The Journal of the Convergence on Culture Technology (JCCT)
Vol. 7, No. 4, pp.807-814, November 30, 2021. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2021.7.4.807
JCCT 2021-11-99

강화학습 기반 주식 투자 웹 서비스

An Implementation of Stock Investment Service based on Reinforcement Learning

박정연*, 홍승식**, 박민규***, 이현****

Jeongyeon Park*, Seungsik Hong, Mingyu Park***, Hyun Lee******

요 약 코로나-19로 인해 경제 활동이 낮아지고 주식 시장이 침체하면서 주식 투자를 통해 또 다른 소득을 마련하기 위해 많은 사람이 주식 시장에 뛰어들고 있다. 사람들의 관심이 높아지면서 더 많은 수익을 얻기 위한 주가 분석 연구가 많이 진행되고 있다. 주가는 종목별 변동의 흐름이 다르므로 각 주가 종목별로 독립적이며 일관적으로 분석할 필요가 있다. 이러한 문제를 해결하고자 본 논문에서는 강화학습 기법 중 하나인 Asynchronous Advantage Actor-Critic(A3C)를 이용하여 주가를 분석할 수 있는 모델 및 서비스를 설계 및 구현하였다. 주식 시장 데이터로 종목별 주가 및 국채, 코스피와 같은 외부 요인들을 반영하였다. 또한 웹페이지 제작을 통해 시각화한 정보를 제공하여 투자자들이 투자 기업에 대한 재무제표를 비롯하여 국내외 경제 및 정치의 흐름을 모두 분석하지 않고도 안전한 투자를 할 수 있도록 서비스를 제공한다.

주요어 : 주가 분석, 강화학습, 딥러닝, 웹서비스

Abstract As economic activities decrease, and the stock market decline due to COVID-19, many people are jumping into stock investment as an alternative source of income. As people's interest increases, many stock price analysis studies are underway to earn more profits. Due to the variance observed in the stock markets, it is necessary to analyze each stock independently and consistently. To solve this problem, we designed and implemented models and services that analyze stock prices using a reinforcement learning technique called Asynchronous Advantage Actor-Critic(A3C). Stock market data reflected external factors such as government bonds and KOSPI (Korea Composite Stock Price Index) as well as stock prices. Our proposed work provides a web service with a visual representation of predictions of stocks and stock information through which directions are given to investors to make safe investments without analyzing domestic and foreign stock market trends.

Key words : Stock Price Analysis, Reinforcement Learning, Deep Learning, Web Service

*정회원, 선문대학교 컴퓨터융합전자공학과 석사과정 (제1저자)
**준회원, 선문대학교 컴퓨터공학부 학부생 (참여자자)
***정회원, 선문대학교 컴퓨터융합전자공학과 석사과정 (참여저자)
****정회원, 선문대학교 컴퓨터공학부 부교수 (교신저자)
접수일: 2021년 10월 19일, 수정완료일: 2021년 10월 27일
게재확정일: 2021년 11월 2일

Received: October 19, 2021 / Revised: October 27, 2021
Accepted: November 2, 2021
*Corresponding Author: mahyun91@sunmoon.ac.kr
Professor, Department of Computer Science and Engineering, Sunmoon University, Korea

Ⅰ. 서 론

코로나19의 전 세계적 유행이 현실화되면서 글로벌 증시의 공통 위험요인으로 작용함에[1] 따라 국내 주식 시장이 급락하였다. 특히 감염병 확산 초기에는 중국수요 위축의 영향이 큰 생활용품, 유통 부문에 타격이 가장 컸으나, 주요국 증시의 동반 급락이 시작된 이후 글로벌 경기에 민감한 부품, 소재, 자동차, 조선 등의 부문의 주가 하락폭이 급격히 확대되는 양상을 보였다[2]. 시간이 지남에 따라 다시 경제 활동이 증가하고 많은 대외요인으로 인해 주식 시장이 높은 변동성에 노출되었다. 이를 계기로 투자를 통한 이익을 얻고자 하는 사람들이 증가하였으며 실제 주식 거래 활동 계좌 수는 계속해서 큰 수치로 증가하고 있다.[3]

주식 투자를 하기 위해서 기본적으로 주가를 분석한 결과에 따라서 투자를 결정하는 것이 일반적이다. 크게 기본적 분석과 기술적 분석 두 가지 유형으로 나누어진다. 기본적 분석의 경우, 기업의 내재적 가치를 평가하여 투자를 결정 내린다. 즉, 투자하고자 하는 시점의 기업의 내재적 가치를 계산하여 현재의 시장가격과 비교 후, 해당 종목의 가격의 움직임을 예상하여 결정하는 방식이다. 기술적 분석은 기본적 분석과 달리 오로지 주식의 변동 가격 움직임 자체만을 연구하여 투자 여부를 판단한다. 종목의 가격에 영향을 주는 요인들은 이미 가격 형성에 포함되어 있다고 생각하여 보이는 결과만을 이용하여 투자를 결정하는 방식이다.

