JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 26 No. 7, pp. 127-132, July 2021
https://doi.org/10.9708/jksci.2021.26.07.127

Forecasting Bulk Freight Rates with Machine Learning Methods

Sangseop Lim*, Seokhun Kim**

*Professor, Div. of Navigation Convergence Studies, Korea Maritime and Ocean University, Busan, Korea
**Professor, Dept. of Electronic Commerce, PaiChai University, Daejeon, Korea

[Abstract]

This paper applies a machine learning model to forecasting freight rates in dry bulk and tanker markets with wavelet decomposition and empirical mode decomposition because they can refect both information scattered in the time and frequency domain. The decomposition with wavelet is outperformed for the dry bulk market, and EMD is the more proper model in the tanker market. This result provides market players with a practical short-term forecasting method. This study contributes to expanding a variety of predictive methodologies for one of the highly volatile markets. Furthermore, the proposed model is expected to improve the quality of decision-making in spot freight trading, which is the most frequent transaction in the shipping industry.

▸ **Key words**: Baltic Freight Rate, BCI, BPI, BDTI, BCTI, Decomposition, ANN

[요 약]

본 논문은 건화물시장과 탱커시장의 운임지수 예측에 관하여 머신러닝을 적용하였으며 신호분해법인 웨이블릿 분해와 EMD분해를 데이터 전처리 과정에 반영하여 시간의 영역의 정보와 주파수 영역의 정보를 모두 반영할 수 있는 운임예측모형을 구축하였다. 건화물 시장의 경우 웨이블릿으로 분해한 예측모형이 우수하였으며 탱커시장의 EMD분해로 예측한 모형이 우수하였으며 실무적으로 각 운송시장 참여자들에게 새로운 단기예측 방법론을 제시하였다. 이러한 연구는 운송시장에서 양적으로 가장 중요한 건화물 시장과 탱커시장에 대한 다양한 예측방법론을 확대하고 새로운 방법론을 제시하였다는 측면에서 중요하며, 변동성이 큰 운임시장에서 과학적인 의사결정 방법에 대한 실무적인 요구를 반영할 수 있을 뿐만 아니라 가장 빈번한 스팟거래에 합리적인 의사결정이 이뤄질 수 있는 기초가 될 것으로 기대된다.

▸ **주제어**: 발틱운임지수, BCI, BPI, BDTI, BCTI, 시계열 분해, 인공신경망

• First Author: Sangseop Lim, Corresponding Author: Seokhun Kim
*Sangseop Lim (limsangseop@kmou.ac.kr), Div. of Navigation Convergence Studies, Korea Maritime and Ocean University
**Seokhun Kim (vambition@daum.net), Dept. of Electronic Commerce, PaiChai University
• Received: 2021. 06. 28, Revised: 2021. 07. 26, Accepted: 2021. 07. 26.

Copyright © 2021 The Korea Society of Computer and Information
http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945
www.kci.go.kr

I. Introduction

세계 해운시장은 크게 건화물시장, 탱커시장, 컨테이너시장 등으로 구분된다. 건화물시장의 화물은 철광석과 석탄이고, 탱커시장은 원유 및 화학제품, 컨테이너시장은 완제품(부품류포함)을 운송하는 시장이다. 2020년 기준, 세계 해상물동량 중 건화물 약 45%, 원유 및 컨테이너가 각각 약 16%씩 비중을 차지하고 있다. 이러한 해상물동량은 2008년 금융위기와 2020년 코로나 팬데믹으로 인한 충격을 감안하더라도 2000년부터 연평균 약 3.3% 성장하고 있다.

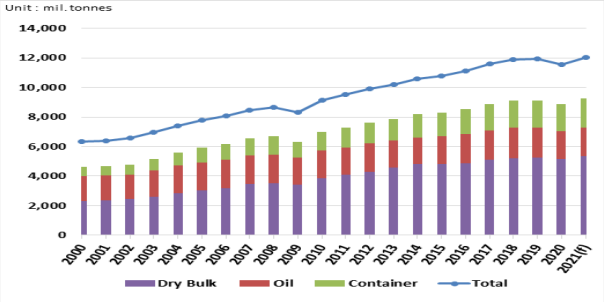

Fig. 1. Trend of World Seaborne Trade
Source: Clarkson Research

해상물동량 중 가장 비중이 큰 건화물 물동량은 철광석과 석탄이며 각각 15억톤(29.1%), 11억톤(22.5%)으로 50%가 상회하는 점유율을 보이며 이시장의 운송시장에는 케이프(Capesize) 선박과 파나막스(Panamax) 선박이 투입된다.

