

人工市場とテキストマイニングの融合による市場分析

Integration of Artificial Market Simulation and Text Mining for Market Analysis

和泉 潔
Kiyoshi Izumi

産業技術総合研究所 デジタルヒューマン研究センター
Digital Human Research Center, AIST
kiyoshi@ni.aist.go.jp, <http://staff.aist.go.jp/kiyoshi.izumi/>

松井 宏樹
Hiroki Matsui

産業技術総合研究所 情報技術研究部門
Information Technology Research Institute, AIST
hiroki.matsui@aist.go.jp

松尾 豊
Yutaka Matsuo

(同上)
y.mastuo@aist.go.jp

keywords: artificial market, social simulation, text mining, multi-agent simulation, complex systems

Summary

We constructed an evaluation system of the self-impact in a financial market using an artificial market and text-mining technology. Economic trends were first extracted from text data circulating in the real world. Then, the trends were inputted into the market simulation. Our simulation revealed that an operation by intervention could reduce over 70 % of rate fluctuation in 1995. By the simulation results, the system was able to help for its user to find the exchange policy which can stabilize the yen-dollar rate.

1. なぜ人工市場とテキストマイニングの融合か?

金融市場のような複雑な社会現象を分析する際に特に困難な点が2つある。1つは対象となる社会現象に影響する外部環境の動向を把握することである。もう1つは、分析者自身が社会現象の参加者である場合に、自分自身の行動が社会現象やその外部環境に与える影響(自己インパクト)を評価することである。例えば金融市場において、特に中央銀行の意思決定者のようなビッグプレイヤーは自分の行動が市場および経済全体に与える影響を意思決定の際に考慮しなければいけない。動向分析と自己インパクト評価は分析の際に分離して取り扱うことはできない。なぜなら、自己インパクトの影響と外部環境の動向は相互に依存する状態にあるからである[塩沢 99]。本研究は金融市場を題材として、両者を統合して分析する枠組みとして、テキストマイニング技術と社会シミュレーションを融合した SEMAS (Socially Embedded Multi Agent Simulation) アプローチを提唱する。

SEMASは図1の2つのステップからなる。まず、ウェブやメディアなどに現時点で流通している膨大な経済指標・記事データから、テキストマイニング技術により、期間毎に注目すべき意思決定に関する動向を抽出する。こ

のように現実世界にある経済ニュースを解析してモデルに取り込むことによって、モデルと現実世界との対応付けを図る。次に、抽出された動向を、仮想的な市場のマルチエージェントモデル(人工市場[和泉 03])に入力し、計算機シミュレーションを行う。取り込んだ経済ニュースを基に自らの経済行動を決定する複数のエージェントが存在する市場のマルチエージェントシミュレーションによって、自己インパクトの評価を行う。シミュレーション結果によって、市場が自分の望ましい状態に近づくような行動を選択する。この2つのステップにより現実の社会動向と自己インパクトのフィードバックループの中に、社会シミュレーションを埋め込むのである。本研究では、この枠組みにより、急激なレート変動で円の最高値を更新した1995年の円ドルレートを安定化するような為替政策を決定することを支援するシステムを構築した。

2. テキストマイニングによる現実の経済動向の抽出

SEMASの第一段階として、テキストマイニングによる現実世界に流通する情報を抽出する手法を紹介する。市場などの現実の社会システムの各参加者は常に様々な情報を受信および発信を行っている。彼らは受信した情

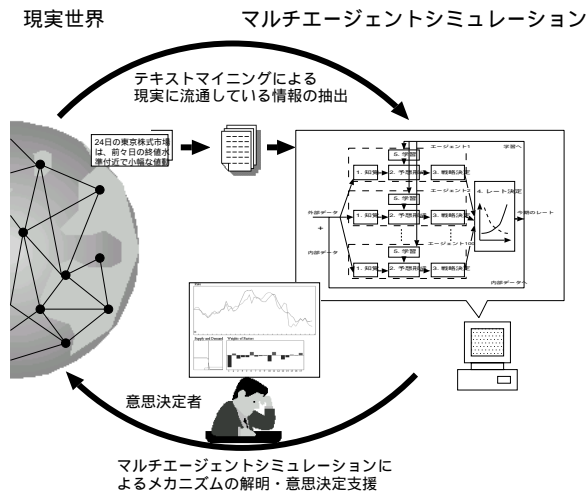


図 1 SEMAS の概要

報を自分なりに認知・解釈を行って、その解釈結果を基に行動を決定している。既存の社会シミュレーションの多くは、外部環境に関する定性的な意味情報を正面から取り扱っていなかった [Edmonds 05]。例えば、今までの人工市場研究では市場外部の影響を全く考慮しないか、考慮したとしても株式への配当のように一つの数値指標に集約して単純化した形でしか取り扱ってこなかった。しかし、外部の多様な環境を考慮しないとどうしても現実世界から乖離したモデルになってしまう。現実世界との対応付けを目指すならば、できるだけ実際の意思決定者の間で流通している情報をモデルに取り込むことが必要である。現在、ウェブ技術の進歩は、この限界を大きく前進させる可能性がある。なぜなら、実際の意思決定者が利用している多くの情報がウェブ上に載るようになってきたことによって、現時点で社会システムの参加者コミュニティ内で流通している情報を取り出す方法が可能になりつつある。本研究では、現実世界にある経済ニュースをテキストマイニングによりニュースのインパクトと意味カテゴリを解析してモデルに取り込むことによって、モデルと現実世界との対応付けを図る。本研究で用いたテキスト解析の手順を図 2 に示す。

