The Journal of the Convergence on Culture Technology (JCCT)
Vol. 8, No. 5, pp.697-703, September 30, 2022. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2022.8.5.697
JCCT 2022-9-87

해외선물 스캘핑을 위한 강화학습 알고리즘의 성능 비교

Performance Comparison of Reinforcement Learning Algorithms for Futures Scalping

정득교*, 이세훈**, 강재모***

Deuk-Kyo Jung*, Se-Hun Lee, Jae-Mo Kang*****

요 약 최근 Covid-19 및 불안한 국제정세로 인한 경기 침체로 많은 투자자들이 투자의 한 수단으로써 파생상품시장을 선택하고 있다. 하지만 파생상품시장은 주식시장에 비해 큰 위험성을 가지고 있으며, 시장 참여자들의 시장에 대한 연구 역시 부족한 실정이다. 최근 인공지능 분야의 발달로 파생상품시장에서도 기계학습이 많이 활용되고 있다. 본 논문은 해외선물에 분 단위로 거래하는 스캘핑 거래의 분석을 위해 기계학습 기법 중 하나인 강화학습을 적용하였다. 데이터 세트는 증권사에서 거래되는 해외선물 상품들 중 4개 상품을 선정해, 6개월간 1분봉 및 3분봉 데이터의 종가, 이동평균선 및 볼린저 밴드 지표들을 이용한 21개의 속성으로 구성하였다. 실험에는 DNN 인공신경망 모델과 강화학습 알고리즘인 DQN(Deep Q-Network), A2C(Advantage Actor Critic), A3C(Asynchronous A2C)를 사용하고, 학습 데이터 세트와 테스트 데이터 세트를 통해 학습 및 검증 하였다. 에이전트는 스캘핑을 위해 매수, 매도 중 하나의 행동을 선택하며, 행동 결과에 따른 포트폴리오 가치의 비율을 보상으로 한다. 실험 결과 에너지 섹터 상품(Heating Oil 및 Crude Oil)이 지수 섹터 상품(Mini Russell 2000 및 Hang Seng Index)에 비해 상대적으로 높은 누적 수익을 보여 주었다.

주요어 : 강화학습, DQN, A2C, A3C, 해외선물, 스캘핑

Abstract Due to the recent economic downturn caused by Covid-19 and the unstable international situation, many investors are choosing the derivatives market as a means of investment. However, the derivatives market has a greater risk than the stock market, and research on the market of market participants is insufficient. Recently, with the development of artificial intelligence, machine learning has been widely used in the derivatives market. In this paper, reinforcement learning, one of the machine learning techniques, is applied to analyze the scalping technique that trades futures in minutes. The data set consists of 21 attributes using the closing price, moving average line, and Bollinger band indicators of 1 minute and 3 minute data for 6 months by selecting 4 products among futures products traded at trading firm. In the experiment, DNN artificial neural network model and three reinforcement learning algorithms, namely, DQN (Deep Q-Network), A2C (Advantage Actor Critic), and A3C (Asynchronous A2C) were used, and they were trained and verified through learning data set and test data set. For scalping, the agent chooses one of the actions of buying and selling, and the ratio of the portfolio value according to the action result is rewarded. Experiment results show that the energy sector products such as Heating Oil and Crude Oil yield relatively high cumulative returns compared to the index sector products such as Mini Russell 2000 and Hang Seng Index.

Key words : Reinforcement learning, DQN, A2C, A3C, Futures, Scalping

*준회원, 경북대학교 인공지능학과 석사과정 (제1저자)
**준회원, 경북대학교 인공지능학과 석사과정 (참여저자)
***정회원, 경북대학교 인공지능학과 조교수 (교신저자)
접수일: 2022년 8월 26일, 수정완료일: 2022년 9월 3일
게재확정일: 2022년 9월 9일

Received: August 26, 2022 / Revised: September 3, 2022
Accepted: September 9, 2022
***Corresponding Author: jmkang@knu.ac.kr
Dept. of A.I, Kyungpook National Univ, Korea

Ⅰ. 서 론

투자는 미래에 기대되는 추가 보상을 얻기 위해 자본을 쓰거나 시간을 쏟는 것을 말하며, 수입을 늘림으로써 삶의 질을 향상시킬 수 있는 수단이 된다. 하지만 수익과 비례해 위험 또한 동반되므로 성공적 투자를 위해 위험관리는 반드시 필요하다. 이를 위해 리스크 헷지(Risk Hedge) 목적으로 파생상품이 많이 활용되고 있다. 뿐만 아니라 파생상품은 투자 수단으로써도 개인에게 활발히 활용되고 있다. 하지만 주식과 다른 고 레버리지(High leverage) 상품이며, 주식시장과 달리 상승 및 하락장 모두 투자가 가능해 투자자들의 손실 역시 크게 증가하고 있다[1].

