시장 예측값을 사용하여 포트폴리오를 위한 재귀 강화학습 알고리즘의 성능 향상을 위한 연구 1)

강문주*, 이주홍*, 안준규*
*인하대학교 컴퓨터공학과
sktel1020@nate.com
juhong@inha.ac.kr
ahnjungyu320@gmail.com

A Study on Performance Improvement of Recurrent Reinforcement Learning Algorithm for Portfolio Using Market Forecast

Moon-Ju Kang*, Ju-Hong Lee*, Jungyu Ahn*
*Dept of Computer Engineering, Inha University

요 약

최근, 자산 매매 및 포트폴리오에 인공지능을 활용한 연구들이 활발히 진행되고 있다. 본 논문은 기존 재귀 강화학습(Recurrent Reinforcement Learning)을 기반으로 한 운용 모델의 성능을 향상시키고자 자산들의 예측값을 사용한다. 예측값 사용 유무에 따른 재귀 강화학습의 성능을 비교분석을 통하여 예측값의 활용이 포트폴리오 운용 성능에 미치는 효과에 대해 분석하였다.

1. 서론

본 논문에서는 기존 포트폴리오를 구현하는 많은 모델 중에서 재귀 강화학습(Recurrent Reinforcement Learning, RRL)을 사용하였다.

Moody(1998)[1]가 소개한 재귀 강화학습은 주식 거래(Stock trading)에 적용되었으며 샤프 지수(Sharpe ratio)를 목표함수로 사용하여 단일 주식 거래와 포트폴리오에 적용되었다.

재귀 강화학습은 과거 일정 기간 동안의 자산들의 데이터와 이전의 상태에서 내린 행동에 대한 정보가 들어가기 때문에 재귀라는 말이 사용되었다. 재귀 강화학습은 연속적인 행동 값을 가지기 때문에 포트폴리오의 자산 배분 비중에 자연스럽게 적용할 수 있고, 목표함수 최적화를 위한 즉각적인 피드백을 받기 때문에 학습속도가 빠르다는 장점이 있다[2].

Moody가 재귀 강화학습을 소개한 뒤로 많은 재귀 강화학습을 사용한 연구가 진행되었는데, 예를 들면, Carl Gold(2003)[3]은 재귀 강화학습을 단일 계층과 다중 계층으로 구현해 비교실험을 하였다. Yue Deng(2017)[4]는 딥러닝을 통해 자산의 과거 데이터에서 특징을 추출하고, 재귀 강화학습에 입력으로 사용한 실험을 하였다. Saud Almahdi(2017)[5]는 재귀 강화학습의 목표함수로 사용된 샤프지수 대신 칼마 지수(Carmal ratio)와 스털링 지수(stering ratio)를 사용하여 비교 실험하였다.

하지만 위와 같은 연구들은 재귀 강화학습의 입력으로 오직 자산들의 과거 데이터를 사용하였을 뿐 포트폴리오의 목표에 도움을 줄 수 있는 자산 예측에 대한 적용은 없었다.

따라서 본 논문은 안정적이면서도 수익을 최대화시키는 포트폴리오의 최적의 자산 배분 비중을 생성하기 위해, 예측값의 활용이 효과적일 것이라는 가설을 세웠다. 본 논문에서는 기존에 Moody가 제시한 재귀 강화학습을 사용하였고, 기존 논문들과 같이 자산들의 과거 데이터와 함께 예측값을 입력으로 함께 사용하였다. 자산들의 과거 데이터만을 사용한 재귀 강화학습과 자산 예측을 추가한 재귀 강화학습의 비교 실험을 통해 자산 예측이 재귀 강화학습을 이용한 포트폴리오 운용에 효과적인지를 실험하였다.

본 논문은, 2장에서 관련 논문을 소개한다. 3장에서는 제안하는 모델이 서술되어있다. 4장에서는 실험 결과가 서술되어있다. 5장에서 결론이 서술되어있다.

