Journal of KIIT. Vol. 17, No. 4, pp. 1-8, Apr. 30, 2019. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2019.17.4.1

DQN을 이용한 트레이딩 예측을 위한 강화학습 모델 구현

하은규*, 김창복**

Model Implementation of Reinforcement Learning for Trading Prediction Using Deep Q Network

Eun-Gyu Ha*, Chang-Bok Kim**

요 약

본 연구는 주가 기본 데이터와 기술 분석 데이터 그리고 주가 변동 요소 데이터를 이용하여, 트레이딩 행동 예측을 위한 강화학습 모델을 구현하였다. 강화학습 모델은 에이전트를 인공신경망으로 사용하였으며, 환경은 현재 상태, 다음 상태, 행동, 보상, 에피소드 종료로 구축하였다. 본 연구는 세 가지 강화학습 모델을 구축하여 학습결과를 비교하였다. 첫 번째 모델은 버퍼의 학습 데이터를 랜덤하게 추출하고, 하나의 신공신경망으로 학습하였다. 두 번째 모델은 버퍼의 데이터를 순서적으로 추출하고, 두 개의 인공신경망으로 학습하였다. 세 번째 모델은 버퍼의 데이터를 랜덤하게 추출하고 두 개의 인공신경망으로 학습하였다. 실험 결과, 세 번째 방법이 근소하게 결과가 좋았으며, 학습 결과가 10배에서 1000배 까지의 이득을 남기는 행동을 하였다. 또한, 학습 결과가 좋은 종목이 테스트 결과도 좋았으며, 이것은 종목별로 주가의 패턴에 기인한 것으로 추정된다.

Abstract

This study implements a reinforcement learning model to predict trading actions efficiently using basic stock data, technical analysis data, and the stock fluctuation factor. The reinforcement learning model uses an artificial neural network as an agent and structures the environment as the current state, next state, action, reward, and termination of the episode. This study compares the training results of three constructed reinforcement training models. The first model extracts the training data of the buffer randomly and learns with a single artificial neural network. The second model extracts the training data of the buffer sequentially and learns with two artificial neural networks. The third model extracts the training data of the buffer randomly and learns with two artificial neural networks. The comparison indicates that the third method was slightly better. The results show a profit of 10 to 1000 times. Additionally, item with good training results have good test results, and estimates that this result is due to the patterns of each item.

Keywords

machine learning, reinforcement learning, deep learning, trading forecast, markov decision process

* 가천대학교 에너지IT학과
- ORCID: http://orcid.org/0000-0002-4436-1751
** 가천대학교 에너지IT학과 교수(교신저자)
- ORCID: http://orcid.org/0000-0002-7155-5033

· Received: Dec. 22, 2018, Revised: Mar. 06, 2019, Accepted: Mar. 09, 2019

· Corresponding Author: Chang-Bok Kim
Department of Energy IT, Gachon University, 1342, Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, Korea
Tel.: +82-32-446-0695, Email: cbkim@gachon.ac.kr

Ⅰ. 서 론

주가는 수요와 공급에 의해 결정되며, 주가 변동은 기술적 요인, 기본적 요인, 심리적 요인 등에 대한 모든 정보가 반영되어 있다. 또한, 잡음, 비정상성, 비 선형성으로 인해 주가 변동 및 기대 수익을 예측하는 것은 매우 어렵다. 따라서 주가를 예측하는 것은 불가능하다는 주장으로 인해, 기존의 분석 방법에 대해 회의적이었다[1][2]. 그러나 주가의 기술 분석 및 기본 분석 데이터와 인공신경망(Artificial Neural Network)과의 결합을 통해 시장 평균을 초과하는 수익의 달성이 가능하다는 연구 결과들을 제시하고 있다[3]-[8].

강화학습(Reinforcement Learning)은 임의의 환경에서 학습주체인 인간의 두뇌에 해당하는 에이전트의 행동 결과에 대한 보상 여부에 따라 행동을 변화시키고 발전시킨다는 이론이다. 즉, 정답은 모르지만, 행동에 대한 보상으로 학습한다. 강화학습은 MDP(Markov Decision Process) 수학적 모델과 DP(Dynamic Programming)를 기반으로, Monte-carlo, Temporal Difference을 거쳐, DQN(Deep Q Network) 등으로 발전되면서 많은 분야에 응용되고 있다[9]-[10].

