

## 科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 5 月 31 日現在

機関番号：33910

研究種目：若手研究（B）

研究期間：2011～2012

課題番号：23700182

研究課題名（和文） ファイナンスにおける時系列データ解析のための強化学習に関する研究

研究課題名（英文） A Study on Reinforcement Learning for Analyzing Temporal Data in Finance

## 研究代表者

松井 藤五郎（MATSUI TOHGOROH）

中部大学・生命健康科学部・講師

研究者番号：90366443

研究成果の概要（和文）：本研究では、強化学習において利益率の複利効果を最大化するために複利型強化学習という新しい強化学習の枠組みを開発した。複利型強化学習を国債銘柄選択、国債取引、株式取引、n 本腕バンディット、ブラックジャックなどに応用し、複利型強化学習がファイナンスやギャンブルのドメインに有効であることを確認した。また、複利型強化学習で導入された投資比率パラメーターをオンライン勾配法によって最適化する手法を開発した。

研究成果の概要（英文）：We developed a new reinforcement learning framework, called compound reinforcement learning to bond selection, bond trading, stock trading, n-armed bandit, and blackjack and confirmed that compound reinforcement learning works well especially in finance and gambling. We also developed a method to optimize bet fraction parameter in compound reinforcement learning using online gradient method.

## 交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	2,800,000	840,000	3,640,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：機械学習、強化学習、複利型強化学習

## 1. 研究開始当初の背景

強化学習は試行錯誤に基づく機械学習の枠組みである。従来の強化学習は将来にわたって獲得する報酬の合計を最大化使用とするが、ファイナンスの分野では、報酬の合計よりも利益率の複利効果を最大化することが求められる。

## 2. 研究の目的

本研究は、ファイナンスにおける時系列データを解析するための強化学習を対象とする。本研究の目的は、強化学習を用いて金融取引戦略を学習するために、時系列データ解析に対する強化学習の問題点を明らかにし、その解決方法として複利型強化学習の枠組みとアルゴリズムを開発することである。

## 3. 研究の方法

(1) 本研究では、利益率に基づいて利益率の複利効果を最大化する複利型強化学習という新しい枠組みを開発した。また、いくつかの複利型強化学習のためのアルゴリズムを開発した。

(2) 開発したアルゴリズムを以下の問題に適用し、その有効性を確認した。

- ① 国債銘柄選択
- ② 国債取引
- ③ 株式取引
- ④ n 本腕バンディット
- ⑤ ブラックジャック

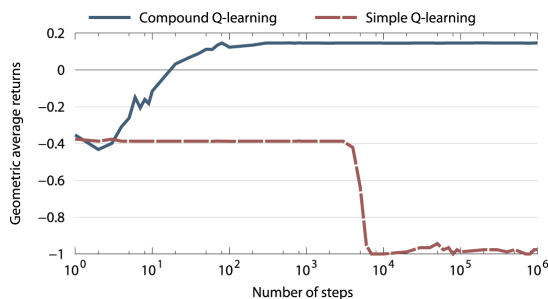


図 1: 2 本腕バンディット問題における幾何平均利益率の推移

(3) 複利型強化学習では、新たに投資比率と呼ばれるパラメーターが導入された。そこで、この投資比率パラメーターをオンライン勾配法によって最適化する方法を開発した。

#### 4. 研究成果

(1) まず、複利型強化学習の理論的枠組みを確立した。複利型強化学習では、報酬の代わりに利益率が観測可能なマルコフ決定過程において二重指数的割引を導入した割引複利利益率の対数の期待値を最大化する。二重指数的割引と対数を導入することによって、従来の強化学習と同様に価値関数の Bellman 方程式を再帰的に表すことを可能とし、従来の強化学習アルゴリズムを複利型強化学習アルゴリズムに拡張することができる。

Q 学習アルゴリズムを複利型に拡張し、2 本腕バンディット問題を用いてその有効性を確認した。図 1 にその結果を示す。従来の Q 学習 (Simple Q-learning) は幾何平均利益率を最大化することができないが、複利型 Q 学習 (Compound Q-learning) は幾何平均利益率を最大化することができた。

(2) 複利型強化学習を国債銘柄選択問題および国債取引問題に適用し、その有効性を確認した。

国債銘柄選択問題では、2010 年末時点における残存期間 5 年の米国債 (利回り 1.929%, デフォルト確率 3.6%), 独国債 (利回り 2.222%, デフォルト確率 5.2%), 英国債 (利回り 2.413%, デフォルト確率 6.4%) について、その利回りとデフォルト確率に基づいて複利型強化学習を用いて利益率の複利効果を最大化するためにどの銘柄を選択すべきかを学習させた。その結果、従来の強化学習がどの国債も大きく変わらない確率で選択したのに対し、複利型強化学習は利回りが高くないがデフォルト確率が低い米国債を高い確率で選択した (図 2)。この結果、複利型強化学習は従来の強化学習を上回る幾何平均リターンを獲得できた (図 3)。