2. 관련 연구

Markowitz(1952)[6]은 포트폴리오를 최적화하는 문제에대한 모델을 소개하였다. 이 모델은 평균-분산 모델로 잘 알려져 있으며, 자산들의 과거 데이터를 사용하는 포트폴리오 수익을 극대화하고 포트폴리오의 위험 최소화가 목적인 모델이다[7].

Moody[1]는 재귀 강화학습을 사용하여 목표함수에 대

1) 이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2017R1D1A1A02018319)

한 포트폴리오의 자산 할당 및 거래 시스템을 최적화하는 방법을 제시했다. 또한 Moody와 Saffell(2001)[8]은 재귀 강화학습과 Q-learning을 실제 데이터를 사용하여 비교실험을 하였고 Q-learning보다 재귀 강화학습이 더 좋은 결과를 보였다고 소개하였다.

Saud Almahdi[5]는 매매 시그널과 자산 배분 비중을 얻기 위해 Calmar Ratio와 RRL 방법을 사용한 모델을 제안했다. 실험은 자주 거래되는 상장 된 펀드로 구성된 포트폴리오를 사용하여, E(MDD) 기반의 목적 함수가 이전에 제안된 RRL 목적 함수와 비교하여 우수한 수익률을 산출한 결과를 발표하였다.

Yue Deng[4]은 딥러닝과 강화 학습의 두 가지 학습 개념에서 영감을 얻은 모델을 제시하였다. 제시한 모델에서 딥러닝 부분은 유익한 기능 학습을 위한 역동적인 시장 상태를 자동으로 감지한다. 그런 다음 강화학습 부분은 딥러닝 부분으로부터 얻은 deep representations과 상호작용하고 알려지지 않은 환경에서 최종 보상을 축적하기 위해 거래 의사 결정을 내렸다. 학습 시스템은 심층 구조와 반복 구조를 모두 나타내는 복합적인 신경망으로 구현했다.

박강희[9]는 포트폴리오 운용 알고리즘으로 Max-Return and Min-Risk(MRMR)을 제안하였고, 주가예측모델을 함께 활용하였다. 또한, 좋은 종목들로 포트폴리오를 구성하는 것과 함께 예측력이 좋은 주가예측 모델을 이용하면 포트폴리오를 통해 추구하는 목표에 근접할 수 있다고 주장했다[9].

3. 제안 방법

구현된 포트폴리오 운용 에이전트(agent)는 주어진 상태(State)에 대해 행동(action)을 선택하고 보상(Reward)인 샤프지수(Sharpe ratio)를 최대화시키기 위해 인공신경망을 학습시킨다. 해당 모델은 몇 가지의 정의들이 사용되었다.

기호 1) 에이전트의 액션 F_t

$\overrightarrow{F_t}$: 시간 t 시점에서 포트폴리오의 m개의 개별자산들에 대한 자산 배분 비중 벡터,

$$\overrightarrow{F_t} = (F_t^1, F_t^2, \ldots, F_t^m) \quad (1)$$

F_t^i: 시간 t 시점에서 포트폴리오의 i 번째의 개별자산에 대한 자산 배분 비중,

$$F_t^i \geq 0 \ , \ \sum_{i=1}^{m} F_t^i = 1 \quad (2)$$

기호 2) 에이전트의 상태 S_t 정의

$\overrightarrow{S_t}$: 시간 t 시점에서 포트폴리오 운용 에이전트의 상태 입력 벡터,

$$\overrightarrow{S_t} = (\overrightarrow{Sa_t^1}, \overrightarrow{Sa_t^2}, \ldots, \overrightarrow{Sa_t^m}) \quad (3)$$

$\overrightarrow{Sa_t^i}$: 시간 t 시점에서 i 번째 개별자산의 과거 l 일동안의 자산 가격 일단위 수익률 벡터와 미래 n일 동안의 예측값 벡터,

$$\overrightarrow{Sa_t^i} = (\overrightarrow{r_t^i}, \overrightarrow{j_t^i}) \quad (4)$$

$\overrightarrow{r_t^i}$: 시간 t 시점에서 과거 l일 동안의 과거 일 단위 수익률 벡터,

$$\overrightarrow{r_t^i} = (r_{t-l+1}^i, r_{t-l+2}^i, \ldots, r_t^i) \quad (5)$$

r_t^i: 시간 t 시점에서 i번째 개별자산의 일 단위 수익률,

$$r_t^i = \frac{z_t^i}{z_{t-1}^i} - 1 \quad (6)$$

z_t^i: 시간 t 시점에서 i 번째 개별자산의 자산 가격.

