

# Monte Carlo 알고리즘에 의한 주가 추세 평가

이재원

성신여자대학교 컴퓨터정보학부  
jwlee@cs.sungshin.ac.kr

## Evaluating stock price trends by Monte Carlo algorithm

Jae-Won Lee

School of Computer Science and Engineering, Sungshin Woman's University

### 요 약

본 논문에서는 환경의 변화에 민감한 시계열 데이터인 주가의 변동 과정을 보다 효과적으로 설명하기 위한 방안의 하나로 강화 학습 모형의 도입을 제안하며, 특정 시점의 주가 추세를 평가하는 기준으로 가치도 함수를 채택한다. 가치도 함수의 계산에는 강화 학습 알고리즘의 일종인 Monte Carlo 알고리즘을 적용하고, 신경망 구현에 의해 구한 근사 가치도의 적절성을 평가하였다. 실험 결과로 볼 때, 여타 강화 학습 알고리즘을 추가적으로 적용할 경우, 주가 변동의 시계열적 특성을 더욱 잘 반영할 수 있을 것으로 판단된다.

### 1. 서론

과거의 주가 추세로부터 미래의 주가 움직임을 예견하려는 노력은 다우, 그린벨, 엘리엇트 등에 의해 오래전부터 시도되었고, 많은 주가 기술 분석 지수들이 개발되어 왔다. 또한 최근에는 컴퓨터의 발달로 인해 인공지능 분야의 기법들을 주가의 기술 분석에 적용한 사례들도 다수 제시되고 있다[1][2]. 그러나 이 사례들에서 주로 채택하고 있는 지도 학습(supervised learning)은 데이터의 시간적 선후 관계가 매우 중요하게 작용하고, 시간에 따라 변화하는 환경에 놓이게 되는(non-stationary) 주가의 변동을 학습하는데 있어서는 비효율적인 측면이 많다.

본 논문에서는 행위자와 환경의 상호작용을 시간적 흐름의 측면에서 분석하는데 보다 효율적인 것으로 알려진 강화 학습(reinforcement learning)[3]의 개념을 도입하여 주가 변동 과정을 설명하였고, 이를 주가 추세의 해석에 활용하기 위해 Monte Carlo 알고리즘을 적용하였다. Monte Carlo 알고리즘은 학습의 대상이 되는 모형(model)의 완전한 확률 분포가 주어지지 않거나 이를 구하기 힘든 경우에 유용한 알고리즘으로, 모형의 모사(simulation)에 의해 실제로 경험하게 되는 상황들만을 사용하여 학습을 진행한다[4]. 강화 학습의 최종 목적은 행위자의 행위 선택에 적용할 최적의 정책을 구하는 것, 즉 '제어(control) 문제'를 해결하는 것이지만, 정책의 학습 과정에서 구하게 되는 각 상태(state)의 '가치도(value)'가 기대되는 '보상(reward)'을 표현할 수 있기 때문에 '예측(prediction) 문제'의 해결에 적용되는 경우도

많다.

Monte Carlo 알고리즘을 주가 추세의 예측 문제에 적용하기 위해, 본 논문에서는 특정 시점에서의 주가 추세를 한정된 수의 특성(feature) 집합에 의해 상태로 표현하고, 신경망에 의해 각 상태의 가치도의 근사값을 구하는 실험을 수행하였다. 실험 결과, 본 논문에서 제안하는 가치도가 실제 주가 변화율에 비례하는 수치를 나타냄으로써 해당 주식의 주가 예측에 직접 활용될 수 있음을 확인하였다.

### 2. Monte Carlo 알고리즘

Monte Carlo 알고리즘은 강화 학습에 사용되는 알고리즘의 하나로 그림 1과 같이 환경(environment)과 상호작용하는 행위자(agent)의 행위 선택에 사용할 최적 정책의 획득에 활용할 수 있다.

