

平成30年 6月29日現在

機関番号：24402

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K01195

研究課題名(和文)人工市場によるアルゴリズム取引の影響分析と制度評価環境の開発

研究課題名(英文)Artificial Financial Market to analyze algorithm trade and institution

研究代表者

中島 義裕 (Nakajima, Yoshihiro)

大阪市立大学・大学院経済学研究科・教授

研究者番号：40336798

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、高頻度アルゴリズム取引(HFT)が証券市場にどのような影響を及ぼしたのかを明らかにするため、オーダーフローを算出するツールを改良し、2007年と2012年のデータを比較する実験枠組みを完成させた。その最初の試みとして、イベント数の比較をおこなった。また、機械学習を用いた株価やオーダーフローの分析を行うため、CNNによる株価の分類や、4種類の機械学習による投資パフォーマンスの比較を行った。

研究成果の概要(英文)：We develop analytical tools of order flow and compare between data of 2007 and 2012. Because TSE Arrowhead started 2010, we can find effects of HFT. At the first step, we count the number of events (including market orders, limit orders and cancel). The number of event increase 5 times from 2007 to 2012, but half of brands decrease the number of events. On the other hand, 22% of the brands increase more than ten times. It suggests that HFT trades mainly in a part of market. Next, We try to use Machine Learning especially Deep learning to apply market analysis. We developed four kinds of models, and observe their profit. Then we find there is no strongest model in these four models.

研究分野：計算機経済学

キーワード：人工市場 高頻度アルゴリズム取引 人工知能 機械学習 深層学習

1. 研究開始当初の背景

(1) **アルゴリズム取引の増加と高速化による時間スケールの多様性の増加**

株式や為替市場などの金融市場において、急速に、アルゴリズム取引が存在感を増している。東京証券取引所では、2010年1月に、ミリ秒単位のレスポンスを行う新株式売買システム Arrowhead を稼働させた。Arrowhead 開始時点のコロケーション（取引所内のサーバ設置サービス）経由の注文件数は10%程度であったが、約1年後には注文件数の35%、2014年には注文件数の60%を超えるまでになっている。欧米でも2005年段階で20%（米）、3%（欧）程度だったものが、2011年には70%（米）、40%（欧）へと増大したことが報告されている。短期間で状況が変化したため、アルゴリズム取引は大きな関心を寄せられているものの、あまり研究は進んでいない。アルゴリズム取引は、市場の混乱や不正取引の疑いなどによって注目された事もあり、当初、規制の観点からの研究が先行した。2010年に、NYダウが瞬時に9%下落し、世界中の多くの市場に混乱が波及した flash crash と呼ばれる事件が発生した。その後、発生、下落、回復の各プロセスが詳細に分析された。類似した現象は、個別の銘柄などで頻発しており、市場の攪乱要因になっていると指摘されている。(SEC, CFTC 2010, Kirilenko et al 2010, 2012 等) また、アルゴリズム取引が瞬時に指し値注文やキャンセルを繰り返し、株価操縦を行っている可能性があること(ESMA 2011)や、情報の遅延を利用して利益を上げていること(Lewis 2014)が報告されるなど、不正な取引が行われているとの批判もある。一方で、(1)主に裁定取引や大口注文の分散化を行っていること(Brogaard2010, Castura et al 2010)や、(2)シェアの拡大前後のデータを比較し、bit-ask spread が狭まる、ボラティリティが低下する、マーケットインパクトが減少するなど、市場の流動性を高め、市場の安定化に寄与している(Hasbrouck et al 2012, Menkveld 2012, 保坂 2014, Vella 2014)など、ポジティブな面も指摘されている。

こうした研究の中で、どのような種類のアルゴリズム取引が市場参加しているのか推測され類型化されつつある。SECは、受動的マーケット・メイク戦略、裁定取引戦略、構造戦略、価格施行性戦略に分けられる(SEC 2010)と報告している。しかし、これまでは価格変動や市場構造、個別の事件の詳細な分析にとどまっており、個々の注文情報（オーダーフロー）から、アルゴリズム取引の特性を見いだす研究は端緒についたばかりである。Brogaardは、こうした方法でマーケット・メイク型/テイク型に分けた上で、市場の流動性へのマイナスの影響とプラスの影響について分析した(Brogaard 2014)。現在、欧米を始めとした金融市場で実施また

は検討されている対策は数多く、サーキット・ブレーカー制度の統一、情報遅延の導入、高頻度バッチ・オークションの導入などが挙げられる。どの対策にせよ、人間の認識/判断能力を遥かに超えた時間・空間スケールで事態が変化することから、アルゴリズムを用いたものにならざるを得ない。これら対策の検証には、人工市場の利用が不可欠であり、そのための実験環境の提供が急務である。しかし、その前提となる人間とアルゴリズム取引が混在する市場を持つ特性そのものも、まだほとんど明らかになっていない。

