Journal of Next-generation Convergence Information Services Technology Vol.10, No.4, August (2021), pp.459-470
ISSN 2384-101X(p) 2672-1163(e) *http://dx.doi.org/10.29056/jncist.2021.08.10*

심층 Q신경망을 이용한 포트폴리오 성능평가

Performance Evaluation of Portfolio using a Deep Q-Networks

이우식[1]

Woosik Lee[1]

요 약

2016년 구글의 알파고가 이세돌과의 바둑대전에서 우승한 이후 금융권에서는 인공지능과 투자자문 전문가의 합성어인 로보어드바이저에 관심이 높아지고 있다. 디지털 자산관리에 대한 중요성이 커지면서 로보어드바이저 산업의 장기적 성장성은 매우 높을 것으로 예상된다. 하지만 로보어드바이저를 이용한 디지털 자산관리가 아직 충분히 검증된 상태가 아니고, 인공신경망을 포함한 강화학습과 금융이 연계된 연구들은 많이 이뤄지지 않고 있는 상태이다. 본 연구에서는 확률적 경사 하강법 알고리즘에 따른 심층 Q신경망 모형기반의 포트폴리오 성능을 비교 분석한다. 분석 결과, 심층 Q-신경망을 이용한 포트폴리오의 샤프지수가 동일비중 포트폴리오보다 낮은 수치를 기록했다. 이는 리플레이 메모리에 저장되는 대부분의 트랜지션에 유익한 보상 지표가 없으며 심층 Q신경망의 수렴 및 훈련에 불충분한 정보가 제공되는 것으로 사료된다. 둘째, 일반적으로 성능이 뛰어나다고 알려진 Adam과 다르게 본 연구에서는 RMSprop의 높은 학습 효과를 확인할 수 있다.

핵심어 : 강화학습, 비즈니스 애널리틱스, 핀테크, 디지털 자산관리, 로보어드바이저

Abstract

After Google's AlphaGo beats Go master Lee Se-dol in 2016, the financial industry has been increasingly interested in Robo-Advisor. As the importance of digital asset management grows, the long-term growth potential of the Robo-advisor industry is expected to be very high. However, digital asset management using Robo-advisors has not yet been sufficiently verified, and many studies related to reinforcement learning including artificial neural networks and finance have not been conducted. This study compares some stochastic gradient descent algorithms of deep Q-network(DQN) to maximize the long-term financial portfolio performance. As a result, the Sharpe ratio of the portfolio using the DQN was lower than that of the equally weighted portfolio. It is considered that most of the transitions stored in the replay memory have no informational reward indicator, and provide insufficient value to the convergence and training of the DQN. Secondly, unlike Adam, which is generally known for its excellent performance, the performance of the DQN with RMSprop outperformed.

Keyword : Reinforcement Learning, Business Analytics, FinTech, Digital Asset Management, Robo-Advisor

1 College of Business and Economics, Gyeongsang National University, Jinju, Korea [Professor]
e-mail: woosiklee@gnu.ac.kr

Received(June 18, 2021), Review Result(1st: July 12, 2021), Accepted(August 13, 2021), Published(August 31, 2021)

1. 서론

한국은행 금융통화위원회가 기준금리 ‘연 0.50%’로 유지하면서 한국에서는 2%를 웃도는 예·적금 금리상품은 거의 존재하지 않고, 계속되는 코로나 19 사태 이후의 경제회복에 대한 전망은 갈리는 것으로 분석이 나오고 있다. 이러한 불투명한 국내·외 경제환경에 목돈 마련과 노후 대비를 위해서 금융투자가 필수가 된 상황이다 [1].

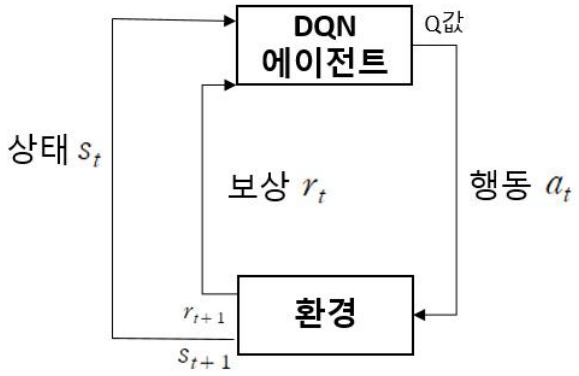
하지만 미국의 소비자물가지수의 증가로 인한 물가 상승, 원·달러 환율의 상승 등 불확실하고 예측이 어려운 금융환경에서 수익률 관리에 매우 어려움이 존재한다. 이러한 금융시장 리스크에 대응하기 위해 금융시장 가격변수의 변동으로 인한 자산 가치 손실을 최소화하는 전략으로 안정적인 수익률 달성을 추구하는 금융자산배분에 관심을 둘 필요가 있다. 금융자산배분은 학문적 뿐만 아니라 실무적으로도 오랫동안 널리 활용되고 있는 투자전략이다. Markowitz 연구 [2]는 수익과 위험 관계의 정량적분석을 통한 금융자산배분을 제시하였고, Brinson 연구 [3]은 금융자산 배분에 관한 전략 성과가 투자시기(Market Timing), 종목 선택(Asset Selection) 그리고 금융자산 배분(Asset allocation) 중 약 91.5%가 자산 배분을 통해 결정된다는 연구 결과를 발표하였다.

