情報処理学会論文誌 Vol.60 No.10 1631–1640 (Oct. 2019)

複利型強化学習を用いたポートフォリオ選択手法についての研究

畠山 卓[1] 澤 亮治[1,a)]

受付日 2019年1月18日, 採録日 2019年7月3日

概要：本論文では，従来の投資比率を最適化する手法からポートフォリオの最適な重みを学習させる手法へと複利型強化学習の拡張を行う．複利型強化学習とは通常の強化学習を，二重指数割引を導入することによって割引複利リターンを最大化するような形に拡張したものである．いくつかの資産の中から 1 つの投資先を決定するような研究はなされてきたが，資産をポートフォリオとして所有していた場合に，それぞれの資産への最適な投資比率を複利型強化学習を用いて学習する手法については明らかにされていない．本研究では各資産への最適な重みを学習させることを目的とする．複利リターンを最大化する投資比率のことをケリー基準と呼ぶが，このケリー基準と一致するような重みの実現を目指す．多腕バンディット問題に本提案手法を適用し，幾何平均リターンで運用成績を評価することによって，有用性の確認を行う．ポートフォリオとして資産運用した方が，複数の投資先の中から 1 つを選ぶ場合よりも高い幾何平均リターンが得られることの確認，および学習した各資産への重みとケリー基準値の比較を行い，最適な重みの学習が可能であることを検証する．

キーワード：複利型強化学習，投資比率，複利リターン，ポートフォリオ

A Study on Portfolio Selection Using Compound Reinforcement Learning

Taku Hatakeyama[1] Ryoji Sawa[1,a)]

Received: January 18, 2019, Accepted: July 3, 2019

Abstract: In this paper, we have expanded the compound reinforcement learning method to optimize portfolio weights from optimizing the investment ratio. Compound reinforcement learning is an extension of ordinary reinforcement learning to maximize discounted compound return by introducing double exponential discount. There are studies that determine one investment destination from multiple assets, but learning methods using compound reinforcement learning with optimal weights for each asset when owning assets as portfolios have not been clarified. Therefore, we aim to learn the optimum weight for each asset of the portfolio. The investment ratio that maximizes compound return is called Kelly criterion. We set out to obtain weight that matches Kelly criteria. For that purpose, we applied this proposed method to multi-armed Bandit problems. After that, we confirmed its application by evaluating investment performance with geometric mean return. As a result, we found that asset management as a portfolio increases geometric mean return than choosing one among multiple investment outlets. Furthermore, since the weight for each asset learned was close to the Kelly criterion, we were confirmed that the optimum weight can be learned by this proposed method.

Keywords: compound reinforcement learning, betting fraction, compound return, portfolio

1. はじめに

学習対象者であるエージェントが現在の状態を観測し，試行錯誤を通じて将来にわたり最も多くの報酬を得られる

[1] 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305–8573, Japan
[a)] rsawa@sk.tsukuba.ac.jp

ような方策を自律的に学習する，機械学習の手法の 1 つを強化学習と呼ぶ．この手法の特長の 1 つに，状態の遷移確率が既知でない状況下でも状態に適応した行動を逐次的に学習できるという点がある．

強化学習をファイナンスの分野へ応用させた例として，文献 [1] や [2] がある．文献 [1] では，強化学習を用いて株の取引戦略の学習を行っている．その結果，急激な環境にも対応可能な，安定した収益率を獲得する取引戦略が学習できることが示されている．また文献 [2] は，強化学習を日本国債の取引に適用させることによって金融市場の取引戦略を学習させている．この研究により，強化学習を用いることで実際の投資家の行動に近い投資戦略が獲得できることが明らかとなった．

ファイナンスにおいては，報酬の大きさそのものよりも投資額に対するリターンが重要となることが多い．また，投資により獲得した利益を次の期に再投資する複利効果も，資産運用する際に重要なファクタである．そこで，複利効果を考慮した投資戦略を学習する研究が近年活発になっている．たとえば，文献 [3] は従来の Q 学習を用いた強化学習から，複利効果を最大化する複利型の強化学習へと拡張する手法を提案した．

文献 [4] は，国債銘柄選択問題に複利型強化学習を適用させることによって，従来の強化学習との比較を行った．その結果，リターンの幾何平均を大きくするためには，従来の強化学習よりも複利型強化学習の方が優れていることを示した．本研究においても，この複利型強化学習を用いて学習を行う．

リターンに 1 を足した値をグロスリターンと呼ぶ．複利型強化学習では，このグロスリターンを二重指数で割り引いた

$$\prod_{k=0}^{\infty}(1+R_{t+k+1})^{\gamma^k} \tag{1}$$

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 のリターンを示している．二重指数とは $f(x)=a^{b^x}$ の形で表される関数のことである．上記の二重指数割引されたグロスリターンの対数をとった値を対数割引複利リターンと呼ぶ．しかし対数割引複利リターンは，賭けた金額をすべて失った場合，すなわちリターンが -1 のときに $\log 0=-\infty$ となり発散してしまうため，それを避けるために投資比率 f を導入する手法が提案されている [5]．このとき，文献 [5] では投資比率は $Rf>-1$ となるように設定されている．なお投資比率とは，保有資産に対する投資額の割合のことであり，過剰投資を避け，破産して資産が 0 にならないようにするために用いられている．

また，複利型強化学習の枠組みの中で，各資産への投資比率をオンライン勾配法によって最適化する方法が文献 [6] によって提案されている．既知のリターン分布から求められる複利リターンを最大化するような，最適な投資比率のことをケリー基準と呼ぶ．文献 [6] はこのオンライン勾配法を用いる手法により，ケリー基準とほとんど一致する，最適な投資比率を学習できることを示した．

文献 [7] では，この最適化手法を用いて株式のポジション調整を行い，さらに投資比率を最適化した場合と固定した場合とを比較することにより，この手法の有効性を示した．また，この最適化によって従来よりも柔軟な取引ルールを獲得できることが分かっている．