근래에 컴퓨터 기술이 발전하고 딥러닝이 주목받음에 따라 주가 분석에서도 딥러닝을 통한 투자 예측 기법이 두각을 나타내고 있다. 컴퓨터 기술을 통한 분석으로 나오는 결과가 매수, 매도 신호이며 이러한 기계적 신호에 따라 투자하는 방식으로 시스템 트레이딩이라고 불린다. 딥러닝 기법을 통한 트레이딩 예측 연구가 활발히 진행됨에 따라 많은 연구가 RNN, LSTM, CNN 등과 같은 모델을 적용[4]하고 있으며, 최근 강화학습을 통한 연구들도 진행되고 있다. Deep Q-network [5], Actor-Critic[6] 또는 Advantage Actor-Critic(A2C) [7]를 적용한 연구들이 존재하고, 좋은 수익률을 결과로 보인다. 하지만 주식은 종목에 따라 상승과 하락의 흐름이 비슷하게 흘러갈 때도 있지만, 대부분 기업의 상황, 분야, 외부적 상황에 따라 다르게 흘러가는 경우가 많다. 종목별 독립적인 흐름을 학습에 반영하기 위해서 본 논문에서는 Asynchronous Advantage Actor-Critic (A3C) 기법을 적용하여 각 종목에 대해 동시에 비동기적이면서 일관적으로 학습한 결과를 분석한다. 더불어 강화학습을 통해 학습된 주가 예측 모델을 이용하여 트레이딩 예측 결과를 사용자에게 제공하는 웹서비스를 구축했다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문과 관련된 연구를 소개한다. 3장에서는 본 논문에서 주가 예측 모델을 학습한 방식과 제안하는 시스템의 구조 및 구현 방법에 관해 기술한다. 4장에서는 모델 실험 결과 및 개발 결과에 관해 기술한다. 5장에서는 결론과 향후 연구 계획을 제안한다.

Ⅱ. 관련 연구

1. Deep Learning 모델 (LSTM)

LSTM(Long Short-Term Memory)[8]는 시퀀스 데이터를 다루는 모델 중 하나이며, RNN(Recurrent Neural Network)의 Vanishing Gradient Problem을 개선한 모델이다. 입력이 오래된 데이터는 기억할 수 없는, 즉 최근에 입력된 데이터를 예측에 더 많이 반영하는 구조인 RNN은 짧은 시퀀스의 데이터에서만 효과를 보이며, 시퀀스가 길어질수록 앞의 정보들이 충분히 뒤로 전달되지 못하는 단점이 존재한다. 이를 장기 의존성 문제라고 하며, 이러한 RNN의 구조를 보완하여 오래된 정보 또한 기억하고 예측에 반영할 수 있도록 한 것이 LSTM이다. LSTM 셀은 입력, 망각, 출력 게이트를 가지고 있다. 입력 게이트는 이전 시점의 셀 데이터와 입력된 데이터를 통해 어느 정도의 데이터를 입력으로 받아들일지를 결정한다. 망각 게이트는 이전 시점의 셀 데이터를 기억해야 하는지를 결정한다. 출력 게이트에서는 연산된 데이터를 내보내야 하는지를 결정한다.

LSTM 셀 내부에서 게이트들을 이용함으로써 긴 시점을 가지는 시퀀스 데이터를 기울기 소실 즉, 정보의 손실 없이 선택적으로 기억하여 처리 가능하다는 장점이 있다. 따라서 주가와 같이 시간의 흐름에 따라 정보가 변동되는 데이터를 다루기에 적절한 모델이다.[9]

2. Reinforcement Learning 모델

강화학습은 에이전트라고 하는 하나의 소프트웨어가 상태를 관측하여 주어진 환경에서 적절한 행동을 취한 후,

행동에 관한 결과로 보상을 받으며, 기대 보상 값을 최대로 할 수 있는 행동을 취하도록 학습하는 알고리즘이다. 이때, 에이전트가 행동을 결정하기 위해서는 정책이라고 불리는 알고리즘이 필요하다. 정책은 무작위성이 포함되어 있으므로 확률적 정책으로 불리기도 한다.

