6N-04

深層強化学習による株式売買戦略の構築

和田裕貴 †　　長尾智晴 †
† 横浜国立大学大学院環境情報学府

1 はじめに

従来から，進化計算や機械学習などの各種工学的手法を用いた株価予測や売買戦略を構築する研究が活発に行われている．しかし，株式市場は投資家の心理や政治的要因などの様々な要因が複雑に絡み合っているため，非線形性が強い．そのため，一般的に株価などの予測は困難である．一方，近年ニューラルネットワークを発展させた深層学習と呼ばれる手法が画像認識や音声認識の分野で，認識性能の高さを示して注目されている．また，深層学習を強化学習に応用した深層強化学習の研究も行われており，適用先であるアーケードゲームにおいては人間より高得点を示したという報告もされている [1]．深層強化学習では，入力が高次元であっても教師なしで過去の経験から将来の報酬を最大化する行動規則を学習することができる．そこで，本稿では深層強化学習を株式売買に応用した手法を提案する．本手法では，株価やテクニカル指標の時系列データ，投資家の状態から，将来の報酬の和を最大化するような方策を学習する．さらに本手法では，従来手法のように全所持金で株を購入し，売却時は全ての株を売るといったような単純な売買戦略ではなく，売買株数の最適化も含めて売買戦略の構築を行う．

2 提案手法

本手法では，深層強化学習を用いることによって株価，テクニカル指標などの時系列情報から売買の判断に有効な特徴量を抽出し，利益率の高い売買戦略を獲得する．

2.1 状態空間

本手法では時刻 t に環境（株式市場）から受け取る株価やテクニカル指標などの時系列情報とエージェント（投資家）の総資産，所持株数などの情報を，エージェントが受け取る状態 s_t と定義する．

2.2 行動空間

本手法において，投資家の行動はおおまかに分けて“buy”,“sell”,“hold” の 3 種類がある．“buy” では所持金での株の購入，“sell” では所持している株の売却，“hold” では所持金や保持している株を翌日に持ち越す．さらに，“buy”,“sell” は扱う株数によってそれぞれ $\mathrm{buy}_1, ..., \mathrm{buy}_N$ と $\mathrm{sell}_1, ..., \mathrm{sell}_N$ の N 種類に分割される．buy_N は所持金の $1/N$ の金額で株を購入し，sell_N は所持している株の $1/N$ を売却する．そのため，全行動数は $2N+1$ となり，時刻 t の行動を $a_t \in (\mathrm{buy}_1, ..., \mathrm{buy}_N, \mathrm{sell}_1, ..., \mathrm{sell}_N, \mathrm{hold})$ と定義する．

2.3 報酬関数

本手法の目的は，売買初日と売買最終日の総資産を比較したときの利益率が高くなる売買戦略を構築することである．しかし，売買最終日にのみ報酬を与えてしまうと投資家が報酬を受け取る間隔が長くなってしまい，学習がうまく進まないことが考えられる．そこで，本手法では以下のように株を売却した時点で，その利益率に応じて報酬 r_t を与える．P_{buy}，P_{sell}，S_{sell}，S_{all} はそれぞれ，平均取得単価，売却時の株価，売却株数，売却前の全所持株数を示している．

$$r_t = \begin{cases} \frac{P_{\mathrm{sell}} - P_{\mathrm{buy}}}{P_{\mathrm{buy}}} \times \frac{S_{\mathrm{sell}}}{S_{\mathrm{all}}} & \text{if } a_t \text{ is sell} \\ 0 & \text{if } a_t \text{ is buy} \\ 0 & \text{if } a_t \text{ is hold} \end{cases}$$

2.4 深層強化学習

深層強化学習 [1] は，強化学習における行動価値関数 Q を多層のニューラルネットワークによって近似する手法（Deep Q-Network;DQN）である．m 次元の状態変数ベクトル $X = (x_1, x_2, ..., x_m)$ で表される状態 s がネットワークに入力されると，最終層のノードから各行動の行動価値 $Q(s, a)$ が出力される．

学習の際は，投資家が株式市場から状態 s_t を受け取り，最も行動価値が高い行動 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} を受け取るという過去の経験 (s_t, a_t, r_t, s_{t+1}) を replay Memory に保存しておき，そこからランダムに選択してミニバッチ学習を行う．提案手法の学習アルゴリズムを Algorithm 1 に示す．