금융자산배분은 투자자의 투자목적과 투자성향 등을 고려하여 금융자산배분을 구성하고, 감내가능한 투자리스크 수준과 투자목적 달성을 위한 금융자산배분 최적화 등의 여러 과정을 포함하는 과학적인 투자전략 [4]으로 최근 주요 자산운용사, 은행, 증권회사, 스타트업(Startup)기업 등은 인공지능 기반의 자산관리 서비스 제공을 위한 로보어드바이저(Robo-Advisor) 서비스를 제공하고 있다. 로보어드바이저를 통해 고객맞춤형 포트폴리오를 제공한다지만 아직 충분히 검증된 상태가 아니고, 인공신경망(Artificial Neural Network)을 포함한 강화학습(Reinforcemen Learning)과 금융(Finance)가 연계된 연구들은 많이 이뤄지지 않고 있는 상태이다. 기존 연구 중 김선웅 [5]의 연구에서는 블랙리터만 모델(Black-Litterman Model)의 투자자 전망을 서포트벡터머신을 활용하여 자산배분 모델을 제시하였고, 이우식 [1]의 연구에서는 바닐라 오토인코더, 잡음제거 오토인코더 그리고 딥오토인코더를 사용하여 시장을 잘 대표할 수 있는 오차가 작은 소수의 종목들로 이루어진 동일비중포트폴리오를 사용하여 목표시장지수의 움직임과 비교 분석하여 자산배분에 머신러닝의 적용 가능성을 제시하였다. García-Galicia외 2인 [6]연구에서는 자산배분관리를 위해 A2C(Advantage Actor-Critic) 강화학습모형을 이용하여 에이전트가 각 금융시장상태 간의 전이 확률(Transition Rate Matrices)을 측정하였다. 본 연구에서는 최적화 알고리즘에 따른 심층강화학습모형기반의 포트폴리오 성능을 비교 분석하고자 한다. 본 논문의 구성은 다음과 같다. 제 1절의 서론에 이어 제 2절에서는 주요 방법론인 AdaGrad, RMSprop 그리고 Adam 최적화 알고리즘에 대한 설명을 소개하였으며, 제 3절에서는 실증분석 및 결과를 확인한다. 마지막으로 제 4절에서는 결론과 시사점을 제시한다.

2. 이론적 배경

2.1 심층 Q신경망

2016년 구글의 알파고가 이세돌과의 바둑대전에서 우승한 이후 게임, 로봇, 자율주행, 주식거래 그리고 국방 등 다양한 분야에서 강화학습(Reinforcement Learning)을 사용하고 있다. 강화학습은 {S, A, P, R, γ}로 구성된 마코프 결정 과정(Markov Decision Process)으로 정의되는 환경으로부터 기대누적보상값(Expected Cumulative Reward)을 극대화하는 최적화 기법이다. 여기서 S는 유한한 크기를 갖는 상태(State) 집합, A는 유한한 크기를 갖는 행동(Action) 집합, $P(s'|s,a)$는 상태 전이(State Transition) 확률로 현재 상태 $s \in S$ 에서 행동 $a \in A$를 취했을 때 다음 상태가 $s' \in S$이 되는 확률 분포(Probability Distribution)를 의미하며, R은 보상(Reward)함수, $\gamma \in (0,1)$은 할인 계수(Discount Factor)를 나타낸다. 강화학습의 학습순서는 [그림 1]과 같이 각 시간단계(Time Step)에서 정의된 에이전트(Agent)가 주어진 환경(Environment)으로부터 현재 상태(State)를 관찰하여 이를 기반으로 행동(Action)을 선택하고, 이 때 환경의 상태가 변화하면서 정의된 에이전트는 행동에 따른 보상(Reward)을 받는다. 에이전트는 학습 초기에 무작위로 행동하지만, 학습이 진행되면서 점차 더 많은 보상을 얻을 수 있는 행동으로 학습하게 된다. 강화학습은 초기 상태로부터 정책(Policy)에 기반을 두고 연속적인 행동(Continuous Action)을 취했을 때의 기대누적보상(Expected Cumulative Reward)을 최대화(Maximize)하는 정책을 발견하는 것을 목적으로 하며 이를 최적 정책(Optimal Policy)이라 하고 π^*로 표현한다.