2.1 ドキュメントの特徴量の抽出

現実世界での経済動向を分析するためのテキストデータとして、本研究では国際金融情報センター (<http://www.jcif.or.jp/>) が毎週発行する市場解説の記事を用いた。市場解説では会員向けの情報提供サービスとして、毎週 A4 紙 1 枚 1000-1500 文字の分量でその週の金融市場の関連したと思われる経済ニュースをまとめたテキストをレポートとして配信していた。この記事は、指標の発表内容など実際に起きた出来事だけでなく、当時の市場の観測に関する情報も記述している。そのため、既存の人工市場シミュレーションとは異なり、本研究では市場が織り込んである期待や見込みといった要素も取り扱

元の経済ニュース・テキストデータ
第 t 週のドキュメント

文 a 「ドル買いが優勢で円は一時 134 円 38 銭まで下落した」
文 b 「その後も日銀総裁が…」
文 c
⋮

(a) 形態素解析

文 a' 「/ドル/買い/が/優勢/で/円/は/一時/134/円/38/銭/まで/下落/した/./」
文 b' 「/その/後/も/日銀/総裁/が/…」
文 c'
⋮

(b) トークン化

文 a'' 「/dollar/buy/ -/strong/ -/yen/ -/temporary/ -/ -/ -/ -/ -/ down/ -/ -/」
文 b'' 「/ -/ -/ -/central_bank/ -/ -/…」
文 c''
⋮

(c) ドキュメントの特徴量の抽出

各キーワードの第 t 週のドキュメントでの tf-idf 値

dollar, 0.041978
strong, 0.018933
yen, 0.081418
⋮

(d) 各週の経済動向の推定

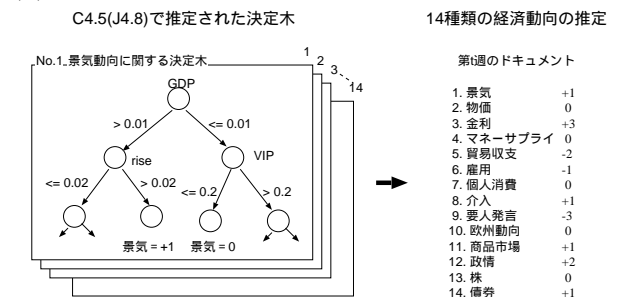


図 2 テキストマイニングの手順

うことができる。

最初にドキュメントの特徴量を抽出するための前処理として、上述の経済ニュース・テキストデータを日本語形態素解析システム「茶筌」(ChaSen) [ChaS] を用いて、各週のテキストを形態素に分解した(図 2a)。形態素解析により各単語の基本形と品詞情報を得ることができる。さらに、あらかじめ用意した辞書とのマッチングにより、テキストデータからその週の経済動向を判断するのに関連しそうな単語のみを抽出し、同義語や類義語を同一のキーワードにまとめる処理(トークン化)を行った*1(図 2b)。キーワード辞書とは表 1 に示すような形式で、形態素解析で得られた単語をグループ化して、経済動向分析に関連がある 183 個のキーワードに分類したものである。この辞書は、1992 年から 1994 年の市場解説記事データの形態素解析結果から、一般的な用語は同義語辞書により金融専門用語と固有名詞は専門用語辞書および金融に関する専門書を参考にしてグループ化や重要性の低い語の削除を行って作成した。キーワードは表 1 の「yen」のような名詞や「rise」のような動詞、「large」のような形容詞・副詞などの種類がある。単語によっては、単語の基本形が下位語に一致するかという条件のみでなく、品詞情報も条件に加えた場合もある。例えば、キーワード yen の下位語「円 !名詞-接尾-助数詞」は、単語の基本形が「円」に一致し、品詞が「名詞-接尾-助数詞」でないという条件を示す。この場合「/円/が/大幅に/上る/」といった場合は条件に一致するが「/105/円/の/大台/」といった数字の後について単位を示す場合には一致しない。

表 1 キーワード辞書の例

キーワード	下位語 1, 下位語 2, ...
yen	円 !名詞-接尾-助数詞
rise	上昇, 急伸, 上げる, 上伸, 続伸, ...
large	大きい, 大量, 大幅, 厚い, ...

