Journal of the Korea Academia-Industrial cooperation Society
Vol. 22, No. 1 pp. 123-130, 2021

https://doi.org/10.5762/KAIS.2021.22.1.123
ISSN 1975-4701 / eISSN 2288-4688

강화학습을 이용한 트레이딩 전략

조현민, 신현준*
상명대학교 경영공학과

Trading Strategies Using Reinforcement Learning

Hyunmin Cho, Hyun Joon Shin*
Department of Management Engineering, Sangmyung University

요 약 최근 컴퓨터 기술이 발전하면서 기계학습 분야에 관한 관심이 높아지고 있고 다양한 분야에 기계학습 이론을 적용하는 사례가 크게 증가하고 있다. 특히 금융 분야에서는 금융 상품의 미래 가치를 예측하는 것이 난제인데 80년대부터 지금까지 기술적 및 기본적 분석에 의존하고 있다. 기계학습을 이용한 미래 가치 예측 모형들은 다양한 잠재적 시장 변수에 대응하기 위한 모형 설계가 무엇보다 중요하다. 따라서 본 논문은 기계학습의 하나인 강화학습 모형을 이용해 KOSPI 시장에 상장되어 있는 개별 종목들의 주가 움직임을 정량적으로 판단하여 이를 주식매매 전략에 적용한다. 강화학습 모형은 2013년 구글 딥마인드에서 제안한 DQN와 A2C 알고리즘을 이용하여 KOSPI에 상장된 14개 업종별 종목들의 과거 약 13년 동안의 시계열 주가에 기반한 데이터세트를 각각 입력 및 테스트 데이터로 사용한다. 데이터세트는 8개의 주가 관련 속성들과 시장을 대표하는 2개의 속성으로 구성하였고 취할 수 있는 행동은 매입, 매도, 유지 중 하나이다. 실험 결과 매매전략의 평균 연 환산수익률 측면에서 DQN과 A2C이 대안 알고리즘들보다 우수하였다.

Abstract With the recent developments in computer technology, there has been an increasing interest in the field of machine learning. This also has led to a significant increase in real business cases of machine learning theory in various sectors. In finance, it has been a major challenge to predict the future value of financial products. Since the 1980s, the finance industry has relied on technical and fundamental analysis for this prediction. For future value prediction models using machine learning, model design is of paramount importance to respond to market variables. Therefore, this paper quantitatively predicts the stock price movements of individual stocks listed on the KOSPI market using machine learning techniques; specifically, the reinforcement learning model. The DQN and A2C algorithms proposed by Google Deep Mind in 2013 are used for the reinforcement learning and they are applied to the stock trading strategies. In addition, through experiments, an input value to increase the cumulative profit is selected and its superiority is verified by comparison with comparative algorithms.

Keywords : Machine Learning, Reinforcement Learning, Trading System, Deep Q-Network, Actor-Critic

1. 연구배경

금융 상품 투자에 있어서 일반적인 투자자들은 크게 기술적 및 기본적 분석을 통한 투자와 그렇지 않은 투자로 나눠진다. 전자의 경우에는 차트(chart) 및 재무제표 등을 통한 가치평가(valuation)를 의미하고, 후자의 경우에는 뇌동매매와 같이 즉흥적인 투자를 의미한다. 기술적 분석(technical analysis)은 주로 차트 분석을 기본으로 하고 과거 주가 및 거래량 등의 자료를 분석하여 미래 주가의 방향성을 예측하는 기법이다. 기본적 분석

본 연구는 2019학년도 상명대학교 교내연구비를 지원받아 수행하였음.
*Corresponding Author : Hyun Joon Shin(Sangmyung Univ.)
email: hjshin@smu.ac.kr
Received October 15, 2020 Revised November 9, 2020
Accepted January 8, 2021 Published January 31, 2021

(fundamental analysis)은 재무제표를 분석하여 기업의 내재적 가치를 산출하고 미래 수익성을 예측한다. 기술적 분석은 1981년 Banz가 시가총액의 추세를 분석하여 투자하는 방법론을 발표한 이후 다양한 기술적 지표를 이용한 투자 방식이 연구되었다[1]. DeBondt and Thaler[2]은 개별 종목의 차트를 분석한 결과 일 별 차트 상 저점에서 거래량이 증가하면서 주가가 상승하면 특정 기간 동안 지속적으로 상승 모멘텀(momentum)을 유지한다는 것을 알 수 있었다. 기술적 분석에 관한 연구는 주로 차트 상 이동평균선(moving average)을 이용한 투자 전략에 관한 연구가 활발하게 진행되었고 지금까지도 향후 주가 방향성을 예측하는데 사용되는 대표적 기술 지표이다[3-4].

