

2023 전기 착수보고서

심층 강화학습을 사용한 주식투자 전략 개발



제출일	2023. 05. 12	전공	정보컴퓨터공학부
		담당교수	유영환 교수님
학번	201824586	이름	정희영
	201824469		박동진
	201824514		신재환

< 목차 >

1. 과제 배경 및 목표	3
1-1. 과제 배경	3
1-2. 과제 목표	3
 2. 과제 내용 및 시나리오	 4
2-1. 과제 내용	4
2-2. 시나리오 및 시스템 흐름도	9
 3. 기대효과	 9
 4. 개발 일정 및 역할 분담	 10
4-1. 개발 일정	10
4-2. 역할 분담	10
 5. 참고 자료	 11

1. 과제 배경 및 목표

1-1. 과제 배경

2022년 11월 처음으로 세상에 공개된 ChatGPT는 미국의 AI 회사인 Open AI에서 개발한 인공지능 서비스로 최근 전세계적으로 이슈가 되고 있다. ChatGPT 이전에도 많은 대화형 인공지능 서비스가 존재해왔지만, ChatGPT는 다른 인공지능과 달리 인터넷에서 답을 찾지 않고 지도학습(Supervised Learning)과 강화학습(Reinforcement learning)을 통해 데이터를 학습한다. 강화학습은 결과를 극대화하기 위해 시행착오 방법을 사용하여 동적으로 학습하는데, 이를 활용해 주식투자에 적용하면 실용적일 것이라 판단되었다. 따라서, 기존 지식에서 학습하여 새로운 데이터셋에 적용할 수 있는 **심층 강화 학습(Deep Reinforcement Learning, DRL)**을 사용하여 **주식투자 전략 개발**을 해보고자 한다.

1-2. 과제 목표

지금까지 개발된 다양한 강화학습 기법과 신경망들을 이용해 강화학습을 수행하여, 주식투자 전략에 가장 적합한 강화학습 기법을 발견한다. 채택된 기법에 대해서는 다양한 hyperparameter를 이용하며 모델을 생성, 평가하여 가장 적합한 매개변수를 판별해 가장 적합한 모델을 찾는다. 가장 적합한 모델의 우열을 가릴 수 없을 경우 ensemble learning과 같은 기법을 통해 여러 모델들을 결합하는 방식으로 최종적인 결과를 내는 방식도 고려할 것이다.

또한 주식투자는 종목, 분야에 따라서 개별적인 변수가 존재하기 때문에 모든 종목에 같은 학습방법을 적용할 수는 없을 것이다. 따라서 종목에 따라 다른 학습방법을 선택하는 것이 필요하다고 판단했다. 하지만 존재하는 모든 종목에 대응하는 모델을 위한 적합한 기법, 신경망, 매개변수를 평가하는 데에는 많은 시간과 자원이 소모된다. 때문에 비슷한 추세를 따르는 종목들을 하나의 섹터로 묶어 실현 가능한 수만큼 모델생성방식을 결정하고자 한다. 각 섹터별로 가장 적합한 학습방법을 찾고, 사용자가 원하는 종목의 섹터에 따른 요소들과 종목의 데이터로 모델을 만들어 낸 후, 해당 모델을 이용해 예측 결과를 사용자에게 전달하는 것이 최종적인 목표이다.

데이터 수집 과정에서 사용자가 요구하는 모든 종목에 대응할 수 있도록 모든 데이터를 보관할 수도 있겠지만, 최신 데이터에 따른 결과를 사용자에게 제공하기 위해서 사용자의 요청과 동시에 증권사 API를 통해 학습에 필요한 데이터를 습득, 가공, 학습하여 적합한 모델을 생성하고, 예측된 결과를 사용자에게 제공한다.

위에서 언급한 모든 백엔드개발이 완료된 후, 해당 시스템을 이용할 수 있는 웹 인

터페이스를 개발한다. 직관적인 UI를 통해 누구나 쉽게 서비스에 접근할 수 있도록 할 것이다.