본 연구는 주가의 기본 데이터와 보조 지표 데이터 뿐 아니라 주가 변동 요소 데이터를 이용하여, 트레이딩 예측을 위한 강화학습 모델을 구현하였다. 기본 데이터는 10년간의 일일 데이터로서 시가, 고가, 저가, 종가, 거래량 등이다. 기술 분석 데이터는 주가 예측에 효율적인 Stochastic, CCI, 등락 구분 5일, 등락 구분 20일 등을 사용하였다. 또한, 주가 변동 요소 데이터는 등락 구분, 환율, 전 산업 생산지수 등을 사용하였다. 이러한 모든 예측 요소는 통계 및 데이터 마이닝을 위한 오픈 소스인 R의 neuralnet 신경망 패키지를 이용하여, 가장 좋은 예측 요소의 조합을 선택하였다. 강화학습 모델은 크게 학습 환경과 에이전트로 구분된다. 환경은 현재 상태, 행동, 다음상태, 보상 그리고 에피소드 종료 여부이다. 에이전트는 DNN(Deep Neural Network)으로 구현하였으며, 학습과 실제 값을 동시에 산출하는 하나의 네트워크와 학습을 위한 에이전트인 메인 네트워크와 실제 값을 산출하기 위한 타겟 네트워크 등 두 개의 네트워크를 사용하였다.

본 논문은 2장에서 관련연구로서 강화학습과 상태 변수에 대해서 서술하였다. 또한, 3장에서 트레이딩 행동 모델을 제안하였으며, 4장에서 구현된 모델의 실험 결과 및 비교 분석을 하였으며, 마지막으로 결론에 대해서 서술하였다.

Ⅱ. 관련 연구

2.1 강화학습

기계학습은 지도학습과 비 지도학습으로 구분된다. 지도학습은 입력 데이터와 정답이 1대 1로 구성되어 있는 훈련 데이터를 사용하여 학습하는 것이다. 비 지도학습은 추론, 분석과 같이 정답이 없이 학습하는 것이다. 강화학습은 정답은 모르지만 행동에 대한 보상으로 학습한다.

강화학습은 MDP의 수학적 모델 정의에 의해 시작되었다. MDP는 $< S, A, P, R, \gamma >$튜플로 정의된다. S는 상태, A는 행동, P는 상태 전이 확률 매트릭스, R은 보상, γ은 할인 계수이다. MDP의 기본 이론은 다음과 같다.

$$P[S_{t+1}R_{t+1} \mid S_0 A_0 R_1 \cdots S_{t-1}A_{t-1}, R_t S_t A_t] \quad (1)$$

$$P[S_{t+1}R_{t+1} \mid S_t A_t] \quad (2)$$

MDP의 기본 이론은 식 (1)과 같이 과거 상태에서 다음 상태 S_{t+1}까지의 확률 P는 식 (2)와 같이 현재 상태에서 다음 상태의 확률과 같다는 것으로, 현재 상태는 과거의 모든 상태를 담고 있어, 다음 상태 S_{t+1}와 보상 R_{t+1}은 현재 상태 S_t와 행동 A_t에 의해서 결정된다.

강화학습은 이러한 수학적 모델 정의를 기반으로, 어떤 환경을 탐색하는 에이전트가 현재의 상태를 인식하여, 어떤 행동을 취하면 환경으로부터 보상을 얻게 되며, 에이전트는 이를 통해 누적 보상을 최대화하는 일련의 행동에 대한 최적정책을 찾는 학습이다. 다음은 Q 학습의 최적정책에 대해서 나

타냈다.

$$\pi^*(S) = \arg max_a Q(s,a) \quad (3)$$

$\arg max_a Q(s,a)$는 현재 상태의 행동들 중에서 가장 보상이 큰 행동을 의미한다. Q 학습은 최적정책으로 각 상태의 행동에 대한 현재 상태의 누적 보상을 다음과 같이 산출할 수 있다.

$$Q(s,a) = r + \gamma \max_{a'} Q(s',a') \quad (4)$$

여기서 $Q(s,a)$는 현재 상태의 행동에 대한 값이며, r은 다음 상태의 보상이다. 또한 γ는 할인계수이며, $\max_{a'} Q(s',a')$는 다음 상태에서 가장 큰 Q 값이다. 즉, 현재 상태의 누적 보상은 다음 상태의 보상과 다음 상태의 행동에서 가장 큰 값으로 구해진다. 그림 1에 강화학습 구성에 대해서 나타냈다.

