

人工市場アプローチにおける強化学習を用いた介入政策の分析

松井 宏樹[†] 永田 裕一[†] 東条 敏[†]

[†] 北陸先端科学技術大学院大学情報科学研究科 〒923-1292 石川県能美郡辰口町旭台 1-1

E-mail: †{hmatui,nagatay,tojo}@jaist.ac.jp

あらまし 本研究では、人工外国為替市場モデルに現実の政府や中央銀行にあたる介入エージェントを組み込み、強化学習を用いてレートの安定に有効な介入政策シナリオを獲得する手法を提案する。最初に和泉らの作成した人工市場モデル AGEDASI TOF をもとに情報交換の仕組みを改良し、より現実的なモデルを構築した。次に介入エージェントをそのモデルに組み込むことで市場の状態に応じた介入を可能にし、強化学習によるレートの安定に有効な介入政策シナリオの獲得を目指した。

キーワード 人工市場, 複雑系, 強化学習, マルチエージェント

An Analysis of Intervention Policies Using Reinforcement Learning in an Artificial Market Approach

Hiroki MATSUI[†], Yuichi NAGATA[†], and Satoshi TOJO[†]

[†] School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai,
Tatsunokuchi-machi, Nomi-gun, Ishikawa, 923-1292 Japan

E-mail: †{hmatui,nagatay,tojo}@jaist.ac.jp

Abstract In this study, we propose the system which acquires effective intervention policies.

Izumi et al. proposed the artificial market model, a multiagent model for a foreign exchange market, named AGEDASI TOF. And then they construct the support system for the government decides exchange rate scenarios. However in the system, the way of intervention: when the government intervenes and quantity of intervention, needs to be determined before simulations. Namely it is intervention without considering the market situation.

First, we improve the way of agents' updating weights of information for forecasting the exchange rate. Second, we incorporate the intervening agent into the model. The intervening agent, which is as the government or the central bank in actual markets, intervenes depending on market situation dynamically and tries to acquire effective intervention policies using reinforcement learning.

Key words artificial market, complex systems, reinforcement learning, multiagent systems

1. はじめに

近年、バブルなどの激しく複雑な変動が市場でいくつも観測されている。これらの変化は個人の思惑を大きく上回り、ときとして市場を混乱に陥れてきた。

このような現象のメカニズムの解析を目的として、和泉らはエージェント群に為替レートの変動要因となる市場条件（景気、物価、金利、介入など）を入力し、それぞれの売買行動から為替レートの推移を得る人工市場モデル AGEDASI TOF^(注1) を

作成し、それを用いた為替政策の意思決定支援システムの構築（為替シナリオの分析）を行っている [1]。

しかし、AGEDASI TOF を用いた為替シナリオの分析には次のような問題点があると考えられる。1つはエージェント間の情報交換についてである。AGEDASI TOF では遺伝的アルゴリズムに基づいた情報交換が行われているが、マイクロレベルすなわちエージェントレベルで見るとレートの予測（適応度）に無関係な変動要因に対する重みづけまでコピーする、あるいは重みづけを交換するなどという不自然な点が見られる。2つめは「介入」についてである。AGEDASI TOF では介入は予想材料の1つにすぎず、市場の需給への直接の影響は全くない。また、予想材料の介入の項目はあらかじめ決めておいた値を入

(注1) : A Genetic-algorithmic Double Auction Simulation in TOKYO Foreign exchange market

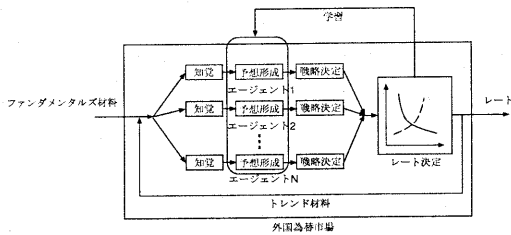


図1 AGEDASI TOF のフレームワーク
Fig.1 The framework of AGEDASI TOF.