「!名詞-接尾-助数詞」は、名詞-接尾-助数詞という品詞情報を含まないという条件を示す。

各週のドキュメントを上述の手順でキーワードの集合に変換し、その週の全ての文章での全キーワードの出現頻度により特徴付けを行った。各週の特徴を表す指標として tf-idf(term frequency - inverse document frequency) 値を用いた*2。キーワード k のある週 t のドキュメントでの tf-idf 値 $\text{tfidf}_{k,t}$ は次の式で計算される。

$$\text{tfidf}_{k,t} = \text{tf}_{k,t} \times \log\left(\frac{N}{\text{df}_k}\right) \quad (1)$$

ただし、 $\text{tf}_{k,t}$ は(キーワード k がその週 t のドキュメントに出現した回数) / (その週 t のドキュメントの全てのキーワードの出現回数の合計) であり、 N は分析対象と

なる全ての週(ドキュメント)の数、 df_k はそのうちキーワード k が 1 回でも出現した週(ドキュメント)の数である。他の週ではあまり出現しないが、その週にだけ出現頻度が高いような単語が、その週の特徴をよく表していると考えられ、tf-idf 値が高くなる。図 2 の(c) で示されているような、183 個の各キーワードに対するその週の tf-idf 値を並べたベクトル D_t がその週のドキュメントの特徴を表していると考えられる。

$$D_t = \{\text{tfidf}_{k,t}\}, \quad k = 1, \dots, 183. \quad (2)$$

2.2 決定木による経済動向の推定

本研究では各テキストの特徴を決定木により分類することによって、テキストデータから自動的に経済動向を推定することを行う。これにより新しく与えられた市場解説記事を利用したり、将来的にウェブ上の記事やニュース等の現実に流通する市場解説記事以外の様々なテキストデータに適用できる可能性がある。

図 2d に示すように、先ほど計算した各週の特徴量 D_t をあらかじめ推定した決定木に入力し、表 2 にある経済指標や政治的なニュースなどの 14 種類の各カテゴリに関する動向を評価する。これらの 14 種類のカテゴリは金融市場の解説本によく現れる関連情報を選んで分類した。決定木の推定は決定木学習 C4.5 の一種である J4.8*3 を用いて行った。本研究では 1995 年の行動に関する意思決定を行うので、それよりも過去の 1992 年から 1994 年の 3 年間(156 週)のデータを用いて決定木の学習を行った。入力データとして、この期間に国際金融情報センターが発行した市場解説記事から抽出した各週の特徴量 D_t を用いた。教師データとして、我々があらかじめこの期間の市場解説記事を見て 14 種類の各カテゴリの動向について $\{0, \pm 1, \pm 2, \pm 3\}$ の 7 段階にコーディングした値を用いた。ここで正のコーディング値は日本の景気上昇など経済学のオーソドックスな理論では円高要因となる情報を表し、負の値は円安要因の情報を表している。コーディングの絶対値は、各カテゴリに関連する単語がその週の解説記事に現われた回数にもとづいて決定された。さらに値は「非常に」「わずかに」のような記事中の記述によっても調整された。例をあげると、「アメリカの失業率は大幅に減少した」という文は「雇用: -3」としてデータ化された。米国の失業率の減少は、経済の基礎的要因(ファンダメンタルズ)の指標として見ると、米国の好景気とドル資本の需要増大を表して円に不利(ドルに有利)な情報であるので、データの符号は負になる。また「大幅に」という記述は大きな影響を示すので、3 の最高点が割り当てられた。もう 1 つの例としては「日本銀行の総裁は円高を容認する発言をした」という文は「要人発言: +1」としてデータ化された。円高の容認は円に有利

*1 経済ニューステキストの処理に関して [Carley 04, Ahmad 05] を参考にした。

*2 tf-idf 値の経済ニュース分析への応用としては例えば [Seo 04] を参照してほしい。

*3 C4.5 アルゴリズムをデータマイニングソフトウェア Weka [Weka, Witten] に実装したもの。

(ドルに不利) な情報であるので、符合は正となる。また要人発言カテゴリに関連する記述は今週の記事にこの文のみ現われただけなので、1の最低点が割り当てられた。

このようにして得られた決定木の性能評価として、各カテゴリの動向に関する決定木による推定結果の正答率を表2に示す。結果として、カテゴリによって多少のばらつきがあるが、決定木による推定値が7段階のコーディング値の符号と値が共に完全に一致した割合(正答率)が、訓練データに対するテストでは平均92.2%、外部データに対する交差検定テスト*4では平均71.9%の成績を収めることができた。特に様々な要素が関連する金利や要人発言、欧州動向のカテゴリの正答率は相対的に低かったが、全体としては実用可能なレベルに達した。

表2 決定木による経済動向推定結果 (ドキュメント数 $N=156$ 。正答率は決定木による推定値が7段階のコーディング値と符号と値が共に一致した割合。訓練データ集合に対するテストと10分割交差検定テスト時の正答率を示した。)

カテゴリ	訓練データ	交差検定
	正答率 (%)	正答率 (%)
1 景気	88.9	62.9
2 物価	94.1	75.9
3 金利	81.1	57.1
4 マネーサプライ	97.4	97.4
5 貿易収支	96.7	85.0
6 雇用	90.9	66.2
7 個人消費	91.5	90.9
8 介入	95.4	66.8
9 要人発言	87.6	45.4
10 欧州動向	88.3	41.5
11 商品市場	99.3	99.3
12 政情	92.2	69.4
13 株	93.5	75.3
14 債券	94.1	74.0
平均	92.2	71.9

