

強化学習エージェントを用いた人工市場の基礎的分析

A Basic Analysis of Artificial Market with Reinforcement Learning Agent

中西 匡弘¹ 橋本 文彦²

Masahiro Nakanishi¹, Fumihiko Hashimoto²

¹大阪市立大学 経済学部

¹Faculty of Economics, Osaka City University

²大阪市立大学大学院 経済学研究科

²Graduate School of Economics, Osaka City University

Abstract: In this paper, we propose a new simple artificial market model based on an idea of the U-Mart system. We designed a reinforcement learning agent which runs on this artificial market with other simple computer agents. We simulated the process that the reinforcement learning agent adapts to the price series of the market. This simulation clearly showed the emergence of market.

1. はじめに

本研究の目的は、可能な限り単純化された人工市場シミュレーションを通して、市場とエージェントの振る舞いを分析することである。

本研究ではこの目的を達成するため、単純な人工市場モデルとエージェントモデルを構築する。2節では市場モデルについて述べる。3節では本研究で用いるランダムエージェント、テクニカルエージェント、そして強化学習エージェントについて述べる。4節において市場とこれらのエージェントを用いたシミュレーションと分析を行う。

2. 人工市場モデル

2. 1 新たな人工市場がなぜ必要か

現在、既にU-Martという人工市場プラットフォームが存在しており、多くのシミュレーション・分析が行われている[1],[2]。U-Martを用いることによって自由度の高い人工市場シミュレーションを行うことが可能である。しかしながら、本研究では、こうした既存の人工市場プラットフォームを用いずに、新たな人工市場モデルを構築する。まずU-Martの概要について述べた後、U-Martの問題点を指摘する形式で、本研究が新たな人工市場モデルを必要とする理由を述べる。

U-Martは株価指数先物取引市場のシミュレーション環境である。U-Martにおけるエージェントは成行注文や指値注文によって株価指数先物の取引を行う。U-Martにおける市場はこれらの注文を集計

して市場価格を決定し、エージェントの注文を約定させる。U-Martを用いた人工市場シミュレーションは、こうしたエージェントによる注文と市場による価格決定・約定を繰り返すことで進行するものである。

これら市場・エージェントの構造についての詳細は省略するが、U-Martは現実の市場制度を高いレベルで反映している人工市場モデルである。その一方で、本研究ではU-Martはある問題点を有していると考えられる。問題点の中心は、U-Martにおける現物価格時系列データの扱いにある。U-Martは株価指数先物を取引対象とするが、このためには現物価格時系列データが必要となる。実際、U-Martを用いたシミュレーションを行う場合は、現物価格時系列データを予め用意しておく必要がある。

U-Martにおいて現物価格時系列データは、エージェントの参照できる情報の一つであり、従って市場で成立する先物価格に影響を与える。市場において成立する先物価格時系列データが現物価格時系列データによってどのような影響を受けたかこれを分離することは非常に困難である。エージェントの意思決定が複雑になる程、現物価格時系列データは先物価格に深く織り込まれていく。このため、シミュレーションによって得られた結果が、市場の制度的側面やエージェント間の相互作用によって得られたものだと確信を持って言えない。

また、U-Martにおけるエージェントは、シミュレーションを通して得られる現物価格時系列と先物価格時系列、板情報を参照した上で、現物価格がどのようなものになるか予想し、これを意思決定に反映させる必要がある。U-Martの構造は複雑な意思

決定をエージェントに要求しているのである。さらに、この現物価格時系列データはシミュレーションを通して変化することがないため、エージェントはこれに全く関与できず、単に予想することしかできない。しかし、現実の市場では現物価格と先物価格が互いに影響を与え合っていることが確認されている[3]。

これらに加えて、妥当な現物価格時系列データをどのように生成するかという問題も存在する¹。

こうした問題点を踏まえ、次節以降において新たな人工市場モデルを構築する。このモデルの理念は、市場の構造とエージェントの構造をできるだけ単純に保ちつつ、市場外部から与えられる情報を限りなく少なくすることである。

2. 2 シミュレーションの概要

まず、シミュレーション全体がどのように進行するのかについて述べる。本研究でのシミュレーションは下記のような順序で進行する。

- 1: シミュレーションの設定パラメーターを与える。
- 2: 全てのエージェントを初期化する。
- 3: 指定された回数、以下の処理を繰り返す。
 - 3-1: 指定された回数、以下の処理を繰り返す。
 - 3-1-1: 全てのエージェントが市場にどのような注文をだすか、あるいは注文ださないか決定する。
 - 3-1-2: 市場はエージェントの注文から市場価格を決定し、約定処理を行う。
 - 3-1-3: エージェントは約定処理の結果ポジションを得る。また、実現損益と未実現損益を計算する。
 - 3-2: 最終的な決済を行い、エピソードにおけるエージェントの最終収益を決定する。