[그림 1] 심층 Q신경망의 구조

[Fig. 1] Structure of deep Q-Network

강화학습 기법 중 큐러닝(Q-Learning)은 현재 상태로부터 행동을 맵핑하는 정책을 도출할 수 있

는 방법론으로 순차적 행동 결정문제에 적용되어 매우 우수한 성과를 나타낼 수 있음이 증명되었다 [7]. 하지만 상태와 행동에 따른 큐값(Q-Value)을 표형식의 큐테이블(Q-Table)을 통해서 정책을 도출하기 때문에 수많은 상태가 존재하는 현실세계의 복잡한 문제에서 학습속도가 매우 느려지고, 희소 큐테이블(Sprase Q-Table)이 도출되는 등 불안정한 학습과정을 겪게 된다. 이런 큐러닝의 문제는 큐러닝의 큐테이블 대신 심층인공신경망을 사용하여 큐값을 근사해낼 수 있도록 학습시키는 심층 Q신경망(Deep Q-Network)을 통해 효과적으로 해결되었다. 즉 심층 Q신경망의 심층인공신경망의 입력층에서는 상태가 이산화(Discretization)된 값의 형식이 아닌 연속적인 값(Continuous Value)의 형식으로 입력되기 때문에 심층 Q신경망은 차원의 저주문제(Curse of Dimensionality Problem)를 해결할 수 있고, 이를 바탕으로 상태의 수가 매우 많은 복잡한 문제에 적용이 가능하다. 컴퓨터 게임 화면의 픽셀(Pixel)값을 입력으로 하는 심층인공신경망을 사용하여 행동 가치 함수(Action-Value Function)를 근사하는 방법을 통해 아타리(Atari) 2600 에 속한 대부분의 게임에서 사람을 뛰어넘는 성능에 도달할 수 있다는 것을 입증했다 [8].

심층 Q신경망은 벨만 방정식(Bellman Equation)을 인공신경망으로 확장하기 때문에 $Q_\theta(s,a)$로 표기한다. 여기서 θ는 인공신경망의 파라미터 벡터를 의미한다. 인공신경망에서 $r+\gamma \max_{a'} Q_\theta(s',a')$을 정답으로 보고 이것과 인공신경망에 의한 추측치인 $Q_\theta(s,a)$사이의 차이를 줄이는 방향으로 학습을 수식(1)과 같이 진행한다. 따라서 심층 Q신경망의 손실함수는 아래와 같이 정의한다.

$$L(\theta) = (r+\gamma \max_{a'} Q(s',a')) - Q_\theta(s,a))^2 \quad (1)$$

[표 1] 심층 Q신경망 학습
[Table 1] Deep Q-network Algorithm

1. Q_θ의 파라미터 θ를 초기화
2. 에이전트의 상태s를 초기화($s \leftarrow s_0$)
3. 에피소드가 끝날 때까지 다음 (a) ~ (e)를 반복
   (a) Q_θ에 대한 $\epsilon - greedy$를 이용하여 액션 a를 선택
   (b) a를 실행하여 r과 s'을 관측
   (c) s'에서 Q_θ에 대한 $greedy$를 이용하여 액션 a'을 선택
   (d) θ 갱신 : $\theta \leftarrow \theta + \alpha(r+\gamma Q_\theta(s',a') - Q_\theta(s,a)) \nabla_\theta Q_\theta(s,a)$
   (e) $s' \leftarrow s$
4. 에피소드가 끝나면 다시 2번으로 돌아가서 θ가 수렴할 때까지 반복

신경망 학습 중 학습자료가 시간(Time)적인 상관(Correlation)관계를 가지고 있다면, 학습에 제약

이 생길 수 있다. 이런 문제를 해결하기 위해 에이전트가 경험을 저장해두었다가 랜덤(Random)하게 일부를 선택해서 인공신경망을 강건(Robust)하게 학습시키는 경험 리플레이(Experience Replay)을 이용한다. 즉 랜덤하게 선택된 경험을 유지함으로써 인공신경망이 다양한 과거 경험으로부터 학습하게 한다.