$\overrightarrow{j_t^i}$: i 번째 개별자산의 시간 t 시점에서 미래 t+n 구간 동안 예측율에 따른 예측값 벡터, 각 예측값들은 상승 예측 시 1, 하락 예측 시 -1로 설정.

$$\overrightarrow{j_t^i} = (j_{t+1}^i, j_{t+2}^i, \ldots, j_{t+n}^i) \quad (7)$$

기호 3) 에이전트의 보상 $U(\theta)$ 정의

$U(\theta)$: 총 운용 기간 T 시간 동안 에이전트의 행동으로 인해 발생한 포트폴리오의 수익률에 대한 샤프 지수,

$$U(\theta) = \frac{E(P_1, P_2, \ldots, P_T)}{\sqrt{Var(P_1, P_2, \ldots, P_T)}} \quad (8)$$

P_t: 시간 t 시점에서 포트폴리오의 수익률,

$$P_t = \sum_{i=1}^{m} W_t^i \bullet r_{t+1}^i \quad (9)$$

W_t^i: 시간 t 시점에서 에이전트의 행동 F_t^i의 포트폴리오 운용에 따라 변경된 자산 배분 비중.

본 논문에서는 기존에 제시된 RRL의 프레임워크를 Long short term Memory (LSTM)을 사용하여 구현하였다. LSTM은 RRL과 같이 이전의 포트폴리오 운용에 대한 정보를 LSTM의 Hidden State와 Cell State를 통해서 이전의 정보를 전달받아 입력으로 받은 현재 시점에서의 상태와 상호작용하여 행동을 결정한다. RRL의 재귀구조를 통한 학습원리는 LSTM의 Backpropagation Through Time (BPTT) 학습방법을 사용하였다[10].

4. 실험

본 논문에서 사용된 개별자산 데이터는 NASDAQ, S&P500, APA, MS, BMY, CPB의 1일 단위 데이터를 사용하였다. 데이터의 기간은 2007년 2월 15일부터 2016년 7월 1일까지의 데이터를 사용하여 실험을 실행하였다. 훈련 데이터의 기간은 2007년 2월 15일부터 2012년 9월 27일까

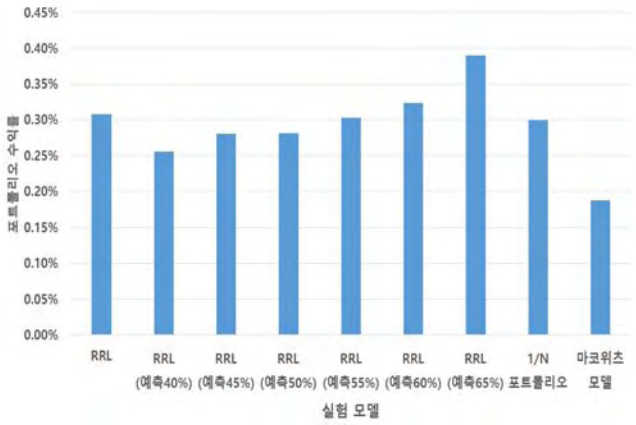
지로 설정하였고 테스트 데이터의 기간은 2012년 9월 28일부터 2014년 7월 28일까지로 설정하였다. 실험환경은 Intel Xeon 3.50Ghz CPU, 128G DRAM과 NVIDIA GTX 1080을 사용하여 진행하였다. 실험 프로그램은 Python과 Tensorflow를 사용하여 구현하였다.