残存期間 10 年の日本国債を対象とした国債取引問題では、学習期間を 3 年間とし、学

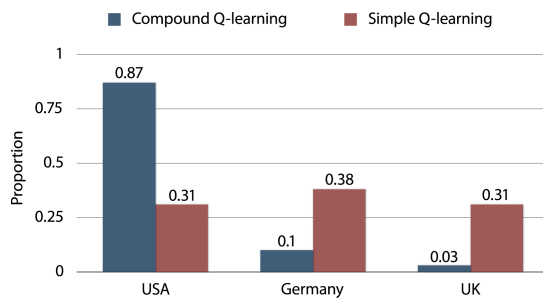


図 2: 国債銘柄選択において学習した選択確率

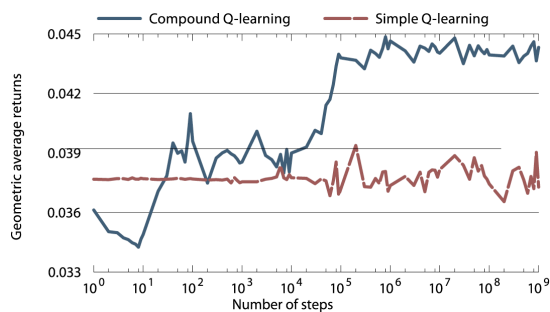


図 3: 国債銘柄選択における幾何平均利益率の推移

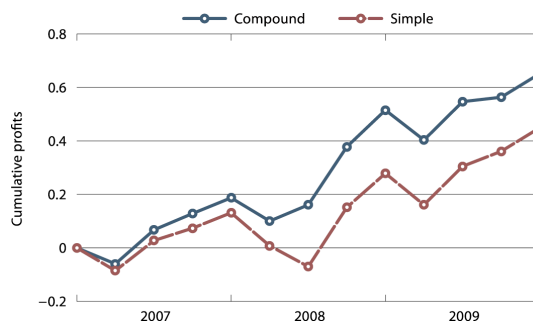


図 4: 日本国債取引問題における累積利益

習した取引戦略を用いてその翌年 1 年間運用したときの累積利益比較した。この結果、複利型強化学習 (Compound) の方が従来の強化学習 (Simple) よりも高い利益を得ることができた (図 4)。

(3) 複利型強化学習で新に導入された投資比率パラメーターをオンライン勾配法によって最適化する方法を開発した。これによって、利用時に投資比率パラメーターを設定する必要がなくなり、複利型強化学習が使いやすくなった。

図 5 に示す 3 本腕バンディット問題を用いて比較を行った。この問題では、平均獲得利益は C が最も高いが、利益率の複利効果を考慮すると A が最も高い利益が獲得できる。投

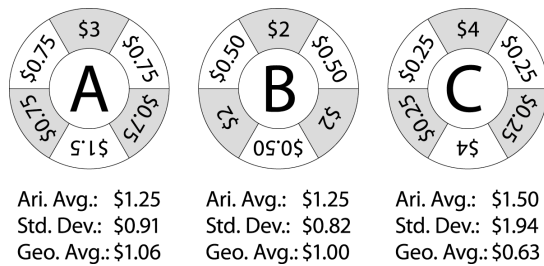


図 5: 3 本腕バンディット問題

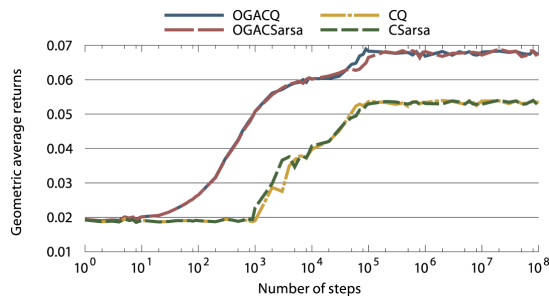


図 6: 3 本腕バンディット問題における幾何平均利益率の比較

資比率が最適化されるため、投資比率最適化付きの複利型強化学習(OGACQ, OGACSarsa)は、投資比率固定の複利型強化学習(CQ, CSarsa)よりも高い幾何平均利益率を獲得することができた(図6)。

株式取引問題では、TOPIX ETF(東証株価指数連動型上場投資信託)を対象として、Javaによる自動取引プラットフォームのカプロボを用いて複利型強化学習によって取引戦略を学習する自動取引プログラムを作成し、2002年から2011年までの10年間を学習期間として10回繰り返し取引戦略を学習した。この結果、投資比率を最適化する複利型強化学習(最適化)は投資比率を0.5に固定した複利型強化学習(固定)よりも高い平均年率換算利回りを獲得できた(図7)。

