강화학습 기반 산업 포트폴리오 트레이딩 모델

최우성[1], 손형오[2], 김태용[3], 류병석[4], 권용현[5], 김영균[6]
[1]강원대학교 행정학전공
[2]강원대학교 정보통계학전공
[3]강원대학교 전기전자공학과
[4]연세대학교 화공생명공학과
[5]한양대학교 데이터사이언스학과
[6]융합소프트웨어랩
cwsu0313@naver.com, daily6003@gmail.com, tkddj50999@gmail.com,
bsryu@yonsei.ac.kr, mrkyh380@hanyang.ac.kr, ygkim-2004@hanmail.net

Reinforcement Learning-based Industry Portfolio Trading Model

Useong Choe[1], Hyoengoh Son[2], Taeyong Kim[3], Byeongseok Ryu[4], YongHyun Kwon[5], YoungGyun Kim[6]
[1]Dept. of Public Administration, Kangwon National University
[2]Dept. of Information and Statistics, Kangwon National University
[3]Dept. of Electrical and Electronics Engineering, Kangwon National University
[4]Dept. of Chemical & Biomolecular Engineering, Yonsei University
[5]Dept. of Data Science, Hanyang University
[6]Convergence Software Lab.

요 약

최근 투자 상품 선호도가 증가함과 동시에 투자 지식 부족으로 인한 채무자가 증가하고 있다. 따라서 본 연구는 시장 지표, 재무제표 및 거시 경제 지표 데이터를 활용하여 사전 지식 없이도 활용 가능한 강화학습 기반 주식 투자 전략 모델을 개발하는 것에 목표를 둔다. 2014년부터 2025년까지의 금융 및 재무 데이터 및 자산을 분석하고 미래 성장 가능성을 평가하는 CNN 기반 정책 신경망인 EIIE를 활용하여 포트폴리오 투자를 진행하였다. 실험 결과, 강화학습 기반 모델은 Buy and Hold 전략 대비 유의미한 수익률 개선을 보였다. 이를 통해 EIIE를 활용한 강화학습 모델이 주식 투자에 효과적으로 활용될 수 있음을 입증하였다.

1. 서론

최근 금융자산 운용 트렌드가 변화하고 있는데, 2022년 투자상품 비중이 23.4%에서 2024년 29.5%까지 증가하며 투자에 대한 관심도가 커지고 있다[1]. 하지만 금융 투자 관련 교육의 부족으로 채무를 지게 되는 인구가 증가하고 있다[2]. 따라서 본 연구에서는 투자 지식이 부족한 투자자도 활용할 수 있는 강화학습 기반 투자 모델 구현에 목표를 둔다. 머신러닝 기법 중 하나인 강화학습은 행동 심리학 연구에서 동물들이 보상을 최대화하고자 환경에 적응하는 데에 있어 자신의 경험을 어떻게 이용하는가에 대한 연구에서 파생되었다[3]. 설정된 환경 안에서 에이전트(Agent)는 현재 상태를 인식하고 선택 가능한 행동 중 보상을 최대화하는 행동 또는 행동 순서를 학습한다[4]. 강화학습에 기반한 주식 거래 모델 구현 시 에이전트가 주식시장을 포함한 금융시장 상태에 대해 올바르게 분석하여 해당 시점의 시장 상태에서 최대 이익을 창출하는 행동을 하도록 만드는 것이 주목표이다. 그러나 금융시장의 데이터는 공시 데이터, 사업 내용 등 비정형 데이터가 포함되고, 투자자들의 심리에 많은 영향을 받는다[5]. 이는 데이터의 노이즈를 증가시키기 때문에 모멘텀, 이동평균과 같이 단순히 주가의 변동만을 반영하는 기술 지표를 활용하여 모델을 구현하는 방법을 사용하였다[6]. 하지만 기술 지표를 활용한 주식 거래의 경우, 과거의 주가 패턴이 반복되지 않고, 주가 변동 패턴의 해석이 분석자에 따라 달라지며, 시장 변동에만 집중하여 변화 원인을 분석할 수 없다는 한계점이 있다[7]. 따라서 기존의 기술적 지표, 가격 변동을 사용하는 방법에서 벗어나, 재무지표와 금리, 환율, 상품 가격과 같은 정형화된 수치 데이터를 활용하여 금융 데이터의 문제를 해결 및 원인 분석 측면을 개선한다.

