

レバレッジを用いた複利型強化学習

塚本 智大*

* 中部大学 工学部 情報工学科

松井 藤五郎†*

† 中部大学 生命健康科学部 臨床工学科

1 はじめに

複利型強化学習 [1] は、報酬の代わりにリターン (利益率) を獲得するリターン型マルコフ決定過程において、試行錯誤を通じてエージェントが将来に獲得する複利リターンを最大化する行動規則を学習する枠組みである。複利式のリターンを考える場合には、従来の強化学習のような期待割引収益の最大化は意味をなさないため、複利型強化学習では期待割引収益の代わりに二重指数関数を用いて割り引いた複利リターンの対数の期待値を最大化する。

「(無分配型の) 投資信託を選択する際にリターンの算術平均ではなくリターンの幾何平均が高い商品を選ぶべきである」というのは、ファイナンスの分野では一般的な考え方であるしたがって、このような場合には、報酬の代わりに複利式のリターンに基づいて学習すべきである。

FX と呼ばれる外国為替証拠金取引では、あらかじめ取引会社に証拠金を預け入れることで、証拠金の数倍から数百倍の金額で取引でき。これをレバレッジという。レバレッジとは、証拠金にける「てこ」のようなもので、少ない資本金でも大きな取引ができる仕組みである。証拠金に対する取引可能金額の比率を、レバレッジの倍率という。

しかしながら、従来の複利型強化学習はレバレッジについて考慮していない。そこで本論文では、複利型強化学習にレバレッジを導入し、レバレッジをかけた場合の複利リターンを最大化する方法について検討する。また、レバレッジをかけると利益も大きくなるが損失も大きくなるため、大きな損失を防ぐための手法について検討する。

2 複利型強化学習とレバレッジの関係

複利型強化学習では、自分が保有する資産うちどれだけ投資するかを表す投資比率パラメーターが導入されている。

複利型強化学習では、エージェントは時刻 t において状態 s_t を観測すると、価値関数 Q から導かれる行動規則 π に基づいて行動 a_t を選択して実行し、その結果としてリターン R_{t+1} を得る。複利型強化学習は、将来にわたって得られる複利リターンを二重指数関数で割り引いたものの対数の期待値を最大化する行動規則を学習ものである。すなわち、複利型強化学習は以下の値を最大化する行動規則を学習する。

$$E \left[\prod_{k=0}^{\infty} (1 + R_{t+k+1} f_{t+k})^{\gamma^k} \right] \quad (1)$$

ここで、 f_t は時刻 t における投資比率を表す。

従来の複利型強化学習は、投資比率が $0 \leq f_t < 1$ であり、レバレッジは考慮されていない。時刻 t における自己資本が p_t 、レバレッジの最大倍率が l 、投資比率が f_t のとき、最大で自己資本の l 倍の $p_t l$ まで投資することができ、投資比率を考慮すると実際の投資額は $p_t l f_t$ となる。

実際の投資額が自己資本を超える、すなわち $p_t l f_t - p_t = p_t (l f_t - 1) > 0$ のとき、自己資本を超える投資金は外国為替証拠金取引業者などから借りているものである。 $l f_t \leq 1$ のときは、投資金を借りる必要がないため、今後は $l f_t > 1$ のときについて考える。

時刻 t に実行した行動の結果として投資金に対するリターン R_{t+1} を獲得したとすると、投資金が $1 + R_{t+1}$ 倍になり、投資金は $p_t l f_t (1 + R_{t+1})$ になる。ここから、借りていた $p_t (l f_t - 1)$ を返済すると、手元には $p_t (1 + R_{t+1} l f_t)$ が残る。自己資本は p_t から $p_t (1 + R_{t+1} l f_t)$ に変化しており、自己資本に対する利益率は $R_{t+1} l f_t$ である。したがって、従来の複利型強化学習における投資比率 f_t を投資比率を考慮したレバレッジ $l_t = l f_t$ に置き換えることで、複利型強化学習にレバレッジを導入することができる。レバレッジが $l_t < 1$ のとき、レバレッジ l_t は従来の複利型強化学習における投資比率 f_t に等しい。

従来の複利型強化学習では、オンライン勾配法を用いて投資比率 f_t を最適化する。レバレッジを導入した複利型強化学習においても、投資比率の最適化と同様にして、以下のようにしてレバレッジ l_t を最適化できる。

$$l_{t+1} = l_t + \eta_t \frac{R_{t+1}}{1 + R_{t+1} l_t} \quad (2)$$

ここで、 η_t はレバレッジの学習率を表す。

Compound Reinforcement Learning with Leverage

*Tomohiro Tsukamoto †Tohgoroh Matsui

*Department of Computer Science, College of Engineering, Chubu University

†Department of Clinical Engineering, College of Life and Health Sciences, Chubu University

