

DQN 강화학습을 이용한 주식 트레이딩에 관한 연구

백지원¹, 서대원², 송주혜³, 정인혁⁴, 이규영⁵

¹광운대학교 전자융합공학과

²광운대학교 정보융합학부

³상명대학교 경영학부

⁴상명대학교 컴퓨터과학과

⁵한국과학기술원 정보보호대학원

bjiwon59@gmail.com, ddww1029wwdd@gmail.com, songddubi58@gmail.com,
inhyukjeong10@gmail.com, leeahn1223@kaist.ac.kr

A Study on Stock Trading using DQN Reinforcement Learning

Ji-Won Baek¹, Dae-Won Seo², Ju-hye Song³, In-Hyuk Jeong⁴, Gyuyoung Lee⁵

¹Dept. of Electronic Convergence Engineering, Kwangwoon University

²Dept. of Information Convergence, Kwangwoon University

³Dept. of Business Administration, Sangmyung University

⁴Dept. of Computer Science, Sangmyung University

⁵Graduate School of Information Security, KAIST

요 약

본 연구는 변동성이 높은 주식시장에서 안정적인 수익창출에 기여할 수 있는 주가예측 강화학습 모델을 제안한다. DQN 알고리즘과 LSTM 신경망을 이용하여 시장의 흐름에 따라 전략을 달리하는 모델을 개발하고, 이를 활용한 주식 트레이딩 시스템의 유용성을 확인하고 발전 방향을 제시한다.

1. 서론

국내 시장의 개인 투자자 10 명 중 9 명은 손해를 보거나 1000 만원 이하의 이익을 낸다[1]. 하지만 이러한 상황과는 대조적으로 개인 투자자 수는 5 년 연속 증가하여 지난해 1400 만명을 넘어섰다[2]. 개인의 트레이딩 방식에 의존해 리스크가 높은 행동을 선택하는 투자자들을 위하여, 정보의 비대칭성에서 기인하는 불균형을 해소할 방법 제시가 절실한 상황이다.

주가 지수는 변동성을 가진 시계열 데이터의 특성을 가져 비정상성과 비선형성을 갖게 되고, 주식 시장의 잡음은 관측에 영향을 주기 때문에 기존의 통계 기반 방법으로는 괄목할 만한 성과를 내기 어려웠다[3]. 하지만 최근에는 인공지능을 기존 분석 기술과 결합하여 시장 수익률을 상회하는 투자가 가능하다는 연구 결과들이 제시되고 있다.

본 연구는 주가기본 데이터, 기술분석 데이터, 주가변동요소 데이터를 학습에 이용하여 주가 예측 강화학습 모델을 개발한다. 해당 모델을 적용한 주식 트레이딩 시스템의 유용성을 탐구해보고 추후 발전 방향을 제시한다.

2. 관련 연구

2.1. 강화학습

강화학습은 주어진 임의 환경에서 학습의 주체가 되는 에이전트가 행동 결과에 따른 보상을 통해 학습을 하며 행동을 변화시키고 발전시킨다는 이론이다.

2.2. DQN

DQN 은 매우 크고 복잡한 환경에서 Q-Learning 으로 학습이 잘 되지 않는 문제를 해결하기 위해, 심층 신경망을 통해서 Q value 를 근사하는 방식으로 해결한 강화학습 알고리즘이다.

2.3. A3C

A3C 는 하나의 전역신경망과 그와 같은 구조를 갖는 여러 개의 에이전트로 구성이 되어 있으며, 에이전트 각각은 가치신경망, 정책신경망을 가지는 강화학습 알고리즘이다.

2.4. LSTM

LSTM 은 기존 RNN 의 은닉층에 셀 상태를 추가한 신경망으로, 셀 상태를 통해서 과거의 학습 정보를 기억하고 이를 새로운 학습 결과에 반영할 수 있다는 특징이 있어 시계열 데이터를 다루기에 적합하다.

2.5. 웨이블릿 변환

금융 시계열은 비정상성을 띠고, 주요 정보와 노이즈가 중첩된다는 특징이 있다. 웨이블릿 변환은 시간

분해능과 주파수분해능을 가지고 있고 다중 스케일 특성도 있어 주가 데이터의 노이즈를 제거할 수 있다.

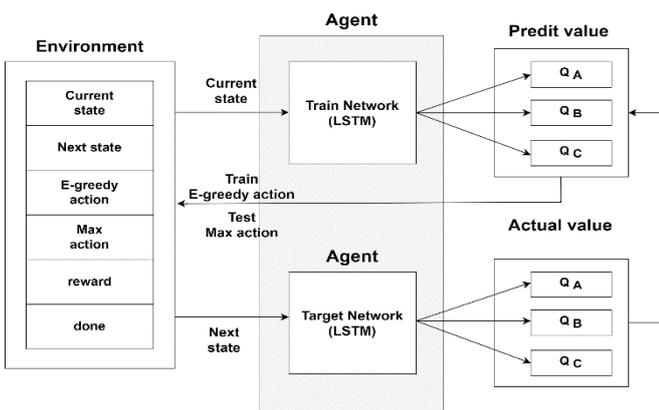
3. 주가 예측 모델 개발

일자별로 날짜, 시가, 고가, 저가, 종가, 거래량의 데이터를 불러와서 Daubechies D8 웨이블릿 필터를 사용하여 변환하고, 임계값을 이용하여 노이즈로 간주되는 값들을 제거한 후에 원래의 시계열 데이터 형태로 되돌렸다. 이후 이동평균 증가 비율, 이동평균 거래량 비율 등을 구하며 데이터 전처리를 실행하였다.