이에 따라, 시장 참여자들은 손실을 최소화하고 수익을 극대화하기 위해 가격 예측을 통한 거래전략 개발에 힘쓰고 있다[2]. 또한, 투자 상품의 분석을 위해 기본적 분석(fundamental analysis)과 기술적 분석(technical analysis)을 활용한다. 기본적 분석은 가격에 영향을 주는 경제적 요인이나 흐름을 분석하여 가격을 예측하는 방법이며[3], 기술적 분석은 가격, 거래량 등 과거와 현재의 시장변수를 분석하여 미래의 가격을 예측하는 방법으로 다우이론(the Dow theory), 엘리어트 파동이론(the Elliott Wave), 이동평균선(Moving average) 등의 방법들이 연구되고 있다[4]. 하지만 실시간으로 시장이 전달하는 많은 정보를 사람이 직접 분석하기에는 한계가 존재한다. 이런 한계를 극복하기 위해 컴퓨터를 이용한 알고리즘 트레이딩이 개발되었고, 덕분에 상대적으로 쉽고 편리하게 많은 정보의 활용이 가능하게 되었다[5].

한편 사람의 뇌 구조를 모방한 인공신경망 및 강화학습의 등장으로 알고리즘 트레이딩에 이를 활용한 연구들도 활발히 진행되고 있다. [6]은 DD-DQN(Dueling Double-DQN) 기반의 제안모델과 DD-DQN, PPO, A2C 총 4개의 알고리즘과 비트코인의 과거 4개년 일별 데이터로 각 에이전트를 학습시켜 샤프지수(위험자산에 투자하여 얻는 초과수익의 정도)를 비교하였고, 제안한 모델이 평균적으로 12.84%의 수익률과 1.41 샤프지수를 예측하여, DD-DQN, PPO, A2C 알고리즘 성능보다 뛰어남을 실험하였다. [7]은 LSTM 신경망에 A3C 알고리즘을 적용하여 증권사에서 실시간으로 전송되는 주가 정보를 통해 에이전트의 매수, 매도, 관망 3가지 행동에 대해 학습시키고 투자자들이 웹을 통해서 에이전트의 예측을 확인하고 추가 투자정보까지 제공할 수 있는 시스템을 제안하였다. [8]은 뉴스 헤드라인 분석을 위한 RCNN(Recurrent Convolutional Neural Network)모델과 학습을 위한 DDPG 강화학습 알고리즘을 통해 주식 트레이딩 봇을 구현하였다.

이처럼 복잡하고 변동성이 높은 금융 시장에서는 에이전트를 통해 환경을 관찰하고, 수익이라는 보상을 통해 스스로 학습해 나가는 강화학습이 많이 활용되고 있다. 이외에도, 파생상품시장의 변동성 및 개인의 자본규모를 고려해 볼 때, 수 분 내로 거래가 완료되는 스캘핑(scalping)은 다양한 거래전략을 시도할 수 있어 개인 투자자들에게 주로 사용되는 방식이지만, 저자가 아는 한 인공지능 기술이 적용된 사례가 없다. 따라서 본 논문에서는 강화학습을 이용한 모델을 구성하고 그에 따른 스캘핑 성능을 실험해 보기로 한다.

논문의 구성은 2장에서는 강화학습을 이용한 선물시장의 기존 연구에 대해 알아보고 3장에서는 선물시장에서의 수익률 최대화를 위해 강화학습 알고리즘을 사용한 연구 모델을 소개한다. 4장에서는 3장에서 소개한 모델의 실험 결과를 보여주고, 결과를 분석하며 5장에서는 연구 모델에 대한 평가 및 향후 과제에 대해 서술한다.

