

可能世界ブラウザ: マッシブデータ解析とエージェントシミュレーション

Would-be Worlds Browser: Massive Data Analysis and Agent-Based Simulation

和泉 潔*1*2

Kiyoshi Izumi

*1 東京大学大学院 工学系研究科

School of Engineering, The Univ. of Tokyo

*2 JST さきがけ

PRESTO, JST

This paper proposes a would-be-worlds browser for socio-economic systems, a new approach integrating a massive-data analysis and agent-based simulation. This approach tries to create interesting simulation path(s) by segmenting and merging simulation paths. The agents' rules in the simulation are extracted from the massive data. Using would-be-worlds browser, we can design dependable socio-economic systems.

1. マッシブデータから社会動向が予測可能か

1.1 スカート丈と株価

1926年に経済学者のジョージ・テイラー氏がヘムライン指数とよばれる経済理論を提唱した[Lewin 08].ヘムライン(hem-line)とはスカート丈のことである.この理論は,スカート丈が短くなると株式市場が上げ相場になり,長くなると下げ相場になると主張している.理論が発表された1920年代から現在に至るまで,様々な検証が行われてきた[Baardwijk 10].

金融市場は世の中の経済活動の活発さを反映しているはずである.だからもし,みんなが持っている平均的な景況感を早く正確に知ることができたら,株価が予測できるはずだ.これが,この理論の根底にある暗黙の仮定である.スカート丈が株価を動かすのではなく,平均的な景況感の優れた指標になると主張している.他にも今までに,ヒット曲の曲調であるとか,甘い食べ物の売り上げや,人気女優の顔の造りまでも指標として提唱されてきた[Lewin 08].金融市場だけでなく,景気指標や選挙結果,商品の売上など様々な社会経済現象に対して,平均的な世の中の意見を表すような色々な指標がこれまでに提唱されてきた.

1.2 スカート丈からウェブデータへ

現代は,みんなの平均的な意見を知るために,スカート丈よりもずっと良い情報がある.Web上のマッシブデータである.ブログ,SNS,Twitterなどのソーシャルメディアサービスが普及したことで,現実世界の様々な情報が,リアルタイムにソーシャルメディア上に反映されるようになってきた.それに伴い,ソーシャルメディア上のデータを分析することで,社会経済現象の動向をリアルタイムに把握したり,将来の方向性を予測することに期待が集まっている.

例えば,ソーシャルメディア上のデータを分析し,株式市場や選挙結果,商品売上,景気動向の予測する技術を研究開発するウェブ関連の会社の取り組みが国内でもすでにある*1.また,金融市場に関してはtwitterやブログ,オンライン経済ニュースを株価予測に利用しているファンドが現れている.研究段階であれば,検索ログから米国のインフルエンザの症例数を予測したり[Ginsberg 09],雇用状況を予測したりしている[Finin 10].

連絡先: 和泉 潔, 東京大学大学院 工学系研究科 システム創成学専攻, 〒113-8656 文京区本郷 7-3-1, izumi@sys.t.u-tokyo.ac.jp

*1 <http://www.hottolink.co.jp/predict/>

Twitter データから株価予測を行ったり[Bollen 11, Ruiz 12], 映画の興行成績を予測する研究もある[Asur 10].

1.3 なぜウェブデータを利用した社会予測の精度が上がらないのか

しかし,これらのソーシャルメディア上のデータを利用した予測モデルは,ある程度の予測精度を示すことができるが,各分野の領域知識を利用した従来のモデルと比べると,さほど優っておらず新しい情報を加えることもないという主張がある[Hagedorn 07].例えば,音楽やビデオゲーム,映画のヒットの予測に関して,ソーシャルメディア上のデータによる予測よりもより単純でよく知られている指標の方が優れていた.