実際に得られた決定木の例として物価カテゴリに関する決定木の一部を図3に示す。見出し語 price の tf-idf 値が低い場合にはその週の物価カテゴリの状況は特に変化なし(状況を表す値がゼロ)と推定したり、見出し語 price と rapid の tf-idf 値が高い場合(「物価が急激に…」など)や price と attitude の tf-idf 値が高い場合(「物価に対する態度を…」など)に物価カテゴリの状況にゼロ以外の値を与えている。このように一般的に見て妥当な決定木が得られていることを確認した。

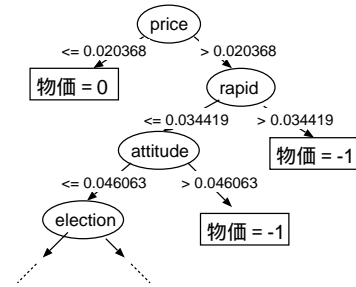


図3 物価カテゴリに関する決定木の一部

3. 人工市場モデルの構築

SEMASの第二段階として、第一段階で取得した現実世界の経済動向を用いてマルチエージェントシミュレーションを行う。ユーザは提示されたシミュレーション結果を基に現実の社会状況における自らの行動を決定して働きかけを行う。本稿では、一つのケーススタディとして、円ドルの為替レート安定化のための行動決定の支援を取り上げる。そのために外国為替市場のマルチエージェントモデル (AGEDASI TOF*5) を構築した*6。AGEDASITOFは100人の仮想的なディーラーからなるコンピュータ上の人工的な市場である(図4)。本モデルにおける仮想的なディーラーは一つのコンピュータプログラムであり、以下エージェントと呼ぶ。モデルの1期間は現実の市場の一週間に対応し、以下の5つのステップよりなる。

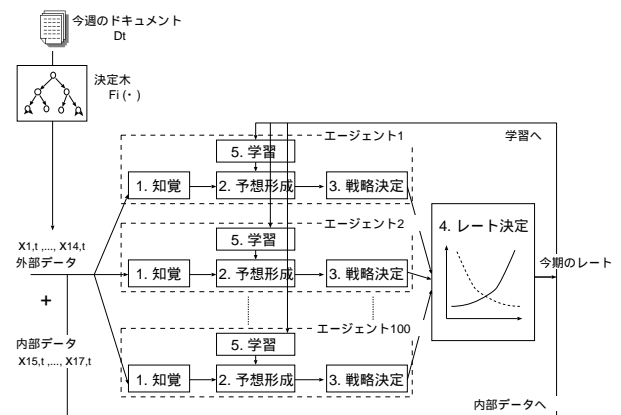


図4 モデルの枠組み

3.1 知覚ステップ

最初に各エージェントは、将来の為替レート予測の材料となるカテゴリ(1. 景気, 2. 物価, 3. 金利, 4. マネーサプライ, 5. 貿易収支, 6. 雇用, 7. 個人消費, 8. 介入, 9. 要人発言, 10. 欧州動向, 11. 商品市場, 12. 政情, 13. 株, 14.

*4 データの集合を10分割し、10分の9を訓練データとして学習し、残りの10分の1を外部データとして学習した処理の信頼度を計算することを全ての組み合わせにおいて行った平均をモデルの評価とする手法 [Witten]。

*5 A Genetic-algorithmic Double Auction Simulation in Tokyo Foreign exchange market.

*6 人工市場モデルの詳細は [和泉 03] を参考にしてほしい。

債券, 15. 短期トレンド (先週の変動), 16. 短期トレンド (変動の変動), 17. 長期トレンド (5 週間の変動)) について各週の動向 $x_{i,t}$ ($i = 1, \dots, 17$) を $\{0, \pm 1, \pm 2, \pm 3\}$ の 7 段階の形で入力される。これらの 17 種類のカテゴリは経済指標や政治的状況に関する経済動向と過去の為替変動に関するチャート動向の 2 種類に分かれる。最初の 14 種類の経済動向は, 2 章のテキストマイニング手法によって評価された値を用いる。

$$x_{i,t} = F_i(\mathbf{D}_t), \quad i = 1, \dots, 14 \quad (3)$$

ただし, $F_i(\cdot)$ は第 i 番目の経済状況に関する決定木の関数, \mathbf{D}_t はテキストデータから (2) 式で計算された第 t 週の特徴ベクトルである。また, 3 種類のチャート動向 $x_{15,t}, x_{16,t}, x_{17,t}$ は人工市場でシミュレートされた過去のレートから長期と短期のトレンドを計算し, それを標準偏差で割って標準された値を用いた。本研究ではインサイダー情報のようなエージェントによって持っている情報に違いがある状況は取り扱わずに, 市場参加者コミュニティで共通して持っている情報のみを扱うこととして, 渡される情報は全てのエージェントで同じである。

