深層強化学習による機会損失を考慮した株式投資戦略の構築

Deep Reinforcement Learning Based Stock Trading Strategy Considering Opportunity Loss

井上 修一[1]　穴田 一[1]

Shuichi Inoue[1], Hajime Anada[1]

[1] 東京都市大学大学院

[1]Tokyo City University Graduate School

Abstract: In recent years, research on financial transactions using deep reinforcement learning, which is one of machine learning, has been actively conducted. In these studies, various approaches have been taken, such as those that consider the number of financial instruments bought and sold, compound interest calculation, and those that use stock price charts for input, but enough profit has not been made in all periods. It is considered that this is because the opportunity loss cannot be taken into consideration. Therefore, in this study, we build a model to learn the optimal buying and selling timings to make a profit in stock investment by incorporating the opportunity loss for each action into the reward in deep reinforcement learning, and show its effectiveness.

はじめに

近年，機械学習を用いた金融取引に関する研究が精力的に行われている．その中には，2016 年に囲碁のプロを打ち負かした AlphaGo で話題になった深層強化学習を用いて金融取引戦略を構築する研究が存在する．ここでの深層強化学習は，エージェントが試行錯誤を通して設定された環境において利益を最大化するための行動を学習させるものである．これらの研究では，金融商品の売買数も含めた最適化[1]に着目したものやなど[2]様々なアプローチがなされているが，すべての期間で十分な利益を上げられているわけではない．これは，売買時における機会損失を考慮した適切な報酬が設定されていないからだと考えられる．そこで，本研究では各行動に対する機会損失を深層強化学習での報酬に組み込み，株式投資において利益を上げるための最適な買いや売りのタイミングを学習するモデルを構築し，その有効性を示す．

深層強化学習

強化学習とはエージェントが試行錯誤を通して目的を達成する方法論である．エージェントは行動を起こし，環境から報酬と次の状態を受け取る．これを繰り返し行い，エージェントは報酬を最大化する最適な行動系列を得るための方策を学習する．深層強化学習[3]の 1 つである Deep Q Network（以下, DQN と略す）は，強化学習のアルゴリズムである Q 学習における行動価値関数 Q に対してニューラルネットワークを関数近似器として用いたものである．本研究の学習の際にはエージェントが株式市場から状態S_tを受け取り，確率εでランダムな行動を，確率$(1-\epsilon)$でニューラルネットワークが出力する行動価値$Q(s,t)$から最も高い行動a_tを選択する．そして環境(株式市場)から報酬r_tと次の状態S_{t+1}を受け取る．この一連の経験(S_t, a_t, r_t, S_{t+1})を Experience Replay Memory に一定数記録しておき，過去の経験をランダムにサンプリングしミニバッチ学習を行う．

提案手法

状態変数

本研究では以下の 4 種類計 6 つの状態変数を使用した．

・所持金（初期保有量:100, 000＄）
・総資産
・株価の増減率（前日比）

・株価移動平均（5 日，25 日，75 日）

急激な上昇や下落に対応するため，株価の前日比を状態変数として設定した．また，短期から中長期における株価の傾向をエージェントに与えるために，株価の 5 日，25 日，75 日移動平均を使用した．移動平均とは代表的なテクニカルチャートのひとつで、価格のトレンドから，相場の方向性を見る手掛かりをつかむために使用される．扱う株価は銘柄の 1 日の終値ベースとした．

行動

エージェントは「買い」，「売り」，「何もしない」の 3 種類の行動から 1 日 1 回 1 つ行動を終値で選択する．「買い」では 100 株をその日の終値で購入する．「売り」はそれまでに保持してきた株をすべて売却する．本研究では株式の現物取引を想定しており，空売りなど信用取引は考慮していない．

報酬

エピソード終了時にまとめて報酬を与えると報酬を受け取るまでの時間が長くなり，学習が進まなくなることを考慮し，先行研究では「売り」行動のみに報酬を与えていた．本研究ではエージェントの選択した「買い」，「売り」の 2 つの行動が，最適な行動であるかを評価するために，t日目の報酬R_tを以下のように定義する．

$$R_t = \begin{cases} S_{all} \times \left(\dfrac{P_{sell} - P_{buy}}{P_{buy}} \right) + S_{all} \times \alpha_t & if\ a_t\ is\ sell \\ B \times \beta_t & if\ a_t\ is\ buy \\ 0 & if\ a_t\ is\ hold \end{cases}$$

ここで，P_{sell}は売却時株価を，P_{buy}は購入時株価を，S_{all}は売却株数を，a_tは t 日目の行動を，Bは購入株式数を表す．α_t，β_tはそれぞれ「売り」と「買い」を選んだことにより発生する機会損失を考慮した適正度を表す項であり，以下のように表される．

$$\alpha_t = \frac{(P_{sell} - max(PL_t)) + (P_{sell} - min(PL_t))}{P_{sell}}$$

$$\beta_t = \frac{(max(PL_t) - P_{buy}) + (min(PL_t) - P_{buy})}{P_{buy}}$$

ここで，PL_tは t 日目における過去 5 日間の終値が格納されているリストで，$max(PL_t)$，$min(PL_t)$はそれぞれPL_tの最大値，最小値を表す．このPL_tに保存された過去の情報を参照することで，「もっと安く買えた」「もっと高く売れた」といった機会損失を表現している．以下の図 1，図 2 に本研究での機会損失のイメージを示す．

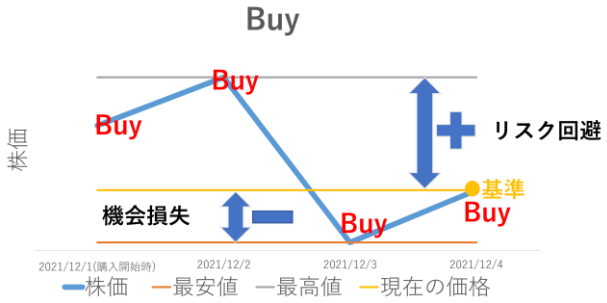

図 1　購入時の機会損失

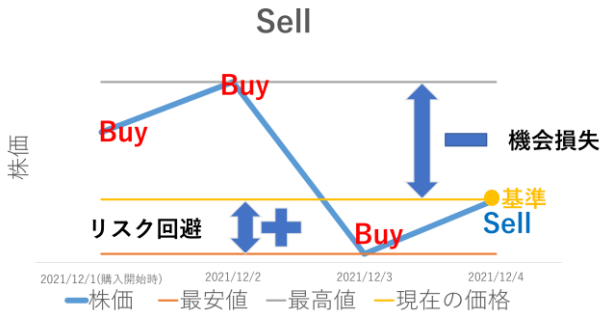

図 2　売却時の機会損失

実験結果

発表時に詳細な結果と考察を述べる．

参考文献

[1] 和田裕貴，長尾智晴．:深層強化学習による株式売買戦略の構築，情報処理学会第 79 回全国大会, Vol.2017, No.1, pp.345-346 (2017)

[2] 近藤巧麻，松井藤五郎．:複雑な環境における複利型深層強化学習を用いた金融取引戦略，人工知能学会全国大会(第 34 回) (2020)

[3] Sutton,R.S.，Barto,A.G.:Reinforcement Learning，MIT

press, (1998)

[４] Jinho Lee, Raehyun Kim, Yookyung Koh, Jaewoo Kang. :Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network, IEEE Access(Volume：7),pp.16726‐167277 (2019)