力しているため、シミュレーションパスが予想と全く違っても介入が行われる期間、介入の大きさは同じである。つまり、市場の状態を全く考慮していない介入と言える。

本研究では、まず情報交換の仕組みを改良し、より現実的なモデルを構築、評価を行う。次にそのモデルに現実世界の政府、中央銀行に相当する介入エージェントを実装することで市場の状態に応じた動的な介入を可能とする市場モデルを構築する。介入エージェントには強化学習を用いた学習機構を持たせることでレートの安定に有効な介入政策シナリオの獲得を目指す。外国為替市場ではマイクロレベル（市場参加者）での相互作用の結果、マクロレベル（レートなど）においてマイクロレベルに直接デザインされていないパターンが創発する。本研究ではこのようなマクロレベルの市場の動向を介入というマイクロレベルの行動で制御することを学習することが目的である。

2. 人工市場モデル AGEDASI TOF

AGEDASI TOF は、マルチエージェントシステムを用いた円、ドルを扱う人工外国為替市場モデルである。AGEDASI TOF では100人のエージェントが参加し、レート決定が1回行われる1期間は以下の5つのステップからなる。また、1期間は現実世界の1週間に対応する。

2.1 知覚ステップ

知覚ステップでは各エージェントが為替レートの変動要因である予想材料を知覚する。入力されるデータは景気、物価、金利、介入など表1の17項目の変動要因を7段階（±3, ±2, ±1, 0）にコーディングしたものである。

2.2 予想形成ステップ

予想形成は、各エージェントがレート（対数）の変動値 $\Delta S(t)$ に対する予測 $E_i[\Delta S(t)]$ を求めることで行われる。各エージェントはそれぞれ独自の市場観を持っており、それは17種類の予想材料に対する重みづけで表される。各エージェントの予測レート変動値は、各予想材料とその予想材料に対する重みづけの積の和として定義される。

表1 予想材料
Table 1 Information for forecasting the exchange rate.

予想材料	もとなる生データ
1 景気	[米][日] GDP etc.
2 物価	[米][日] 消費者物価指数 etc.
3 金利	[米][日] 公定歩合、長期金利
4 マネーサプライ	[米][日] マネーサプライ
5 貿易収支	[米][日] 貿易収支
6 雇用	[米] 失業率 etc.
7 個人消費	[米] 小売売上、個人所得
8 介入	[米][日] 介入
9 要人発言	[米][日] 中銀総裁の発言 etc.
10 マルク	ドル/マルク、円/マルク
11 石油	石油価格
12 政治	政情、国際的な事件
13 株	[米][日] 株価
14 債券	[米][日] 債券価格
15 短期トレンド1	先週の変動値 ($\Delta S(t-1)$)
16 短期トレンド2	変動の変動値 ($\Delta S(t-1) - \Delta S(t-2)$)
17 長期トレンド	5週間の変動 ($S(t-1) - S(t-6)$)

$$E_i[\Delta S(t)] \equiv \alpha \left(\sum_{k=1}^n w_i^k(t) x^k(t) \right) \quad (1)$$

$$\left(\begin{array}{l} S(t): \text{期間 } t \text{ の為替レート (対数)} \\ E_i[\Delta S(t)]: \text{エージェント } i \text{ の } \Delta S(t) \text{ 予測値} \\ x^k(t): \text{期間 } t \text{ の変動要因 } k \text{ の値} \\ w_i^k(t): \text{エージェント } i \text{ の変動要因 } k \text{ に} \\ \text{対する重みづけ} \\ (\pm 3, \pm 1, \pm 0.5, \pm 0.1, 0 \text{ の } 9 \text{ 段階}) \\ \alpha: \text{スケール係数} \end{array} \right)$$

2.3 戦略決定ステップ

戦略決定ステップでは各エージェントが各自の予想に基づき期待収益を最大にする最適ドル資産保有高を計算し、ドルの売買要求量を決定する。

各エージェントの売買戦略は、期間 t における最適ドル資産保有高が期間 $t-1$ のドル資産保有高よりも大きく（小さく）、かつレートが自分の予想したレートよりも安く（高く）有利な場合には、その差の分だけドルを買って（売って）保有高を最適値に近づけようとする。

