

強化学習を用いた金融市場取引戦略分析システムの試作

A Prototype of Financial Trading Strategy Analysis System with Reinforcement Learning

松井藤五郎^{1*} 後藤卓² 和泉潔³ 大和田勇人¹
Tohgoroh Matsui¹, Takashi Goto², Kiyoshi Izumi³, Hayato Ohwada¹

¹ 東京理科大学

¹ Tokyo University of Science

² 三菱東京 UFJ 銀行

² The Bank of Tokyo-Mitsubishi UFJ, Ltd.

³ 産業技術総合研究所

³ National Institute of Advanced Industrial Science and Technology

Abstract: This paper describes a prototype of financial trading strategy analysis system. It uses reinforcement learning to acquire a trading strategy and visualize the strategy. We illustrate how to use this system using an example for trading in the Japanese government bond market.

1 はじめに

近年、ソフトウェア・エージェントをファイナンスに応用する研究が注目を集めている [7]。強化学習は試行錯誤に基づく機械学習の枠組みであり、ソフトウェア・エージェントの行動戦略学習に適している。我々は、これまでに、自動株式取引ソフトウェア・コンテストのカブロボ [6] と日本国債の取引戦略の獲得に強化学習を適用してその有効性を確認している [8, 9, 10, 11, 12]。

本研究は、テクニカル指標に基づく取引戦略の分析をサポートするために強化学習を用いて取引戦略を獲得することを目的としている。強化学習によって獲得された戦略を視覚化して実務家に提示することで、それぞれの市場に適している戦略はどれか、それぞれの戦略が得意とする（または不得意とする）市場はどれかなどを分析することができるようになると考えられる。

本論文では、試作した強化学習を用いた金融市場取引戦略分析システムを紹介する。本システムは、価格（または金利）の時系列データからテクニカル指標を計算し、そこから強化学習によって取引戦略を獲得することができる。また、獲得した取引戦略を可視化することができる。本システムを用いることによって、投資家が

暗黙のうちに獲得している知識を可視化することができるようになり、新たな視点から取引戦略を分析することが可能となる。

2 強化学習を用いた金融市場取引戦略の獲得

本節では、日本国債の取引を例に、強化学習を用いて金融市場における取引戦略を獲得する方法 [9] を紹介する。

株式においては価格が上昇すると運用利回りが上昇するが、日本国債を含む債券において価格が上昇すると運用利回りが減少する。したがって、金利が高い時に債券を買って、金利が低い時に債券を売るのが良い取引方法である。債券を買っている状態をロング・ポジションといい、債券を売っている（信用売りしている）状態をショート・ポジションという。

ここでは、各取引日において市場が閉まる直前に金利を観測でき、その値に応じてすぐに注文するとそのままの金利ですぐに取引が成立するものとする。そこで、市場が閉まる直前の金利を終値で近似することとする。また、取引手数料はかからないものとする。

