# 深層強化学習によるFXの最適化に関する研究

平石研究室

S20168

諏訪航平

# 深層強化学習とは

- 深層強化学習とは, 深層学習を取り入れた強化学習の手法であり, 強化学習よりも複雑な環境での動作が可能である.
- 強化学習は, 一般的にはマルコフ決定過程に従うベルマン方程式をモデル化したシステムであり, 直前の状態と行動のペアのみに依存して現在の状態が決まる.
- エージェントが行う行動は方策に基づき, 方策は将来得られる報酬が最大になるように更新する. 最適な方策を見つけることが, 報酬の最大化につながる.
- 学習の主体となるエージェントは, 環境からの情報である状態変数から方策を学習する.

# 深層強化学習とは

- 価値関数は割引報酬和の期待値であり, これをすべての状態において最大になるように方策を学習することが目標となっている.

# PPOアルゴリズムとは

- PPOは, Proximal Policy Optimizationの略であり,

強化学習の一種であり, 価値ベース（価値関数を基に方策を決定する手法）と方策ベース（直接, 方策を更新する手法）の両方を取り入れた手法である.

# 使用するライブラリとAPI

- 今回はstablebaselines3 1)を使用する. これは強化学習ライブラリで, pytorchベースで作成されているものである. Stablebaselines 2)という機能が似通ったライブラリが存在するが, これはtensorflow1.xベースで作成されており, google colabolatoryで実行可能なのはstablebaselines3である.

# 今回作成するシステム

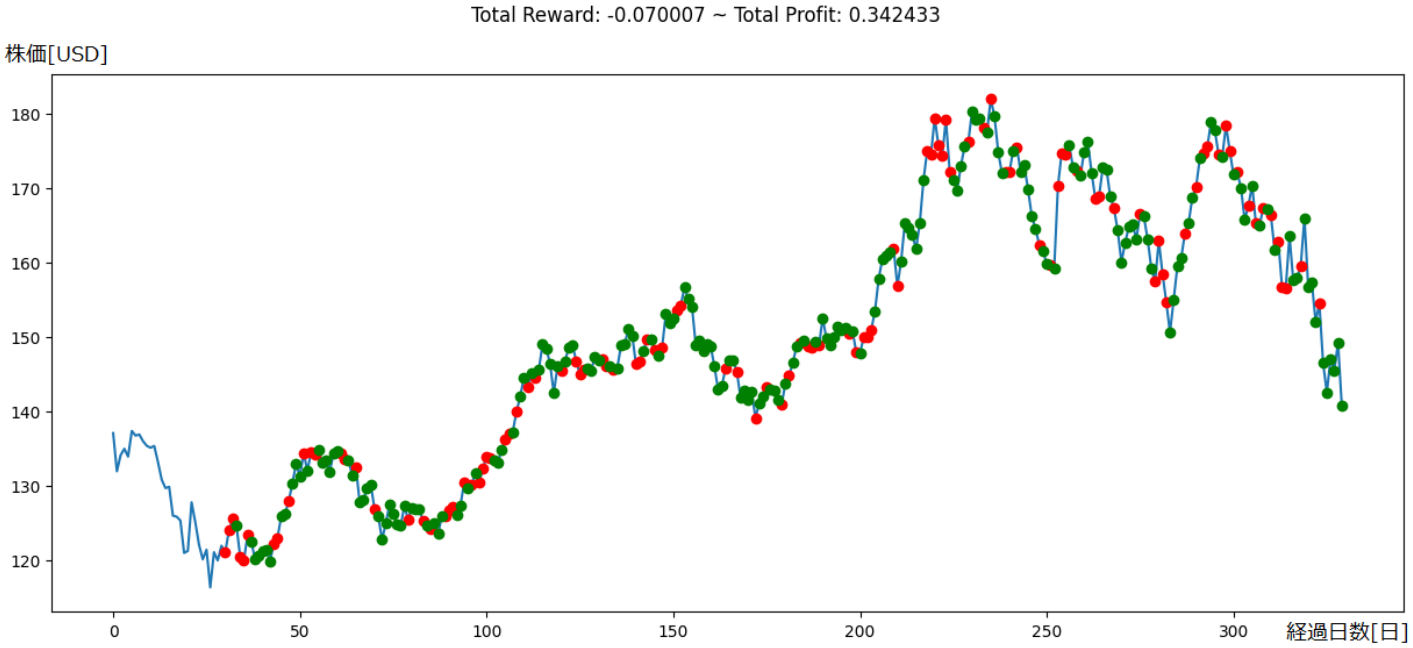
- 今回はMLP（多層パーセプトロン）とPPOアルゴリズムを組み合わせた深層強化学習手法を使用.
- Google Colaboratory上で実行する.
- アップルの株価300日分を使用.
- 状態変数としてはlongとshort, その他, 始値, 終値, 高値, 安値, 出来高, 調整後終値をとる.

# 今回作成するシステム

- 行動はbuyとsellの2種類. の出力されるグラフでは緑の点がbuyの行動, 赤の点がsellの行動を表している.

# 実行結果

- 実行結果はこのようになり, 累計報酬は-0.070007, 総利益は0.342433で, 使用した原資のうち34%程度しか回収ができていないことが分かった.



# まとめ

- 今回はアップルの株価に対するMLPとPPOでの実行のみとなったため, 今後はFXに対して行い, LSTMや別の強化学習アルゴリズム, そして, 状態変数を増やすことや, ハイパーパラメータの変更等を試して収益を上げたいと考えています.

# 参考文献


- 1)GitHub – DLR-RM/stable-baselines3,
- https://github.com/DLR-RM/stable-baselines3,
- (access 2023.7.20).

- 2)GitHub – hill-a/stable-baselines,
- https://github.com/hill-a/stable-baselines,
- (access 2023.7.20).