(2) **ディープラーニングの衝撃とAIブーム**

現在証券市場ではHFTが席卷しており、それらにディープラーニングが取り入れられつつある。従来のHFTは、概ね極めて単純なアルゴリズムであり、HFTがなぜその注文を出したのか、その理由を調べることができる。また、HFTの性能を向上させるためには、人間がアルゴリズムを変更する必要がある。それに対してディープラーニングの特徴は表現学習にある。経験を通して自ら学習するという人間の習熟に似ており、データや人工知能の規模（ディープラーニングを構成するノード数や階層数）に応じて、性能を引き上げられる。また、学習はノード間ネットワークの結びつきの強さという形でなされるため、自律的AIトレーダーはなぜ、そのような発注をするのか人間が理解するのは困難である。このような自律的なAIトレーダーが、自身で性能を向上させたときに何が生じるのか。この問題にアプローチするためには、実際にシミュレーションしてみるしかない。すでに現実の方が先行しており、その予想や検証がなされていないという現状は極めて危険である。

2. 研究の目的

本研究の目的は、まだアルゴリズム取引、取分けミリ秒単位で注文を出すHFTが存在していない2007年以前と東証アローヘッドが稼働して本格的にHFTが参加し始めた2010年以降のオーダーフローを実証的に比較分析すると共に、両者の市場状況を人工市場に再現するという構成論的方法により、板ダイナミクスのモデル化を行うことにある。モデル化を通じて両者の相違を明らかにすることで、人間の相互作用と、アルゴリズムの相互作用がもたらす違いと、その主従関係が交代してもなお保持される市場の特性を明らかにすることを目的とする。

また、実証分析を行う際に様々な機械学習を用いることにより、新しい観点からの市場の特性を明らかにする。

3. 研究の方法

データとしては日経NEEDS社が提供する株式ティックデータを用いる。このデータを加工しオーダーフローを求める。株価やオーダーフローの分析に機械学習を用いる。

#### 4. 研究成果

##### (1) オーダーフローを求めるツールの開発と2007年と2012年の市場の変化

日経 NEEDS 社が提供する株式ティックデータは、市場（東証1部、2部、マザーズ）毎に1日の全ての板の状態が1つのファイルに収められている。このデータから個々の銘柄のオーダーフローを切り出すには、銘柄毎の板情報に分けた上で、板の変化と出来高の変化を調べた上で、約定による板の変化と注文の取り消しによる板の変化を区別し、それらが矛盾しないように分類することが必要である。私たちは、これらの処理を行うツールを開発した。

図1は、東証1部の全銘柄について1日に発生したイベント総数（成行注文、指値注文、キャンセルをトータルしたもの）について2007年の1年間の平均値と2012年の平均値を求め、プロットしたものである。X軸は2007年の1日当たりの平均イベント数、Y軸は2012年の平均イベント数を示している。

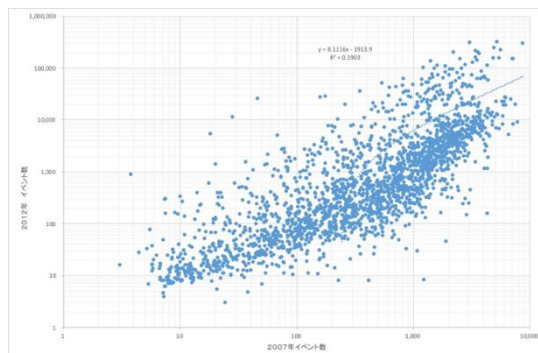


図1：2007年と2012年のイベント数の変化

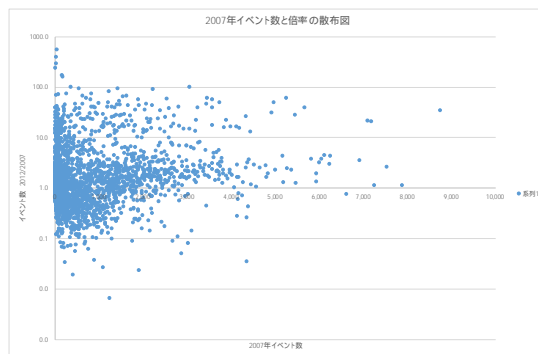


図2：2007年と2012年のイベント数の変化（倍率）

2007年当時、東証では注文が執行されるまでの時間は最短で2秒程度のオーダーであった。それがアローヘッド稼働後の2010年には2/1000秒、2012年には1/1000秒になっている。実に2000倍も処理能力が増加している。それに伴って、この処理速度に合わせて発注、キャンセルなどを行うHFTが増える事で、1日のイベント数も大幅に増加すると予想される。2007年の東証1部の全銘柄の1