3.2 予想形成ステップ

各エージェント j は自分独自に各予想材料 i に対する重要度を持っており, さきほど知覚されたニュース $x_{i,t}$ の加重平均で将来の為替レート (対数値) の変動値 ΔS_t を予想する。

$$\mathbf{E}_t^j[\Delta S_t] \equiv f\left(\sum_{i=1}^{17} w_{i,t}^j x_{i,t}\right),$$

ただし, f は小数点以下を切り捨てる関数である。また各エージェントの予想の確信度 (分散の逆数) は以下のように計算される。

$$\left(\text{Var}_t^j[\Delta S_t]\right)^{-1} \equiv \sqrt{|(wx_+)^2 - (wx_-)^2|},$$

ここで, wx_+ は $w_{i,t}^j x_{i,t} > 0$ の和, wx_- は $w_{i,t}^j x_{i,t} < 0$ の和である。確信度は円高の要因 (予想材料と重要度の積が正であるもの) の和と, 円安の要因 (予想材料と重要度の積が負であるもの) の和との間の差から計算され, どちらかの要因に片寄っていれば分散は小さく (確信が強く), 両方の要因が均等であれば分散が大きく (確信が弱く) なる。

3.3 戦略決定ステップ

各エージェント j は期待利得を最大にするように最適なドル資産保有高 q_t^{j*} を決定する。

$$q_t^{j*} = \frac{1}{a} \frac{\mathbf{E}_t^j[\Delta S_t]}{\text{Var}_t^j[\Delta S_t]},$$

ただし, a は危険回避度である。自分のドル資産保有高が最適になるように, 自分の予想したレートで市場に注文をする。

$$\text{注文量 } \Delta q_t^{j*} \equiv q_t^{j*} - q_{t-1}^j, \quad \text{注文レート} \equiv \mathbf{E}_t^j[\Delta S_t].$$

3.4 レート決定ステップ

100 個のエージェントの売買戦略を市場全体で集積して, 需要と供給が一致するような値に人工市場モデルの今期の円ドルレートが決定される。

3.5 学習ステップ

遺伝的アルゴリズム (以下, GA) を用いて各エージェントの予想形成部を適応させる。具体的には, 17 種類の材料に対する重要度を一列に並べたものを個体とし, 予想レートと実際のレートの差にマイナスをかけたものを利得として, 淘汰や交叉, 突然変異の操作を行った*7。GA のパラメータは 1986 年から 1993 年のデータを用いた外挿予測テストの結果より, 自然淘汰の確率 80% とし, 淘汰を受ける個体に交叉の確率を 30%, 突然変異の確率を 0.3% とした [和泉 03]。

$$\text{個体 } \mathbf{w}_t^j \equiv (w_{1,t}^j, w_{2,t}^j, \dots, w_{n,t}^j).$$

$$\text{個体 } \mathbf{w}_t^j \text{ の利得} = -|\mathbf{E}_t^j[\Delta S_t] - \Delta S_t|.$$

これらの操作により, 実際のレートに近い予想をした重要度の組合せの市場全体での頻度が増え, 予想が外れた組合せの頻度が減る。このことは, 各エージェントが予測力の高い予想方式をまねして取り入れたと解釈できる。又, 他人との相互作用や自分一人によって新しい予想方式をつくっていく。このようにして次期に用いる予想方式が用意される。

4. 人工市場シミュレーションによる行動ルールの決定

3 章で構築したマルチエージェントモデルの計算機シミュレーションにより, 現実世界に働きかけを行うユーザの金融・為替政策に関する意思決定の支援を目指す。今回の意思決定のゴールは, これから来る 1995 年の円ドルレートの変動をある一定の幅の中におさめて安定化できる行動ルールを, 過去のデータを用いたシミュレーションにより決定することである。本研究における意思決定者とは金利, 介入や要人発言に関する政策に対して決定もしくは提言を行える政府・日銀関係者を想定している。行動ルールは「IF(条件)THEN(行動)」の形式であり, 行動部は意思決定者が操作可能な金利・介入・要人発言に関する操作であり, 条件部は表 2 の 17 種類のカテゴリのうち操作対象である 3 つのカテゴリを除いた 14 種類のカテゴリに関する今週の動向である。つまり, 1995 年の 1 年間に経済やチャート動向に応じてどのように金利・介入・要人発言を操作するかというルールである。1992 年から 1994 年の過去 3 年間のデータを用いて人工市場シミュレーションを行って, レートを最も安定化できそうな条件部と行動部をそれぞれ決定する。

*7 GA の操作は Goldberg [Goldberg 89] をもとにした

4.1 条件部の決定

市場の安定化のためには、最近の市場が敏感に反応しているカテゴリに関する状況の変化があったときに、金利・介入・要人発言に関する操作を行った方が効果的である。そのような市場が敏感なカテゴリを条件部として選ぶために、1992年から1994年の過去3年間のデータを用いて、次の手順による人工市場シミュレーションを行った。まず最初に初期化として、100個のエージェントの持っている重要度の初期値はランダムに決定され、ポジションはスクエア(円資本のみ)から始まった。次に、1992年から1994年の3年間の実際の市場における17種類のファクターのデータと実際の円ドルレートのデータを用いて、各エージェントに重要度の値を学習させた。条件部の決定のためのシミュレーションでは人工市場はレート決定を行わず、現実の円ドルレートを知覚ステップにおけるチャートトレンドと学習ステップにおける各エージェントの適合度の計算に用いた。上記の手順を乱数の種を変えて100回シミュレーションを行い、全てのエージェントが各カテゴリに対して与えた重要度の平均値を求めた。重要度の平均が大きいカテゴリほど、この期間で市場が敏感に反応したカテゴリである。