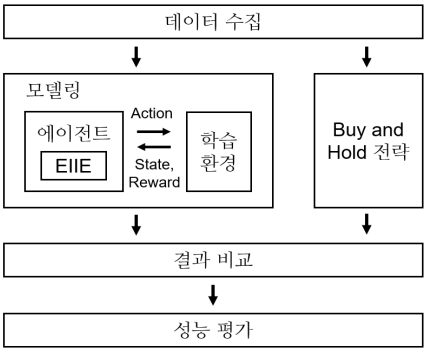
2. 연구설계

2.1 강화학습 모델

본 연구는 강화학습 기반 주식 투자 모델 구현을 목표로 하며, 데이터 수집 후 Python FinRL 라이브러리를 활용하여 모델을 개발하고 동일한 비율로 주식 매수 후 유지하는 Buy and Hold(매수 후 보유) 전략과 결과 비교를 통해 성능을 평가한다. 전체 흐름도는 그림 1과 같다.

강화학습은 기계학습의 한 영역으로 행동심리학을 기반으로 하며, 에이전트가 무지인 상황에서 임의의 행동(Action)을 했을 때 보상(Reward)을 극대화할 수 있는 선택을 하는 과정이다[8]. 강화학습은 마르코프 결정 과정(Markov Decision Process, MDP)을 기반으로 모델링하며, 이는 상태(State), 행동, 보상, 전이 확률(Transition Probability)로 구성된다. 상태는 에이전트가 환경에서 경험하는 상황으로 본 연구에서는 주식의 현재 가격을 포함한 시장 상태로 설정한

다. 행동은 에이전트가 취할 수 있는 특정 상태에서의 선택지이며, 매수, 매도, 유지로 정의한다. 보상은 에이전트가 환경으로부터 받는 특정 행동에 대한 피드백으로, 이는 이익과 손실로 나타낼 수 있고, 전이 확률은 에이전트가 현재 상태에서 임의의 행동을 취했을 때 다른 상태로 이동할 확률로 특정 주식의 매수 후 가격 상승과 하락으로 설정한다[9].



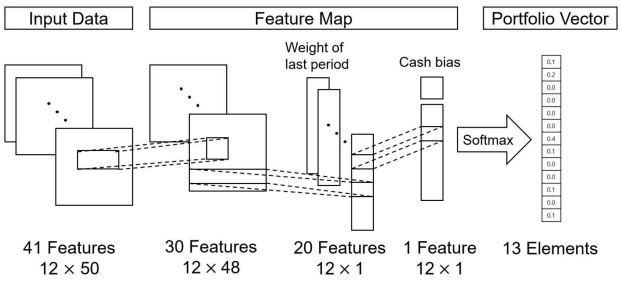
(그림 1) 전체 흐름도

에이전트는 이 마르코프 결정 과정 모델을 기반으로 환경과 상호작용을 하며 각 상태에서 취할 행동을 결정하는 정책(Policy)을 학습한다. 정책을 학습할 때 사용할 강화학습 기법으로 PG(Policy Gradient) 알고리즘을 사용하였다. PG 알고리즘은 가치 기반(Value-based) 강화학습을 사용할 때, 연속적 또는 무한 상태 등의 이유로 최적의 정책을 발견하지 못하는 경우를 보완한 방법이다[10]. PG 알고리즘은 최적의 값을 구하기 위한 Policy 자체를 파라미터로 사용하며, Objective Function을 최대화하는 Theta라는 정책의 파라미터 벡터를 찾아내는데 목표를 두어, 가치 기반 방법에 비해 최적의 값에 더 효과적으로 수렴한다[11].

