4I1-OS-16-4

複利型強化学習の株式取引への応用

Stock Trading Using Compound Reinforcement Learning

後藤 卓[*1] 松井 藤五郎[*2] 大澄 祥広[*2]
Takashi Goto Tohgoroh Matsui Yoshihiro Osumi

[*1]三菱東京 UFJ 銀行 [*2]中部大学
Bank of Tokyo-Mitsubishi UFJ, Ltd. Chubu University

This paper describes stock trading using compound reinformcent learning with optimizing bet fraction and function approximation. We applied compound reinforcement learning to acquire trading rules for TOPIX ETF in Kaburobo and analysed the acquired trading rule in terms of a financial trader.

1. はじめに

金融業界では，株式取引に代表される金融商品の取引は担当者の経験に基づいて行われており，どのようなときにどのように売買すればいいかについては，「もうはまだなり，まだはもうなり」に代表される相場格言や，評価額が 10%下落したら決済して損切りするというような簡単なルールで表現されてきた．このような取引ルールを過去のデータから自動的に獲得するために，試行錯誤に基づく機械学習アルゴリズムである強化学習を用いて金融商品の取引規則を学習する研究が行われてきた [松井 07, 松井 09].

一方で, エージェントが獲得するリターンに基づく複利リターンを将来にわたって最大化する行動規則を試行錯誤を通じて学習する枠組みである複利型強化学習が提案されている [Matsui 12, 松井 11a, 松井 11b]．これまでに，複利型強化学習は国債銘柄選択問題や国債取引問題に適用され，その有用性が示されている．

また，複利型強化学習では，エージェントが自分の資産のうちのどれだけを投資するかを表す投資比率パラメーター f が導入されており，この投資比率をオンライン勾配法を用いて最適化する手法が提案されている [松井 13]．ただし，従来の方法は離散状態を対象としたものであるため，連続状態に適用するためにはこれを拡張する必要がある．

そこで，本論文では，複利型強化学習をカブロボの自動取引エージェントに適用し，株式の取引ルールを学習する．また，学習した取引ルールを金融の実務家の観点から分析し，提案手法の有効性を考察する．本論文では，カブロボ [カブ 07] を用いることで，手数料や取引ルールを考慮した実際的な取引ルールを学習することを可能としている．

2. 複利型強化学習

複利型強化学習は，割引複利リターン

$$\begin{aligned}&(1+R_{t+1}f)(1+R_{t+2}f)^{\gamma}(1+R_{t+3}f)^{\gamma^2}\dots\\&=\prod_{k=0}^{\infty}(1+R_{t+k+1}f)^{\gamma^k}\end{aligned}\tag{1}$$

の期待値を最大化するような行動規則を学習する．ここで，R_t は時刻 t に観測されたリターン，γ は割引率パラメーター，f は投資比率パラメーターを表す．割引複利リターンは，対数を取ることで従来の強化学習と同じ形で表すことができるため，複利型強化学習では，行動価値を割引複利リターンの対数の期待値と定義する．すなわち，行動規則 π の下での状態 s における行動 a の価値 $Q^{\pi}(s,a)$ は次のように表される．

$$Q^{\pi}(s,a)=\mathrm{E}_{\pi}\left[\log\prod_{k=0}^{\infty}(1+R_{t+k+1}f)^{\gamma^k}\middle|\, s_t=s, a_t=a\right]\tag{2}$$

複利型強化学習では，すべての s, a に対してこの $Q^{\pi}(s,a)$ を最大化するような行動規則 π を学習する．

本論文では，強化学習のアルゴリズムとして，複利型 OnPS [松井 11b] を用いる．複利型 OnPS は，オンライン型の Profit Sharing である OnPS [Matsui 03] を複利型に拡張したアルゴリズムであり，最適解への収束は保証されないが，Q 学習のような最適化アルゴリズムに比べて過学習が起こりにくいという特徴を持つ．

従来のオンライン勾配法による投資比率最適化 [松井 13] は離散状態を対象としているが，株式取引では状態空間が連続であるため，本論文では，動径基底関数（RBF）を用いた線形関数近似を導入する．