2.4 レート決定ステップ

各エージェントの売買戦略を市場全体で集積して、需要と供給が均衡するような値にモデルの今期のレートを決定する。

2.5 学習ステップ

学習ステップでは simple GA [2] をもとにした遺伝的アルゴリズムを用いて各エージェントは情報交換、学習を行い、予想材料に対する重みづけの列を変化させる。学習ステップで行われるオペレーションは淘汰、交叉、突然変異の3つである。

a) 淘汰

まず、エージェント i の期間 t における適応度 $F_i(t)$ を以

(注2) 厳密には、エージェント i の期間 t にとった予想方式（予想材料に対する重みづけの列）の適応度

下のようにレートの子測と実際のレートの子れをもとに定義する。

$$F_i(t) = -|E_i[\Delta S(t)] - \Delta S(t)| \quad (2)$$

確率 G (Generation Gap) で選ばれたエージェント i (全エージェント数 $N \times G$ 人) は、適応度 $F_j(t)$ に比例する確率で選ばれた相手 j から重みづけの列をコピーする。

つまり、各エージェントは自分の予想が正確であれば変更せず、予想が不正確であれば他のエージェントが持つ適応度の高い重みづけの列に入れ替えるのである。

これは予測に失敗したディーラーが成功したディーラーと情報交換を行うことによって自らの予想方式を変更することにあたる。

b) 交 叉

ランダムに全エージェントをペアにし、それぞれのペアの重みづけの列に対し、確率 P_{cross} で一点交叉を行う。

これはディーラー間のコミュニケーションによりお互いの意見を交換しあい、その結果予測方針を変更することにあたる。

c) 突然変異

エージェント i がある重みづけ w_i^k を、確率 $P_{mutation}$ でランダムに変化させる。

これはディーラーがある変動要因に対して、思いつきで新たな価値を設定することにあたる。

3. 情報交換の仕組みの改良

人工市場モデルを用いてシミュレーションを行うことのメリットとしてマイクロレベルすなわちエージェントの状態、行動を分析できることがあげられる。しかし、AGEDASI TOF ではエージェントの学習に simple GA [2] を用いているためにその行動が不自然になってしまっている。そこで本研究ではまず以下の情報交換の仕組みの改良を行うことで、その問題を改善することを試みた。

3.1 提案モデル

a) 交 叉

AGEDASI TOF の交叉にあたるものは行わない。すでに予測の成功、失敗が決まっている段階で、ディーラーが同等に重みづけを交換することは考えにくいからである。

b) 淘 汰

AGEDASI TOF における淘汰の操作は適応度の低いエージェントが適応度の高いエージェントの予想材料に対する重みづけをすべてコピーするというものであった。適応度を高める(予

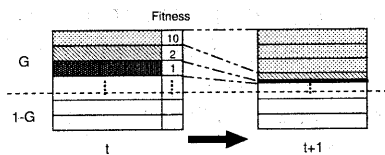


図2 淘汰

Fig. 2 The selection.

予想材料 $x^k(t)$	1	-3	0	...	2	0	
agent i $w_i^k(t)$	3	-0.1	0.5	...	0	-3	適応度:高
agent j $w_j^k(t)$	0	1	3	...	1	0.5	適応度:低

↓

予想材料 $x^k(t)$	1	-3	0	...	2	0	
agent i $w_i^k(t+1)$	3	-0.1	0.5	...	0	-3	
agent j $w_j^k(t+1)$	3	-0.1	3	...	0	0.5	

図3 コピーする対象を限定した淘汰の例

Fig. 3 Examples of the limited selection.

測を誤った原因を修正する) 為に情報交換するのであれば、適応度に関係な重みづけをコピーするのは不自然である。そこで本研究では、入力データ $x^k(t) = 0$ に対する重みづけ $w_i^k(t)$ は、式 (1), (2) より適応度に関係ないことに着目し、すべての重みづけをコピーするのではなく $x^k(t) (\neq 0)$ に対する重みづけだけ (間違った原因だけ) コピーの対象とする。

図3に適応度の低いエージェント j が適応度の高いエージェント i から重みづけをコピーする例を示す。この例で、3つめの予想材料と最後の予想材料は0でありレート予想に関係ないため、エージェント j はそれらに対する重みづけに関してはコピーを行わず、自分の認識を変えていない。この操作では、適応度に関係ない重みづけをむやみにコピーしないことで各エージェントが個性を失うことを防いでいるとも言える。