ソーシャルメディア上のデータによる予測精度が上がらない理由として,Gayoらは,ソーシャルメディアの利用ユーザ集団が一般人口の統計的な代表になっていないという問題を挙げている[Gayo-Avello 11b].ソーシャルメディアの利用ユーザの分布は一般人口に対して固定した偏りがあるので,今までの無作為抽出による調査手法と比べて結果にバイアスが出てくる.そのため彼らは,単に大規模な集団から生じたデータというだけで,その分析結果を何も考えなしで社会全体を代表していると考えることにに対して警告している[Gayo-Avello 11a].

本稿では,上述の代表性の問題以外に,ソーシャルメディア上のデータによる予測精度が上がらない理由として,社会経済現象におけるミクロ・マクロ関係のダイナミクスを挙げる.そして,その問題を解決する一つのアプローチとしてのエージェントシミュレーションの可能性を論じる.

2. 可能世界ブラウザとしてのマルチエージェントシミュレーション

2.1 ミクロ・マクロ・ループによる社会経済現象の構造変化

社会経済現象の予測が困難である原因は,社会経済現象にミクロ・マクロ・ループが存在するからである.社会経済現象は個人という要素から構成されている.社会全体でみられる現象の原因すべてを,各個人の意図や行動に還元することはできない.つまり社会経済現象は構成要素である個人の単純な総和ではない.したがって社会経済現象は個人より上位のレベルにあるという枠組みでモデルを作る方が妥当であろう.構成要素である個人に関することをミクロなレベルとする.また,個人の行動が集積した結果現れる社会経済現象全体の動きに関する

ことをマクロなレベルとする。

社会経済現象におけるマイクロ・マクロ・ループとは、個人の行動の集積が社会全体の動きを生成し、さらに社会全体の動きが個人の行動を変化させていくような循環のことを意味している。社会全体というマクロなレベルで観察すると、社会の挙動パターンは期間ごとに大きく構造を変化させているように見える。例えば、ある期間の社会経済現象はある特定の経済情報に対して敏感に反応して変動が大きくなる構造をもっていたりするが、別の期間の社会経済現象は情報には全く反応せず変動がとても小さい構造だったりする。このような構造の変化に適応しようとして、個人は学習により自分達の行動ルールを変化させていく。個人の行動ルールが変化することによって、社会経済現象全体の挙動はさらに新たな構造へと変化していく。このようなマイクロとマクロがお互いに相手を変化させながら自らも変化していく関係が、マイクロ・マクロ・ループである。

このようなマイクロ・マクロ・ループが存在する社会経済現象では、全ての期間に共通してマクロな挙動を予測できるような方程式を見つけることはとても困難である。そのため、ある特定時点のスナップショット的なマッシュデータから抽出した指標と、社会経済現象のマクロな動きとの間にある程度の規則性があるとしても、それらの規則性は時間とともに有効性が変動する。つまりマイクロ・マクロ関係を含んだモデルを構築しない限り、過去のマッシュデータから抽出した規則の単純な外挿だけでは、社会経済現象の予測は困難である。

2.2 マルチエージェントシミュレーションによるマイクロ・マクロ関係の分析

エージェントと呼ばれる社会における個人の役割をする計算機プログラムが多数集まって社会的な相互作用を行う、マルチエージェントモデルで社会経済現象のシミュレーション実験を行う研究がある。これらの研究は、マイクロ・マクロ関係が引き起こす構造変化のメカニズムを理解することを目指している。社会経済現象のエージェントシミュレーション研究の多くは、現象をリアルに再現することだけを主な目的とするシミュレーションとは異なり、構成論的シミュレーションと呼ばれるアプローチをとっている [Hashimoto 08]。現実の社会現象は、個人行動に対応するマイクロレベルや社会経済全体を表すマクロレベルの様々な要素が複雑に絡み合っている。そのため、マクロな挙動の原因を、少数のマイクロな要素のみに還元することは多くの場合不可能である。そこで、対象となる社会現象をシミュレーションで再現できるような、個人の行動・相互作用ルール・外部環境などの要素の条件を発見する。現象を生成できたマイクロな要素やその相互作用が、実際のマクロな社会現象の原因となっている可能性が高いと考えるのである。このようにして、社会経済システムにおける構造変化のメカニズムを、マイクロ・マクロ関係を観点から理解するのである。

