectional Sales Data: A Multivariate Non-Parametric Approach, *Maritime Economics and Logistics*, Vol. 9, No. 2, pp. 105-118.
- [2] Beenstock, M.(1985), A theory of ship prices, *Maritime Policy and Management*, Vol. 12, No. 3, pp. 215-225.
- [3] Bennell, J. A. and C. M. Sutcliffe(2003), Black-Scholes Versus Artificial Neural Networks in Pricing FTSE 100 Options, *SSRN Electronic Journal*, Vol. 12, No. 4, pp. 243-260.
- [4] Choi, J. S., K. H. Lee, and J. S. Nam(2015), A Ship Valuation Model Based on Monte Carlo Simulation, *Journal of Korea Port Economic Association*, Vol. 31, No. 3, pp. 1-14.
- [5] French, N. and L. Gabrielli(2005), Discounted Cash Flow: Accounting for Uncertainty, Working Paper of IUAV Venice University.
- [6] Jeff, H.(2008), Introduction to Neural Networks for Java, Heaton Research, p. 129.
- [7] Kim, C. Y., D. K. Ryoo, and J. K. Kim(2009), Real Option Analysis on Ship Investment Valuation, *Journal of Navigation and Port Research*, Vol. 33, No. 7, pp. 469-476.
- [8] Lee, S. Y.(2017), An Empirical Study on the Influence Factors of Ship Valuation, *Journal of Finance and Public Assets*, Vol. 1, No. 2, pp. 33-60.
- [9] Li, J. and M. G. Parsons(2006), Forecasting tanker freight rate using neural networks, *Maritime Policy and Management*, Vol. 24, No. 1, pp. 9-30.
- [10] Lim, S. S., K. H. Lee, H. J. Yang, and H. S. Yun(2019), Panamax Second-hand Vessel Valuation Model, *Journal of Navigation and Port Research*, Vol. 43, No. 1, pp. 72-78.
- [11] McNelis, P. D.(2005), *Neural Networks in Finance: Gaining Predictive Edge in the Market*, Elsevier Academic Press.
- [12] Mostafa, M. M.(2006), Forecasting the Suez Canal traffic: a neural network analysis, *Maritime Policy and Management*, Vol. 31, No. 2, pp. 139-156.
- [13] Roger, B., F. Valentine, and T. Wee-Hyong(2015), *Predictive Analytics with Microsoft Azure Machine Learning Second Edition*, Apress, p. 156.
- [14] Thalassinos, E. I. and E. D. Politis(2014), Valuation Model for a Second-hand Vessel: Econometric Analysis of the Dry Bulk Sector, *Journal of Global Business and Technology*, Vol. 10, No. 1, pp. 1-17.
- [15] Tsolakis, S. D., C. Cridland, and E. H. Hercules(2003), Econometric Modelling of Second-Hand Ship Price, *Maritime Economics and Logistics*, Vol. 5, No. 4, pp. 347-377.
- [16] Tvedt, J.(2006), Valuation of VLCCs under Income uncertainty, *Maritime Policy and Management*, Vol. 2, pp. 159-174.
- [17] Tvedt, J.(2003), A new perspective on price dynamics of the dry bulk market, *Maritime Policy and Management*, Vol. 30, No. 3, pp. 221-230.

Received : 2020. 10. 21.

Revised : 2020. 12. 09. (1st)

: 2020. 12. 29. (2nd)

Accepted : 2021. 02. 25.