重要度の平均が大きかったカテゴリの上位5個は、大きい順に短期チャートトレンド $x_{15,t}$ 、債券市場 $x_{14,t}$ 、チャートの2次変化 $x_{16,t}$ 、マネーサプライ $x_{4,t}$ 、政情 $x_{12,t}$ であった。つまり、この期間において人工市場のエージェントはこれらのカテゴリに関する動向に対して特に敏感だったことが分かった。従って、これらの状況に変化があったとき、特に為替レートは大きく変動して不安定になりやすいということである。よって、この5つのカテゴリを条件部と定義して、その変化に応じて金利・介入・要人発言の制御可能なカテゴリのどれを操作して安定化するかで、図5の7つの為替政策の行動ルールの候補を用意した。操作はその週だけで完結して翌週以降に操作を継続しないよう、できるだけ短期的で最大の効果が起こるように、制御可能なカテゴリを操作するときは最大の+3または最小の-3に操作するように設定した。

4.2 自己インパクト評価による行動部の決定

図5の7つの行動ルールの候補について、意思決定期間である1995年の一番近くの過去1年間(1994年)のデータを用いたシミュレーションにより、それぞれの行動が市場に与えるインパクトを評価した。その中で最もレートが安定化した行動ルールを、次の1995年のレート安定化に対して最善なルールとして意思決定者に提示する。

シミュレーションの手順はまず4.1節と同様に初期化を行い、1992年から1993年の2年間について現実の経済動向データおよび為替レートを使って各エージェントに重要度の学習を行わせた。さらに、得られた重要度を初期値として、1994年の外挿シミュレーションを行った。

5つのカテゴリ(マネーサプライ・政情・債券市場・短期チャートトレンド・チャートの2次変化)に関する状況の変化があったときに ($|\sum_{i=4,12,14,15,16} x_{i,t}| > 0$)、

(a) 金利をニュースと逆の方向で大きさ3に操作する。

$$x_{i,t} = \begin{cases} 3 & (\sum_{i=4,12,14,15,16} x_{i,t} < 0) \\ -3 & (\sum_{i=4,12,14,15,16} x_{i,t} > 0) \\ 0 & (\sum_{i=4,12,14,15,16} x_{i,t} = 0) \end{cases} \quad (4)$$

ただし、 $i = 3$ 。

- (b) 介入をニュースと逆の方向に大きさ3で行う。
式(4)で $i = 8$ 。
- (c) 為替政策に関する発言をニュースと逆の方向に大きさ3で行う。
式(4)で $i = 9$ 。
- (d) 金利と介入をニュースと逆の方向で大きさ3に操作する。
式(4)で $i = 3, 8$ 。
- (e) 金利と発言をニュースと逆の方向で大きさ3に操作する。
式(4)で $i = 3, 9$ 。
- (f) 介入と発言をニュースと逆の方向で大きさ3に操作する。
式(4)で $i = 8, 9$ 。
- (g) 金利と介入と発言をニュースと逆の方向で大きさ3に操作する。
式(4)で $i = 3, 8, 9$ 。

図5 為替政策の行動ルールの候補

外挿シミュレーション中はテキストマイニングで得られた経済動向データのみを人工市場に与え、実際の為替レートを与えずに人工市場のレート決定ステップでレートが決定される。さらに、金利・介入・要人発言のカテゴリに関しては図5の各行動ルールで操作された値を人工市場に与える。7つの行動ルール間の比較には、実際の1994年の円ドルレートの変動値からの各行動ルールでのレートの変動値の減少率を安定化の指標とした。

シミュレーション結果を図6に示す。介入のみの単独で操作を行う行動ルール(b)が最もレートの変動値を減少させることができた。介入による操作でのレート変動値は100回のシミュレーションの平均で0.02310であり、実際の1994年のレート変動値0.03793よりも約40%の変動を減少することができた。介入の操作を含む行動ルール(b)(d)(f)は全て30%以上の変動を減少することができた。しかし、複数のカテゴリの操作による市場へのインパクトが大きすぎて、介入単独の安定化の効果を損なうことが見られた。特に金利・介入・要人発言の3つ全てのカテゴリで操作を行った行動ルール(g)では、約20%しか変動を減少することができなかった。