このように，複利型強化学習に関する研究は数多くなされ，その有効性が示されてきた．しかし，それらの多くは投資対象が 1 つであったり，いくつかの資産の中から 1 つの投資先を決定する場合であり，資産をポートフォリオとして所有していた場合の投資比率最適化については明らかにされていない．一般的にリスクのある資産に投資する場合には，ポートフォリオを構築することにより単一資産に投資するよりも資産価値の下がる確率を抑えることができる [8]．そこで本研究では，既存の研究手法をポートフォリオ選択問題へと拡張する．なお本研究におけるポートフォリオの最適化とは，資産家の将来の期待複利リターンを最大化するような各資産への重みを求めることとする．

ポートフォリオ最適化の研究は，数理計画モデルによる研究が行われている．特に本研究で検討しているような多期間にわたる投資行動を考慮したポートフォリオ選択問題は，多期間確率計画モデルによる研究が活発に行われている [9], [10], [11], [12]．これらはリターンの分布が既知の状況での多期間にわたる最適戦略を解く手法であり，本研究で検討するリターンに応じて適応的に学習する環境とは異なっている．また，複利型強化学習以外のポートフォリオ選択の適応的学習手法として，遺伝的アルゴリズムや遺伝的プログラミングなどの進化的手法がある [13], [14], [15], [16], [17], [18]．これらの手法も，投資戦略を表現した遺伝子の適合度計算時にリターン分布を既知と仮定している場合が多い [13], [14], [15]．また，遺伝的アルゴリズムなど進化的手法による解探索に関しては，理論的な最適解の保証は一般的にはなされていない [15]．そのため，すべての状況に進化的手法が適用可能かは明らかではなく，強化学習や他の学習法を用いたポートフォリオ選択研究を試行する意義は少なくない．複利型強化学習は逐次的に学習を行うため，各資産のリターンの確率分布が未知の状況下で投資を行いながら，リターンに応じて逐次投資戦略を調整する．これにより動的なポートフォリオ構築が可能となり，動的な環境変化にも対応できることが予想される．

本研究では多腕バンディット問題に，複利型強化学習によるポートフォリオ最適化を適用し，性能を評価する．シミュレーション実験より，ポートフォリオ構築を組み合わせた本学習法は，投資先を 1 つに限定した場合よりも高い

幾何平均リターンが得られることが確認された．また，学習した各資産への重みはケリー基準に近い値であり，本提案手法により最適な重みが学習できることが確認された．

2. 複利型強化学習

従来の強化学習では，割引利益の期待値を最大化する行動を学習することを目的としていた．割引利益は，以下のように定義される．

$$r_{t+1}(a)+\gamma r_{t+2}(a)+\gamma^2 r_{t+3}(a)+\cdots = \sum_{k=0}^{\infty}\gamma^k r_{t+k+1}(a) \tag{2}$$

ここで γ は割引率，$r_t(a)$ は時刻 t，方策 a の下での報酬である．割引率とは，将来価値を現在価値に換算するために割り引くパラメータである．したがって，従来の強化学習の目的関数は以下である．

$$\mathrm{E}\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}(a)\right] \tag{3}$$

この目的関数を最大化するような方策 a を学習する．

しかし，ファイナンスにおいては一般的に前の期で得た利益も次の期の掛け金に組み入れるため，利益よりも複利リターンの方が重視される．ここで，ある 1 つの投資先に総資産のある割合 f だけ投資することを考える．投資に対するネットのリターンを R で表すと，毎期間資金は $(1+Rf)$ 倍されていくため，複利リターンは次式で表される．

$$(1+R_{t+1}f)(1+R_{t+2}f)(1+R_{t+3}f)\cdots = \prod_{k=0}^{\infty}(1+R_{t+k+1}f) \tag{4}$$

文献 [3] は，この複利リターンに割引の概念を導入し，以下の式で表される割引複利リターンを提案した．

$$(1+R_{t+1}f)(1+R_{t+2}f)^{\gamma}(1+R_{t+3}f)^{\gamma^2}\cdots = \prod_{k=0}^{\infty}(1+R_{t+k+1}f)^{\gamma^k} \tag{5}$$

この割引複利リターンに対数をとることで，式 (2) と同じ形で表すことができる．したがって，複利型強化学習の目的関数は次式のようになる．

$$\mathrm{E}\left[\log\prod_{k=0}^{\infty}(1+R_{t+k+1}f)^{\gamma^k}\right] = \mathrm{E}\left[\sum_{k=0}^{\infty}\gamma^k\log(1+R_{t+k+1}f)\right] \tag{6}$$

すなわち，式 (6) を最大化する行動を学習することを目的とする．ここで，R_t は時刻 t のリターン，f は投資比率である．また，時刻 1 から $t+1$ までの複利リターンを以下で表す．

$$G_{t+1}=\prod_{k=1}^{t+1}(1+R_k f) \tag{7}$$

なお，複利リターンの対数をとり，その期待値（算術平均）を最大にするような投資比率は資産の成長率（複利リターンの幾何平均）を最大にする投資比率であることが知られている [19]．これより，γ が十分に 1 に近い条件の下では，式 (6) を最大化する解は式 (5) で表現される資産の成長率を最大化させる．

本研究では上記の複利型強化学習のモデルを，N 個の投資可能資産が存在する場合のポートフォリオ選択問題へと拡張する．すなわち，式 (7) を N 個の資産を保有していた場合のポートフォリオの複利リターンとして次のように書き換える．

$$\prod_{k=0}^{\infty}\left(1+\sum_{i=1}^{N}R_{i,k}w_i\right) \tag{8}$$

ここで，$R_{n,t}$ は資産 n の時刻 t でのリターン，w_n はポートフォリオに対する資産 n の重みを表す．N 番目の資産はつねにリターンが 0 となるような無リスク資産とする．つまり，すべての時刻 k について $R_{N,k}=0$ とする．また，$R_{i,k}>-1$ をすべての時刻 k について仮定する．