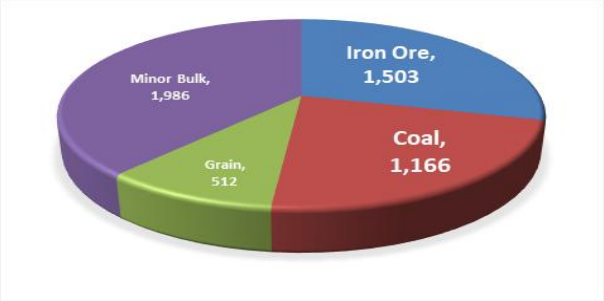

Fig. 2. Dry bulk commodities(million ton)
Source: Clarkson Research

또한 원유를 운반하는 탱커시장은 주로 VLCC(very large crude oil carrier)가 투입되다. 화물 운송하는 해운시장에서 운송서비스 이용 및 제공에 대한 운임을 거래하며 화물을 가진 화주의 측면에서는 운임이 비용이 되며 선박을 가진 선주 측면에서는 수익이 된다. 이를 운임거래라고 한다.

해상운임거래는 크게 일정기간을 기준으로 하는 기간용선거래(Time-Charter, T/C)와 항차당 기준인 스팟거래(Spot)로 나눌 수 있으며 아래 표와 같이 거래의 빈도상 스팟거래가 파나막스선형의 경우 약 11배, 케이프사이즈 선형은 약 24배에 이를 정도로 용선거래보다 압도적으로 빈번하게 일어난다. 이러한 현상은 탱커시장에서 더 심하게 나타나는데 2018년이후 기간용선거래는 55건만 성약이 되었으나 스팟거래는 5,066건의 거래가 이뤄져 약 90배 이상 빈번하게 거래되는 것을 알 수 있다.

Table 1. Fixtures Volume of Spot and T/C contracts in Bulk sector

Year	Capesize		Panamax	
	T/C	Spot	T/C	Spot
2015	39	1,316	139	1,308
2016	79	2,127	306	2,317
2017	115	2,044	223	2,400
2018	110	1,837	259	2,463
2019	47	1,320	111	2,391
2020	15	1,379	104	2,431
2021	33	697	120	1,135
Sum	438	10,720	1,262	14,445

Source: Clarkson Research (May/2015~May/2021), search date: 2021.06.24.

위의 데이터에서와 같이 스팟거래는 일반적으로 20~50일에 해당하는 항차를 기준으로 운임계약이 이뤄지는 반면에 기간용선거래는 짧게는 3개월에서 길게는 3~5년, 극단적으로는 선박의 선체수명에 해당하는 10년이상의 긴 용선거래가 있다. 스팟거래가 기간용선거래에 비해 단기거래이기 때문에 빈번하게 일어나며 그만큼 해운경영에서 가장 중요한 의사결정이라고 할 수 있다[1].

단기거래의 특징을 보이는 스팟거래에서 운임의 단기예측에 기반한 의사결정이 중요하며 이를 기준으로 화주는 거래 체결을 연기하거나 조기에 시행하여 운송비용을 최소화하는 경영전략을 수립할 수 있으며 선박을 가진 선주나 용선주의 경우 화주와 반대로 수익을 극대화할 수 있는 전략을 구사할 수 있다.