2.3 マッシュデータから抽出した個人のルールをエージェントシミュレーションでテストする

それでは、個人のどの行動ルールが、対象になる社会現象と関係がありそうかというアイデア(仮説)は、どこから来るのであろうか。そのような思いつき自体を計算機が自動的に生成することはかなり困難である。最初の思いつきはやはり人間が見つかるしかない。マッシュデータの解析は、社会経済現象に関連しそうな個人行動のパターンを見つける時に大いに助けになる。年の情報通信技術の進歩により、社会経済全体のマクロなデータの取得だけでなく、社会経済的な様々な場面での人間の行動に関する詳細なデータも大量に集積できるようになった。例えば、個人の購買履歴に関しては、スーパーや

小売店舗で商品を購入すれば POS データとして記録・蓄積したり、RFID などのデバイスにより店舗内の回遊行動を計測できたりする。オンラインで買い物すれば、実際に買ったものに関するデータだけでなく、興味を持ったが実際には購入にまで至らなかった商品に関する行動まで記録可能である。また、金融取引の電子化によって、機関投資家やオンライン取引を利用する個人投資家に関して、金融商品の注文や成立した取引のデータを、秒単位で取得することが可能となっている。これらの行動データを分析した結果を、社会シミュレーションのエージェントの行動ルールを決定するのに利用できるになった。

マッシュデータの解析から得られた行動パターンと社会経済現象の関係に関するアイデアの有効性について、テストを行う必要がある。アイデアに基づいて社会モデルを決めてしまえば、後は計算機の膨大な処理能力を利用して、そのモデルで対象となる社会現象を再現できるか試すことが可能である。モデルの設定を変えて複数の条件でのテストを高速に行うことができ、モデルに入力するデータやモデルが出力した大量のデータも計算機で分析することができる。そしてエージェントシミュレーションは、現実世界では起こらなかった環境を設定することが可能であり、そのような架空の条件での社会経済現象の振る舞いも調べることができる。つまりエージェントシミュレーションは、社会現象に関して人間の思いついたアイデアの検証・テストを行うことができる。マイクロ・マクロ・ループによって構造を絶え間なく変化させている社会経済現象は、マクロなデータを分析しているだけでは予測困難であり、エージェントシミュレーションのようなアプローチが分析に必要な不可欠である。これらのマイクロとマクロの両レベルでの実データの間にある未発見の関係性(ミッシングリンク)を見つけ出すのが、エージェントシミュレーション研究の目的である。

2.4 可能世界ブラウザによる dependability の実現

社会経済現象について、マッシュデータから得られたマイクロルールを含むエージェントシミュレーションの使い方について、構成論的なアプローチ [池上 07] に基づいて議論したい。既存の社会シミュレーション研究では、モデルのパラメータを複数用意してシミュレーションを実行することが多い。そして例えば、シミュレーション結果全体の中で特定の結果がどれくらいの割合で含まれるか分析したり、複数条件でのシミュレーション結果の平均値を比較したりすることが、標準的な分析手法である。社会経済現象のエージェントシミュレーションの本当の価値は、このような結果全体の分析よりも、「ひとつでも面白い一本の軌道」を見つけてその軌道での「可能性」を論じる構成論的アプローチにあると考えている。

前節までの話だと、エージェントシミュレーションはマッシュデータの解析で得られたルールの集合をただ実行して現れる挙動だけを得るだけだと思われてしまうかもしれない。抽出したルールからシミュレーションの次の状態が決定論的に決められるという意味では、実際そのとおりである。しかし、大抵の社会モデルの中には非線形性が存在し、そのためシミュレーション結果に偶然性が生まれることが多い。その決定論的な確率の生成機構がカオスと呼ばれるものである。シミュレーションを同じ条件で複数回実行した時でも、カオスによって、社会モデルの挙動がいくつか異なるパス(軌道)として現れる。構成論的なアプローチとは、そのうちのある一つの興味深いパスを取り上げて、そこから何か新しい知見を得る手法を取ることが多い。シミュレーションの全てのパラメータを調べて、パス全体が作りだす性質を調べる鳥瞰図的な手法とは異なる。