2.2 연구모형

본 연구에서는 심층 Q-신경망을 이용한 금융자산배분의 성능평가를 위해 여러 가지 초매개변수(Hyperparameter)중에서 최적화 알고리즘인 경사하강법(Gradient Descent)에 중점을 두고 자료를 분석하였다. 경사하강법은 인공신경망을 학습하는데 손실 함수 혹은 비용 함수를 최소화되도록 가중치를 조절하기 위해 사용하는 알고리즘이다. 하지만 경사하강법의 단점은 한 번 가중치를 갱신하기 위해 전체 데이터를 사용하는 것이고 이로 인해 비효율적이고 학습에 오랜 시간이 걸린다. 이런 문제를 해결하기 위해 확률적 경사하강법을 이용한다. 즉 데이터를 샘플링하여 적은 데이터를 학습하고, 이 과정을 여러 번 반복하면서 확률적으로 전체데이터를 학습하는 것과 비슷한 효과가 나타나게 한다는 것이다. 그러나 경사하강법과 확률적 경사하강법은 학습 도중에 극소값에 빠져 최소값을 발견할 수 없는 경우가 발생할 수 있는 단점이 존재한다. 이에 본 연구에서는 Adaptive Gradient(AdaGrad), Root Mean Square Propagation(RMSProp) 그리고 Adaptive Moment Estimation(Adam) 최적화 경사하강법 알고리즘을 통해 심층 Q-신경망을 학습한다.

2.2.1 Adaptive Gradient(AdaGrad) 최적화 알고리즘

실제 인공신경망 학습에서 데이터의 특성에 따른 효과적인 학습률(Learning Rate)을 선택하는 것은 중요하다. 학습률이 너무 작으면 최적점을 찾아가기 위해 많은 학습 횟수가 필요하고, 학습률이 너무 크면 오차를 줄이는 방향으로 가지 못하고 발산해서 학습이 제대로 이루어지지 않는다. 이런 문제를 학습률 규제(Learning Rate Decay)를 통해 해결했는데, AdaGrad알고리즘은 가중치가 갱신됨에 따라 학습률도 자동으로 조정되도록 설계되었다. AdaGrad알고리즘의 업데이트 방법은 다음과 같다.

$$G = G + \nabla_w J(w) \odot \nabla_w J(w) \tag{2}$$

$$w = w - \frac{\eta}{\sqrt{G} + \epsilon} \odot \nabla_w J(w) \tag{3}$$

여기서 $\odot$는 행렬의 원소별 제곱을 의미하고, ϵ는 분모가 0이 될 때의 오류를 제한하기 위한 작은 값이다.

2.2.2 Root Mean Square Propagation(RMSProp) 최적화 알고리즘

RMSProp알고리즘은 AdaGrad알고리즘을 개선한 알고리즘으로 AdaGrad알고리즘에서 G는 현재 시간까지의 변화량의 합으로 정의되기 때문에 시간이 지날수록 증가하게 되고 학습률은 감소하지만, RMSProp알고리즘은 이전의 변화량과 현재의 변화량의 지수 평균(Exponential Average)으로 정의되기 때문에 학습률이 급격하게 감소하는 현상을 방지할 수 있다 [9]. 즉 최신에 학습한 데이터가 가중치 변경에 좀 더 많은 영향을 미치도록 설계되었다.

$$G = \gamma G + (1-\gamma)\nabla_w J(w) \odot \nabla_w J(w) \tag{4}$$

$$w = w - \frac{\eta}{\sqrt{G+\epsilon}} \odot \nabla_w J(w) \tag{5}$$

2.2.3 Adaptive Moment Estimation(Adam) 최적화 알고리즘

Adam 알고리즘은 가중치 변경값에 관성(Momentum) 을 추가한 Momentum알고리즘과 RMSProp알고리즘의 특성을 모두 추가한 알고리즘으로, 최적화에 의한 갱신 경로를 변경하는 운동량(Momentum) 개념을 기반으로 과거의 갱신 정보를 일정 부분 반영하면서, 새로 계산된 경사 방향과의 조합으로 최종적인 갱신크기를 계산한다. 즉 운동량처럼 현재까지 계산해온 기울기의 지수 평균을 저장하고, 아래과 같이 RMSProp알고리즘처럼 기울기의 제곱 값의 지수 평균(Exponential Average)을 저장한다.

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_w J(w) \tag{6}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)(\nabla_w J(w))^2 \tag{7}$$

위 식에서 m_t과 v_t는 각각 1차 운동량과 2차 운동량의 추정값을 나타내고, β_1와 β_2는 감쇠 비율(decay rate)로 운동량 측정에 사용되는 고정 변수로서 일반적으로 β_1는 0.9 β_2는 0.999가 사용된다. 이러한 갱신 방식만을 이용할 경우, 운동량의 초기값이 0이기 때문에 초반 갱신 크기가 매우 작아지는 문제가 야기된다. 따라서 이러한 문제를 보정하기 위해 Adam 알고리즘에서는 아래와 같이 1차와 2차 운동량 추정치의 편향 보정한다.