3.2 提案モデルの評価

提案モデルによる情報交換の仕組みの有効性の評価を行う。

1986年1月から1998年12月までの現実の予想材料とレートの子データを用いた予測誤差の測定を行った。初期化したディーラーエージェントに対し1986年1月から2年間の現実のレートの子データを用いたトレーニングを20回繰り返して最適化を行ったのち、トレーニング期間終了後52週先までシミュレーションを行う。このトレーニング20回、シミュレーションという一連の操作を100回行い、その平均をそのシミュレーションの子データとする。次にトレーニング期間、シミュレーション期間を26週間ずらしシミュレーションを行う。これをデータがなくなるまで繰り返した。その後、各シミュレーションの平均データのシミュレーション開始後1, 4, 13, 26, 52週先のレートについて現実のレートとの対象期間に対する平均絶対誤差 (MAE) を求め、AGEDASI TOF と比較した。この評価方法は和泉ら [1] が用いている方法である。

MAE(k)

$$= \frac{\sum_{s=0}^{N-1} |\tilde{S}(t+s \times 26+k) - S(t+s \times 26+k)|}{N} \quad (3)$$

$S(t)$: 期間 t の現実の為替レート (対数)
 $\tilde{S}(t)$: 期間 t のシミュレーションモデルの
 為替レート (対数)
 k : 誤差計測対象期間 (1, 4, 13, 26, 52)
 N : 全試行回数 (= 21)

表2 提案モデルの評価
Table 2 The evaluation of our model.

	MAE (×10 ³)	
	AGEDASI TOF	提案モデル
k = 1	1.78	0.98
k = 4	2.53	1.82
k = 13	4.96	4.27
k = 26	5.41	5.74
k = 52	8.53	8.36

結果を、表2に示す。提案モデルはほぼすべての期間でAGEDASI TOFよりも平均絶対誤差(MAE)が小さかった。これらのことから提案モデルによる情報交換の仕組みがAGEDASI TOFのものよりモデルの精度においてもより現実的な仕組みを実装していると考えられる。

4. 介入エージェントの導入

本研究では前節で述べた、市場モデルにレートと安定を目指す介入エージェントを導入する。これは現実世界では政府、中央銀行に相当するものである。

AGEDASI TOFでは、予想材料の介入の項目はあらかじめ決めておいた値を入力しているため、シミュレーションパスが予想と全く違っても介入が行われる期間、介入の大きさは同じである。本研究では、介入を行う機関をエージェントとして系に組み込むことで市場の状態に応じた介入を可能とするモデルを構築する。また、AGEDASI TOFにおける介入は予想材料の1つにすぎず市場の需給には影響しない。本研究のモデルでは介入エージェントも市場に参加し売買を行うので、介入が行われた場合、それ自体が市場全体の需給に影響する。

4.1 強化学習による介入政策シナリオの獲得

AGEDASI TOFを用いた為替政策の意思決定支援システムでは、シミュレーションから得られたデータを分析し戦略シナリオを作成する必要があった[1]。本研究では介入エージェントに強化学習を用いた学習機構を組み込むことで為替レートと安定させる介入政策シナリオを自動的に獲得することを目的とする。

強化学習のアルゴリズムとして、本研究ではProfit Sharing[3]を用いた。Profit Sharingとは、報酬に至るエピソードにおける状態Stateと行動aの対からなるルール系列を記憶しておき、報酬が得られたら系列上のルールを強化していくというものである。本研究での状態Stateは、次式で定義する。

$$\text{State} \equiv \left(\sum_{k=1}^{17} \bar{w}^k x^k, \bar{w}^8, x^{14}, p \right) \quad (4)$$

ここで x^k はk番目の予想材料、 \bar{w}^k は予想材料 x^k に対する全ディーラーエージェントの重みづけの平均値である。よってStateの第1項は市場のレートの変動予測に対する平均を表す。 \bar{w}^8 は介入に対する全ディーラーエージェントの重みづけの平均値、 x^{14} は1期間前のレートの変動値(短期トレンド)である。またpは1期間前のレートのポジションであり、目標とする

