강화학습을 이용한 포트폴리오 투자 프로세스 최적화에 대한 연구

손형진 [1], 임동휘 [1], 한영우 [2]

[1] 경희대학교 산업경영공학 학부생
[2] 한국예탁결제원

fromson99@gmail.com, ukedonghui@gmail.com, ywhan6@gmail.com

Reinforcement learning portfolio optimization based on portfolio theory

Hyeong-Jin Son[1], Lim Donhui[1], Young-Woo Han[2]
[1]Dept. of Industrial and Management Engineering, Kyunghee University
[2]Korea Securities Depository

요 약

포트폴리오 구성문제는 과거부터 현재까지 많은 연구가 이루어지고 있다. 현재는 강화학습을 통해 포트폴리오를 구성하는 연구가 많이 진행되고있다. 포트폴리오를 구성함에 있어 종목선택과 각 종목을 얼만큼 투자할 것인지는 둘 다 중요한 문제이다. 본 연구에서는 과거부터 많이 사용해오던 방식을 차용하여 강화학습 방법과 접목시켰고 이를 통해 설명력이 높은 모델을 만들려고 노력하였다. 강화학습에 사용한 모델은 PPO(Proximal Policy Optimization)을 기본으로 하였고 인공신경망은 LSTM을 활용하였다. 실험결과 실험 기간 동안(2023년 3월 30일 부터 108 영업일 까지)의 코스피 수익률은 5%인데 반해 본 연구에서 제시한 모델의 수익률은 평균 약 9%를 기록했다.

1. 서론

여러가지 자산군을 선택해 포트폴리오를 만들고 각 자산군의 투자 비율을 정하는 행위는 금융 투자의 중요한 프로세스이며 현재까지 다각도에서 분석되고 있다. 최근에는 강화학습을 적용하여 각 자산군의 투자 비중을 정하거나[1], 팩터 투자 관련 연구가 증가하고 있다[2]. 하지만 두 가지 방법을 모두 강화학습 기반으로 적용한 사례는 현재까지 매우 드물었다. 이번 연구에서는 기존의 팩터투자 이론을 바탕으로 포트폴리오 투자 프로세스를 강화학습을 통해 구현해 성능과 설명력을 높이려고 노력했다.

2. 본론

2.1 포트폴리오 이론

포트폴리오를 구성함에 있어 어떠한 자산군을 가지고 포트폴리오를 구성할지 각 자산을 얼마만큼 투자할지는 중요한 문제이다. 본 연구에서는 전자는 팩터투자 이론을 통해 후자는 포트폴리오 최적화 방법을 사용 해결한다.

팩터투자에는 여러가지 방법이 있지만 본 연구에선 퀄리티(ROE, GPA, CFO), 밸류(PER, PBR, PSR, PCR, DY), 모멘텀(K-Ratio) 지표들을 사용해서 종목들을 선정했고, 강화학습을 통해 각각의 지표들의 비중을 결정했다. 또한 팩터투자 모델을 통해 선정된 주식들을 가지고 일일투자를 진행 하면서 포트폴리오 최적화를 진행하였다.

기존에 투자하는 종목을 변경하거나 각 종목의 투자 비중을 조정하는 것을 재조정이라고 하는데 팩터 모델의 출력을 통해 팩터 지표들의 가중치를 결정하고 이를 통해 종목을 재선정 한다. 자산 투자비중은 포트폴리오 모델의 출력 값을 통해 재조정한다. 본 논문에서 팩터 모델은 50 영업일 간격으로 포트폴리오 모델은 하루 간격으로 재조정을 진행했다.

2.2 강화학습

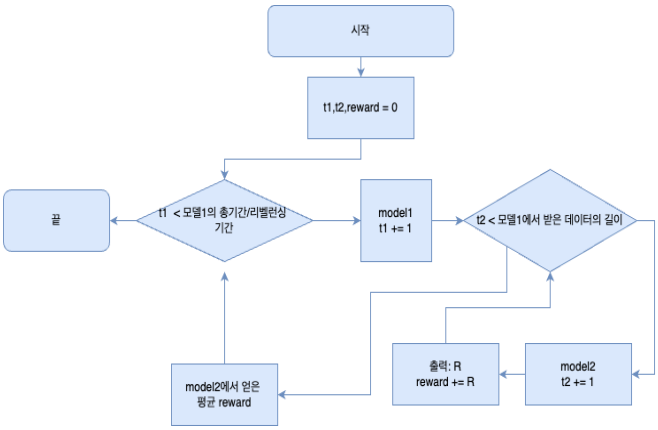
팩터투자모델(이하 모델1)과 포트폴리오투자모델(이하 모델2)에 적용된 강화학습 알고리즘은 PPO(Proximal Policy Optimization)이다. 강화학습에서 중요한 요소는 환경과 보상이다. MDP(Markov Decision Process) 이론에 기반한 강화학습은(state - action

- reward) (state' …)으로 구성된다.