반면 기본적 분석은 1939년에 벤자민 그레이엄(Benjamin Graham)의 재무 자료를 이용한 기업 가치 분석법이 시초이며 이후 많은 관련 연구들이 진행되고 있다[5]. Fama and French는 기업의 개정 항목을 바탕으로 다요인 모형(multi factor model)을 통해 동종 기업보다 저평가된 기업들을 발굴하는 연구를 하였다[6]. 또한, 기본적 분석에 필요한 요인들을 자료포락분석(data envelopment analysis)을 통해 정량적으로 기업의 가치를 산출하는 연구도 진행되었는데 이는 특정 요인들을 정성적으로 판단하는 것보다 수익률 측면에서 우수했다[7].

이처럼 기술적 및 기본적 분석은 투자 전략을 수립하기 위해 대중적으로 활용되는 분석 기법이다. 그러나 이러한 투자 방식은 예측하기 힘든 다양한 시장 변수가 발생하면 정성적 판단을 하면서 기존의 투자 원칙을 상실하고 잘못된 오판으로 인해 큰 손실이 발생한다. 예컨대 2008년 서브프라임 모기지 금융위기(subprime mortgage crisis)와 최근 코로나 바이러스로 인한 경제 위기는 투자자의 감정을 붕괴시키면서 효율적인 투자 판단 및 적절한 대응 전략을 수립하는데 큰 어려움을 갖게 된다. 또한 만기가 긴 펀드를 운용하는 투자자의 경우에는 장기적인 관점에서 금융 시장이 호황 및 불황일 때 시장 수익보다 초과 수익을 달성할 수 있는 매매 전략을 수립하는 것이 무엇보다 중요하다.

앞서 설명한 투자 방식은 갑작스러운 시장 위험이 발생했을 때 투자자의 심리적 부분이 크게 반영되면서 효과적인 투자 의사결정을 할 수 없다는 한계점을 내포하고 있는 반면 시스템 매매는 투자자의 심리적 반영을 최소화하고 사전에 수립한 규칙에 의해서 매매를 진행하기 때문에 투자자의 감정을 최소화할 수 있다. 그러나 시스템 매매의 경우는 투자자가 사전에 투자 규칙을 개발해야하는데 잘못된 투자 전략을 수립할 경우에 이를 동적으로 개선하는데 한계가 있기 때문에 시장의 변화에 즉각적으로 반응하여 업데이트 하는 것이 어렵다[8]. 반면 인공지능 기반의 시스템 매매는 규칙 기반의 시스템 매매와 달리 데이터를 기반으로 강화학습 모델이 스스로 학습을 통해 규칙을 정의한다. 강화학습은 최적의 의사결정 시 보상(reward)이라는 신호를 주고 이런 과정에서 받은 보상의 총합, 즉 누적 보상(cumulative reward)을 최대화하는 것을 의미하는데 이런 강화학습 기법은 계속 진화하고 있다. 본 연구에서는 금융 상품의 가치 변화를 예측하기 위해 강화학습의 DQN(deep q-network)과 A2C(advantage actor critic) 모형을 적용한다. 또한, 각 모형에 대한 예측력을 분석하고 실효성을 입증하고자 한다.

2. 선행연구

최근 수많은 연산을 짧은 시간에 처리할 수 있게 되면서 인공신경망(artificial neural network), SVM(support vector machine) 등의 기계학습(machine learning) 기법을 이용하여 금융 상품의 미래 가치를 예측하는 기술에 많은 관심을 갖고 있다[9]. 1987년 Lapedes와 Farber이 기계학습 모델을 이용하여 과거 약 2달의 주가 자료를 바탕으로 미래 2달 후의 주가를 예측하는 연구가 발표되면서 2000년 이후에는 기계학습을 통해서 주가를 예측하는 연구가 많이 진행되고 있다[10].

이모세와 안현철[11]은 KOSPI 지수를 예측하기 위해 CNN(convolution neural network)을 적용하였는데 영업일 기준으로 5일씩 분할 차트를 생성한 결과 예측력이 가장 높았다. 주일택과 최승호[12]는 LSTM(long short term memory)을 이용해 국내 개별 종목의 주가 방향성을 예측하는데 입력 셀의 개수별로 시뮬레이션 실험을 통해 예측력을 입증하였다. 또한 A2C(advantage actor-critic) 모형을 포트폴리오 관리 방안에 적용하면 일반적인 ETF(exchange traded fund) 성격의 상품에 비해서 우수한 성과를 보였다[13].

Kanas[14]는 인공신경망 학습 모델을 통해서 S&P500 지수를 예측하였는데 지수의 변동성이 높은 구간에서 우수한 예측력을 보였다. Yoon and Swales[15]는 다변량 판별분석(multivariate discriminant analysis)을 이용해 주가 예측능력을 분석하였고 그 결과 다변량 판별분

석은 학습 기간에는 약 74 %이며 예측 기간에서는 약 65 %의 예측력을 나타냈다. Wong[16]은 퍼지(fuzzy) 시스템과 인공신경망 학습 모델을 결합시킨 주식 예측 모형을 개발하였는데 입력 자료(input data)를 전문가의 지식으로 전환시킬 수 있는 규칙을 활용하여 퍼지 시스템으로 가공한 후에 인공신경망 학습 모델에 규칙을 입력하는 방식을 적용하였다.