1) Deep Q-network (DQN)

DQN 알고리즘[10]은 model-free 강화학습 알고리즘 중 하나인 Q-learning 알고리즘을 심층망을 적용하여 학습하고, 적절한 Q값을 찾도록 근사하는 알고리즘이다. 기존의 Q-learning 알고리즘은 주어진 상태에서 취한 행동의 가치인 Q 값을 Q 테이블을 이용해 업데이트하며 최적의 값을 찾아가는 방식이다. 이는 경우에 따라 메모리 소모 및 학습에 좋지 않기 때문에 이를 보완하여 가중치를 가지는 심층망을 통해 근사한다. 예를 들어 [11]은 CNN 모델 학습 과정에서 DQN을 적용했다. 각 회사의 증시 차트 이미지를 상태로 입력받아 행동을 취한 후, 행동에 대한 보상을 받는다. 이때, 계산되는 보상을 시간의 흐름에 따라 누적하여 CNN 모델의 훈련에 이용한 연구이다.

2) Actor-Critic

Actor-Critic 알고리즘[12]은 강화학습 알고리즘 중 하나로 Actor network, Critic network를 통해 정책을 개선한다. Actor network의 경우 주어진 상태에 대한 행동을 결정하고, Critic network의 경우 actor가 결정한 행동에 따른 변화된 상태에 대한 가치를 평가한다.

예를 들어, [13]는 최종 결정을 위해 Actor-Critic 알고리즘 중 Actor network만을 사용하는 것이 아니라, Critic network도 같이 사용하여 최종 결정을 안정화한 연구이다. 이때, value network와 policy network에 LSTM 모델을 공유하여 action 결정을 안정화하였다.

3) A2C (Advantage Actor-Critic)

A2C 알고리즘[14]은 Actor의 기대 출력값에 Advantage를 사용한 알고리즘이다. 즉, Actor-Critic 알고리즘에서 주어진 상태에서 취한 행동의 가치인 상태-행동 가치에서 주어진 상태에서 정책을 따랐을 때의 가치인 상태 가치를 뺀 Advantage를 이용해 Critic 함수를 학습한다.

[7]은 DNN 모델에 A2C를 적용하여 주식 투자 예측 모델을 훈련한 연구이다. Policy Gradient, DQN을 적용한 모델보다 더 좋은 수익성을 보장한다는 것을 보여주었다.

4) A3C (Asynchronous Advantage Actor-Critic)

A3C 알고리즘[14]은 A2C알고리즘에 비동기식을 적용한 알고리즘으로 병렬 학습에 초점을 맞췄다, 여러 개의 A2C 알고리즘을 각각의 에이전트에 적용한다. 각 에이전트는 다양한 환경에서 독립적으로 수행되며 공유 신경망을 학습시킨다.

III. 본 론

1. 강화학습을 통한 주가 예측

본 논문에서 설계한 강화학습 모델은 LSTM 모델에 Asynchronous Advantage Actor Critic(A3C) 알고리즘을 적용하여 학습을 진행하였다. 전반적인 구조는 그림 1과 같다. 학습 데이터로 준비된 차트 데이터를 종목별로 구분하고, 각 종목별로 에이전트를 비동기적으로 훈련했다. 또한, 비동기적으로 에이전트를 훈련하는 과정에서 공유네트워크인 LSTM 모델을 업데이트하는 방식을 통해 주가를 예측하고 투자를 결정하는 모델을 구축하였다.

강화학습을 통한 주가 예측을 위해 에이전트를 투자자로 에이전트가 놓인 환경을 주가 데이터로 설정하였다. 상태는 에이전트의 주식 보유 비율과 포트폴리오 가치 비율로 두어 환경인 주가 데이터에 따라 행동으로 매수, 매도, 관망을 결정한다. 학습 초기에는 탐험을 통해 무작위로 행동을 결정함으로써 경험을 쌓고, epoch가 증가함에 따라 탐험 비율(ϵ)을 줄이면서 무작위로 행동을 결정하는 것이 아니라, 신경망을 통해 계산된 결과로 행동을 결정한다. 결정된 행동을 에이전트가 취함으로써 얻는 다음 상태 즉, 변화된 주식 보유 비율과 포트폴리오 가치 비율을 통해 보상을 계산한다. 보상은 즉시 보상과 지연 보상으로 나뉘는데, 즉시 보상(R)은 에이전트의 행동에 대한 수익률로 pv가 포트폴리오 가치일 때, 현재 시점 t에 대해 다음 수식(1)과 같이 계산된다.