レートの範囲($\text{rate}_1 \sim \text{rate}_2$)に対しレートSが、 $S < \text{rate}_1$, $\text{rate}_1 \leq S \leq \text{rate}_2$, $\text{rate}_2 < S$ のどの状態かを表す。この状態表現を用いることでレートが目標値より高くかつ上がりトレンドならドル売り介入を行う(目標為替相場仮説, 非対称的leaning-against-the-wind政策[4])というような行動を学習することが可能になる。

次に行動aは介入量で、次式で表される。

$$a = q \times n \quad (n = \pm 3, \pm 2, \pm 1, 0) \quad (5)$$

qは介入の基準量である。nの値により行動aは7種類で、その行動は予想材料の介入 x^8 (表1の介入)にフィードバックされる(注3)。行動選択は各ルールの重みづけによるルーレット選択とした。

報酬は試行全体と各期間ごとの2つの方法で評価する場合を考えた。

a) 試行全体の評価

対象期間全体が目標とするレートの範囲に入っていれば報酬r(>0)を得る。そのときエピソード上のi番目のルールに対する重みづけ $W(\text{State}_i, a_i)$ は次式により一律に強化する。

$$W(\text{State}_i, a_i) \leftarrow W(\text{State}_i, a_i) + r \quad (6)$$

b) 各期間の評価

各期間において決定したレートが目標とする範囲に入っているかどうかでその期間にとったルールに対する重みづけを変化させる。

レートが目標範囲内の場合

$$W(\text{State}, a) \leftarrow W(\text{State}, a) + r' \quad (7)$$

レートが目標範囲外の場合

$$W(\text{State}, a) \leftarrow W(\text{State}, a) - r' \quad (8)$$

すなわちレートが範囲内なら報酬を得るが範囲外の場合は負の報酬により重みづけが小さくなる。ここでは $r' = r/2$ とした。

4.2 実験の設定

10月に1週間で20円もドル-円レートが下落した1998年を対象にシミュレーションを行う(図4)。まず、3.2と同様に1996年1月から1997年12月の2年間の現実のレートを用いたディーラーエージェントの最適化のためのトレーニングを20回行ったのち、対象期間である1998年のシミュレーションを行い介入エージェントに学習させた。この一連の操作を30000回繰り返す。ただし、各一連の操作に置いて介入エージェントの各行動に対する重みづけは保持するものとする。

a) 目標とするレートの範囲

単純化のために、tによらず常に116~136円とした。

(注3): nの値がそのまま予想材料 x^8 となる

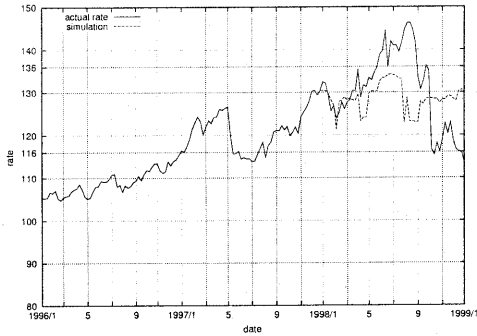


図4 現実のレートと介入が成功したシミュレーション

Fig. 4 The actual rate and a simulation that the intervention policy succeeded.

トレーニング期間: 1996年1月第1週~1997年12月最終週

シミュレーション期間: 1998年1月第1週~1998年12月最終週

b) 介入の基準量

現実の介入においてはシグナル効果^(注4)の重要性が指摘されており、ポートフォリオ・バランス効果^(注5)はシグナル効果に比べるとかなり小さいという事例も報告されている[5]。そこで本研究では介入量をディーラー1人の売買要求量と同程度になるようにした。これは介入のコストを押さえるためでもある。

4.3 評価方法

シミュレーション期間のレートが常に介入エージェントの目標とする範囲に収まっているパス(図4)を安定グループとする。その安定グループがシミュレーション100回中に占める割合で評価を行った。

4.4 実験結果

まず、介入エージェントが存在しない市場のシミュレーションを行った。その内安定グループに属するシミュレーションパスは16%であった。

次に介入エージェントが存在する市場のシミュレーションを行った。学習回数に対する安定グループの割合を図5に示す。また、シミュレーション期間におけるディーラーエージェントの学習を行わずに行った場合の結果を図6に示す。

