유가증권 시장에서 이중강화학습을 이용한 종목선택과 포트폴리오 최적화

김태윤[[1]](#footnote-1)

전남대학교 통계학과

요약

주식 투자와 자산 관리에서 포트폴리오 분배와 최적화는 위험을 관리하고 수익률을 극대화하기 위해 필수적인 부분으로 금융분야에서 해결해야 할 전통적인 문제였다. 한편 최근 딥러닝이 많은 연구가 이루어지고 큰 성과를 이루었고 그와 함께 강화학습 또한 큰 발전을 이루고 있다. 본 논문에서는

1.서론

강화 학습은 에이전트가 환경과 상호작용하며 시행 착오를 통해 최적의 의사결정을 학습하는 기계학습 알고리즘이다(Sutton & Barto, 2018). 강화학습은 지난 10년간 비디오 게임(Mnih et al., 2013; Mnih et al., 2015), 보드게임(Bard et al., 2020; Silver et al., 2016), 로봇 제어(Lillicrap et al., 2015) 등에서 큰 성과를 거두었고 이에 따라 현재 활발한 연구가 진행 중이다. 그러나 금융분야에서 강화학습은 큰 성과를 보여주지 못하고 있다(Yu, Lee, Kulyatin, Shi, & Dasgupta, 2019).

금융 분야에서 포트폴리오 최적화는 오랜 기간동안 해결하지 못한 주된 과제였다(Markowitz, 1959). 금융 투자에서 투자자산 분산을 통해 포트폴리오를 구성하고 비체계적 위험을 축소하면서 수익을 극대화할 수 있다. 지금까지 포트폴리오 이론들은 주로 계량경제학적인 방법으로 접근했고(Blume, 1970; Elton & Gruber, 1997) 이런 방법에는 투자 자산들의 기대수익률과 변동성, 투자 자산간의 상관계수 등의 추정치를 핵심적인 파라미터로 써야 하고 이런 추정치를 입력하는 것에 대한 문제점도 지적되어 왔다(Bawa, Brown, & Klein, 1979; Jobson & Korkie, 1981; Michaud, 1989).

그에 따라 금융 분야에 강화 학습을 적용시키기 위한 연구가 이루어졌으며(Cumming, Alrajeh, & Dickens, 2015; Dempster & Leemans, 2006; Deng, Bao, Kong, Ren, & Dai, 2016), 강화학습을 이용해 포트폴리오를 관리하기 위한 연구 또한 이루어졌다(Guo, Fu, Shi, & Liu, 2018; Jiang, Xu, & Liang, 2017; Kim, Heo, Lim, Kwon, & Han, 2019; Yu et al., 2019). 이런 연구들에서 이루어진 방법들은 주로 총 수익률에 따라 강화학습을 수행하는 에이전트에 보상을 줌으로써 피드백이 이루어지는 것이다. 하지만 이 연구들은 주로 암호화폐, 외환, 파생상품 등을 거래함으로써 소수점거래의 제한을 받지 않았다(Dempster & Leemans, 2006; Deng et al., 2016; Jiang et al., 2017; Kim et al., 2019). 하지만 한국 주식시장에서는 소수점거래가 제한적이며 이에 따라 새로운 강화학습 방법을 적용할 필요가 있다.

Deep Q-learning(Mnih et al., 2013; Mnih et al., 2015), Double Deep Q-learning(Van Hasselt, Guez, & Silver, 2015), Dueling Deep Q-leaning(Wang et al., 2016)과 같은 가치 기반 방법들은 가중치와 같은 연속형 행동을 취할 수 없고

2.관련 연구

포트폴리오 최적화를 위해 강화학습을 이용한 연구로 Jiang et al에서는 컨볼루션 신경망과 순환신경망을 이용해 3차원 가격 데이터를 이용하는 EIIE와 이전 가중치에 대한 메모리를 저장하여 가중치를 구할 때 이용하는 PVM 그리고 시계열 자료에 적절한 배치 학습을 적용하기 위한 OSBL을 제안하였다. 또한 여기에서 xx, xx, xx를 도입해 모델 기반 강화학습을 구현한 연구가 있었다.(Yu et al., 2019)  
암호화폐의 포트폴리오 구성에서 a3c를 이용해 에이전트의 강화 학습을 멀티쓰레드로 구성해 학습속도를 향상 시킬 수 있다는 것이 확인 되었다(Kim et al., 2019).

RRL을 이용함으로써 동적 포트폴리오 구성으로 기대 MDD값을 최소화시켜 위험을 관리해 헤지펀드의 벤치마크를 상회할 수 있다는 것이 확인되었다(Almahdi & Yang, 2017).

3.환경 설정

에이전트의 학습이 이루어지기 위한 강화학습 환경을 주가 데이터를 통해 구성하였다. 에이전트는 해당 환경에서 상태로써 주식 가격데이터를 받고 행동으로 가중치를 실행하면 다음 시점의 주식 데이터를 상태로, 수익을 보상으로써 받는다.

3.1 주가 데이터

본 연구에서 주가의 데이터는 코스피200 종목 중 거래일이 1000일이 넘는 종목에서 시가총액 상위 10개의 종목을 선택하였다.