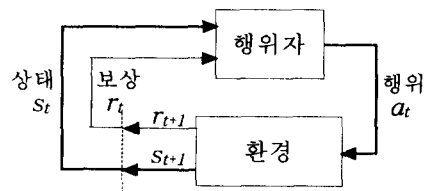


그림 1. 행위자와 환경의 상호 작용

위 그림의 모형을 마르코프 의사결정 프로세스(Markov decision process, MDP)<sup>1)</sup>에 기반하여 설명하자면 다음과 같다. 행위자는 현재 상태  $s_t$ 를 감지하고, 현재 행위  $a_t$ 를 선택하여 수행한다. 환경은 응답으로 행위자에게 보상  $r_{t+1}=r(s_t, a_t)$ 를 주고 다음 상태  $s_{t+1}=\delta(s_t, a_t)$ 를 생성한다. 상태의 집합을  $S$ , 행위의 집합을  $A$ 라고 할때, 행위자의 목적은 현재 관찰된 상태  $s_t$ 에 기반하여 다음 행위  $a_t$ 를 선택하기 위한 정책  $\pi:S \rightarrow A$ (즉,  $\pi(s_t)=a_t$ )를 학습하는 것이다.

임의의 초기 상태  $s_t$ 로부터 임의의 정책  $\pi$ 에 의해 얻게되는 누적값  $V^\pi(s_t)$ 를 아래 식과 같이 정의하며, 상태  $s_t$ 의 가치도라고 부른다.

$$V^\pi(s_t) \equiv r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots$$

$$\equiv \sum_{i=0}^{\infty} \gamma^i r_{t+i} \quad (1)$$

행위자의 학습은 모든 상태  $s$ 에 대해  $V^\pi(s)$ 를 최대화하는 정책  $\pi$ 를 학습하는 것이며 이러한 정책을 '최적 정책(optimal policy)'이라고 부르며  $\pi^*$ 로 표기한다.

$$\pi^* \equiv \operatorname{argmax}_{\pi} V^\pi(s), (\forall s) \quad (2)$$

$\pi^*$ 를 얻기 위해서는 그림 2와 같은 일반화된 정책 순환(Generalized Policy Iteration, GPI) 과정이 필요하다. 이 과정은 현재의 정책  $\pi$ 를 사용하여 각 상태의 가치도를 계산함으로써 정책을 평가하고, 갱신된 가치도를 다시 정책의 개선에 사용하는 순환 과정으로, 이에 의거하여 강화 학습 알고리즘들을 분류할 수 있다.

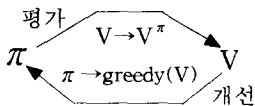


그림 2. 일반화된 정책 순환 과정

Monte Carlo 알고리즘은 모형의 완전한 확률 분포를 사용하는 대신 경험(학습 예제)만을 사용한다는 점에서 동적 프로그래밍(Dynamic Programming)과 구분되며, 각 시간 단계마다 가치도를 갱신하지 않고 최종적인 보상을 알게되는 에피소드<sup>2)</sup>의 끝에서만 가치도를 갱신한다는 점에서 TD(Temporal Difference) 알고리즘과 구분된다 [5].

결국 Monte Carlo 알고리즘의 궁극적인 목적은 최적의 정책을 구하는 것이라 할 수 있지만, 전문가에 의해 일련의 최적 행위가 예제로 주어지는 경우, 즉  $\pi^*$ 가 주어진다든 전제하에서는, Monte Carlo 알고리즘을 '평가(evaluation)' 단계에 적용하여 각 단계의 가치도를 구

하고 이를 '예측(prediction)' 문제에 적용할 수 있다. 그림 3은 이러한 상황에 적용할 수 있는 Monte Carlo 알고리즘을 기술하고 있다. 여기서 Returns는 하나의 에피소드의 끝에서 얻게 되는 최종적인 보상을 나타낸다.

```

Initialize:
  π ← policy to be evaluated
  V ← an arbitrary state-value function
  Returns(s) ← an empty list, for all s ∈ S
Repeat forever:
  (a) Generate an episode using π
  (b) For each state s appearing in the episode:
      R ← return following the first occurrence of s
      Append R to Returns(s)
      V(s) ← average>Returns(s)
    
```

그림 3. 예측 문제를 위한 Monte Carlo 알고리즘

### 3. 주가 추세의 가치도 평가

주가의 과거 추세가 미래에 영향을 미친다는 경제학 분야의 연구 결과를 믿는다면, 시계열 데이터인 주가의 변동을 시간적 선후 관계를 고려할 수 있는 Monte Carlo 알고리즘에 의해 학습해보는 것이 유의미한 시도일 수 있다. 이를 위해 먼저 주가의 추세 변화에 내포된 요소들을 앞서 설명한 강화 학습 모형의 각 요소에 대응시키고 이를 기반으로 하여 특정 시점의 주가 추세에 대한 가치도의 평가 방안을 제시한다.