$$R = (pv_t - pv_0)/pv_0 \qquad (1)$$

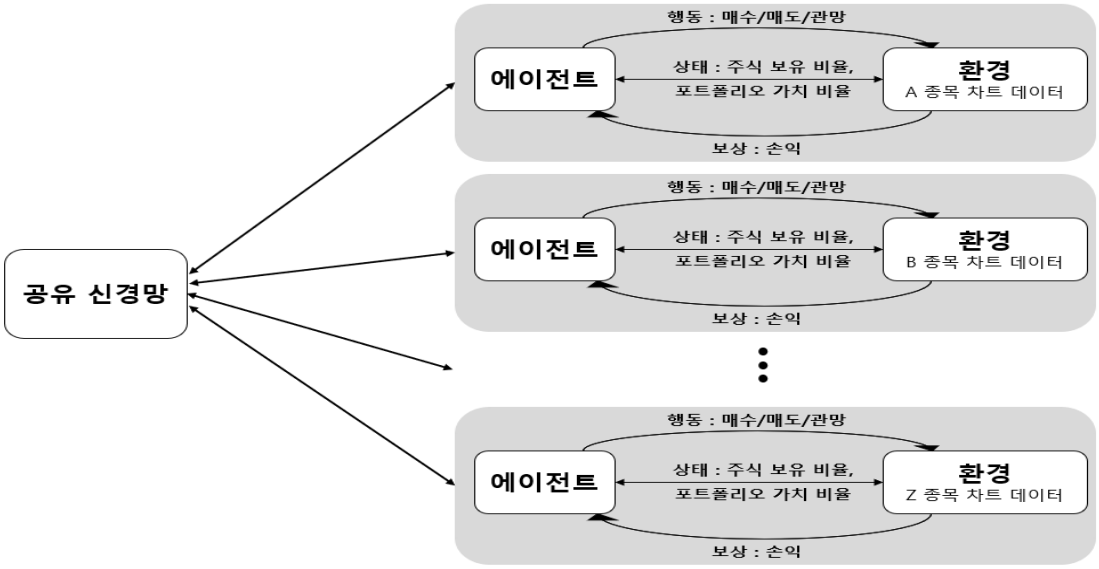

그림 1. A3C를 통한 주식 투자 강화학습
Figure 1. Reinforcement Learning to stock investment through A3C

지연 보상은 임계 값을 넘었을 때의 손익률로써, 0 또는 즉시 보상 값이다. 만약 지연 보상 임계치보다 즉시 보상이 크다면 즉시 보상 값을 지연 보상으로 두고, 작다면 0으로 설정한다. 지연 보상이 0이 아니면 학습이 진행되어 임계치를 초과하는 수익이 난 경우에 긍정적으로 학습하도록 하였다.

제안한 모델을 기반으로 다음 장에서 실험을 통한 결과를 비교 분석하고자 한다.

2. 웹 서비스 설계

본 서비스는 1에서 학습된 강화학습 모델을 이용하여 주식 투자 정보를 제공하는 서비스를 웹 기반으로 제공하고자 했다. Django 기반의 웹 서버를 구축하여 현재 주식 정보와 함께 투자 전략 정보를 제공하는 서비스를 구현했다.

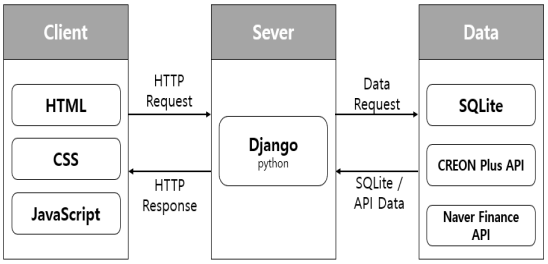

그림 2. 제안 시스템 구조
Figure 2. Proposal System Architecture

그림 2와 같이 제안하는 시스템은 Django 서버를 기반으로 사용자에게 웹 서비스를 제공하는데, SQLite 데이터베이스를 연동한다. 클라이언트로부터 가입자, 로그인, 게시글 정보를 입력받고 SQLite 데이터베이스 서버는 가입, 로그인에서의 사용자 정보를 저장하거나 참조, 게시글을 조회하고 작성하는 과정에서 게시글 정보를 저장하거나 참조한다. 또한, 요청에 따라 상황에 따라 증권사에서 제공하는 CREON Plus API, Naver Finance API로부터 정보를 파싱한다.