본 논문에서 사용한 PG 알고리즘은 DDPG(Deep Deterministic Policy Gradient)에서 영감을 받았으나 몇 가지 핵심적인 차이점을 보인다. DDPG는 행동을 결정하는 액터와 그 행동으로 발생한 보상을 통해 행동을 평가하는 크리틱 신경망을 동시에 사용하여 정책과 가치 함수를 추정하는 반면, 본 방법은 크리틱 신경망 없이 포트폴리오 가치를 직접 가치 함수로 활용하여 정책을 갱신한다. 그리고 DDPG는 학습 시 행동에 노이즈를 추가해 탐색성을 확보하지만, 본 알고리즘은 설정된 목표 함수에 집중한 행동 선택 방식을 취한다. 더 나아가, DDPG가 리플레이 버퍼에서 무작위로 경험을 샘플링하는 것과 달리, 본 방법은 시간 순으로 연속된 배치의 경험을 수집하여 해당 배치 내 포트폴리오 가치의 변동을 계산하고, 이를 바탕으로 정책 신경망의 파라미터를 경사상승법으로 갱신한다[12].

2.2 EIIE

EIIE(Ensemble of Identical Independent Evaluators)는 자산의 과거 데이터를 검사하고, 해당 자산의 미래 성장 가능성을 평가하는 CNN 기반 정책 신경망이다. 각 자산의 평가 점수는 자산의 포트폴리오 내 가중치 변화 크기에 의해 조정되고, 향후 거래 기간의 새로운 포트폴리오 가중치로 나타나 강화학습 에이전트의 시장 행동을 정의하게 된다. 목표 가중치가 증가한 자산은 추가 금액으로 매수되고, 가중치가 감소한 자산은 매도되는 방식으로 작용한다. EIIE는 과거 수익성이 좋지 못한 주식을 배제하는 타 정책 신경망의 단점을 보완하기 위해 최근 시점만을 반영하도록 설계되었다. 이를 통해 과거 시점에서는 낮은 수익성을 보였으나 높은 잠재성을 지닌 주식에 투자하여 더욱 높은 수익을 얻을 수 있다[13]. EIIE의 흐름도는 그림 2와 같다.



(그림 2) EIIE 흐름도

2.3 데이터 수집 및 전처리

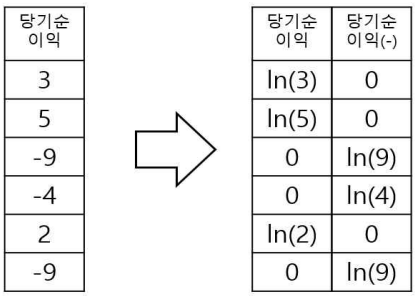
데이터는 FnGuide에서 제공하는 DataGuide를 통해 수집하였다. 투자 대상은 KRX(한국거래소)에서 산출한 산업지수를 기반으로 하는 ETF(상장지수펀드)이다. 해당 산업지수를 구성하는 주식들의 시가총액과 K-IFRS 연결재무제표 기준 17개의 재무상태표 항목, 4개의 포괄손익계산서 항목의 총합을 수집하였다. 또한 7개의 주요 상품 가격 정보와 5개의 채권 수익률 및 3개의 환율 지표를 수집하였다. 데이터 기간은 대부분의 증권업과 보험업 상장사의 결산 월이 12월로 변경된 뒤 첫 공시일인 2014년 4월 2일부터 2025년 1월 31까지의 일별 영업일 데이터이다. 투자 대상 산업은 2014년 4월 2일부터 ETF가 상장된 산업에 한하였다. 건설, 기계 장비, 반도체, 방송·통신, 보험, 에너지화학, 운송, 은행, 자동차, 증권, 철강, 헬스케어로 12개의 산업이 선정되었다.