本研究では、3-1-1 から 3-1-3 の処理を1つのステップとする。また、3-1 から 3-2 の処理を1つのエピソードとする。つまり、1つのエピソードは複数のステップと最終的な決済によって構成され、1つのシミュレーションは複数のエピソードによって構成される。

1 その他にも問題点は残されている。U-Mart では同一指値かつ同一期間の注文について、ランダムに優先順位を付けることで公平さを保っている。これによって市場そのものが不確実性を有することになる。またこの公平さを保つ仕組みそのものは小口注文によって回避できる[2]。

2. 3 市場とエージェント

本モデルにおいて取引されるモノは、価格になんの根拠も持たない。このモノは、エージェントの需給によってのみ価格が決定される。エージェントは意思決定の際に、既に成立した市場価格の時系列を参照することができる。これは、新たな市場価格が、市場の構造とエージェントの意思決定というフィルタを通した過去の市場価格時系列によって決定されるということでもある。

エージェントの資産は無限であるとする。また 2.2 節の 3-1-1 において、全てのエージェントが意思決定するまでシミュレーションは次の処理段階に進まないため、モデル上のエージェントは無限の時間を意思決定に費やすことできる。もちろん、実際のシミュレーションにおいては現実的な範囲で意思決定を行う必要がある。本モデルのエージェントは板情報の参照が不可能であるため、意思決定に費やす時間によって、エージェント及び市場は何の影響も受けない。

2. 4 注文

本研究のエージェントは、2.2 節で述べた 3-1-1 において、「買う」「売る」「なにもしない」の3種類の意思決定を行うものとする。空売りは無制限に可能であるとする。エージェントが「買う」および「売る」を選択した場合、市場に注文情報が送信される。この注文情報には、注文種類と共に、注文枚数が含まれている。注文枚数はどのエージェントも常に1であるとする。

また注文において取引手数料はかからないとする。また、1ステップにつき1注文しかだせず、注文が有効な期間は1ステップに限るものとする。

2. 5 価格決定メカニズム

本モデルの市場価格決定メカニズムは板寄せ方式に類似した独自の方式を用いている。まず、市場に集められた買い注文の合計枚数を q_b 、売り注文の合計枚数を q_s とする。また i 番目のステップにおける市場価格を p_i とすると、以下のようにして市場価格が決定される。

$$p_i = \begin{cases} p_{i-1} & (q_b = q_s \text{ or } q_b = 0 \text{ or } q_s = 0) \\ p_{i-1} + 1 & (q_b > q_s) \\ p_{i-1} - 1 & (q_b < q_s) \end{cases}$$

ただし初期価格 p_0 はシミュレーション開始時にパラメータとして与える。

この価格決定メカニズムを用いる場合、市場価格の変動は -1, 0, +1 の三種類しか存在しない。

この価格決定メカニズムには二つの利点が存在する。第一の利点は、エージェントが暗黙のうちに持

たなければならぬ価格の分散についての情報が必要なくなることにある。エージェントは次のステップの価格変動が1か10かあるいは1000になるか考慮しなくても良い。第二の利点として、価格変動の範囲を限定し、欠損値を存在させないようにすることでエージェントの意思決定の負担を単純化するという点を挙げることができる。より具体的には、後述する強化学習アルゴリズムにおける状態数削減に貢献している。

2. 6 約定

買い注文と売り注文のそれぞれの合計枚数が等しい場合、全ての注文は約定され、エージェントはポジションを持つことになる。それぞれの合計枚数が等しくない場合は部分約定を行う。例えば、エージェントAが買い注文1枚、エージェントBが売り注文1枚、エージェントCが売り注文1枚を市場に出した場合、エージェントAは1枚の買い持ちポジション、エージェントA,Bはそれぞれ0.5枚の売り持ちポジションを持つものとする。このように注文枚数は分割可能であるとする。

約定の結果、エージェントにポジションが発生し、かつこれまでに反対ポジションがある場合、それぞれのポジションを閉じてエージェントは収益を得る。また、エピソードの最終ステップにおいてその時点の市場価格で全てのポジションを解消する。このようにエージェントが収益を得るタイミングは二種類存在する。