Construction of stock trading strategy using Deep Reinforcement Learning
†Yuki Wada　†Tomoharu Nagao
†Graduate School of Environment and Information Sciences, Yokohama National University

Algorithm 1 提案手法の学習アルゴリズム

Input:銘柄リスト M，初期総資産，初期所持金，初期所持株数
サイズ N の replay memory D を初期化
ランダムな重みで行動価値関数 Q を初期化
for $epoch$ = 1 to E **do**
  銘柄リストをランダムに並び替え
  for $symbol$ = 1 to M **do**
    総資産，所持金，所持株数を初期化
    for t = 1 to T **do**
      確率 ϵ でランダムに行動 a_t を選択
      そうでなければ $a_t = max_a Q(s_t, a; \theta)$ を選択
      行動 a_t を実行し，報酬 r_t と状態 s_{s+1} を受け取る
      総資産，所持金，所持株数を計算
      遷移 (s_t, a_t, r_t, s_{t+1}) を D に保存
      D からランダムに遷移 (s_j, a_j, r_j, s_{j+1}) のミニバッチを取得
      $$\text{Set } y_j = \begin{cases} r_j & \text{if } s_{j+1}\text{が終端} \\ r_j + max_{a'} Q(s_{j+1}, a_j; \theta) & \text{if otherwise} \end{cases}$$
      $(y_j - Q(s_j, a_j; \theta))^2$ に勾配降下法を実行
    end for
  end for
end for

3 売買シミュレーション実験

3.1 実験設定

本稿では東証一部 2000 銘柄を使用して提案手法の学習を行った．また，状態変数の時系列情報として売買中の銘柄の株価，出来高，テクニカル指標をそれぞれ過去 20 日分使用する．売買は 1 日 1 回，終値ベースで行う．比較手法にテクニカル指標を用いて売買する手法（Buy&Hold;B&H, ゴールデンクロス&デッドクロス;GD&DC,RSI）とランダムトレード (10000 回平均) を行う手法を用いた．訓練期間は 2000 年〜2008 年，テスト期間は 2009 年〜2010 年とした．また，同時に売買する銘柄は常に 1 つまでとする．行動分割数については N = 1〜4 の 4 パターンで実験を行った．評価は 100 銘柄における訓練期間・テスト期間の平均利益率で比較した．

3.2 実験結果

表 1 に訓練期間，テスト期間の 100 銘柄の平均利益率を示す．提案手法は訓練期間，テスト期間において他の比較手法と比べて大きく利益を伸ばすことができた．表 2 にテスト期間における提案手法のリスクを示した．ここでのリスクは利益率の標準偏差を表している．提案手法では N を大きくするほど利益率のばらつきを小さくしてリスクを抑えることができ，安定して高い利益をだすことができた．以上の結果から，深層強化学習によって利益率の高い売買戦略を構築することができたと言える．

表 1: 実験結果：平均利益率（%）

	B&H	GD&DC	RSI	random
訓練期間	38.6	62.7	29.6	34.3
テスト期間	14.5	17.3	11.2	8.4
	Proposed $N=1$	Proposed $N=2$	Proposed $N=3$	Proposed $N=4$
訓練期間	86.84	80.19	81.19	**93.07**
テスト期間	24.42	**30.67**	29.65	29.49

表 2: 提案手法のリスク評価

	Proposed $N=1$	Proposed $N=2$	Proposed $N=3$	Proposed $N=4$
リスク	42.33	36.48	30.69	**25.46**

4 まとめ

本稿では，深層強化学習を用いた株式売買戦略の構築手法を提案した．売買シミュレーション実験では，提案手法は他の手法と比較し，訓練期間，テスト期間ともに高い利益率を出すことができた．今後は，actor-critic[3] のような連続値行動が可能な強化学習を用いた，より細かな売買株数の最適化などを行い，提案手法の改良を行っていきたい．

参考文献

[1] Mnih, Volodymyr, et al. “Human-level control through deep reinforcement learning.” Nature 518.7540 (2015): 529-533.

[2] Ioffe, Sergey, and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” arXiv preprint arXiv:1502.03167 (2015).

[3] Mnih, Volodymyr, et al. “Asynchronous methods for deep reinforcement learning.” arXiv preprint arXiv:1602.01783 (2016).