3. オンライン勾配法による投資比率の最適化の線形関数近似

本論文では，オンライン勾配法による投資比率の最適化に線形関数近似を導入する．

リターン R_{t+1} を受け取ったとき，時刻 $t+1$ までの複利リターンは次のように計算される．

$$G_{t+1}=\prod_{k=1}^{t+1}(1+R_k f)\tag{3}$$

本論文では，投資比率 f をパラメーター・ベクトル $\vec{\psi}$ と特徴ベクトル $\vec{\phi}$ の内積として表す．

$$f_t=\vec{\psi}_t^{\mathrm{T}}\vec{\phi}_{s,a}=\sum_{i=1}^{n}\psi_t(i)\phi_{s,a}(i)\tag{4}$$

式 (4) を式 (3) へ代入し，両辺の対数を取ると次のようになる．

$$\log G_{t+1}=\sum_{k=1}^{t+1}\log(1+R_k\vec{\psi}_t^{\mathrm{T}}\vec{\phi}_{s,a})\tag{5}$$

連絡先: 後藤卓, takashi_6_gotou@mufg.jp

この両辺を $\vec{\psi}_t$ で偏微分する．

$$\frac{\partial}{\partial \vec{\psi}_t} \log G_{t+1} = \sum_{k=1}^{t+1} \frac{\partial}{\partial \vec{\psi}_t} \log\left(1 + R_k \vec{\psi}_t^{\mathrm{T}} \vec{\phi}_{s,a}\right) \tag{6}$$

$$= \sum_{k=1}^{t+1} \frac{R_k \vec{\phi}_{s,a}}{1 + R_k \vec{\psi}_t^{\mathrm{T}} \vec{\phi}_{s,a}} \tag{7}$$

したがって，オンライン勾配法による $\vec{\psi}$ の更新式は次のようになる．

$$\vec{\psi}_{t+1} = \vec{\psi}_t + \eta \frac{R_{t+1} \vec{\phi}_{s,a}}{1 + R_{t+1} \vec{\psi}_t^{\mathrm{T}} \vec{\phi}_{s,a}} \tag{8}$$

ここで，η はオンライン勾配法の学習率パラメーターを表す．

4. 複利型強化学習を用いた取引ルールの学習

複利型強化学習における状態は，相対終値と相対移動標準偏差で表現する．本論文では，移動標準偏差の算出期間は 50 営業日とし，相対化 [Matsui 12] は以下のように行った．

$$o_t = \frac{v_t - \mu_{t,50}}{4\sigma_{t,50}} \tag{9}$$

ここで，v_t は t における値，$\mu_{t,50}$ は t の直近 50 個のデータから求めた平均値，$\sigma_{t,50}$ は時刻 t の直近 50 個のデータから求めた移動標準偏差を表す．すなわち，$[\mu_{t,50} - 4\sigma_{t,50}, \mu_{t,50} + 4\sigma_{t,50}]$ の範囲を $[-1, 1]$ に正規化している．

カブロボ（エージェント）の行動は買いと売りの 2 種類である．株式を購入している状態をロング・ポジション，株式を信用売りしている状態をショート・ポジションという．複利型強化学習では，投資比率 f が導入されているため，投資比率によってポジションの大きさを調整する．

複利型強化学習では，報酬の代わりに利益率（リターン）を用いて学習する．利益率は，以下の式によって計算する．

$$R_{t+1} = \begin{cases} \frac{asset_t - money_{t-1}}{stock_{t-1}} - 1 & (money_t \geq money_{t-1} \text{ のとき}) \\ \frac{stock_t}{asset_{t-1} - money_t} - 1 & (\text{そうでないとき}) \end{cases} \tag{10}$$

ここで，$stock_t$ は t における株式評価額，$money_t$ は t における取引余力（現金），$asset_t$ は t における資産評価額（株式評価額 $stock_t$ と取引余力 $money_t$ の合計）を表す，この式は，現在と前営業日との取引余力（現金）を比較し，少ない方を投資しなかった資産として除いた投資金額の増加率を求めるものである．

5. 実験

5.1 実験方法

2002 年から 2011 年までの 10 年間の TOPIX ETF（証券コード 1306）のデータを 10 回繰り返し用いて取引戦略を学習した．また，2012 年のデータを用いて学習した取引戦略を評価した．2002 年から 2012 年までの TOPIX ETF の終値の推移を図 1 に示す．