강화학습은 학습 환경과 에이전트로 구분된다. 환경은 현재 상태, 행동, 다음 상태, 보상 그리고 에피소드 종료 여부이다. 즉, 현재 상태에서 에이전트가 행동을 하면 다음 상태로 전이되며 다음 상태에서 보상을 받는다.

에이전트는 인공신경망으로 구현된다, DQN은 학습과 실제 값을 동시에 산출하는 하나의 네트워크와 학습을 위한 에이전트인 메인 네트워크와 실제 값을 산출하기 위한 타겟 네트워크 등 두 개의 네트워크로 구현할 수 있다. 여기서 메인 네트워크의 입력은 현재 상태이며, 출력은 상태에 대한 예측 값이다. 타켓 네트워크의 입력은 다음 상태이며 실제 값을 구하기 위한 Q 값이다. 다음은 예측 값, 실제 값, 오차 함수에 대해서 나타냈다.

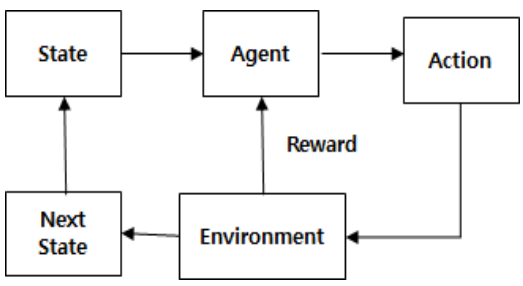


그림 1. 강화학습 구성
Fig. 1. Constitution of reinforcement learning

$$Ws = Q(s,a|\theta) \quad (5)$$

$$Q^*(s,a|\bar{\theta}) = r + \gamma \max Q(s',a'|\bar{\theta}) \quad (6)$$

$$\cos t(W) = \frac{1}{m}\sum_{i=1}^{m}(Q(s,a|\theta)^i - Q^*(s,a|\bar{\theta})^i)^2 \quad (7)$$

여기서 θ는 예측 값과 학습을 위한 메인 네트워크이며 $\bar{\theta}$은 실제 값을 산출하는 타겟 네트워크이다. 식 (5)는 메인 네트워크에서 예측 값을 산출하는 것이고 식 (6)은 타겟 네트워크의 Q 값을 이용하여 실제 값을 산출하는 것이다. 식 (7)은 실제 값과 예측 값의 오차 함수이며, 오차 함수를 이용하여 메인 네트워크를 학습한다.

2.2 상태 변수

상태 변수는 강화 학습을 위해 에이전트에 입력되는 변수로서 예측 값과 실제 값을 산출하는 요소이다. 특히, 주식 데이터는 많은 불확실한 요소로 인해 학습에 어려움이 있다. 따라서 시가, 고가, 저가, 종가, 거래량 등 기본적인 주가 데이터 뿐 아니라 기술 분석을 통한 보조 지표 데이터 그리고 주가 변동 요인이 되는 요소를 상태 변수로 사용하였다. 기술 분석은 전통적인 주가 예측 방법으로 과거 주가나 거래량을 이용하여, 주가 변화의 패턴을 분석하여 주가를 예측하는 방법이다[11]-[13]. 표 1에 주가 예측에 많이 사용하는 보조 지표에 대해서 나타냈다.

이동 평균은 특정 기간 동안의 방향성을 수치화한 것이며, 여기서 이동 평균선의 기울기와 이동 평균선 간 거리를 수치화한 것이다. 지수 이동 평균은 최근의 값에 좀 더 큰 가중치를 부여하는 방법이다.

MACD(Moving Average Convergence and Divergence)는 지수 이동 평균 값 간의 거리를 나타낸다. 이격도는 주가와 이동 평균선과의 거리이다. RSI는 현재의 주가 추세 강도를 백분율로 나타내어, 추세 전환을 예측하는 지표이다. CCI(Commodity Channel Index)는 이동 평균으로부터 주가의 변동성을 측정하는 지표로 추세의 강도와 방향을 표시한다. Stochastic은 과열 및 침체를 나타내는 지표이다.