Ⅱ. 선행 연구

강화학습이 위험을 관리하고 수익을 얻는 트레이딩 시스템의 훈련 방법으로 제안된 이후, 선물시장에서의 강화학습 연구 역시 활발히 이어지고 있다[9].

[10]은 S&P500, NASDAQ, Hang Seng Index와 같은 지수 상품의 시계열 데이터를 선형과 비선형으로 구분하여 선형은 ARIMA, 비선형은 RNN모델로 학습하는 하이브리드 모델을 구성하였다. 실험 결과는 시계열 예측 지표인 MSE, MAE, MAPE를 통해 비교하였고, 모든 종목에서 ARIMA, RNN을 단일 모델로 사용할 경우보다 뛰어난 결과를 보였다. 그러나 ARIMA와 RNN 모델은 각각 단기 예측에 적합한 모델로 장기 예측에 적합한 LSTM을 이용한 추가 연구가 필요함을 확인하였다.

또 다른 연구는 지수, 채권, 환율, 에너지 등의 섹터 내 25개의 상품을 정하여 DQN만을 활용해 학습한다.

25개 상품의 수익률을 정규화해 입력으로 사용하며, 5년간의 데이터, GBM(Geometric Brownian Motion), VG(Variance Gamma)로 상태를 확장하여 S&P500 지수와의 수익률을 비교한다. 실험의 결과를 통해 훈련 데이터의 생성 방식 및 훈련 데이터의 부족으로 인한 과적합의 중요성에 대해 확인한다[11]. 하지만 연구의 성능을 높이기위해서는 DQN 외에 정책 기반의 알고리즘인 PPO, A2C와 같은 모델을 추가하여 실험이 필요할 것으로 보인다. [12]는 선물시장 상품들의 1일 및 5일간의 수익률을 입력 데이터로 활용해 DDQN(Double Deep Q-Network) 알고리즘을 사용하였다. 성능 비교를 위해 4개의 모델을 구성해 각 모델 별로 입력 데이터를 달리하였고, 모델 1에서 시장의 수익률을 상회하는 결과를 얻었다. 연구 결과에서는 좀 더 복잡한 입력 데이터를 사용하는 모델의 성능이 저조하다고 나타났고, 일부 모델에서는 과적합이 되는 경향을 보였다.

III. 연구 방법

1. 데이터 세트(Data Set)

본 논문에서 사용된 해외선물 데이터는 키움증권 해외선물 HTS(Home Trading System)의 차트 데이터를 이용하여 수집(가격정보를 엑셀로 Export)하였다. 또한 많은 데이터 세트를 확보하여 학습을 용이하기 위해 1분봉과 3분봉 데이터를 각각 활용하였으며, A3C 알고리즘은 1분봉과 3분봉 데이터를 병렬 학습시켰다. 이런 설정으로 상품, 알고리즘과 성능의 연관성을 보고자 하였다. 학습 데이터는 2021. 1. 1. ~ 6.30 의 1분봉 및 3분봉 데이터를 사용하였고, 테스트 데이터는 2022. 1. 1 ~ 6.30까지의 1분봉 및 3분봉 데이터로 선정하였다. 수집된 데이터 정보는 날짜, 시간, 시가, 고가, 저가, 종가, 거래량이다.

주식, 파생상품 등의 시계열 데이터는 큰 비정상성(Non stationary)으로 인해 강화학습 모델을 통한 학습에 어려움이 있다. 그러므로 명확한 알고리즘별 성능 비교를 위해 적합한 상품 선정 및 데이터 전처리가 필요하다.

첫 번째로 표 1과 같이 현재 키움 증권에서 거래 가능한 전체 해외선물 상품(약 90여개) 중 2021년의 1분봉 차트 종가 데이터(평균 약 25만건)로 각 상품의 통계량 정보를 구하였다. 이후 에너지, 지수 섹터에서 분산, 왜도 등에서 갭이 있는 상품들을 아래와 같이 각각 선정하였다.

