

<제목 페이지>

안녕하세요. 8조 정원균, 최동현입니다. bitcoin trading Agent with Reinforcement Learning에 대해 발표 시작하겠습니다.

<목차>

발표 진행은 프로젝트 소개를 거쳐 어떻게 design하고 setting하였는지. 그리고 결과와 결론으로 이어가겠습니다.

<introduction>

2017년부터 비트코인을 비롯한 암호화폐에 대한 관심이 높아져 2017년 초 800만원 대였던 비트코인 가격이 2021년 8000만원대까지 최소 10배 이상 올랐습니다. 최대 50% 이상의 가격 변동성을 기준으로 다음 상태의 비트코인의 가격을 쉽게 예측할 수 없는 경우가 많습니다.

비트코인을 왜 강화학습으로 학습할까요? RNN 또는 CNN을 사용한 비트코인 미래 가격 예측 시도가 많이 있음을 확인할 수 있었습니다. 하지만 이러한 미래 가격의 단순한 예측은 암호화폐 환경의 복잡성과 변동성을 감안할 때, 가격 예측이 쉽지 않았습니다. 한 가지 가능한 해결책으로 clever한 deep neural network 옵티마이징 정책과 함께 강화 학습을 사용하는 것입니다.

<Dataset>

저희는 암호화폐에서도 메인인 비트코인 Data를 활용하였고, Data는 Yahoo Finance에서 다룬 받았습니다.

2014년 11월 28일부터 2022년 3월 1일까지의 일일 단위 00시 기준 데이터를 활용하였습니다. 가격은 시작가인 open column의 데이터를 사용했습니다.

오른쪽 데이터는 pyupbit library를 통해 최근 200시간의 한 시간 단위 데이터를 불러온 결과입니다. US달러 단위로 환산하여 두 데이터를 맞춰주었고, 이 데이터는 추후에 result 부분에서 최근 비트코인 가격을 예측하는데 활용하겠습니다.

<design1>

저희 프로젝트에서는 Deep-q-learning 알고리즘으로 학습시켰습니다. 큐-러닝은

- agent가 취할 수 있는 state의 수가 많은 경우, Q-table 구축에 한계를 가지게 되고,
- 순차적인 sample data 간의 correlation으로 인해 학습이 어렵고,

또한 큐-러닝은 미래에 대한 평가가 과평가 되는 경우가 있습니다.

///따라서 이를 보완하는 딥큐러닝을 사용하여 학습하였습니다.

<design2>

오른쪽 다이어그램은 trading decision 모델 간략도입니다. Buy, Hold, Sell 총 3가지의 Decision을 하게 되며, 비트코인을 구매할 땐 가능한 많이 사고 자금이 없으면 사지 않습니다. Hold 과정에서는 아무것도 하지 않고, 비트코인을 팔 때 모든 코인을 다 판매하거나 가진 비트코인이 없으면 아무것도 하지 않습니다.

action에서의 변수 m 은 남아있는 현금을 의미하고, r 은 수수료의 비율을 의미하고, P 는 각 암호화폐의 가격을 의미합니다. N 은 각 암호화폐의 주식수입니다.

저희의 리워드 모델은 다음 state의 보유한 비트코인과 남은 자산의 합에서 이전 state의 합을 뺀 값으로 리워드를 주었습니다.

<질문1>

-비트코인을 사고파는 것에 대한 기준을 알 수 있을까요? 예를 들면 사고 파는 모델이 시간 인건가요?

답 : 비트코인을 사고파는 기준은 설계한 trading decision 다이어그램을 기반으로 작동합니다. reward model과 window size 변수를 적용하여

레코드 모델에 이전 일수를 취하는 window size 변수를 적용했고,

적용한 window size를 기반으로

창 크기는 레코드 모델이 조치를 취하는 것으로 간주하는 이전 일수를 나타냅니다.

<질문2>

딥큐에서 experience replay를 사용하는데, 실제로 과평가 하지 않고, 옵티멀하게 잘 수렴할 수 있도록 하기 때문에 로컬 옵티멀에 빠지지 않기 때문에 그래서 딥큐 사용

<질문3>

추가적으로 저희가 고민한 부분이 있는데,

-비트코인이 가격 변동성이 크지만 초기 가격과 요즘 가격의 차이가 크기 때문에 장기투자 모델의 경우 어떻게 설계할 수 있는지. 고민해봤습니다.

1. 저희는 현재 단기기간의 optimal한 리워드 모델을 기반으로 작동하는데, 장기적으로는 현재 얻는 리워드 보다는 미래에 얻는 리워드를 더 크게 보는 방법이 있고.

2. window size와 exploration계수를 조정해서

우리가 찾은 optimal한 것보다 좋은 optimal이 있을 수 있기 때문에 미래를 보는 방향으로 모델을 설계하면 될 것 같다.