構成論的アプローチに基づく社会シミュレーションの分析手法の一つとして、「可能世界ブラウザ」という手法を提唱した

い。エージェントシミュレーションの結果に偶然、興味深いパスが出てくるのを受動的に待っているだけでなく、積極的に興味あるパスを創りだそうとする試みである。まず最初に、初期状態がわずかに異なる複数条件のシミュレーションを行う。それにより、異なるパスを待つシミュレーション結果が作り出される。その中からある特定時点までの挙動が興味深いパス集団を選択する。当たり前だったり見たいこととの関係性が薄いパスは削除する。そして、残した部分パスを、先ほどの時点を初期条件として新たにシミュレーションパスを足していく。このようなシミュレーション結果の選択と融合により、興味深いパスをシミュレーションにより作り出していく。これにより例えば、金融市場シミュレーションで金融パブルが起きるパスを発生させたり、マーケティングシミュレーションである商品のマーケットシェアが急激に伸びるパスを発生させることが出来るかもしれない。このように、実世界では百万回に1回しか発生しないような、特異なケースや最悪・最善のケースを、積極的にシミュレーションで発生させるのである。

ここで重要なのは、シミュレーションの結果を作り出す力学系は、現実のマッシュデータから抽出したマイクロルールによって規定されていることである。これにより、単なるランダムウォークで偶然できたパスを選択しているのとは、全く意味が異なるのである。マッシュデータから規定される力学系から生じることが可能な世界をブラウジングさせるのが、ここでのエージェントシミュレーションの役目である。特定のマイクロルールの集合によって、自分の興味あるパスがどうしても発生させることが出来なければ、その興味ある現象は現在用いているマイクロルールからは構成できないのかもしれない。つまり、それらのマイクロルール集合はマクロ現象の原因となり得ないと言えるだろう。また、もし特定現象をシミュレーションで発生させることができたならば、エージェントシミュレーションの内部で何が起きたのかを調べることによって、ある特定現象が起こるとしたらどのようなプロセスで起こりうるのかということの詳細に調べることが可能である。

この可能世界ブラウザの考えは、もともと物理現象のCGで提唱された手法 many-worlds browsing [Twigg 07] を参考にしている。物理シミュレーションにおいて、パラメーターをいじることで期待したシミュレーションパスを作っていくのだが、広大なパラメータ空間で期待通りのパスを見つけることは困難である。そこで、数多くの物体の動きを少しだけ初期状態を変えて同時シミュレートする。生成されたシミュレーションパス集合を、ユーザの指示に応じてインタラクティブに補正・融合させ、最終的なアニメーションを生成する手法が提唱されている。これと同様に、社会シミュレーションのパスも合成しようという考えである。

可能世界ブラウザが特に有望な分野は、「ブラックスワン」現象の事前分析だろう。ブラックスワンとは、デリバティブトレーダーから認識論の研究者に転身した N・タレブが提唱している概念である [Taleb 08]。もともとブラックスワンとは黒鳥のことであり、1697年にオーストラリアで「黒い白鳥」が発見されたことが、世の中に大きな衝撃をもたらしたことを例えとしている。このことからタレブは、非常にまれで予測不可能だが甚大な影響をもたらす事態の事をブラックスワン現象と名付けた。特に社会経済現象では、金融市場のクラッシュなどのブラックスワン現象が様々な分野で見られる。ブラックスワン現象は現実世界で起こることは稀であるので、実際に起きたブラックスワン現象のデータを分析しても、個別的な状況にばかり気を取られ、その原因が後付的で個別的な説明のみに終始してしまうことが多い。実社会でのブラックスワン現象の事