학습기 모듈 개발 시에는 학습데이터를 2012년 1월에서 2021년 12월까지의 주가로 정해 3개 종목을 DQN과 A3C 두 알고리즘으로 모델링하고 성능을 비교하였다. 평균적인 수익 안정성이 높은 사유로, DQN 알고리즘을 채택하여 본 프로젝트에 적용하였다.

에이전트 모듈은 학습을 위한 에이전트인 메인 네트워크와 실제값을 구하기 위한 타겟 네트워크를 LSTM으로 구현하였다. 활성화함수와 optimizer를 변경할 때 수익을 비교하였고, 일반적으로 softsign 활성화 함수와 adamax와 nadam에 optimizer를 조합하여 사용할 경우 최적의 성능을 이끌어 내는 것을 확인하였다.

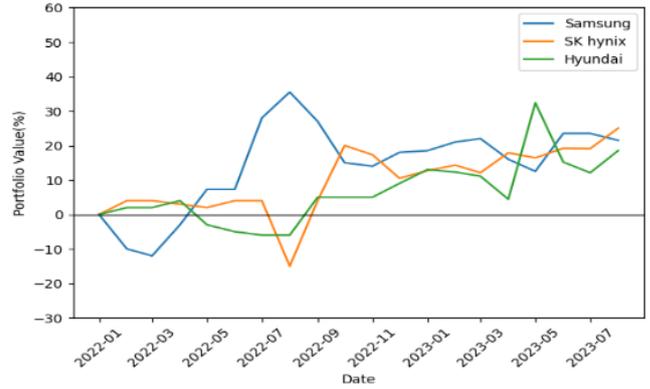
최종 모델은 그림 1와 같다. 거래의 주체인 에이전트의 상태는 초기 자본금과 현금 잔고, 주식 잔고로 정의되고 주어진 상태에서 에이전트 매수-매도-관망 행동을 할 수 있다. 주가 데이터를 환경으로 하여 행동에 따른 보상을 최대화하는 방식으로 학습을 하게 되며, 보상은 배치 데이터 내 마지막 손익률에서 행동을 수행한 시점에서의 손익률을 뺀 값으로 주어지게 된다. 학습을 마친 에이전트는 행동 예측 과정에서 주식데이터를 가치신경망에 넣어 매도, 매수를 했을 때 향후 획득할 수 있는 수익률을 예측한다. 매수가 매도보다 가치가 높으면 매수를, 반대의 경우는 매도를, 매도와 매수의 가치 차이가 1.5% 이하인 경우에는 관망을 하게 된다.



(그림 1) 최종적인 강화학습 모델 구조

제작한 모델을 사용하여 2022년 1월 ~ 2023년 8월 주가를 바탕으로 3개 종목에 대한 투자를 진행했을 때, 최종 PV(Portfolio value)값이 평균적으로 21.3% 상승하였고, PV값의 하락이 있을 때도 평균적으로 10% 대의 손실에 그쳐 비교적 안정적인 투자 결

과를 보였다.



(그림 2) 학습한 모델을 사용할 때의 PV 변화

4. 결론 및 향후 연구

주식 트레이딩은 실시간으로 즉각적인 의사결정이 요구되기 때문에, 머신러닝을 통한 예측 알고리즘이 다양한 도메인에서 널리 사용되고 있음에도, 직접 적용하는 것에는 한계가 있다. 즉각적인 매도와 매수 의사결정을 고려할 때 보상과 행동을 모델링할 수 있는 강화학습은 효과적인 대안이 될 수 있다[4]

본 연구는 주가 기본 데이터, 기술분석 데이터, 주가 변동 요소 데이터를 이용하여, 주식 트레이딩 예측을 위한 강화학습 모델을 구현하였다. 실험 결과 강화학습이 트레이딩에 적합한 결과를 도출하였고, 수익률을 높이는 데 도움을 줄 수 있다는 가능성을 확인하였다. 개발한 모델을 보완하여 투자자동화 기능을 보유한 트레이딩 시스템을 개발한다면, 그 활용 가치는 더욱 높아질 것으로 예상된다.

본 프로젝트에서는 결론적으로 DQN 알고리즘만 사용했지만, 다양한 강화학습 알고리즘을 앙상블 기법을 통해 활용한다면 안정적인 투자 전략을 구사할 수 있을 것으로 예측된다. 또한 구조적 변경과 하이퍼파라미터 튜닝 등을 통해 모델 자체의 성능을 보다 향상시킬 수 있을 것이다.

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

참고문헌

[1] 한국조세재정연구원, 20년 06월 연구보고서
 [2] 한국예탁결제원, 22년 12월 결산 상장법인 주식 소유자 현황, 보도자료, 2022.03.16
 [3] 권희석·이민혁, (2023). Wavelet Transform LSTM 과 Multi-Head Attention 의 결합모형을 이용한 주가 지수 예측. 경영과학, 40(2), 97-112
 [4] 황호현·김용훈·이영훈, (2021). 강화학습 기반 주식 자동 매매 모델 전략 제안. 대한산업공학회지, 47(4), 399-405.