

複利型強化学習における投資比率学習法の改善

羽根田 拓哉*

* 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

ファイナンスの分野では一般的な強化学習のような利益の最大化を考えるのではなく、報酬の代わりとして複利式のリターン（利益率）を最大化する必要がある。なので、ファイナンスの分野においてリターンの算術平均よりも幾何平均が重視される。複利式のリターンを学ぶためには幾何平均リターンが重要となり、利益率及び投資比率について考える必要がある。この、報酬の代わりとして複利式のリターンを最大化する強化学習が複利型強化学習 [1] である。

一般的な強化学習では報酬 r を受け取り、指数関数による割引収益の期待値を最大化する行動規則を学習する。これに対し、複利型強化学習では報酬の替わりとしてリターン R を受け取り、二重指数関数による割引複利リターンの対数期待値を最大化するような行動規則を学習する。時刻 t における複利リターンは投資比率 f を考慮すると

$$G_t = \prod_{k=1}^t (1 + R_k f) \quad (1)$$

となり、割引複利リターンは

$$(1 + R_{t+1}f)(1 + R_{t+2}f)^\gamma(1 + R_{t+3}f)^{\gamma^2} \dots \\ = \prod_{k=0}^{\infty} (1 + R_{t+k+1}f)^{\gamma^k} \quad (2)$$

となる。ここで、 γ は割引率を示す。複利型強化学習では、すべての状態 s と行動 a に対して価値 $Q^\pi(s,a)$ を最大化するような行動規則 π を学習する。

複利型強化学習では、学習の発散を防ぐために投資比率が使われている。投資比率は複利リターンに大きく影響し、また投資比率によって複利リターンを最大化する最適な行動も変化するので重要となる。複利型強化学習では、複利リターンを最大化する規則を学習する際にオンライン勾配法を用いて投資比率を同時に学習する。オンライン勾配法では獲得したリターン R_{t+1}

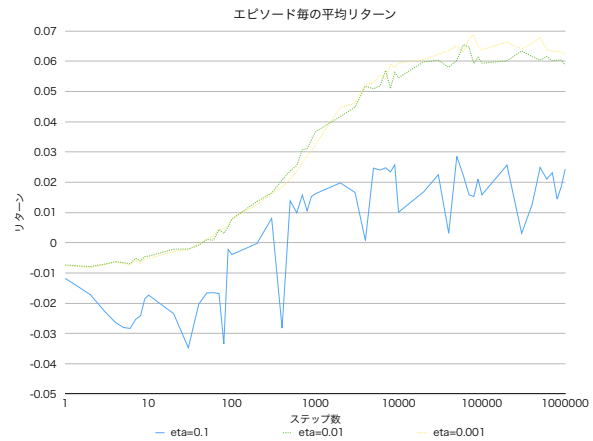


図 1: 投資比率最適化付き Q 学習の平均複利リターン

を用いて、 f を次のように更新する。

$$f_{t+1}(s_t, a_t) = f_t(s_t, a_t) + \eta_t \frac{R_{t+1}}{1 + R_{t+1}f_t(s_t, a_t)} \quad (3)$$

ここで、 η_t は投資率の学習率である。

2 投資比率最適化と学習のリセット

複利型強化学習における投資比率は、 $0 < f < 1$ でなければならない。このため、従来手法では投資比率を更新して $f \leq 0$ または $f \geq 1$ になったときは、学習をリセットし、最初から学習をやり直す。しかしながら、学習のリセットを行うと、学習が遅くなってしまいうという問題と繰り返しリセットが生じて学習できないことがあるという問題が生じる。特に、後者の問題は投資比率の学習率 η が大きいときに顕著である。

これを確認するために、カジノにおけるカードゲームの一つであるブラックジャックを用いて、実験を行った。その結果を図 1 と図 2 に示す。この結果は、100 万ステップの学習を 30 回行い、その平均を比較したものである。図中の η は f の更新の際に用いられる学習率 η_0 であり、投資比率を更新する際の更新量を決めるものとなる。図 1 の結果から η が 0.1 の時にはうまく学習をしていないが、それ以外の場合には 6-7%ほど利益を上げることが出来ているのが分かる。これは、現在の方式では投資比率を更新した際に更新後の投資比率の値が 0 以下になるとき、投資比率と Q の学習の結果をリセットし初めからやり直すからである。 η の