2. 과제 내용 및 시나리오

2-1. 과제 내용

■ 수집할 데이터

A. 기본적 분석 지표

- 주가수익비율(Price Earning Ratio, PER)

PER은 주가와 회사의 순이익의 비율이다. PER이 낮을수록 회사의 순이익에 비해 주가가 저평가돼 있다고 볼 수 있다. 그러나 업종마다 평균 PER의 차이가 크기 때문에 PER이 높다, 낮다라고 절대적으로 판단할 기준은 없다.

$$\text{PER} = \frac{\text{주가}}{\text{EPS (주당순이익)}} = \frac{\text{주가}}{\frac{\text{당기순이익}}{\text{주식수}}}$$

그림 1 주가수익비율(PER) 계산공식

- 주가순자산비율(Price Book-value Ratio, PBR)

PBR은 자산 대비 주가의 비율이다. PBR이 낮을수록 회사의 가치 대비 주가가 낮다는 것이므로 해당 종목이 저평가돼 있을 가능성이 있다.

$$\text{PBR} = \frac{\text{주가}}{\text{BPS (주당순자산)}} = \frac{\text{주가}}{\frac{\text{총자산-총부채}}{\text{발행주식수}}}$$

그림 2 주가순자산비율(PBR) 계산공식

- 주당순자산가치(Bookvalue Per Share, BPS)

BPS는 기업이 활동을 중단한 뒤 그 자산을 모든 주주들에게 나눠줄 경우 1주당 얼마씩 배분되는가를 나타내준다. 기업의 총자산에서 부채를 빼면 기업의 순자산이 남는데, 이 순자산을 발행주식수로 나눈 수치이다. BPS가 높을수록 수익성 및 재무건전성이 높아 투자가치가 높은 기업이라 생각할 수 있지만 BPS 지표를 사용하기보다는 주가와 청산 가치 사이의 비율을 알 수 있는 PBR 값을 이용하는 것이 더 효과적이다.

$$\text{BPS} = \frac{\text{주가}}{\text{PBR}} = \frac{\text{총자산-총부채}}{\text{발행주식수}}$$

(주당순자산가치) (주당순자산비율)

그림 3 주당순자산가치(BPS) 계산공식

- 주당순이익(Earning Per Share, EPS)

EPS는 기업의 순이익을 발행한 주식 수로 나눈 값을 의미한다. PER과 함께 가치투자의 대표적인 지표이다. EPS가 주가보다 낮다고 가정하면, EPS 값이 높아질수록 주가에 수렴하기 때문에 고평가되었던 종목이 저평가 단계로 다가가기 때문에 투자가가 높아 가치투자자 입장에서는 매력적으로 볼 수 있다. EPS가 높다는 것은 그만큼 경영실적이 양호하다는 뜻이고, 배당 여력도 많으므로 주가에 긍정적인 영향을 미친다.

$$\text{EPS} = \frac{\text{주가}}{\text{PER}} = \frac{\text{주가}}{\text{발행주식수}}$$

(주당순이익) (주가수익비율)

그림 4 주당순이익(EPS) 계산공식

- 자기자본이익률(Return On Equity, ROE)

ROE가 높을수록 자산 대비 이익이 크다는 의미이다. 그러므로 ROE가 높은 종목은 회사의 성장 가능성이 높다고 볼 수 있다.

$$\text{ROE} = \frac{\text{당기순이익}}{\text{자기자본}} \times 100$$

(자기자본이익률)

그림 5 자기자본이익률(ROE) 계산공식

B. 기술적 분석 지표

- 전일 종가 대비 당일 시가 비율(open/last close)

당일 시가를 전일 종가로 나눈 값이다. 주로 전일 종가에 비해 당일 시가가 상당히

높으면 '갭상승'이라 하고, 반대로 전일 종가보다 당일 시가가 많이 떨어져 있으면 '갭하락'이라 부른다.

- 당일 종가 대비 전일 종가 비율(close/last close)

당일 종가를 전일 종가로 나눈 값이다. 일반적으로 주가가 올랐다, 떨어졌다는 얘기할 때 당일 종가에서 전일 종가를 뺀 다음에 다시 전일 종가로 나누는데, 수식으로는 $(\text{close} - \text{last close}) / \text{last close}$ 가 된다.

- 전일 거래량 대비 당일 거래량 비율(volume/last volume)

당일 거래량을 전일 거래량으로 나눈 값이다. 주가가 상승하면서 거래량이 늘면 주가 상승에 긍정적인 신호로 여기는데, 이를 포착하고자 이 값을 학습 데이터의 특징으로 추가했다.