これらは、銀行の債券取引部門など、実際に日本国債の取引を行っている現場でも同じように取引ができることから、このように単純化しても大きな問題はない。

本研究では、強化学習アルゴリズムには OnPS[2] を

* 連絡先：東京理科大学 理工学部 経営工学科
千葉県野田市山崎 2641
E-mail: matsui@ia.noda.tus.ac.jp

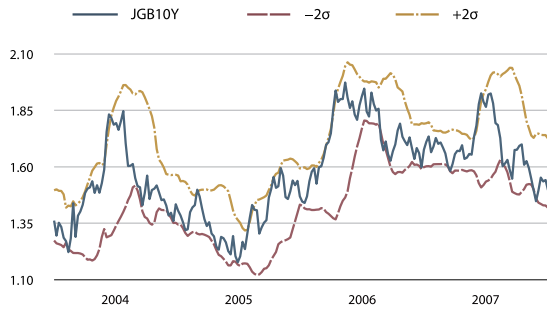


図1 2004年から2007年までの10年国債の金利とそのボリンジャー・バンドの推移。-2σはバンドの下端，+2σはバンドの上端を表す。

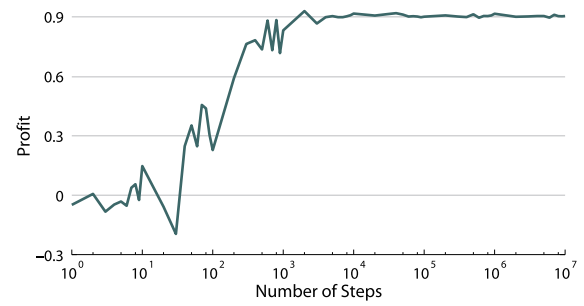


図2 学習曲線 (10回の実験の平均)。

用いている。OnPSは、少ない試行錯誤から学習でき、中間報酬を扱うことができることから、金融市場取引戦略の学習に適した強化学習アルゴリズムである。

強化学習エージェントは、2つのテクニカル指標を観測し、ロング・ポジションかショート・ポジションを行動として選択する。そして、ポジションを開いてから閉じるまでの一連の行動を一つのエピソードとし、評価損益の増分を報酬として強化学習を行う。

ここでは、テクニカル指標としてボリンジャー・バンドを用いた例を紹介する。ボリンジャー・バンドは、価格変動の大きさ（ボラティリティ）を標準偏差によって評価する手法である。2004年から2007年までの10年国債の2σのボリンジャー・バンドを図1に示す。

2σのバンド幅 w_t の変化量とバンド内での相対価格を観測し、次のように表す。

$$\begin{aligned} o_{t,1} &= N_{t,14} [w_t - w_{t-1}] \\ o_{t,2} &= N_{t,14} [p_t] \\ w_t &= 4\sigma_{t,14}(p) \end{aligned}$$

ここで、 p_t は価格（金利）を表し、 $\sigma_{t,k}$ は時刻 t からの直近 k 個のデータから求めた標準偏差を表す。また、 $N_{t,m}$ はそれぞれの値が正規分布に従うと仮定して $\pm 2\sigma$ の範囲を $[-1, 1]$ に変換して正規化することを表す。

図2は、2004年から2007年の金利データを繰り返し用いて強化学習を行った結果を表しており、縦軸は4年間の平均運用利益を表している。強化学習の特色の一つが遅延報酬であることから、遅延報酬を考慮せずに即時報酬だけから学習した場合と比較したところ、遅延報酬を考慮した場合の方が良い結果となった。また、Welchの t 検定を用いて比較したところ、有意水準1%で平均に差があることが確認できた。

図3は、獲得した最も良い戦略においてエージェントがショート・ポジションをとる確率を表している。ボリンジャー・バンドを用いた取引戦略においては、バン

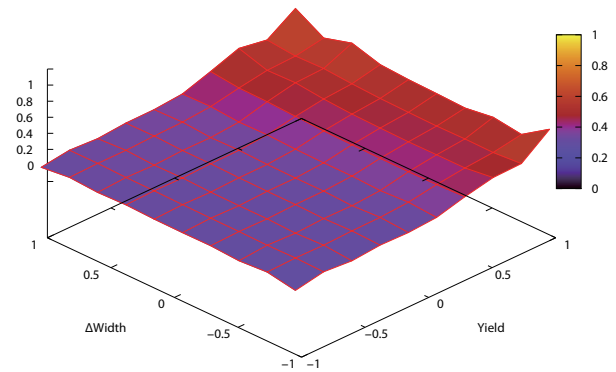


図3 獲得した取引戦略。ショート・ポジションを選択する確率を表す。

ド幅が広がって価格（金利）がバンドを上抜けしたときにその後価格が上昇すると予測して順張りするのが一般的である。しかしながら、強化学習エージェントが獲得した戦略は、バンド幅の広がりとは無関係に価格がバンドの上方にあるときに順張りする戦略であると解釈できる。