ただし、カジノゲームの一種であるブラックジャックのように正の利益率を得る行動規則の学習が難しい問題では、投資比率を最適化する際に投資比率が0に収束してしまうケースが生じることが確認された。そこで、そのような場合には投資比率をリセットして学習し直す方法を開発した。この方法(提案手法)は、従来手法に比べて高い幾何平均利益率を獲得することができた(図8)。

## 5. 主な発表論文等

[雑誌論文](計3件)

- (1) 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

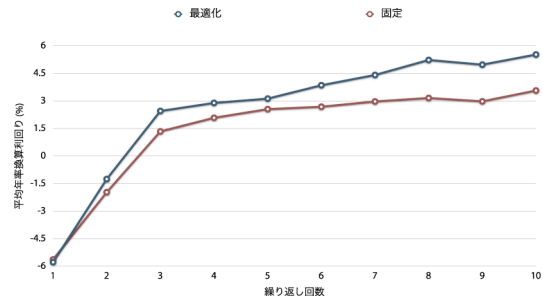


図 7: 株式取引における平均年率換算利回りの推移

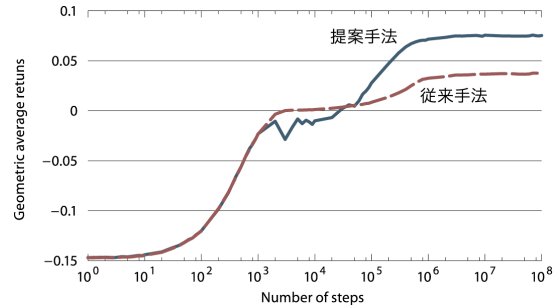


図 8: ブラックジャックにおける幾何平均利益率の推移

適化, 人工知能学会論文誌, Vol. 28, pp. 267-272, 10.1527/tjsai.28.267 (2013), 査読有り

- (2) Tohgoroh Matsui, Takashi Goto, Kiyoshi Izumi, Yu Chen: Compound Reinforcement Learning: Theory and an Application to Finance, Lecture Notes in Computer Science, Vol. 7188, 10.1007/978-3-642-29946-9\_31 (2012), 査読有り
- (3) 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, pp. 3300-3308, <http://id.nii.ac.jp/1001/00079533/> (2011), 査読有り

[学会発表](計9件)

- (1) 松井藤五郎, 落合宏旭: 複利型強化学習による危険回避行動の学習, 第27回人工知能学会全国大会(JS AI2013), 2013年6月4日~7日, 富山
- (2) 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(JS AI2013), 2013年6月4日~7日, 富山
- (3) 松井藤五郎, 後藤卓, 和泉潔: 複利型強化学習における投資比率最適化手法の検討, 第9回人工知能学会ファイナンスにおける人工知能応用研究会

- (SIG-FIN), 2012年11月17日, 横浜
- (4) 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, 第26回人工知能学会全国大会 (JSAI2012), 2012年6月12日~15日, 山口
  - (5) 後藤卓, 松井藤五郎, 和泉潔: 最急降下法を用いたCDSプロキシヘッジにおける最適ヘッジ比率の学習, 第26回人工知能学会全国大会 (JSAI2012), 2012年6月12日~15日, 山口
  - (6) 松井藤五郎: 強化学習のファイナンスへの応用, 電子情報通信学会名古屋大学学生ランチ講演会, 招待講演, 2012年2月9日, 名古屋
  - (7) 松井藤五郎, 後藤卓, 和泉潔, 陳ユ: オンライン勾配法による投資比率最適化付き複利型強化学習, 第8回人工知能学会ファイナンスにおける人工知能応用研究会 (SIG-FIN), 2012年1月28日, 東京
  - (8) Tohgoroh Matsui, Takashi Goto, Kiyoshi Izumi, Yu Chen: Compound Reinforcement Learning: Theory and An Application to Finance, European Workshop on Reinforcement Learning (EWRL2011), 2011年9月11日, アテネ (ギリシャ)
  - (9) 松井藤五郎, 後藤卓, 和泉潔, 陳ユ: 複利型強化学習を用いた国際銘柄選択, 第25回人工知能学会全国大会 (JSAI 2011), 2011年6月1日~3日, 盛岡

[その他]

ホームページ等

<http://とうごろう.jp>

<http://1056lab.org>

## 6. 研究組織

### (1) 研究代表者

松井 藤五郎 (MATSUI TOHGOROH)

中部大学・生命健康科学部・講師

研究者番号: 90366443