関数近似を行うため，15×15 の動径基底関数を格子状に配置した．割引率は $\gamma = 0.9$，オンライン勾配法の学習率は $\eta = 0.1$ とした．行動選択には，温度 $\tau = 0.2$ のボルツマン選択を用いた．強化学習の内部パラメーター θ の初期値は 0，投資比率の内部パラメーター ψ の初期値は 0.1 とした．

以上の実験を，乱数のシードを変えて 10 回行い，その平均を求めた．

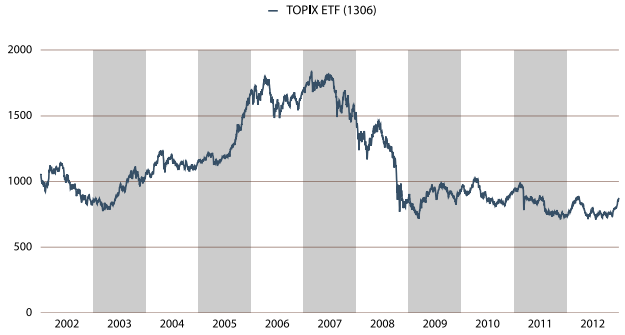

図 1: TOPIX ETF（証券コード 1306）の終値の推移

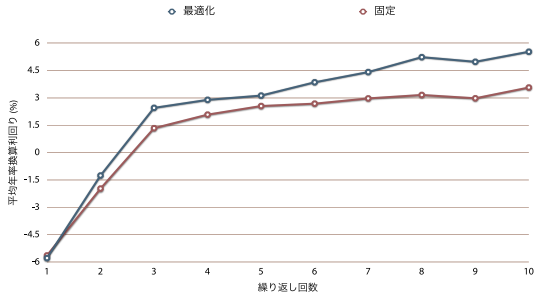

図 2: 訓練期間における平均年率換算利回りの推移

5.2 実験結果

訓練期間における平均年率換算利回りの推移を図 2 に示す．学習が進むにつれて年率換算利回りが向上した．10 回目の学習における平均年率換算利回りは，投資比率を 0.5 に固定した場合が約 3.6%だったのに対して，投資比率の最適化を行う提案手法が約 5.5%であった．

乱数のシードを変えて行った 10 回のうち，訓練期間における 10 回目の学習における年率換算利回りが最も高かった取引ルールを図 3 に示す．RCP は終値の相対値，RMSD は移動標準偏差の相対値を表す．Z 軸はロング・ポジションの選択確率を表しており，選択確率が高いところではロング・ポジションを取りやすく，低いところではショート・ポジションを取りやすい．

また，図 4 と図 5 に，それぞれ，ショート・ポジションの投資比率とロング・ポジションの投資比率を示す．Z 軸は投資比率を表しており，選択される行動が同じところでも，投資比率が異なることがわかる．

学習した取引ルールを用いて 2012 年の 1 年間の取引を行ったところ，提案手法の平均年率換算利回りは 2.9%であった．

6. 考察とまとめ

これまでの複利型強化学習を用いた研究では，行動は「買い」または「売り」の選択のみでポジション量の概念が無かったため，常に一定のポジション量を保有する状況となっていた．

しかし，実際には，「買い」を選択すべき状態と「売り」を選択すべき状態の境界付近では「買い」と「売り」の成功確率はほぼ同じであると想定される．ところが，従来手法のように一定のポジション量を保有した場合，失敗した際の損失が大きくなってしまう．金融業界に従事する実際のディーラーは，このように判断に悩む局面では，少しずつポジションをテイク

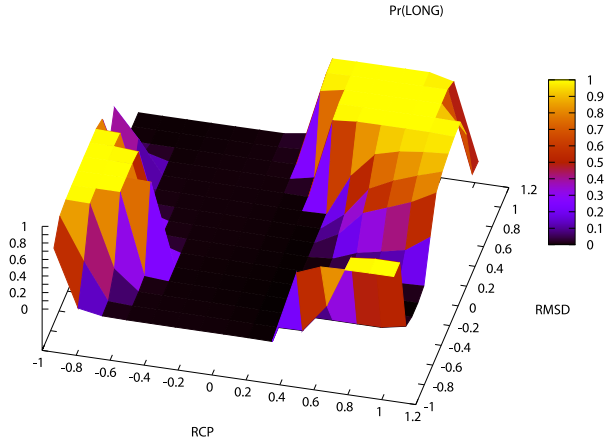

図 3: 学習した取引ルール（行動選択確率）

し，思った方向に動いた場合にポジション量を拡大していくという行動をとる．

提案手法では，同じ行動であっても，投資比率によってポジションを調整することができる．従来手法と提案手法によるポジションの違いを図 6 に示す．投資比率最適化と複利型強化学習に組み合わせた新手法では，前述したディーラーの行動を再現できていると考えられる．