- Heating Oil(에너지)
- Crude Oil(에너지)
- Hang Seng Index(지수)
- Mini Russell 2000(지수)

표 1. 선물 상품 통계정보
Table 1. Derivatives Statistical Information

섹터	상품명	평균	분산	표준편차
에너지	Heating Oil	2.06	0.07	0.26
	Gasoline	2.05	0.08	0.28
	Natural Gas	3.76	1.10	1.05
	Brent Crude Oil	70.90	63.04	7.94
	Crude Oil	67.98	67.93	8.24
지수	Mini Russell 2000	2240.30	5533.48	74.39
	Mini S&P 500	4263.96	82660.8	287.51
	Mini NASDAQ	14474.08	1300868	1140.56
	Hang Seng Index	27051.66	4177394	2043.87

두 번째로 가격은 '달러', 거래량은 '계약'과 같이 항목별 사용하는 단위가 다르고 가격은 Hang Seng의 경우 Crude Oil 상품에 비해 높은 값을 이용한다. 이런 경우 계산이 복잡해지기 때문에 수집된 데이터를 각 비율값으로 계산하는 전처리 과정을 표 2와 같이 거쳤다. 또한 기술적 분석 중 이동평균선과 볼린저 밴드[13]를 지표로 활용하여 전처리 후 총 21개의 데이터를 입력으로 사용한다.

표 2. 입력 데이터 전처리
Table 2. Input data preprocessing

내 용	수 식	설 명
종가 비율	$\frac{D_{price} - D_{close}}{D_{close}}$	종가와 전일 종가, 시가, 저가, 고가를 이용한 비율
이동평균선	$\frac{D_{close} - D_{MA}}{D_{MA}}$	종가와 5, 20, 60, 120, 240일의 이동평균선을 이용한 비율
볼린저밴드	$\frac{D_{close} - D_{BB}}{D_{BB}}$	종가와 볼린저밴드(12,2), (20,2)를 이용한 비율

2. 강화학습 모델

알고리즘별 성능 비교를 위해 그림 1과 같이 DNN 모델을 이용해 DQN, A2C, A3C 알고리즘을 각각 적용하고, 성능 최적화를 위해 10회 이상의 실험을 거쳐 하이퍼 파라미터(Hyper-parameter)의 값을 튜닝 하였고, 그 값은 아래와 같이 모든 모델에 동일하게 설정하였다.

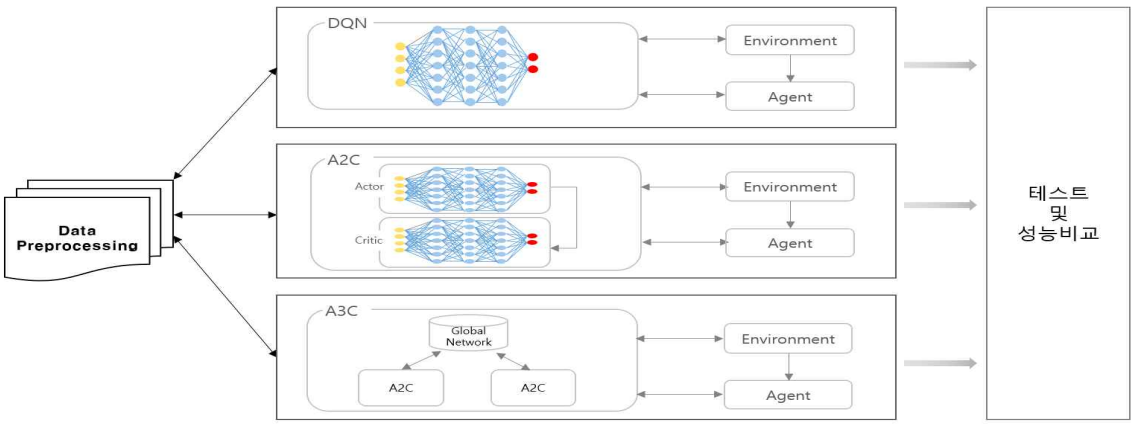

그림 1. 강화학습 모델
Figure 1. Reinforcement learning Model

- 학습률(Learning rate) : 0.0001
- 에포크(Epoch) : 100회
- 할인율(Discount factor) : 0.9
- 은닉층 활성함수 : Relu(Rectified linear unit)
- 출력층 활성함수 : sigmoid
- 손실함수 : 평균제곱오차(Mean Square Error)
- 옵티마이저 : Adam(Adaptive Moment Estimation)