표 1. 기술 분석 및 보조 지표
Table 1. Technical analysis and sub-indicators

Index	Sub-Indicator	Calculation Formula
1	Moving Average	$MA_t^n = \frac{1}{n}\left(\sum_{i=t-n+1}^{t} close_i\right)$
2	Moving Average Line Slope	$g_t^n = \frac{\left(MA_t^n - MA_{t-1}^n\right)}{MA_{t-1}^n}$
3	Moving Average Line Distance	$d_t^{m,n} = \frac{\left(MA_t^n - MA_t^m\right)}{MA_t^m}$, (m > n)
4	Exponential Moving Average	$EMA(m)_t = close_i \times EP + close_{i-1} \times (1-EP)$ $EP(\mathrm{Exponential\ Percentage}) = \frac{2}{m+1}$
5	MACD(Moving Average Convergence & Divergence)	$MACD = \sum_{i=t-9}^{t} (EMA(12)_t - EMA(26)_t)$
6	Disparity	$disparity = \frac{close_t}{MA_t^n} \times 100$
7	RSI(Relative Strength Index)	$RSI = \frac{AU_t^n}{AU_t^n + AD_t^n} \times 100$
8	CCI (Commodity Channel Index)	$CCI = \frac{M_t - MA_t^n}{D \times 0{,}015}$, $M_t = \frac{high_t + low_t + close_t}{3}$, $MA_t^n = \frac{1}{n}\left(\sum_{i=t-n+1}^{t} M_t\right)$, $D = \frac{1}{n}\left(\sum_{i=t-n+1}^{t} \lvert M_t - MA_t^n \rvert\right)$
9	Stochastic1	$\%K(m) = \frac{close_t - MIN(t:t-m)}{MAX(t:t-m) - MIN(t:t-m)} \times 100$

이외에 주가 변동에 요인이 되는 등락구분 5일, 등락구분 20일, 환율, 환율 이동 평균 5일, 전 산업 생산 지수 등의 예측 요소를 사용하였다.

이러한 모든 예측 요소는 통계 및 데이터 마이닝을 위한 오픈 소스인 R의 neuralnet 신경망 패키지를 이용하여, 가장 좋은 예측 요소의 조합을 선택하였다.

III. 강화학습 모델

그림 2에 강화학습 구조에 대해 나타냈다. 강화학습 모델은 기본적으로 DQN 모델을 이용하였으며, 학습을 위한 에이전트인 메인 네트워크와 실제 값을 구하기 위한 타겟 네트워크 등으로 구성하였다. 또한 버퍼를 사용하여, 현재 상태, 다음 상태, 행동, 보상, 에피소드 종료 여부를 저장하여, 임의의 샘플을 추출하여 학습하였다.

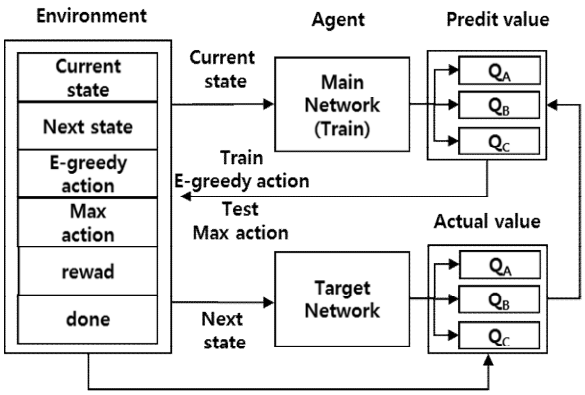


그림 2. 강화학습 모델
Fig. 2. Reinforcement learning model

에이전트의 상태 변수는 보다 안정적인 학습을 위해 시가, 고가, 저가, 종가, 거래량 등 기본적인 주가 데이터 뿐 아니라 기술 분석을 위한 보조 지표 데이터 그리고 변동 요인이 되는 요소를 상태 변수로 사용하였다. 표 2에 상태 변수에 대해서 나타냈다.