3 強化学習を用いた金融市場取引戦略分析システム

3.1 システム構成

本研究で試作した取引戦略分析システムは、金融市場における取引戦略を強化学習を用いて獲得し、それを視覚化するためのものである。システム全体の外観を図4に示す。

本システムは、図5に示すように、次の3つのモジュールから構成されている。

1. GUI モジュール
2. 市場シミュレーター・モジュール
3. 強化学習エージェント・モジュール

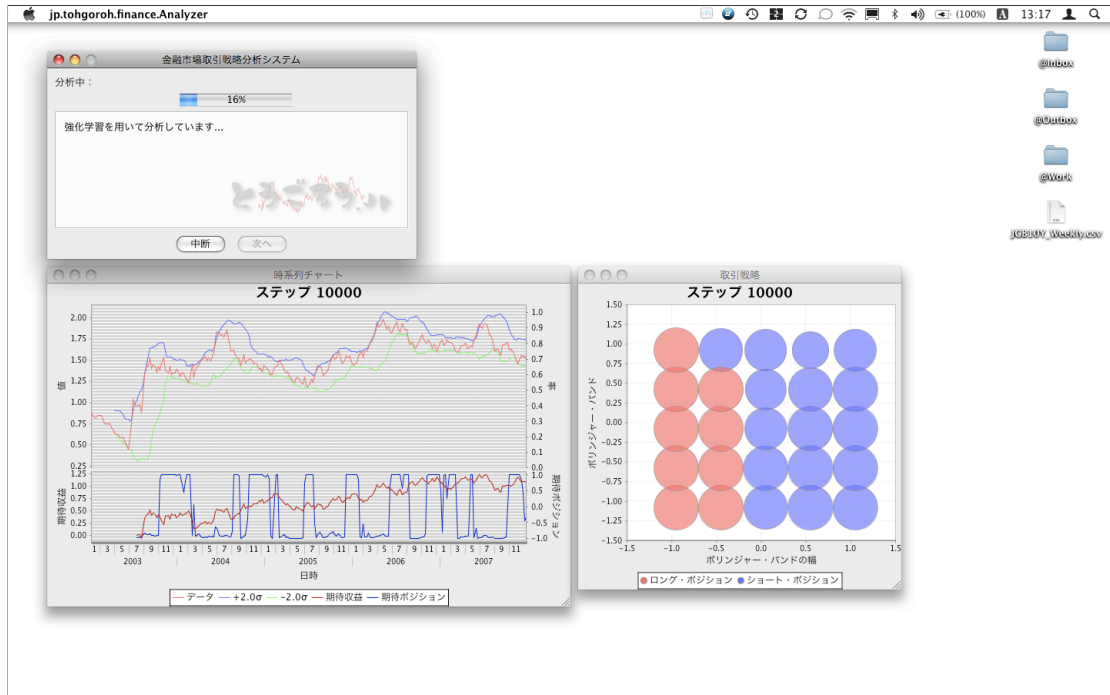


図4 Mac OS X 上での実行画面。

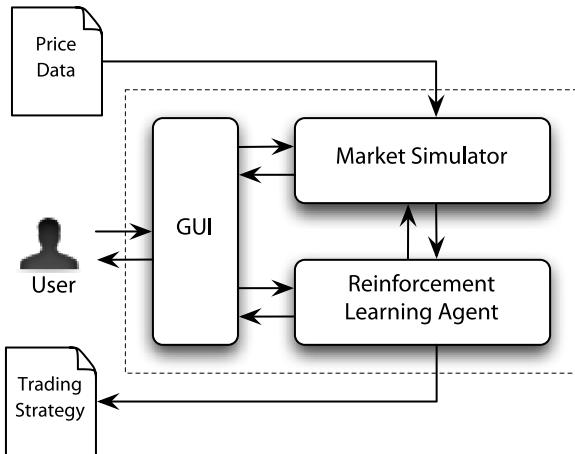


図5 システム構成図。

GUI モジュールは取引戦略分析システムの GUI を提供するためのものである。システムを操作するコントロール・パネル・ウィンドウ、分析対象の時系列チャートを表示する時系列チャート・ウィンドウと、獲得した取引戦略を表示する取引戦略ウィンドウの三つのウィンドウから成る。

市場シミュレーター・モジュールは価格（または金利）の時系列データが記録されたファイルを読み込み、エージェントからの取引注文に基づいて評価損益と実現損益を計算し、エージェントに伝える。また、その結果を

GUI モジュールを通してユーザに示す。

強化学習エージェント・モジュールは、市場シミュレーター・モジュールを環境として市場の状態を観測し、その観測に基づいて注文を決定し、その結果としての評価損益と実現損益を受け取る。これらを通して前節に説明した方法で強化学習を行い、取引戦略を学習する。