■ 데이터 수집방법

A. 증권사 HTS 사용

- 해당 증권사의 계좌 개설 필수
- 데이터를 엑셀로 저장가능
- 실시간으로 데이터를 가져오긴 힘들다

B. 증권사 API 사용

	이베스트증권 xingAPI	대신증권 CYBOS / CREON Plus	키움증권 KOA+	한국투자증권 eFriend Expert
API 연결방식	COM, DLL	COM	OCX	-
요청 제한	10/s	4/s	5/s	-
수수료	금액/수수료/제세금 매도50백 /7,556/116,066	금액/거래비용 매수49백/4,294 매도49백/117,373	금액/수수료+세금 매도45백/117,520	금액/수수료/세금 매수39백/1,407 매도39백 /1,412/89,931
HTS 스탑로스	서버주문 지원 잔고자동편입주문 지원	서버주문 지원(수수료) 잔고자동편입주문 지원x	서버주문지원 잔고자동편입주문 지원	-
HTS 복기차트	-	분단위	틱단위	-
고객지원 재접속 시간	활성화 07:10	활성화 06:00	활성화 06:50	1:1 문의만 가능 -

■ 학습 방법

A. 강화학습 알고리즘

1. DQN (Deep Q-Network)

Q-learning과 딥러닝을 합친 것인데 Q-learning에서는 Q-table을 이용해 Q-value를 저장한다. 이때 state space와 action space의 크기가 커지게 되면 모든 Q-value를 저장하기 위해선 많은 메모와 탐색 시간이 요구된다. 이 문제점을 막기 위해서 딥러닝을 사용한다. 즉, 딥러닝을 이용해 Q-table에 해당하는 Q-function을 근사시켜서 Q-value를 구하는 것이다.

· DQN의 특징

i) Experience Replay

샘플 간의 correlation 문제를 해결하기 위해서 sample data들을 replay buffer에 저장하고 random하게 추출해서 agent가 학습하고 신경망을 update한다.

ii) Target Network

하나의 network를 사용하지 않고 분리된 target network를 두어 update하는 동안에 target을 계산하는 과정에서 계산에 사용되는 parameter를 고정하는 것이다. 일정 step동안엔 target network는 업데이트 하지 않고 학습을 진행하고 일정한 step마다 학습이 진행된 네트워크로 target network를 업데이트한다.

2. Actor-Critic

2개의 네트워크 즉, Actor 네트워크와 Critic 네트워크를 사용한다. Actor는 상태에 따라 행동을 결정하고, Critic은 상태의 가치를 평가한다. Replay Buffer를 사용하지 않고, 매 step마다 얻어진 상태, 행동, 보상, 다음 상태를 이용해서 모델을 학습시킨다.

3. A2C (Advantage Actor-Critic)

Actor-Critic의 Actor의 기대 출력으로 Advantage를 사용하면 A2C가 된다. 여기서 Advantage는 예상했던 것보다 얼마나 더 좋은 값인지를 판단하는 값이다.

$$A(s, a) = Q(s, a) - V(s)$$

4. A3C (Asynchronous Advantage Actor-Critic)

여러 개의 에이전트를 각각 다른 환경에서 학습을 진행시키고 비동기적으로 공유 네트워크를 업데이트 한다. 쉽게 말해서 A2C 에이전트 여러 개를 독립적으로 실행시키고 얻은 샘플들을 통해 공유 네트워크인 Global network와 학습 결과를 주고받으면서 Global network를 업데이트하는 것이다.

각 에이전트가 다른 환경에서 학습을 진행하기 때문에 샘플 간의 correlation 문제를 극복할 수 있다는 장점이 있다.

B. 모델(신경망)

1. LSTM (Long Shor-Term Memory) 순환 신경망

기존의 RNN은 단기기억에는 적합하지만 장기기억 성능이 떨어지는데 이를 개선하고, RNN에 기울기 정보 크기를 조절하기 위한 gate를 추가한 모델이다. RNN과 동일하게 hidden state를 가지지만 활성화 함수를 직접 거치지 않는 상태인 cell-state가 추가되었고 cell-state는 출력으로 전달되지 않고 cell에서 cell로 순환만 되는 정보이다.

· LSTM gate종류

- i) forget gate : 과거 정보를 얼마나 유지할 것인지 결정
- ii) input gate : 새로 입력된 정보를 얼마만큼 유지/활용할 것인지 결정
- iii) output gate : 입력 정보(forget gate + input gate)를 계산하여 나온 출력 정보를 다음 셀로 얼마만큼 넘겨줄 것인지 결정

2. GRU (Gated Recurrent Unit)

GRU는 LSTM와 마찬가지로 장기기억에도 좋은 성능을 보이고 LSTM보다 학습 속도가 더 빠르지만 성능은 비슷하다.

