

딥러닝 기반의 건화물선 시황예측 연구

A Study on the Prediction of Dry Bulk Market
based on Deep Learning

2019. 11.

윤희성 | 최건우 | 황수진 | 박동원



한국해양수산개발원
KOREA MARITIME INSTITUTE

연구진	윤희성 한국해양수산개발원 해운해사연구본부 부연구위원 최건우 한국해양수산개발원 해운해사연구본부 전문연구원 황수진 한국해양수산개발원 해운해사연구본부 전문연구원 박동원 한국해양수산개발원 해운해사연구본부 연구원
-----	--

보고서 집필 내역	
연구책임자	윤희성 연구총괄, 제1장, 제5장

연구진	최건우 제2장 제1절~제2절, 제4장 제2절 일부 황수진 제2장 제3절, 제3장 제1절 일부 박동원 제3장 제1절 일부, 제2절, 제4장 제1절, 제2절 일부
-----	--

발간사

해운업을 영위함에 있어 선박투자, 용대선, 특정 사업영역에의 진출 입 등 다양한 의사결정의 문제에 직면하게 된다. 의사결정은 합리성을 갖추어야 하고 합리적인 의사결정은 데이터에 기반을 둔 과학적 분석을 통해 이루어질 수 있다.

이 연구는 해운의 의사결정에 핵심적인 역할을 하는 예측의 문제를 다루고 있다. 예측이 위에서 언급한 데이터와 과학적 분석이 가장 필요한 영역 중의 하나임에도 불구하고 실무에서 경험, 직관 또는 타 기관의 분석을 종합하는 수준의 예측에 의존하는 경우를 종종 보아 왔다. 물론 경험이나 직관을 통한 예측도 의미 있는 접근이지만 시장을 움직이는 기제에 복잡성이 더해지고 불규칙한 변동요인들이 작용하면서 예측에 대한 과학적인 접근의 중요성이 더욱 강조되고 있다.

과학적인 접근방법으로 최근에 각광을 받는 영역은 빅데이터와 인공지능을 활용한 분석이다. 특히 인공지능을 활용한 방법론은 다양한 예측의 영역에서 기존의 시계열 분석이 갖는 한계를 극복할 수 있는 단초를 제공함으로써 관심을 끌고 있다.

이 연구에서는 다층 퍼셉트론 구조를 갖는 딥러닝 모델을 적용하여 단기시황예측을 시도하였다. 특히 입력변수의 선정에 있어서 시계열적인 접근이 아닌 기술적 분석의 방법론을 채택한 것은 흥미로운 접근이

라고 할 수 있다. 또한 비교모델로 서포트벡터머신(SVM)과 랜덤포레스트를 사용하여 딥러닝모델과 예측성능을 비교하였다. 그 결과 딥러닝 모델이 가장 성과가 좋은 것으로 나타났다.

우리 원에서는 인공지능과 빅데이터의 활용을 위하여 수년 전부터 대학의 인공지능연구실과 업계가 공동으로 연구를 진행하여 왔다. 이 연구는 그러한 노력의 결실이며 연구결과는 무엇보다도 예측에 목말라 하는 산업계에 중요한 인사이트를 제시하는 측면에서 가치 있는 평가를 받을 것으로 기대된다. 또한 예측결과를 활용할 수 있는 정부, 조선소, 선박금융기관의 수요에도 부응할 것이다.

앞으로 연구의 범위를 확장하여 해운시장의 다른 영역과 중장기 예측, 조기경보 등에서도 성과를 내기를 기대한다.

2019년 11월
한국해양수산개발원
원장 장 영 태

목차

❖ 요약 · i

❖ Executive Summary · v

제 1 장 서론 · 1

제1절 연구 배경 및 필요성	1
제2절 연구 목적	2
제3절 연구 내용 및 방법	3
1. 연구 범위 및 주요 내용	3
2. 연구 방법	4
제4절 선행연구 검토	5

제 2 장 데이터 및 변수선정 · 9

제1절 데이터	9
제2절 기술적 분석	9
1. 기술적 지표	11
제3절 변수탐색 및 선정	27
1. 기술적 변수 선정	27
2. 기술적 통계 분석	29

제 3 장 딥러닝모델 분석 · 37

	제1절 딥러닝	37
	1. 다층퍼셉트론	38
	2. 과대적합(Overfitting)	44
	제2절 모델구조 및 학습방법	46
	1. 학습데이터	46
	2. 모델구조	46
	3. 학습방법	46

제 4 장	분석결과 · 51
제1절 딥러닝 모델 분석결과	51
1. 모델 평가지표	51
2. 모델 분석결과	55
제2절 모델성능비교	62
1. 비교모델	62
2. 모델성능비교 결과	64

제 5 장		결론 · 69	
제1절 연구결과 요약		69	
1. 딥러닝모델 분석 결과		69	
2. 비교모델 분석 결과		70	
제2절 연구의 기여		70	
제3절 추가연구		71	

❖	참고문헌 · 73
----------	------------------

표 목차

◆	
〈표 1-1〉 연구 범위	4
〈표 1-2〉 선행연구 요약	7
〈표 2-1〉 기술적 변수 요약	27
〈표 2-2〉 케이프 기술통계	29
〈표 2-3〉 파나막스 기술통계	30
〈표 2-4〉 수프라막스 기술통계	31
〈표 2-5〉 상관분석 결과	34
〈표 3-1〉 다층퍼셉트론 학습 5단계	43
〈표 4-1〉 혼동행렬 구성	51
〈표 4-2〉 케이프 모델 학습결과	55
〈표 4-3〉 파나막스 모델 학습결과	56
〈표 4-4〉 수프라막스 모델 학습결과	58
〈표 4-5〉 케이프 혼동행렬	59
〈표 4-6〉 파나막스 혼동행렬	59
〈표 4-7〉 수프라막스 혼동행렬	59
〈표 4-8〉 선형별 평가지표	60
〈표 4-9〉 케이프 모델평가	65
〈표 4-10〉 파나막스 분석결과	66
〈표 4-11〉 수프라막스 분석결과	67

그림 목차



〈그림 2-1〉 케이프 5TC, 장·단기 이동평균	12
〈그림 2-2〉 케이프 5TC, MACD, Signal	13
〈그림 2-3〉 케이프 5TC, MACD Oscillator	14
〈그림 2-4〉 케이프 5TC, 모멘텀과 변화율	15
〈그림 2-5〉 케이프 5TC, 상대강도지수	17
〈그림 2-6〉 케이프 5TC, 윌리엄스 %R	18
〈그림 2-7〉 케이프 5TC, 스토캐스틱	19
〈그림 2-8〉 케이프 5TC, 백분위	20
〈그림 2-9〉 케이프 5TC, 회귀계수	21
〈그림 2-10〉 케이프 5TC, Log Return	22
〈그림 2-11〉 케이프 5TC, 표준편차	23
〈그림 2-12〉 케이프 5TC, 장·단기 용선료	25
〈그림 2-13〉 케이프 5TC, FFA 당월물	26
〈그림 2-14〉 케이프 운임 변동성	33
〈그림 2-15〉 파나막스 운임 변동성	33
〈그림 2-16〉 수프라막스 운임 변동성	33
〈그림 3-1〉 다층퍼셉트론 구조	39
〈그림 3-2〉 활성화함수	42
〈그림 3-3〉 역전파 알고리즘 기본원리	44
〈그림 3-4〉 모델 과대적합, 과소적합	45
〈그림 3-5〉 학습 모델 구조	49
〈그림 4-1〉 ROC 곡선	54
〈그림 4-2〉 케이프 예측결과	61
〈그림 4-3〉 파나막스 예측결과	61

〈그림 4-4〉 수프라막스 예측결과	61
〈그림 4-5〉 서포트 벡터 머신	63
〈그림 4-6〉 랜덤 포레스트	64
〈그림 4-7〉 케이프 ROC 곡선	65
〈그림 4-8〉 파나막스 ROC 곡선	66
〈그림 4-9〉 수프라막스 ROC 곡선	67

요약

딥러닝 기반의 전화물선 시황예측 연구

1. 연구의 목적

- ▶ 해운업에 있어 예측이 의사결정에서 중요한 역할을 함에도 불구하고 이에 대한 과학적인 접근보다는 직관이나 경험에 의존하는 경향이 있음. 이에 인공지능 기반 시황예측 모델 개발을 통해 해운기업의 의사결정을 지원하고 나아가 정책당국의 정책적 판단 정확도를 높여 해운위기 재발을 막는 것이 본 연구의 목적임

2. 연구의 방법 및 특징

1) 연구방법

연구흐름	세부내용	연구방법
↓	↓	↓
서론	• 연구개요, 내용 및 연구 차별성	• 선행연구검토

연구흐름	세부내용	연구방법
↓	↓	↓
데이터 및 변수선정	<ul style="list-style-type: none"> • 데이터 설명 • 기술적 분석 기법 <ul style="list-style-type: none"> - 추세지표, 모멘텀지표, 기타지표 • 기술적 통계 분석 	<ul style="list-style-type: none"> • 문헌조사 • 통계분석
↓	↓	↓
딥러닝모델 분석	<ul style="list-style-type: none"> • 딥러닝개요 <ul style="list-style-type: none"> - 모델 구성요소 및 특징 • 모델구조 및 학습방법 	<ul style="list-style-type: none"> • 문헌조사
↓	↓	↓
분석결과	<ul style="list-style-type: none"> • 딥러닝 모델 분석결과 • 모델성능 비교 <ul style="list-style-type: none"> - 서포트 벡터 머신, 랜덤 포레스트 	<ul style="list-style-type: none"> • 통계분석
↓	↓	↓
결론	<ul style="list-style-type: none"> • 결론 	-

2) 연구의 특징

- ▶ 본 연구는 해운시황 예측을 위해 기존 수급구조를 바탕으로 한 계량 기법을 활용하지 않고 시장 추세와 변동성을 알 수 있는 기술적 지표를 입력변수로 사용하여 재무나 자원 등의 분야에서 기존 모델 보다 좋은 성능을 기록하고 있는 인공지능 기반 모델을 적용하는 연구임
- ▶ 자료의 가용성과 연속성을 고려하여 벌크선에 한정하여 케이프, 파나막스, 수프라막스 3가지 선형에 한정하여 주간 단위로 방향성 예측을 실시함
- ▶ 모델 성능 평가를 위해 다양한 평가 지표를 활용 하였으며 다양한 문제 해결을 위해 사용되는 머신러닝 모델들과 비교를 통해 모델의 객관적인 성능을 제시함

3. 연구 결과

1) 연구 결과 요약

- ▶ 기술적 지표들과 주간 시황 변동성 상관분석을 실시한 결과 대부분의 기술적 지표와 시황 변동성간에는 통계적으로 유의미한 상관관계가 있는 것으로 나타남
- ▶ 케이프, 파나막스, 수프라막스 선형별로 20회의 파라미터 추정과정을 반복하여 이들의 출력치를 결합한 앙상블 모델을 적용하였음
- ▶ 그 결과 케이프선에서는 정확도가 70.9%, 파나막스에서는 72.3%, 수프라막스에서는 79.7%로 나타남
- ▶ 모델 성능 비교를 위해 서포트 벡터 머신과 랜덤 포레스트를 사용한 결과 케이프 선에서는 각각 61.5%, 64.2%를 기록하였으며, 파나막스에서는 64.2%, 74.3%, 수프라막스는 66.2%, 75.7%를 기록함
- ▶ 세 선형을 종합적으로 살펴보면 딥러닝 모델의 성과가 가장 좋게 나타남
- ▶ 시장의 유동성이 작은 대형선의 경우 운임 변동폭이 크고 단기적인 공급의 쏠림현상이나 기상이변의 영향이 크기 때문에 대형선의 정확도가 상대적으로 낮은 것은 시장참여자의 직관과 일치하는 결과임

2) 연구의 기여

- ▶ 본 연구는 해운 시황의 단기방향성 예측에 대한 기술적 지표를 활용한 인공지능 모델의 유효성을 검증한 데 있음
 - 재무나 자원 등 다양한 영역에서 인공지능기반 모델이 기존 계량 분석보다 우수한 성과를 보이는 연구결과가 지속적으로 나오고 있

- 음. 특히, 중장기 예측에 어려움을 보이는 계량분석 모델을 인공지능 기반의 새로운 방법론 예측연구로 확대 가능한 기반을 마련함
- ▶ 해운기업이 용대선 의사결정에 해운 시황의 단기방향성 예측결과를 참고하여 의사결정의 질을 높이고 경영성과를 개선할 수 있을 것임

3) 추가연구

- ▶ 본 연구에서는 벌크선의 단기시황예측으로 범위를 한정하여 진행되었음. 이 연구를 3가지 방향으로 확장하는 것을 고려할 수 있음
- ▶ 첫째, 연구 분야를 탱커와 컨테이너선 까지 확장, 둘째, 모델에 활용 가능한 다양한 입력변수를 확보, 셋째, 예측연구 대상기간의 확장임

EXECUTIVE SUMMARY

A Study on Predictions of the Dry Bulk Market based on Deep Learning

1. Purpose

- ▶ Market forecasting plays an important role in decision-making within the shipping industry. Despite such significance, however, there is a tendency to rely on intuition and experience rather than utilizing a scientific approach. The primary purpose of this study is to develop an artificial intelligence-based market prediction model to support the decision-making processes of shipping companies. Furthermore, it aims to prevent the recurrence of shipping crisis by enhancing the accuracy of policy decisions carried out by policy authorities.

2. Methodology and Feature

1) Methodology

Work flow	Specific contents	Study methods
↓	↓	↓
Introduction	<ul style="list-style-type: none"> • Study overview and differentiation in terms of contents 	<ul style="list-style-type: none"> • Review of preceding studies
↓	↓	↓
Selecting Data and Variables	<ul style="list-style-type: none"> • Data explanation • Technical analysis methods <ul style="list-style-type: none"> – Trend indicators, momentum indicators, other indicators • Technical statistical analysis 	<ul style="list-style-type: none"> • Literature review • Statistical analysis
↓	↓	↓
Analysis of Deep learning model	<ul style="list-style-type: none"> • Overview of deep learning <ul style="list-style-type: none"> – Components and features of the model • Model structure and learning methods 	<ul style="list-style-type: none"> • Literature review
↓	↓	↓
Analysis result	<ul style="list-style-type: none"> • Analysis results of deep learning model • Comparison of model functions <ul style="list-style-type: none"> – Support Vector Machine, Random Forest 	<ul style="list-style-type: none"> • Statistics analysis
↓	↓	↓
Conclusion	<ul style="list-style-type: none"> • Conclusion 	–

2) Feature

- This study uses technical indicators that are able to understand market trends and volatility as input variables, instead of utilizing the existing demand-supply based quantitative method in predicting shipping market conditions. Therefore, it applies

an AI-based model, which has shown better performance in areas like finance and resources.

- ▶ Taking the availability and continuity of the data into consideration, the study has narrowed its scope to bulk carriers with three types of ships including Capesize, Panamax and Supramax, forecasting the market direction on a weekly basis.
- ▶ Various assessment indicators are utilized to evaluate the performance of the model. In addition, this study makes comparisons with machine learning models used for solving a variety of problems, presenting an objective performance of the model.

3. Results

1) Summary

- ▶ The result of the correlation analysis between technical indicators and the weekly volatility of market conditions shows that there is a significant correlation between most of the technical indicators and the volatility of market conditions.
- ▶ This study applies an ensemble model, which combines outputs of the processes repeated around 20 times of parameter estimation by type of ships including Capesize, Panamax and Supramax.
- ▶ The results showed that the accuracy of the model varied slightly

in the different types of ships with Capesize having an accuracy of 70.9%, Panamax having an accuracy of 72.3% while Supramax recorded an accuracy of 79.7%.

- ▶ This study utilized the Support Vector Machine and Random Forest to compare the performance of the models. The accuracy of these models in capsize were 61.5% and 64.2% respectively, with Panamax showing an accuracy of 64.2% and 74.3% and Supramax having an accuracy of 66.2% and 75.7%.
- ▶ A comprehensive analysis of the three types of vessels showed that the performance of deep learning model was the best.
- ▶ Large carriers, which has small market volatility have a high range of fluctuation in freight rates, while largely impacted by the herd behaviors of short term supply and weather changes. Therefore, a relative low accuracy of large carriers is in accord with the intuition of market participants.

2) Contributions of the study

- ▶ This study verifies the validity of artificial intelligence model tapping into technical indicators in predicting the short-term direction of the shipping market.
 - A number of studies have continued to conclude that artificial intelligence-based models have showed more outstanding performance than the existing quantitative analysis in various areas such as finance and resources. In particular, this study builds up a foundation on which quantitative analysis models

having difficulties in making mid-to-long term predictions can be expanded to new AI-based forecasting methodologies.

- By referring to shipping market forecasts in the decision-making process of chartering or hiring of vessels, shipping companies will be able to enhance the quality of their decision making, and improve business results.

3) Following Studies

- The present study has narrowed its scope to short-term predictions of market conditions of bulk carriers. This study can be expanded further in three possible directions.
- First, the area of the study can be expanded to tankers and containerships. Second, following studies can secure a variety of input variables applicable to models. Third, the period of the prediction study can be further extended.

제 1 장

서론

제1절 연구 배경 및 필요성

해운업에 있어 예측이 의사결정에 중요한 역할을 함에도 불구하고 이에 대한 접근이 과학적이지 못해 직관이나 경험에 의존하는 경향이 있었다. 특히 브로커 계열의 연구기관이나 독립연구기관이 잘 발달되지 않은 우리나라에서는 외국의 시황분석 보고서를 적당히 조합하여 전략의 수립에 활용하는 경우도 비일비재했다.

과학적인 접근이 활성화되지 않은 데에는 지금까지의 계량적 예측의 성과가 그다지 좋지 않았던 것도 영향을 미쳤다. 해운계의 저명한 학자인 Martin Stopford도 그의 저서 해운경제학에서 계량적 예측이 지금까지 변변치 않은 성과를 낳은 것과 계량분석이 한계를 가졌던 것을 적시하고 있다.¹⁾

이런 상황에서 1970년대에 등장한 이래 구현의 한계 때문에 담보

1) Martin Stopford(2009), 양창호 외 번역(2015), pp.841-849, pp.894-898

상태에 있던 기계학습 기반의 모델들이 컴퓨터의 연산능력의 획기적인 발달에 힘입어 다시 부각되기 시작하였다. 이들 모델들은 다양한 영역에서 활용되기 시작했으며 예측영역에서는 재무나 자원 등의 분야에서 이미 성과를 입증하기 시작하였다. 해운에 인공지능 기반의 예측이 적용된 것은 비교적 최근의 일로서 이를 활용한 연구에 대한 산업계의 요구가 점차 강해지고 있다. 일례로 근래 일본 3사가 유수의 분석기관에 의뢰하여 시황예측 모델의 개발을 추진하였으며 중국의 상해국제해운연구소(Shanghai International Shipping Institute)에서는 항만관련 예측과 분석에 인공지능 모델을 적용하는 연구를 활발히 진행하고 있다.

해운산업의 현황은 예측을 통한 의사결정능력의 제고를 어느 때보다도 필요로 하고 있다. 2008년 말에 찾아온 금융위기는 해운산업의 붕괴로 이어졌으며 국내 최대의 컨테이너 선사였던 한진해운이 불황을 견디지 못하고 2016년에 파산하였다. 우리 해운업계는 아직도 국제 해운시장에서 경쟁할 수 있는 체력을 회복하지 못하고 있으며 앞으로도 정책적으로 그리고 기업의 입장에서 많은 지원과 개선을 필요로 하고 있다.