以上の結果より、1995年の円ドルレートの安定化の

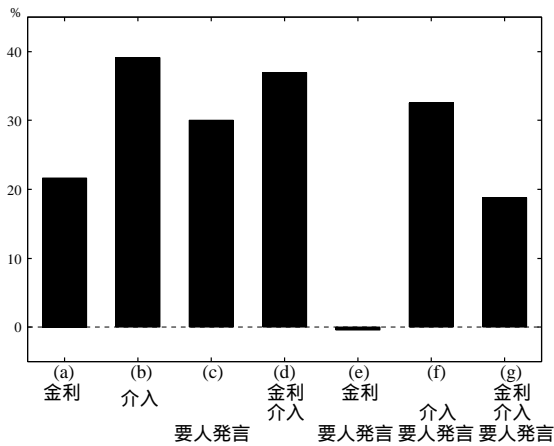


図 6 各行動ルールのレート変動の安定化の割合

ために、以下の行動ルールを本アプローチがユーザに対して最善のルールとして提示を行った。

IF 今週にマネーサプライ・政情・債券市場・短期チャートトレンド・チャートの2次変化に関する状況の変化があったならば、
THEN 上記の経済動向の反対の方向に作用するような介入操作を行う。

5. 得られた行動ルールのテスト

過去のデータを用いて提示された行動ルールを、もし実際に1995年の1年間の為替政策に用いたらどれほどレートを安定化することができたかテストを行った。シミュレーションの手順はまず4.1節や4.2節と同様に初期化を行い、1992年から1994年の4年間について現実の経済動向データおよび為替レートを使って各エージェントに重要度の学習を行わせた。さらに、得られた重要度を初期値として、1995年の外挿シミュレーションを行った。4.2節と同様に外挿シミュレーション中は経済動向データのみを人工市場に与え、人工市場でレートが決定される。また、金利・介入・要人発言のカテゴリに関しては、行動ルール(b)の介入単独操作によって決定された値と国際金融情報センターの相場解説記事を見て実際の操作をコーディングした値の2種類を用意した。この2種類の操作でのレート変動値の比較により本研究で提示された行動ルールの評価を行う。

図7の実線aは介入単独操作を行った場合での100回の人工市場シミュレーションの円ドルレートの平均パス、破線bは実際の操作をコーディングした値を入力した場合のシミュレーションの平均パスである。点線cは実際の1995年の円ドルレートのパスを示す。介入単独操作は実際の操作よりもレートを大幅に安定化することに成功している。実際の操作の場合のシミュレーションでは、変動の幅は小さいが現実のレートのパスと同様に大きな

円高の変動パターンを示していることが分かる。それに対して、介入単独操作を行った場合には、1995年の全期間においてほぼフラットに推移しレートが安定している。介入単独操作は実際の操作を行った場合のパスに比べて70%以上のレート変動を減少することができた。さらに、実際の円ドルレートのパスと比較すると変動の減少率がほぼ80%に達した。これらのテスト結果より、過去のデータを使った評価で提示された行動ルールが、1995年の為替レート安定化に対して有効であることが分かった。

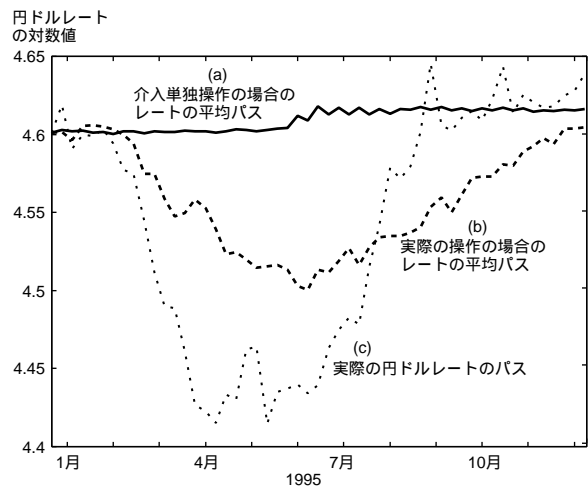


図 7 介入単独操作と実際の操作の場合のシミュレーションパスの比較。シミュレーションパスは100試行の平均。

操作の詳細を見ると、1995年の1年間の外挿シミュレーションでは1試行あたり平均9.5回の介入操作が行われた。ほとんどの介入操作は連続せずに1期間ごとの単発的な操作で、2期間以上の連続した操作はまれであった(図8)。現実世界では1992年から1995年の期間では、年間で平均2兆5739億円、42.75回の介入が行われた[財務]。このうち、1000億円以上の大規模介入を年間に3-10回程度行い、残りは100-400億円の中小規模の介入を行っていた。つまり本システムは実際の大規模介入のタイミング決定を支援していることに相当していた。

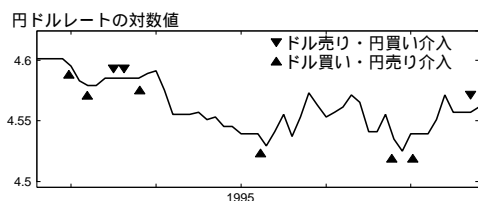


図 8 シミュレーションパスと介入操作の例

6. お わ り に

本研究では、現実の情報を取り込んだ人工市場シミュレーションによって、円ドルレート安定化のための為替政策決定を支援するシステムを構築した。実証テストではシステムによって提示された行動ルールの高い有効性が示された。今回のシステムでは市場参加者の情報行動のみをモデル化し、資産の保有量に関しては直接には取り扱わなかった。図 7 で実際の操作の場合のシミュレーションが実際のパスよりも変動幅が小さかったのは、この点が関係していると思われる。つまり、現実には損切り等の資産保有量の調整がレート変動をさらに急激なものとしている可能性があり、今回の手法そのままでは意思決定支援の有効性が低下する可能性がある。この要素を取り組んだモデル化を行うことを今後の発展の方向性として考えている。