박재연, 유재필 그리고 신현준[17]은 SVM(support vector machines)과 라쏘 회귀분석(lasso regression) 등을 이용해서 KOSPI 지수를 예측하였다. 그 결과 학습 데이터에서는 SVM이 인공신경망에 비해서 더 높은 예측력을 보였고 실험 데이터에서는 인공신경망의 예측력이 더 우수했다.

Hamid and Zahid[18]의 연구에서는 옵션모형, 선형모형 그리고 인공신경망 학습 모형을 이용하여 S&P500 지수 선물의 변동성을 예측한 결과, 인공신경망 학습 모형의 예측력이 우수하다는 것을 입증하였다. Hadavandi[19]는 퍼지 시스템과 인공신경망을 이용하여 주가를 입력 값으로 설정하고 향후 움직임을 예측한 결과 기술적 분석의 방향성 예측 기법에 비해 우수한 성과를 보였다. Zhiqiang, Guo[21]는 SVM을 이용해 미래 주가 움직임을 예측하였는데 입력 변수를 정의할 때 POS(particle swarm optimization) 알고리즘을 적용하면 일반적인 시장 데이터를 적용했을 때보다 예측 성능이 더욱 우수했다.

3. 강화학습 모형

3.1 강화학습

강화학습은 기계학습의 한 영역으로써 행동심리학의 이론을 기본으로 한다. Fig. 1에서 보는 바와 같이 에이전트(agent)가 무지인 상황에서 어떠한 선택(action)을 했을 때 보상(reward)을 극대화할 수 있는 선택을 하는 과정을 강화학습이라 한다. 즉 강화학습은 에이전트가 현재 처한 환경(environment)에서 스스로 시행착오(experience)를 경험하여 최적의 행동(action)을 찾아간다는 점에서 기존의 모든 경험치 결과를 바탕으로 학습 및 설계되는 지도학습(supervised learning)과 구분된다.

에이전트는 현재 환경에서 자신의 상황을 관찰하고 이런 상황에서 자신이 기대하는 보상의 합이 가장 큰 방향으로 행동을 선택한다. 선택 후 보상을 받으면 보상의 강도에 따라 에이전트가 보유한 정보를 스스로 수정한다. 이런 과정을 새로운 의사결정 문제의 환경 속에서 반복한다. 강화학습을 주식 매매전략에 적용할 경우, 에이전트는 과거 주식시장의 시계열 데이터를 학습하여 어떤 상태일 때 어떤 행동을 취하는 것이 최선인지 학습을 진행한다. 그 다음 에이전트가 주어진 현재 주식시장의 상태를 관찰하여 최선의 행동을 수행하게 된다. 여기서 상태가 변하게 되면 그에 따라 최선의 행동도 달라지고, 상태 변화에 따른 행동의 순서(action sequence)는 무한히 많아지며 이 중에서 가능한 한 최선의 행동순서를 결정해 나가는 것이 강화학습의 목표다.

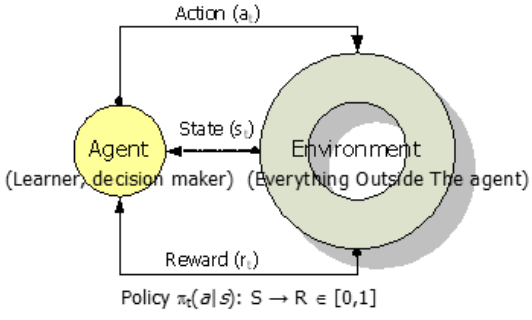

Fig. 1. Principles of reinforcement learning

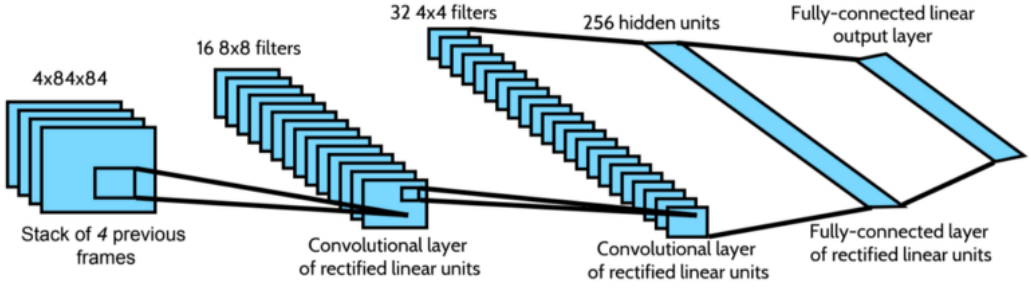

Fig. 2. Architecture of Deep Q-Network[23]