式 (5) と同様の割引複利リターンを用いて，本研究の最適化問題を以下のように定式化する．

[制約条件付き最大化問題]

$$\max_{w\in\mathbb{R}^N}\mathrm{E}\left[\log\prod_{k=0}^{\infty}\left(1+\sum_{i=1}^{N}R_{i,k}w_i\right)^{\gamma^k}\right] \tag{9}$$

$$\text{制約条件}\quad \sum_{i=1}^{N}w_i=1 \tag{10}$$

$$w_i\geq 0\quad (i=1,\ldots,N) \tag{11}$$

式 (9) のパラメータ γ は 1 に近い値を選び，式 (8) の精度の良い近似となるようにする．シミュレーションでは，$\gamma=0.9$ を用いた．なお資産 n に対する重み w_n は，最急降下法を用いて各時刻 t で以下のように更新を行う．

$$\vec{w}^{t+1}=\vec{w}^{t}+\eta\nabla_{\vec{w}}\log G_{t+1} \tag{12}$$

ここで，

$$\nabla_{\vec{w}}\log G_{t+1}=\left[\frac{\partial\log G_{t+1}}{\partial w_1},\cdots,\frac{\partial\log G_{t+1}}{\partial w_N}\right]^T \tag{13}$$

である．

強化学習では選択した行動を実行した結果のみ観察可能であることが一般的だが，本研究では各離散時間で各資産のリターンが観察できる資産市場を仮定する．この仮定

は，式 (12) の計算に必要となる．一般的な仮定より強い仮定であるが，たとえば株式市場においては，未投資の銘柄の値動きも通常は観察できるため，資産市場という限定した状況ではある程度現実性を持つ仮定と考えられる．

3. 投資比率最適化法

式 (7) の両辺に対数をとり，w_n で偏微分することにより次式が得られる．

$$\begin{aligned}\frac{\partial}{\partial w_n}\log G_{t+1} &= \frac{\partial}{\partial w_n}\log\prod_{k=1}^{t+1}\left(1+\sum_{i=1}^{N}R_{i,k}w_i\right)\\ &= \sum_{k=1}^{t+1}\frac{\partial}{\partial w_n}\log\left(1+\sum_{i=1}^{N}R_{i,k}w_i\right)\\ &= \sum_{k=1}^{t+1}\frac{R_{n,k}}{1+\sum_{i=1}^{N}R_{i,k}w_i}\end{aligned} \tag{14}$$

したがって，資産 n に対する重みは式 (12) を使い，次式で更新していく．

$$w_{n,j+1} = w_{n,j} + \eta\sum_{k=1}^{t+1}\frac{R_{n,k}}{1+\sum_{i=1}^{N}R_{i,k}w_{i,j}} \tag{15}$$

ここで η は学習率である．なお，学習率とは更新幅を表すパラメータのことである．式 (15) では記憶しなければならないパラメータ数が多いため，従来の複利型強化学習の手法に則りオンライン勾配法を用いて資産 n の重み w_n を以下のように更新する．

$$w_{n,t+1} = w_{n,t} + \eta\frac{R_{n,t+1}}{1+\sum_{i=1}^{N}R_{i,t+1}w_{i,t}} \tag{16}$$

しかし，この更新により式 (10) を満たさなくなってしまうため，調整の必要がある．本研究では，以下のようにして毎期間，逐次的に調整を行うことを提案する．

$$\hat{w}_{n,t+1} = \frac{w_{n,t+1}}{\sum_{i=1}^{N}w_{i,t+1}} \tag{17}$$

時刻 $t+1$ での学習では，$\hat{w}_{n,t+1}$ を式 (16) の更新前重みとして扱う．

また運用成績の評価として，幾何平均リターンを用いる．n 期間の幾何平均リターンは以下のように定義されている．

$$\bar{G}_n = \left(\prod_{t=1}^{n}\left(1+\sum_{i=1}^{N}R_{i,t}w_i\right)\right)^{\frac{1}{n}} - 1 \tag{18}$$

この幾何平均リターンは，投資成績の評価に用いられている尺度であり，平均的に，1 期間ごとに資産が $1+\bar{G}_n$ 倍になることを意味している．

4. ケリー基準

ケリー基準とは複利リターンを最大化するような投資比率のことである．式 (7) に従い，t 期までの複利リターンを以下で表す．

$$G_t = \prod_{k=1}^{t}(1+R_k f) \tag{19}$$

ここで，t 期間に R というリターンが $a(t)$ 回発生したとする．すると，$\lim_{t\to\infty}\frac{a(t)}{t}$ は R の発生確率に近づく．発生しうるリターンが m パターンあるとすると，複利リターンは以下のように書くことができる．

$$G_t = \prod_{k=1}^{m}(1+R_k f)^{a_k} \tag{20}$$

ここで，a_k は t 期間までの R_k の発生回数とする．したがって，両辺に t 乗根をとると

$$g(f) = \prod_{k=1}^{m}(1+R_k f)^{P_k} \tag{21}$$

となり，これがケリー基準の目的関数である．ただし P_k は R_k の発生確率であり，$\lim_{t\to\infty}\sqrt[t]{G_t} = g(f)$.

これを最大化するため，式 (21) の両辺に対数をとり f で偏微分すると，

$$\begin{aligned}\frac{\partial}{\partial f}\log g(f) &= \frac{\partial}{\partial f}\log\prod_{k=1}^{m}(1+R_k f)^{P_k}\\ &= \frac{\partial}{\partial f}\sum_{k=1}^{m}P_k\log(1+R_k f)\\ &= \sum_{k=1}^{m}\frac{P_k R_k}{1+R_k f}\end{aligned} \tag{22}$$

となる．すなわち，ケリー基準とは

$$\sum_{k=1}^{m}\frac{P_k R_k}{1+R_k f} = 0 \tag{23}$$

を満たす投資比率 f のことである．ただし，m の数が多い場合計算で解くことは難しいため，通常は解析ソフトなどを用いて求める．

N 資産ポートフォリオを考える場合，ケリー基準の目的関数は発生しうるすべての場合を掛け合わせる必要があるため，以下のようになる．

$$g(w) = \prod_{k_1=1}^{m_1}\prod_{k_2=1}^{m_2}\cdots\prod_{k_N=1}^{m_N}\left(1+\sum_{i=1}^{N}R_{i,k_i}w_i\right)^{P_k} \tag{24}$$

