4M-08

深層強化学習による機会損失を考慮した金融取引戦略の構築

井上　修一†　　穴田　一‡

東京都市大学　総合理工学研究科†　　東京都市大学　情報工学部‡

1. はじめに

近年，機械学習を用いた金融取引の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．これらの研究では，金融商品の売買数や複利計算を考慮したものなど様々なアプローチがなされているが，安定的な利益を上げられていない．これは，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．

2. 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利用したものである．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を行う．そして株式市場から報酬r_tと次の状態S_{t+1}を受け取り，過去の経験(S_t, a_t, r_t, S_{t+1})をreplay Memoryに保存しランダムに取り出しミニバッチ学習を行う．

3. 提案手法

3.1. 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．

3.2. 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．

3.3. 報酬

取引最終日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，その時点での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を以下のように示す．

Creating a financial transaction strategy with consideration to the opportunity loss using deep reinforcement learning.
† Shuichi Inoue † Graduate School of Integrative Science and Engineering, Tokyo City University
‡ Hajime Anada ‡ Faculty of Information Technology, Tokyo City University

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

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．

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

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

ここで，$price_list$は最初の購入から売却を行うまでの日ごとの終値が格納されるリストで，$max(price\_list)$，$min(price_list)$はそれぞれ$price_list$の最大値，最小値を表す．この$price_list$に登録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．

4. 実験結果

提案手法の有効性を確認するため，実際の株式データを利用した実験を行った．対象は NASDAQ 市場の銘柄であるアメリカン航空グループとし，2014 年 1 月～2015 年 12 月の 2 年間を学習期間，2016 年 1 月～2017 年 12 月の 2 年間をテスト期間とした．ニューラルネットワークのパラメータはそれぞれバッチサイズ 20，メモリーサイズ 200，隠れ層 3，隠れニューロン数 100，最適化手法には Adam optimizer，活性化関数には ReLU 関数と softMax 関数を用いた．学習期間とテスト期間の最終資産と利益率を表１に示す．

表１ 学習期間とテスト期間の最終資産と利益率

	最終資産(＄)	利益率(%)
学習期間	**175320**	**75.32**
テスト期間	**135210**	**35.21**

表１より，テスト期間において約 35%の利益を生みだしていることがわかる．

5. 考察

行動時系列を確認すると，「買い」行動について学習期間，テスト期間それぞれで適切に行動することが確認できた．「売り」行動については学習期間においては株価が高いときに売れているが，テスト期間では高くなる直前で売ってしまうことや，連続して売り行動を選択してしまうことがみられた．また 2016 年の４月～5 月にかけての急速な株価の下落の途中に売りを行っていることが分かった．このことから，急速な下落において損切をすることで大幅な資産減少のリスクを回避するような行動をとることが確認できた．

6. 今後の展望

本研究では機会損失を考慮した深層強化学習による金融取引戦略を構築し，実際の株式銘柄に対して売買実験を行い，その有効性を示した．「買い」と「売り」の報酬に機会損失を組み込んだことで，より安く買い，高く売ることをエージェントに学習させることができた．今後は，さらに資産を増やすため，高値の時に「買い」行動を選択するための報酬設計や，「何もしない」行動についても何らかの報酬を与えること，多数の状態変数を扱える深層強化学習の特徴を活かすため，出来高やボリンジャーバンドなど新たな状態変数の導入の検討する必要があると考えている．また，様々な株価時系列に提案手法が対応することを確認するために，複数の銘柄に対しても実験を行い汎用性の高い金融取引戦略の構築を行っていきたい．

参考文献

[1] Sutton,R.S.,
Barto,A.G.:Reinforcement Learning,
MIT press, (1998)

[2] 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)

[3] 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)

[4] 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)