〈표 1〉 수집데이터

변수구분	변수명
시장지표	ETF 수정주가
	시가총액
재무상태표	총금융부채
	단기금융부채
	단기금융자산
	당좌자산
	매출채권
	무형자산
	비유동부채
	비유동자산
	유동부채
	유동자산
	유형자산
	장기금융부채
	재고자산
	총부채
	총자본
	총자산
	현금및현금성자산
포괄손익계산서	매출액
	영업이익
	당기순이익
	총포괄이익

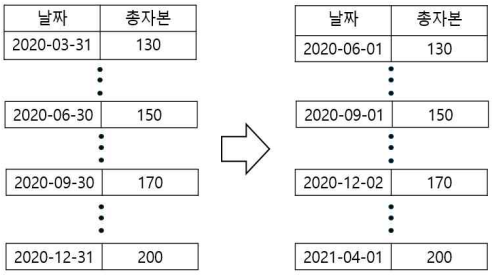
주요 상품 가격	DDR3 4Gb 512M×8 eTT(USD)
	DDR4 16G (2G×8) eTT MHZ(USD)
	금(선물)($/ounce)
	니켈(선물)($/ton)
	DUBAI(ASIA1M)($/bbl)
	소맥(최근월물)(₡/bu)
	전기동(선물)($/ton)
채권 수익률	국채금리_미국국채(10년)
	국채금리_미국국채(1년)
	CD유통수익률
	국고10년
	회사채(무보증3년AA-)
환율	미국(달러)(통화대원)
	일본(100엔)((100)통화대원)
	중국(위안)(통화대원)

우선, ETF 가격을 시작 값으로 정규화를 해주었다. 그 후, 나머지 변수에 대하여 로그 변환을 수행하였다. 총포괄손익, 영업이익, 당기순이익의 경우에는 음수 값이 존재하는데, 0 이하의 값에는 로그변환을 적용할 수 없다. 따라서 양수인 경우 로그 변환을 수행하고 음수인 경우 0으로 설정하였고, 세 변수에 대해 부호를 변환한 파생변수를 생성하여 파생변수에도 양수인 경우 로그 변환을 수행하고 음수인 경우 0으로 설정하였다.



(그림 3) 포괄손익계산서 전처리 예시

여기에 더해 재무제표 정보의 경우, 해당 분기의 정보 발표까지 시차가 존재한다. 따라서 공시 제출 기한을 연장할 경우까지 고려하여 해당 정보가 1분기는 6월, 2분기는 8월 첫 영업일에 확인되도록 하였고 3분기는 12월 두 번째 영업일, 4분기는 4월 첫 영업일에 확인되도록 시차를 두었다. 2014년 4월 2일부터 2021년 3월 31일까지를 학습데이터로 설정하였고 2021년 4월 1일부터 2025년 1월 31일까지를 테스트 데이터로 설정하였다.



(그림 4) 날짜 전처리 예시

2.4 모델링 및 평가

모델은 2014년 4월 1일부터 2021년 3월 31일까지 과거 데이터를 학습한 후, 향후 2021년 4월 1일부터 2025년 1월 31일 동안의 데이터를 통해 주식 매매를 진행하도록 설계하였으며 모델의 파라미터는 표 2와 같다.

〈표 2〉 모델 파라미터

Time Window	15
입력 피처 수	41
중간 피처 수	30
최종 피처 수	20
커널 사이즈	3
배치 사이즈	128
에피소드	200
보상 스케일	0.99995
거래비용	0.0012

전략의 유효성은 〈표 3〉의 성과지표를 이용해 Buy and Hold 전략과 비교하여 평가를 진행하였다. Sharpe 비율은 보편적인 포트폴리오 성과지표로 계산 방법은 식 1과 같다. R_a는 자산 수익률, R_b는 무위험 자산 수익률, $E[R_a - R_b]$는 초과 수익률, σ_a는 자산 수익률의 표준편차(변동성)를 의미한다. 연구에선 무위험 수익률을 연 2%로 가정한다.