· GRU gate종류

- i) reset gate : 이전 상태를 얼마나 반영할 것인지 결정 즉, 이전 hidden state의 값을 얼마나 활용할 것인지에 대한 정보이다.
- ii) update gate : 이전 상태와 현재 상태를 얼마만큼의 비율로 반영할 것인지 결정

LSTM에서의 cell-state와 hidden state가 GRU에서는 hidden state로 통합되었고, GRU에서의 update gate가 LSTM에서의 forget gate와 input gate를 제어한다.

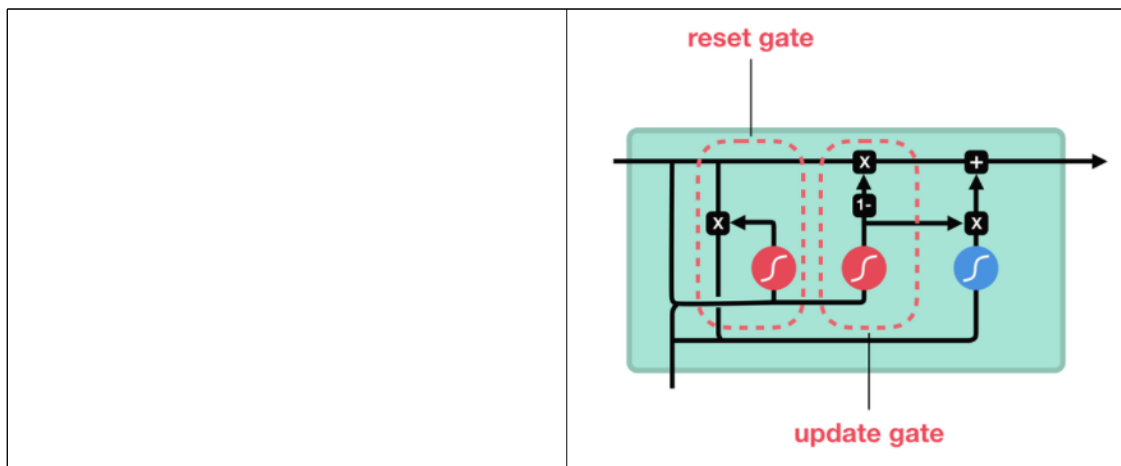


그림 6 LSTM(左) GRU(右) 구조

2-2. 과제 시나리오 및 시스템 흐름도

- ① 사용자가 웹 인터페이스에 접속하여 조언을 구할 종목을 입력한다.
- ② 시스템은 증권사 API에 요청하여 해당 종목과 관련된 데이터를 획득한다.
- ③ 획득한 데이터에서 필요한 데이터를 선별하고, 학습에 적합한 형태로 가공한다.
- ④ 가공된 데이터를 이용하여 모델을 학습한다. 강화학습 방법과 신경망, 결과도출방법과 같은 모델의 상세한 정보는 과제를 진행하면서 가장 적합한 것으로 결정할 것이다.
- ⑤ 학습된 모델을 이용하여 해당 주식에 대한 투자조언을 사용자에게 제공한다.