표 2. 상태 변수
Table 2. State variable

start price	close price	high price	low price
X1	X2	X3	X4
trading volume	Stochastic	CCI	fluctuations_5
X5	X6	X7	X8
fluctuations_20	exchange rate	exchange rate moving avg.	production index
X9	X10	X11	X12

상태 변수는 시가(X1), 종가(X2), 고가(X3), 저가(X4), 거래량(X5), Stochastic(X6), CCI(X7), 등락구분 5일(X8), 등락구분 20일(X9), 환율(X10), 환율 이동평균 5일(X11), 전 산업 생산지수(X12) 등 12개로서 각 네트워크 입력으로 하였으며, 출력은 매도, 매수, 보유로 하였다.

환경은 현재 상태, 다음 상태, 행동, 행동의 횟수, 이득 계산, 보상 그리고 에피소드 종료 여부 등이다. 상태는 현재 상태와 다음 상태로 당일 데이터와 익일 데이터이다. 당일 데이터는 메인 네트워크에 입력되어 예측 값과 행동을 추출한다. 익일 데이터는 타겟 네트워크에 입력되어 보상과 Q 값을 이용한 실제 값을 추출한다. 행동은 주식의 매도, 매수, 보유 등 3개의 행동으로 구분된다. 특히 학습 과정에서의 행동은 학습 초기에는 랜덤하게 행동하다 점차로 가장 큰 Q 값으로 행동하는 E-greedy 행동을 사용하였다. 이득 계산은 다음과 같이 각 행동의 횟수와 종가를 이용하였다.

$$profit = profit + (profile \times close) \tag{8}$$

강화학습에서 학습을 위해 가장 중요한 요소인 보상은 익일 이득에 당일 이득을 감산하였다.

$$reward = Nextprofit - Currentprofit \tag{9}$$

구현된 강화학습 모델의 학습 및 테스트 단계는 표 3과 같다.

그림 3에 에이전트로 사용할 인공신경망의 구성에 대해서 나타냈으며, 표 4에 최적화된 강화학습 파라메터에 대해서 나타냈다. 에이전트에서 상태 변수가 입력되는 입력층은 12노드, 중간층은 6노드로 하였으며, 출력층은 매도, 매수, 보유 등 3 노드로 구성하였다. 또한, 활성화 함수는 ReLU(Rectified Linear Unit)를 사용하였으며, 최적화 알고리즘은 오차 감소속도가 빠른 AdamOptimizer 알고리즘을 사용하였다. 실제 값을 추출하기 위해 할인 계수는 0.009로 하였으며, 버퍼의 사이즈는 5000개로 하여 랜덤하게 학습 데이터를 추출하여 학습하였다.

표 3. 학습 및 테스트 단계
Table 3. Training and testing steps

step	contents
1	Extracts the Q value with the current state as an input, multiplies the action by a scalar and outputs the predicted value.
2	Determines the action using the e-greedy method.
3	Obtains the reward by subtracting the current day's gain from the next day's gain.
4	Insert the current state, next state, action, reward, and termination of the episode into buffer.
5	The data of the buffer is extracted randomly, and the actual value is extracted using the next state and reward.
6	Computes the error with the actual value and predicted value, and learns the main network using the error.
7	After Training a certain stage, copy the weights from the target network to the main network.
8	After Training 50 times, tests the model using the test data.

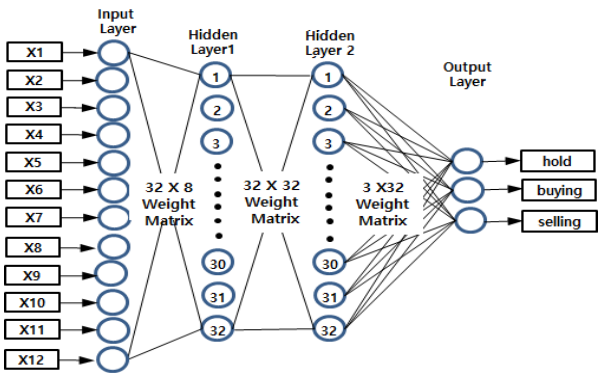


그림 3. 인공신경망
Fig. 3. Artificial neural network

표 4. 강화학습 파라메터
Table 4. Reinforcement learning parameters

parameters		parameters	
input nodes	12	learning rate	0.0001
hidden nodes	6	discount factor	0.009
hidden layer	2	replay buffer size	5000
output node	3	batch size	64