本研究では、各単語の単独の出現頻度からドキュメントの特徴付けを行った。この単純な手法でも、今回は相場解説という限定された領域のテキストであり多くのカテゴリは比較的少数のキーワードに対応していたので、トレンドの推論に有効であった。しかし金利・要人発言・欧州動向などの関連するキーワードの種類が多くなり多様なパターンの文章で解説されているカテゴリに関しては、トレンドの推論の正答率が他のカテゴリに比較すると高くなかった。今後一般的な新聞記事やウェブ記事に処理対象を拡張した場合に、より詳細な文章構造の解析による状況の特定といった手法が必要になる。

本稿で提案した SEMAS は、現実社会で流通している情報のセンシングと、モデルを用いたシミュレーション・アプリケーションによるユーザの行動決定支援、そしてユーザの現実社会への行動といったフィードバックループの中で「使える社会シミュレーション」をツールとして提供することを目指している。さらに今後は、本村ら [本村 06] の提唱するように、現実の社会現象に関して、研究者・研究成果を利用する者・使う現場にいる者などでデータやモデル、それらを利用した結果を共有して新たな社会シミュレーションモデルを創造する知識循環型の体制を作っていくことを目指す。

謝 辞

本研究の成果は一部、科学研究費補助金 基盤研究 (B)(2) 課題番号 16300047 の支援を受けました。

◇ 参 考 文 献 ◇

- [Ahmad 05] Ahmad, K., Gillam, L., and Cheng, D.: Textual and Quantitative Analysis: Towards a new, e-mediated Social Science, in *Proc. of the 1st International Conference on e-Social Science* (2005)
- [Carley 04] Carley, K., Diesner, J., and Tsvetovat, M.: An Integrated Approach to the Collection and Analysis of Network Data, in *Proceedings of NAACOS04* (2004)

- [ChaS] ChaSen ホームページ: <http://chasen.naist.jp/hiki/ChaSen/>
- [Edmonds 05] Edmonds, B. and Moss, S.: From KISS to KIDS - An 'Anti-simplistic' Modelling Approach., in Davidsson, P., Logan, B., and Takadama, K. eds., *Multi-Agent and Multi-Agent-Based Simulation, Joint Workshop MABS 2004*, pp. 130-144, Springer (2005)
- [Goldberg 89] Goldberg, D.: *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley Publishing Company (1989)
- [和泉 03] 和泉 潔: 人工市場: 市場分析の複雑系アプローチ, 森北出版 (2003)
- [本村 06] 本村 陽一, 西田 佳史: オープンライフマトリックス: 日常生活環境における人間行動理解の研究基盤, 2006 年度人工知能学会全国大会論文集 (2006)
- [Seo 04] Seo, Y.-W., Giampapa, J. A., and Sycara, K.: Financial News Analysis for Intelligent Portfolio Management, Technical Report CMU-RI-TR-04-04, Carnegie Mellon University (2004)
- [塩沢 99] 塩沢 由典: ミクロ・マクロ・ループについて, 京都大学経済学会・経済論叢, Vol. 164, No. 5, pp. 463-535 (1999)
- [Weka] Weka ホームページ: <http://www.cs.waikato.ac.nz/~ml/weka/>
- [Witten] Witten, I. H. and Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, second edition
- [財務] 財務省: 外国為替平衡操作の実施状況, <http://www.mof.go.jp/1c021.htm>

〔担当委員: 高間 康史〕

2006 年 12 月 12 日 受理

著 者 紹 介



和泉 潔 (正会員)

1993 年東京大学教養学部基礎科学科第二卒業。1998 年同大学院博士課程修了。博士 (学術)。同年より、電子技術総合研究所 (現 産業技術総合研究所) 勤務。マルチエージェントシミュレーション、特に社会シミュレーションに興味がある。情報処理学会会員。



松井 宏樹

2000 年京都産業大学理学部物理学科卒業。2006 年北陸先端科学技術大学院大学情報科学研究科博士後期課程修了。博士 (情報科学)。2005 年より産業技術総合研究所情報技術研究部門勤務。現在、産業技術総合研究所特別研究員。マルチエージェントモデルによる社会シミュレーションの研究に従事。



松尾 豊 (正会員)

1997 年 東京大学工学部電子情報工学科卒業。2002 年同大学院博士課程修了。博士 (工学)。同年より、産業技術総合研究所情報技術研究部門 勤務, 2005 年 10 月よりスタンフォード大学客員研究員。人工知能、特に高次 Web マイニングに興味がある。情報処理学会, 言語処理学会, AAAI, INSNA の各会員。