해운시장은 가장 변동성이 심한 시장으로 알려져 있으며 일례로 BDI기준으로 2008년 5월 11,793포인트를 기록하다 그해 12월 663포인까지 불과 6개월만에 약 95% 수준으로 급락한 것을 보면 변동성이 얼마나 심한 시장인지 알 수 있다. 그러나 해운시장에서 주요 참여자들, 특히 해운기업들은 주요 의사결정을 내리는데 있어 경영자들의 오랜 경험과 통찰력에 상당히 의존적인 경향이 있다. 시황변동위험에 제대로 대응하지 못한 과거의 실패 사례를 반복하지 않기 위해서는 다양한 과학적인 의사결정방법에 대한 연구와 실무적인 요구가 절실한 상황이다[2].

www.kci.go.kr

따라서 본 논문은 해운시장에서 가장 규모가 큰 건화물시장과 탱커시장을 대상으로 하였으며 케이프벌크선형과 파나막스선형의 스팟거래를 위한 운임단기예측에 기계학습모형을 이용하여 과학적인 의사결정 방법을 제시함으로써 합리적인 의사결정모형에 대한 실무적 필요를 채우고자한다.

기계학습 모델이나 하이브리드 모델을 운임예측에 활용한 연구사례는 종종 수행되어 왔는데 탱커운임지수[3,4], 파나막스운임지수(BPI)[5], 벌크운임지수[6,7]등이 대상이었다. 기존의 연구들은 하이브리드모델이 단순 기계학습모델에 비해 우수하다고 밝힌 기존 연구의 결과는 하이브리드 모델이 단순 기계학습모델에 비해 우수하다고 밝히고 있다.

단순한 기계학습 모델보다 하이브리드 형태의 모델이 시계열 예측에 더 적합한 이유는 대부분의 시계열이 비선형적인 특성을 가지기 때문이다. 따라서 데이터 전처리과정을 거치면 기계학습의 학습효과를 극대화시킬 수 있을 것이다. 운임지수의 경우 많은 경제 시계열에서 보이는 불안정 특성을 가지며 이를 분해하여 분석할 수 있는 과정이 포함되면 예측성능을 개선하는데 도움이 된다고 여러 선행 연구결과에서 확인된 바 있다[8,9,10].

본 논문에서는 데이터 전처리 과정으로 분해기법으로 알려진 웨이블릿(wavelet)분해와 경험적모드분해(EMD) 각각을 기계학습모델을 결합한 하이브리드모델을 적용할 것이다. 또한 건화물 시장을 대표하는 지수로서 Baltic Capesize Index(BCI), Baltic Panamax Index(BPI)가 있으며 탱커시장은 Baltic Dirty Tanker Index(BDTI), Baltic Clean Tanker Index(BCTI)가 있고 이 중 BCI와 BDTI만을 연구의 데이터로 활용하고자 한다. 기존 연구에서 두 시장을 광범위하게 조사된 연구가 없으며 제시된 모델들을 종합적으로 비교하여 각 시장에 적합한 연구모델을 제시하고자 한다.

II. Data and Modelling

1. Data

본 연구는 건화물 및 원유 운송지수를 대상으로 하였으며, Clarkson Research에서 제공하는 발틱운임지수인 BCI, BPI, BDTI, BCTI의 월별 데이터를 이용하고자 한다[11]. 아래 Fig. 3과 Fig. 4를 보면 2008년 세계금융위기 이후 해운시장의 장기침체에 있으며 이후로 상당히 안정화 되어있는 점을 반영하여 2010년 이후의 데이터를 활용하고자 한다.

Fig. 3. Baltic Dry Index(BCI, BPI)
Source: Clarkson Research

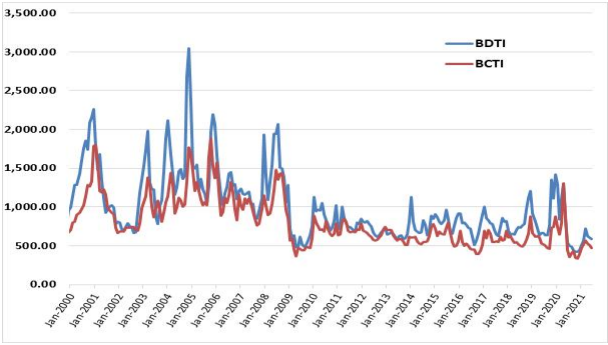

Fig. 4. Baltic Tanker Index(BDTI, BCTI)
Source: Clarkson Research

인공신경망 모형의 학습과 예측성능 평가를 위해 Table 2와 같이 데이터를 훈련샘플과 검증샘플 약 9:1 비율로 나누어 수행하였다.