1) DQN

Q-learning은 테이블과 같이 유한한 상태를 가지는 환경에서는 문제 해결에 뛰어나지만 현실세계에서는 대부분의 경우 무수히 많은 상태를 가지는 문제들이 발생되기 때문에 해결에 한계가 있다. 따라서 이를 인공신경망을 통해 해결한 것이 DQN[14] 이다.

$$Q(s,a,\theta) = R_s^a + \gamma \max Q(s',a',\theta') \quad (1)$$

식(1)은 다음 상태(s')에서 선택할 수 있는 행동들 중 행동가치함수 Q를 최대화하는 행동(a')과 인공 신경망의 가중치를 나타내는 파라미터θ'로 함수 근사를 통해 Q의 가치를 계산한다.

2) AC, A2C, A3C

AC 알고리즘은 행동을 선택하는 인공신경망 Actor와 선택된 행동의 가치를 평가하는 인공신경망 Critic으로 구성된다[15]. 그리고 에이전트의 학습 과정에서 정책π와 행동가치함수Q를 모두 학습한다.

A2C 알고리즘[16]은 AC알고리즘과 유사하나 차이점으로는 Critic의 비용함수를 식(2)와 같이 Actor의 결괏값인 행동가치함수 Q의 값에서 상태가치함수를 Baseline으로 채택하여 이를 행동가치함수로부터 뺀 값인 Advantage를 사용한다.

$$A(s,a) = Q(s,a) - V(s) \quad (2)$$

A3C 알고리즘[16]은 비동기적 A2C 알고리즘으로, 여러 개의 A2C 에이전트를 병렬로 학습하는데 사용하는 방법이다. 병렬적으로 학습을 진행하면 Replay Buffer를 대체하여 데이터의 상관관계를 없애고 분산을 낮춰주는 효과가 있다.

3. 에이전트

환경으로부터 상태를 관찰하는 에이전트는 인공신경망에 상태를 입력하고 출력된 값으로 행동을 결정한다. 본 논문에서는 스캘핑의 목적을 달성하기 위해 Output Layer에서 매수와 매도 2가지 상태 값만을 출력한다. 다른 연구[17-18]와 달리 보류(Hold)에 대한 출력을 없앰으로써 빈번한 거래가 발생할 수 있게 한 것이다.

또한, 거래를 위한 초기 자본금은 50,000 USD로 정의하고, 에이전트 행동에 대한 보상은 초기의 자본금으로 현재의 PV(Portfolio value)를 나눈 수익률을 보상으로 한다. 강화학습 모델은 epoch가 반복될수록 학습 데이터의 매수, 매도 행동을 수정하면서 수익률을 최대화시키는 방향으로 인공신경망을 업데이트한다.

Ⅳ. 실험 및 결과

본 논문에서 제안한 해외선물 상품과 모델을 이용해 누적 수익을 최대화하는 성능 비교를 실험하였다. 실험 결과는 표 3과 같이 각각 1분봉, 3분봉으로 학습 및 테스트한 결과로써, Heating Oil과 Crude Oil은 모든 실험에서 누적 수익이 발생했고 Hang Seng은 A3C에서 모두 손실로 나타났다.

상품의 통계 정보 중 분산과 편차가 가장 적었던 Heating Oil은 그림2와 같이 1분봉에서 알고리즘별 수익 차이가 많이 발생했고. A3C의 누적 수익이 가장 높았고, A2C의 누적 수익이 가장 낮게 나타났다. 3분봉에서는 DQN과 A2C의 성능이 높게 나타났다.