3.2 수익과 보상함수

강화학습의 에이전트는 시행 결과로부터 정책을 최적화시키며 이 과정에는 보상이 필요하다.

포트폴리오를 구성하는 종목 수가 N 일 때, 에이전트의 행동은 시점 t에서 각 종목에 대한 가중치 로써 이뤄진다. 일 때, 에이전트는 초기 자산 가치 를 부여받으며 에이전트는 이 초기 자산을 가중치 와 주식 가격에 따라 분배해서 를 얻는다.

여기에서 는 Hadamard product이며 는 의 Hadamard inverse이다. 따라서 는 정수 값을 가진 벡터이며 각 종목의 보유 주식 수를 나타낸다.

에이전트가 t시점에서 전체 포트폴리오의 가치에서 각 자산에 대해 적용시킬 비율 w를 구하면 종목별로 P\_n의 가격에 할당된 자산 내에서 살 수 있는 주식의 개수를 구한다. 각 종목에 할당된 금액에서 살 수 있는 주식을 뺀 금액이 잔여 가치이다. 그리고 t+1시점에 포트폴리오의 가치는 각 종목의 t+1시점의 가격과 잔여 가치의 합이다.

3.3 할인계수

강화학습에서 할인계수 는 불필요한 행동으로 보상이 지연되기보다 최대한 빠른 시간 내에 보상을 얻기 위해 적용된다. 이 할인계수는 환경과 목적에 맞게 하이퍼파라미터로 설정해야 하지만 금융 문제에서는 객관적이고 신뢰성 있는 할인계수로써 금리 을 사용할 수 있다. 이 금리 은 해당 기간 동안에 적용되는 기준금리를 사용하였으며 연속복리로 결정된다. 본 연구에서는 시간 단위 를 1영업일로 하였고 1년을 250영업일로 하였다.

3.4 가정

본 연구에서 학습된 에이전트의 백테스트가 실행된 환경은 아래의 세가지 가정이 충족된다고 가정한다.

1. 거래비용 없음: 시장에는 유동성이 충분하여 종가에 거래할 수 있고 거래세와 거래수수료는 없다고 가정한다. 이 가정은 매매 빈도가 낮고 법인세법의 적용을 받는 법인 거래라면 정당화된다.
2. 시장 충격 없음:

구성

Jiang et al에서는 정해진 날짜와 포트폴리오 구성 자산 수를 높이와 너비로 하고 특징을 깊이로 갖는 3차원 텐서를 인풋으로 컨볼루션 신경망을 구성할 것을 제안하였다.

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.

Bard, N., Foerster, J. N., Chandar, S., Burch, N., Lanctot, M., Song, H. F., . . . Hughes, E. (2020). The hanabi challenge: A new frontier for ai research. *Artificial Intelligence, 280*, 103216.

Bawa, V. S., Brown, S. J., & Klein, R. W. (1979). Estimation risk and optimal portfolio choice. *NORTH-HOLLAND PUBL. CO., N. Y., 190 pp*.

Blume, M. E. (1970). Portfolio theory: a step toward its practical application. *The Journal of Business, 43*(2), 152-173.

Cumming, J., Alrajeh, D., & Dickens, L. (2015). An investigation into the use of reinforcement learning techniques within the algorithmic trading domain. *Imperial College London: London, UK*.

Dempster, M. A., & Leemans, V. (2006). An automated FX trading system using adaptive reinforcement learning. *Expert Systems with Applications, 30*(3), 543-552.

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems, 28*(3), 653-664.

Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking & Finance, 21*(11-12), 1743-1759.

Guo, Y., Fu, X., Shi, Y., & Liu, M. (2018). Robust log-optimal strategy with reinforcement learning. *arXiv preprint arXiv:1805.00205*.

Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*.

Jobson, J. D., & Korkie, R. M. (1981). Putting Markowitz theory to work. *The Journal of Portfolio Management, 7*(4), 70-74.

Kim, J.-B., Heo, J.-S., Lim, H.-K., Kwon, D.-H., & Han, Y.-H. (2019). Blockchain Based Financial Portfolio Management Using A3C. *KIPS Transactions on Computer and Communication Systems, 8*(1), 17-28.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., . . . Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Markowitz, H. (1959). Portfolio selection. *Investment under Uncertainty*.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is ‘optimized’optimal? *Financial Analysts Journal, 45*(1), 31-42.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529-533.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., . . . Lanctot, M. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*(7587), 484-489.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*: MIT press.

Van Hasselt, H., Guez, A., & Silver, D. (2015). Deep reinforcement learning with double q-learning. *arXiv preprint arXiv:1509.06461*.

Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., & Freitas, N. (2016). *Dueling network architectures for deep reinforcement learning.* Paper presented at the International conference on machine learning.

Yu, P., Lee, J. S., Kulyatin, I., Shi, Z., & Dasgupta, S. (2019). Model-based deep reinforcement learning for dynamic portfolio optimization. *arXiv preprint arXiv:1901.08740*.

1. 광주광역시 북구 용봉동 전남대학교, 석사과정 [↑](#footnote-ref-1)