의사결정의 질을 높이기 위한 이러한 과학적인 연구는 경영의 합리성을 높이고 성과를 개선하는 측면에서 큰 기여가 될 것이다.

제2절 연구 목적

이 연구는 두 단계의 목적을 갖는다. 첫 번째 목적은 시황예측 모델을 개발함으로써 해운기업의 의사결정을 지원하고 기업의 경영성

과를 제고하는 것이다. 시황예측은 해운기업의 투자의사결정과 용대선 의사결정의 기초가 되는 중요한 기능이다. 반복되는 예측오류는 원가곡선을 높이고 수익곡선을 낮추는 결과를 초래하여 기업의 지속가능성을 위협하게 된다.

2000년대 중반 이후 시황의 급등락을 경험하면서 시황이 높았던 시기의 잘못된 의사결정이 뒤이은 하락기에 파국적인 결과를 초래하는 것을 경험하였다.

궁극적인 목적은 정책당국의 입장에서 정책적 판단의 정확도를 높여 기업활동을 효과적으로 지원함으로써 해운위기의 재발을 막고 대응비용을 최소화하는 것이다. 정책은 사전적일 때 작은 비용으로 큰 성과를 낼 수 있어 예측의 성과가 크게 작용하는 영역이다. 이 연구는 단기예측이기 때문에 아직은 이런 측면의 효과가 크지 않을 수 있다. 하지만 단기예측에서 출발한 연구는 장기예측과 조기경보 등의 영역으로 이끄는 출발점으로서의 의미를 갖는다.

제3절 연구 내용 및 방법

1. 연구 범위 및 주요 내용

이 연구는 벌크선 분야에 한정하여 진행되었는데 벌크선을 대상으로 한 데는 몇 가지 이유가 있다. 우선 컨테이너는 수급의 변동이 비교적 정적이기 때문에 단기 시황예측의 중요성이 크지 않다. 또한 정기선의 특성 상 가격결정이 GRI(General Rate Increase)와 같이 시장 외적인 요인에 좌우되는 경우가 많다. 탱커는 국내선사의 노출 비중

이 크지 않고 벌크선을 대상으로 한 연구방법이 그대로 적용될 수 있어 우선순위가 높지 않다.

또한 인공지능 모델의 적용을 위해서는 자료의 가용성과 연속성이 확보되어야 하는데 벌크선이 이 조건을 가장 잘 갖추고 있는 것도 고려되었다. 벌크선 분석 시 시장에서 중요한 역할을 하는 케이프, 파나막스, 수프라막스의 3가지 선형을 모두 분석대상에 포함하였다.

예측의 범위는 주간시황의 예측으로 결정하였다. 장기로 갈수록 예측의 정확도가 하락하는 한계가 있어 우선 단기예측으로 성과를 낸 후 다음 단계로 장기예측이나 조기경보의 영역으로 이행하는 것이 타당한 것으로 판단된다. 단기예측과 장기예측은 입력변수의 선정에서부터 모델의 설정까지 다르게 접근해야 하고, 특히 입력변수의 선정에 있어서 시차(lag)에 대한 세밀한 사전분석이 필요하다.

〈표 1-1〉 연구 범위

	컨테이너	벌크선	탱커
단기			
장기			

2. 연구 방법

벌크선 시장을 대상으로 주간 단위의 시황을 예측하기 위하여 운임시리즈에서 기술적 분석(technical analysis)에 활용되는 입력변수들을 추출하였다. 기술적 분석은 시장움직임(market action) 즉, 가격과 거래량만으로 변화의 동인을 찾아내는 것이다. 주로 이동평균을 기반으로 한 다양한 지표들이 활용되는데 이 연구에서는 기술적

분석 지표에 추가하여 분산, 백분위 수 등 간단한 통계지표와 운임선도거래(FFA) 시장의 당월물 거래가격을 추가하였다.

연구방법으로는 은닉층(hidden layer)이 복수인 인공신경망 즉, 딥러닝 모델이 적용되었다. 비교모델로는 서포트벡터머신, 랜덤포레스트를 적용하여 딥러닝 모델과 성과를 비교하였다.

제4절 선행연구 검토

Lyridis(2004)²⁾는 원유수요, 선대규모, 원유 생산량, 원유 가격, 용선료, 신조선가, 중고선가, 병커유가, 해체가 등의 변수를 활용하여 인공신경망(ANN; Artificial Neural Networks) 모델을 활용해 VLCC 운임 예측모델을 연구하였다. 1개월, 3개월, 6개월, 9개월, 12개월 후의 시장에 대한 예측을 시도하여 예측기간이 장기화 될수록 예측이 불안정해지는 결과를 관찰했다.

모수원(2010)³⁾은 다변량 모형보다 단일변량모형이 더 우수한 예측 성과를 보임과 BDI를 설명하는 변수들이 BDI 자료보다 시차를 두고 발표되는 것을 근거로 단일변량분석기법인 자기회귀-이동(ARIMA)모형과 장기순환가정을 도출하는 HP(Hodrick-Prescott)필터 기법을 활용하여 분석을 진행했다.

이우식(2017)⁴⁾은 코스피 주가지수 방향성 예측문제에 기술적 분석

2) Lyridis 외(2004), "Forecasting Tanker Market Using Artificial Neural Networks", *Maritime Economics & Logistics*, Vol.6(2), pp.93-108

3) 모수원(2010), 「2010년 BDI의 예측 - ARIMA모형과 HP기법을 이용하여」, 『한국항만경제학회지』, Vol. 26(1), pp.222-233

4) 이우식(2017), 「딥러닝분석과 기술적분석지표를 이용한 한국 코스피주가지수 방향성 예측」, 『한국데이터정보과학지』, Vol. 28(2), pp.287-295

지표와 의사결정나무(decision tree), 서포트 벡터 머신(SVM; Support Vector Machine), 인공신경망 알고리즘을 결합한 모형들을 비교하였다. 본 연구에서는 상대강도지수, 모멘텀, 변화율, 지수이동평균 등의 기술적 지표를 활용하여 75%의 정확도를 보였다.

Shynkevich(2017)⁵⁾은 S&P 500 주가지수 방향성 예측을 위해 기술적 분석 지표와 머신러닝 모형을 활용하였다. 본 연구에서는 방향성 예측 기간으로 1일부터 30일까지 다양한 기간을 예측하였다. 이동평균, 지수가중이동평균, 변화율, 상대강도지수, 윌리엄스 %R, 스톡스틱 등의 기술적 변수를 활용하였다. 머신러닝 모델은 인공신경망, 서포트 벡터 머신, k-최근접 이웃(k-nearest neighbor)을 사용했다. 연구결과 70% 이상의 모델 정확도를 기록하여 기술적 변수 사용이 예측 모델 개발에 도움이 되는 것을 확인하였다.

임상섭(2018)⁶⁾은 벌크시장 수급구조의 복잡성으로 인해 시장 예측에 어려움이 있어 수급 변수 대신 기술 지표를 활용하여 문제를 해결하고자 하였다. 수프라막스에 한정하여 단기 방향성 예측을 위한 인공신경망 분석을 실시하였다. 기술적지표로는 이동평균비율, 상대강도지수(RSI; Relative Strength Index), 볼린저밴드(BB; Bollinger Band), 기울기(slope), 로그 수익률(log return), 수프라막스 케이프 운임 비율, 장·단기 용선료 비율, 신조선·중고선가 비율을 활용하였다. 분석결과 현물시장 방향성을 70% 이상 정확도로 분류하는 모델을 개발하여 단기 시황예측에 기술적 지표 활용을 통한 인공신경망 모델 예측의 유효성을 보였다.

5) Shynkevich 외(2017), "Forecasting price movements using technical indicators: Investigating the impact of varying input window length", *Neurocomputing*, Vol. 264, pp. 71-88

6) 임상섭 외(2018), 「Supramax Bulk Carrier Market Forecasting with Technical Indicators and Neural Networks」, 『한국항해항만학회지』, Vol. 42(5), pp.341-346

〈표 1-2〉 선행연구 요약

구 분		선행연구와의 차별성		
		연구목적	연구방법	주요 연구내용
주요 선행 연구	1	<ul style="list-style-type: none">과제명: Forecasting Tanker Market Using Artificial Neural Networks연구자(연도):Lyridis 외(2004)연구목적: 인공신경망을 활용한 탱커 스팟운임 예측	<ul style="list-style-type: none">인 공 신 경 망 (ANN)을 활용한 예측 분석수요, 공급, 운임 데이터 활용	<ul style="list-style-type: none">비정상 및 비선형 시계열 처리에 적합한 인공신경망을 활용하여 스팟운임 예측(1개월, 3개월, 6개월, 9개월, 12개월)예측개월마다 다양한 변수의 조합에 따른 모델설정으로 최적의 예측 모델 선택인공신경망 모델과 시계열 모델 예측력 비교
	2	<ul style="list-style-type: none">과제명: 2010년 BDI의 예측 - ARIMA모형과 HP기법을 이용하여연구자(연도):모수원(2010)연구목적: 시계열 모형을 활용한 2010년 BDI 예측	<ul style="list-style-type: none">ARIMA, HP Filter (HP) 모델을 통한 BDI 예측단 변량(BDI)만 사용한 시계열 모델 예측	<ul style="list-style-type: none">시계열모형 (ARIMA, IARIMA, HP)을 통한 2010년 월평균 BDI 예측모델별 예측성능 비교실제 2010년 BDI값의 변화를 잘 반영하지 못하는 것으로 나타남
	3	<ul style="list-style-type: none">과제명: 딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측연구자(연도):이우식(2017)연구목적: 기술적지표와 딥러닝분석기법을 활용한 주가 방향성예측	<ul style="list-style-type: none">의사결정나무, 서포트벡터머신, 인공신경망을 활용한 주가 예측 분석상대강도지수, 모멘텀, 변화율 지표 등의 기술적지표 활용	<ul style="list-style-type: none">주가의 추세와 강도를 나타내는 기술적 지표를 바탕으로 주가의 변화 예측 모델 학습예측 모형의 성능 분석 및 평가
	4	<ul style="list-style-type: none">과제명: Forecasting price movements using technical indicators: Investigating the impact of varying input window length연구자(연도):Shynkevich 외 (2017)연구목적: 기술적지표와 머신러닝기법을 활용한 주가 방향성 예측	<ul style="list-style-type: none">인공신경망, 서포트벡터머신, k-최근접 이웃을 활용한 주가 방향성 예측이동평균, 윌리엄스 %R, 스톡캐스틱 등 기술적 변수 활용	<ul style="list-style-type: none">기술적 지표와 머신러닝을 통해 1일~30일 후 주가 방향성 예측 모델 개발'상승', '하락' 두 가지 범주를 가지는 모델과 '상승', '보합', '하락' 세 가지 범주를 가지는 모델 개발

구 분		선행연구와의 차별성		
		연구목적	연구방법	주요 연구내용
	4	<ul style="list-style-type: none"> • 과제명: Supramax Bulk Carrier Market Forecasting with Technical Indicators and Neural Networks • 연구자(연도):임성섭 외(2018) • 연구목적:기술적지표와 인공신경망을 활용한 수프라막스 운임 방향성 예측 	<ul style="list-style-type: none"> • 인공신경망을 활용한 예측 모델 • 수급 변수를 활용하지 않고 기술적 변수 활용 	<ul style="list-style-type: none"> • 해운시장 수급구조의 복잡성으로 수급변수 대신 기술적 변수를 활용하여 해운시황을 예측 • 인공신경망 모델을 활용하여 70% 이상 정확도를 가지는 모델 개발
	본 연구	<ul style="list-style-type: none"> • 수급 변수를 활용하여 분석한 연구는 데이터 생성주기가 길어 딥러닝 분석에 적합하지 않다는 한계가 있음 • 본 연구에서는 수급 구조를 통한 운임 예측에서 벗어나 가격만을 활용하여 데이터 양을 확대하였으며 기술적 분석을 통해 유의미한 지표를 추출함 	<ul style="list-style-type: none"> • 기술적 분석을 통해 시황예측에 적합한 변수 도출 • 딥러닝 기반의 스팟운임 예측 모델 학습 • 다른 모델들과 예측성능 비교 	<ul style="list-style-type: none"> • 기술적 지표를 활용한 변수탐색 • 인공신경망 바탕의 모델 학습 및 분석 • 머신러닝 기법들과 예측성능 비교

제 2 장

데이터 및 변수선정

제1절 데이터

본 연구에서는 건화물선 선종별 차주 운임 방향성 예측을 위해 볼틱 해운 거래소(Baltic Exchange), 클락슨(Clarkson)에서 제공하는 스팟(spot)운임, 기간용선료, 운임선도거래(FFA; Forward Freight Agreement) 자료를 활용하였다. 학습에 사용된 데이터의 기간은 2009년 9월 10일부터 2019년 8월 5일까지이며, 스팟운임과 운임선도거래는 일 단위로 2,472개, 용선료는 주 단위로 518개의 자료를 활용하였다.

제2절 기술적 분석

기술적 분석은 증권시장에서 주로 사용되는 분석방법으로 시장의 가격 그 자체에만 관심을 갖고 가격을 움직이게 하는 원인에는 관심

을 두지 않는다. 기술적 분석은 과거부터 현재까지 가격흐름만을 분석하여 미래의 가격이 어떻게 움직일지를 예측하고자 한다. 기술적 분석에는 이론적인 근거는 거의 없으며, 분석의 유효성은 시장 경험을 통해서 대부분 검증된다. 기술적 분석은 다음 3가지 가정을 토대로 한다.

첫 번째, 시장움직임은 모든 것을 반영한다. 기술적 분석은 가격에 영향을 미칠 수 있는 모든 요인이 실제 시장가격에 반영된다고 믿는다. 따라서 가격 형성에 영향을 미치는 수요와 공급의 변화를 파악하지 않더라도 시장이 어떤 방향으로 움직일 것인지를 예측할 수 있다.⁷⁾

두 번째, 가격움직임은 추세를 이룬다. 추세란 개념은 기술적 접근에 있어서 필수적이다. 가격움직임이 추세를 이룬다는 것은 기존 추세를 역행하기 보다는 유지하려는 성질을 가지고 있음을 의미한다. 이는 어떤 추세는 전환하기까지 같은 방향으로 계속 움직이는 것을 뜻한다. 따라서 기술적 분석은 차트에서 기존 추세를 파악하고 추세가 전환되는 시기를 찾아내는 것을 목표로 한다.⁸⁾

세 번째, 역사는 스스로 반복된다. 기술적 분석 연구의 많은 부분은 인간심리 연구와 관련이 있다. 기술적 분석은 차트의 패턴을 분석하여 시장의 강세심리 약세심리를 파악한다. 이러한 패턴분석이 과거에 잘 맞았기 때문에 미래에도 계속 잘 맞을 것이라는 가정을 가지고 미래예측에 활용한다. 이는 기술적 분석이 변화하지 않으려는 경향이 있는 인간심리를 바탕으로 하고 있기 때문이다.⁹⁾

7) John J. Murphy(1999), 최용석 번역(2000), p.25

8) John J. Murphy(1999), 최용석 번역(2000), p.26

9) John J. Murphy(1999), 최용석 번역(2000), p.27

1. 기술적 지표

1) 추세지표

추세 지표는 현재 추세의 방향성을 알기 위해 사용하는 지표이다. 시장추세는 상승추세, 하락추세, 횡보추세로 크게 나누어 볼 수 있다. 대표적인 추세지표로 이동평균, MACD(Moving Average Convergence Divergence)가 있다.

(1) 이동평균

이동평균은 기술적 지표 가운데 가장 널리 사용되는 것 중 하나이다. 이동평균은 시장의 움직임을 따라가며 추세가 시작되거나 끝나는 지점의 신호를 확인하거나, 추세의 진행과정을 확인하는 데 사용된다.

이동평균을 계산하기 위해 사용되는 기간이 단기일수록 시장변화에 민감하게 반응하며, 기간이 장기일수록 시장변화에 둔감하게 반응한다. 민감한 이동평균, 즉 단기 이동평균을 사용할 경우 추세가 반전할 때 빠르게 신호를 나타낸다. 하지만 시장에서 일시적으로 발생하는 임의적 변화를 추세로 받아들일 수 있는 문제가 있다. 반대로 장기 이동평균은 지속적인 추세에 대한 신호를 잘 나타낸다. 하지만, 추세 반전이 발생했을 때 신호를 포착하는 데 비교적 오랜 시간이 필요하다. 따라서 한 가지 이동평균을 사용하기 보다는 장기, 단기 두 가지 이동평균을 사용하는 것이 시장의 변화를 포착하는 데 유용하다.

단기 이동평균이 장기 이동평균을 상회하고 있으면 시장이 상승추세에 있음을 의미하고 단기 이동평균이 장기 이동평균을 하회하고 있으면 시장은 하락추세에 있음을 의미한다.

별도로 단기 이동평균이 장기 이동평균을 상향 돌파 하는 현상을 골든 크로스(golden cross), 반대로 단기 이동평균이 장기 이동평균을 하향 돌파 하는 현상을 데드 크로스(dead cross)라고 한다. 골든 크로스는 시장이 상승세로 전환되는 시기에 나타나며, 데드크로스는 시장이 하락세로 전환되는 시기에 나타난다.

〈그림 2-1〉 케이프 5TC, 장 · 단기 이동평균



자료: Clarkson, 저자 작성

(2) MACD¹⁰⁾

제럴드 애플(Gerald Appel)에 의해 개발된 MACD(Moving Average Convergence and Divergence)는 기간이 다른 2개의 이동평균 사이의 관계에서 추세변화 신호를 찾으려는 지표이다. MACD 지표는 이동평균의 차이로 계산된 MACD선과 MACD선의 이동평균으로 구성되며 다음과 같이 표현된다.

10) John J. Murphy(1999), 최용석 번역(2000), pp.274-278을 참고하여 작성됨

$$MACD = short - long$$

$Signal = MACD$ 의 이동평균

$short$ = 단기이동평균, $long$ = 장기이동평균

MACD 지표는 이동평균과 마찬가지로 빠른 MACD선이 느린 신호선과 교차를 통해 신호를 나타낸다. MACD 선이 신호선을 상향교차하면 상승신호로 하향교차하면 하락신호로 해석된다. 일반적인 이동평균 방법과 다른 점은 MACD는 0선을 기준으로 오르내리는 점이다. 0선으로부터 거리가 멀어졌을 때 시장 전환신호가 나타난다. 특히, 0선으로부터 거리가 멀고 두 선이 교차하면 시장전환이 발생할 것을 강하게 나타낸다. 뿐만 아니라 0을 기준으로 위 아래로 교차하는 것을 통해서도 시장의 변화를 포착할 수 있다.