Ⅳ. 실험 및 결과

본 논문에서는 주식 예측을 위해 LSTM 신경망에 강화학습 기법의 하나인 A3C 알고리즘을 적용하여 투자자인 에이전트가 수익을 얻기 위해 적절한 매수, 매도, 관망 타이밍을 결정하도록 학습시켰으며, 이를 시각화하여 투자 정보를 확인할 수 있는 웹서비스 플랫폼을 구현했다.

1. 데이터셋

주식 예측 학습에 필요한 데이터는 기본적으로 대신증권에서 제공하는 API인 CREON Plus API를 통해 수집했다. CREON Plus API에서 날짜별 종목별 주가, 거래량을 이용했다.

날짜별 주가 정보만을 이용하면 단기적인 관점에서만 학습 가능하므로 중장기적인 관점에서의 데이터 또한 다뤄야 하며, 주가의 등락에 영향을 줄 수 있는 외부 데이터인 국채, 코스피, 주가수익비율(Price Earning Ratio, PER), 주가순자산비율(Price Book-value Ratio, PBR), 자기자본이익률(Return On Equity, ROE) 관련 정보 또한 학습에 이용해야 한다. 따라서 국고채 통합정보시스템의 3년간 일별 최종호가 수익률, 네이버 금융 API를 통한 연도별, 분기별 PER, PBR, ROE를 수집하였다.

하지만 수집된 데이터들은 종목별로 모두 다른 범위를 가지고 있고 편차도 크기 때문에 표준화 작업을 통해 같은 스케일을 가지면서 종목의존도를 낮추는 과정이 필요하다. 일자별 주가 데이터의 경우 전일 대비 당일 비율 및 당일 종가 대비 당일 비율 등으로 전처리를 거쳤으며 그 외의 코스피, 국채 등의 데이터들은 5, 10, 20, 60, 120일 간격의 평균 종가 대비 당일 종가 비율로 조정하여 단기, 중기, 장기적 관점에서의 방향성을 의미할 수 있도록 조정하였다.

2. 강화학습 모델 학습

A3C 모델 학습에 있어 적절한 Optimizer와 하이퍼파라미터를 찾기 위해 학습 횟수(epoch) 200회, num_steps를 5회로 고정한 이후 learning rate와 optimizer를 변경해가며 학습을 진행했다. Optimizer의 경우 3개의 optimizer를 사용하였다.

미니배치를 통해 학습을 진행하는 SGD (Stochastic Gradient Descent), 기울기를 단순 누적하지 않고 지수가중 이동 평균을 사용하여 최근에 사용된 기울기들이 더 크게 반영되도록 한 RMSProp, 마지막으로 Momentum과 RMSProp을 융합한 Adam을 이용하였다. 또한 learning rate의 경우 0.001, 0.0001, 0.00001 총 3개의 값을 이용하여 테스트를 진행하였다. 그 결과는 표 1과 같다.

표 1. optimizer, learning rate에 따른 수익률 비교
Table 1. Comparison of yields for optimizer and learning rate

옵티마이저	학습률	횟수	수익률(%)
RMSprop	0.001	1	19.57
		2	10.12
		3	8.84
	평균 수익		12.84
	0.0001	1	11.12
		2	9.19
		3	3.26
	평균 수익률		7.86
	0.00001	1	5.26
		2	7.42
		3	11.30
	평균 수익률		7.99
	평균 수익률 (옵티마이저)		**9.57**
SGD	0.001	1	2.51
		2	1.21
		3	3.10
	평균 수익률		2.27
	0.0001	1	5.40
		2	6.85
		3	8.99
	평균 수익률		7.08
	0.00001	1	9.75
		2	6.94
		3	0.27
	평균 수익률		5.65
	평균 수익률 (옵티마이저)		**5.00**
Adam	0.001	1	5.06
		2	7.33
		3	3.08
	평균 수익률		5.16
	0.0001	1	4.15
		2	7.44
		3	10.84
	평균 수익률		7.47
	0.00001	1	5.28
		2	7.29
		3	7.28
	평균 수익률		6.61
	평균 수익률 (옵티마이저)		**6.42**