ここで，m_n は資産 n で発生する可能性のあるリターンのパターン数である．R_{i,k_i} は資産 i で発生しうる k_i 番目のリターンを表している．P_k はリターンの組み合わせ $(R_{1,k_1},\cdots,R_{N,k_N})$ の発生確率を示す．ただし，$k=(k_1,\cdots,k_N)$ とする．なお式 (9) の目的関数と形が違うのは，理論的には発生するすべての可能性を考慮しなければならないが，実際に発生するリターンの実現値は，1 期につき 1 つのみだからである．

式 (24) も先程と同様に両辺に対数をとり，w_n で偏微分することにより次式が得られる．

$$\begin{aligned}
&\frac{\partial}{\partial w_n}\log g(w)\\
&=\frac{\partial}{\partial w_n}\log\prod_{k_1=1}^{m_1}\prod_{k_2=1}^{m_2}\cdots\prod_{k_N=1}^{m_N}\left(1+\sum_{i=1}^{N}R_{i,k_i}w_i\right)^{P_k}\\
&=\frac{\partial}{\partial w_n}\sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\cdots\sum_{k_N=1}^{m_N}P_k\log\left(1+\sum_{i=1}^{N}R_{i,k_i}w_i\right)\\
&=\sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\cdots\sum_{k_N=1}^{m_N}\frac{P_kR_{n,k_n}}{1+\sum_{i=1}^{N}R_{i,k_i}w_i} \qquad (25)
\end{aligned}$$

したがって，N 資産ポートフォリオのケリー基準とは，

$$\sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\cdots\sum_{k_N=1}^{m_N}\frac{P_kR_{n,k_n}}{1+\sum_{i=1}^{N}R_{i,k_i}w_i}=0 \qquad (26)$$

を，w_1 から w_N まで偏微分して得られた N 個の方程式を解くことにより求められる，各資産への最適な重みの集合 $(w_1,\cdots,w_N)$ のことである．しかし，これで得られた値は式 (10) を満たすとは限らない．そこで本研究では，投資比率の制約条件である式 (10) を満足するような各資産の重みの組合せの中で，ケリー基準の目的関数である式 (24) を最大にする重みの組合せを選ぶという，条件付最大化問題の解を疑似的な最適解とした．またそれは，式 (9) を最大化するような値となっている．

5. シミュレーション

5.1 シミュレーション方法

図 1 および**図 2** はそれぞれ 1 つのポートフォリオ選択問題を表し，図内の 4 つのルーレットが各問題での投資先を表している．各ポートフォリオ選択問題において，提案手法と従来の複利型 Q 学習 [6] との幾何平均リターンによる運用成績の比較を行う．ただし，文献 [6] はポートフォリオとして投資比率を最適化している訳ではなく，Q 値に基づきどのルーレットを選択するかを決定し，そのうえで 1 つのルーレットのみに対する最適な投資比率を求めている．そのため単純に比較することはできないが，ポートフォリオとして資産運用した場合に，1 つに投資する場合と比べてどの程度の差が出るかの参考値として示している．また，これについては元の文献 [6] と同じパラメータ設定によりシミュレーションを行った．

図 1，図 2 のルーレットの各値はネットのリターンを表している．例として，図 1 のルーレット A に 100 円投資して「2.0」という結果となった場合には，元金の 100 円に加えて 200 円が手に入る．ただし，自分の所持金を全額投資している訳ではないため，合計した所持金が必ずしも 300 円になるとは限らない．なお，それぞれの出目が出る確率は 1/6 である．

また，ルーレット D および D′ はすべてのリターンが 0

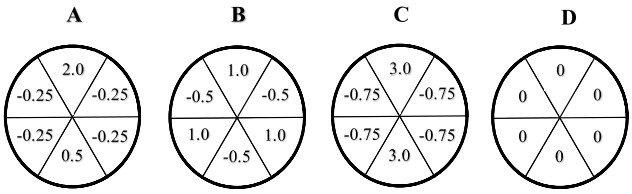

図 1 ポートフォリオ 1
Fig. 1 Portfolio 1.

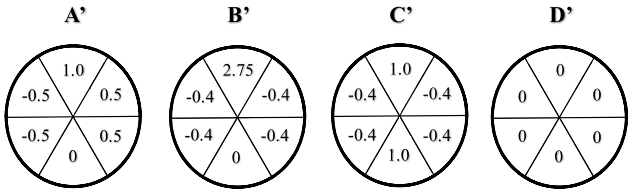

図 2 ポートフォリオ 2
Fig. 2 Portfolio 2.

表 1 ポートフォリオ 1 の各ルーレットの統計量
Table 1 Statistics of each roulette in Portfolio 1.

	A	B	C	D
算術平均	0.25	0.25	0.5	0
幾何平均	1.061	1	0.630	1
標準偏差	0.829	0.75	1.768	0

表 2 ポートフォリオ 2 の各ルーレットの統計量
Table 2 Statistics of each roulette in Portfolio 2.