#### 3.1 주식 시장의 강화 학습적 정의

주식 시장에서의 행위는 시장에 참여하는 매매 주체들에 의해 집단적으로 이루어진다. 물론 이러한 행위에는 무수한 변인이 작용하며 심지어는 시장이 속한 사회 전반의 정치, 경제, 사회, 문화적 요인이 총체적으로 작용하기도 한다. 그러나 엘리엇은 이러한 매매 행위에 가장 결정적인 영향을 미치는 것이 이들 매매 주체의 군중적 심리이며, 이것이 주가의 과거 추세에 투영될 뿐만 아니라 미래 추세의 진행이 피보나치 수열과 황금 분할 비율로 표현되는 자연의 질서를 벗어나지 않으므로 미래의 주가 파동(wave) 예측에 과거의 주가 추세를 활용할 수 있다고 주장하였고[6], 엘리엇 사후 60여년이 지난 지금에도 그의 '파동 이론'은 매우 유용한 기술 분석 이론으로 받아들여지고 있다.

파동이동에 의하자면, 결국 주식 시장에서의 매매는 특정 시점에서 주가 추세(상태)에 대한 매매자 집단(행위자)의 행위 선택 과정이며, 그 결과로 주가의 변동(보상)이 이루어지는 것이 된다. 본 논문에서는 이러한 관점에서 상태를 전일까지의 주가 추세로, 행위자를 당일 시장에 참여한 매매자 집단으로, 행위는 각 집단의 투자 원칙이라는 정책에 의한 주식의 매매로, 그리고 보상은 전일 대비 당일 주가 변화율로 정의한다. 이 정의에 의하면, 일반적인 주가 데이터(차트)는 상태와 보상의 연속적인 사례를 보여줄 뿐 구체적인 행위와 정책에 대하여는 아무런 정보도 제공하지 않는다. 그러나 하루의 주가 데이터를 그림 3에서의 (a) 단계로 간주하면(이 경우

1) Monte Carlo 알고리즘이 직접적으로 MDP를 표현하지는 않지만, 강화 학습은 일반적으로 MDP를 가정한다.  
 2) 행위자와 환경의 상호작용을 분할하는 단위로 상호작용의 일 단락을 의미한다.

에피소드의 길이는 1이 된다), 상태와 보상에 대한 일련의 사례만으로도 Monte Carlo 알고리즘의 적용이 가능하게 된다. 따라서 본 논문에서는 정책과 행위는 학습 입력의 일부로만 간주하고, 학습의 목표를 주가 추세의 현재 상태에 대한 가치도의 평가로 국한한다.

3.2 주가 추세의 상태 표현

주가 추세를 표현하기 위해 주가 데이터로부터 추출 및 계산 가능한 추세 관련 수치 특성 요소들로 구성된 (3)의 특성 벡터를 사용한다. 표 1은 이러한 특성들의 예이다. 각 특성의 가능한 모든 조합을 상태 집합으로 정의한다.

$$\vec{\phi}_s = (\phi_s(1), \phi_s(2), \dots, \phi_s(n))^T \quad (3)$$

표 1: 특성의 예

전일 대비 주(종)가(거래량) 변화율
5일 이동평균값 기준 (거래량) 이격
10일 이동평균값 기준 (거래량) 이격
20일 이동평균값 기준 (거래량) 이격
60일 이동평균값 기준 (거래량) 이격
...
EOM(Ease of Movement) 지수
EQUIVOLUME 지수
William's %R 지수
MACD 지수
...