$$\widehat{m_t} = \frac{m_t}{1-\beta_1^n} \tag{8}$$

$$\hat{v_t} = \frac{v_t}{1-\beta_2^n} \tag{9}$$

여기서 $\hat{m_t}$과 $\hat{v_t}$은 보정된 운동량 추정치이다. 따라서 RMSProp알고리즘의 특성에 추가적으로 운동량 개념을 적용한 Adam 알고리즘은 식은 수식(10)과 같다.

$$w = w - \frac{\eta}{\sqrt{\hat{v_t}} + \epsilon} \tag{10}$$

3. 실증분석

3.1 자료의 구성

본 연구에서 사용할 표본은 미국 다우존스 산업평균 지수로 미국 시장 전체를 대표할 수는 없지만, 지수를 구성하는 종목 수가 우량기업 30개로 상대적으로 쉽게 추종 지수 움직임에 따른 포트폴리오를 민들 수 있고, 이를 통해 금융 포트폴리오의 다변화를 꾀하는 국내 투자자가 미국 주식시장에 투자할 수 있는 간단한 투자수단에 부합한다고 볼 수 있다. 실험을 위해 2010년 1월 4일부터 2019년 12월 31일까지 일별 수정 종가데이터를 수집했다.

다우존스 산업평균지수 일별 수정종가에 대한 평균, 표준편차, 왜도 및 첨도에 대한 기초통계량은 [표 2]에서 살펴볼 수 있다. 지수에서 나타나는 음의 첨도는 꼬리가 정규 분포보다 얇음을 나타내고, 지수 수익률의 왜도가 음의 값이라는 것은 부정적인 극단 현상이 나타날 가능성이 정규분포보다 높음을 의미한다 [1].

$$\text{주가지수변화율} = \ln(\text{주가지수}(t) \ / \text{주가지수}(t-1)) \tag{11}$$

[표 2] 기초통계량

[Table 2] Descriptive Statistics

	다우존스 산업평균지수	다우존스 산업평균지수 변화율(%)
평균	17,602.39	0.000433
중앙값	17,006.77	0.00056
최대값	28,645.26	0.049846
최소값	9,686.48	-0.055464
표준편차	5,143.45	0.008857
왜도	0.425716	-0.399456
첨도	-0.969152	3.927072

3.2 모형의 추정 및 분석

심층 Q신경망에 의한 금융자산배분 분석에 각 행동 선택에 따른 기대 보상값의 비교와 검증을 통하여 최종적으로 최적의 모형을 확인하였다. 즉 AdaGrad, RMSprop 그리고 Adam 최적화 알고리즘로 이루어진 각각의 심층 Q신경망의 학습을 위해 미국다우지수를 구성하고 있는 종목의 일별 수익률과 이에 대한 상관행렬을 상태변수로 사용하고 금융자산비중에 따른 기대 샤프지수의 비교와 검증을 실시하였다. 이때 활성함수에서 발생할 수 있는 기울기 소실문제를 해결하기 위해 활성화함수로 엘루(ELU)함수를 이용하였고 심층 Q신경망의 최적화 함수와 손실함수는 각각 AdaGrad, RMSprop, Adam 그리고 평균제곱오차(Mean Squared Error)로 설정하였다. 더불어 공매도를 허용하였고, 증권거래세 등의 거래비용은 고려하지 않았다.

인공신경망을 구성할 때, 은닉층의 수, 은닉층에 존재하는 유닛 개수, 활성화 함수(Activation function), 신경망의 초기화 기법 그리고 최적화 등이 적절히 최적화하지 못하면 높은 결과를 이끌어 낼 수 없다. 이러한 인공신경망의 초매개변수들의 모든 가능한 조합(Combination)을 통한 최적의 조합을 찾는 것은 매우 어려운 문제이며 많은 계산량이 요구된다. 여러 초매개변수 중 학습과 가장 직접적으로 관련된 최적화 알고리즘은 모형의 손실 값을 최소화 할 수 있도록 가중치를 조정하는데 가장 중요한 역할을 담당한다 [9]. 동일한 구조와 동일한 초매개변수를 갖는 인공신경망이라도 최적화 알고리즘에 따라 성능이 다르게 나타날 수 있으므로 적합한 최적화 알고리즘를 찾는 것은 매우 중요하다. 이에 본 논문에서는 AdaGrad, RMSprop 그리고 Adam 최적화 알고리즘이 심층 Q신경망기반의 금융자산배분에 미치는 영향을 측정하는 연구를 수행했다. 그 결과 [표 3] 에서와 같이 모든 실험에서 정(+)의 샤프지수를 보여 이는 위험대비 투자수익이 발생했음을 의미한다. 하지만 AdaGrad, RMSprop 그리고 Adam 최적화 알고리즘을 적용한 심층 Q신경망 에이전트는 동일비중 포트폴리오보다 낮은 성과를 보여주었다. 즉 샤프지수가 가장 높은 때는 동일비중을 사용할 때(0.0090)로 나타냈으며 뒤이어 RMSProp알고리즘을 사용하고 학습진행이 5000번일 때(0.0046)가 위험 대비 높은 투자 성과를 보여 주었다. 이는 학습이 진행됨에 따라서 동일한 상태에서 동일한 행동을 수행했을 때 얻는 보상의 비정상적(Non-Stationary)으로 인해 심층 Q-신경망의 경험 리플레이 메모리(Experience Replay Memory)에 저장된 과거의 경험들이 그대로 활용하기 어렵다는 문제와 보상이 드물게 혹은 지연되어 주어지는 환경에서는 학습이 성공적으로 이루어지기 어렵다고 사료된다 [10]. 하지만 학습진행이 증가함에 따라 RMSprop, AdaGrad 그리고 Adam 최적화 알고리즘을 순으로 심층 Q신경망 에이전트는 주어진 상태에서 점점 더 많은 보상 혹은 샤프지수를 추구하는 쪽으로 행동을 취하는 경향성을 보임을 확인할 수 있다.