〈그림 2-2〉 케이프 5TC, MACD, Signal

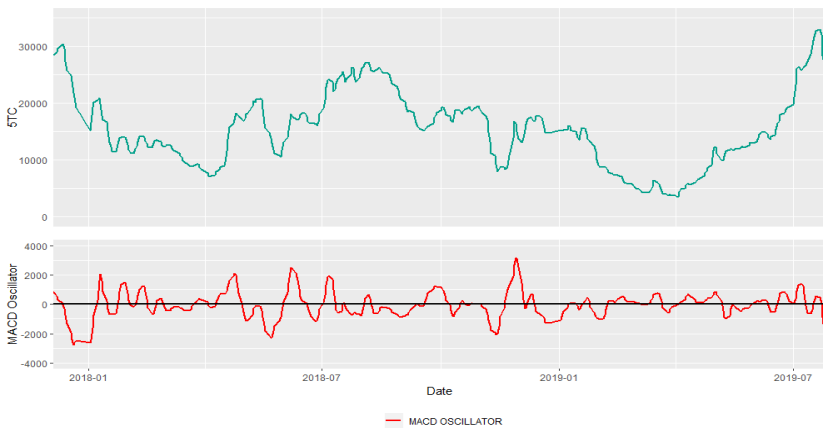


자료: Clarkson, 저자 작성

MACD에서는 보조지표로 MACD Oscillator를 활용하기도 한다. MACD Oscillator는 MACD선과 신호선의 차이를 나타낸다. MACD

Oscillator는 0을 기준으로 위아래로 움직인다. 이 선을 통해 MACD 선과 신호선의 간격이 넓어지거나 좁아지는 시점을 파악할 수 있다. MACD Oscillator 선이 양의 위치에서 0선을 향해 하락할 때 상승추세가 약해짐을 의미하고, 반대로 음의 위치에서 0선을 향해 상승할 때 하락추세가 약해짐을 의미한다. 이를 통해 0선과의 교차 이전에 시장 추세 변화를 미리 파악할 수 있어 조기 이탈신호를 파악하는 데 활용된다.

〈그림 2-3〉 케이프 5TC, MACD Oscillator



자료: Clarkson, 저자 작성

2) 모멘텀 지표

모멘텀 지표는 운동에너지를 이용하여 주가 추세의 변곡점을 찾는 데 사용되는 지표를 의미한다. 모멘텀 지표는 현재 가격과 과거 가격 차이 혹은 비율로 계산된다. 이를 통해 현재 추세가 가속되고 있는지 감속되고 있는지를 판단할 수 있다. 모멘텀(Momentum), 변화율(Rate of Change), 상대강도지수(RSI: Relative Strength Index), 윌리

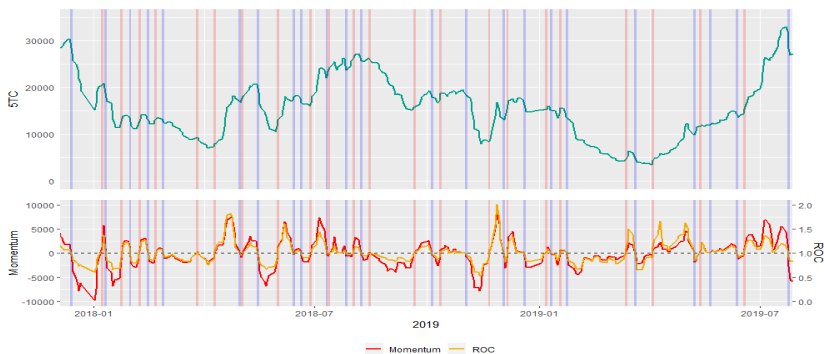
업스 %R(Williams %R), 스토캐스틱(Stochastic) 등이 대표적인 모멘텀 지표이다.

(1) 모멘텀과 변화율¹¹⁾

모멘텀(Momentum)은 가격변화 속도를 측정하는 지표이다. K일 모멘텀은 현재 시장가격과 K일 전 시장가격의 차이로 계산된다. 모멘텀 지표는 0을 기준으로 위에 있으면서 상승추세에 있음을 의미하고, 0을 기준으로 아래에 있으면 하락추세를 보임을 의미한다. 만약 상승추세에 있던 모멘텀이 0선을 향해 가까워지면 이는 상승속도가 하락함을 의미한다.

변화율(ROC; Rate of Change) 또한 모멘텀과 유사하게 가격변화 속도를 측정하는 지표이다. 변화율은 K일 변화율은 최근 시장가격을 K일 전 가격으로 나눈 비율로 측정된다. 이 경우에는 1을 기준으로 상승추세와 하락추세를 나누어 볼 수 있다.

〈그림 2-4〉 케이프 5TC, 모멘텀과 변화율



자료: Clarkson, 저자 작성

11) John J. Murphy(1999), 최용석 번역(2000), pp.250-256을 참고하여 작성됨

(2) 상대강도지수(RSI)¹²⁾

상대강도지수(RSI; Relative Strength Index)는 현재 추세에 강도를 백분율로 나타내어 추세 변경시점을 예측하는 데 유용한 지표로 활용된다. 상대강도지수는 일정기간동안 시장 변동 중에서 시장가격 상승의 비중으로 계산되어 지며 상승추세에서 얼마나 강한 상승세인지, 반대로 하락추세에서 얼마나 강한 하락세인지를 나타낸다.

상대강도지수를 수식으로 표현하면 다음과 같다.

$$RSI = \frac{U}{U+D} \times 100$$

U = n일간 시장가격 상승폭 합계

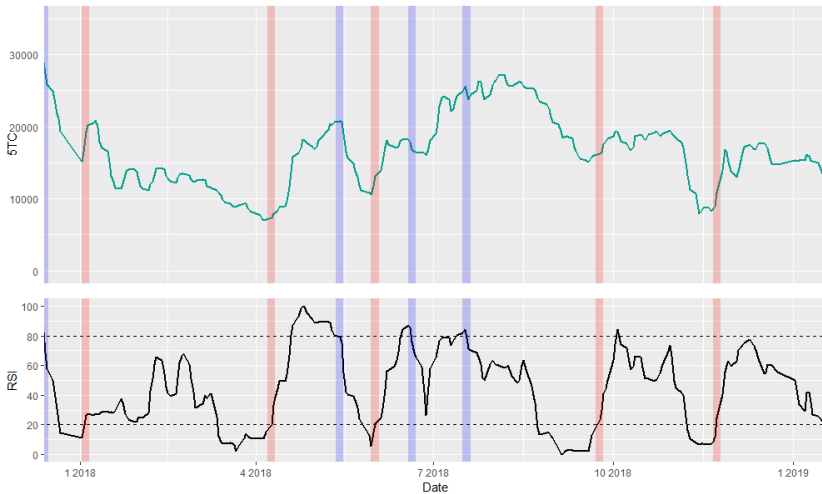
D = n일간 시장가격 하락폭 합계

상대강도지수를 개발한 와일더(Wilder)는 기본적으로 지표 산정기간으로 14일을 사용할 것을 권장했다. 기간이 짧을수록 지표는 변화에 더욱 민감해지고 강도가 강해지게 되며, 반대로 기간이 길수록 변화에 둔감하고 강도가 약하게 나타나게 된다.

상대강도지수는 0에서 100까지의 백분율로 그려지는데, 70~80일 때 과매수 상태, 20~30일 때 과매도 상태로 규정한다. 상대강도지수가 20선을 하향돌파한 후 다시 20선을 교차 상향하게 되면 추세가 상승 전환되는 신호를 나타낸다. 반대로 상대강도지수가 80선을 상향돌파한 후 다시 80선을 교차하여 하향하면 추세가 하락 전환되는 신호를 나타낸다.

12) John J. Murphy(1999), 최용석 번역(2000), pp.260-267을 참고하여 작성됨

〈그림 2-5〉 케이프 5TC, 상대강도지수



자료: Clarkson, 저자 작성

(3) 윌리엄스 %R¹³⁾

래리 윌리엄스(Larry Williams)의 %R은 주어진 기간 동안 가격변동과 현재 시장의 관계를 측정하는 지표이다. 윌리엄스 %R은 일정 기간 동안 고가와 현재 가격 차이를 가격변동으로 나눈 값에 마이너스를 취하여 계산된다. 수식으로는 다음과 같이 표현된다.

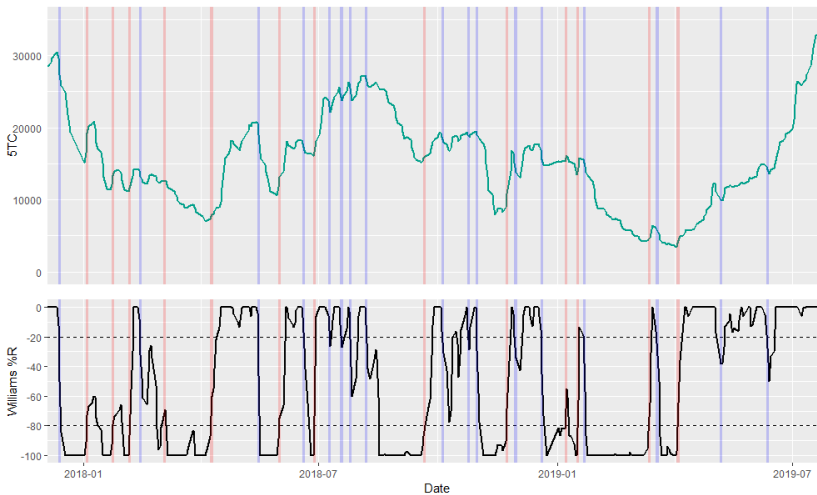
$$Williams \% R = - \frac{H_X - C}{H_X - L_X} \times 100$$

동 지표는 스토캐스틱 'Slow_%K'와 상당히 유사한 형태를 보인다. 일정기간의 고점을 기준으로 계산되기 때문에 시장이 고점에 가까우면 0의 값을, 시장이 저점에 가까우면 -100의 값을 가지게 된다.

13) John J. Murphy(1999), 최용석 번역(2000), pp.270-272을 참고하여 작성됨

윌리엄스 %R이 -80선을 상향돌파하면 상승시장으로 전환될 가능성이 있는 것으로 해석하며 -20선을 하향돌파하면 하락시장으로 전환될 가능성이 있는 것으로 판단할 수 있다.

〈그림 2-6〉 케이프 5TC, 윌리엄스 %R



자료: Clarkson, 저자 작성

(4) 스토캐스틱(Stochastic)¹⁴⁾

스토캐스틱은 시장 가격이 일정기간 동안의 가격변동 속에서 어느 정도 수준에 있는지를 백분율로 나타내는 지표로 사용된다. 스토캐스틱은 'Fast %K', 'Slow %K', 'Slow %D' 3가지로 구성되며, 'Slow %K'와 'Slow %D' 두 선을 통해 시장의 변화추세 신호를 포착한다.

우선 'Fast %K' 선은 다음과 같은 식으로 계산되어 진다.

14) John J. Murphy(1999), 최용석 번역(2000), pp.268-270을 참고하여 작성됨

$$Fast \%K = \frac{C - L_X}{H_X - L_X} \times 100$$

C = 현재가, H_X = 최근 X 일 중 최고가, L_X = 최근 X 일 중 최저가

계산된 'Fast %K' 선의 Y 일 이동평균을 통해 'Slow %K' 값이 계산되며, 다시 'Slow %K' 값을 Z 일 이동평균하게 되면 'Slow %D' 값을 얻게 된다. 이동평균 방법을 사용하기에 측정 기간이 짧을수록 단기 시장변화를 잘 포착하며, 측정 기간이 길수록 장기시장변화 포착에 유리하게 사용된다.

〈그림 2-7〉 케이프 5TC, 스토캐스틱



자료: Clarkson, 저자 작성

3) 기타지표

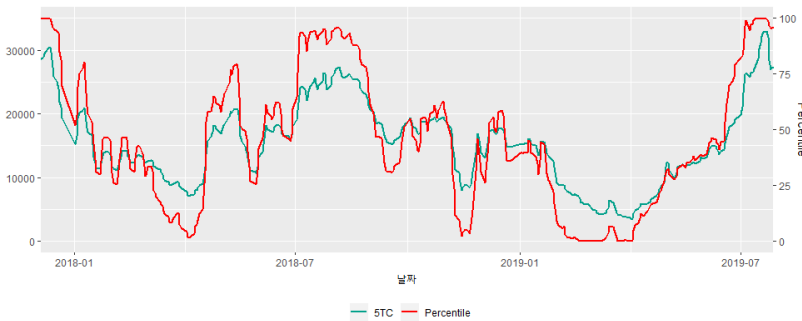
(1) 백분위(Percentile)

백분위는 집단에서 특정 관측치보다 낮은 관측치의 비중을 나타내

는 방법이다. 본 연구에서는 특정 날짜를 기준으로 1년 전까지 운임을 기준으로 백분위를 계산하였다.

운임 상승 혹은 하락세가 지속되면 백분위는 극값(0, 100)에 가까워지게 된다. 보통 극값은 단기간 유지된 후 시장이 반대로 움직인다. 따라서 백분위는 추세변동의 신호를 포착하는 데 활용할 수 있다. 하지만, 최고, 최저 운임이 지속적으로 갱신되는 경우에는 백분위를 통한 신호포착에 어려움이 있을 수 있다.

〈그림 2-8〉 케이프 5TC, 백분위



자료: Clarkson, 저자 작성

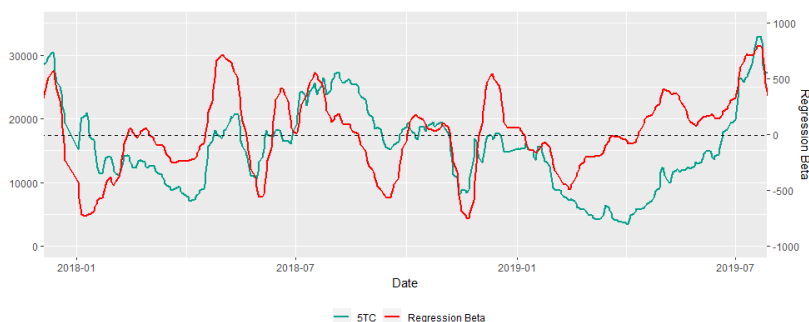
(2) 회귀계수(Regression Beta)

회귀계수는 회귀분석에서 종속변수를 설명하기 위해 독립변수에 곱해지는 계수(coefficient), 즉, 기울기를 말한다. 본 연구에서는 최근 20영업일 기준 운임 데이터를 활용하여 회귀계수를 산출하였다.

회귀계수가 양수인 경우, 최근 20영업일 동안 시장이 상승추세에 있었음을 나타내며 반대의 경우 시장이 하락추세에 있었음을 나타낸다. 그리고 회귀계수의 크기(절대값)는 추세의 강도를 나타낸다.

시장이 상승세에서 하락세로 전환하게 되면 회귀계수의 크기는 작아지게 되어 고점을 형성하며, 반대로 시장이 하락세에서 상승세로 전환할 때에는 저점을 형성한다. 형성된 변곡점에서 회귀계수 크기 변동 크기를 통해 시장 추세변화 속도를 포착할 수 있다.

〈그림 2-9〉 케이프 5TC, 회귀계수



자료: Clarkson, 저자 작성

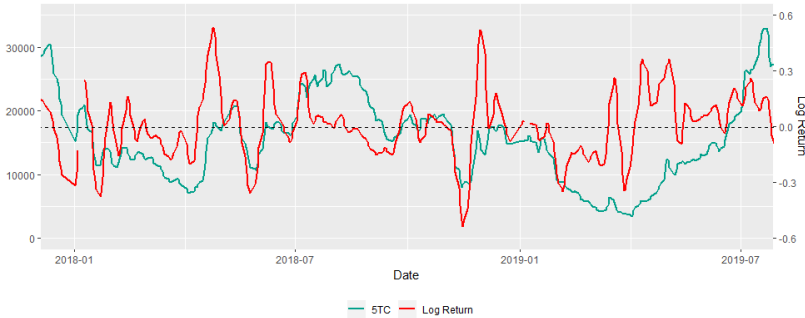
(3) 로그 수익률(Log-Return)

금융 분야에서는 수익률을 계산할 때 계산상의 편의를 위해 주로 로그 수익률을 사용한다. 로그 수익률은 시장가격 비율에 자연로그를 취해 계산되는데 이는 가격에 로그를 취하여 그 차이를 계산한 것과 동일하다.

$$\text{로그 수익률} = \ln\left(\frac{y_t}{y_{t-k}}\right) = \ln(y_t) - \ln(y_{t-k}), \quad y_t = t\text{시점 시장 가격}$$

본 연구에서는 최근 일주일 평균 운임과 전주 일주일 평균운임을 활용하여 로그 수익률을 계산하였다. 로그 수익률의 부호는 시장 추세 방향성을 의미하며, 크기는 추세의 강도를 의미하며 단기간의 변화를 나타내는 지표로 사용할 수 있다.

〈그림 2-10〉 케이프 5TC, Log Return



자료: Clarkson, 저자 작성

(4) 표준편차(Standard Deviation)

표준편차는 집단에서 자료값이 평균을 중심으로 얼마나 떨어져 있는지를 나타내는 수치값이다. 본 연구에서는 시장 변동성을 나타내기 위해 최근 20영업일을 기준으로 표준편차를 계산하였다.

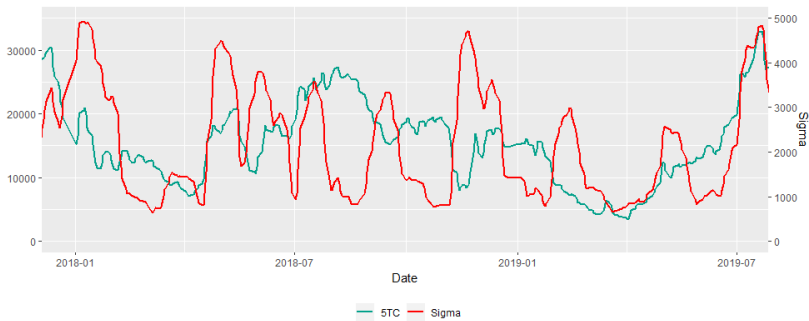
높은 표준편차는 시장이 큰 폭으로 변화될 때 나타나며, 낮은 표준편차는 시장의 변화가 크지 않은 횡보상태에 나타난다. 즉, 표준편차가 클수록 시장이 일정한 추세를 가지고 빠른 속도로 시장이 변하고 있음을 의미한다. 반대로 표준편차가 작으면 시장은 변화 속도가 느리거나 큰 변화가 없음을 의미한다.

대체적으로 운임은 고점을 형성한 후 높은 운임이 지속되기 보다는 하락세로 전환하는 경향을 많이 보인다. 반대로 운임이 저점을 형성한 후에는 상승세로 바로 전환하기 보다는 낮은 운임이 일정 기간 지속되는 경우가 많다. 따라서 일반적으로 표준편차는 시황 고조기에서 크고 시황 저조기에는 작게 나타난다.

표준편차가 고점에서 반전하락을 기록하거나, 저점에서 반전상승

할 때 시장추세가 반전되는 국면으로 진입함을 예측할 수 있다. 그리고 표준편차선이 고점에서 하락하는 기울기를 나타내면 시장이 안정세를 보이는 것으로 판단할 수 있다.

〈그림 2-11〉 케이프 5TC, 표준편차



자료: Clarkson, 저자 작성

(5) 장·단기 용선료 비율

해운 운임시장은 항해용선(voyage charter)과 기간용선(time charter) 2가지 형태로 나뉘어진다. 항해용선 계약은 화주의 화물을 운송하고 화주로부터 협정된 톤당 운임을 지불 받는다. 기간용선은 선박을 필요로 하는 용선주에게 일정기간 선박을 대선하고 일정 용선료를 수취한다.¹⁵⁾

스팟(spot)계약은 즉시 거래가 가능한 수송계약으로 항해용선계약(voyage charter), 1항차 기간용선(trip time charter)이 대표적인 계약형태이다. 스팟계약은 단기간 거래형태로서 짧은 기간에 걸쳐 화물을 수송 혹은 용선하므로 단기 수급구조에 영향을 받아 가장 민감하게 반응한다. 시장 상승 모멘텀이 있을 때 스팟운임이 가장 먼저

15) Martin Stopford(2009), 양창호 외 번역(2015), p.218

상승하게 되며 하락 모멘텀이 있을 때 가장 먼저 하락하게 된다.