3. エージェントモデル

3. 1 ランダムエージェント

本研究におけるランダムエージェントは確率 $1/2$ で「買い」あるいは「売り」の注文を選択する。

3. 2 テクニカルエージェント

本研究におけるテクニカルエージェントは、過去の価格変動が三連続して上昇ならば「売り注文」、三連続で下落ならば「買い注文」をだす。その他の価格変動パターンの場合、ランダムエージェントのように、確率 $1/2$ で「買い注文」、また「売り注文」をだす。

3. 3 強化学習エージェント

本節では強化学習エージェントについて述べる[4]。3.1節、3.2節で述べたランダムエージェントおよびテクニカルエージェントは、シミュレーションを通して自らの振る舞いを変化させない。こうした静的なエージェントに対して、強

化学習エージェントは、試行錯誤を繰り返しながら、自らの収益を最大化させるように動的に自らの振る舞いを変化させるものである。

強化学習には多様なアプローチが存在する。本研究では、方策オン型のTD(0)法による方策評価と ϵ -Greedy 法による方策改善を組み合わせた単純な価値関数ベースの強化学習エージェントを採用した。この強化学習エージェントは、ステップ毎に環境に属する状態を知覚し、これに応じて行動し、その結果得られる報酬を観測し、価値関数を推定する。この処理を繰り返すことで試行錯誤的に総報酬を最大化することを目的とする。従って、強化学習エージェントを設計するにあたり、本モデルにおける「環境」「状態」「報酬」「行動」を定義する必要がある。

3. 3. 1 環境・状態の定義

強化学習エージェントは、エージェント外部の環境との相互作用によって学習を進める。本モデルでは、他のエージェントや市場に加えて、エージェント自らのポジションについての情報も環境に含まれる²。本モデルでは、環境に属する状態は以下の情報から構成されるとする³。

- 直近3の市場価格変動の時系列データ
- 自らのポジションについての情報 (売り持ち, 買い持ち, ポジション無し)
- ポジションを市場価格で評価した未実現損益についての情報 (正, 負, ゼロ)

3. 3. 2 行動の定義

強化学習エージェントが選択できる行動は「買う」「売る」「何もしない」の三種類いずれかである。

3. 3. 3 報酬の定義

本研究における報酬は「2.6節において述べた収益」と同一であるとする。 r_{t+1} を t 番目のステップの後に得られる報酬とする。また最終ステップを T とする。本モデルにおける強化学習エージェントはこの報酬を足しあわせた以下に示す収益

- 2 エージェントは、自らのポジションを完全に制御することができない。そのため、本モデルにおける「エージェント」と「市場」の境界と、強化学習の枠組みでいう「エージェント」と「環境」の境界は異なる[4]。
- 3 一つの価格変動は3パターンであり、これを直近3までみると価格変動の情報には $3^3=27$ パターンとなる。またポジション・未実現損益についての情報はそれぞれ3パターンとなる。本モデルの状態はこれらの組み合わせで表現されるため、 3^5 のパターンを持つ。

$$R_t = \sum_{i=0}^T r_{t+i+1}$$

の期待値を最大化することを目的とする。

3. 3. 4 学習の流れ

状態 s において、行動 a を選択する確率を方策と呼び π で表す。方策 π に従い、状態 s において行動 a を選択することによって得られる期待収益を行動価値関数として以下のように表す⁴。

$$Q^\pi(s, a) = E_\pi\{R_t \mid s_t = s, a_t = a\} = E_\pi\left\{\sum_{i=0}^T r_{t+i+1} \mid s_t = s, a_t = a\right\}$$

強化学習エージェントは、ステップ t における状態 s_t の下で行動 a_t を選択する。その後、得られる報酬 r_{t+1} と遷移した状態 s_{t+1} を観測し、次のステップの行動 a_{t+1} を選択する。そして、これら $s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1}$ を用いて、次の規則により行動価値関数を更新する⁵。ただし α は学習率とする。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

本研究における強化学習エージェントが状態 s において選択する行動は、 ϵ -Greedy 法によって決定される。すなわち、確率 $1-\epsilon$ で最も高い行動価値を示す行動を選択し、確率 ϵ でランダムに行動を選択する。

強化学習エージェントは、以上に述べたような行動価値関数の更新と行動選択をステップ毎に繰り返すことによって、期待収益を最大化させる方策を学習する。

4. シミュレーション

4. 1 シミュレーション設定

本研究では、以下の3パターンからなるエージェントの構成を用いてシミュレーションを行う。

- (a) ランダムエージェント: 10体
- (b) ランダムエージェント: 10体
強化学習エージェント: 1体
- (c) ランダムエージェント: 5体
テクニカルエージェント: 5体
強化学習エージェント: 1体