Table 2. Description of Data

Data	Periods(monthly)	
• Baltic Capesize Index • Baltic Panamax Index • Baltic Dirty Tanker Index • Baltic Clean Tanker Index	2010.1 ~ 2021.6 (138)	
	Training Set	Test Set
	2010.1~2020.6 (126)	2020.5~2021.6 (12)

Source: Clarkson Research

2. Modelling

본 연구에서는 운임시계열을 분해하였는데 기존의 시계열분석은 시간영역의 정보만을 추출하는 반면에 웨이블릿(wavelet)분해나 EMD 분해는 주파수영역의변동에 대한 정보들을 파악하여 모형에 반영할 수 있다. 이러한 분해법을 이용하여 운임시계열의 데이터 전처리과정을 거친 후 이를 입력변수로하는 인공신경망 모형을 구축할 것이다. 본 연구에서 인공신경망을 적용하는 이유는 전통적인 시계열모델들은 선형모델로서 연구 대상이 되는 시계열 데이터의 불안정성(non-stationarity) 제약이 있는데 이를

www.kci.go.kr

안정화 단계에서 많은 정보의 소실이 생기는 제약이 따른다[12]. 연구의 대상이 되는 시계열 데이터들은 불안정적이기 때문에 위와 같은 위의 인공신경망 모델을 적용하여 운임을 예측하고자 한다[13,14].

첫 번째 분해 방법은 경험적모드분해로 원래 시계열을 IMF(intrinsic mode function)라고 불리는 몇 개의 모드(mode)로 분해하는 방법이며 Table 3과 같다. 이러한 방법은 원시계열의 고유한 속성을 보존할 수 있는 장점이 있다[15].

Table 3. Procedure of Calculating IMF

(1) Initial default $i=1, s=y(t)$
(2) Identify local extremums in sereis s
(3) Create upper envelope(s_u) with local maximum value and lower envelope(s_l) by cubic spline respectively
(4) Calculate mean envelope(s_m)
(5) Extract IMF_i by $s-s_m$
(6) Repeate (1)~(5) until the rest has only one extremum.

웨이블릿(wavelet)이론은 시간의 신호를 주파수 영역으로 변환하여 분석하는 이론이다. 푸리에 변환은 원 신호에 대한 시간의 정보를 해석할 수 없는 반면에 웨이블릿(wavelet) 변환은 시간과 주파수에 대한 정보를 모두 다 반영할 수 있는 장점이 있어 시계열 분해에 이용되어 왔다[16]. 다음식은 웨이블릿 변환 함수이다.

$$W(a,b)=\frac{1}{\sqrt{a}}\int x(t)\Psi(\frac{t-b}{a})dt$$

a는 규모 매개변수(scale parameter), b는 위치 매개변수(location parameter), $x(t)$는 입력신호, Ψ는 모 웨이블릿함수(mother wavelet)이다.

최적의 인공신경망 모형 구축하는데 있어 훈련 데이터를 이용하여 최적의 파라미터를 도출하여야 하며 모형의 예측성능을 좌우하기 때문에 상당히 중요한 과정임에도 불구하고 이론적 기준이 없으며 반복실험과 연구자의 경험에 의존해야하는 제약이 있다[13]. 본 논문에서는 최적 파라미터를 찾기위해 교차검증(n-fold cross-validation)을 수행하여 검증단계에서 최소오차를 보이는 파라미터를 선정하였다. 인공신경망 모델의 경우 이러한 시계열 모델의 한계를 극복하여 모형화가 가능하며 다음 Fig. 4와 같이 구조화할 수 있다.