기간용선은 한번 계약이 성약되면 선택이 용선되는 기간 동안 용선료가 고정된다. 용선료는 현재 시황에도 일부 영향을 받지만 선택이 용선되는 기간, 즉 미래에 대한 전망이 큰 영향을 미친다. 미래 시황에 대한 전망이 긍정적일 경우 용선료는 상승하며 부정적일 경우 용선료는 하락하게 된다. 따라서 선택이 용선되는 기간이 길어질수록 현재 시장변화에는 상대적으로 둔감하게 반응한다.

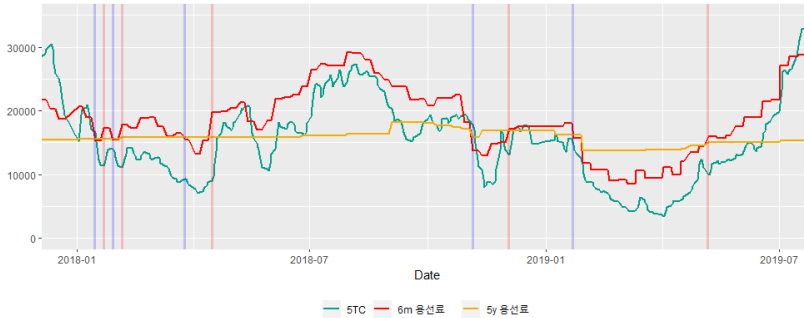
이와 같은 스팟운임과 용선운임의 특성을 활용하면 현재 시장 방향성을 확인할 수 있다. 시황이 상승기일 때는 스팟 운임이 가장 먼저 영향을 받으며 단기 용선료, 장기 용선료 순서로 운임이 상승하게 된다. 반대로 시장 하락기에는 스팟운임이 가장 먼저 하락하고 이후 단기 용선료, 장기 용선료가 하락한다. 즉 상승시장에서는 스팟운임, 단기 용선료, 장기 용선료 순으로 운임이 높으며 하락시장에서는 장기 용선료, 단기 용선료, 스팟운임 순으로 운임이 높게 나타난다.

본 연구에서는 장·단기 용선료 비율을 측정하여 현재 시황을 나타내는 변수로 활용하였다. 용선료 비율은 다음과 같이 계산되었다.

$$\text{용선료 비율} = \frac{\text{단기용선료}}{\text{장기용선료}} - 1$$

단기용선료가 장기용선료 보다 클 경우 용선료 비율은 양의 값을 가지게 되며 시장이 상승기임을 나타낸다. 반대로 장기용선료가 단기용선료보다 크면 용선료 비율은 음의 값을 가지게 되며 시장은 하락기임을 알 수 있다. 용선료 비율이 음수에서 양수 혹은 양수에서 음수로 전환되는 것을 통해 시장 방향성이 바뀌는 신호를 포착할 수 있다.

〈그림 2-12〉 케이프 5TC, 장·단기 용선료



자료: Clarkson, 저자 작성

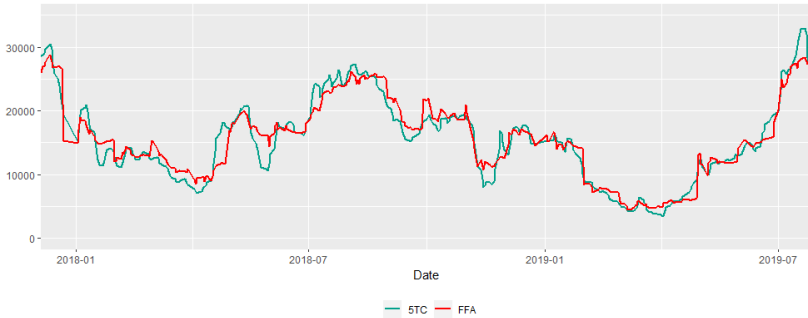
(6) 운임선도거래(FFA)

운임선도거래는 해상운임 파생거래의 일종으로 기초자산은 볼틱 해운거래소에 등록된 주요 독립 해운중개 기업으로 구성된 평가단으로부터 취합된 주요 노선별 운임으로 산정된다.¹⁶⁾ 운임선도거래는 미래 운임에 대한 시장참여자들의 합리적 예측을 바탕으로 산정되기에 미래 운임에 대한 불편추정치(unbiased estimator)로 생각할 수 있다.

운임선도거래 당월물 가격은 해당 월 운임평균에 대한 시장참여자들의 추정치로 계산된다. 월이 시작될 때는 당월 운임에 대한 정보가 없기에 완전한 추정치가 되지만 기간이 지날수록 이전 운임을 고려하여 당월물 가격이 형성되게 된다. 따라서 당월물 가격에서 이전 운임이 차지하는 부분을 제거하면 이 달 남은 기간에 대한 시장참여자들의 예상가격을 알 수 있다.

16) 정봉민(2015), p.128

〈그림 2-13〉 케이프 5TC, FFA 당월물



자료: Clarkson, Baltic, 저자 작성

이달 남은기간에 대한 시장참여자들의 추정 운임이 현재 운임보다 높다면 시장이 상승할 것으로 예상할 수 있으며, 반대로 추정 운임이 현재 운임보다 낮다면 시장이 하락할 것으로 예상할 수 있다. 추정 운임과 현재 운임의 차이가 클수록 시장의 변화를 더욱 강하게 예상하는 것으로도 해석할 수 있다.

하지만 이러한 효과는 월이 바뀔 때는 실제 해당 월의 운임정보가 없고 예측해야할 기간이 길어 변동성이 커지기 때문에 부정확할 수 있다. 특히, 당월물과 익월물 가격차이가 큰 경우에는 월 바뀔때로 인해 추정 운임과 현재 운임의 차이가 크게 변동할 수 있다. 이를 해결하기 위해 월 바뀔때 발생할 때는 당월물을 통해 계산된 추정운임과 익월물의 가중평균을 이용하여 사용했다.

제3절 변수탐색 및 선정

1. 기술적 변수 선정

본 연구에서 선정된 기술적 변수들의 파라미터는 <표 2-1>에 제시되었다. 주간 방향성 예측을 위해 추세 및 모멘텀 확인에는 단기로는 5일, 장기로는 20일을 파라미터로 사용하였다. 상대강도지수, 스톡캐스틱, 윌리엄스 %R은 일반적으로 주식시장에서 널리 사용되는 파라미터를 적용하였다. 백분위는 최근 1년간 운임을 비교하기 위해 1년 영업일인 250일을 사용하였고, 단기간 시황의 움직임과 변동성을 파악하기 위해 회귀계수와 표준편차 계산에는 20일을 사용했다. 마지막으로 초단기의 시황변화를 알기 위해 5일 로그 수익률을 사용했다.

<표 2-1> 기술적 변수 요약

변수	수식	파라미터
이동평균비율	$\frac{\frac{1}{s} \sum_{j=0}^{s-1} TC_{t-j}}{\frac{1}{l} \sum_{i=0}^{l-1} TC_{t-i}}$	$(s, l) = (5, 20)$
MACD	$\frac{1}{s} \sum_{j=0}^{s-1} TC_{t-j} - \frac{1}{l} \sum_{i=0}^{l-1} TC_{t-i}$	$(s, l, m) = (5, 20, 5)$
Signal	$\frac{1}{m} \sum_{i=0}^{m-1} MACD_{t-i}$	
오실레이터	$MACD - Signal$	
모멘텀	$TC_t - TC_{t-k}$	$k = 5$
변화율	$\frac{TC_t}{TC_{t-k}}$	

변수	수식	파라미터
상대강도지수	$\frac{U}{U+D}$ $U = \sum_{i=0}^{k-1} \max(TC_{t-i} - TC_{t-i-1}, 0)$ $D = \sum_{i=0}^{k-1} \max(TC_{t-i-1} - TC_{t-i}, 0)$	$k = 14$
Slow %K	$H_t(k) = \max(TC_{t-k+1}, TC_{t-k+2}, \dots, TC_t)$ $L_t(k) = \min(TC_{t-k+1}, TC_{t-k+2}, \dots, TC_t)$ $Fast\%K_t = \frac{TC_t - L_t(k)}{H_t(k) - L_t(k)}$	$(k, m, n) = (10, 6, 6)$
Slow %D	$Slow\%K_t = \frac{1}{m} \sum_{i=0}^{m-1} Fast\%K_{t-i}$ $Slow\%D = \frac{1}{n} \sum_{i=0}^{n-1} Slow\%K_{t-i}$	
윌리엄스 %R	$-\frac{H_t(k) - TC_t}{H_t(k) - L_t(k)}$	$k = 14$
백분위	$\frac{1}{k} \sum_{i=0}^{k-1} I(TC_t \geq TC_{t-k})$	$k = 250$
회귀계수	$\frac{S_{xy_t}(k)}{S_{xx(k)}}$ $S_{xy_t}(k) = \sum_{i=0}^{k-1} (x_i - \bar{x})(TC_{t-i} - \overline{TC_t}(k))$ $x_i = i + 1$ $\overline{TC_t}(k) = \frac{1}{k} \sum_{i=0}^{k-1} TC_{t-i}$	$k = 20$
로그 수익률	$\log(T_t) - \log(T_{t-k})$	$k = 5$
표준편차	$\frac{1}{k} \sqrt{\sum_{i=0}^{k-1} (TC_{t-i} - \overline{TC_t}(k))^2}$	$k = 20$

자료: John J. Murphy(1999), 최용석 번역(2000) 참고, 저자 재작성

2. 기술적 통계 분석

본 연구에서는 운임, 용선료, FFA 가격은 천 달러를 단위로 사용하였다. 각 선형별 독립변수들의 기초통계량은 <표 2-2>, <표 2-3>, <표 2-4>와 같이 나타났다.

<표 2-2> 케이프 기술통계

변수	최소	1분위수	중앙값	평균	3분위수	최대값
5TC	1.985	7.060	11.425	13.659	17.674	47.498
6개월 용선료	3.750	10.750	15.375	16.250	20.000	41.000
5년 용선료	8.600	13.550	15.500	16.097	18.750	28.000
FFA 당월물	1.748	7.455	11.826	13.717	17.072	43.441
FFA 익월물	2.382	8.410	12.832	13.986	17.616	40.805
장·단기 용선료 비율	0.436	0.781	0.962	0.982	1.143	1.870
FFA-스팟 비율	0.282	1.089	1.225	1.293	1.440	3.127
장·단기 이동평균 비율	0.443	0.877	0.996	1.008	1.123	1.646
MACD	-14.610	-1.340	-0.039	0.000	1.503	11.112
Signal	-13.332	-1.283	-0.003	-0.001	1.471	10.801
오실레이터	-4.726	-0.415	0.027	0.001	0.453	3.433
모멘텀	-17.723	-1.343	-0.073	-0.004	1.408	11.820
변화율	0.453	0.885	0.991	1.021	1.117	2.018
상대강도지수	0.000	0.215	0.489	0.482	0.737	1.000
윌리엄스 %R	-1.000	-1.000	-0.557	-0.526	-0.050	0.000
Slow %K	0.000	0.071	0.457	0.472	0.860	1.000
Slow %D	0.000	0.114	0.461	0.473	0.821	1.000
백분위	0.004	0.232	0.532	0.522	0.796	1.000
회귀계수	-1.630	-0.163	-0.002	0.000	0.197	1.506
로그 수익률	-0.683	-0.104	-0.004	0.001	0.096	0.595
표준편차	0.095	0.915	1.539	2.075	2.784	10.321

변수	최소	1분위수	중앙값	평균	3분위수	최대값
2일 이동평균	2.000	7.063	11.419	13.661	17.634	47.423
3일 이동평균	2.010	7.079	11.415	13.662	17.597	47.308
4일 이동평균	2.012	7.089	11.432	13.663	17.603	47.161
5일 이동평균	2.010	7.086	11.460	13.663	17.574	47.041
6일 이동평균	2.014	7.068	11.425	13.663	17.577	46.981
7일 이동평균	2.024	7.092	11.429	13.663	17.591	46.945
8일 이동평균	2.033	7.063	11.398	13.663	17.589	46.906
9일 이동평균	2.046	7.062	11.389	13.663	17.546	46.843
10일 이동평균	2.063	7.077	11.453	13.663	17.524	46.739

〈표 2-3〉 파나막스 기술통계

변수	최소	1분위수	중앙값	평균	3분위수	최대값
4TC	2.260	6.360	8.738	9.304	11.994	27.329
6개월 용선료	4.625	9.000	11.000	11.728	13.750	31.500
5년 용선료	7.550	9.850	11.750	11.606	13.225	19.375
FFA 당월물	2.515	6.415	8.930	9.412	11.840	26.939
FFA 익월물	2.595	6.961	9.375	9.800	12.209	27.886
장·단기 용선료 비율	0.597	0.821	1.007	0.988	1.120	1.651
FFA-스팟 비율	0.436	1.077	1.205	1.267	1.387	2.757
장·단기 이동평균 비율	0.647	0.927	0.999	1.002	1.072	1.458
MACD	-3.555	-0.598	-0.007	-0.018	0.624	3.624
Signal	-3.425	-0.583	-0.009	-0.019	0.599	3.559
오실레이터	-1.323	-0.178	0.009	0.002	0.179	1.238
모멘텀	-3.787	-0.603	-0.039	-0.016	0.551	3.480
변화율	0.677	0.932	0.995	1.005	1.068	1.601
상대강도지수	0.000	0.141	0.504	0.491	0.828	1.000
윌리엄스 %R	-1.000	-1.000	-0.543	-0.510	0.000	0.000
Slow %K	0.000	0.021	0.477	0.487	0.953	1.000
Slow %D	0.000	0.071	0.477	0.486	0.894	1.000
백분위	0.004	0.256	0.584	0.545	0.836	1.000

변수	최소	1분위수	중앙값	평균	3분위수	최대값
회귀계수	-0.484	-0.079	0.000	-0.003	0.079	0.479
로그 수익률	-0.361	-0.060	-0.004	0.001	0.064	0.417
표준편차	0.081	0.393	0.644	0.759	0.963	2.900
2일 이동평균	2.260	6.364	8.732	9.306	11.976	27.321
3일 이동평균	2.265	6.373	8.752	9.309	11.978	27.232
4일 이동평균	2.274	6.365	8.740	9.311	11.965	27.181
5일 이동평균	2.282	6.359	8.747	9.312	11.946	27.066
6일 이동평균	2.293	6.367	8.756	9.314	11.956	26.970
7일 이동평균	2.302	6.377	8.757	9.315	11.920	26.820
8일 이동평균	2.313	6.397	8.767	9.316	11.932	26.636
9일 이동평균	2.321	6.392	8.775	9.317	11.911	26.465
10일 이동평균	2.331	6.406	8.766	9.317	11.913	26.257

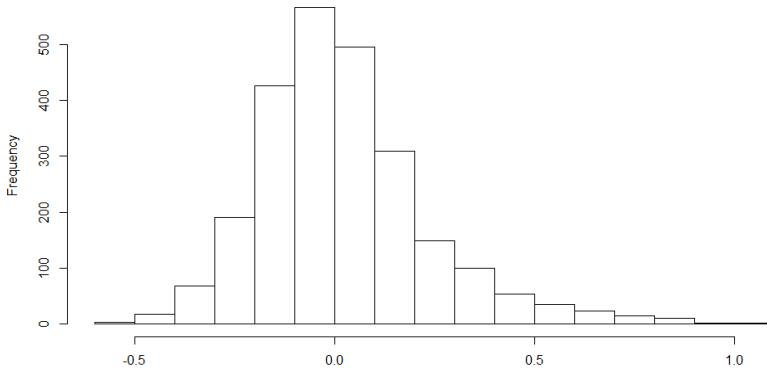
〈표 2-4〉 수프라맥스 기술통계

변수	최소	1분위수	중앙값	평균	3분위수	최대값
6TC	2.544	7.554	9.507	9.878	11.464	21.395
6개월 용선료	4.250	9.500	11.500	11.739	13.750	26.500
5년 용선료	7.250	9.594	11.000	11.152	12.500	17.500
FFA 당월물	2.620	7.740	9.548	9.910	11.423	21.894
FFA 익월물	3.015	8.050	9.780	10.123	11.790	23.991
장·단기 용선료 비율	0.548	0.967	1.045	1.042	1.125	1.514
FFA-스팟 비율	0.814	1.048	1.167	1.246	1.360	2.441
장·단기 이동평균 비율	0.703	0.960	1.000	0.999	1.039	1.294
MACD	-2.309	-0.366	0.002	-0.031	0.386	1.755
Signal	-2.230	-0.370	0.008	-0.033	0.371	1.717
오실레이터	-0.704	-0.088	0.001	0.003	0.084	0.805
모멘텀	-2.546	-0.295	0.010	-0.014	0.318	1.824
변화율	0.782	0.968	1.001	1.001	1.032	1.384
상대강도지수	0.000	0.088	0.508	0.506	0.925	1.000
Slow %K	0.000	0.004	0.552	0.512	0.996	1.000

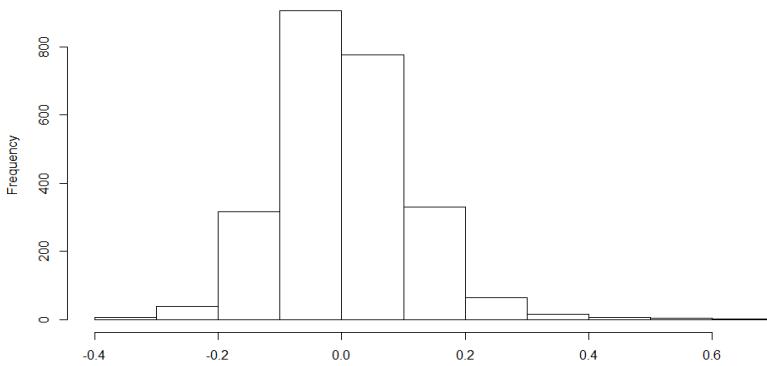
변수	최소	1분위수	중앙값	평균	3분위수	최대값
Slow %D	0.000	0.037	0.538	0.510	0.969	1.000
월리엄스 %R	-1.000	-1.000	-0.466	-0.493	0.000	0.000
백분위	0.004	0.228	0.540	0.536	0.844	1.000
회귀계수	-0.302	-0.050	0.001	-0.004	0.049	0.234
로그 수익률	-0.235	-0.032	0.002	-0.001	0.030	0.297
표준편차	0.038	0.202	0.369	0.453	0.611	1.790
2일 이동평균	2.549	7.546	9.511	9.880	11.463	21.379
3일 이동평균	2.559	7.554	9.513	9.882	11.462	21.368
4일 이동평균	2.565	7.544	9.515	9.883	11.471	21.360
5일 이동평균	2.575	7.551	9.518	9.884	11.466	21.346
6일 이동평균	2.586	7.541	9.515	9.886	11.481	21.325
7일 이동평균	2.596	7.546	9.520	9.887	11.469	21.300
8일 이동평균	2.611	7.550	9.519	9.889	11.479	21.267
9일 이동평균	2.625	7.565	9.524	9.891	11.478	21.239
10일 이동평균	2.645	7.570	9.522	9.893	11.473	21.310

동 연구에서 종속변수로는 차주 운임의 ‘상승’, ‘하락’을 이용했다. 선종별로 차주 운임 변동성을 살펴보면 <그림 2-14>, <그림 2-15>, <그림 2-16>과 같이 나타난다. 케이프, 파나막스, 수프라막스의 1주일 하락 변동 최대 크기는 각기 -54.7%, -32.3%, -21.8%로 나타났으며 상승 변동 최대 크기는 101.8%, 60.1%, 38.4%로 나타났다. 평균 변동 크기는 15.7%로 케이프선이 가장 컸으며 파나막스가 8.4%, 수프라막스가 4.5%의 변동성을 가지는 것으로 나타났다.