4 本研究の強化学習エージェントの行動価値関数は単純なテーブルとして表現している。このテーブルは状態のパターン 3^5 と行動のパターン 3 の組み合わせについて、行動価値を保持する。関数近似などの手法は用いていない。

5 ここでは更新の際に割引を用いていない。

全ての構成について、シミュレーションのエピソード数は 300、ステップ数は 500 とした。また、(b) および(c)の構成で用いる強化学習エージェントのパラメーターは、学習率 $\alpha=0.2$ 、 ϵ -Greedy 法のパラメーター $\epsilon=0.05$ とした。

4. 2 結果

4. 2. 1 構成(a)

構成(a)は全てのエージェントがランダムエージェントである。

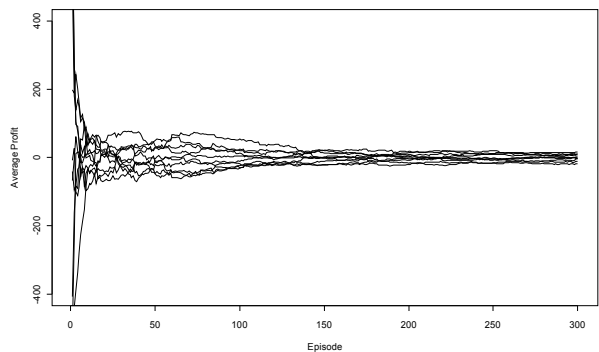


図 1 ランダムエージェントの平均収益

まず、エージェントの側面からシミュレーション結果を述べる。図 1 は、ランダムエージェントが最終的に得たエピソード毎の平均収益の推移である。ランダムエージェントの平均収益は 0 に収束していくことが分かる。

次に、市場において成立する価格時系列の側面からシミュレーション結果を述べる。

まず、分析手法について述べる。本構成の市場価格変動の分布はエージェントの総数に影響を受ける。2.5 節で述べたように、本市場モデルでは、買い注文の枚数と売り注文の枚数が等しい場合、また買い注文、売り注文の枚数のいずれかがゼロの場合、市場価格は変動しない。ランダムエージェントの数が多くなればこうした価格変動がゼロとなる場合は少なくなる。そこで、この影響を無視するために市場価格変動から上昇(+1)、下降(-1)のみを抽出する。その結果 1 と -1 からなる二進数列を得ることができる。

ランダムエージェントにより構成される市場から得られる二進数列はランダムな数列になるだろう。シミュレーション結果から得られた二進数列がランダムか否か次の三つの基準を用いて判断する[5]⁶。

6 ランダムなエージェントは擬似乱数を用いて行動を決定しているため、この二進数列もまた擬似乱数の一つであると考えられる。こ

数列の長さを T , t 番目の価格変動を X_t とする.

1. T が大きくなれば, $X_t=1$ となる回数と, $X_t=-1$ となる回数がほぼ等しい.
2. ある系列 $\langle X_t \rangle$ において, 1が n 個連続して, その両側が-1で挟まれている部分を長さ n の「1の連」と呼ぶ. $run(n, m)$ を系列 $\langle X_t \rangle$ に含まれる長さ n の「 m の連」の個数とすると, $\lim_{T \rightarrow \infty} \frac{run(n+1, m)}{run(n, m)} = \frac{1}{2}$ となる.
3. 系列 $\langle X_t \rangle$ の自己相関関数は次のようになる.

$$R(s) = \lim_{T \rightarrow \infty} \frac{1}{T-s} \sum_{i=1}^{T-s} X_i X_{i+s} = \begin{cases} 1, (s=0) \\ 0, (s \neq 1) \end{cases}$$

構成(a)のシミュレーションから得られた価格変動の二進数列は, これらの基準に照らし合わせてランダムなものであった. 次節以降で用いるため, ここでは特に基準2の連を用いた分析を行う.