[표 3] 포트폴리오 성과분석

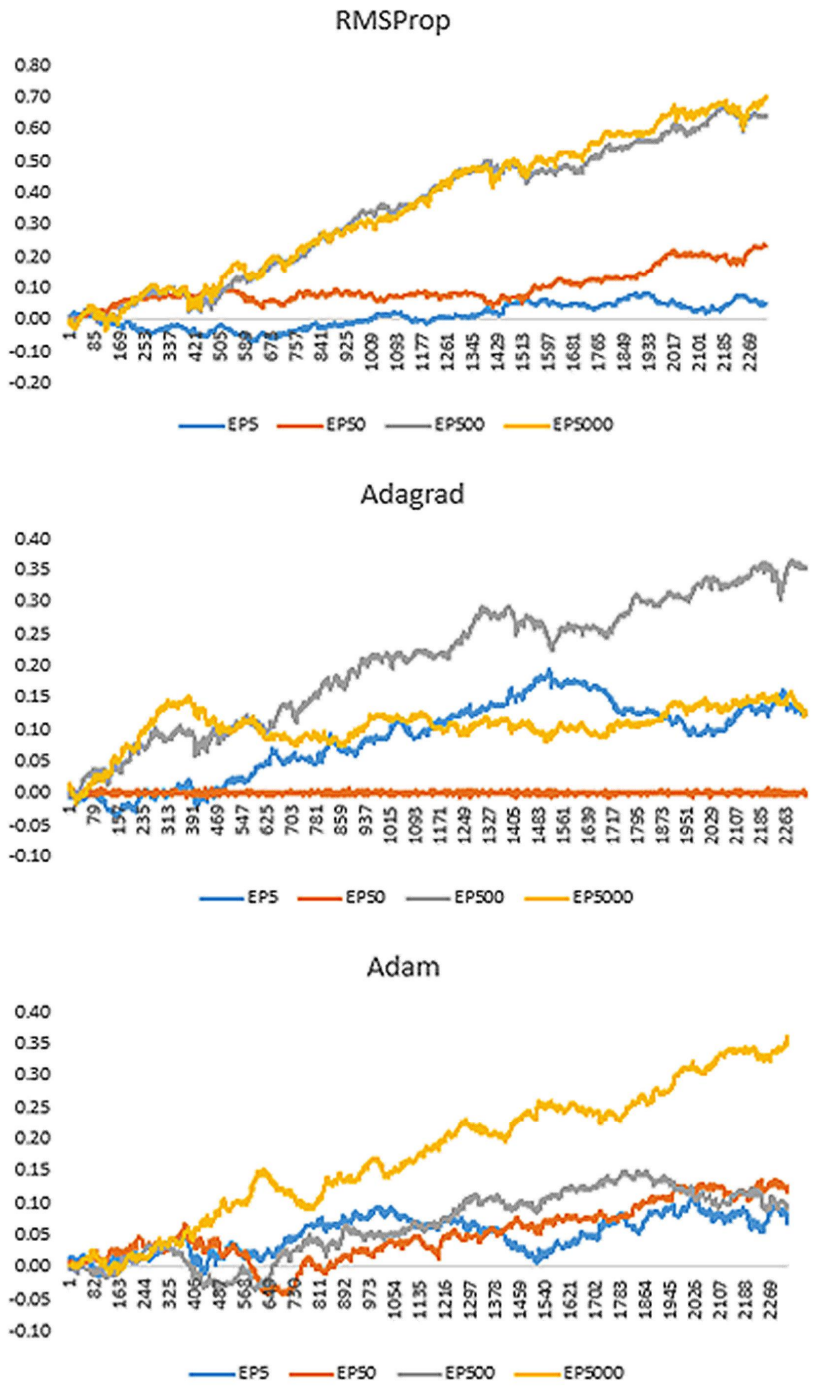
[Table 3] Performance of Portfolio

	수익률	표준편차	샤프지수
동일비중	0.0007	1.1601	0.0090
학습진행 5			
RMSProp	0.0000	0.1668	0.0020
Adagrad	0.0010	0.4105	0.0021
Adam	0.0000	0.1850	0.0025
학습진행 50			
RMSProp	0.0001	0.6562	0.0024
Adagrad	0.0001	0.9367	0.0022
Adam	0.0001	0.4210	0.0021
학습진행 500			
RMSProp	0.0003	1.1960	0.0036
Adagrad	0.0002	0.8638	0.0028
Adam	0.0000	0.2946	0.0021
학습진행 5000			
RMSProp	0.0003	1.0493	0.0046
Adagrad	0.0001	0.4151	0.0022
Adam	0.0002	1.1761	0.0021

심층 Q-신경망 학습 성능을 살펴보면 [그림 2]에서와 같이 상대적으로 오랜 학습진행이 필요했던 AdaGrad와 Adam과 다르게 RMSprop 최적화 알고리즘의 경우, 학습진행이 500과 5000 일 때와 비슷한 성능을 보여주었다. 이를 통해 특히 RMSProp 최적화 알고리즘이 직·간접적으로 복잡하게 얽혀 있는 여러 변수들에 의해 불규칙하게 변화하는 주식시장 환경에서 보다 잘 작동한다 높은 학습 효과를 확인할 수 있다. 마지막으로 심층 Q신경망에 운동량(Momentum)개념과 적응형 방법(Adaptive Method)을 이용한 Adam 최적화 알고리즘을 이용한 학습이 RMSprop과 AdaGrad에 비해 비교적 학습진행에 따른 학습 성과가 안정적으로 증가하는 현상을 보인다.

4. 결론

미국 내 Wealthfront, Betterment등 약 200 여개의 로보어드바이저회사가 이미 존재하고 우리나라도 2016년 구글의 알파고가 이세돌과의 바둑대전에서 우승한 이후 금융권에서는 인공지능과 투자자문 전문가의 합성어인 로보어드바이저(Robo-Advisor)에 관심이 높아지면서 인공지능 기반의 자산관리 서비스 제공을 위한 스타트업(Startup)기업들이 생겨나고, 주요 은행들은인공지능 금융 서비스 고도화를 위한 업무협약이 활발하게 진행되고 있다.