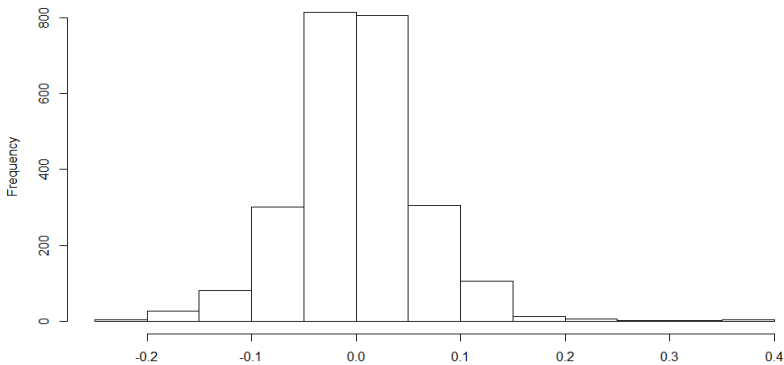
〈그림 2-14〉 케이프 운임 변동성



〈그림 2-15〉 파나마맥스 운임 변동성



〈그림 2-16〉 수프라맥스 운임 변동성



다음으로 선형별로 차주 운임 변동성과 변수들의 상관성 분석을 실시한 결과는 아래 <표 2-5>와 같이 나타났다. 대부분의 기술변수들은 선형과 상관없이 차주 운임 변동성과 큰 상관성을 가지는 것으로 나타났다. 특히, 장·단기 이동평균 비율, 오실레이터, 모멘텀, 변화율, 상대강도지수, 윌리엄스 %R, Slow %K는 모든 선형에서 차주 운임 변동성과 60% 이상의 상관성이 있는 것으로 나타났다.

<표 2-5> 상관분석 결과

변수	케이프		파나막스		수프라막스	
	R^2	$p-value$	R^2	$p-value$	R^2	$p-value$
6TC	-.152***	.000	-.102***	.000	-.055*	.010
6개월 용선료	-.087***	.000	.001	.955	-0.006	.770
5년 용선료	.030	.160	.078***	.000	0.054*	.011
FFA 당월물	-.041	.056	-.022	.293	-0.006	.776
FFA 익월물	-.156***	.000	-.115***	.000	-0.113***	.000
장·단기 용선료 비율	-.165***	.000	-.081***	.000	-0.092***	.000
FFA-스팟 비율	.144***	.000	.008	.709	-0.11***	.000
장·단기 이동평균 비율	-.616***	.000	-.623***	.000	-0.741***	.000
MACD	-.587***	.000	-.585***	.000	-0.705***	.000
Signal	-.396***	.000	-.364***	.000	-.545***	.000
오실레이터	-.664***	.000	-.735***	.000	-.651***	.000
모멘텀	-.884***	.000	-.915***	.000	-.935***	.000
변화율	-.946***	.000	-.982***	.000	-.992***	.000
상대강도지수	-.627***	.000	-.608***	.000	-.655***	.000
윌리엄스 %R	-.737***	.000	-.750***	.000	-.731***	.000
Slow %K	-.630***	.000	-.646***	.000	-.674***	.000
Slow %D	-.392***	.000	-.387***	.000	-.502***	.000
백분위	.263***	.000	.274***	.000	.250***	.000
회귀계수	-.447***	.000	-.399***	.000	-.562***	.000
로그 수익률	-.824***	.000	-.879***	.000	-.932***	.000
표준편차	.174***	.000	.111***	.000	.165***	.000

변수	케이프		파나막스		수프라막스	
	R^2	$p-value$	R^2	$p-value$	R^2	$p-value$
2일 이동평균	-.127***	.000	-.081***	.000	-.039	.065
3일 이동평균	-.098***	.000	-.058**	.006	-.023	.280
4일 이동평균	-.067**	.002	-.034	.111	-.006	.762
5일 이동평균	-.036	.095	-.010	.641	.010	.641
6일 이동평균	-.006	.776	.013	.539	.026	.224
7일 이동평균	.020	.353	.034	.108	.041	.053
8일 이동평균	.042*	.049	.053*	.013	.056**	.009
9일 이동평균	.061**	.004	.070**	.001	.069**	.001
10일 이동평균	.078***	.000	.084***	.000	.081***	.000

* : $p \leq 0.05$, ** : $p \leq 0.01$, *** : $p \leq 0.001$

마지막으로 각 선형별 차주 운임 상승, 하락 횟수를 살펴보면 케이프는 상승이 1,194번, 하락이 1,273번 있었다. 파나막스는 상승 1,198번 하락 1,269번, 수프라막스는 상승 1,241번 하락 1,226번이 기록 되었다.

제3장

딥러닝모델 분석

제1절 딥러닝

머신러닝(machine learning)은 인공지능을 구성하는 한 부분으로 연구되었는데 ‘컴퓨터가 데이터를 통해 스스로 학습하는 것처럼 하는 기술’을 뜻한다. 딥러닝(deep learning)은 심층 학습이라고도 불리는 머신러닝의 한 방법으로 분류(classification), 군집(clustering), 기계번역(machine translation), 이상탐지(anomaly detection) 등 다양하고 복잡한 문제를 해결하는 데 사용된다.

딥러닝은 여러 분야에서 기존 다른 머신러닝 기법보다 좋은 성능을 보이며 주목 받고 있다. 딥러닝이 주목받는 점은 좋은 성능 외에도 주어진 데이터에서 좋은 성능을 발휘하는 특징을 자동으로 찾아낸다는 점이다. 예를 들어, 이미지를 분류하는 문제를 해결할 때 기존 머신러닝은 색을 구분할 것인지, 경계선을 구분할 것인지 등을 미리 정해 학습을 수행한다. 하지만 딥러닝은 데이터만 주어진다면 자

동으로 어떤 기준으로 판단할지를 학습한다. 그래서 사람이 생각하지 못한 특징을 발견하는 면에서 뛰어난 성능을 보여준다.¹⁷⁾

1. 다층퍼셉트론

다층퍼셉트론(MLP; Multilayer Perceptrons)으로 불리는 심층 신경망(Deep feedforward neural networks)은 딥러닝의 기본이 되는 모델이다.

다층퍼셉트론의 기본적인 구조는 <그림 3-1>과 같다. 행렬 표기법을 활용하여 l 개의 층을 가지는 다층퍼셉트론 구조를 표현하면 아래와 같다.

$$\begin{cases} \mathbf{W}_i : m_i \times m_{i-1} \text{ 가중치 행렬} \\ \mathbf{b}_i : m_i \times 1 \text{ 편향 벡터} \\ m_i : i\text{층의 노드 개수} \end{cases} \quad i = 1, \dots, l$$

$$\mathbf{z}_i = \mathbf{W}_i \mathbf{h}_{i-1} + \mathbf{b}_i, \quad i = 1, \dots, l, \quad \text{여기서 } \mathbf{h}_0 = \mathbf{x}, \quad \mathbf{h}_l = \hat{\mathbf{y}}$$

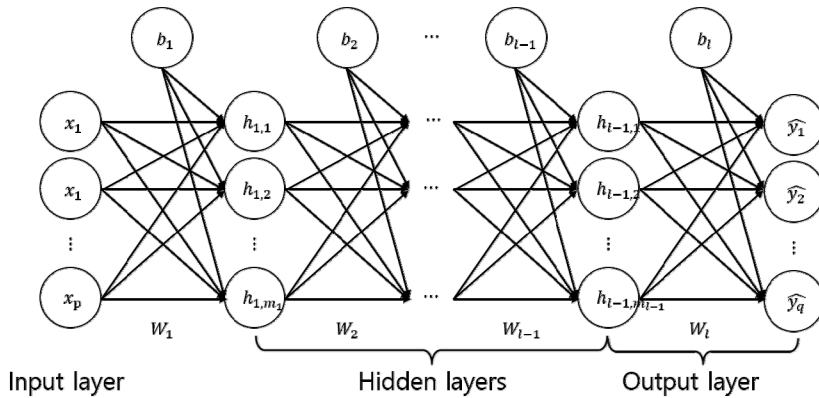
$$\mathbf{h}_i = f_i(\mathbf{z}_i), \quad f_i \text{는 } i\text{층의 활성화 함수}$$

다층퍼셉트론 모델은 학습하기 위해 기존의 머신러닝 모델들과 동일하게 최적화 방법, 손실함수를 선택해야 한다. 다른 머신러닝 모델과 구분되는 특징으로는 복수의 은닉층(hidden layers) 사용이 있다. 복수의 은닉층에서 다음 층으로 정보를 효과적으로 전달하기 위해 신경망 계열에서 활용되는 활성화 함수를 사용한다.

다층퍼셉트론은 은닉층 개수 선택, 손실함수 정의, 활성화 함수 선택에 따라 모델의 성능이 크게 달라지게 된다.

17) 야마시타 타카요시(2016), 번역 심효섭(2017), p.6

〈그림 3-1〉 다층퍼셉트론 구조



자료: 저자 작성

1) 손실함수(Loss functions)

손실함수는 모델의 학습을 통해 얻은 데이터의 추정치와 실제 데이터의 차이를 나타내는 지표로 쓰인다. 다층퍼셉트론 모델 설계에 있어서 손실함수 설정은 중요하다. 손실함수를 설정하게 되면 모델은 손실함수를 최소화하는 방향으로 학습을 진행하게 된다.

손실함수는 비용함수(cost function)로도 불린다. 대표적인 손실함수로는 평균 제곱 오차(MSE)와 크로스 엔트로피(cross-entropy)가 있다.

(1) 평균 제곱 오차(MSE; Mean Squared Error)

평균 제곱 오차는 종속변수가 연속형일 때 유의미하게 사용되는 손실함수로 다음과 같이 정의된다.

$$L(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

평균 제곱 오차는 기본적인 회귀분석에서부터 널리 활용되어 모수를 추정하는 손실함수로 활용된다. 다층퍼셉트론 모델에서 평균 제곱 오차 손실함수를 사용하게 되면 y_i 와 \hat{y}_i 의 차이를 작게 만드는 방향으로 모델이 학습된다.

(2) 크로스 엔트로피(Cross-Entropy)

종속변수가 K개의 범주를 가질 때 사용되는 크로스 엔트로피는 다음과 같이 정의된다.

$$L(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^K y_{ij} \log \hat{y}_{ij}$$

만약 Y가 범주 k에 속한다면 $y_{ij} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$ 의 값을 가진다. \hat{y}_{ij} 은 i번째 입력변수가 주어졌을 때 j번째 범주에 속할 확률로 0에서 1사이의 값을 가지며 모든 확률의 합은 $\sum_{j=1}^K \hat{y}_{ij} = 1$ 이 된다.

동 손실함수는 y_{ij} 의 값이 0일 경우에 $-0 \cdot \log(\hat{y}_{ij})$ 으로 \hat{y}_{ij} 과 무관하게 항상 0의 값을 가지게 된다. 반대로 y_{ij} 값이 1인 경우, 즉 j번째 범주인 경우 $-1 \cdot \log(\hat{y}_{ij})$ 의 손실이 발생하게 되는데 \hat{y}_{ij} 이 0에 가까우면 해당 값은 작아지고 1에 가까울수록 손실은 0이 된다.

따라서 동 손실함수를 활용한 다층퍼셉트론 모델은 입력변수가 주어졌을 때 올바른 범주를 선택할 확률을 1에 가깝게 올바르게 옳은 범주를 선택할 확률은 0에 가깝게 학습하게 된다.

2) 활성화 함수(Activation functions)

활성화 함수는 사람의 신경세포와 유사하게 입력된 신호를 선형 혹은 비선형으로 변환하여 다음 층으로 전달하는 역할을 한다. 신경망에서는 활성화 함수로 비선형 함수를 사용해야 하는데, 이는 선형 함수를 사용할 경우 신경망 층을 깊게 하는 의미가 없어지기 때문이다. 선형 함수의 문제는 층을 아무리 깊게 해도 은닉층이 없는 네트워크로 똑같이 만들 수 있기 때문이다.¹⁸⁾

대표적인 활성화 함수로는 시그모이드(sigmoid), 하이퍼볼릭 탄젠트(hyperbolic tangent), ReLU(Rectified Linear Unit), 소프트맥스(softmax)가 있다.

시그모이드 함수는 다음과 같이 정의된다.

$$f(z) = \frac{1}{1 + e^{-u}}$$

동 함수를 사용하면 0에서 1사이의 값을 출력한다. 입력된 신호가 클 경우 1에, 입력된 신호가 작을 경우 0에 가까운 값을 출력하여 다음 노드로 전달되는 신호의 세기를 조절한다.

다음으로 하이퍼볼릭 탄젠트 함수는 다음과 같이 정의된다.

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

하이퍼볼릭 탄젠트 함수는 입력된 신호를 -1에서 1사이로 변환시킨다. 시그모이드 함수와 유사하게 입력된 신호가 클 경우에는 1에, 입력된 신호가 작을 경우는 -1에 가까운 값을 출력하여 다음 노드로 신호를 전달한다.

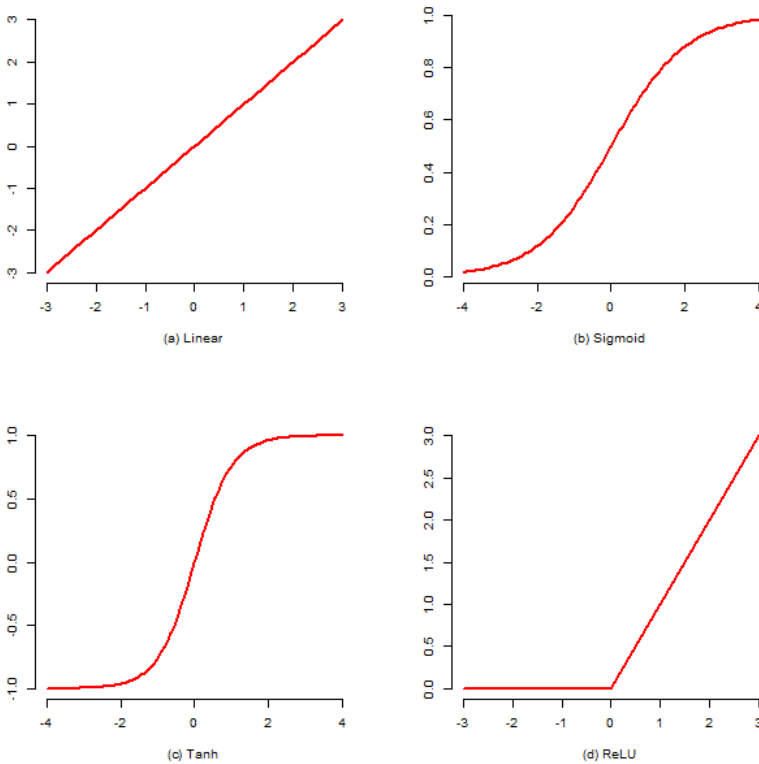
18) 사이토 고키(2016), 번역 김태현(2017) p.75

ReLU는 비교적 최근 주목받는 활성화 함수로 다음과 같이 정의된다.

$$f(z) = \max(0, z)$$

입력된 신호 값이 0보다 작을 경우에는 0을, 0보다 큰 신호 값이 입력될 경우 신호를 그대로 전달한다. 앞의 두 활성화 함수보다 비교적 빠른 연산이 가능한 장점이 있다.

〈그림 3-2〉 활성화함수



자료: 저자 작성

3) 역전파

다층퍼셉트론은 모델의 학습을 위해 역전파(backpropagation) 알고리즘을 활용한다. 역전파는 모델을 통해 계산된 결과와 실제 결과 사이의 손실을 줄이는 방향으로 가중치를 조금씩 조정하는 알고리즘이다. 역전파를 활용한 모델 학습 과정은 네트워크 초기화, 피드포워드, 손실 평가, 역전파, 조정 총 5단계로 구성된다.¹⁹⁾

〈표 3-1〉 다층퍼셉트론 학습 5단계

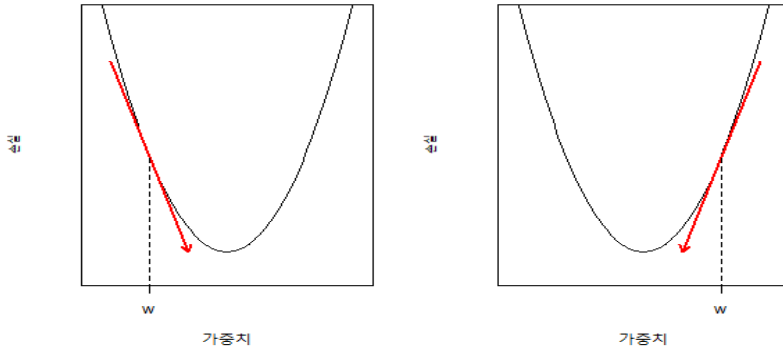
단계		설명
1	네트워크 초기화	• 각 층 가중치, 편향 초기화 (일반적으로 임의로 진행)
2	피드포워드	• 입력층, 은닉층, 출력층까지 신호를 전달
3	손실(오차) 평가	• 사전에 정의한 손실함수에 대해 출력된 결과와 실제 결과를 입력해 오차 평가
4	역전파	• 평가된 오차를 역으로 각 층으로 전파해 경사(gradient) 계산
5	조정	• 오차를 줄이는 방향으로 각 층의 가중치와 편향을 조정

자료: Deep Learning with R, Lewis(2016), 이원상 번역(2017) p.43

네트워크를 학습시키는 기본적인 아이디어는 〈그림 3-3〉과 같다. 계산된 손실함수의 경사(편미분)가 음수인 경우에는 가중치를 증가시키며, 반대로 양수인 경우에는 가중치를 감소시킨다. 따라서 학습을 반복 수행하게 되면 가중치는 손실을 감소하는 방향으로 점차 조정된다.

19) Lewis(2016), 이원상 번역(2017), p.43

〈그림 3-3〉 역전파 알고리즘 기본원리



자료: 저자 작성

2. 과대적합(Overfitting)²⁰⁾

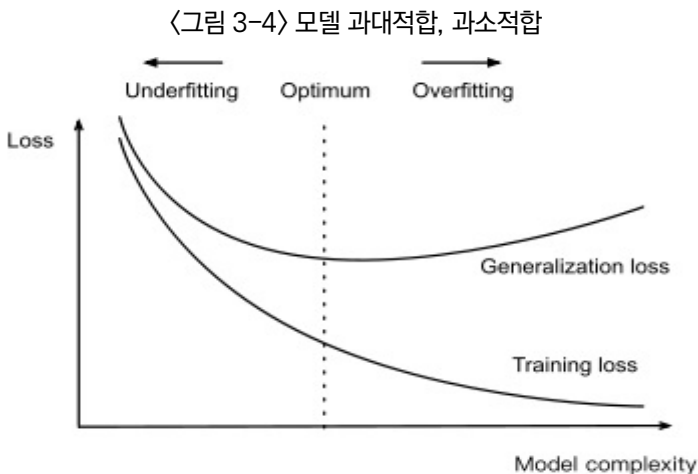
딥러닝 모델의 목적은 훈련 데이터를 통한 학습을 통해 훈련 데이터뿐만 아니라 학습에 사용되지 않은 새로운 데이터에 대해서도 뛰어난 성능을 가지는 것이다. 주어진 훈련 데이터의 훈련 오류(training error)를 최소화하는 것을 최적화문제(optimization problem)라고 한다. 딥러닝은 최적화와 달리 훈련 손실을 최소화하는 것이 아닌 일반화 오류(generalization error) 혹은 시험 오류(test error)를 가능한 작게 만드는 것이다.

