

유가증권 시장에서 강화학습을 이용한 종목선택과 포트폴리오 최적화

김태윤¹, 고봉균²

¹전남대학교 통계학과 · ²전남대학교 통계학과

접수 0000년 0월 1일, 수정 0000년 0월 0일, 게재확정 0000년 0월 0일

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. 이에 따라 최근 포트폴리오 관리에 강화학습 방법론을 적용하려는 시도가 이루어졌지만 연구의 대부분은 거래 규모가 큰 암호화폐에 한정되어 이루어진 것이 대부분이다. 본 논문에서는 유가증권시장의 상위 종목 중 대표성이 높은 종목으로 선정되는 KOSPI200을 구성하는 종목 중 투자 대상 주식을 선정하는 selector 네트워크와 선정된 주식을 배분하는 allocator 네트워크 두가지를 통해 포트폴리오를 구성하는 신경망을 강화학습을 통해 구현하였다.

주요용어: 강화학습, 포트폴리오 이론, 딥러닝, 코스피

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 상관관계수 등의 추정치를 핵심적인 파라미터로 써야 하고 이런 추정치를 입력하는 것에 대한 문제점도 지적되어 왔다(Bawa, Brown, & Klein, 1979; Jobson & Korkie, 1981; Michaud, 1989).

¹ (61186) 광주광역시 북구 용봉로 77 자연대 1호관 238, 전남대학교 통계학과, 석사과정. E-mail: 196350@jnu.ac.kr

² (61186) 광주광역시 북구 용봉로 77 자연대 1호관, 전남대학교 통계학과, 교수. E-mail:

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 위의 연구들은 주로 거래량이 많은 암호화폐, 외환, 파생상품 등을 거래함으로써 소수점거래의 제한을 받지 않았고 종목선택이 자의적으로 이루어졌다(Dempster & Leemans, 2006; Deng et al., 2016; Jiang et al., 2017; Kim et al., 2019). 하지만 한국 주식시장에서는 종목선택의 문제가 있고 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-learning(Wang et al., 2016)과 같은 가치 기반 방법들은 가중치와 같은 연속형 행동을 취할 수 없고 특정 상태에서 취할 수 있는 개별적인 행동에 대한 가치를 정확히 추정하는데 목표를 둔다.

반면 정책 기반 방법들은(Schulman, Levine, Abbeel, Jordan, & Moritz, 2015; Schulman, Wolski, Dhariwal, Radford, & Klimov, 2017; Silver et al., 2014) 보상을 극대화시키는 것을 목표로 정책을 최적화시키는 알고리즘이다. 정책 기반 방법은 연속형이나 확률분포형으로 행동을 취할 수 있어 로봇제어, 비디오 게임 등에 적합하다.

본 연구에서는 대상 주식을 선정하는 것에 컨볼루션 신경망을 사용하는 selector 네트워크를 사용하며, selector 네트워크를 통해 선정된 주식을 적절한 비율로 배분하는 것에 정책기반 강화학습 allocator 네트워크를 사용한다.

2 FRACAS 개요

내용이 여기에 옴.

Table 2.1 표의 제목

열제목1	열제목2	열제목3	열제목4	열제목5	열제목6
1	123	88	33	aa	999

내용이 여기에 옴.

$$h(y) = \sin\left(P_1(\beta; y) + R_2(\gamma; y)\right),$$

내용이 여기에 옴.

$$h(x) = \log\left(\int_{i,j} \alpha_{[i][j]} x_{ij} x_{i(j+1)} + \beta_{[i][j]} y_{ij} m u_{(i+1)j}\right), \quad (2.1)$$

내용이 여기에 옴.

내용이 여기에 옴.

정리 2.1 정리 내용

증명: 증명이 여기에 옴.

□

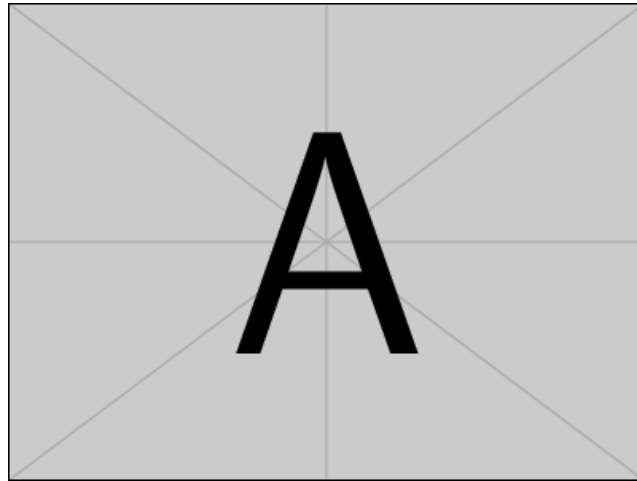


Figure 2.1 그림 제목

참고문헌

Kang, S. M. and Kim, J. H. (2008). Statistical model of effective impact speed based on vehicle damages in case of rear-end collisions. *Journal of the Korean Data & Information Science Society*, **19**, 463-473.

부록 A: A 절 제목

내용이 여기에 옵니다.

정리 **A.1** 정리의 내용이 여기에 옵니다.

증명: 정리의 증명이 여기에.

□

Article title[†]

Author First¹. Author Second²

Department of Statistics, Azzzz University
Department of Management, Bzzzz University

Received 1 0000, revised 0 0000, accepted 0 0000

Abstract

Some sentences come here.

Keywords: key1, key2.

[†] This reserach bla bla bla....

¹ Professor, Department of Statistics, Gyeongbuk 712-749, Korea.
E-mail: zzzzaong@azzzz.ac.kr

² Professor, Department of Management, Seoul 100-100, Korea.
E-mail: zzzzaong@bzzzz.ac.kr