日の平均イベント数は910イベントであり、2012年では5488イベントであった。個々の銘柄について、2012年のイベント数が2007年に比べて何倍になったのかを調べてみると平均は5.1倍であった。また、全2186銘柄の中で、イベント数が10倍以上になった銘柄が485(22.2%)あったのに対し、918銘柄(42.0%)は2012年のイベント数が2007年に比べて下落していた。

通常より、市場内では銘柄による人気の違いが大きく、絶えず注文が入り価格が変動し続ける銘柄もあれば、閑散としていて滅多に新しい注文が入らない銘柄もあることが観測される。この意味で、処理速度が高速化し、HFTが多く参加したことにより、イベント数の意味で目に見える影響があったのは市場の20%程度の銘柄に過ぎないことがわかった。もちろん、閑散としている銘柄においても、多くの取引アルゴリズムが監視しており、何らかのミスプライシングが生じた際には、即座に注文が入るといったような影響はある。今後は、個々の銘柄の違いを詳細に観察し、HFTの影響を分析する必要がある。

##### (2) 深層学習による2007年と2012年の価格変動の識別

先の分析はイベント数という定量的な変化を扱ったが、HFTの影響を調べるためには価格変動のパターンに違いが生じているのが調べる必要がある。ここでは、株価チャートを「図」として見たときに、画像識別で良いパフォーマンスを見せている畳み込みニューラルネットワーク(Convolutional Neural Network: CNN)によって、それが2007年のものであるのか、2012年のものであるのか識別可能かどうかを調べた。

各銘柄の1分足を1日分プロットした株価チャートを20枚(約1ヶ月分に相当)シート状に並べて1枚の画像とした。12ヶ月分のシートの中からランダムに9枚選んで学習し、残りの3枚を判別テストに用いる「交差検定」を行った。比較的イベント数が多い大型株、中型株、小型株としてソフトバンク、レカム、ガーラの3銘柄を選んだ。最初に各年のデータを用いて、この3銘柄の違いが識別できるのか調べた。その結果、学習データに関する識別は100%正答し、テストデータに関する識別率は58.3%であった。これは、ベースラインである1/3を大幅に上回っていることから、ある程度は識別可能であることを示している。一方で、同じ銘柄について、そのグラフが2007年のものが2012年のものを識別して見た。これも学習データについては100%の識別率であったが、テストデータについては50%前後の識別率であり、識別には失敗した。これは、学習データによる過学習が生じているためであると考えられる。

##### (3) 様々な機械学習による株価分析

先の実験により機械学習、特に深層学習の株価分析への応用の可能性と困難さが同時に得られた。CNN は画像を対象とした機械学習であるが、時系列データを学習する LSTM (Long Short Term Memory) など、多くのアルゴリズムが提案されている。本格的な分析に入る前に、それらのアルゴリズムが株価分析にどのように有効であるのか、計測する必要がある。

そこで、上記の CNN と LSTM に加えて従来から広く用いられている SVR (Support Vector Regression) と MLP (Multi-Layer Perceptron) の4つを比較した。それらのアルゴリズムによる株価予測を行うが、株価予想の場合、正答率としての予測精度よりも、その予想に基づく仮想売買を行い、パフォーマンスを比較する方が、株価の特徴やアルゴリズムの特徴を把握しやすい。そのため、上昇トレンドにあると予想される間、株を保有し、下落トレンドにあると予想される間は、現金を保有することにした。

本実験ではソフトバンクグループ(株)(銘柄コード:9984) の分足データを実験に用いた。訓練データとしては 2017/7/3 から 7/19 の期間のデータ、テストデータとしては 2017/7/20、7/24 のデータを用いた。

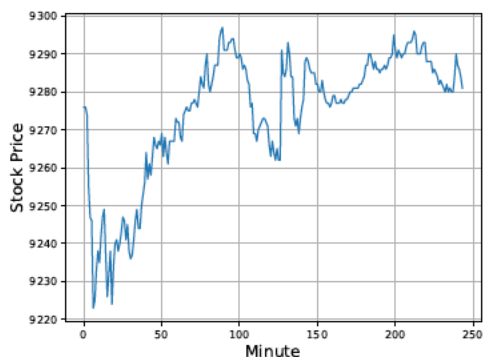


図3 2017/7/20 のソフトバンクの1分足

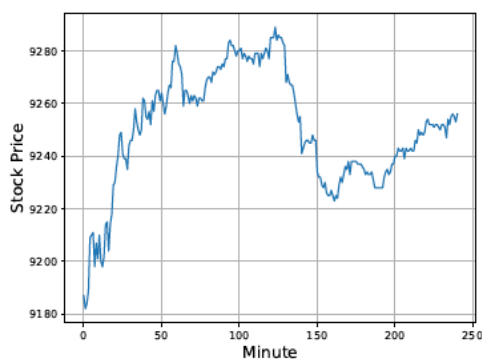


図4 2017/7/24 のソフトバンクの1分足

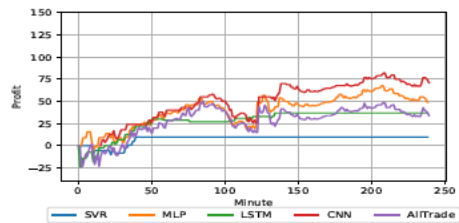


図5 獲得金額の推移(2017/7/20のデータ)