일반적으로 모델을 학습하기 위해서 주어진 데이터의 임의추출을 통해 훈련 셋(training set)과 시험 셋(test set)으로 나눈다. 추출을 통해 나뉜 훈련 셋을 학습하여 모델 모수를 조절함과 동시에 시험 셋을 통해 일반화 오류를 측정한다. 이 과정을 통해서 훈련 오류를 작게 만들면서 훈련 오류와 일반화 오류의 차이를 가능한 작게 만들고

20) Goodfellow et al. (2016), pp.108-114 참고하여 작성

자 한다. 훈련 오류가 크다면 과소적합(underfitting)이 발생하게 된다. 그리고 훈련 오류가 작지만 일반화 오류와 크기 차이가 클 경우에는 과대적합이 발생하게 된다.

딥러닝 모델은 동일한 데이터를 반복적으로 학습하기 때문에 학습이 지속되면서 아래 <그림 3-4>와 같이 훈련 오류는 계속해서 감소하지만 일반화 오류는 오히려 증가하는 경향을 보이게 된다. 이를 해결하기 위한 다양한 방법을 사용할 수 있다. 가장 간단한 방법은 학습에 사용되는 데이터 수를 늘려 실제 데이터와 유사하게 만드는 것이다. 하지만 이는 시간과 비용, 자료 확보 문제로 현실적으로 사용하기는 어렵다. 또 다른 방법으로는 학습 중에 특정한 제약조건을 활용하여 훈련 오류가 일정 이하로 떨어지지 않게 하는 것이다. 이 외에도 별도의 검정 셋(validation set)을 활용하여 과대적합이 발생하기 전에 학습을 완료하는 학습 조기종료(early stopping), 학습 시 일부 노드를 제거하는 드롭아웃(dropout) 방법이 널리 사용되고 있다.



자료: Dive into Deep Learning(검색일: 2019. 10. 3)

제2절 모델구조 및 학습방법

1. 학습데이터

학습에 사용된 데이터는 2009년 9월 10일부터 2019년 8월 5일까지인 2,472개 중 모든 변수가 사용가능한 2,205개로 구성되었다. 동 학습 데이터는 모델 평가를 위해 훈련 셋과 시험 셋으로 나누어 구성하였다. 훈련 셋은 2009년 9월 10일부터 2018년 12월 31일까지 2,057개, 시험 셋은 2019년 1월 1일부터 2019년 8월 5일까지 148개로 구성 하였다.

2. 모델구조

본 연구에서는 3개의 은닉 층을 가지고 있는 다층퍼셉트론 모델을 활용했다. 각 은닉 층 노드 개수는 20개로 동일하게 구성하였으며 활성화 함수로는 ReLU 함수를 사용하였다. 출력 층에서는 소프트맥스 함수를 활용하여 차주 운임의 상승 확률, 하락 확률을 예측하였으며, 크로스 엔트로피 손실함수를 통해 모델 손실을 평가하여 손실을 최소화하는 방향으로 모델 학습을 진행하였다. 그리고 과대적합 문제를 해결하기 위해 드롭아웃 기법을 사용하였다.

3. 학습방법

1) 데이터 정규화(normalization)

본 연구에서 사용되는 입력변수들은 0~1 사이의 값을 가지는 변수

가 있는가 하면 20~50 사이의 넓은 범위를 가지는 입력 변수도 있다. 딥러닝은 학습을 시작하기 전에 가중치 행렬을 임의로 초기화 한다. 일반적으로 가중치 행렬의 각 요소들은 일정한 분포에서 작은 값으로 생성된다. 학습 시 가중치 값들이 유사하다면 입력 변수의 값이 클수록 다음 층으로 정보를 전달하는 데 더 큰 역할을 하게 된다. 따라서 모델 학습이 올바르게 진행되지 않는 문제가 발생할 수 있다.

이를 해결하기 위해서는 데이터 학습이전에 변수들을 정규화 해주어야 한다. 본 연구에서는 범위가 고정되어 있는 변수인 상대강도지수, Slow %K, Slow %D, 백분위, 윌리엄스 %R을 제외한 모든 변수의 범위를 [-1, 1]로 정규화하기 위해 아래 식을 이용했다.

$$\frac{2x - (\max - \min)}{\max - \min}$$

max는 주어진 변수의 최대값을 의미하면 min은 주어진 변수의 최소값을 의미한다. 각 변수별로 정규화 과정을 거치게 되면 모든 변수들이 [-1, 1]로 동일한 범위를 가지게 되어 모델의 학습이 쉽게 된다.

2) 페이스팅

본 연구에서는 입력 변수로 기술적 지표들이 사용되기 때문에 시장의 변화가 크지 않을 때 일부 지표들이 유사한 값을 가지게 된다. 따라서 시장의 변화가 크지 않은 시점의 데이터가 연속적으로 학습에 사용되게 되면 상대적으로 해당 패턴을 학습하는 비중이 높아지게 되며, 이에 따라 학습되지 않은 데이터에 대해서 정확한 예측에 실패할 가능성이 있다.

이 문제를 해결하기 위해 페이스팅(pasting) 기법을 활용하였다.

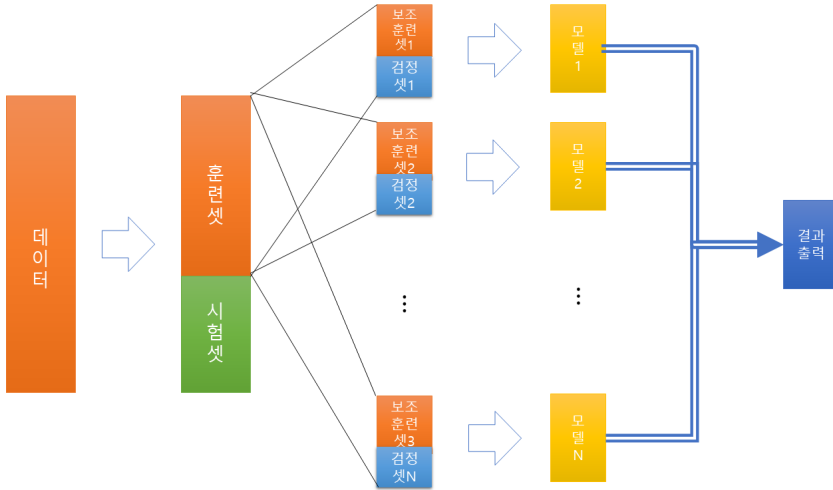
페이스팅 기법은 훈련 데이터를 무작위 비복원 추출하는 과정을 반복하여 다수의 보조 훈련 셋을 구성하고 이를 통해 각기 다른 모델을 만드는 방법을 말한다. 페이스팅 기법을 활용하면 각 모델들은 학습 데이터에 따라 다양한 패턴을 학습하게 된다. 최종적으로 페이스팅 기법을 통해 생성된 다양한 모델들을 앙상블(ensemble)하여 출력결과로 사용함으로써 과대적합 문제를 해결할 수 있다.

3) 모델학습

본 연구에서는 모델 학습에 사용된 훈련 셋에 페이스팅 기법을 활용하여 임의 추출을 통해 보조 훈련 셋 1,600개와 검정 셋(validation set) 457개를 20번 반복 생산하여 학습에 활용하였다. 생성된 보조 훈련 셋을 기준으로 데이터 정규화를 실시하여 변수들의 범위를 [-1, 1]로 동일하게 생성하였으며, 이 때 각 변수들의 최대값 최소값을 기억하여 다른 데이터를 적용할 때 이 변수들을 활용해 정규화를 실시하였다.

각 모델들은 전체 데이터 셋을 10,000번 반복 학습하였으며 학습이 반복 될 때 마다 셋 별로 크로스 엔트로피와 모델 정확도를 계산하였다. 학습이 종료된 후 크로스 엔트로피와 모델 정확도가 가장 좋을 때의 모수를 최종 모델의 모수로 선택해 차주 운임 방향성 예측모델로 활용하였다.

〈그림 3-5〉 학습 모델 구조



자료: 저자 작성

제 4 장

분석결과

제1절 딥러닝 모델 분석결과

1. 모델 평가지표²¹⁾

분류 모델은 성능평가를 위해 아래 그림과 같은 혼동행렬(confusion matrix)을 활용한다. 혼동행렬은 실제 값과 예측 값을 측정해 <표 4-1>과 같이 작성된다.

<표 4-1> 혼동행렬 구성

		실제 정답	
		0	1
분류 결과	0	True Positive(TP)	False Positive(FP)
	1	False Negative(FN)	True Negative(TN)

자료: 저자 작성

21) Tistory 블로그(검색일:2019. 9. 30) 참고 작성

0과 1을 분류하는 문제를 예를 들면 'True', 'False'는 실제 값을 올바르게 예측한 것을 의미하고 'Positive', 'Negative'는 예측 값이 0인지 1인지를 의미한다. 즉, True Positive(TP)는 0이라고 예측했는데 실제 값이 0인 경우이며, False Negative(FN)은 실제 값이 0인데 1로 예측, False Positive(FP)는 실제 값이 1인데 0으로 예측, True Negative(TN)은 실제 값이 1인데 1로 예측한 경우를 나타낸다.

혼동행렬은 훈련 셋이 아닌 시험 셋을 이용하여 작성된다. 혼동행렬을 바탕으로 여러 지표를 통해 모델성능을 평가하게 된다. 대표적인 지표로 정확도, F1 score, AUC가 있다.

1) 정밀도(Precision), 재현율(Recall), 정확도(Accuracy)

정밀도, 재현율, 정확도는 모델 성능을 평가하는 직관적인 지표로 많은 논문에서 사용된다. 정밀도는 모델이 0으로 분류한 것 중 실제 0인 비율로 Positive 정답률이라고도 불린다. 재현율은 실제 0인 것 중 모델이 0으로 예측한 비율로 민감도(Sensitivity)라고도 한다. 수식으로는 다음과 같이 표현할 수 있다.

$$(\text{Precision}) = \frac{TP}{TP + FP},$$

$$(\text{Recall}) = \frac{TP}{TP + FN}$$

정밀도는 모델 입장에서 재현율은 실제 입장에서 정답을 맞춘 것을 바라본다. 만약 모든 예측을 0으로 하게 되면 재현율은 1이 되지만 정밀도는 재현율 대비 낮은 값을 가지게 된다. 한 가지 지표만 사용한 모델 평가는 한계가 있기에 정밀도와 재현율을 동시에 확인해야 한다.

정확도는 0을 0으로 1을 1로 맞춘 비율로 식으로는 다음과 같이 나타난다.

$$(\text{Accuracy}) = \frac{TP + TN}{TP + FN + FP + TN}$$

정확도는 가장 직관적으로 모델의 성능을 나타낼 수 있는 평가 지표이다. 하지만, 실제 값이 편향(bias)되어 있다면 해당 값만 예측 했을 때 정확도는 높게 나오지만 모델 성능이 높다고는 말할 수 없다. 따라서 정밀도, 재현율, 정확도를 함께 활용하여 모델 성능을 평가해야 한다.

2) F1 score

F1 score 지표는 정밀도와 재현율의 조화평균으로 계산된다.

$$(\text{F1-score}) = 2 \times \frac{1}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 score는 실제 값이 한쪽으로 치우쳐 있을 때, 모델의 성능을 정확하게 평가할 수 있는 지표로 알려져 있다.

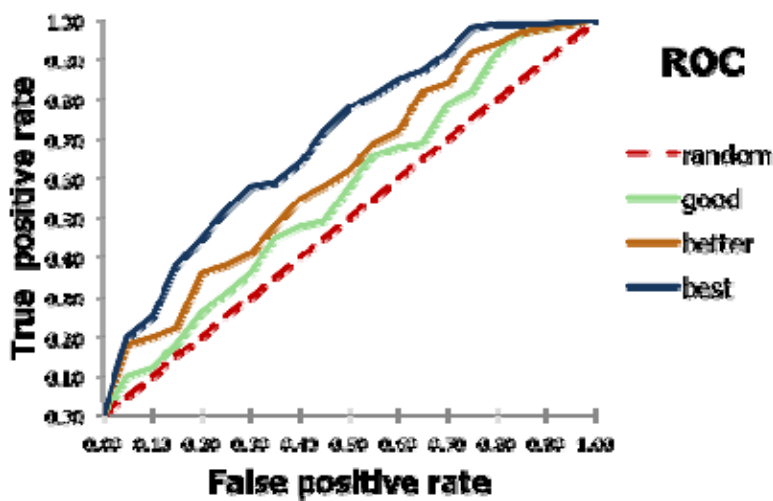
3) ROC와 AUC

FPR(False Positive Rate)는 실제 1인 데이터 중에서 모델이 0이라고 예측한 비율 즉 잘못 예측한 것으로 수식으로는 다음과 같이 표현된다.

$$FPR = \frac{FP}{TN + FP}$$

ROC(Receiver Operating Characteristic) 곡선은 여러 임계값(Threshold)을 기준으로 TPR(재현율)과 FPR의 변화를 시각화한 곡선이다. ROC 곡선은 $y=x$ 그래프보다 상단에 위치하고 있어야 어느 정도 모델 성능이 있는 것으로 말할 수 있다. 모델 간 성능 비교 시에는 ROC 곡선이 왼쪽 위 모서리와 가까울수록 성능은 좋은 것으로 평가할 수 있다.

〈그림 4-1〉 ROC 곡선



자료: OpenEye Python Cookbook(검색일: 2019. 9. 30)

2. 모델 분석결과

페이스팅 기법을 통해 학습된 케이프 차주 운임 방향성 예측은 아래 <표 4-2>와 같이 나타난다. 페이스팅 기법을 통한 20개 모델 손실 평균은 훈련 셋, 검정 셋, 시험 셋에서 각기 0.232, 0.288, 0.324로 나타났으며, 정확도는 0.781, 0.695, 0.646으로 나타났다. 동 모델들을 앙상블한 결과는 페이스팅 모델의 평균치보다 개선되었으며, 훈련 셋과 시험 셋의 손실차이가 줄어들어 과대적합문제를 다소 해결하였음을 알 수 있다.

<표 4-2> 케이프 모델 학습결과

구 분	훈련 셋				시험 셋	
	보조 훈련 셋		검정 셋			
	손실	정확도	손실	정확도	손실	정확도
model1	.255	.745	.300	.678	.296	.706
model2	.204	.816	.260	.729	.351	.650
model3	.211	.816	.263	.735	.380	.510
model4	.238	.774	.274	.726	.319	.650
model5	.218	.813	.289	.685	.336	.615
model6	.228	.800	.292	.698	.301	.734
model7	.233	.781	.302	.678	.313	.706
model8	.246	.764	.282	.716	.329	.671
model9	.308	.643	.327	.600	.307	.643
model10	.212	.808	.275	.711	.311	.664
model11	.233	.777	.312	.698	.316	.650
model12	.222	.785	.276	.713	.320	.601
model13	.206	.819	.275	.720	.317	.664
model14	.241	.773	.298	.663	.355	.587
model15	.213	.816	.300	.670	.330	.608
model16	.200	.834	.267	.705	.379	.587
model17	.239	.784	.283	.696	.325	.671

구 분	훈련 셋				시험 셋	
	보조 훈련 셋		검정 셋			
	손실	정확도	손실	정확도	손실	정확도
model18	.241	.768	.266	.724	.306	.622
model19	.248	.766	.298	.696	.287	.720
model20	.250	.748	.311	.667	.313	.650
평균	.232	.781	.288	.695	.324	.646
양상블	.218	.850	.268	.738	.290	.709

자료: 저자 작성

다음으로 파나막스 차주 운임 방향성 모델 결과는 아래 <표 4-3>과 같이 나타난다. 페이스팅 기법을 통한 20개 모델 손실평균은 훈련 셋, 검정 셋, 시험 셋에서 각기 0.175, 0.229, 0.309로 나타났으며, 정확도는 0.846, 0.784, 0.708로 나타났다. 파나막스선의 경우에도 케이프와 마찬가지로 양상블 결과가 페이스팅 모델들의 평균치보다 좋게 나타났으며, 훈련 셋과 시험 셋의 손실차이도 줄어들어 과대적합 문제를 일부 해결했음을 알 수 있다.

<표 4-3> 파나막스 모델 학습결과

구 분	훈련 셋				시험 셋	
	보조 훈련 셋		검정 셋			
	손실	정확도	손실	정확도	손실	정확도
model1	.148	.881	.190	.834	.426	.713
model2	.191	.826	.241	.759	.311	.692
model3	.156	.860	.206	.818	.324	.699
model4	.188	.833	.229	.788	.256	.762
model5	.175	.843	.255	.757	.217	.776
model6	.200	.817	.232	.772	.260	.748
model7	.175	.841	.248	.740	.403	.608
model8	.168	.856	.204	.832	.288	.720
model9	.172	.853	.238	.775	.327	.657

구 분	훈련 셋				시험 셋	
	보조 훈련 셋		검정 셋			
	손실	정확도	손실	정확도	손실	정확도
model10	.164	.869	.196	.821	.217	.797
model11	.209	.798	.267	.744	.296	.713
model12	.186	.829	.253	.742	.334	.650
model13	.163	.874	.219	.783	.340	.657
model14	.151	.883	.243	.770	.347	.720
model15	.156	.859	.202	.794	.327	.671
model16	.191	.832	.209	.821	.225	.748
model17	.185	.829	.248	.770	.225	.776
model18	.180	.841	.230	.814	.400	.650
model19	.170	.851	.226	.799	.367	.678
model20	.169	.854	.251	.742	.298	.727
평균	.175	.846	.229	.784	.309	.708
양상블	.163	.879	.211	.810	.271	.723

자료: 저자 작성

마지막으로 수프라맥스 차주 운임 방향성 모델 결과는 아래 <표 4-4>와 같이 나타난다. 페이스팅 기법을 통한 20개 모델 손실평균은 훈련 셋, 검정 셋, 시험 셋에서 각기 0.143, 0.188, 0.298로 나타났으며, 정확도는 0.878, 0.828, 0.794로 나타났다. 수프라맥스 모델의 경우 훈련 손실과 시험 손실이 다른 두 선형에 비해 큰 차이를 나타내 훈련 셋에 더욱 과대적합 됐음을 알 수 있었다. 동 모델도 양상블 결과가 페이스팅 모델 평균보다 좋게 나타났으며 과대적합 문제도 다소 해결했음을 알 수 있다.