수익률의 경우 RMSProp일 때의 수익률이 다른 optimizer에 비해 좋은 결과를 보였으므로 RMSProp optimizer를 이용하여 최종적으로 학습한 결과와 테스트 데이터를 통한 결과는 다음 그림 3, 4와 같다. 그림에서의 첫 번째 그래프는 학습한 주식 종목의 일봉 차트, 두 번째 그래프는 에이전트가 보유한 주식 수이며 파란색 부분은 매도, 빨간색 부분은 매수한 것을 의미한다. 세 번째 그래프는 Critic의 출력으로 파란 선은 매도에 대한 가치, 빨간 선은 매수에 대한 가치를 의미하며 네 번째 그래프는 Actor의 출력으로 각 선은 Critic 그래프의 선과 의미가 같다. 마지막 다섯 번째 그래프는 에이전트의 포트폴리오 가치를 보여주는 것으로 투자 손익을 보여주며 노란 선이 의미하는 것은 지연 보상이 발생했다는 것이다.

그림 4을 통해 테스트 데이터에 대한 에이전트의 투자 결과를 확인할 수 있다. 하락장에서의 손실을 1, 2, 3번 시점에서 계속해서 매수하여 가치를 상승시킨 것을 확인할 수 있다.

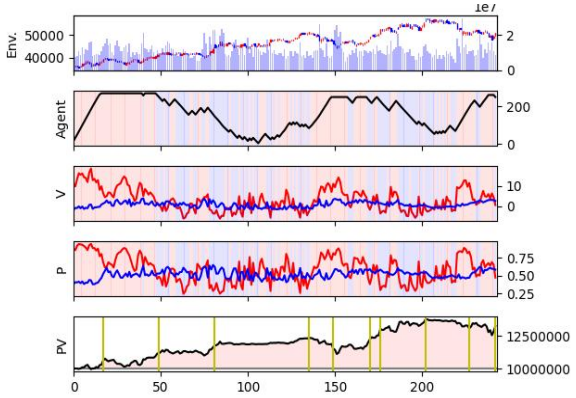

그림 3. 학습 결과 그래프
Figure 3. Result Graph of Training

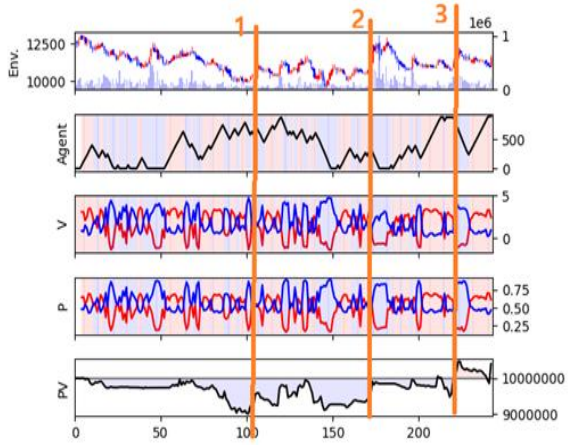

그림 4. 테스트 데이터에 대한 그래프
Figure 4. Graph for test data

3. 제안 웹 서비스

주식 투자 정보를 제공하는 웹 서비스 특징상 인터페이스를 단순하게 적용하고, 사용자 정보인 나이, 성별, 검색 기록을 분석하여 맞춤 종목을 추천하며 2 과정에서 학습시킨 모델을 통한 주가 분석 예측 결과를 매수, 매도 확률로 제공하는 페이지로 구성하였다.

웹에서 제공하는 서비스, 기능들을 직관적으로 확인함과 동시에 간편하게 메뉴 간 이동을 할 수 있도록 내비게이션 바를 구성하였다.

1) 주식 검색

주식 검색 페이지는 그림 5와 같이 구성되어있다. 좌측 상단은 인기 검색 종목으로, 네이버 증권의 인기 검색 주식 종목을 파싱해 동일하게 보여주는 기능을 하지만, 추가로 본 웹 서비스에서 제공하는 종목과 일치하는 종목 중 상위 10가지만을 제공한다.

우측 상단은 맞춤 종목으로 로그인한 회원의 성별, 나이, 검색 데이터를 분석하여 적절한 종목을 보여주는 기능을 한다. 이때, 로그인하지 않는 사용자의 경우 추천 종목을 임의로 보여주도록 설정됐다.

그림 5. 주식 검색 화면
Figure 5. Stock Search Page

2) 주식 상세

주식 검색 화면에서 종목을 검색한 후, 상세 페이지로 이동하게 되면 그림 6과 같은 화면을 제공한다. 우측 상단에는 검색한 종목명과 주가를 그래프로 정보를 제공한다. 우측 하단에는 검색 종목과 관련된 기사를 제공한다. 반면, 좌측 상단의 경우 강화학습을 통해 학습된 딥러닝 모델을 통해 검색 종목의 최근 5일 치 데이터로 예측한 매수, 매도 확률을 제공한다. 또한 좌측 하단에는 검색 종목의 상세 정보로, 현재 주가, 시가 총액, ROE, PBR, PER를 제공하도록 설정했다.