本論文では，TOPIX ETF を対象とした株式取引に，投資比率最適化を行う複利型強化学習を適用し，取引ルールを学習した．投資比率を最適化することによって，同じ行動でもポジションを調整することができる柔軟な取引ルールを学習することができた．

留意事項

本論文は三菱東京 UFJ 銀行の公式見解を表すものではありません．

参考文献

[Matsui 03] Matsui, T., Inuzuka, N., and Seki, H.: On-Line Profit Sharing Works Efficiently, *KES 2003*, 317–324 (2003)

[Matsui 12] Matsui, T., Goto, T., Izumi, K., and Chen, Y.: Compound Reinforcement Learning: Theory and An Application to Finance, *EWRL 2011*, 321–332 (2012)

[松井 07] 松井 藤五郎：カブロボへの招待–人工知能を用いた株式取引–, 人工知能学会誌, 22(4):540–547 (2007)

[松井 09] 松井 藤五郎, 後藤 卓：強化学習を用いた金融市場取引戦略の獲得と分析, 人工知能学会誌, 24(3):400–407 (2009)

[松井 11a] 松井 藤五郎：複利型強化学習, 人工知能学会論文誌, 26(2):330–334 (2011)

[松井 11b] 松井 藤五郎, 後藤 卓, 和泉 潔, 陳 ユ：複利型強化学習の枠組みと応用, 情報処理学会論文誌, 52:(12):3300–3308 (2011)

[松井 13] 松井 藤五郎, 後藤 卓, 和泉 潔, 陳 ユ：複利型強化学習における投資比率の最適化, 人工知能学会論文誌, 28(3):267–272 (2013)

[カブ 07] カブロボ公式サイト (2007), http://kaburobo.jp/

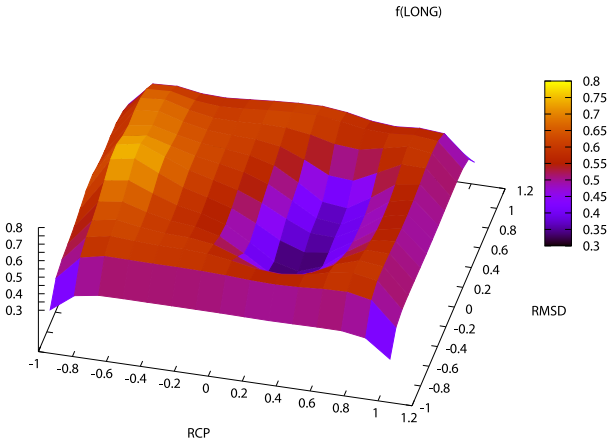

図 4: 学習したロング・ポジションの投資比率

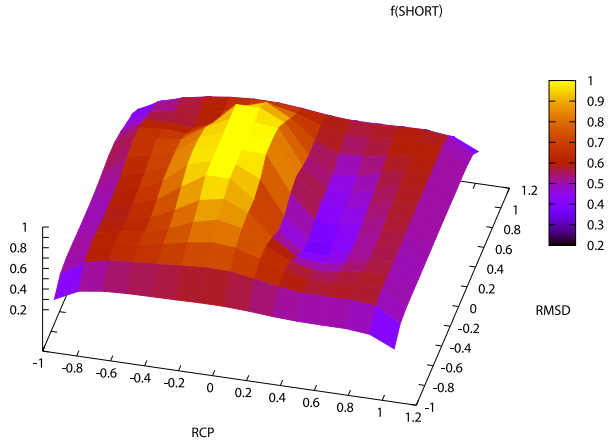

図 5: 学習したショート・ポジションの投資比率

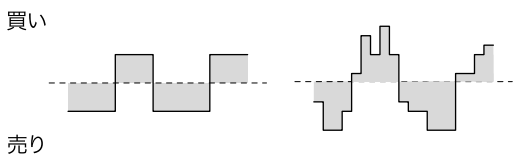

図 6: 従来手法と提案手法によるポジションの違い