〈표 4-4〉 수프라maks 모델 학습결과

구 분	훈련 셋				시험 셋	
	보조 훈련 셋		검정 셋			
	손실	정확도	손실	정확도	손실	정확도
model1	.151	.873	.195	.821	.363	.769
model2	.141	.885	.174	.862	.422	.762
model3	.141	.879	.175	.851	.272	.811
model4	.159	.860	.202	.821	.237	.818
model5	.143	.876	.207	.788	.479	.755
model6	.137	.881	.186	.825	.225	.839
model7	.134	.879	.190	.829	.327	.762
model8	.135	.890	.181	.825	.227	.797
model9	.127	.893	.162	.858	.396	.790
model10	.143	.878	.180	.838	.212	.846
model11	.135	.896	.161	.856	.210	.832
model12	.180	.834	.220	.796	.271	.762
model13	.141	.886	.184	.825	.276	.762
model14	.132	.891	.192	.829	.315	.804
model15	.148	.856	.186	.818	.329	.797
model16	.130	.889	.199	.816	.224	.797
model17	.138	.883	.185	.823	.290	.790
model18	.149	.872	.208	.832	.232	.811
model19	.177	.855	.208	.801	.395	.769
model20	.123	.899	.171	.856	.254	.811
평균	.143	.878	.188	.828	.298	.794
양상블	.133	.900	.177	.840	.218	.797

자료: 저자 작성

양상블 모델을 기반으로 한 차주 방향성 예측결과와 실제결과를 통한 혼동행렬은 아래 표와 같이 구성된다.

〈표 4-5〉 케이프 혼동행렬

훈련 셋		실제		
		상승	하락	
분류	상승	767	136	903
	하락	209	945	1,154
		976	1,081	

시험 셋		실제		
		상승	하락	
분류	상승	53	17	70
	하락	26	52	78
		79	69	

자료: 저자 작성

〈표 4-6〉 파나막스 혼동행렬

훈련 셋		실제		
		상승	하락	
분류	상승	831	139	970
	하락	131	956	1,087
		962	1,095	

시험 셋		실제		
		상승	하락	
분류	상승	73	21	94
	하락	20	34	54
		93	55	

자료: 저자 작성

〈표 4-7〉 수프라막스 혼동행렬

훈련 셋		실제		
		상승	하락	
분류	상승	906	102	1,008
	하락	131	918	1,049
		1,037	1,020	

시험 셋		실제		
		상승	하락	
분류	상승	65	15	80
	하락	15	53	68
		80	68	

자료: 저자 작성

혼동행렬을 바탕으로 선형별 모델 평가지표를 통해 각 모델들을 비교해 보겠다. 차주 운임이 상승한 경우를 기준으로 지표를 계산하게 되면 〈표 4-8〉과 같은 결과를 얻을 수 있다. 모든 지표가 수프라막스 모델의 성능이 가장 좋으며 이어서 파나막스, 케이프 순으로 나타났다. 또한, 훈련 셋과 시험 셋에서 각 지표들의 차이가 수프라막스에서 가장 작게 나타났는데 이는 수프라막스 선형에서 과대적합문제가 가장 적었음을 알 수 있다.

〈표 4-8〉 선형별 평가지표

지표	모델	케이프		파나막스		수프라막스	
		훈련	시험	훈련	시험	훈련	시험
정밀도		.786	.671	.864	.785	.874	.812
재현율		.849	.757	.857	.777	.899	.812
정확도		.832	.709	.869	.723	.887	.797
F1 Score		.816	.711	.860	.781	.886	.812
AUC		.915	.784	.944	.797	.963	.880

자료: 저자 작성

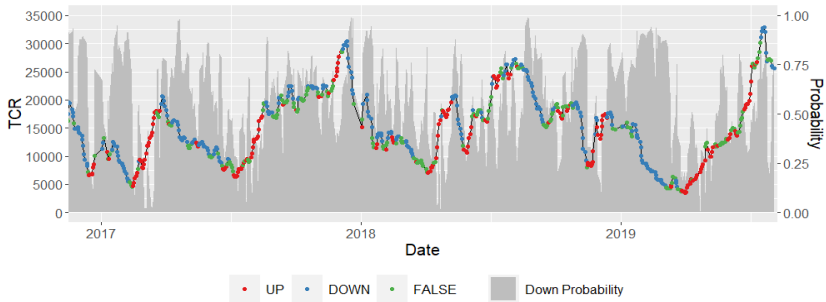
선형별 예측결과를 차트로 나타내면 〈그림 4-2〉, 〈그림 4-3〉, 〈그림 4-4〉와 같다. 실제 운임 방향성과 예측 결과가 모두 상승이면 빨간색, 모두 하락이면 파란색으로 표시되며 실제와 예측 결과가 다를 경우에는 초록색으로 표현된다.

케이프선의 경우 상대적으로 운임의 변동성이 크게 나타나 다른 선형에 비해 오차가 크게 나타났다. 특히 시장이 위아래로 움직이며 큰 방향성을 가질 때 상승 확률과 하락 확률이 비슷하여 이에 대한 분류가 정확하게 되지 않았음을 알 수 있다. 하지만 운임의 큰 방향성이 바뀌는 시기에는 정확하게 운임 방향성을 예측하는 것을 알 수 있다.

파나막스선은 케이프선에 비해 시장이 위아래로 움직이는 국면에서도 높은 정확도를 보여주었다. 파나막스 예측 모델은 순간 운임 변화에 따라 차주 운임 상승, 하락 확률이 급격하게 변화는 지점이 많은 것을 확인할 수 있다.

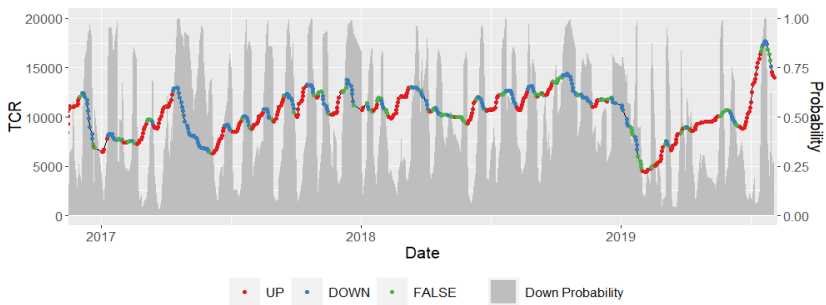
수프라막스선은 운임의 방향성이 일정하고 변동 폭이 크지 않아 잘못된 예측을 하는 경우가 가장 적음을 확인할 수 있다. 수프라막스의 경우 추세가 정해져 있을 때 높은 확률로 방향성을 맞추고 있다. 방향성을 맞추지 못하는 경우는 일부 변곡점에서 발생했는데, 짧은 기간 내로 올바른 방향을 예측하는 것을 볼 수 있다.

〈그림 4-2〉 케이프 예측결과



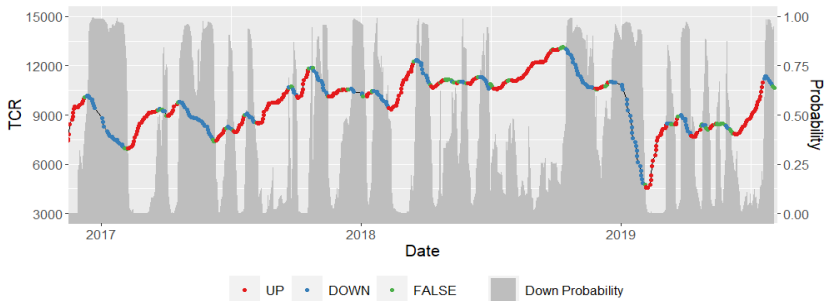
자료: 저자 작성

〈그림 4-3〉 파나마스 예측결과



자료: 저자 작성

〈그림 4-4〉 수프라maks 예측결과



자료: 저자 작성

제2절 모델성능비교

1. 비교모델

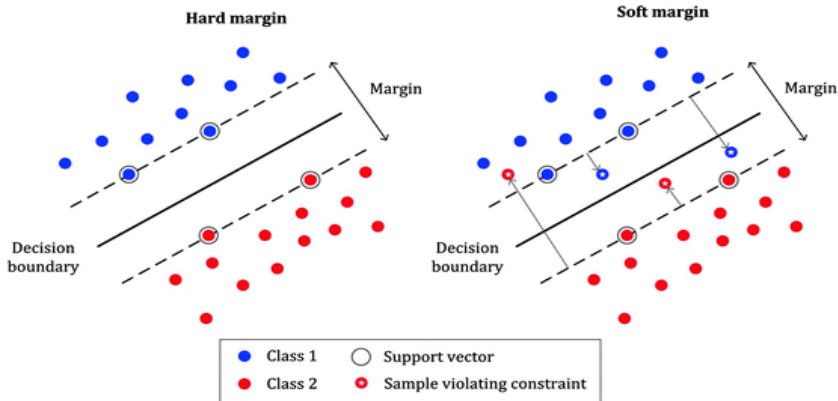
본 연구에서는 심층신경망 기반 예측 모델 성능을 평가하기 위해 대표적인 기계학습 모델인 서포트 벡터 머신과 랜덤 포레스트를 비교해보았다.

1) 서포트 벡터 머신(SVM)

서포트 벡터 머신(SVM; Support Vector Machine)은 주어진 데이터 집합에서 데이터를 가장 잘 분리하는 최적의 초평면(hyperplane)을 찾는다. 초평면을 기준으로 가장 가까운 각 범주의 가장 가까운 데이터를 서포터 벡터(support vector)라고 부르는데 이 서포트 벡터와 초평면 사이 직선거리의 2배를 마진(margin)이라고 한다. 서포트 벡터 머신은 마진을 최대한 크게 만드는 초평면을 통해 데이터를 분류한다. 즉, 명확한 경계선을 만드는 초평면을 찾는 것이다.

서포트 벡터 머신은 기본적으로 두 범주를 완벽하게 분류하는 초평면을 찾고자 하지만 하나의 평면으로 데이터를 완벽하게 구분할 수 없는 경우가 있다. 이를 위해 $\text{cost}(C)$ 로 부르는 하이퍼파라미터(hyperparameter)를 도입하여 오차를 허용하는 정도를 결정한다. C 값이 커질수록 오차를 적게 허용하여 주어진 데이터에 과대적합되는 문제가 발생하며 C 값을 작게 하면 주어진 데이터에 과소적합되는 문제가 있어 적절한 C 값을 선택하는 것이 중요하다.

〈그림 4-5〉 서포트 벡터 머신



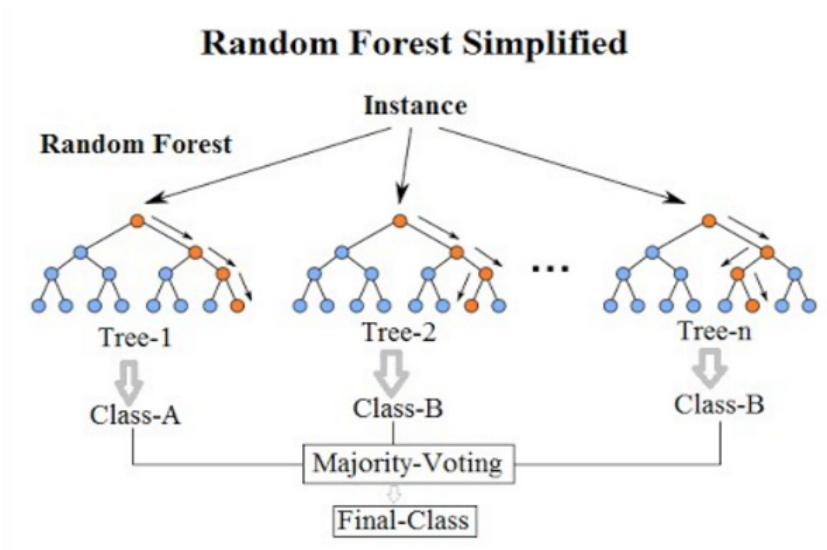
자료: Medium 블로그(검색일: 2019. 10. 6)

2) 랜덤포레스트(Random Forest)

의사결정나무(decision tree)는 의사결정 규칙을 나무 구조로 나타내 전체 데이터를 몇 개의 집단으로 분류하는 방법이다. 의사결정나무는 단순한 방법으로 설명변수와 종속변수의 영향력을 쉽게 파악할 수 있는 장점이 있다. 하지만, 간단한 방법으로 인해 예측력이 떨어지며 학습데이터가 아닌 새로운 데이터에 대해 낮은 성능을 보이는 과대적합이 발생하는 단점이 있다.

랜덤포레스트는 의사결정나무가 가지고 있는 이러한 단점을 극복할 수 있는 모델이다. 랜덤포레스트는 간단하게 설명하면 여러 개의 의사결정트리를 통해 집단을 분류하는 방법이다. 랜덤포레스트 학습 방법은 다음과 같다. 주어진 학습 셋에서 무작위 복원추출로 n 개 데이터를 선택하고 해당 데이터에서 d 개의 변수만 활용하여 의사결정나무를 만든다. 동 과정을 k 번 반복하여 생성된 의사결정나무들을 다수 투표(majority vote)를 통해 최종 범주를 분류한다.

〈그림 4-6〉 랜덤 포레스트



자료: Medium 블로그(검색일: 2019. 10. 6)

2. 모델성능비교 결과

비교모델은 다층퍼셉트론 모델과 동일하게 페이스팅 기법을 활용하여 20개의 모델을 생성하고 이를 앙상블하여 결과를 출력 하였다.

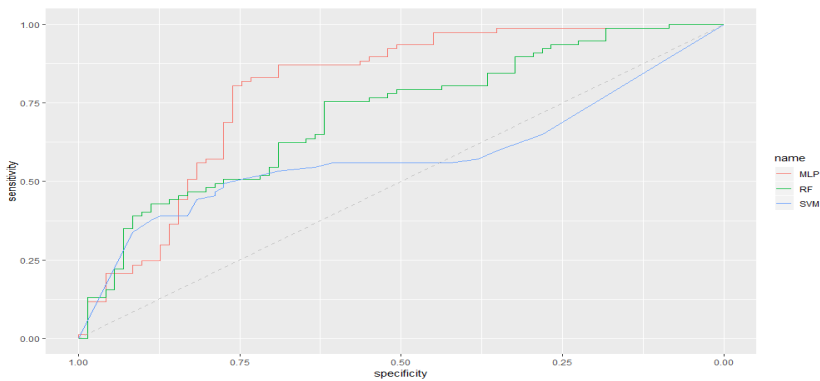
케이프션의 경우 훈련 셋에 있어서는 랜덤포레스트가 가장 성능이 좋았으며 서포트 벡터 머신, 다층퍼셉트론 순으로 모델 성능이 좋은 것으로 나타났다. 하지만 시험 셋에 대해서는 다층퍼셉트론을 사용한 경우 모든 지표에서 가장 좋은 모델로 나타났다. 특히 다층퍼셉트론은 세 모델 중 훈련 셋과 시험 셋의 차이가 가장 작았다. 이는 다층퍼셉트론이 비교적 훈련 데이터에 과대적합 되지 않았음을 나타낸다. 또한, ROC 곡선을 보았을 때에도 다층 퍼셉트론이 가장 높은 성능을 보임을 알 수 있다.

〈표 4-9〉 케이프 모델평가

지표	모델	다층퍼셉트론 (MLP)		서포트 벡터 머신 (SVM)		랜덤포레스트 (RF)	
		훈련	시험	훈련	시험	훈련	시험
정밀도		.786	.671	.977	.481	.998	.608
재현율		.849	.757	.972	.704	1	.686
정확도		.832	.709	.976	.615	.999	.642
F-score		.816	.711	.975	.571	.999	.644
AUC		.915	.784	.992	.561	1	.712

자료: 저자 작성

〈그림 4-7〉 케이프 ROC 곡선



자료: 저자 작성

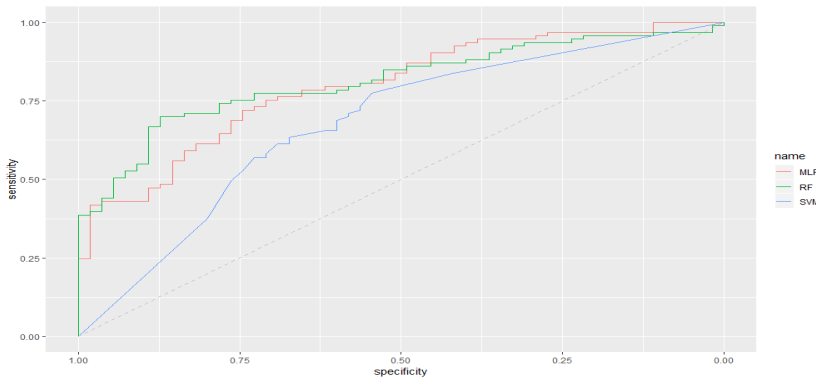
파나마스선의 분석 모델의 경우 전체적으로 케이프선보다 모델 성능을 나타내는 지표들의 값이 상승했다. 훈련 셋에 대해서는 케이프와 마찬가지로의 결과를 기록하였으며, 시험 셋에서는 정밀도와 F-score 지표에서 다층퍼셉트론이 우세한 것으로 나타났다. 랜덤포레스트 모델은 정밀도와 재현율의 차이가 큰 것을 통해 치우쳐진 분류를 하는 것을 알 수 있다. ROC 곡선을 보았을 때 랜덤포레스트 모델과 다층퍼셉트론 모델은 유사한 결과를 나타낸 것으로 확인된다.

〈표 4-10〉 파나막스 분석결과

지표	모델	다층퍼셉트론 (MLP)		서포트벡터머신 (SVM)		랜덤포레스트 (RF)	
		훈련	테스트	훈련	테스트	훈련	테스트
정밀도		.864	.785	.983	.656	1	.656
재현율		.857	.777	.975	.744	1	.910
정확도		.869	.723	.981	.642	1	.743
F-score		.860	.781	.979	.697	1	.763
AUC		.944	.797	.995	.679	1	.812

자료: 저자 작성

〈그림 4-8〉 파나막스 ROC 곡선



자료: 저자 작성

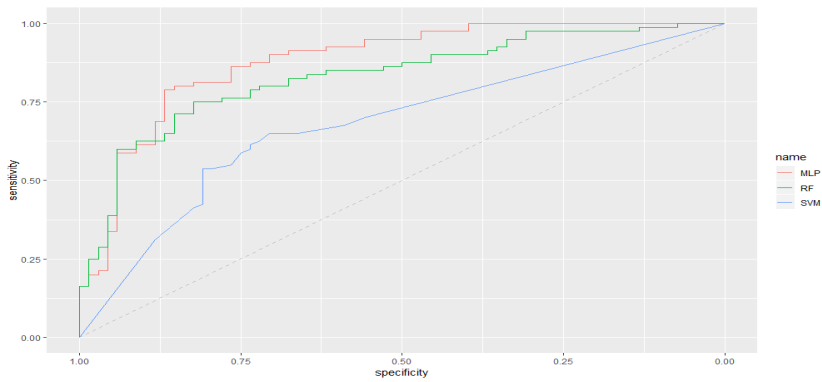
수프라막스 분석 결과는 케이프선의 결과와 동일하게 훈련 셋에서는 랜덤포레스트가 가장 우수했으며, 시험 셋에서는 다층퍼셉트론 모델이 가장 우수한 것으로 나타났다. ROC 곡선을 통해서도 다층퍼셉트론 모델이 다른 두 모델보다 위에 위치함을 확인할 수 있다.