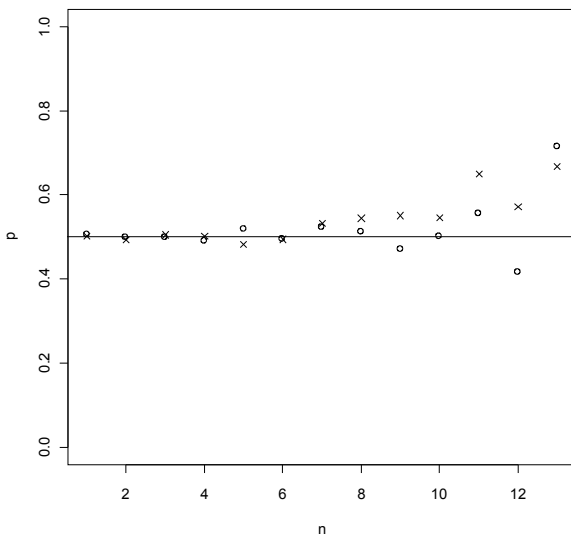


図 2 連の分析. ○としてプロットされた点は1の連, ×は-1の連を表す.

図2の横軸は連の長さ, 縦軸は基準2を参考に求めた $run(n, m) / \sum_{n=1}^{\infty} run(n, m)$ である. この図において, プロットされた点の縦軸の値が0.5より大きければ, 長さ n の連が出た後に, ランダムな場合に

ここに挙げた三つの基準は, M系列と呼ばれる擬似乱数が満たす性質である[5]. この基準を全て満たしたとしても, 真にランダムな系列であるとはいえないことに注意する必要がある.

期待されるよりも多く価格が反転したことを意味する. 反対に0.5より小さければ, 価格が一定方向に伸びたことを意味する.

連の長さ n が大きくなれば連そのものの個数が減るためにばらつきが大きくなるが, 概ね0.5周辺に固まっている. すなわち, 価格は反転しやすい傾向があるわけでも, また一定方向にトレンドを持つわけでもない.

以上より, エージェントと市場の両面から見て, このシミュレーション結果はランダムなものであるといえる.

4. 2. 2 構成(b)

構成(b)は構成(a)に強化学習エージェントを1体加えたものである.

2.6節で述べたように, 本モデルのエージェントが得る収益には二つの種類がある. まず一つは, エピソード中において反対売買を行い, ポジションを解消することによって得られる収益 $\sum_{i=0}^{T-1} r_{i+1}$ である. もう一つはエピソードの最終ステップにおいて強制的にポジションが解消されることにより得られる収益 r_{T+1} である. ただし r_{i+1} を t 番目のステップの後に得られる収益, また最終ステップを T とする.

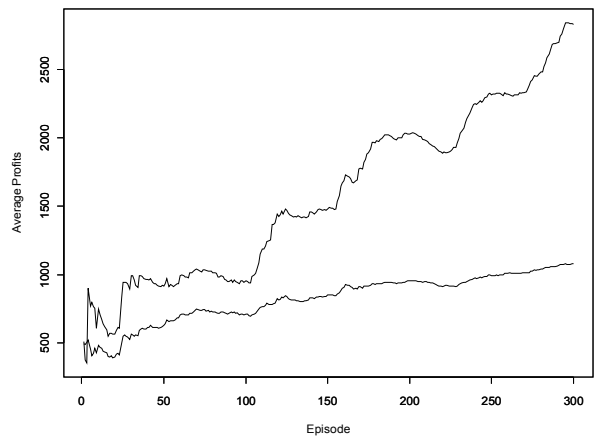


図 3 強化学習エージェントの平均収益

図3において上部に位置する系列はこれら二種類の収益を足した, $\sum_{i=0}^{T-1} r_{i+1}$ つまり「最終的に強化学習エージェントが得た全収益」の平均である. 図の下部に位置する系列は, 強化学習エージェントの得た収益のうち $\sum_{i=0}^{T-1} r_{i+1}$, すなわち「エピソード中の反対売買によって得られた収益」の平均である.

まず、図の上部の系列を見ることにより、強化学習エージェントはエピソードが進むに連れてより大きな収益を得ていることが分かる。

これにあわせて、図の下部の系列を見る。強化学習エージェントの収益全体 $\sum_{i=0}^T r_{i+1}$ に占める

r_{T+1} の割合は、エピソードが進むに連れてより拡大していることが分かる。この結果からエージェントは反対売買を繰り返し収益を確定させるよりも、最終ステップにおける収益を最大化するように市場を一定の方向に制御しようとしているのではないかと推測できる。市場価格時系列の分析によってこの推測が正しいか確かめる。