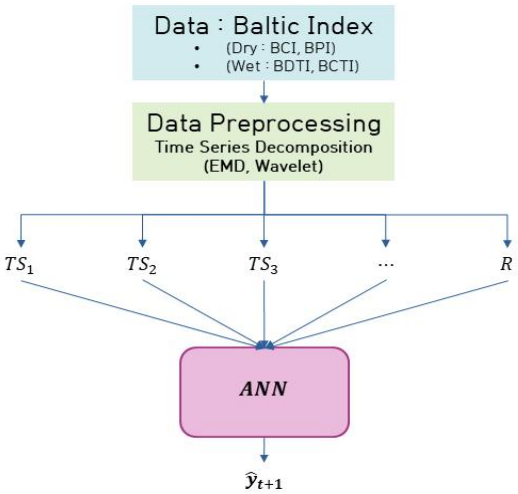

Fig. 5. Modelling Process

제시된 모형의 성능평가와 관련하여 RMSE(root mean squared error, $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}$)를 기준으로 사용하였다. 여러 예측기준 중에 특별히 우수한 기준을 선정하기 어려우나 RMSE의 경우 일반적인 연구에서 받아들여지고 있는 지표들이다.

III. Empirical Results

전술한 바와 같이 Fig. 5의 모델링 과정에서 BCI와 BDTI 원시계열 데이터를 웨이블릿분해와 경험적모드분해법으로 분해하여 데이터를 전처리하였으며 분해결과는 각각 Fig. 6과 Fig.7에서와 같다.

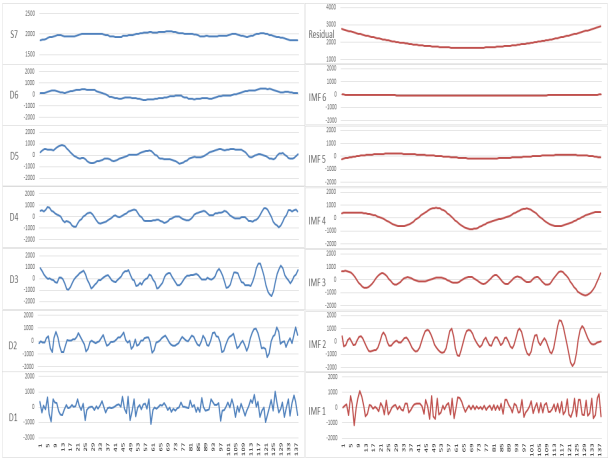

Fig. 6. Wavelet(L) and EMD(R) of BCI Decomposition

www.kci.go.kr

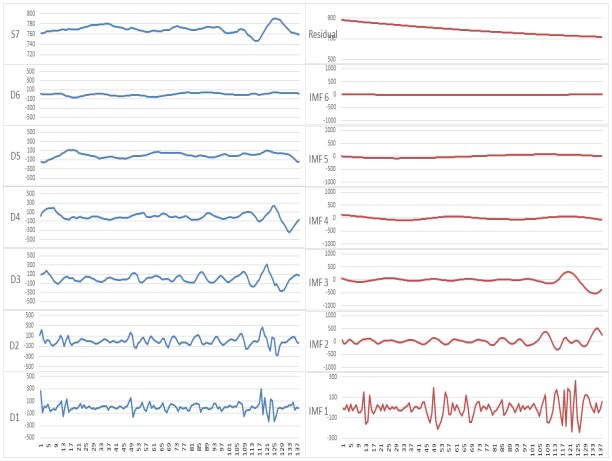

Fig. 7. Wavelet(L) and EMD(R) of BDTI Decomposition

분해된 시계열 데이터의 훈련데이터셋을 입력변수로 하여 인공신경망을 학습하였으며 최적의 파라미터를 선정하였다. 은닉층은 1개로 하였으며 은닉노드수와 weight decay값은 5차 교차검증방법으로 선정하였으며 Table 4과 같다.

Table 4. Parameters of ANN

Decomposition		BCI	BDTI
Wavelet	Hidden Node	11	11
	Weight Decay	10	10
EMD	Hidden Node	10	11
	Weight Decay	10	10

위 Table 4와 같이 시계열을 분해한 다음 인공신경망 학습을 통해 얻어진 파라미터를 선정하여 최적 예측모형을 구축하였으며 검증데이터로 모델의 예측성능을 평가하였다. 예측결과는 BCI의 경우 Fig 8과 같으며 RMSE를 기준으로 EMD로 분해한 인공신경망 모형이 1048.1이며 웨이블릿 분해는 985.8로 나타났다. 결과적으로 케이프운임데이터에는 웨이블릿 분해가 더 우수한 예측성능을 보임을 알 수 있다.