3.2 Deep Q-Network(DQN)

구글 딥마인드에서 제안한 알고리즘인 DQN은 기존의 강화학습 알고리즘인 Monte Carlo Methods와 Temporal Difference에 심층신경망(DNN: Deep Neural Network, 이하 DNN)을 적용하여 DNN의 장점을 극대화했다. DQN은 지도학습과 달리 출력 층의 정상 출력 값을 알 수 없어서 가중치(weight)를 갱신하기 위한 오차를 계산할 수 없는 반면 추정값인 부트스트랩(bootstrap value)을 이용해 오차를 추정하여 가중치를 갱신하고 오차를 하위 단으로 전달한다[22]. Fig. 2는 DQN의 구조를 표현하고 있는데 입력 자료의 다운사이징(downsizing)을 통해서 4개의 프레임(frame) stack으로 구성한다[23]. 본 연구에서는 Volodymyr Mnih[22]가 제안하고 있는 심층신경망에 기반을 둔 DQN 알고리즘을 기반으로 모형을 구성한다.

3.3 Advantage Actor-Critic(A2C)

DQN의 경우에는 심층신경망의 출력값은 각 의사결정에 대한 보상이고 이를 바탕으로 정책망(policy network)을 추정하는데 A2C 알고리즘은 Actor 네트워크와 Critic 네트워크로 구성해 딥 러닝으로 추정한다. Actor는 Eq. (1)에서 gradient ascent을 통해 θ를 설정하는데 π 값이 모수이기 때문에 $Q^{\pi}(s,a)$도 알 수 없다. 따라서 본 연구에서는 TD(temporal difference) 기법을 통해서 실제 $Q^{\pi}(s,a)$ 값을 추정하는데 이를 Critic 네트워크로 학습하는 방식을 취한다.

$$\nabla_{\theta} J(\pi_{\theta}) = E_{s \sim \rho^{\pi}, a \sim \pi_{\theta}}[\nabla_{\theta} \log \pi_{\theta}(a|s) Q^{\pi}(s,a)] \quad (1)$$

DQN 및 A2C 알고리즘은 파이썬(python)과 R 소프트웨어를 통해서 구현하고 A2C 알고리즘의 경우에는 David Silver[24]가 제안한 알고리즘을 기반으로 모형을 구성한다.

4. 실험계획 및 분석

본 장에서는 DQN 및 A2C 모형을 이용해 12개 업종에 속한 개별주식들의 약 13.75년 동안의 주가관련 시계열 데이터를 토대로 매매전략을 학습하고, 학습에 이용되지 않은 12개 업종 내 개별주식들에 대한 테스트 결과를 분석하고자 한다. 본 실험에서 공매도(short selling)는 허용하였고, 증권거래세 등의 거래비용은 고려하지 않았다.

4.1 실험계획

실험 데이터는 Table 1과 같이 KOSPI시장에 상장된 주식들에 대해 12개의 산업업종(industries)을 대표하는 종목들로 구성하였으며, 각각 학습을 위한 데이터와 테스트를 위한 종목으로 분류하였다. 주가 수집기간은 2007년 1월 1일 ~ 2020년 9월 30일까지의 종목 당 약 3360일에 해당한다.

학습 및 테스트 데이터 세트의 상태(state)는 10개의 벡터로 구성된다. 즉, 데이터 한 세트는 5개의 주가 관련 데이터(Open, High, Low, Close, Volume) 및 3개의 기술적지표(ATR, Boll, MACD) 그리고 2개의 주가지수(KOSPI, S&P500)로 구성된다.

보상에 해당하는 리워드(reward)는 주가수익률로 설정하였고, 여기서 주가수익률은 로그수익률로 계산하였다. 종료시점(terminal)에서는 보유 포지션을 강제 청산하지만 이때의 수익률은 누적 수익률에 포함하지 않았다. 또한, 10개의 항목으로 구성된 시계열 데이터 세트는 모두 정규화(normalization)를 통해 전처리하였다. 14개의 종목을 통합(aggregation)하여 하나의 데이터 세트로 학습하였고, 테스트는 13개의 개별 종목들에 대하여 시행하였다.

Table 1. Data for Experiments

Industries	Stocks for Training	Stocks for Test
Construction	GS E&C	Hyundai Engineering & Construction
Finance	Hana Financial Group	Korea Investment Holdings
Mechanic	Doosan Infracore	Doosan Heavy Industries & Construction
Services	GS, Kangwon Land	LG, NCSoft
Transportation Apparatus	Samsung Heavy Industries	Hyundai Mobis
Distribution	Daewoo International	Hotel Shilla
Automobile	KIA Motors Corp.	Hyundai Motor Company
Electric/Gas	KOGAS	KEPCO
Electrical Electronics	LG Display, LG Electronics	SK Hynix, Samsung SDI
Steel/Metal	POSCO	Hyundai Steel
Communication	KT	LG U+
Chemistry	S-Oil	Lotte Chemical
Periods	2007.01.01 ~ 2020.09.30	
Data set	Open, High, Low, Close, Volume, ATR(Average True Range), Boll(Bollinger Bands), MACD, KOSPI, S&P500	