IV. 결과 및 비교 분석

실험환경은 윈도우 10 기반으로 파이선 3.6 기반 텐서플로우(Tensorflow) CPU Version을 이용하였다. 주가 데이터는 2009년 2월에서 2016년 12월 기간의 가격 안정성이 높은 KOSPI 200 종목에서 4 종목을 선택하였다. 학습 데이터는 시가, 종가, 고가, 거래량, Stochastic, CCI, 등락 구분 5일, 등락 구분 20일, 환율, 환율 이동 평균 5일, 전 산업 생산지수 등의 예측 요소 변수로 구성된 1966개의 레코드이다. 또한, 테스트 데이터는 10개, 20개, 30개의 레코드이다. 학습은 모든 주가 데이터가 50회 반복 후에는 변화가 없었기 때문에, 50회 반복 실험 하였다. 학습과 테스트는 보유금액은 100으로 시작하였다. 본 연구는 세 가지 강화학습 모델을 구축하여 학습결과를 비교하였다. 첫 번째 모델은 버퍼의 학습 데이터를 랜덤하게 추출하고, 하나의 인공신경망으로 학습하였다. 두번째 모델은 버퍼의 데이터를 순서적으로 추출하고, 두개의 인공신경망으로 학습하였다. 세 번째 모델은 버퍼의 데이터를 랜덤하게 추출하고 두개의 인공신경망으로 학습하였다. 모든 실험결과 큰 차이가 없었으나 세 번째 방법의 학습 결과가 가장 좋았다. 표 5에 실험결과를 나타냈다.

표 6에 세 번째 모델의 실험 결과를 나타냈다. 학습 결과는 50회 반복 후의 전체 평균이며, 5회 반복 실험 후 평균을 나타냈다. 표에서 test1은 테스트 데이터 10개의 결과, test2는 테스트 데이터 20개의 결과, test3은 테스트 데이터 30개의 결과이다.

학습결과 A 종목의 학습 평균은 1265.093이며, B 종목은 715.428이다. 또한, C 종목은 70099.96이며, D 종목은 11444.79이다. 전반적으로 학습 결과가 우수했으며, 특히 C 종목과 D 종목의 학습 결과가 좋았다. 10개 레코드의 테스트 결과 A 종목의 테스트 평균은 101.65이며, B 종목은 100.33이다.

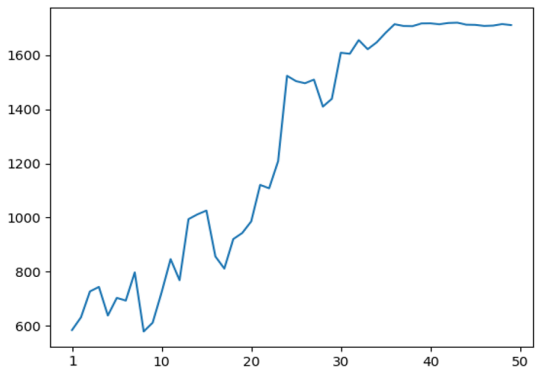
또한, C 종목은 104.6이며, D 종목은 105.65이다. 20개 레코드의 테스트 결과 A 종목의 테스트 평균은 102.05이며, B 종목은 100.485이다. 또한, C 종목은 122.5이며, D 종목은 115.6이다. 30개 레코드의 테스트 결과 A 종목의 테스트 평균은 103.15이며, B 종목은 101.27이다.

표 5. 강화학습 모델의 학습 결과
Table 5. Training result of reinforcement learning model

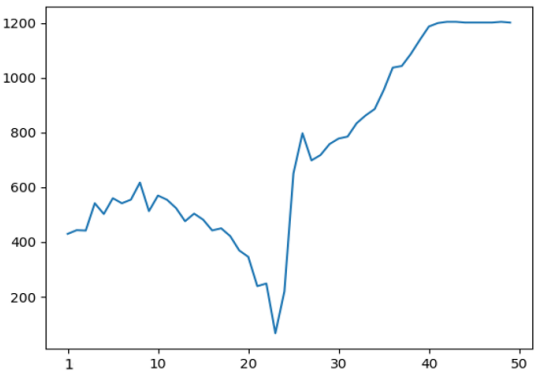
Item	model 1	model 2	model 3
A	1243.0	1255.08	1275.68
B	724.88	715.42	755.07
C	65914.32	68309.69	70229.97
D	11277.18	11341.96	12134.5