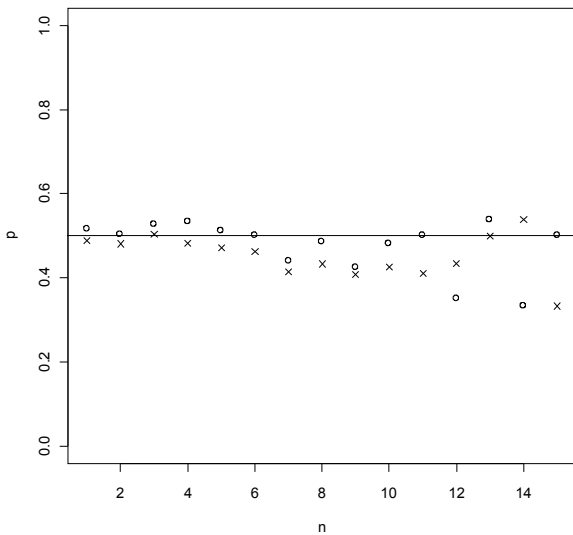


図 4 連の分析

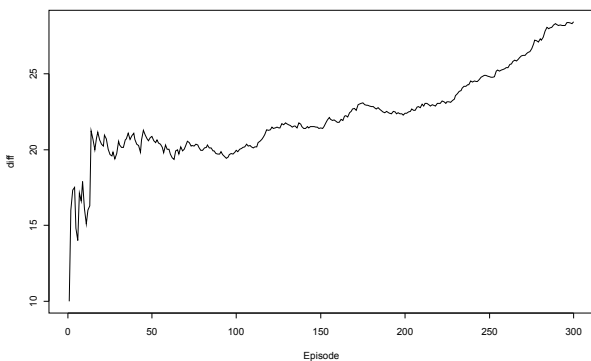


図 5 価格差の絶対値の平均

図 4 は 4.2.1 節で取り上げた連の分析によるものである。特に長さ 6 以上の連が出やすい傾向にあるように見えるがはっきりとしない。そこで、図 5 において、エピソード毎の初期価格と最終価格の差の

絶対値を平均したものを示す。これを見ると明らかであるが、エピソードが進むにつれて、初期価格からより離れるように、すなわち一定のトレンドを持つように価格が推移することが分かる。

本構成における強化学習エージェントは反対売買によって収益を確定すると共に、最終ステップにおいて得られる収益を拡大させるように市場を一定の方向に動かしている。そして、学習が進むにつれて後者の市場制御による利益をより拡大させていくのである。

市場全体の取引数量に占める強化学習エージェントのそれは相対的に小さなものである。しかし、他のエージェントが全てランダムエージェントである場合、強化学習エージェントは確率的に市場価格を操作できる。

4. 2. 3 構成(c)

構成(c)は、構成(b)のランダムエージェントのうちの半数をテクニカルエージェントに変えたものである。

まず、4.2.2 節と同じように、強化学習エージェントの平均収益とエピソード中に確定された平均収益の系列を図 6 に示す。構成(b)と同様、構成(c)の場合もまた、強化学習エージェントは学習を進めることによってより収益を得ていることが分かる。ただし構成(b)と比較してやや収益は低い。