	A'	B'	C'	D'
算術平均	0.167	0.192	0.067	0
幾何平均	1.020	0.887	0.896	1
標準偏差	0.553	1.153	0.660	0

となっているが，これは安全な投資先を表しており，このルーレットに投資する場合，資金の増減はない．したがって，ルーレット D および D′ に投資するということは，投資をせずに保有していたことと同義であるため，$1-w_D$ および $1-w_{D'}$ は保有資産に対する投資比率を表す．この安全な投資先という設定により，本提案手法はそれぞれの資産に対する最適な重みだけでなく，最適な投資比率も同時に学習できると考えられる．

表 1 および**表 2** は，それぞれポートフォリオ 1 およびポートフォリオ 2 の各ルーレットの統計量について示したものである．各表の算術平均は，リターンの算術平均を表している．幾何平均はリターンではなく，リターンに 1 を加えたグロスリターンの幾何平均を表す．たとえばルーレット A なら，

$$\sqrt[6]{3\times1.5\times0.75^4}=1.06066\cdots$$

である．この値は，それぞれの出目が出る確率が等しいな

らば，当該ルーレットに全資産を賭けた場合，自分の資産が 1 期後に何倍になるかの期待値を表している．

表 3 および**表 4** の最適投資比率（個別投資）とは，各資産に個別投資した場合のケリー基準による最適な投資比率のことである．ただし，ルーレット D およびルーレット D′ は投資をしたとしても資産が減ることはない無リスク資産である．そのような無リスク資産についてケリー基準を適用することはできない．一般的に，最適投資比率が高いほど運用成績が良いルーレットであることが知られている．表 3・表 4 の期待リターン（個別）とは個別投資したときのケリー基準の目的関数のことであり，最適投資比率で資産運用した場合の 1 期後の期待グロスリターンを表している．これは，総資産が 1 であるときの 1 期後の総資産の期待値と等しい．つまり，たとえばポートフォリオ 1 の A のみに最適投資比率である全資産の 69.3%を投資した場合，1 期後の総資産の期待値は 1.070 である．表 3・表 4 の最適重み（分散投資）とは，4 つの資産をポートフォリオとして持ったときの各ルーレットの最適な重みのことであり，期待リターン（分散）はそのときの期待グロスリターンである．

ルーレット C および B′，C′ に着目すると，投資比率 1 で運用した場合の成績（幾何平均）は 1 を切っているにもかかわらず，最適な投資比率で投資した場合の成績は 1 を上回っている．したがってあくまでも長期的には，最適な投資比率で運用すると，通常なら負ける賭けであってもプラスの収益へと変えられる可能性がある．これが最適な投資比率を求めることの有用性である．また表 3，表 4 の個別投資と分散投資の期待リターンを比較すると，分散投資は個別投資のどの数値よりも期待リターンが高くなっていることが分かる．したがって，最適な重みを学習可能ならば，資産はポートフォリオとして所持した方がよい．

ここで，図 1 に示したポートフォリオ 1 のルーレット A は幾何平均が最大であり，ルーレット B は標準偏差が最小であり，ルーレット C は算術平均が最大である．複利を考慮せず，期待報酬を最大化するならばルーレット C の投資比率が高くなり，リスク（分散）を最小化するならばルーレット B への投資比率が高くなると考えられる．しかし，幾何平均リターンを最大にする学習がなされていれば，表 3 の最適重み（分散投資）に示されているようにルーレット A への投資比率が最も高くなるはずである．この点を検証する．図 2 に示したポートフォリオ 2 の特徴は，ポートフォリオ 1 では 0 であった無リスク資産が正の最適重みを持つことである．無リスク資産へ正の割合で投資する行動が学習可能かを検証する．また，ルーレット B′ は算術平均・幾何平均共にルーレット C′ よりも低く，かつ標準偏差（リスク）が C′ よりも高い投資対象となっている．しかし，ポートフォリオを構築した場合の最適な重みは C′ よりも高くなっている．一般的な基準では B′ より C′ が選

表 3 最適投資比率と最適重みおよび期待リターン（ポートフォリオ 1）

Table 3 Optimal investment ratio and optimal weight and expected return of portfolio 1.

	A	B	C	D
最適投資比率 (個別投資)	0.693	0.5	0.222	−
期待リターン（個別）	1.070	1.061	1.050	1
最適重み（分散投資）	0.454	0.363	0.183	0.000
期待リターン（分散）	1.170			

表 4 最適投資比率と最適重みおよび期待リターン（ポートフォリオ 2）

Table 4 Optimal investment ratio and optimal weight and expected return of portfolio 2.

	A’	B’	C’	D’
最適投資比率 (個別投資)	0.580	0.209	0.167	−
期待リターン（個別）	1.048	1.018	1.005	1
最適重み（分散投資）	0.561	0.196	0.149	0.094
期待リターン（分散）	1.070			

択されそうであるが，このような場合でも最適重みに従った学習が可能かを検証する．

シミュレーションでは，学習率 $\eta = 0.001$，割引率 $\gamma = 0.9$ を用いた．この値は，文献 [6] と同様の設定である．まず，ルーレット D の重みの初期値が $w_{D,0} = 1$ の場合と，それぞれの重みの初期値が $w_{n,0} = 1/4$ の場合についてシミュレーションを行った．また，図 2 のリターンを変えたルーレットについても，同様のシミュレーションを行った．シミュレーションはステップ数を 100 万回とし，100 回の平均により評価した．

5.2 シミュレーション結果および考察

図 3 は，ポートフォリオ 1 においてルーレット D への重みの初期値が 1 の場合の幾何平均リターンの推移であり，**図 4** は初期値として各ルーレットに 1/4 ずつ投資したときの幾何平均リターンの推移を表している．

A のみに投資するだけでは理論的に期待リターンは 7.0%程度しかないが，ポートフォリオとして運用することで 17%以上ものリターンが期待できる．図 3 と図 4 のいずれの場合でも，今回の提案手法を用いることで高い幾何平均リターンを生み出す重みを学習できた．このことから，本提案手法が投資の初期状態によらず最適な学習が行える手法であることが確認できた．また，ルーレットをポートフォリオ 2 に変えて図 4 と同じシミュレーションを行った結果について示したものが**図 5** である．ルーレット 2 でも同様の良好な学習の結果が得られた．

図 3，図 4 および図 5 を見ると，従来手法は幾何平均リターンがそれぞれ約 7%，7%，5%弱という結果になってい

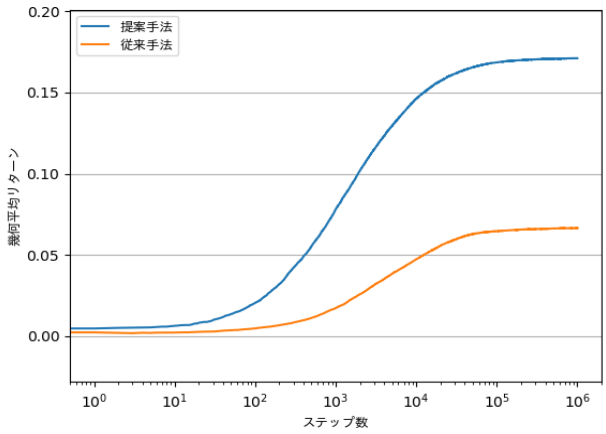

図 3 ポートフォリオ 1 における $w_{D,0} = 1$ のときの幾何平均リターンの推移

Fig. 3 Transition of geometric mean return in case of $w_{D,0} = 1$ in Portfolio 1.