3) 커뮤니티

투자자들 간의 정보 공유 및 질의응답을 위해 커뮤니티 기능을 제공한다. 커뮤니티의 분류는 크게 주식, 자유, 공지로 나누어지며 공지글은 그림 7과 같이 빨간색 글씨로 강조되어 최상단에 게시된다. 글쓰기 버튼을 통해 게시글을 작성하게 되면 분류와 제목, 작성자, 작성일, 조회수 정보를 리스트로 제공하며 각 게시글을 선택하면 작성된 내용을 조회할 수 있다.

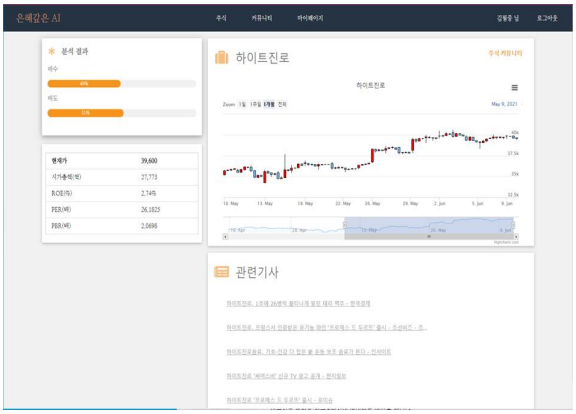

그림 6. 주식 상세 화면
Figure 6. Stock Detail Page

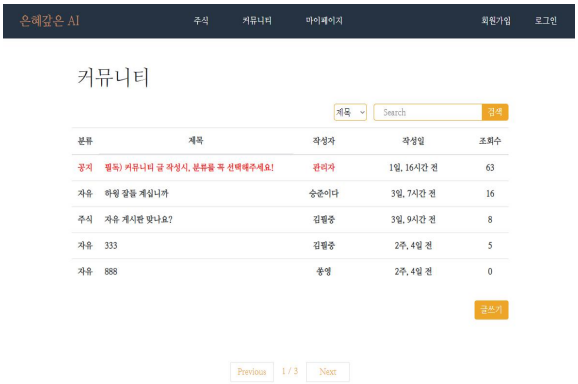

그림 7. 커뮤니티 화면
Figure 7. Community Page

V. 결 론

주식 데이터의 특성상 종목별로 분야, 회사, 정세 등 외부요인에 따라 상승과 하락의 흐름이 반영되기 때문에, 이러한 비동기적인 특성을 주가 데이터에 반영할 수 없는 단점이 존재한다. 기존의 Policy Gradient, AC, A2C와 같은 강화학습 알고리즘을 사용한 연구는 많이 존재 하였으나, 초보자 또는 관련 전공자가 아닌 사람이 사용하기에는 많은 어려움이 존재한다. 본 논문은 A3C를 이용하여 웹 서비스를 제공해줌으로써 사용자가 웹 브라우저를 이용하여 쉽게 주가 정보를 조회할 수 있고, 조회 시점에서 어떤 투자 전략이 좋을지에 대한 정보를 시각화 해서 한 눈에 쉽게 확인 하고, 투자에 반영할 수 있도록 구현하고 제안하였다. 향후, 제안한 A3C 주가 예측 모델의 성능을 PG, A2C, DDPG, TRPO와 같은 기존의 강화학습 알고리즘을 적용한 주가 예측 모델과의 수익률을 비교, 평가, 분석하고 이를 웹 서비스 내에서 여러 가지 투자 전략으로 제공하고자 한다. 하나의 모델에 대한 투자 전략보다는 여러 개의 모델로부터 도출된 투자 전략들을 제안한다면 사용자들에게도 신뢰성 있는 정보를 제공할 수 있을 것으로 기대된다.

References

[1] Verma, P., Dumka, A., Bhardwaj, A. et al, “A Statistical Analysis of Impact of COVID19 on the Global Economy and Stock Index Returns”, SN Computer Science, Vol 2, Jan 2021. DOI: https://doi.org/10.1007/s42979-020-00410-w

[2] Z. Xu, A. Elomri, L. Kerbache and A. El Omri, "Impacts of COVID-19 on Global Supply Chains: Facts and Perspectives,", IEEE Engineering Management Review, Vol. 48, No. 3, pp. 153-166, Sept. 2020, DOI: 10.1109/EMR.2020.3018420.