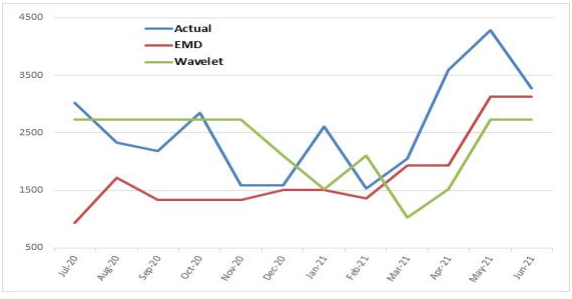

Fig. 8. Forecasting performance for BCI

BDTI의 경우 Fig 9과 같으며 RMSE를 기준으로 EMD로 분해한 인공신경망 모형이 108.9이며 Wavelet분해는 124.1로 나타났다. 결과적으로 탱커운임데이터에는 EMD 분해법이 더 우수한 예측성능을 보임을 알 수 있다.

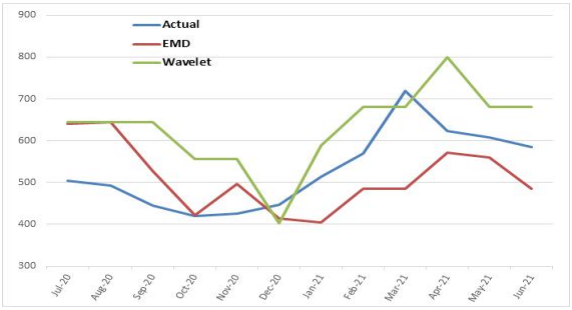

Fig. 9. Forecasting performance for BDTI

IV. Conclusions

본 연구는 해운시장에서 운임거래에서 절대적인 비중을 차지하고 있는 합리적인 스팟거래를 위해 과학적인 단기 예측방법을 제시하였다. 특히, 시계열모형과 기계학습모형을 비교한 연구가 주류를 이루는 기존의 연구와 비교하여 본 연구는 시계열의 데이터 전처리과정에서 EMD 분해와 웨이블릿 분해법을 적용하였으며 인공신경망 모형에 입력변수로 활용하여 예측력을 비교하였다는 측면에서 학문적인 성과가 있다. EMD와 웨이블릿 분해는 시간영역의 정보뿐만 아니라 주파수영역에서의 정보를 충분히 반영할 수 있다는 측면에서 활용한만한 가치가 있다. 또한 본 연구의 결과는 건화물시장에는 웨이블릿 분해가 적합하였으며 탱커시장의 경우 EMD분해가 더 적합함을 알 수 있었다. 따라서, 실무적으로 각 시장별로 가장 빈번하게 일어나는 스팟운임거래 의사결정에 단기예측에 과학적인 방법을 제시하였으며 이를 통해 합리적인 의사결정이 이뤄질 수 있으리라 판단된다.

마지막으로 벌크와 탱커 운임지수는 선형별로 구분되어 있다. 본 논문에서 분석한 데이터의 경우 해운시장의 대표적인 선형의 운임지수를 대상으로 하였다. 따라서 본 논문에서 제시한 운임지수예측 모형을 다른 선형의 지수예측에는 별도의 연구가 추가적으로 필요한 한계가 있을 수 있으며 향후 연구과제로 남겨둔다.

www.kci.go.kr

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5B5A07073372)

REFERENCES

[1] S. Lim, and H. Yun, “Forecasting Tanker Indices with ARIMA-SVM Hybrid Models,” Korean Journal of Financial Engineering, Vol.17, No.4, pp.79-98, 2018. DOI: 10.35527/kfedoi.2018.17.4.004

[2] S. Lim, and H. Yun, “An Analysis on Determinants of the Capesize Freight Rate and Forecasting Models,” Journal of Navigation and Port Research, Vol.42, No.6, pp.539-545, 2018. DOI: 10.5394/KINPR.2018.42.6.539