표 6. 학습 및 테스트 결과
Table 6. Training and testing results

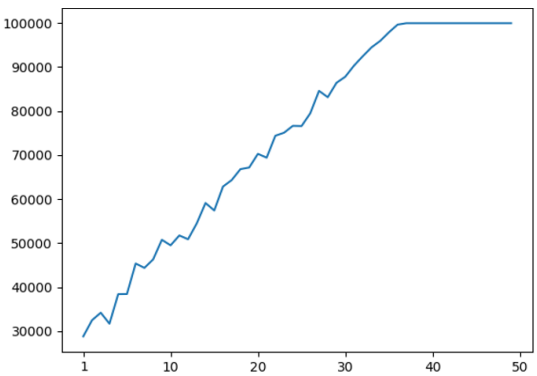
Item		1	2	3	4	5	Average
A	train	1254.7684	1243.629	1296.3746	1260.2824	1270.4104	1265.0930
	test1	101.65	101.65	101.65	101.65	101.65	101.65
	test2	102.05	102.05	102.05	102.05	102.05	102.05
	test3	103.15	103.15	103.15	103.15	103.15	103.15
B	train	671.9992	735.3846	677.0087	770.9809	721.7667	715.428
	test1	100.33	100.33	100.33	100.33	100.33	100.330
	test2	100.485	100.485	100.485	100.485	100.485	100.485
	test3	101.27	101.27	101.27	101.27	101.27	101.27
C	train	69840.458	70146.988	69969.43	70001.307	70541.626	70099.96
	test1	104.6	104.6	104.6	104.6	104.6	104.6
	test2	122.5	122.5	122.5	122.5	122.5	122.5
	test3	209.5	209.5	209.5	209.5	209.5	209.5
D	train	11879.648	11529.477	11350.737	11216.353	11247.733	11444.79
	test1	105.65	105.65	105.65	105.65	105.65	105.65
	test2	115.6	115.6	115.6	115.6	115.6	115.6
	test3	116.95	116.95	116.95	116.95	116.95	116.95



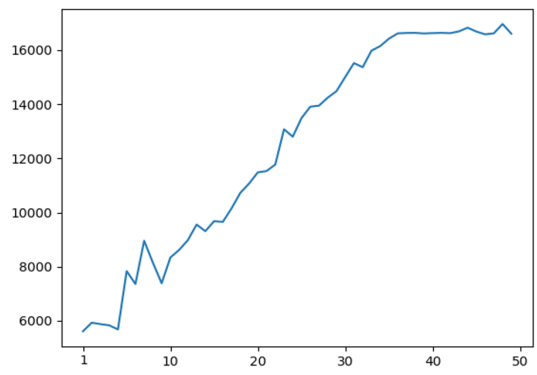
(a) A stock



(b) B stock



(c) C stock



(d) D stock

그림 4. A 종목, B 종목, C 종목, D 종목의 학습 과정
Fig. 4. Training process of A stock, B stock, C stock and D stock

또한, C 종목은 209.5이며, D 종목은 116.95이다. 테스트 결과 종목마다 많은 차이가 있었으며, 테스트 레코드가 많을수록 테스트 결과 값이 좋았다. 특히 C 종목과 D 종목의 테스트 결과가 좋았다.

그림 4에 학습 과정을 나타냈다. 그림에서 x좌표는 반복 횟수이며, y좌표는 이득을 나타낸다. 모든 종목이 40회 반복 학습 후에는 이득이 증가하지 않았다. A 종목은 이득의 변동 폭이 크게 나타났으며, 약 17배의 이득을 보였다. B 종목은 10회 반복에서 22회 반복 학습에서 큰 폭으로 이문이 큰 폭으로 감소하였으나, 최종적으로 약 12배 정도의 이득을 보였다. C 종목과 D 종목은 선형적으로 이득이 증가하였으며, 약 1000배와 160배의 높은 이득을 보였다.