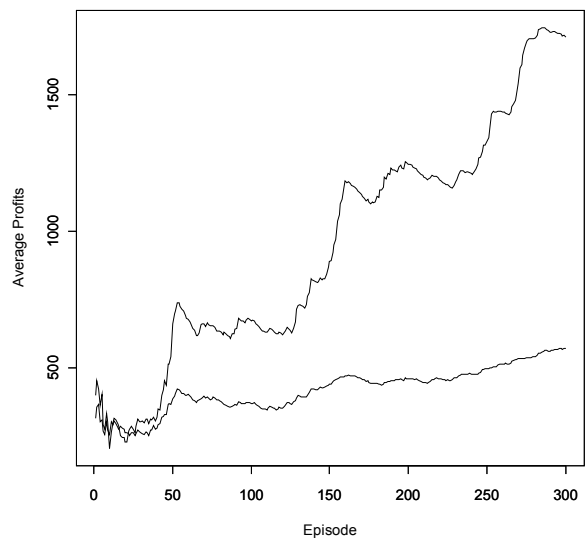


図 6 強化学習エージェントの平均収益

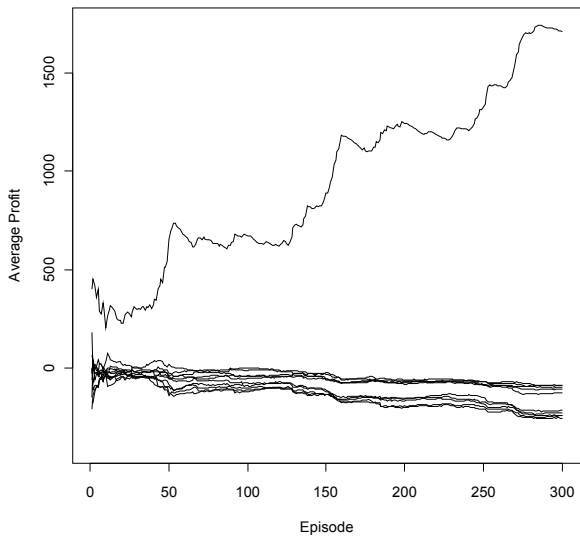


図 7 エージェント毎にみた平均収益

図 7 は、エージェント毎の平均最終収益を表している。強化学習エージェントの平均収益は図上部の系列で示されており、もっと高い平均収益を得ていることが分かる。図の下部では、ランダムエージェント及びテクニカルエージェントの平均収益が二群に分かれている。このうち損失の大きなグループはテクニカルエージェント群であり、損失の小さなグループはランダムエージェント群である。強化学習エージェントの学習が進むにつれて、テクニカルエージェントの損失はランダムのエージェントの損失以上に拡大していく。次に市場価格時系列を見ることで、この問題をより詳しく見たい。

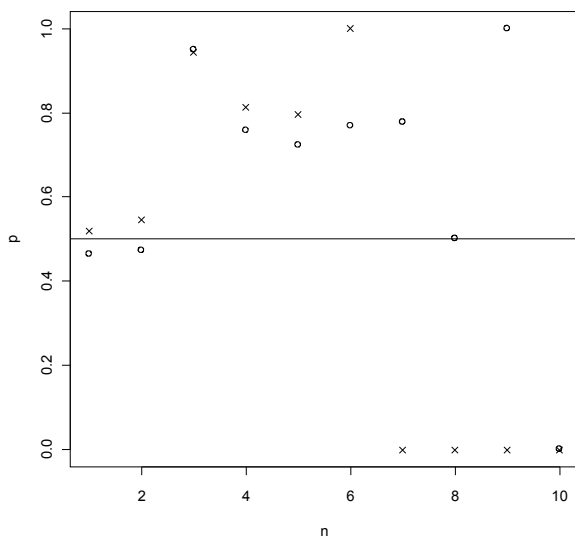


図 8 連の分析

図 8 はこれまでに述べた連の分析を示すものである。これを見ると分かるように、長さ 3 以上の連において価格上昇及び下降は反転しやすくなる。これはテクニカルエージェントの影響であるといえる。テクニカルエージェントの取引数量が市場全体に占める割合はほぼ半分であり、価格変動が三連続して同じであった場合、次のステップにおいて成立する価格はそれまでと高い確率で反転する。

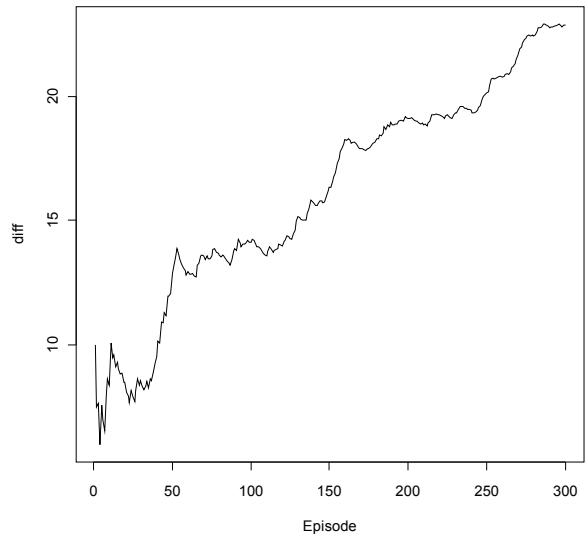


図 9 価格差の絶対値の平均

図 9 は 4.2.2 節においても述べた初期市場価格と最終市場価格の差の絶対値の平均の推移である。この図から、エピソードが進むに連れて初期価格と最終価格の差が拡大していることが分かる。これは強化学習エージェントの影響である。エピソードを通して振る舞いを変化させ、市場に影響を与えることができるエージェントは強化学習エージェントしか存在しない。また、図 5 と比較すると初期価格と最終価格の差が構成 (b) に比べて抑制されていることが分かる。これは先に述べたテクニカルエージェントの影響である。