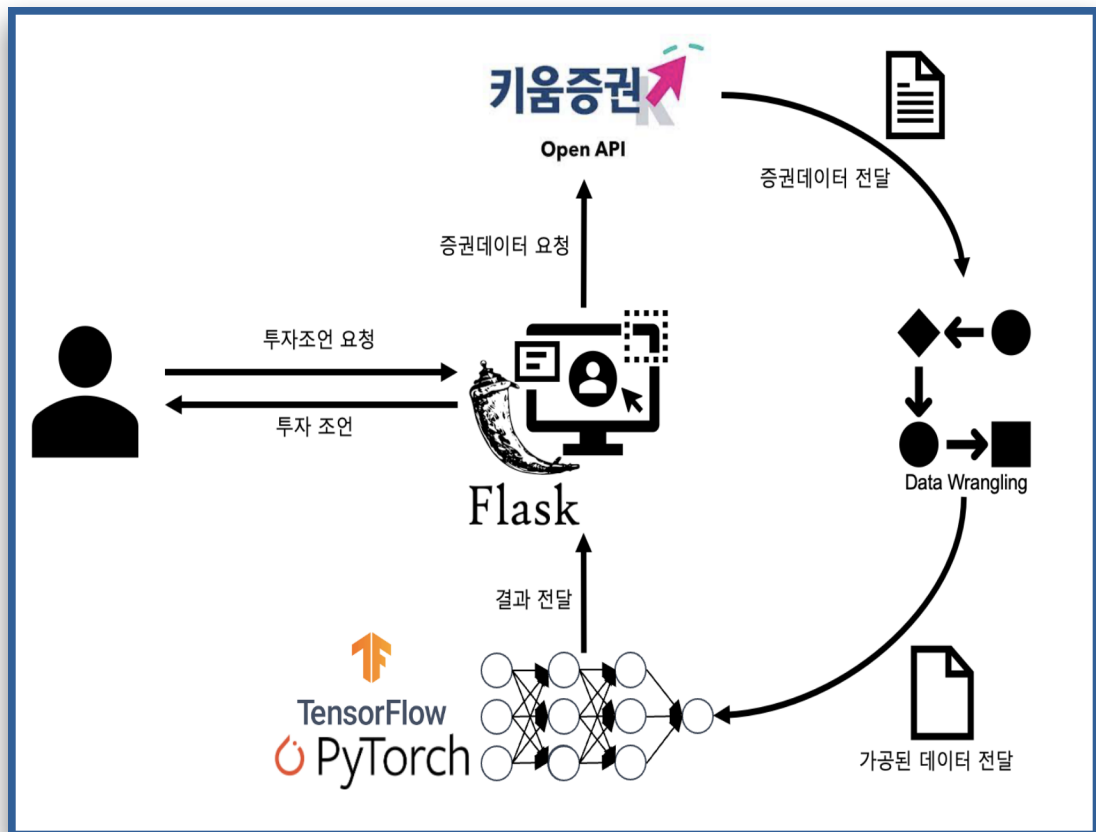


그림 7 시스템 흐름도

3. 기대효과

최근 들어 많은 사람들이 재테크에 관심을 가지고 특히 주식에 관심을 갖고 입문하는 사람들이 늘고 있다. 처음 주식을 시작하는 사람들은 투자에 대한 정보도 부족하고 자신만의 투자 기준이 정립되어 있지 않으므로 성공적인 투자를 하기에는 어려움이 있다. 이러한 입문자들을 위한 주식 투자 길라잡이 역할을 수행할 것으로 기대한다. 또한, 심층 강화학습이 주식투자에도 활용될 수 있다는 것을 보임으로써 앞으로 다양한 분야에서도 유용하게 활용될 것으로 예상된다.

4. 개발 일정 및 역할 분담

4-1. 개발 일정

6월		7월					8월					9월				
4주	5주	1주	2주	3주	4주	5주	1주	2주	3주	4주	5주	1주	2주	3주	4주	5주
데이터 수집방법 모색																
	데이터 수집															
		데이터 가공														
			적합한 기법/신경망 선별													
			중간보고서 작성													
							적합한 기법/신경망 구현									
										모델 학습 및 성능평가						
												UI 제작 및 최적화				
														최종보고서 작성 및 발표준비		

4-2. 역할 분담

이름	역할
정희영	<ul style="list-style-type: none"> - 적합한 강화학습기법/신경망 선별 - 학습에 적합한 hyperparameter 탐색 - 학습에 필요한 데이터로 변환
박동진	<ul style="list-style-type: none"> - 사용자 인터페이스 개발 - 적합한 강화학습기법/신경망 선별 - 학습에 적합한 hyperparameter 탐색
신재환	<ul style="list-style-type: none"> - 파이프라인을 위한 csv파일 변환 - 실시간 데이터 수집 방법 개발 - 수집할 주식투자 데이터 선별
공통	<ul style="list-style-type: none"> - 필요한 지식 습득 - 보고서 작성 - 발표 및 시연 준비

5. 참고 자료

1. 쿼티랩. (2022). 파이썬을 이용한 딥러닝/강화학습 주식투자(개정2판) : 파이토치와 케라스를 활용한 인공지능 퀀트 투자 시스템. 위키북스.
2. 전병조, & 노동건. (2021). 다양한 강화학습 기법을 이용한 업종별 주가예측 비교. 한국통신학회 학술대회논문집, 457-458.
3. 박정연, 홍승식, 박민규, 이현. (2021). 강화학습 기반 주식 투자 웹 서비스. 문화기술의 융합, 7(4), 807-814.
4. 조현민, & 신현준. (2021). 강화학습을 이용한 트레이딩 전략. 한국산학기술학회 논문지, 22(1), 123-130.
5. 김영숙. “기본적/기술적 분석을 통한 주식투자전략에 관한 연구.” 국내석사학위논문 한양대학교 산업경영디자인대학원, 2008. 서울