ディーラーエージェントが学習する場合、しない場合ともに学習回数が大きくなっても安定グループの割合の増加がなく、また収束する様子も見られない。

5. 考 察

現段階では、図5、6に見られるように学習がうまく進んでいない。これには「不完全知覚問題」が大きな原因として考えられる。本研究のモデルにおいて介入エージェントが知覚している状態 State は市場の状態を過度に縮約していると考えられる。例えば状態 State の第1項として全ディーラーエージェン

(注4)：中央銀行の介入から市場参加者が将来の金融政策に関する意図を読み取り、それに基づいて将来の為替相場に関する予想を変化させることによる為替相場への影響

(注5)：介入が外国通貨の需給に影響を与え、それを反映して為替相場が変化するというもの

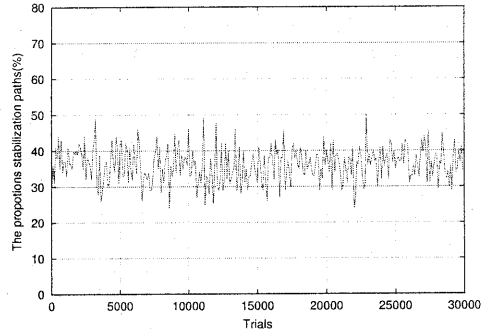


図5 介入エージェントの学習回数に対する安定グループの割合

Fig. 5 The proportions of stabilization paths to learning times of intervention agent.

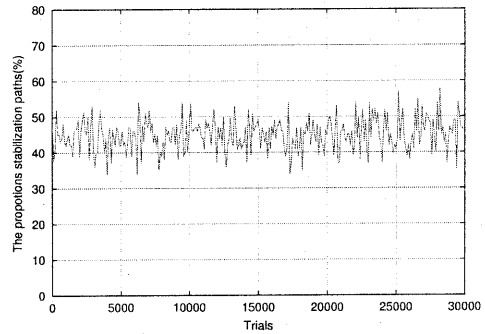


図6 介入エージェントの学習回数に対する安定グループの割合(ディーラーエージェント学習なし)

Fig. 6 The proportions of stabilization paths to learning times of intervention agent. (without dealer agents' learning)

トのレート予測の平均値を持っているが、個々の予測や売買要求次第によってはまったく異なるレートにもなりうる。Profit Sharingを用いた学習では状態数の爆発を防ぐためにある程度状態を縮約する必要があるが、的確な縮約を行わないと学習が進まないという問題がある。本研究の問題のように的確な縮約を行うことが難しい場合、Classifier Systemなど大きな状態を扱える学習法を用いることによる改善が考えられる。

またディーラーエージェントが学習する場合、「同時学習問題」も原因と考えられる。ディーラーエージェントの状態が学習によって変化するために、介入エージェントの行動の結果がディーラーエージェントの状態の変化によって予測しがたくなっている。

6. ま と め

本研究では AGEDASI TOF をもとに情報交換により現実的な仕組みを採用した。これにより各エージェントの状態の変化を見ても不自然さが少なくなり、マイクロレベルでの解析に寄与できると考える。

また、系に介入エージェントを組み込み強化学習を用いてレートの安定に有効な介入政策シナリオの獲得を目指した。しかし、学習が進む様子が見られなかった。原因として「不完全

知覚問題」,「同時学習問題」が考えられそれらの対処のためにはより適切な状態表現の模索, これらの問題に強い学習方法の採用が必要であると思われる。

文 献

- [1] 和泉潔, 植田一博, 人工市場アプローチによる為替シナリオの分析, コンピュータソフトウェア, vol.17, no.5, pp.47-54, 2000
- [2] D.E.Goldberg, "Genetic algorithms in search, optimization, and machine learning," Addison-Wesley Publishing Company, 1989
- [3] J.J.Grefenstette, "Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms," Machine Learning, vol.3, pp.225-245, 1988
- [4] 渡辺努, 市場の予想と経済政策の有効性, 東洋経済新報社, 1994
- [5] Katharyn M. Dominguez and Jeffrey Frankel, "Does Foreign Exchange Intervention Matter? Disentangling the Portfolio and Expectations Effects for Mark," American Economic Review, NBER Working Paper # 3299, vol.83, no.5, pp.1356-1369, 1993