すなわち、構成 (c) の強化学習エージェントは構成 (b) の場合と同様に、反対売買によって収益を確定しつつ、市場価格を一定の方向に動かすことで最終的に得る収益を最大化しているといえる。ただし構成 (b) の場合とは異なり、テクニカルエージェント群はこうした価格操作の動きに反応し、これを抑制しようと働く。その結果、最終的に強化学習エージェントが得る収益は構成 (b) と比較して小さくなる。ここで図 7 をもう一度参照すると、テクニカルエージェント群は三種類のエージェントの中で最も損失の大きく、またこの損失は強化学習の収益に転

じている。

結論として、テクニカルエージェントの振る舞いは強化学習エージェントの収益を全体的に低下させると共に、強化学習の収益に貢献するといえる。テクニカルエージェントの特定のパターンに沿う行動は、市場価格の初期値からの乖離を抑制し、強化学習エージェントの収益拡大を妨害する。その反面、特定の時系列パターンに沿うという特徴を（価格時系列を通して）強化学習エージェントに学習されることにより、テクニカルエージェント自身の収益は下がってしまうと考えられる。

5. まとめ

構成(a)のシミュレーションを通して、ランダムエージェントのみが参加する市場の価格時系列変動はランダムになること、またランダムエージェントの平均収益は0に収束することを確認した。

構成(b)のシミュレーションを通して、市場全体から見て小さな取引数量しかない強化学習エージェントが市場全体を制御し、その結果収益を上げることを示した。

構成(c)のシミュレーションでは、特定の価格変動パターンに反応するテクニカルエージェントを投入した。この場合においても、強化学習エージェントは収益を上げることを示した。

構成(b)、構成(c)の場合共に、強化学習エージェントは平均的に収益を改善し続けており、またその様子も類似している。しかし、その意味は異なっている。構成(b)の市場において成立する価格は強化学習エージェントの強い影響下にある。強化学習エージェントは自身の影響下にある価格時系列を参照し学習することで収益を上げている。これに対して、構成(c)の市場において成立する価格は強化学習エージェントのみならず、テクニカルエージェントの影響も受けている。このテクニカルエージェントの影響によって、価格時系列の初期値からの乖離はある程度抑制され、強化学習エージェントの平均収益は低下する。しかしながら、この振る舞いによってテクニカルエージェントの損失は拡大し、この損失は強化学習エージェントの収益のうち多くを占めることになる。

このようにテクニカルエージェント群が市場価格に影響を与え、強化学習エージェントの収益水準を低下させながら、なおかつ強化学習エージェントの得る収益に寄与するということが同時に成り立つ。当然ながら、こうした現象は強化学習エージェントのみ、あるいはテクニカルエージェントのみのシミュレーションでは観測できない。

ユレーションでは観測できない。

本研究で用いた市場モデルは非常に単純である。またエージェントについても、時系列を参照しないランダムエージェント、時系列を参照するが静的な行動しかとれないテクニカルエージェント、時系列を参照し動的に振る舞いを変化できる強化学習エージェントの三種類しか存在しない。にもかかわらず、エージェントの組み合わせによって複雑な結果が生じうる。この複雑な結果の要因を探るにあたって本研究における市場モデルの単純性は、一定の役割を果たしたといえるのではないだろうか。

とはいえ、多くの問題は残されている。本研究のモデルは、単純性と引き換えに現実から乖離したものになった。本研究における結果はそもそも市場において成立する価格の根拠が存在しないからこそ得られた結果である。また、本研究の強化学習エージェントの振る舞いは、パラメータを調整した結果、得られたものである。 ϵ -Greedy法のパラメータが非常に小さいと「なにもしない」という行動に陥り、なかなか学習が進まない。かといって、パラメータを上げ過ぎると、結果は不安定になってしまう。

市場とエージェントの両面において考慮すべき点は数多く残されており、これらは今後の課題とする。

参考文献

- [1] 塩沢由典・中島義裕・松井啓之・小山友介・谷口和久・橋本文彦. 知的エージェントで見る社会3 人工市場で学ぶマーケットメカニズム——U-Mart 経済学編. 共立出版, 2006.
- [2] 喜多一・森直樹・小野功・佐藤浩・小山友介・秋元圭人. 人工市場で学ぶマーケットメカニズム——U-Mart 工学編. 共立出版, 2009.
- [3] 宇野淳. 価格はなぜ動くのか——金融マーケットの謎を解き明かす. 日経BP社, 2008.
- [4] Richard S. Sutton and Andrew G. Barto (三上貞芳・皆川雅賞). 強化学習. 森北出版, 2008.
- [5] 伏見正則. 乱数. 東京大学出版会, 2004.