V. 결론 및 향후 과제

본 연구는 주가의 기본 데이터와 보조 지표 데이터 뿐 아니라 주가 변동 요소 데이터를 이용하여, 주식 트레이딩 예측을 위한 강화학습 모델을 구현하였다. 실험 결과 강화학습이 트레이딩에 적합한 행동을 하였음을 확인할 수 있었다. 또한, 종목별로 학습 결과와 테스트 결과가 큰 차이가 있었으며, 학습 결과가 좋은 종목이 테스트 결과도 좋았다. 이것은 종목별로 주가의 패턴에 기인한 것으로 생각된다. 또한 안정적이고 선형적으로 학습이 된 종목이 테스트 결과도 좋았다. 따라서 강화 학습 결과의 패턴을 고려하여 테스트 결과를 추정할 수 있을 것으로 기대된다. 향후 기술 분석 뿐 아니라 기업의 기본 분석을 통해 기업의 내재 가치 분석을 상태 변수로 사용한다면 보다 효율적인 트레이딩 행동이 가능할 것이다.

References

[1] K. Y. Kim and K. R. Lee, "A study on the prediction of stock price using artificial intelligence system", The Korean Journal of Business Administration, Vol. 21, No. 6, pp. 2421-2449, Dec. 2008.

[2] E. J. Lee, C. H. Min, and T. S. Kim, "Development of the KOSPI (Korea Composite Stock Price Index) forecast model using neural

network and statistical methods", The Journal of the Institute of Electronics Engineers of Korea-CI, Vol. 45, No. 5, pp. 95-101, Sep. 2008.

[3] C. Hsu, "A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming", Expert Systems with Applications, Vol. 38, No. 11, pp. 14026-14036, Oct. 2011.

[4] J. W. Lee, "A stock trading system based on supervised learning of highly volatile stock price patterns", The Journal of Korean Institute of Information Scientists and Engineers : Computing Practices and Letters, Vol. 19, No. 1, pp. 23-29, Jan. 2013.

[5] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market prediction", The European Journal of Operational Research, Vol. 270, No. 2, pp. 654-669, Oct. 2018.

[6] H. J. Song and S. J. Lee, "A study on the optimal trading frequency pattern and forecasting timing in real time stock trading using deep learning: focused on KOSDAQ", The Journal of KAIS, Vol. 27, No. 3, pp. 123-140, May 2018.

[7] J. W. Lee, "Short-term stock price prediction by supervised learning of rapid volume decreasing patterns", Korean Institute of Information Scientists and Engineers Transactions on Computing Practices, Vol. 24, No. 10, pp. 544-553, Oct. 2018.

[8] J. M. Won, H. S. Hwang, Y. H. Jung, and H. D. Park, "Stock Price Prediction Technique Using Technical Analysis Index and Deep Running", Conference of Korean Institute of Information Technology, pp. 404-405. Nov. 2018.

[9] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning : A survey", The Journal of Artificial Intelligence Research, Vol. 4, pp. 237-285, May. 1996.

[10] V. Mnih, et al. "Human-level control through deep reinforcement learning", Nature, Vol. 518, No. 26, pp. 529-541, Feb. 2015.

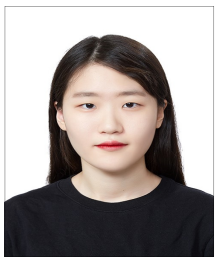
[11] D. H. Shin, K. H. Choi, and C. B. Kim, "Deep Learning Model for Prediction Rate Improvement of Stock Price Using RNN and LSTM", Journal of KIIT, Vol. 15, No. 10, pp. 9-16, Oct. 31, 2017.

[12] T. J. Hsieh, H. F. Hsiao, and W. C. Yeh, "Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm", Applied Soft Computing, Vol. 11, No. 2, pp. 2510-2525, Mar. 2011.

[13] K. H. Park and H. J. Shin, "Stock price prediction based on time series network", Korean Management Science Review, Vol. 28, No. 1, pp. 53-60, Mar. 2011.

저자소개

하 은 규 (Eun-Gyu Ha)



2016년 3월 ~ 현재 : 가천대학교 IT대학 에너지 IT학과 재학

관심분야 : 딥러닝, 빅 데이터, AI, 사물인터넷

김 창 복 (Chang-Bok Kim)

1986년 2월 : 단국대학교 전자공학과(공학사)

1989년 2월 : 단국대학교 전자공학과(공학석사)

2009년 2월 : 인천대학교 컴퓨터 공학과(공학박사)

1994년 ~ 현재 : 가천대학교 IT대학 에너지 IT학과 교수

관심분야 : 데이터 마이닝, 딥러닝, 강화학습, 사물인터넷