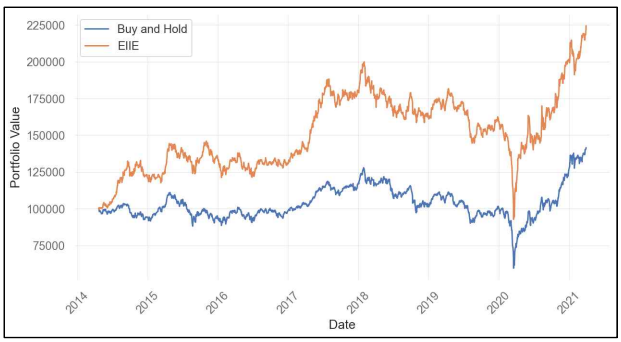
$$Sharp_a = \frac{E[R_a - R_b]}{\sigma_a} \quad \text{(식 1)}$$

MDD(Max Drawdown)는 투자 기간 동안 포트폴리오 가치가 도달한 최고점 대비 최고점 이후의 최저점의 비율을 의미하며, 해당 주식의 위험성을 판단하는데 중요한 지표로 활용된다[14]. MDD를 통해 각 전략이 일정 기간동안 얼마나 안정적인 투자를 진행했는지를 확인할 수 있다. Peak Value는 주어진 기간 내 가장 높은 값을 의미하고, Trough Value는 Peak Value 이후의 가격 중 가장 낮은 값을 의미한다. 계산 방법은 식 2와 같다.

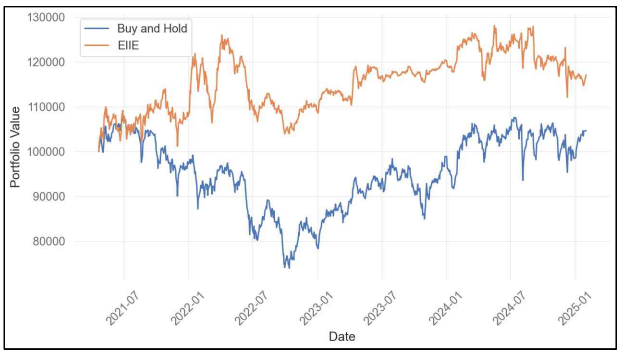
$$MDD = \frac{Peak\ Value - Trough\ Value}{Peak\ Value} \quad \text{(식 2)}$$

Omega 비율은 투자 전략의 수익률 분포를 기반으로 한 위험-수익률 평가 지표이다. $F(x)$는 수익률 x에 대한 누적 확률 분포 함수를 뜻하며, 계산 방법은 식 3과 같다. 수익률 분포 전체를 반영하여 기준 수익률(Threshold, θ) 이상과 이하의 누적 확률을 각각 고려한다. 연구에선 θ를 연 2%로 설정한다.

$$\Omega(\theta) = \frac{\int_{\theta}^{\infty} [1 - F(x)]dx}{\int_{-\infty}^{\theta} F(x)dx} \quad \text{(식 3)}$$



(그림 5) 학습 기간 누적 가치



(그림 6) 테스트 기간 누적 가치

그림 5와 그림 6에서 Portfolio Value는 100,000원을 투자했을 시 각 전략에 따라 투자한 ETF의 누적 가치를 나타낸다. 그림에서 확인할 수 있듯이 학습 기간과 테스트 기간에서 모두 EIIE의 누적 가치가 더 높음을 확인할 수 있다.

〈표 3〉 성과지표

기간 구분	성과지표	EIIE	Buy and Hold
학습기간	Sharpe	**0.488**	0.173
	Omega	**1.092**	1.033
	MDD	−0.535	**−0.534**
	누적수익률	**2.244**	1.414
	연평균성장률	**0.083**	0.0350
테스트 기간	Sharpe	**0.162**	−0.042
	Omega	**1.030**	0.992
	MDD	**−0.175**	−0.312
	누적수익률	**1.171**	1.046
	연평균성장률	**0.029**	0.008

〈표 3〉에서 확인할 수 있듯, 강화학습 기반 포트폴리오 트레이딩 모델이 Buy and Hold 전략에 비해 학습기간의 MDD를 제외한 지표에서 더 높은 성과를 기록하였다. 특히 테스트 기간에서 Buy and Hold 전략이 음의 Sharpe 비율과 1보다 작은 Omega 비율을 보이며 저조한 성과를 거둔 것에 반해 EIIE는 약세장으로 보이는 테스트 기간에서 Buy and Hold 전략보다 더 좋은 수익률을 기록하고 더 낮은 MDD를 보여 위험도 또한 개선됨을 확인할 수 있다.