Table 2는 DQN과 A2C 알고리즘을 이용한 주식 매매전략의 우수성을 평가하기 위한 실험계획으로, 매입후 보유(Buy and hold, 이하 BH) 및 매도후유지(Sell and wait, 이하 SW)전략과 비교함으로써 단순 우상향 또는 단순 우하향하는 시황에서도 제시한 강화학습 매매전략들의 초과수익달성 가능 여부를 확인하고자 한다. 누적수익률을 성능지표로 하였고, 보상함수는 수익률(log return)로 상태(state)에는 데이터 세트에 포함된 10개의 속성 외에 포지션 유지기간 및 현재 포지션 상태(buy, sell, hold)를 추가하였다. 또한 강화학습 알고리즘이 취할 수 있는 행동(action)은 매입(buy), 매도(sell), 유지(hold) 중 하나이다.

Table 2. Experimental design

Settings		Values
Performance		Cumulative profit
Reward function		profit(= sell price - buy price)
State		Data set(Open, High, Low, Close, Volume, ATR, Boll, MACD, KOSPI, S&P500) + Holding time, Current inventory
Action		Buy, Sell, Hold
Comparative methods		DQN, A2C, Buy&Hold, Sell&Hold
Hyper parameters & basic configurations	DQN	Alpha=0.0001, Gamma=0.95, Epsilon=0.1, 2 hidden layers Target network, Experience replay memory Activation functions for input, hidden & output layers: Relu, Relu, Linear, respectively Loss function: MSE, Optimizer: Adam
	A2C	Alpha=0.0001, Gamma=0.95, Max Entropy=0.05, 1 hidden layer 〈Actor network〉 Activation functions for input & output layers: Relu, Softmax Optimizer:Adam 〈Policy network〉 Activation functions for input & output layers: Relu, Linear Loss function: MSE, Optimizer: Adam

4.2 실험결과

본 연구에서 제안한 DQN과 A2C 기법을 이용한 매매전략의 성능 비교를 위해 앞 절에서 기술한 실험계획에 따라 실험한 결과는 각각 Table 3-4와 같다. 각 테이블은 테스트 데이터 세트에 포함된 종목들에 대해 연 환산수익률(annualized profit), 평균 보유기간(average holding time), 매매 횟수(#of tradings) 그리고 BH와 SW에 대한 상대값(comparative value; 이하 CV, Eq. (2))을 나타낸다.

$$CV = \frac{(profit_1 - profit_2)}{profit_2} \times 100\% \quad (2)$$
$$where\ \ 1: DQN \text{ or } A2C,\ \ 2: BH \text{ or } SW$$

DQN의 연 환산수익률은 평균 31.3%이었고, 평균 보유기간이 상대적으로 긴 262일로 1년에 약 1.2회 매매하는 결과를 보여주었다. A2C의 연 환산수익률은 평균 12.9%이었고, 평균보유기간이 상대적으로 짧은 71.1일로 1년에 약 3회 매매하는 결과를 나타냈다. BH와 SW에 대한 상대값은 평균으로 보면 DQN과 A2C이 모두 2~5배 가까이 높은 수익률을 보인다. 즉, DQN과 A2C를 이용한 매매 의사결정이 더 효과적임을 알 수 있다.

Table 3. Results with DQN

Stocks for Test	Annualized profit	Average holding time (days)	# of Tradings	CV:Buy & Hold (BH)*	CV:Sell & Wait (SW)*
Hyundai Engineering & Construction	64%	66.65	27	2530.6%	2330.6%
Korea Investment Holdings	14%	114.86	15	-1.1%	198.9%
Doosan Heavy Industries & Construction	42%	64.63	28	736.3%	536.3%
LG	29%	97.47	20	0.3%	200.3%
NCSoft	11%	173.8	16	-65.2%	134.8%
Hyundai Mobis	34%	227.09	12	135.5%	335.5%
Hotel Shilla	35%	360.22	10	158.5%	358.5%
Hyundai Motor Company	43%	251.83	13	137.8%	337.8%
KEPCO	16%	136.08	14	236.1%	36.1%
SK Hynix	9%	1640.5	3	-48.7%	151.3%
Samsung SDI	29%	156	18	-26.0%	174.0%
Hyundai Steel	49%	84.96	24	1986.1%	1786.1%
LG U+	33%	136.25	13	426.3%	226.3%
Lotte Chemical	31%	157.81	17	228.1%	428.1%
Average	31.3%	262.0	16.4	459.6%	516.8%

*: *comparative value*

Fig. 3-4는 현대건설(Hyundai engineering & construction)과 한국전력(KEPCO) 종목에 대해 각각 DQN과 A2C 기법을 적용하여 매매한 결과이며, 매매일별 누적수익률을 나타내고 있다. DQN을 적용했을 때 현대건설의 경우 총 매매일은 27일 및 890%의 누적수익률을 보였고, A2C를 적용했을 때 한국전력의 경우 총 매매일은 47일 및 250%의 누적수익률을 나타냈다. 마찬가지로 Fig. 5-6은 약 3360일 동안의 정규화된 주가의 움직임에 따른 매매(매수-보유-매도 또는 매도-유지-매수)기

록, 즉 언제 매수하여 언제 매도하였는지 또는 언제 매도하여 언제 매수(청산)하였는지 자세히 보여주고 있다.