Improving the bet-fraction learning method for Compound RL

*Haneda Takuya†*Tohgoroh Matsui

*Information Engineering, College of Engineering department, Chubu University

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

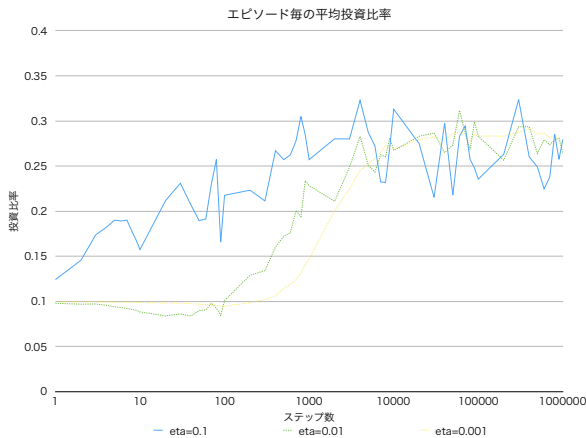


図 2: 投資比率最適化付き Q 学習の平均投資比率

値を大きくすると更新量が大きくなるのでリセットの数が増え上手く学習が出来なくなってしまう。

3 リセットの回避

本論文では、投資比率が $f \leq 1$ または $f \geq 1$ となって学習がリセットされるのを回避するため、投資比率の学習率 η を次のように変動させることを提案する。

$$\eta_t = \left(\frac{\eta_0}{\sqrt{t}} \right)^k \quad (4)$$

ここで、 η_0 は初期学習率、 k は $0 < f_{t+1} < 1$ となる最小の正の整数である。これによって、投資比率は必ず $0 < f < 1$ となるため、学習をリセットする必要がなくなる。

学習のリセットを回避する単純な方法として $f \leq 1$ または $f \geq 1$ となる場合には投資比率 f を更新しないという手段も考えられるが、 f が 1 に近く、かつ、 R が -1 に近いときに f が更新できなくなってしまう。これに対し、提案手法では必ず f が更新されるため、このような問題は生じない。

4 実験結果と考察

提案手法の有効性を確認するために、初期学習率を $\eta_0 = 0.1$ として、提案手法を用いて上と同じ実験を行った。その結果を図 3 と図 4 に示す。提案手法を導入する前のリセットの回数は平均 64,521 回であったが、提案手法を導入することによって、リセットを生じないようにすることができた。また、提案手法を導入することによって、 η が大きい時でも正しく学習を行えるようになった。

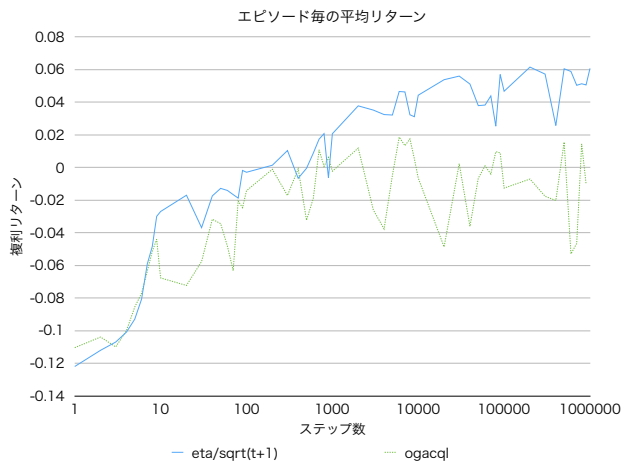


図 3: リセット回避型投資比率最適化付き Q 学習の平均複利リターン

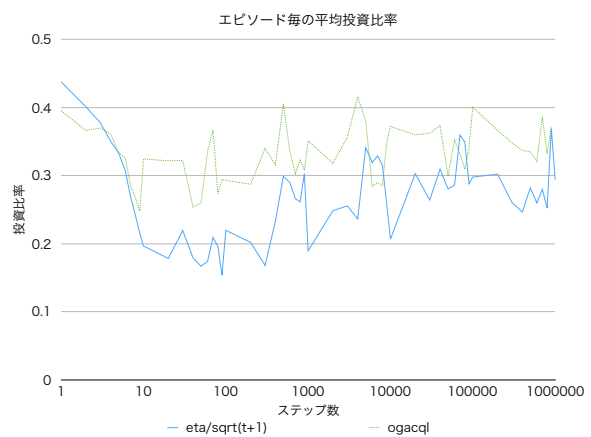


図 4: リセット回避型投資比率最適化付き Q 学習の平均投資比率

5 まとめ

本論文では、複利型強化学習において重要となる投資比率の学習法に対して、学習のリセットを回避する方法を提案した。既存の学習方法ではリセットが多発して学習に時間がかかる場合や複利リターンをプラスまでもっていけない場合があったが、提案手法を用いることでリセットが回避できるため、学習速度の向上や複利リターンの増加が期待できる。

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, Vol. 52, No. 11, pp. 1022–1027, 2013.