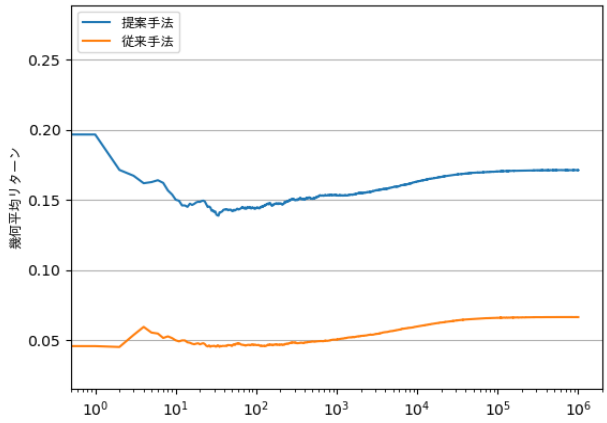

図 4 ポートフォリオ 1 における重みの初期値がそれぞれ $w_{n,0} = 1/4$ のときの幾何平均リターンの推移

Fig. 4 Transition of geometric mean return in case of default weight values are $w_{n,0} = 1/4$ respectively in Portfolio 1.

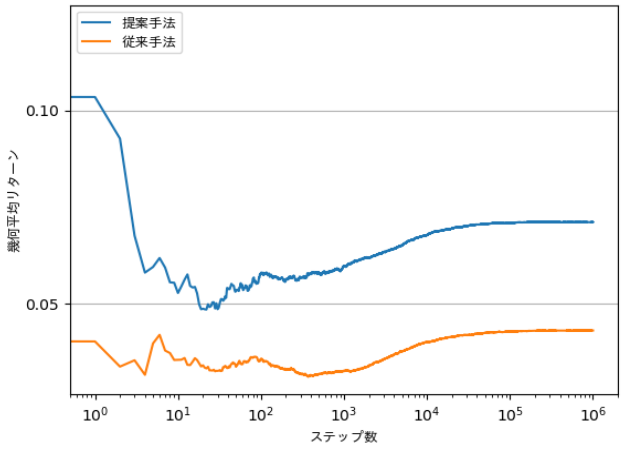

図 5 ポートフォリオ 2 における重みの初期値がそれぞれ $w_{n,0} = 1/4$ のときの幾何平均リターンの推移

Fig. 5 Transition of geometric mean return in case of default weight values are $w_{n,0} = 1/4$ respectively in Portfolio 2.

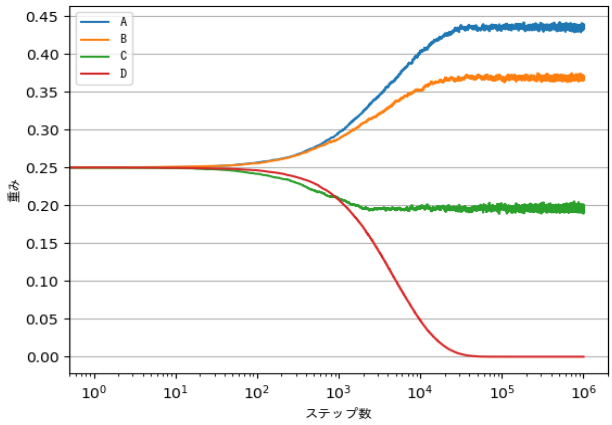

図 6 ポートフォリオ 1 における重みの初期値が各 $w_0 = 1/4$ のときの提案手法の重みの推移

Fig. 6 Transition of weight in the proposed method in case of default weight values are $w_0 = 1/4$ respectively in portfolio 1.

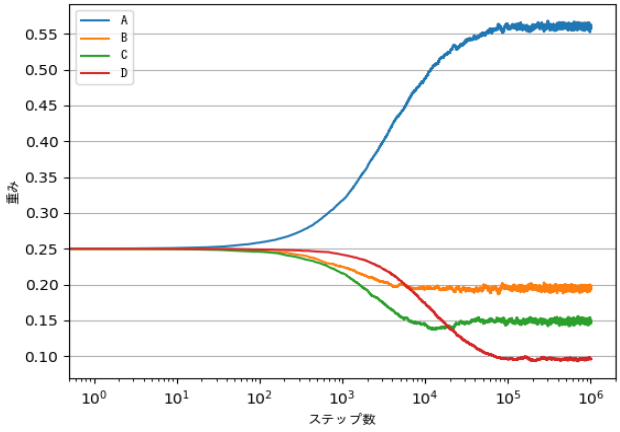

図 7 ポートフォリオ 2 における重みの初期値が各 $w_0 = 1/4$ のときの提案手法の重みの推移

Fig. 7 Transition of weight in the proposed method in case of default weight values are $w_0 = 1/4$ respectively in portfolio 1.

る．これは，期待リターンの一番高い A または A′ を多く選択するという行動規則を正しく学習できていることを示唆している．また，7%および 5%弱という幾何平均リターンは表 1 および表 2 の示す A または A′ の投資に対する最適な期待リターンと一致しており，従来手法では投資先を 1 つに限定した状況での最適な投資比率が学習されていることが確認された．一方で，図 3〜図 5 よりつねに提案手法の方が従来手法よりも高い幾何平均リターンを実現できていることが確認できる．表 3，表 4 が示すように資産はポートフォリオとして運用した方が有利であり，提案手法により最適なポートフォリオの組み合わせの学習が進んだことが確認された．

図 6 および**図 7** は，重みの初期値としてそれぞれルーレット 1，2 の各ルーレットに 1/4 ずつ投資したときの，提案手法における重みの推移を表している．

表 5 ポートフォリオ 1 の学習した重みと期待リターン
Table 5 Learned weight and expected return on assets for portfolio 1.