3 環境が変化した時への対応

強化学習では、未知の環境における行動や状態において実際に行動して得られた報酬を元に次の行動を推測し決定する方法で学習する。しかし、実際の学習をする場面では報酬の期待値や状態は時間で変化することが多い。また、レバレッジをかけたまま新たな最適解を探す場合、学習を終えるまでの時間に今までの学習を元に投資をしてしまうので最悪の場合自己資本がなくなり破産してしまう可能性がある。したがって、環境が変化した際にはレバレッジを一度下げる必要がある。

Noda は、通常の強化学習において環境が変化することを想定し、行動価値の学習率 α を適応させる RASP という手法を提案している [2]。RASP は、観測値の系列 $\{x_t\}$ に対して、以下のように定義される再帰的指数移動平均 (REMA) を用いて、 α を最適化する手法である。

$$\xi_t^{(0)} = x_t \quad (3)$$

$$\xi_{t+1}^{(1)} = \tilde{x}_{t+1} = (1 - \alpha)\tilde{x}_t + \alpha x_t \quad (4)$$

$$\xi_{t+1}^{(k)} = (1 - \alpha)\xi_t^{(k)} + \alpha\xi_t^{(k-1)} \quad (5)$$

RASP では、行動の学習率 α を最適化し、学習が進むと徐々に α が小さくなり、環境が変化して学習が必要になると α が大きくなる。そこで、本論文では、レバレッジを導入した複利型強化学習におけるレバレッジ l_t の最適化において、その学習率 η を RASP を用いて最適化することを提案する。これによって、環境が変化してリターンが変わると、特に、損失が発生してレバレッジをかけていると破産する恐れが生じた時に、素早くレバレッジ l_t を変化させて環境の変化に対応できると考えられる。

4 予備実験と考察

レバレッジと投資比率の関係を確認するために、以下のような予備実験を行った。図 1 に、予備実験に用いた 2 本腕バンディット問題を示す。左を円盤 A、右を円盤 B とする。円盤に記されている値は、グロスのリターンを表しており、リターンに 1 を加えた値である。例えば、B に 100 ドルを投資して 0.9 が出た時、90 ドルが払い戻され、このときのリターンは -0.1 である。

この 2 本腕バンディット問題に対して、レバレッジを導入し、幾何平均リターンを調べた。図 2 にその結果を示す。

横軸はレバレッジを、縦軸は幾何平均リターンを表している。ケリー基準では、幾何平均リターンを最大化する投資比率を解析的に求めるが、レバレッジを導入した場合でも、同様に幾何平均リターンを最大化す

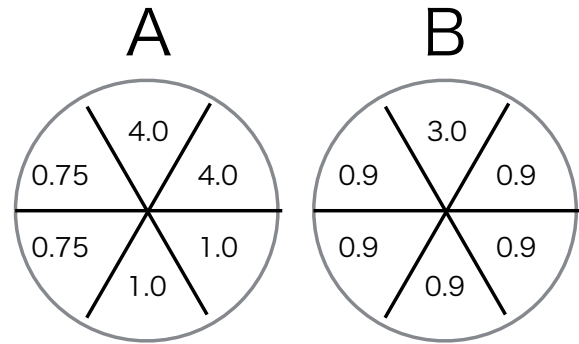


図 1: 2 本腕のバンディット問題

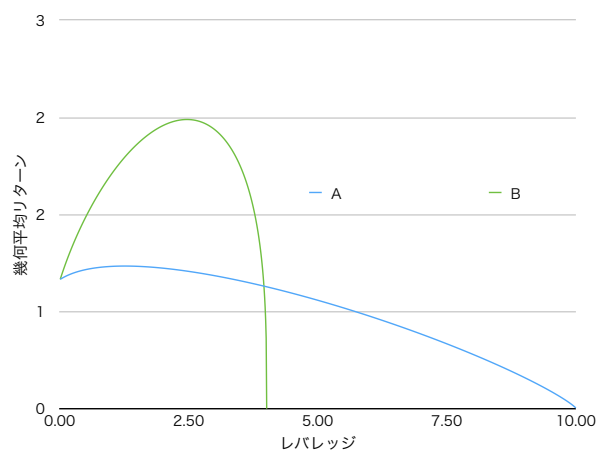


図 2: 投レバレッジと幾何平均リターンの関係

るレバレッジが存在することが確認できる。したがって、複利型強化学習にレバレッジを導入し、レバレッジを投資比率と同じようにオンライン勾配法で最適化すると、最適なレバレッジを学習できると考えられる。

しかしながら、実際の金融市場においては、リターンの分布は状況に応じて変化しており、最適なレバレッジもそれに依拠して変化する。RASP を用いてレバレッジの学習率 η の最適化することによって、このような状況にも対応できるようになることが期待できる。今後、実験によりその有効性を確認したい。

参考文献

- [1] 松井 藤五郎. 複利型強化学習—強化学習のファイナンスへの応用—. 計測と制御, 52, (11): pp. 1022–1027, 2013.
- [2] Itsuki Noda. “Adaption of Stepsize Parameter Using Newton’s Method.” *AGENTS IN PRINCIPLE, AGENTS IN PRACTICE*, 7047:349–360, 2011.