3.3 신경망에 의한 가치도의 근사

3.2절의 상태 정의에 의하면 상태의 수가 유한하지 않기 때문에, 함수 근사 방식에 의한  $V^r(s)$ 의 일반화가 요구된다. 이를 위해서는 여러 가지 방법이 가능하지만, 본 논문에서는 (4)로 표현되는, 특성 벡터에 대한  $V^r(s)$ 의 선형 근사 함수  $V_r(s)$ 를 구하기 위해, (5)에 의해 연결 강도를 갱신하는 신경망을 사용한다.

$$V_r(s) = \vec{\theta}_r^T \vec{\phi}_s = \sum_{i=1}^n \theta_r(i) \phi_s(i) \quad (4)$$

$$\vec{\theta}_{r+1} = \vec{\theta}_r + \alpha [R - V_r(s)] \vec{\phi}_s \quad (5)$$

4. 실험

3장에서 제시한 가치도 함수의 효용성을 평가해보기 위해, 한국 거래소 시장 상장 주식중 KOSPI200 지수 산출에 포함되는 종목들로 구성된 데이터를 사용하였으며, 그 내용은 표 2와 같다.

표 2: 실험 데이터

학습 데이터		테스트 데이터	
종목수	100	종목수	50
기간	최근 2년	기간	최근 1년

$V_r(s_r)$ 의 크기를 32단계로 구분하고(이는 한국 주식

시장의 하루 최대 주가 변동율이 -15%~15%인 점을 고려한 것이다), 32 개의 신경망 출력 노드 중 해당 노드의 목표값을 1로 나머지는 모든 노드의 목표값을 모두 0으로 하여 실험하였다. 기본적으로 주가의 변동이 연속적 작업(continuous task)임에도 불구하고, 길이가 1인 에피소드로 간주한 본 논문의 실험에서도 가치도와 실제 주가 변화율이 비례하는 것으로 나타났으며 세부적인 내용은 표 3과 같다.

표 3: 가치도 등급과 실제 주가 변화율

가치도 등급	일일 주가 변화율	데이터수	가치도 등급	일일 주가 변화율	데이터수
1	-5.90	34	17	0.44	782
2	-5.13	79	18	1.89	667
3	-5.40	161	19	1.19	575
4	-5.18	136	20	-0.09	539
5	-4.66	140	21	-1.44	399
6	-4.53	207	22	2.24	391
7	-3.21	230	23	3.13	368
8	-3.66	257	24	2.70	276
9	-1.95	249	25	3.17	189
10	-2.32	322	26	3.02	207
11	-2.53	493	27	2.18	211
12	-2.07	437	28	5.08	161
13	-0.96	391	29	4.04	172
14	-1.25	506	30	5.23	92
15	0.15	549	31	5.40	46
16	-0.33	690	32	5.18	59

5. 결론

본 논문에서는 주가 추세의 현재 상태를 평가하기 위해 강화 학습 모형을 도입하였고, Monte Carlo 알고리즘의 신경망 구현에 의해 가치도 함수의 효용성을 검증하였다. 이 방식에 의해 산출된 가치도는 그 값 자체가 해당 주식의 기대 변화율에 비례한다는 점에서, 이진 매체 신호(signal)의 발생에 주력하는 기존의 경제학적 기술 지수와 구별된다. 이후 주식 시장에서의 행위와 최적 정책에 대한 모형화, TD 알고리즘과 같은 다 강화 학습 알고리즘의 적용, 개선된 신경망 구조의 사용 등이 유의미한 연구 과제가 될 수 있을 것으로 판단된다.

6. 참고 문헌

- [1] R. J. KUO, A Decision Support System for the Stock Market through Integration of Fuzzy Neural Networks and Fuzzy Delphi. *Applied Artificial Intelligence*. 1998.
- [2] BaBa, N., and M. Kozaki. An Intelligent Forecasting System of Stock Price using Neural Networks. *Proc. IJCNN*, Baltimore, Maryland. IEEE Press. 1992.
- [3] T. M. Mitchell. *Machine Learning*, The McGraw-Hill Companies, Inc. 1997.
- [4] Kalos, M. H., and Whitlock, P. A. *Monte Carlo Methods*, Wiley, New York. 1986.
- [5] R. S. Sutton, and A. G. Barto. *Reinforcement Learning*. The MIT Press. 1998.
- [6] T. S. Eliot. *The nature's law(The Secret of the Universe)*. 1946.