[그림 2] 누적수익률

[Fig. 2] Cumulative Returns

더불어 대한민국 정부도 핀테크(FinTech)를 더욱 활성화하는 내용의 전자금융거래법 개정을 추진하고 있다. 이처럼 디지털 자산관리(Digital Asset Management)에 대한 중요성이 커지면서 로보어드바이저의 성장성은 매우 높을 것으로 예상된다.

본 연구에서는 심층 Q-신경망을 이용한 금융자산배분의 성능평가를 위해 AdaGrad, RMSProp 그리고 Adam 최적화 경사하강법 알고리즘 기반의 심층강화학습성능을 분석하였다. 분석 결과는 다음과 같다. 첫째, 동일비중 포트폴리오의 샤프지수가 심층 Q-신경망을 이용한 포트폴리오에서보다 더 높은 수치를 기록했다. 이는 학습이 진행됨에 따라서 동일한 상태에서 동일한 행동을 수행했을 때 얻는 보상의 비정상적(Non-Stationary)으로 인해 심층 Q-신경망의 경험 리플레이 메모리(Experience Replay Memory)에 저장된 과거의 경험들이 그대로 활용하기 어렵다는 문제와 보상이 드물게 혹은 지연되어 주어지는 환경에서는 학습이 성공적으로 이루어지기 어렵다고 사료된다 [10]. 둘째, 일반적으로 성능이 뛰어나다고 알려진 Adam과 다르게 본 연구에서는RMSprop, AdaGrad 그리고 Adam 최적화 알고리즘을 순으로 심층 Q신경망 에이전트는 주어진 상태에서 점점 더 많은 보상 혹은 샤프지수를 추구하는 쪽으로 행동을 취하는 경향성을 보임을 확인할 수 있다. 셋째, 오랜 학습진행이 필요했던 AdaGrad와 Adam과 다르게 RMSprop 최적화 알고리즘의 경우, 학습진행이 500과 5000 일 때와 비슷한 성능을 보여주었다. 이를 통해 특히 RMSProp 최적화 알고리즘이 직간접적으로 서로 복잡하게 얽혀 있는 수많은 변수들에 의해 불규칙하게 변화하는 주식시장 환경에서 보다 잘 작동한다 높은 학습 효과를 확인할 수 있다. 마지막으로 심층 Q신경망에 운동량(Momentum)개념과 적응형 방법(Adaptive Method)을 이용한 Adam 최적화 알고리즘을 이용한 학습이 RMSprop과 AdaGrad에 비해 비교적 학습진행에 따른 학습 성과가 안정적으로 증가하는 현상을 보인다.