本システムは Java を用いて構築されている。したがって、プラットフォームを選ばず、Windows, Mac OS, Linux などでも実行できる。

3.2 取引戦略の可視化と評価

本システムは、獲得した取引戦略をバブル・チャートを用いて可視化する。取引戦略は、期待ポジション——すなわち、ショート・ポジションを取る確率とロング・ポジションを取る確率の和として、それぞれの特徴座標を中心としたバブル (円) によって表される。このとき、ロング・ポジションは赤色、ショート・ポジションは青色とし、半径がポジションの絶対値の大きさを表す。獲得した取引戦略を可視化した例を図 6 に示す。

獲得した取引戦略の評価には、この期待ポジションを用いる。ポジションの選択確率に基づいてランダムに行動を選択すると、一回のシミュレーションでは正確な評価ができない。そこで、それぞれの取引可能時刻において、期待ポジションを取るものとし、その時点での評価

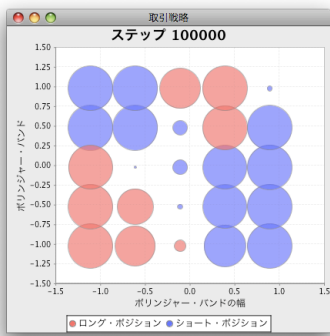


図6 獲得した取引戦略の可視化.

損益とその時点までの実現損益の和を期待損益とする。これにより、一回のシミュレーションで獲得した取引戦略の評価を可能としている。

4 実行例

本節では、日本国債を対象とした債券取引戦略の分析を例に、本システムの実行例を示す。

本システムを起動すると、コントロール・パネル・ウィンドウが現れる。このコントロール・パネルの画面遷移図を図7に示す。以下、画面遷移の流れに沿って本システムの動作を説明する。

最初は、図8に示すようなデータ・オープン・パネル(Data Open Panel)が表示される。ユーザーはファイル選択ダイアログを呼び出し、入力するデータ・ファイルを選択する。データ・ファイルは、一組の日時と価格(または金利)が一行に書かれたCSV形式である。

システムにファイルが読み込まれると、図9に示された時系列チャート・ウィンドウが新しく表示される。上のチャートは価格(または金利)とテクニカル指標を表し、下のチャートはエージェントの観測値を表している。

このとき、コントロール・パネルには図10に示されたテクニカル指標選択パネル(Technical Index Chooser Panel)が表示される。現在のシステムでは、ボリンジャー・バンド、ボリンジャー・バンドの幅、単純移動平均、単純移動平均乖離率の4種のテクニカル指標が選択できる。他のテクニカル指標については、必要に応じて順次追加する予定である。また、詳細設定ボタンを押すと、それぞれのテクニカル指標についてのパラメータを選択するテクニカル指標詳細設定パネル(Technical Index Preferences Panel)が表示される。例えば、ボリンジャー・バンドの場合、幅の広さと分析期間をこのパネルで変更できるようになっている。

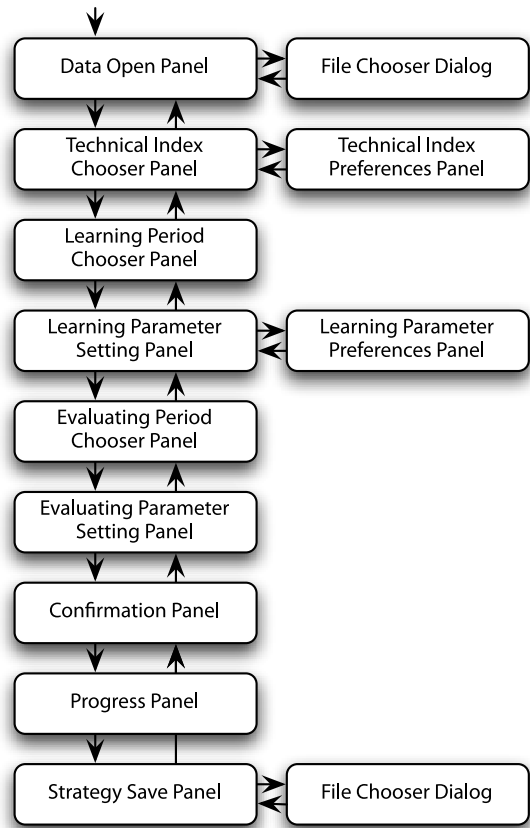


図7 画面遷移図.