後分析の多くが、特定の細かい原因探しに終わってしまい、システム全体で何が起きたのかという議論が少ないことから、事後分析の難しさがわかるだろう。

エージェントシミュレーションならば、実世界では百万回に1回しか発生しないようなブラックスワン現象を、何千・何万回も発生させることが可能である。そして、仮想世界で生じたブラックスワン現象のデータを解析することにより、ブラックスワンを事前に見ることが出来る。これによって、ブラックスワンの原因がどこから生じる可能性があるのか、ブラックスワンが起きたときにどのような対応を行えばその後の状況が変わるのかを論じることが出来る。このように、マッシュデータに基づくエージェントシミュレーションによって、ブラックスワンに対応する社会経済システムのデザインを行うことが出来るのではないだろうか。そして、「想定外のリスク」を想定することによって、社会経済システムの安全性や信頼性 (dependability) を向上するためことが期待できると信じている。

参考文献

- [Asur 10] Asur, S. and Huberman, B. A.: Predicting the Future with Social Media, in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT '10*, pp. 492-499 (2010)
- [Baardwijk 10] Baardwijk, van M. and Franses, P. H.: The hemline and the economy: is there any match?, Technical report, Econometric Institute, Erasmus University (2010), <http://publishing.eur.nl/ir/repub/asset/20147/EI%202010-40.pdf>
- [Bollen 11] Bollen, J. and Mao, H.: Twitter Mood as a Stock Market Predictor, *IEEE Computer*, Vol. 44, No. 10, pp. 91-94 (2011)
- [Finin 10] Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., and Dredze, M.: Annotating named entities in Twitter data with crowdsourcing, in *Proceedings of the NAACL Workshop on Creating Speech and Text Language Data With Amazon's Mechanical Turk*, Association for Computational Linguistics (2010)
- [Gayo-Avello 11a] Gayo-Avello, D.: A warning against converting social media into the next literary digest, in *Communications of the ACM* (2011)
- [Gayo-Avello 11b] Gayo-Avello, D., Metaxas, T., and Mustafaraj, E.: Limits of electoral predictions using twitter, in *International AAAI Conference on Weblogs and Social Media (posters)* (2011)
- [Ginsberg 09] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L.: Detecting influenza epidemics using search engine query data., *Nature*, Vol. 457, No. 7232, pp. 1012-1014 (2009)
- [Hagedorn 07] Hagedorn, B. A., Ciaramita, M., and Atserias, J.: World knowledge in broad-coverage information

filtering, in *Proceedings of the 30th annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 801–802 (2007)

[Hashimoto 08] Hashimoto, T., Sato, T., Nakatsuka, M., and Fujimoto, M.: Evolutionary Constructive Approach for Studying Dynamic Complex Systems,, in Petrone, G. and Cammarata, G. eds., *Recent Advances in Modelling and Simulation*, pp. 111–136, I-Tech Books (2008)

[池上 07] 池上 高志 : 動きが生命をつくる:生命と意識への構成論的アプローチ, 青土社 (2007)

[Lewin 08] Lewin, T.: The hemline index, updated, *International Herald Tribune* (2008)

[Ruiz 12] Ruiz, E. J., Hristidis, V., Castillo, C., Gionis, A., and Jaimes, A.: Correlating financial time series with micro-blogging activity, in *Proceedings of the fifth ACM international conference on Web search and data mining, WSDM '12*, pp. 513–522, ACM (2012)

[Taleb 08] Taleb, N.: *The black swan: the impact of the highly improbable*, Penguin Books (2008), (望月衛 訳, ブラック・スワン: 不確実性とリスクの本質, ダイヤモンド社,2009)

[Twigg 07] Twigg, C. D. and James, D. L.: Many-worlds browsing for control of multibody dynamics, *ACM Trans. Graph.*, Vol. 26, No. 3 (2007)