본 실험에서 전체적으로 DQN에 의한 성능이 A2C 보다 우수하였는데, 이는 A2C 기법이 DQN과 같이 replay buffer를 활용하지 않고 탐색 데이터를 즉시 학습에 이용하기 때문에 잘못된 경로로 학습했을 경우 결과값에 좋지 않은 영향을 미친 것으로 판단된다.

Table 4. Results with A2C

Stocks for Test	Annualized profit	Average holding time (days)	# of Tradings	CV:Buy & Hold (BH)*	CV:Sell & Wait (SW)*
Hyundai Engineering & Construction	15%	72.22	41	663.9%	463.9%
Korea Investment Holdings	27%	64	44	96.3%	296.3%
Doosan Heavy Industries & Construction	21%	65.11	46	416.5%	216.5%
LG	13%	55.55	52	-53.3%	146.7%
NCSoft	-4%	75.89	39	-113.1%	86.9%
Hyundai Mobis	-4%	82.57	38	-127.0%	73.0%
Hotel Shilla	-8%	70.72	40	-156.4%	43.6%
Hyundai Motor Company	3%	75.51	38	-82.9%	117.1%
KEPCO	18%	61.83	47	248.8%	48.8%
SK Hynix	7%	76.57	36	-61.0%	139.0%
Samsung SDI	4%	61.32	45	-90.4%	109.6%
Hyundai Steel	15%	72.22	41	663.9%	463.9%
LG U+	76%	79.75	37	860.6%	660.6%
Lotte Chemical	-2%	82.55	34	-120.3%	79.7%
Average	12.9%	71.1	41.3	153.3%	210.4%

*: *comparative value*

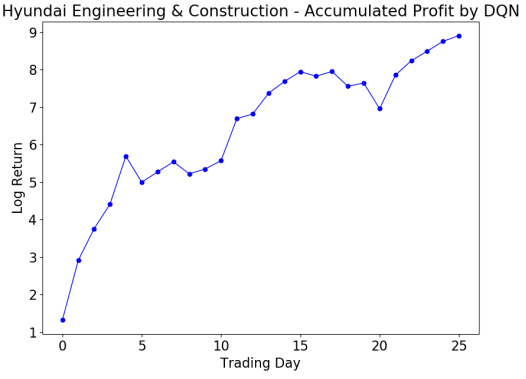

Fig. 3. Accumulated profit by DQN (e.g. Hyundai Engineering & Construction)

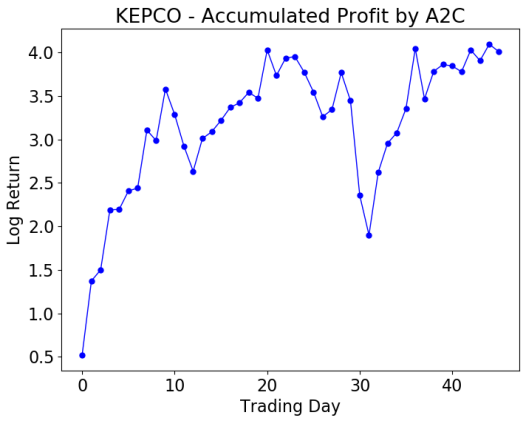

Fig. 4. Accumulated profit by A2C (e.g. KEPCO)

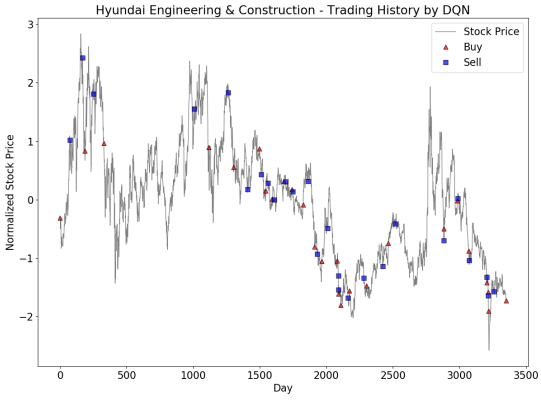

Fig. 5. Trading history by DQN (e.g. Hyundai Engineering & Construction)

