

深層強化学習による FX の最適化に関する研究

S20168 諏訪航平

1. はじめに

投資銀行や証券会社などにおいて、様々な金融商品の取引が行われているが、近年では、自動化が進み、AI がこれを担っているといえる。今回は、比較的情報の入手がしやすい外国為替取引(FX)の自動化、そして、利益の最大化を目指すにあたり、サンプルコードを活用しながらプログラムの制作を進めた。

2. 使用するライブラリと API

今回は強化学習ライブラリとして、`stablebaselines3`¹⁾と`stablebaselines`²⁾を利用した。これらは、ほぼ同じように関数呼び出しができる等、操作は非常に似通っている。ただし、後者のほうが対応するモデルの種類が豊富である。前者は `pytorch` ベース、後者は `tensorflow1.x` ベースで作成されているため、クラウドの実行環境である `Google Colaboratory` で実行できるのは、現時点では `stablebaselines3` のみであり、本研究では、`stablebaselines3` により実験を実施した。

後期からは `investing.com` から経済指標等と呼び出して使用する予定だが、今回のサンプルコードにおいては、`yahoo finance` からの取得した 1 日ごとのアップルの株価データを用いた。これは `dataframe` 型のデータである。

3. 使用するモデル

今回のプログラムで使用するモデルは PPO と MLP を組み合わせたモデルである。PPO は、Proximal Policy Optimization の頭文字をとったもので、価値ベースと方策ベースの両方の特性を持った強化学習のアルゴリズムである。MLP は多層パーセプトロンのことであり、深層学習を担う部分である。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。`long` はロングポジションの略で株価の上昇により利益を得られる状態で、`short` はショートポジションのことで株価の下落により利益を得られる状態である。また、行動は `buy` と `sell` の二つになっている。

4. プログラムの実行結果

今回の実験では MLP と PPO を組み合わせたモデルで、2020 年 1 月 1 日から 300 日間のデータを使用した。ウィンドウサイズを 30 日、タイムステップを 120000 とし、`Google Colaboratory` 上で実行した。

結果は図 1 のようになり、赤い点がエージェントの `sell` の行動、緑の点が `buy` の行動を意味する。ウィンドウサイズが 30 日に設定したため、最初の 30 日間分はエージェントの状態として扱われるため、エージェントの行動はない。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。

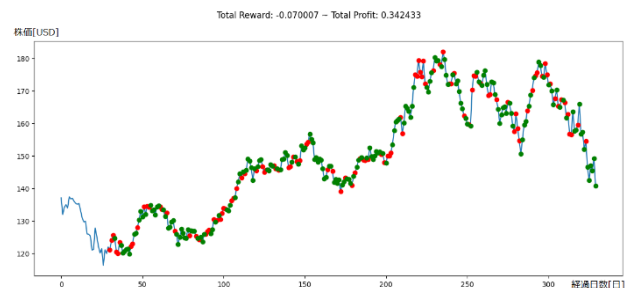


図 1 MLP と PPO での実行結果

5. まとめ

本研究では、サンプルプログラムによる実験を行ったが、ハイパーパラメータの設定が上手くできていないためか、全く利益を出すことが出来なかった。今後、ハイパーパラメータの調整などを行い、株価ではなく FX での実行を行なっていく予定である。

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).