〈표 4-11〉 수프라막스 분석결과

지표	모델	다층퍼셉트론 (MLP)		서포트벡터머신 (SVM)		랜덤포레스트 (RF)	
		훈련	테스트	훈련	테스트	훈련	테스트
정밀도		.874	.812	.972	.588	1	.775
재현율		.899	.812	.970	.734	1	.775
정확도		.887	.797	.971	.662	1	.757
F-score		.886	.812	.971	.653	1	.775
AUC		.963	.88	.993	.678	1	.831

자료: 저자 작성

〈그림 4-9〉 수프라막스 ROC 곡선



자료: 저자 작성

제 5 장

결론

제1절 연구결과 요약

1. 딥러닝모델 분석 결과

별크선의 3대 주요 선형인 케이프, 파나막스, 수프라막스에 대하여 기술적 분석 지표를 중심으로 입력변수를 추출하고, 딥러닝 모델을 적용하여 주간단위의 시황변동을 예측했다. 파라미터 추정은 20회 반복되었으며 최종적으로 복수 추정치를 결합한 앙상블 모델이 적용되었다.

그 결과 케이프선에서는 정확도가 70.9%, 파나막스에서는 72.3%, 수프라막스에서는 79.7%로 나타났다. 시장의 유동성이 작은 대형선 일수록 운임의 변동 폭이 크고 단기적인 공급의 쏠림현상, 기상이변 또는 사고에 의한 추세이탈이 크기 때문에 정확도가 상대적으로 낮게 나타난 것은 시장참여자의 직관과 일치하는 것이다.

2. 비교모델 분석 결과

비교모델로는 머신러닝 기법으로 활발히 적용되고 있는 서포트 벡터 머신(SVM)과 랜덤 포레스트(RF)가 사용되었다. 비교모델의 경우에도 딥러닝 모델과 마찬가지로 페이스팅 기법으로 데이터를 추출하여 20개의 모델을 생성하고 이를 앙상블 하여 결과를 도출하였다.

케이프선의 경우 SVM과 RF의 정확도가 각각 61.5%, 64.2%에 그쳤으나 딥러닝 모델의 경우 70.9%로 가장 높았다. 파나막스의 경우에는 RF의 정확도가 74.3%로 딥러닝 모델보다 다소 높았으며 SVM의 성과는 64.2%에 그쳤다. 수프라막스의 성과는 케이프와 마찬가지로 딥러닝 모델이 79.7%로 가장 높았고 RF가 75.7%, SVM이 66.2%로 가장 낮은 결과를 보여주었다.

세 선형을 종합해서 볼 때 딥러닝 모델의 성과가 가장 좋고 그 다음이 RF, SVM의 순이었다. SVM이나 RF가 인공지능망 계열의 모델보다 우수한 성과를 보인 연구도 있으나 이 연구에서는 딥러닝 모델의 성과가 우수한 것으로 나타났다 이는 입력변수의 선정과 데이터 샘플링 방법에 따라 결과에 차이가 있을 수 있음을 의미하는 것으로 해석된다.

제2절 연구의 기여

이 연구는 시황의 단기방향성 예측에 대한 인공지능모델의 유효성을 검증한 것이다. 해운의 용대선 의사결정에 있어 단기예측은 매우 중요하다. 단기 운임시황의 변동성이 다른 영역에 비해 크기 때문에

단기적인 시점의 이동이 가져오는 성과의 차이도 매우 크다. 예를 들어 케이프의 경우 스팟 항차를 수행하기 위해 선박을 확보하는 의사결정을 내릴 때 선박을 오늘 용선할 수도 있고 1주일 후에 용선할 수도 있다. 만일 용선 데스크에서 1주일 정도의 시황을 정확하게 예측할 수 있다면 계약 시점을 조정함으로써 용선비용에 상당한 차이를 가져올 수 있을 것이다.

또한 해운시황의 분석에 인공지능 기반의 새로운 기법이 적용되는 것은 매우 바람직한 현상이다. 재무나 자원 등 다른 영역에서 인공지능기반의 모델이 기존의 시계열 분석보다 우수한 성과를 보인 연구가 지속적으로 나오고 있다. 특히 1년 이상 중장기 예측에서 기존 시계열분석 모델의 성과가 부진함에 비추어 기계학습 기반의 새로운 방법론이 예측연구의 지평을 확대할 가능성이 크다고 할 수 있다.

이 연구는 해운기업의 용대선 의사결정의 질을 높여 경영성과를 개선하는 데 직접적인 기여를 할 것으로 생각된다. 아울러 이 연구가 중장기적인 예측 연구로 확장될 경우 기업의 선박투자 의사결정, 해운 금융기관의 금융 의사결정, 정부의 정책 의사결정에도 긍정적인 영향을 줄 것이다.

제3절 추가연구

이 연구는 벌크선의 단기시황예측으로 범위를 한정하여 진행되었다. 이 연구의 확장은 다음의 3가지 측면에서 고려해 볼 수 있다.

첫째, 연구의 분야를 벌크선에서 탱커와 컨테이너로 확장할 수 있다. 탱커의 경우 현재는 우리나라 선사의 노출 선대규모가 작아 연구

우선순위가 높지 않지만 향후 우리 해운산업이 균형적으로 발전하게 된다면 유효성이 증대될 것이다. 컨테이너는 운임의 변동성이 상대적으로 작으며 시계열 자료의 한계로 머신러닝 기법을 적용하는 데 한계가 있으나 다른 변수의 활용 등을 통해 한계를 극복할 가능성을 모색할 수 있다.

둘째, 입력변수의 다양성을 확보하는 것이다. 이 연구는 주로 기술적 분석에 사용되는 입력변수를 사용하여 진행되었지만 예측의 기간을 길게 할수록 적절한 외생변수의 중요성이 더욱 커지게 될 것이다. 변수선정을 위한 탐색적 연구가 활발히 진행되어 유효한 외생변수가 확보되어야 예측연구의 범위를 확장할 수 있다.

셋째, 예측연구의 대상기간을 확장하는 것이다. 해운기업에서는 대체로 단기예측은 조직 하위의 전술적인 의사결정이고 장기예측은 상위의 전략적인 의사결정 즉, 투자, 사업포트폴리오와 관련된 의사결정이다. 또한 이러한 전략적인 의사결정은 해운기업 뿐만 아니라 해운금융기관, 정부 등 장기적이고 정책적인 의사결정을 내리는 주체들에게도 중요한 의사결정이 된다.

참고문헌

〈국내 문헌〉

- 모수원. 「2010 년 BDI 의 예측: ARIMA 모형과 HP 기법을 이용하여」. 『한국항만경제학회지』 26.1, 2010: pp.222-233.
- 이우식. 「딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측」. 『한국데이터정보과학회지』 28.2, 2017: pp.287-295.
- 임상섭 외, 「Supramax Bulk Carrier Market Forecasting with Technical Indicators and Neural Networks」, 『한국항해항만학회지』 42.5, 2018: pp.341-346.
- 정봉민, 『해운경제학』, 한국해양수산개발원, 2015, p.128
- Lewis, Nigel Da Costa. 이원상 옮김, 『R로 하는 쉬운 딥러닝』, 2017, p.43
- Murphy, John J. 최용석 옮김, 『금융시장의 기술적 분석』, 국일증권경제연구소, 2000.
- Stopford, Martin, 양창호 외 옮김, 『해운경제학』, 2015
- 齋藤康毅. 김태현 옮김, 『밑바닥부터 시작하는 딥러닝』, 한빛미디어, 2017, p.75
- 山下隆義, 심효섭 옮김, 『그림과 수식으로 배우는 통통 딥러닝』, 제이펍, 2017, p.6

〈국외 문헌〉

- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.
- Lewis, Nigel Da Costa, *Deep learning made easy with R: A gentle introduction for data science*, AusCov, 2016.
- Murphy, John J, *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*, Penguin, 1999.
- Lyridis, D. V., et al, “Forecasting tanker market using artificial neural networks”, *Maritime Economics & Logistics* 6.2, 2004: pp.93-108.

Shynkevich, Yauheniya, et al. "Forecasting price movements using technical indicators: Investigating the impact of varying input window length". *Neurocomputing* 264, 2017: pp.71-88.

Stopford, Martin. *Maritime economics 3e*. Routledge, 2009.

斎藤康毅. 『ゼロから作る Deep Learning—Python で学ぶディープラーニングの理論と実装』, 오라이리ー · ジャパン, 2017

山下隆義, 『イラストで学ぶディープラーニング』, 講談社, 2016

〈인터넷 자료〉

Dive into Deep Learning 홈페이지(https://ko.d2l.ai/chapter_deep-learning-basics/underfit-overfit.html)(검색일:2019. 10. 3)

Medium 블로그(<https://medium.com/@ankitnitjsr13/math-behind-svm-support-vector-machine-864e58977fdb>)(검색일:2019. 10. 6)

Medium 블로그(<https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>)(검색일:2019. 10. 6)

OpenEye Python Cookbook 홈페이지(<https://docs.eyesopen.com/toolkits/cookbook/python/plotting/roc.html>)(검색일:2019. 9. 30)

Tistory 블로그(<https://sumniya.tistory.com/26?category=818582>)(검색일:2019, 9, 30)

〈데이터〉

볼틱해운거래소 홈페이지(<https://www.balticexchange.com>)(검색일:2019. 8. 5)

클락슨 홈페이지(<https://sin.clarksons.net>)(검색일:2019. 8. 5)

기본연구보고서 발간목록

2019년

01	한반도 평화 체제 수립 대비 해양 관련 법제도 정비 방안	최지현
02	수산부문 전망모형「KMI-FOSIM」구축 연구(2차년도)	이현동
03	4차 산업혁명 시대의 스마트 어촌 구축방안 연구	박상우
04	해양수산분야 예견적 위험분석 기반구축 연구	최석우
05	연안토지의 공공적 관리 강화를 통한 연안관리 효과 제고방안	윤성순
06	선박 배출저감기술의 국내 보급·활용 촉진을 위한 제도개선 방안 연구	안용성
07	해양레저관광서비스 산업 육성방안 연구	최일선
08	주요국의 해양유전자원 관리체계 강화에 대응한 정책방안 연구	박수진
09	하천·해양 수질의 연계·통합 관리 체계 연구	장원근
10	총허용어획량(TAC) 기반 수산자원관리 강화 방안 연구	이정삼
11	우리나라 수산업·어촌의 공익적 가치 평가 및 보존에 관한 연구	류정곤
12	우리나라 원양산업의 사회적책임 실천 강화를 위한 정책연구	정명화
13	수산업 노동시장 구조 분석과 노동정책 영향분석 연구	한광석
14	우리나라 정기선 해운정책 개선 방안 연구	김태일
15	국내외 해운금융 비교를 통한 국내 해운금융 역량강화 방안 연구	전형진
16	해상 사이버 보안 대책방안 연구	박한선
17	해운산업 재건을 위한 해운기업 역량강화방안 연구	윤희성
18	국내 항만산업이 지역경제에 미치는 영향 분석: 부산항 및 주요항만을 중심으로	하태영
19	동북아 공동배출규제해역 지정 필요성 및 추진방안 연구	이기열
20	수출기업의 글로벌 공급사슬협력 수준 분석과 물류정책 개선방안 연구	김은수
21	온도조절 컨테이너의 국제물류 분석과 대응방안 연구	신수용
22	IMO 규제기반 해사산업의 글로벌 지속발전방안 연구(3/3)	박한선
23	환동해권 경제개발 및 협력방안 연구(3/5)	최나영환

2018년

01	해양환경영향평가제도의 실효성 확보를 위한 개선 연구	박수진
02	공유수면매립 정책의 개선과 전환에 관한 연구	윤성순
03	도서지역 해양관광 발전전략 연구 : 관광행태 분석을 통한 수요 대응을 중심으로	홍장원
04	양식장 해양쓰레기 자원순환 방안 연구	김경신
05	수산부문 전망모형 「KMI-FOSiM」 구축 연구(1/2) - 양식산업 전망모형 구축을 중심으로 -	이현동
06	수산업 주요 연관산업의 글로벌 경쟁력에 관한 연구	임경희
07	인구소멸 시대의 어촌사회 정책 연구	박상우
08	양식어장 환경개선 방안에 관한 연구	마창모
09	어업여건 변화에 대응한 연근해어업의 허가제도 개선 방안	엄선희
10	과학적해양사고조사체계도입 및 구축기반 연구	박한선
11	컨테이너 해상물동량 예측 모형 연구 - 베이지안 방법론을 중심으로	고병욱
12	해운기업 지배구조 및 비즈니스 유형과 경영성과 관계 분석	황진희
13	건설용 모래 공급사슬관리(SCM) 방안 - 바다모래 가치 재산정을 중심으로	이연경
14	항만산업 종합통계 연구	하태영
15	항만 위험물 컨테이너 공급사슬 관리방안 연구	최나영환
16	국내 컨테이너항만의 비용함수 추정과 효율성 연구	최석우
17	환동해권 국내기업 진출사례 분석 및 개발 추진전략- 극동 러시아 지역 사례 중심으로-	이기열
18	IMO 규제기반 해사산업의 지속발전방안 연구 - 신규 해사산업 진흥을 위한 법제도 개선방안(2차년도) -	박한선

수시연구보고서 발간목록

2019년

01	일본의 수산물 수입 구조 분석과 對일본 수출 전략 연구	임경희
02	수산물 위생안전을 위한 저온유통체계 구축방안 연구	장홍석
03	사업체 총조사 MD를 활용한 2016,2017 어촌지역 고용통계	한광석
04	외래 해양부착생물에 대한 선제적 대응을 위한 정책방향 연구	박수진

2018년

01	선화주조선 상생발전을 위한 해운산업투자 확대방안 연구	윤재웅
02	낙시관리 실행력 제고 방안 연구	이정삼
03	국내 항만의 대기오염물질 관리정책 및 제도 개선방안	안용성
04	농·축산물 등의 관세행정 제도 개선을 통한 항만형 자유무역지역(FTZ) 활성화 방안	조지성
05	최저임금 상승이 항만연관산업에 미치는 영향 및 대응방안	이기열
06	수산업·어촌분야 사회적 경제 활성화 방안	류정곤
07	새로운 어선등록제 도입을 위한 제도 개선연구 방안	염선희
08	유엔 BBNJ 신 공해질서에 대응한 정책방안 연구	박수진

현안연구보고서 발간목록

2019년

01	해양관광 경쟁력 지수 개발 연구	최일선
02	고등어류 수급통계 개선방안	이남수
03	해수욕장의 사회악자 포용성 증진 방안	정지호
04	수산물품질관리사 제도 활성화 방안	황규환
05	항만지역 대기질 개선을 위한 데이터의 연계·활용방안 연구	안용성
06	한국산 수산물 글로벌 브랜드 가치 제고 방안 연구	한기욱
07	해양 플라스틱 대응을 위한 산업계 협력 방안 구상	이윤정
08	USMCA 수산분야 협정문 분석 및 시사점	안지은
09	해운 조세제도 개선 방안 연구	안영균
10	어촌뉴딜 300사업 성과평가 체계 구축	박상우
11	지역 해양수산 역량지표 개발을 위한 기초 연구	황재희
12	국동려 진출 우리 화주·물류 기업의 물류 애로사항 분석 및 개선 방안	박성준

2018년

01	미얀마 국립수산대학 설립 추진 방안	정명화
----	---------------------	-----

02	연안여객 해상교통의 대중교통체계 구축 방안 연구	김태일
03	해조류 국제양식규범확산에 따른국내 김산업 수용태세 분석- ASC 인증을 중심으로-	이상철
04	청년일자리 창출을 위한 해양수산 인력양성 방안 연구- 국제물류 청년인력을 중심으로 -	박광서
05	컨테이너 화물 해상운송 계약 개선방안 연구	윤재웅
06	청년 물류인력 해외진출 정주 지원사업 구상 연구	김은수
07	자율운항선박 도입 관련 대응정책 방향 연구	박혜리
08	연안토지매입 동향과 도입 가능성 검토	윤성순
09	강릉선 KTX 개통에 따른 해양관광 수용태세 개선방안 연구	최일선
10	선박관리산업 육성을 통한 청년 일자리 창출 방안 연구	최영석
11	시민 참여형 연안·해양정보 생산·활용방안	정지호
12	부산항 배후지 부가가치활동 확대를 통한 일자리 창출 연구	하태영
13	해양바이오 산업화를 위한 국가연구개발사업 추진 방향 연구	한기원
14	근로시간 단축제도 시행에 따른 해양수산업 고용영향 분석	장정인
15	어린물고기 남획실태 및 보호정책 연구	이정삼
16	남북관계 개선에 따른 항만물류 부문의 협력사업 발굴	이기열
17	한국 제조기업의 SCM 역량 강화 방안에 관한 연구	신수용
18	제3자 물류산업 발전방향 연구 - 대기업 물류자회사의 역할을 중심으로	고병욱
19	신남방정책 이행 위한 우리나라-베트남 수산분야 개발협력 방안 연구	정명화
20	신 해양도시 조성 필요성 연구	최지연
21	지역 해양수산 혁신사업 발굴과 추진방향	최지연
22	학교급식의 수산물 이용 활성화 방안	문석란
23	주민주도의 에너지 자립형 어촌마을 모델 연구	박상우
24	페루 수산업 실태 및 수산 협력 방향	엄선희
25	경제 총 조사 마이크로 데이터를이용한 어촌지역 고용지표 통계생산	한광석
26	서해평화수역 조성을 위한 정책방향 연구	남정호
27	해양치유산업 육성을 통한 해양휴양·복지서비스 도입방안	홍장원
28	어촌지역재생을 통한 청년일자리 창출방안 연구	박상우
29	해상풍력클러스터 조성을 위한 항만 및 배후단지 활용방안 연구	이종필
30	IMO 배출가스 규제 강화에 대비한 국내 해운산업 대응 전략	이호춘
31	외래병해충에 대한 항만분야 방제체계 개선방안	최나영환
32	수산물 생산·유통분야 블록체인 기술의 도입가능성과 도입방안	고동훈

일반연구보고서 발간목록

2019년

01	양식수산물의 경영비 조사체계 구축에 관한 연구	백은영
----	---------------------------	-----

2018년

01	전라남도 작은 SOC사업 활성화 방안	김근섭
02	2019 국내 물류기술 수요조사 및 분석	최상희
03	러시아 Arctic LNG-2 사업 참여 방안 연구	김민수
04	물류기술 R&D 추진성과 분석 연구	강무홍
05	우리나라 기업의 일대일로 물류분야 협력방안 연구	김형근
06	북한 동해 해양관광 활성화 방안	윤인주
07	해양수산 분야 복극권 4차 산업기술 수요조사 및 분석 연구	김민수
08	경상북도 논 생태양식 육성 방안	이상철
09	남북관계 변화에 따른 해운협력 기반 구축 방안	황진희
10	영토 권원 이론의 현대적 발전과 한계	김원희
11	동해표기 문제 연구	현대송
12	우리나라 기업의 창장물류 활용방안 연구	김형근
13	태국의 수산물 소비트렌드 분석 및 수출 전략연구	임경희
14	제주 해상풍력발전지구 제도 개선 연구	육근형
15	해양수산 분야 9브릿지 구축방안 연구	박성준
16	강원도 어촌 신활력 제고방안	박상우

딥러닝 기반의 건화물선 시황예측 연구

인 쇄 | 2019년 11월 5일 인쇄

발 행 | 2019년 11월 7일 발행

발 행 인 | 장 영 태

발 행 처 | 한국해양수산개발원

49111 부산시 영도구 해양로 301번길 26(동삼동)

연 락 처 | 051-797-4800 (FAX 051-797-4810)

등 록 | 1984년 8월 6일 제313-1984-1호

조판·인쇄 | 애드원플러스

판매 및 보급: 정부간행물판매센터 Tel : 394-0337
정가 6,000원