3. 결론

본 연구에서는 시장, 재무지표 및 거시 경제 지표 데이터를 이용하여, 강화학습 기반 포트폴리오 트레이딩 모델을 구현하였다. 실험 결과, Buy and Hold 전략과 비교하여 Sharpe 비율은 0.204, Omega 비율은 0.038, MDD는 0.137, 누적수익률은 0.125, 연평균성장률은 0.021 정도 개선되었음을 확인하였다. 이는 구현 모델이 정책에 따른 동적 주식 투자를 통해 보다 안정적이고 높은 수익률을 달성했음을 의미하며, 연구 목표인 뛰어난 성능의 투자 모델 제공 가능성을 시사한다. 향후 연구에서는 다른 강화학습 기법 도입 및 새로운 재무, 경제 지표 추가를 통해 더 높은 수익률을 달성하여 실무적 용이성을 제고하고자 한다.

참고문헌

[1] 하나금융연구소, “대한민국 금융소비자 보고서 2025”, p.6, 2025

[2] 송기영, “[청년빈곤시대]⑥ 20대 금융이해력 49점… 범죄·사기 노출된 금융문맹 청년층”, 조선일보, 2024, https://biz.chosun.com/stock/finance/2024/01/27/BOMZDHF2NREVVPMFMDFBRMM5EM/

[3] 김영삼, “강화 학습을 이용한 단어 감정 값 및 진술문 상태 값 측정법 연구”, 서울대학교 대학원 박사학위 논문, p.2, 2018

[4] 푸 빼잉 송 외 5인, “모바일 에지 컴퓨팅 네트워크의 연합 심층 강화 학습을 활용한 서비스 블록 오프로딩 결정 기법”, 정보과학회 컴퓨팅의 실제 논문지 제31권 제1호, p.6, 2025

[5] 김정우, “주식 시장에서의 인공지능 활용 및 연구”, 정보과학학회지 제42권 제3호, p.47, 2024

[6] Yawei Li, Peipei Liu, and Ze Wang, “Stock Trading Strategies Based on Deep Reinforcement Learning”, Scientific Programming, Volume 2022, p.1, 2022

[7] 조현, “기술적 지표를 이용한 상해와 심천 주식시장의 효율성에 관한 실증 연구”, 우석대학교 대학원 석사학위 논문, pp.9-10, 2014

[8] 조현민, 신현준, “강화학습을 이용한 트레이딩 전략”, 한국산학기술학회논문지 22권 1호, p.125, 2021

[9] 이보미, “강화학습을 이용한 주가 예측”, 한양대학교 대학원 석사학위논문, pp.8-11, 2018

[10] 이요셉, 김효진, 이창환, “Policy Gradient 강화학습을 이용한 대화생성에서 다양한 리워드 함수의 적용”, 한국컴퓨터종합학술대회 논문집, p.2084, 2018

[11] 김주희 외 5인, “Cart-pole 모델에서 DDQN과 Policy gradient 강화학습 방법 비교”, 한국통신학회 추계종합학술발표회, p.381, 2019

[12] Zhengyao Jiang, Jinjun Liang, “Cryptocurrency portfolio management with deep reinforcement learning”, Intelligent systems conference, p.4, 2017

[13] Zhengyao Jiang and Dixing Xu and Jinjun Liang, “A Deep reinforcement Learning Framework for the Financial Portfolio Management Problem”, pp.2-3, 2016

[14] 김대환, “Relevance of Maximum Drawdown in the Investment Fund Selection Problem When Utility is Nonadditive”, Journal of Economic Research Vol.16 No.3, pp.258-259, 2011