	A	B	C	D
学習した重み	0.443	0.364	0.192	0.000
理論値との差	0.011	0.001	0.009	0
期待リターン	1.170			

表 6 ポートフォリオ 2 の学習した重みと期待リターン
Table 6 Learned weight and expected return on assets for portfolio 2.

	A'	B'	C'	D'
学習した重み	0.558	0.198	0.148	0.096
理論値との差	0.003	0.002	0.001	0.002
期待リターン	1.071			

重みの合計が 1 であることを条件とした場合のポートフォリオのケリー基準による最適な重みは，ポートフォリオ 1 は $w_A \approx 0.454$，$w_B \approx 0.363$，$w_C \approx 0.183$，$w_D \approx 0$ であり，ポートフォリオ 2 は $w_{A'} \approx 0.561$，$w_{B'} \approx 0.196$，$w_{C'} \approx 0.149$，$w_{D'} \approx 0.094$ であった．

表 5 と**表 6** は，それぞれポートフォリオ 1 およびポートフォリオ 2 の学習した重み，学習した重みと表 3，表 4 の理論値との差（絶対値），および学習により得られた期待リターンを表にしたものである．これらを見ると，どちらのルーレットも最適な重みに近い値が獲得できていることが分かる．

ルーレット 2 では安全な投資先である D′ への最適な重みの理論値が 0 より大きく，それを正しく学習できていることも読み取れる．したがって，最適な投資比率も学習されたと考えられる．また，ルーレット B′ の最適重みは C′ よりも高くなっているが，表 6 よりその最適重みが学習できているため，複利型強化学習では幾何平均の相対的な大きさに基づいた重みではなく，ケリー基準に沿うような最適重みが学習できることが分かった．さらに，期待リターンについても理論値とほとんど同じ数値となっているため，本提案手法によってポートフォリオ最適化が可能となっていることが確認された．

5.3 ポートフォリオが最適化されないケース

ここでは，本手法ではポートフォリオが最適化されないケースを紹介し，その原因を考察する．確率 1 でリターン 1.0 を得られるルーレット A*と，確率 1 でリターン 0.1 を得られるルーレット B*のポートフォリオ選択問題を考える．これをポートフォリオ 3 とする．この問題では A*に 1.0，B*に 0 の重みで投資するのが最もよいことが明らかである．このルーレットについてシミュレーションした結果を**図 8** に示す．

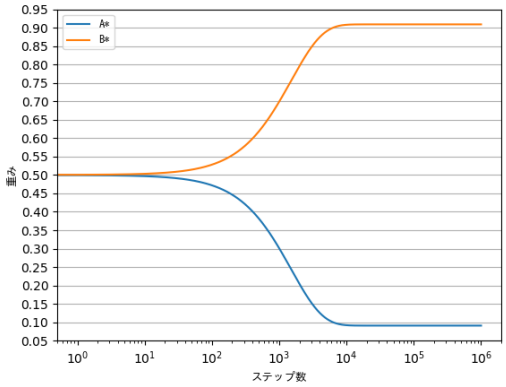

図 8 ポートフォリオ 3 における重みの推移
Fig. 8 Transition of weight in portfolio 3.

A*の重みが 0.909，B*が 0.091 で投資するというシミュレーション結果となり，実際の最適な重みとは異なる結果となった．この設定では，式 (16) の $R_{n,t+1}$ 値がルーレット A*では必ず $R_{n,t+1} = 1$ となり，ルーレット B*では必ず $R_{n,t+1} = 0.1$ となる．そのため，10 : 1 という限界的な収益の比率で重みが学習されてしまっていると考えられる．

式 (26) に基づきケリー基準による最適重みを求めると，

$$\begin{cases} \dfrac{1}{1 + 1.0w_{A^*} + 0.1w_{B^*}} = 0 \\ \dfrac{0.1}{1 + 1.0w_{A^*} + 0.1w_{B^*}} = 0 \end{cases}$$

となり，解が求められないことが分かる．このようなケースでは，本論文の複利強化学習による手法ではポートフォリオが最適化されない可能性がある．

また，重み更新式の式 (16) では式 (10) の制約条件が考慮されていない．そのため，制約条件なしで幾何平均リターンを最大化させる式 (16) の勾配と，制約条件内で幾何平均リターンを最大化させる勾配が一致しない場合，最適化されないことが考えられる．

6. まとめ

本論文では，従来の複利型強化学習をポートフォリオ選択問題へと拡張し，今まであまり研究がなされていなかったポートフォリオの最適な重みを複利型強化学習により学習させることを試みた．シミュレーション結果より，本提案手法によってケリー基準の理論値に近い最適な重みを学習できることが明らかとなった．また，学習された重みから計算される期待リターンも同様に理論値にかなり近い値となり，複利リターンを最大にするような運用成績を出せることが確認された．

最後に今後に検討すべき課題と現時点で考えられる対策をまとめる．本論文では自分の資産以上に投資できない条件の下で計算したケリー基準と提案手法の比較を行っている．しかし，ケリー基準本来の式である式 (26) ではレ