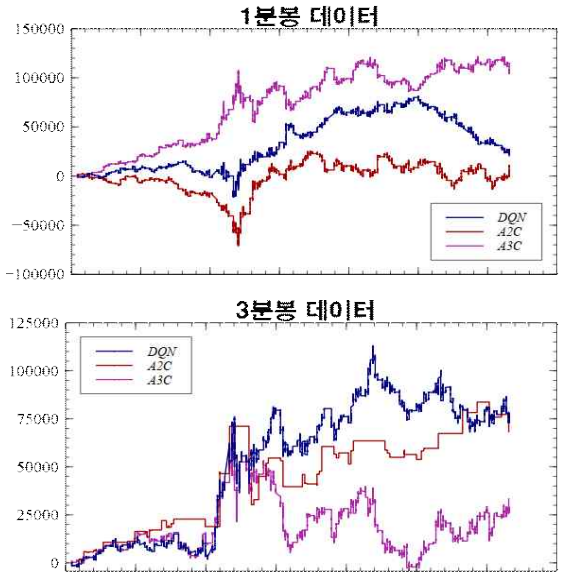

그림 2. Heating Oil 테스트 결과
Figure 2. Results of testing for Heating Oil

Crude Oil은 그림 3과 같이 1분봉에서 모두 높은 누적 수익을 나타냈지만 구간 별로 급등락 및 횡보 등의 변동성이 나타났다. 3분봉에서는 DQN이 가장 높은 누적 수익 나타냈지만 급락으로 인해 손실이 가장 크게 발생한 구간 역시 발생했다.

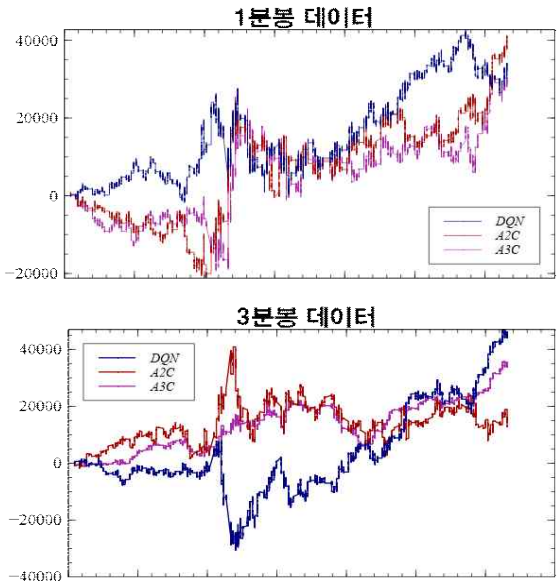

그림 3. Crude Oil 테스트 결과
Figure 3. Results of testing for Crude Oil

그림 4의 Mini Russell 2000 경우 A3C에서 모두 높은 누적 수익을 냈고, 3분봉에서 DQN이 수익으로 전환하지 못하고 손실 상태를 유지한 채로 테스트가 종료되었다.

지수 상품 중 가장 분산과 표준편차가 컸던 그림 5의 Hang Seng Index 상품은 DQN의 수익이 가장 높게 나타났지만 다른 상품과 비교해 누적 수익이 커지지 않고

표 3. 입력 데이터 전처리
Table 3. Input data preprocessing

내 용		Heating Oil			Crude Oil			MIini Russell 2000			Hang Seng Index		
		DQN	A2C	A3C	DQN	A2C	A3C	DQN	A2C	A3C	DQN	A2C	A3C
1분봉	Buy Signal	58572	42067	101102	87741	100837	82224	85858	111144	3812	56998	25941	98234
	Sell Signal	57992	74497	15462	87939	74843	93456	85531	60245	167577	56665	87722	15429
	청산횟수	38873	32587	12527	58681	43741	54718	57167	42935	2764	37984	22176	12347
	수익률*	41.1	18.6	**209**	62.2	**76.7**	59.2	30.8	18.3	**44.4**	**18.5**	11.7	-38
3분봉	Buy Signal	39313	48238	38306	28634	55374	26381	29084	777	4691	133	481	33101
	Sell Signal	9009	84	10016	30065	3325	32318	29389	57696	53782	38112	37764	5144
	청산횟수	7757	84	5568	19543	3022	15941	19439	482	2782	133	423	4101
	수익률*	**150**	136	66.6	**89**	26.1	69.3	-16.2	35.9	**53.4**	21.6	**39.2**	-22.3

* 누적수익 / 초기자본금(단위 : %)

박스권을 형성하였다. 또한 A3C의 성능이 1분봉과 3분봉 모두 저조하게 나타났다.