인공신경망을 구성할 때, 은닉층의 수, 은닉층에 존재하는 유닛 개수, 활성화 함수(Activation function), 신경망의 초기화 기법 그리고 최적화 등이 적절히 최적화하지 못하면 높은 결과를 이끌어 낼 수 없다. 이러한 인공신경망의 초매개변수들의 모든 가능한 조합(Combination)을 통한 최적의 조합을 찾는 것은 매우 어려운 문제이며 많은 계산량이 요구된다. 여러 초매개변수 중 학습과 가장 직접적으로 관련된 최적화 알고리즘은 모형의 손실 값을 최소화 할 수 있도록 가중치를 조정하는데 가장 중요한 역할을 담당한다 [9]. 동일한 구조와 동일한 초매개변수를 갖는 인공신경망이라도 최적화 알고리즘에 따라 성능이 다르게 나타날 수 있으므로 적합한 최적화 알고리즘를 찾는 것은 매우 중요하다. 이에 본 논문에서는 AdaGrad, RMSprop 그리고 Adam 최적화 알고리즘에 따른 심층 Q신경망기반의 학습 성능에 미치는 영향을 측정하는 연구를 수행했다. 하지만 본 논문에도 향후 몇 가지 보완할 점이 필요하다. 가치 기반의 강화학습모형 외에 정책기반모형 등을 추가한 비교연구가 필요하다. 더불어 본 연구에서 제시하였던 심층강화학습모형에 다양한 경제지표 등을 추가하여 좀 더 높은 성과를 기대할 수 있도록 하는 방안에 대해 연구하고자 한다.

References

[1] W. S. Lee, “A Study on the Index Reconstruction Error Analysis Using Neural Networks”, The Korean Society of Management Consulting, vol. 21, no. 1, February 2021, pp. 1-8, doi: 10.29056/jncist.2018.12.04.

[2] H. Markowitz, “Portfolio Selection”, Journal of Finance, vol. 7, no. 1, February 1952, pp. 77-91, doi: 10.2307/2975974.

[3] G. P. Brinson, P. L. Randolph-Hood, G. L. Beebower, “Determinants of Portfolio Performance”, Financial Analysts Journal, vol. 51, no. 1, December 1955, pp. 133-138, doi: 10.2469/faj.v42.n4.39.

[4] I. Bajeux-Besnainou, J. V. Jordan, R. Portait, “Dynamic Asset Allocation for Stocks, Bonds, and Cash”, The Journal of Business, vol. 76, no. 2, December 2003, pp. 263-288, doi: 10.1086/367750.

[5] S. Kim, “Robo-Advisor Algorithm with Intelligent View Model”, Journal of intelligence and information systems, vol. 25, no. 2, December 2019, pp. 39-55, doi: 10.13088/jiis.2019.25.2.039.

[6] M. García-Galiciaab, A. A. Carsteanuab, J. B. Clempnerab, “Continuous-time reinforcement learning approach for portfolio management with time penalization”, Expert Systems with Applications, vol. 7, no. 1, September 2019, pp. 27-36, doi: 10.1016/j.eswa.2019.03.055.

[7] J. Clifton, E. Laber, “Q-Learning: Theory and Applications”, Annual Review of Statistics and Its Application, vol. 129, no. 1, March 2020, pp. 279-301, doi: 10.1146/annurev-statistics-031219-041220.

[8] V. Minih, K. Kavukcuoglu, D. Silver, “Human-level control through deep reinforcement learning”, Nature, vol. 518, no. 1, February 2015, pp. 529-533, doi: 10.1038/nature14236.

[9] G. Joo, C. Park, H. Im, “Performance Evaluation of Machine Learning Optimizers”, Journal of IKEEE, vol. 24, no. 3, December 2020, pp. 766-776, doi: 10.7471/ikeee.2020.24.3.766.

[10] S. Y. Jang, H. J. Park, N. S. Park, “Research Trends on Deep Reinforcement Learning”, Electronics and Telecommunications Trends, vol. 34, no. 4, December 2019, pp. 1-14, doi: 10.22648/ETRI.2019.J.340401.