[3] Stock trading accounts exceeded 50 million... 10 milion more in 5 months, http://www.m-i.kr/news/articleView.html?idxno=851354

[4] Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. "Stock price prediction using LSTM, RNN and CNN-sliding window model“, *2017 international conference on advances in computing, communications and informatics (ICACCI)*. IEEE, pp. 1643-1647. 2017. DOI: 10.1109/ICACCI.2017.8126078

[5] Carta, S., Ferreira, A., Podda, A. S., Recupero, D. R., & Sanna, A, "Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting", *Expert systems with applications*. Vol. 164, February 2021. DOI:https://doi.org/10.1016/j.eswa.2020.113820

[6] Li, Jinke, Ruonan Rao, and Jun Shi. "Learning to trade with deep actor critic methods." *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*. Vol. 2. pp. 67-71, IEEE, 2018. DOI:10.1109/ISCID.2018.10116

[7] H.M. Cho, H.J.Shin, “Trading Strategies Using Reinforcement Learning”, *Journal of Korea Academia-Industrial cooperation Society*, Vol. 22, No. 1, pp. 123-130, January 2021. DOI:https://doi.org/10.5762/KAIS.2021.22.1.123

[8] Hochreiter, S., Schmidhuber, J. "Long short-term memory." *Neural computation*, Vol. 9, No. 8, November 1997. DOI:10.1162/neco.1997.9.8.1735.

[9] D.H. Shinm K.H. Choi. C.B. Kim, “Deep Learning Model for Prediction Rate Improvement

of Stock Price Using RNN and LSTM", Journal of Koriean Institute of Information Technology. Vol. 15, No. 10, pp. 9-16, October, 2017. DOI:https://doi.org/10.14801/jkiit.2017.15.10.9

[10]Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., "Playing atari with deep reinforcement learning." *arXiv preprint* arXiv:1312.5602 2013.

[11]Lee, Jinho, et al. "Global stock market prediction based on stock chart images using deep Q-network.", *IEEE Access*, Vol. 7, pp. 167260-167277, 2019 DOI:10.1109/ACCESS.2019.2953542.

[12]Konda, Vijay R., and John N. Tsitsiklis. "Actor-critic algorithms." *Advances in neural information processing systems*. pp. 1008-1014, 2000.

[13]Li, Jinke, Ruonan Rao, and Jun Shi. "Learning to trade with deep actor critic methods." *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*. Vol. 2, pp. 66-71, IEEE, 2018.

[14]Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Kavukcuoglu, K. "Asynchronous methods for deep reinforcement learning." *International conference on machine learning. PMLR*, pp. 1928-1937, June 2016.

[15]I.T. Joo, S.H. Choi, "Stock Preidciton 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, April, 2018. DOI: https://doi.org/10.17661/jkiiect.2018.11.2.204

[16]H.H. Hwang, Y.H. Kim, Y.H. Lee, "Suggestion of Strategy for the Automation of Stocks Based on Reinforcement Learning", Journal of the Korean Istitute of Industrial Engineers, Vol. 27, No. 4, pp. 399-405, August 2021.

[17]Li, Yuming, Pin Ni, and Victor Chang. "Application of deep reinforcement learning in stock trading strategies and stock forecasting." *Computing*, Vol. 102, No. 6, pp. 1305-1322, 2019, DOI: https://doi.org/10.1007/s00607-019-00773-w

[18]S.H. Moon, "Analysis of AI-Applied Industry and Development Direction", Journal of the Convergence on Culture Technology (JCCT), Vol. 5, No. 1, pp. 77-82, 2019, DOI: http://dx.doi.org./1-0.18005/JCCT.2019.5.1.67

[19]Ts. Tengis, L. Uurtsaikh, A. Batminkh, "Balancing a seesaw with reinforcement learning", International Journal of Advanced Culture Technology (IJACT), Vol. 8, No. 4, pp. 51-57, 2020, DOI: https://doi.org/10.17703/IJACT.2020.8.4.51

[20]H.Y.Jung, "The Effect of Capital Adequacy Requirements on the Profitability of Korean Banks", The Journal of the Convergence on Culture Technology (JCCT), Vol. 7, No. 1, pp. 511-517, February 2021, DOI: http://dx.doi.org/10.17703/JCCT.2021.7.1.511

※ 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음