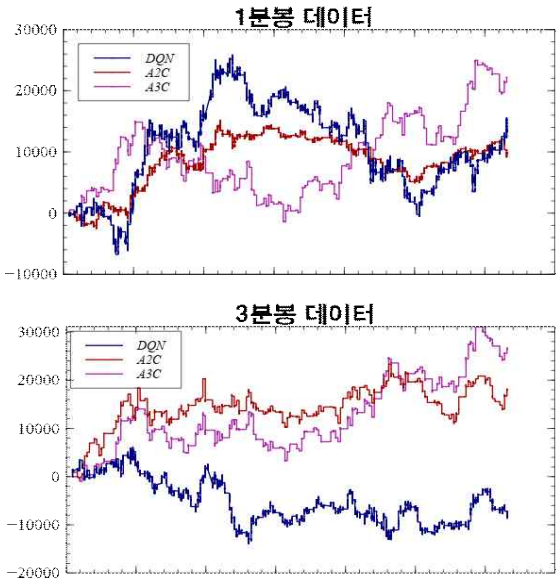

그림 4. Mini Russell 2000 테스트 결과
Figure 4. Results of testing for MIni Russell 2000

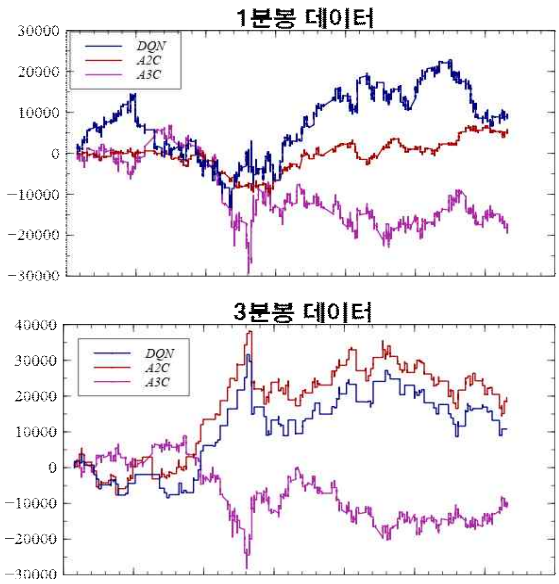

그림 5. Hang Seng Index 테스트 결과
Figure 5. Results of testing for Hang Seng Index

전체적인 실험 결과 각 상품에서 1분봉과 3분봉간의 가시적 유사성이 나타나지 않았고, 좀 더 많은 학습 데이터로 패턴을 학습한 A3C의 성능이 좋을 것으로 예상했으나, 유의미한 결과가 도출되지 않았다. 알고리즘별 성능을 살펴보면 DQN은 Russell 3분봉을 제외한 실험에서 모두 수익을 보였고, A2C는 최종 수익률의 편차는 가장 작지만 모든 실험에서 수익을 거뒀다. A3C는 실험 중 가장 높은 수익률인 209%를 나타내는 등 누적 수익이 발생한 상품에선 좋은 성능을 보였지만, Hang Seng의 경우 1분봉, 3분봉 모두 손실이 발생했다.

V. 결 론

본 논문에서는 차트 데이터와 이동평균선과 볼린저 밴드 지표로 전처리 과정을 거치고 DQN, A2C, A3C를 이용해 해외선물 중 일부 상품의 스캘핑 성능을 비교하였다. 4장의 실험 결과처럼 에너지 섹터 상품의 경우 높은 누적 수익을 보였고, 지수 섹터 상품의 경우 누적 수익이 크게 높지 않고, 특정 알고리즘들은 지속적으로 손실을 나타내는 경우도 있었다. 향후 에너지 섹터의 상품들에서는 DQN과 보조지표들을 추가한 모델을 통해 누적 수익의 극대화를 예상할 수 있다. 또한 A2C 알고리즘은 모든 상품에서 누적 수익(최소 11.7%, 최대 136%)을 보였으므로, 다양한 환경에서 추가 검증을 거친다면 시장에서의 활용 가능성이 높을 것으로 보인다.

하지만, 단기간의 데이터를 활용해 학습을 진행함으로써, 잡음(Noise)으로 인한 과적합(over-fitting)[19] 등으로 모델의 정확도와 신뢰성을 떨어트릴 수 있다. 추후 연구에서는 이런 잡음을 제거할 수 있는 차분, 로그 변환등의 통계적 방법을 통한 전처리 과정을 추가하고, PPO(Proximal Policy Optimization), DDPG(Deep Deterministic Policy Gradient)등 최신의 강화학습 알고리즘을 활용하여 누적 수익의 극대화를 도모할 것이다. 또한 수수료, 세금의 문제를 추가하여 실제 환경에서 스캘핑이 가능한 실제 트레이딩 모델을 구축하는 것을 목표로 할 것이다.