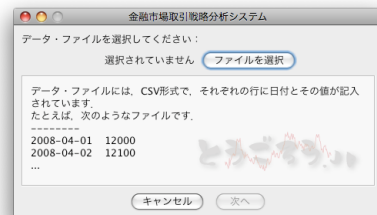


図8 データ・オープン・パネル.

使用するテクニカル指標を決定すると、コントロール・パネルには図11に示された学習期間選択パネル(Learning Period Chooser Panel)が表示される。ここでは、取引戦略の学習に使用するデータの期間を指定する。入力ファイルに記録されている期間からテクニカル指標の計算ができない期間を除いた期間から任意に選択することができる。

続いて、図12に示された学習パラメータ設定パネル(Learning Parameter Setting Panel)で強化学習のパラメータを設定する。ここで設定するのは、割引率パラ

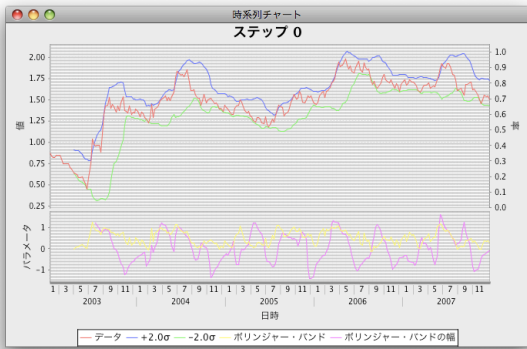


図9 時系列チャート・ウィンドウ.

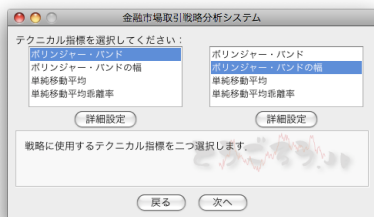


図10 テクニカル指標選択パネル.



図11 学習期間選択パネル.

メータと温度パラメータ,そして学習ステップ数である.

割引率パラメータは, OnPSにおける信用割当の割引率 [2] であり, この値を大きくすると過去の行動に対して報酬を多く分配し, 小さくすると過去の行動への報酬分配を小さくする働きを持つ. 温度パラメータは, Boltzmann 分布に基づくソフトマックス選択における温度を表し, 温度が大きいかほど一様選択に, 温度が小さいほどグリーディ選択に近づく. 学習ステップ数は強化学習を行うステップ数である.

学習パラメータ選択パネルの詳細設定ボタンを押すと, 学習パラメータ詳細設定パネル (Learning Parameter Preferences Panel) が表示される. ここでは, 状態

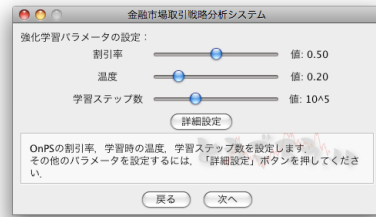


図12 学習パラメータ設定パネル.



図13 評価パラメータ設定パネル.

空間を線形近似 [5] する際の特徴数を設定できる.

この後, 学習期間選択パネルと同様の評価期間選択パネル (Evaluating Period Chooser Panel) で評価期間を選択する. 評価期間は学習期間とは独立に選択することができ, イン・サンプルだけでなくアウト・オブ・サンプルによる評価もできる. さらに, 図13に示された評価パラメータ設定パネル (Evaluating Parameter Setting Panel) において, 評価時の温度パラメータを (学習時の温度パラメータとは独立に) 設定できる.

最後に, 確認パネル (Confirmation Panel) で学習の開始を指示し, 強化学習を開始する. 強化学習を行っている間は, コントロール・パネル・ウィンドウに進行パネル (Progress Panel) が表示され, 図14のように, 時系列チャート・ウィンドウの下のチャートが獲得した戦略の評価結果となる. このチャートは, 前節で説明したように, 期待ポジションに基づいて取引したときの期待損益を表している.

強化学習エージェントが取引戦略を獲得すると, 取引戦略ウィンドウが現れ, 前節で説明したように取引戦略をバブル・チャート用いて表示する. これにより, ユーザーはエージェントが獲得した取引戦略を理解することができる. そして, コントロール・パネル・ウィンドウの取引戦略保存パネル (Strategy Save Panel) を通して獲得した取引戦略を保存する.