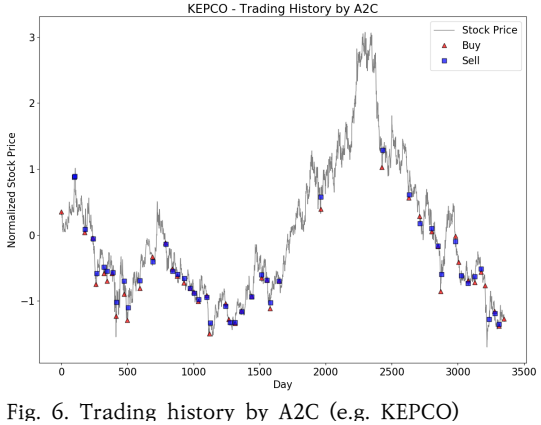

Fig. 6. Trading history by A2C (e.g. KEPCO)

5. 결론

본 논문에서는 강화학습 알고리즘 중 DQN과 A2C 기법을 이용하여 매매전략을 수립하고, 2007년 1월부터 2020년 9월까지 12개 업종에 속하는 14개 종목(학습 및 테스트 총 28 종목)의 주가관련 데이터를 이용하여 학습 데이터 세트를 구성하여 학습한 후 테스트를 통해 그 성능을 비교 평가하였다. 실험결과 DQN과 A2C를 이용한 매매전략의 평균 연 환산수익률이 각각 31.3% 및 12.9%로 동일 기간 국내채권형펀드 연 환산수익률(3.7%) 및 가치주식형펀드 연 환산수익률(6.0%) 대비 우수하다고 판단된다. 또한 BH 및 SW 전략과의 비교에서도 모두 2~5배 가까이 높은 수익률을 나타냈고 이를 통해 DQN과 A2C를 이용한 매매 의사결정이 매매전략에 더 효과적임을 보였다.

제안한 DQN 및 A2C 매매전략의 성능을 향상시키기 위한 추가 연구로는 주가, 기술적지표 및 시장지수 외에 더욱 다양한 특성을 입력데이터인 상태(state)에 반영하는 것을 고려하고 있다. 예를 들어, 현재 주가의 방향성(상승, 하락, 보합)을 나타내는 지표 값 또는 주가 데이터의 집합을 입력데이터에 반영하는 방안 등이 있다. 또한 알고리즘 성능 개선을 위해 DDQN(Double DQN) 및 A3C(Asynchronous A2C) 등의 기법을 적용한 실험과 CNN(Convolutional Neural Networks), LSTM(Long-short Term Memory network) 등과의 결합 알고리즘에 대한 평가도 의미있는 추후 연구로 판단된다.

References

[1] Banz, R. W., "The Relationship between Return and Market Value of Common Stocks", *Journal of Financial Economics*, Vol.9, No.1, pp.3-18, 1981.
DOI: http://dx.doi.org/10.1016/0304-405X(81)90018-0

[2] DeBondt, W. and R. Thaler, "Does the Stock Market Overreact?", *Journal of Finance*, Vol.40, No.3 pp.793-805, 1985.
DOI: https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

[3] Han, Y., K. Yang, and G. Zhou, "A New Anomaly: The Cross-Sectional Profitability of Technical Analysis," *Journal of Financial and Quantitative Analysis*, Vol.48, No.5, pp.1433-1461, 2013.
DOI: http://dx.doi.org/10.2139/ssrn.1656460

[4] Zhu, Y. and G. Zhou, "Technical Analysis: An Asset Allocation Perspective on the Use of Moving Averages," *Journal of Financial Economics*, Vol.92, No.3, pp.519-544, 2009.
DOI: http://dx.doi.org/10.2139/ssrn.1656460

[5] Benjamin Graham, David L. Dodd, Security Analysis, p.258, Natl Book Network, 2003, pp.77-120.

[6] Fama, E. F. and K. R. French, "Common Risk Factors in the Returns on Bonds and Stocks", *Journal of Financial Economics*, Vol.33, No.1, pp.3-56, 1993.
DOI: http://dx.doi.org/10.1016/0304-405X(93)90023-5

[7] Jae Pil Ryu, Chang Hoon Hahn, and Hyun Joon Shin, "Portfolio Construction Strategy for IT Companies Using DEA Method", *The Journal of Information Technology and Architecture*, Vol.14, No.2, pp.139-146, 2017.
https://www.earticle.net/Article/A345705

[8] H. G. Shong, *Multimodal Reinforcement Learning based Stock Trading System combined with CNN and LSTM*, Master's thesis, Kwangwoon University of Science and Technology, Seoul, Korea, pp.8-15, 2018.