<그림1>은 강화학습 기반 포트폴리오 최적화 흐름도이다. 모델 1과 모델2는 입력 받는 시간간격이 서로 다르기에 보상을 기존 연구들과 다르게 설계하였다. 본 연구에선 모델 1의 보상(reward)를 계산하기 위해 모델2 학습을 진행했다.

<표1>은 본 연구에서 제안한 알고리즘에 대한 의사순서도이다. 모델 1에서는 총기간/리벨런싱 값만큼 기간만큼 순환한다. 모델2에서는 리벨런싱 기간만큼 순환한다. 모델1의 보상(reward)는 모델2에서 얻은 reward의 평균값이다.

<그림1 강화학습 기반 포트폴리오 최적화 흐름도>



<표1 알고리즘 의사순서도>

의사코드(PSUEDO CODE)

RL1

한 마르코프 과정 (s_t, a_t, r_t, s'_t)

For episode = 1,E do

For all model1 rebalancing time t = 1,T do

Step: Input data가 model1에 들어가 팩터 가중치 출력: a_t

해당 출력값을 통해 20 개의 종목을 선정

25 개의 종목을 통해 RL2 모델을 학습

RL2 모델의 전체 에피소드의 평균 점수: r_t

20 일 이후 데이터로 model1 학습진행: s'_t

RL1 모델 파라미터 업데이트

RL2

For episode = 1,E do

For all model2 rebalancing time t = 1,T do

RL1에서 받은 각 종목의 가중치 출력: a_t

포트폴리오의 가중치로 weighted return 계산: r_t

하루 이후 데이터를 가지고 model2 진행: s'_t

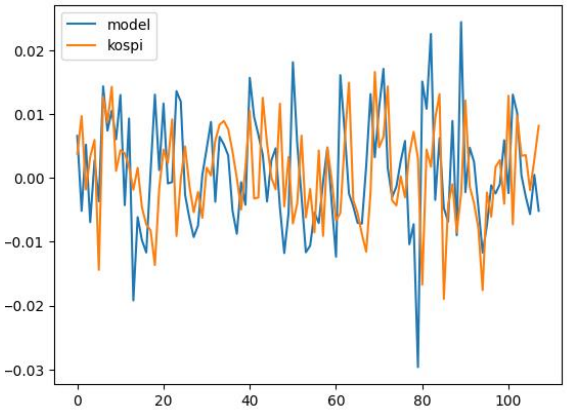
2.3 실험

2.3.1 데이터셋

전체 데이터 셋은 2023년 3월 30일 부터 108 영업일 까지 코스피(KOSPI)에 상장된 종목들과 그 종목들의 종가를 사용하였다.

모델1에는 54 영업일 데이터가 들어가며 이를 통해 팩터의 가중치를 구한다. 가중치를 통해 20개의 종목을 선택하면 20일 데이터를 가지고 모델2에 넣는다. 모델2의 모든 에피소드들의 점수의 평균을 모델 1의 한 step의 보상(reward)로 설정한다.

2.3.2 실험 결과



<그림2 108영업일간 수익률,X축 영업일,Y축 수익률>

<그림2>결과에 따르면 실험 기간동안(2023년 3월 30일 부터 108 영업일 까지) 코스피 수익률은 5%인데 반해 본 연구에서 제시한 모델의 수익률은 평균약 9%를 기록했다.

3. 결론

3.1 성능

벤치마크인 코스피의 수익률은 3% 모델 수익률은 9% 정도 였으며 코스피와의 상관계수는 약 0.1 이였다.

3.2 한계점 및 발전방향

투자기간을 108 영업일을 기준으로 잡았기에 기간을 늘리면 성능이 차이 날 수 있다. 강화학습을 두개를 동시에 진행하면서 과최적화가 많이 일어났다. 향후 해당 문제에 대한 추가 연구가 필요해보인다.

참고문헌

[1] Amine Mohamed Aboussalah "Continuous control with Stacked Deep Dynamic Recurrent Reinforce ment Learning for portfolio optimization" Expert Systems with Applications, Volume 140, 2020, 11 2891

[2] Jennifer Bender, Remy Briand, Dimitris Melas, Raman Aylur Subramanian "Foundations of Factor Investing" MSCI, Research Insight, December 2013

※본 프로젝트는 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.