(그림 1) 개별 자산들의 수익률 그래프

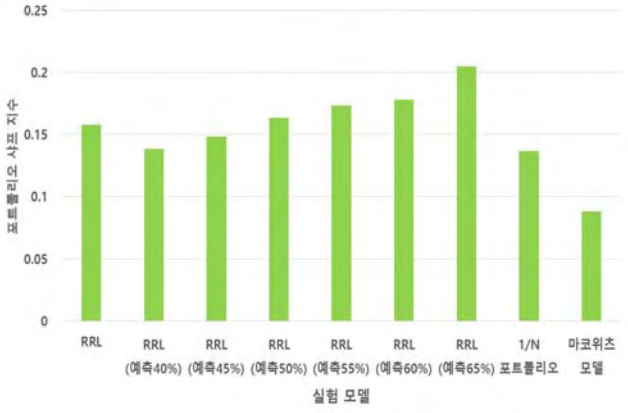
실험은 예측값을 적용하지 않은 포트폴리오 운용모델과 예측율을 40%~65%까지 5%단위마다 적용한 포트폴리오 운용모델의 포트폴리오 가치와 샤프지수를 비교분석하였다. 또한, 대표적인 포트폴리오 운용 모델인 마코위츠 방법과 1/N 포트폴리오 방법을 함께 비교분석 하였다. 표 1과 그림2, 그림3은 실험결과이다.

<표 1> 실험 모델에 따른 샤프 지수와 포트폴리오 수익률 결과

실험 모델	샤프 지수	포트폴리오 수익률
RRL	0.158	0.31%
RRL (예측 40%)	0.138	0.26%
RRL (예측 45%)	0.148	0.28%
RRL (예측 50%)	0.167	0.29%
RRL (예측 55%)	0.173	0.34%
RRL (예측 60%)	0.178	0.41%
RRL (예측 65%)	0.205	0.42%
1/N 포트폴리오	0.137	0.32%
마코위츠 모델	0.088	0.23%



(그림 2) 포트폴리오 가치



(그림 3) 포트폴리오 샤프 지수

5. 결론

본 논문은 재귀 강화학습을 사용한 포트폴리오의 성능을 향상시키기 위해 자산 예측값을 사용한 모델을 제안하였다. LSTM을 사용하여 재귀 강화학습을 구현하고 예측률에 따른 실험을 통하여 자산 예측이 포트폴리오의 성능 향상을 확인하였다. 향후, 자산을 예측하는 여러 딥러닝 또는 기계학습을 방법을 사용하여 본 모델에 추가할 계획이다.

참고문헌

[1] Moody J et al 1997 Performance functions and reinforcement learning for trading systems and portfolios J. Forecasting at press

[2] Lu, David W. "Agent Inspired Trading Using Recurrent Reinforcement Learning and LSTM Neural Networks." arXiv preprint arXiv:1707.07338 (2017).

[3] Gold, C. (2003). FX trading via recurrent reinforcement learning. Proceedings of the IEEE International Conference on Computational Intelligence in Financial Engineering 2003, op.cit., 363 - 371

[4] Deng, Y., Bao, F., Kong, Y., Ren, Z., and Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems.
[5] Almahdi, S., & Yang, S. Y. (2017). “An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown”. Expert Systems with Applications”, 87, 267 - 279.
[6] Harry Markowitz. Portfolio Selection:Efficient Diversification of Investment. New York: John Wiley & Sons, 1959.
[7] Z. Mashayekhi, H. Omrani, An integrated multi-objective Markowitz-DEA crossefficiency model with fuzzy returns for portfolio selection problem, Appl. Soft Comput. 38 (2016) 1 - 9.
[8] Moody, J. , & Saffell, M. (2001). Learning to trade via direct reinforcement. IEEE Trans- actions on Neural Networks, 12 (4), 875 - 889 .
[9] 박강희, 신현정, “포트폴리오 최적화와 주가예측을 이용한 투자 모형”, 대한산업공학회지 , 제39권, 제6호(2013), pp.535-545.
[10] Moody, J. , Wu, L. , Liao, Y. , & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and portfolios. Science, 17 (February 1997), 441 - 470 .