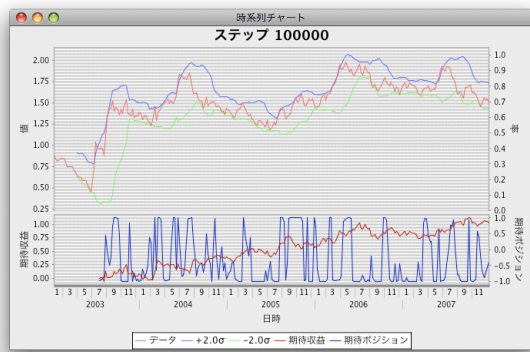


図 14 取引戦略の評価.

5 まとめ

本論文では、試作した強化学習を用いた金融市場取引戦略分析システムを紹介した。

本システムは、価格（または金利）の時系列データからテクニカル指標を計算し、その指標に基づいた取引戦略を強化学習を用いて獲得することができる。また、獲得した取引戦略を可視化することによって、市場の局面に応じた取引戦略の違いやテクニカル指標の有効性などを投資家が判断することができる。

強化学習を用いて金融市場取引戦略を学習する研究には、Sherstov らの PXS (Penn Exchange Simulator) を用いた人工市場の中での取引戦略を学習する研究 [4]、O らの強化学習を用いて銘柄と投資比率を決定する戦略を学習する研究 [3]、Lee らのマルチエージェント強化学習を用いてポートフォリオ・マネジメントを行う研究 [1] などがある。これらの研究では、取引戦略を獲得すること自体を目的としており、その取引戦略を分析することを目的としていない。したがって、これらの研究によって獲得された取引戦略がなぜ有効に働くのかを分析することは困難である。

一般に、複雑な取引戦略を獲得するには観測するパラメータの種類が多い方が良いが、獲得された取引戦略も複雑なものとなり理解することが難しくなってしまう。また、過学習の危険性が大きくなってしまふ。

本研究では、使用するテクニカル指標の数を二つに限定することやエージェントが観測する値を正規化することによって、獲得した取引戦略を可視化し、投資家が獲得した取引戦略を理解することを可能としている。また、同時に、過学習の危険性が小さくなり、より一般的な取引戦略が獲得されることが期待できる。

参考文献

- [1] J. W. Lee, J. Park, J. O, J. Lee, and E. Hong. A multiagent approach to Q-learning for daily stock trading. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, Vol. 37, No. 6, pp. 864–877, 2007.
- [2] T. Matsui, N. Inuzuka, and H. Seki. On-line profit sharing works efficiently. In *Proc. of KES-2003, Part 1, LNAI 2773*, pp. 317–324, 2003.
- [3] J. O, J. Lee, J. W. Lee, and B.-T. Zhang. Adaptive stock trading with dynamic asset allocation using reinforcement learning. *Information Science*, Vol. 176, pp. 2121–2147, 2006.
- [4] A. A. Sherstov and P. Stone. Three automated stock-trading agents: A comparative study. In *Proc. of AMEC-2004, LNAI 3435*, pp. 173–187, 2005.
- [5] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 1998. 三上貞芳, 皆川雅章 共訳. 強化学習. 森北出版, 2000.
- [6] カブロボ・コンテスト, 2004. <http://kaburobo.jp/>.
- [7] 和泉潔, 鳥海不二夫, 松井藤五郎. 人工市場とアルゴリズム・トレード. *日本知能情報ファジィ学会誌*, Vol. 21, No. 4, pp. 609–615, 2008.
- [8] 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*, Vol. 22, No. 4, pp. 540–547, 2007.
- [9] 松井藤五郎, 後藤卓, 和泉潔, 大和田勇人. 強化学習を用いた債券取引戦略の獲得. 2008 年度人工知能学会 (第 22 回) 全国大会, 2C3-1, 2008.
- [10] 松井藤五郎, 大和田勇人. 株式取引エージェントへの強化学習の応用. 2005 年度人工知能学会 (第 19 回) 全国大会, 1D4-1, 2005.
- [11] 松井藤五郎, 大和田勇人. 強化学習を用いた株式取引エージェントの評価. 2006 年度人工知能学会 (第 20 回) 全国大会, 3C1-6, 2006.
- [12] 松井藤五郎, 大和田勇人. 強化学習を用いた株式取引エージェントにおける汎用政策の学習. 2007 年度人工知能学会 (第 21 回) 全国大会, 3D9-5, 2007.