References

[1] “‘22 Capital market risk analysis report”, the Financial Supervisory Service, pp. 82-87, 2022.
[2] Zhou, Feng, et al. “EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction.” Expert Systems with Applications, Vol 115, pp. 136-151, 2019. DOI:https://doi.org/10.1016/j.eswa.2018.07.065
[3] Mukherji, Sandip and Dhatt, Manjeet S and Kim,

Yong H "A fundamental analysis of Korean stock returns." Financial Analysts Journal, Vol 53, No. 3, pp. 75-80, 1997. DOI:https://doi.org/10.2469/faj.v53.n3.2086

[4] Achelis, S. B, "Technical Analysis from A to Z.",McGraw Hill, 2001.

[5] Orlando, J. M. "Algorithmic presentation to european central bank-BNP Paribas.", 2007.

[6] El Akraoui, Bouchra, and Cherki Daoui. "Deep Reinforcement Learning for Bitcoin Trading." International Conference on Business Intelligence. Springer, Cham, pp. 82-93, 2022.DOI: 10.1007/978-3-031-06458-6_7

[7] J.Y. Park, S.S. Hong, MG.. Park, H. Lee, "An Implementation of Stock Investment Service based on Reinforcement Learning", The Journal of the Convergence on Culture Technology (JCCT), Vol. 7. No. 4, pp. 807-814. November 2021, DOI:https://doi.org/10.17703/JCCT.2021.7.4.807

[8] Azhikodan, Akhil Raj, Anvitha GK Bhat, and Mamatha V. Jadhav. "Stock trading bot using deep reinforcement learning." Innovations in Computer Science and Engineering, pp. 41-49, 2019. DOI: 10.1007/978-981-10-8201-6_5

[9] Moody, John, et al. "Performance functions and reinforcement learning for trading systems and portfolios.", Journal of Forecasting, Vol. 17, No. 5-6, pp. 441-470, 1998. DOI:https://doi.org/10.1002/(SICI)1099-131X(1998090)17:5/6<441::AID-FOR707>3.0.CO;2-%23

[10]Shui-Ling, Y. U., and Zhe Li. "Stock price prediction based on ARIMA-RNN combined model." 4th International Conference on Social Science (ICSS 2017), pp 1-6, 2017.

[11]Hirsa, Ali, et al. "Deep reinforcement learning on a multi-asset environment for trading." arXiv preprint arXiv:2106.08437, 2021, DOI:https://doi.org/10.48550/arXiv.2106.08437

[12]Zejnullahu, Frensi, Maurice Moser, and Joerg Osterrieder. "Applications of Reinforcement Learning in Finance--Trading with a Double Deep Q-Network." arXiv preprint arXiv:2206.14267, 2022. DOI:https://doi.org/10.48550/arXiv.2206.14267

[13]Bollinger, John. "Using bollinger bands." Stocks & Commodities, Vol. 10, No. 2, pp. 47-51, 1992.

[14]Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602, 2013. DOI:https://doi.org/10.48550/arXiv.1312.5602

[15]Konda, Vijay, and John Tsitsiklis. "Actor-critic algorithms." Advances in neural information processing systems, Vol. 12, 1999.

[16]MNIH, Volodymyr, et al. "Asynchronous methods for deep reinforcement learning." In: International conference on machine learning. PMLR, pp. 1928-1937, 2016.

[17]Xiong, Zhuoran, et al. "Practical deep reinforcement learning approach for stock trading". arXiv preprint arXiv:1811.07522, 2018. DOI:https://doi.org/10.48550/arXiv.1811.07522

[18]Liu, Xiao-Yang, et al. "FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance." arXiv preprint arXiv:2011.09607, 2020. DOI:https://doi.org/10.48550/arXiv.2011.09607

[19]Kim, Sung-Hyeock, et al. "Influence on overfitting and reliability due to change in training data". International Journal of Advanced Culture Technology(IJACT), Vol 5. No. 2, pp 82-89, June 2017, DOI: https://doi.org/10.17703/IJACT.2017.5.2.82

※ 본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다(No. 20224000000150).

※ 본 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2020R1I1A3073651)의 지원을 받아 작성되었음.