[9] Jae Pil Ryu, Hyun Joon Shin, "Portfolio Selection Strategy Using Deep Learning", *The Journal of Information Technology and Architecture*, Vol.15, No.1, pp.43-50, 2018.
https://www.earticle.net/Article/A345667

[10] P. R. Burrell and B. O. Folarin, "The impact of neural networks in finance", *Neural Computing & Applications*, Vol.6, pp.193-200, 1997.
http://link.springer.com/article/10.1007/BF01501506

[11] M. S. LEE and H. C. Ahn, "A Time Series Graph based Convolutional Neural Network Model for Effective Input Variable Pattern Learning : Application to the Prediction of Stock Market", *Journal of Intelligence and Information Systems*, Vol.24, No.1, pp.167-181, 2018.
http://dbpia.co.kr/journal/articleDetail?nodeId=NODE07408509

[12] I. T. Joo and S. H. Choi, "Stock Prediction Model based on Bidirectional LSTM Recurrent Neural Network", *Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol.11, No.2, pp.204-208, 2018.
http://dbpia.co.kr/Journal/articleDetail?nodeId=NODE07424401

[13] García-Galicia, Mauricio, Alin A. Carsteanu, and Julio B. Clempner, "Continuous-time reinforcement learning approach for portfolio management with time penalization", *Expert Systems with Applications*, Vol.129, No.2, pp.27-36, 2019.
DOI: http://dx.doi.org/10.1016/j.eswa.2019.03.055

[14] Angelos Kanas, "Non-linear forecasts of stock returns", *Journal of Forecasting*, Vol.22, No.4, pp.299-315, 2003.
DOI: http://dx.doi.org/10.1002/for.858

[15] Yoon, Y., Swales G., "Predicting stock price performance: a neural network approach", *Proceedings of the Twenty-Fourth Annual Hawaii*

International Conference on System Sciences, IEEE, HI, USA, pp.156-162, August 2002
https://ieeexplore.ieee.org/abstract/document/184055

[16] Wong, L. K., Leung, F. H. F. and Tam P. K. S, "An Improved Lyapunov Function Based Stability Analysis Method for Fuzzy Logic Control Systems", *Electronics Letters*, Vol.36, pp.1085-1086, 2000.
DOI: http://dx.doi.org/10.1109/FUZZY.2000.838698

[17] Jae Yeon Park, Jae Pil Ryu and Hyun Joon Shin, "Predicting KOSPI Stock Index using Machine Learning Algorithms with Technical Indicators", *The Journal of Information Technology and Architecture*, Vol.13, No.2, pp.331-340, 2016.
https://www.earticle.net/Article/A346152

[18] Hamid, Shaikh A. and Iqbal, Zahid, "Using neural networks for forecasting volatility of S&P 500 Index futures prices", *Journal of Business Research, Elsevier*, Vol.57, No.10, pp.1116-1125, 2004.
http://ideas.repec.org/a/eee/jbrese/v57y2004i10p1116-1125.html

[19] Hadavandi, E., H. Shavandi, and A. Ghanbari, "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting", *Knowledge-Based Systems*, Vol.23, No.8, pp.800-808, 2010.
DOI: https://doi.org/10.1016/j.knosys.2010.05.004

[20] Ping Feng Pai and Chih Sheng Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting", *Omega*, Vol.33, No.6, pp.497-505, 2005.
DOI: https://doi.org/10.1016/j.knosys.2010.05.004

[21] Zhiqiang, Guo, Wang Huaiqing, and Liu Quan, "Financial time series forecasting using LPP and SVM optimized by PSO", *Soft Computing*, Vol.17, No.5, pp.805-818, 2013.
DOI: http://dx.doi.org/10.1007/s00500-012-0953-y

[22] Volodymyr Mnih, "Playing Atari with Deep Reinforcement Learning", DeepMind Technologies, United States, pp.2-4, 2013.
https://arxiv.org/pdf/1312.5602.pdf

[23] Nando de Freitas, Reinforcement learning : OXFORD University p.28, 2019, pp.20,
http://www.cs.ox.ac.uk/people/nando.defreitas/machinelearning/lecture12.pdf

[24] David Silver, Lecture7:Policy Gradient, p.41, 2020, pp.10-3,
www.davidsilver.uk/wp-content/uploads/2020/03/pg.pdf

조 현 민(Hyunmin Cho) [준회원]

- 2020년 2월 : 상명대학교 경영공학과 (공학사)
- 2019년 9월 ~ 현재 : 상명대학교 경영공학과 학석사연계과정

〈관심분야〉
품질경영, 스마트제조, 인공지능, 빅데이터

신 현 준(Hyun Joon Shin) [종신회원]

- 1997년 2월 : 고려대학교 산업공학과 (공학석사)
- 2002년 2월 : 고려대학교 산업공학과 (공학박사)
- 2002년 5월 ~ 2004년 4월 : Texas A&M Univ. Post-Doc.
- 2004년 5월 ~ 2005년 2월 : ㈜삼성전자 책임연구원
- 2005년 3월 ~ 현재 : 상명대학교 경영공학과 교수

〈관심분야〉
금융공학, 인공지능, 빅데이터, 스마트제조