[3] Li, J. and M. G. Parsons , “Forecasting Tanker Freight Rate Using Neural Networks,” Maritime Policy and Management, Vol.24, No.1, pp.9-30. 1997. DOI: 10.1080/03088839700000053

[4] Lyridis, D., P. Zacharioudakis, P. Mitrou and A. Mylonas, “Forecasting Tanker Market Using Artificial Neural Networks,” Maritime Economics & Logistics, Vol.6, pp.93-108, 2004. DOI: 10.1057/palgrave.mel.9100097

[5] Yang, Z., L. Jin and M. Wang, “Forecasting Baltic Panamax Index with Support Vector Machine,” Journal of Transportation Systems Engineering and Information Technology, Vol.11, No.3, pp.50-57, 2011. DOI: 10.1016/S1570-6672(10)60122-5

[6] Zeng, Q. and C. Qu, “An Approach for Baltic Dry Index Analysis Based on Empirical Mode Decomposition,” Maritime Policy & Management, Vol.41, No.3, pp.224-240, 2014. DOI: 10.1080/03088839.2013.839512

[7] Zeng, Q., C. Qu, A. K. Y. Ng and X. Zhao, “A New Approach for Baltic Dry Index Forecasting Based on Empirical Mode Decomposition and Neural Networks,” Maritime Economics & Logistics, Vol.18, No.2, pp.192-210, 2016. DOI: 10.1057/mel.2015.2

[8] Guo, Z., W. Zhao, H. Lu, and J. Wang, “Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model,” Renewable Energy, Vol.37, No.1, 241-249, 2012. DOI: 10.1016/j.renene.2011.06.023

[9] Yu, L., S. Wang, and K. K. Lai., “Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm” Energy Economics, Vol.30, No.5, pp.2623-2635, 2008. DOI: 10.1016/j.eneco.2008.05.003

[10] Hasan, N., “A Methodological Approach for Predicting COVID-19 Epidemic Using EEMD-ANN Hybrid Model,” Internet of Things, Vol.11, pp.1-9, 2020. DOI: 10.1016/j.iot.2020.100228

[11] Clarkson Research, https://sin.clarksons.net/

[12] Ha, H., and K. Ha, “Forecast of the Stock Market Price using Artificial Neural Network and Wavelet Transform.” Vol.46, No.12, pp.1249-1261, 2019, DOI: 10.5626/JOK.2019.46.12.1249

[13] Fan, S., T. Ji, W. Gordon and B. Rickard, “Forecasting Baltic Dirty Tanker Index by Applying Wavelet Neural Networks,” Journal of Transportation Technologies, Vol. 3(January), pp.68-87, 2013. DOI: 10.4236/jtts.2013.31008

[14] Zhang, G., B. E. Patuwo and M. Y. Hu, “Forecasting with Artificial Neural Networks: The State of the Art,” International Journal of Forecasting, Vol.14, No.1, pp.35-62, 1998. DOI: 10.1016/S0169-2070(97)00044-7

[15] Flandrin, P., G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," IEEE Signal Process. Lett., Vol.11, No.2, pp.112-114, Feb. 2004. DOI: 10.1109/LSP.2003.821662

[16] Lee, H., J. Lee, and C. Yoo, “Selecting a mother wavelet for univariate wavelet analysis of time series data,” Journal of Korea Water Association, Vol.52, No.8, pp.575-587, 2019. DOI: 10.3741/JKWRA.2019.52.8.575

Authors

Sangseop Lim received the B.S. degree in ship engineering and M.A. and Ph.D. degrees in shipping management from Korea Maritime and Ocean University, Korea, in 2007, 2014 and 2018, respectively.

Since 2020, Dr. Lim is currently a Professor in the Division of Navigation Convergence Studies at Korea Maritime and Ocean University, Busan, Korea. He is interested in shipping finance, shipping market forecasting and market risk management.

Seokhun Kim received the M.S and Ph.D. degree in Computer Engineering from Hannam University in 2003 and 2006. He is an assistant professor Mobile Media at Suwon Women’s University in from 2012 to 2017.

Dr. Kim is currently an assistant professor in the Electronic Commerce at Paichai University. His teaching and research specialties are in the fields Mobile computing, Web-App programming, E-commerce System.

www.kci.go.kr