バレッジを考慮したポートフォリオを前提としており，本手法ではレバレッジの対応はされていない．これについては，[20] の手法を応用し，式 (10) の条件を 1 以上に緩和する形でレバレッジを考慮した最適な投資戦略の学習が可能かを確認することが考えられる．また，本手法では学習率 (式 (16) の η) をつねに一定としている．逐次型の学習であるため，市場の緩やかな変化に対する動的なポートフォリオ変更は可能であるが，学習率が一定であるため市場の急激な変化には対応が遅れることが予想される．これについては，市場の急激な変化を検知する機構を追加し，変化検知時には学習率を一時的に増加させるなどの対策が考えられる．たとえば，文献 [21] では市場に急激な変化が起きた際の予測性能を向上する機構を提案している．また，文献 [22] は市場に急激な変化を起こしうる情報が開示された際に影響を受けやすい投資対象を分析している．個々の投資家の行動までモデル化し，市場の動学を分析する研究には文献 [23] などがある．これらの論文で提案されている予測・情報の影響・市場動学の分析手法などと組み合わせることで，市場の変化に対応した学習率の設定の可能性が出てくると思われる．加えて，連続的なリターンの場合にも応用できるかなど多角的な視点からの検証が今後の課題としてあげられる．

謝辞 本研究は JSPS 科研費 18K12740 の助成を受けたものです．また，メタ査読者および 2 名の査読者からの貴重なコメントに謝意を表します．

参考文献

[1] 中原孝信：強化学習を用いたブーム検知型株トレーディングシステムの構築，人工知能学会，第 11 回金融情報学研究会，SIG-FIN, Vol.11, No.4 (2013).
[2] 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，Vol.24, No.3, pp.400–407 (2009).
[3] 松井藤五郎：複利型強化学習，人工知能学会論文誌，Vol.26, No.2, pp.330–334 (2011).
[4] Matsui, T., Goto, T., Izumi, K. and Chen, Y.: Compound: Reinforcement Learning: Theory and An Application to Finance, in Sanner, S. and Hutter, M. (Eds.), *Recent Advanced in Reinforcement Learning: Revised and Selected Papers of the European Workshop on Reinforcement Learning 9* (*EWRL 2011*), Vol.7188 of Lecture Notes in Computer Science, pp.321–332 (2012).
[5] 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，Vol.52, No.12, pp.3300–3308 (2011).
[6] 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，情報処理学会論文誌，Vol.28, No.3, pp.267–272 (2013).
[7] 後藤 卓，松井藤五郎，大澄祥広：複利型強化学習の株式取引への応用，第 27 回人工知能学会全国大会論文集，4I1-OS-16-4 (2013).
[8] 枇々木規雄：ポートフォリオ最適化入門，オペレーションズ・リサーチ，Vol.61, No.6, pp.335–340 (2016).
[9] 枇々木規雄：コンパクト表現によるシミュレーション型多期間確率計画モデルの定式化，日本オペレーションズ・リサーチ学会論文誌，Vol.45, No.4, pp.529–549 (2002).
[10] Hibiki, N.: Multi-period stochastic optimization models for dynamic asset allocation, *Journal of Banking and Finance*, Vol.30, No.2, pp.365–390 (2006).
[11] Takano, Y. and Gotoh, J.: Constant Rebalanced Portfolio Optimization under Nonlinear Transaction Costs, *Asia-Pacific Financial Markets*, Vol.18, No.2, pp.191–211 (2011).
[12] Takano, Y. and Gotoh, J.: Multi-Period Portfolio Selection Using Kernel-Based Control Policy with Dimensionality Reduction, Expert Systems with Applications, Vol.41, No.8, pp.3901–3914 (2014).
[13] 和多田淳三，水沼洋人，松田 浩：遺伝的アルゴリズムを用いた投資銘柄数限定型ファジィ・ポートフォリオ・セレクション，日本経営工学会論文誌，Vol.49, No.2, pp.91–99 (1998).
[14] Lin, C.-C. and Liu, Y.-T.: Genetic algorithms for portfolio selection problems with minimum transaction lots, *European Journal of Operational Research*, Vol.185, No.1, pp.393–404 (2008).
[15] Chang, T.-J., Yang, S.-C. and Chang, K.-J.: Portfolio optimization problems in different risk measures using genetic algorithm, *Expert Systems with Applications*, Vol.36, No.7, pp.10529–10537 (2009).
[16] 柿木秀文，木村周平，松村幸輝：進化的アルゴリズムによるリスク管理を目的とした投資戦略最適化，情報処理学会研究報告，Vol.2009-AL-126, No.9, pp.1–4 (2009).
[17] 伊庭斉志：金融工学のための遺伝的アルゴリズム，オーム社，240pp (2011).
[18] 海野一則，山田隆志，寺野隆雄：機械学習を用いたポートフォリオの最適化，人工知能学会全国大会論文集，Vol.27, pp.1–4 (2013).
[19] Latané, H.A.: Criteria for Choice Among Risky Ventures, *Journal of Political Economy*, Vol.67, No.2, pp.144–155 (1959).
[20] 塚本智大，松井藤五郎：レバレッジを用いた複利型強化学習，情報処理学会第 78 回全国大会講演論文集，Vol.1, pp.355–356 (2016).
[21] 鳥海不二夫，石井健一郎：人工市場を用いた予測市場の予測メカニズムの分析，人工知能学会論文誌，Vol.27, No.6, pp.346–354 (2012).
[22] 岡田克彦，東 高宏，中元政一，羽室行信：証券アナリストの格下げ記事により価値を失う企業の特徴分析，人工知能学会論文誌，Vol.27, No.6, pp.355–364 (2012).
[23] Kanazawa, K., Sueshige, T., Takayasu, H. and Takayasu, M.: Derivation of the Boltzmann Equation for Financial Brownian Motion: Direct Observation of the Collective Motion of High-Frequency Traders, *Physical Review Letters*, Vol.120, No.13, 138301 (2018).

畠山 卓

2018 年筑波大学理工学群社会工学類卒業．現在，筑波大学博士課程前期システム情報工学研究科在学中．機械学習および取引戦略に関する研究に従事．

澤 亮治 （正会員）

1998 年慶應義塾大学理工学部電気工学科卒業．2012 年 Wisconsin 州立大学 Madison 校経済学博士課程修了，経済学博士．会津大学コンピュータ理工学部文化研究センター准教授を経て，2016 年より筑波大学システム情報系社会工学域准教授．ゲーム理論，社会学習理論に関する研究に従事．