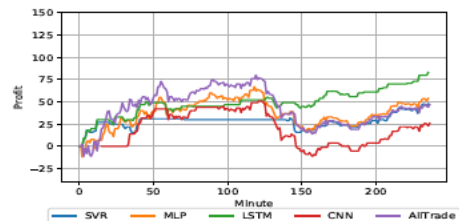


図6 獲得金額の推移(2017/7/24のデータ)

図5と図6を比較すると、例えばCNNは7/20のデータでは、最も利益を上げているのに対し、7/24では、最も利益が少ない。また、SVRは、7/20も7/24も直線部分が長い(すなわち下落すると予測し、株を保有していない)という傾向があることがわかる。また、時系列データに強いと言われているLSTMであるが、7/24のように大きなトレンドの違いが出るような場合には、そのトレンド変化を適切に捉えていることがわかる。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

(1) Romic Ivan, Nakajima Yoshihiro, "Ecosystem engineering as an energy transfer process: a simple agent-based model", *Theoretical Ecology*, Vol. 10, 2017, pp. 1-13, <https://doi.org/10.1007/s12080-017-0357-9>

〔学会発表〕(計16件)

(1) 小野功、森直樹、中島義裕「U-Mart Toolkit: 人工先物市場から汎用市場シミュレータへ」計測自動制御学会 第15回社会システム部会研究会, 2018年3月16日, 沖縄県 沖縄産業振興センター

(2) 中島義裕、小野功、森直樹「文駆動型市場上のシステミック・リスク・シミュレーション」計測自動制御学会 第15回社会システム部会研究会, 2018年3月16日, 沖縄県 沖縄産業振興センター

(3) K. Koshida, N. Mori, K. Matsumoto, and Y. Nakajima, Application of Machine Learning to the Stock Price Forecasting Problem, 6th Asian Conference on Information Systems, 2017, 12-14th December 2017, Sokha Hotel Phnom Penh, Cambodia

(4) 渡邊順一郎, 森大典, 森直樹, 松本啓之亮, 「機械学習を用いた株式取引エージェントの進化」電気学会 産業計測制御研究会, 2017年12月7日, 岩手県 岩手県立大学

(5) 渡邊順一郎, 森大典, 森直樹, 松本啓之亮, 「深層強化学習を用いた株式取引エージェントの進化」, 電気学会 電子・情報・システム部門大会, 2017年9月6日, 香川県 サポートホール高松

(6) 越田恭平, 森直樹, 松本啓之亮, 「LSTMを用いた株価変動予測手法の提案」, 電気学会 電子・情報・システム部門大会, 2017年9月6日, 香川県 サポートホール高松

(7) 越田恭平, 森直樹, 松本啓之亮, 「機械学習を用いた株価の変動予測における有効な指標の選別」, 計測自動制御学会関西支部・システム制御情報学会 若手研究発表会, 2017年01月13日, 大阪府吹田市 大阪大学

(8) 渡邊順一郎, 森大典, 森直樹, 松本啓之亮, 「遺伝的プログラミングを用いたデイトレード戦略獲得手法の拡張」, 計測自動制御学会関西支部・システム制御情報学会 若手研究発表会, 2017年01月13日, 大阪府吹田市 大阪大学

(9) 住田和也, 松本啓之亮, 森直樹, 「株式市場における人工市場と現実市場の類似度指標についての考察」, 電気学会 C 部門大会, 2016年08月31日, 兵庫県神戸市 神戸大学

(10) 渡邊順一郎, 森直樹, 松本啓之亮, 「遺伝的プログラミングを導入した株式取引エージェントの応用」, 電気学会 C 部門大会, 2016年08月31日, 兵庫県神戸市 神戸大学

(11) 渡邊順一郎, 三浦秀之, 森直樹, 松本啓之亮, 「デイトレード戦略獲得のための遺伝的プログラミングの応用」, 第60回システム制御情報学会研究発表講演会(SCI'16), 2016年05月25日, 京都府京都市 京都テルサ

(12) 住田和也, 松本啓之亮, 森直樹, 「株価データの解析における DeepLearning の導入」, 第60回システム制御情報学会研究発表講演会(SCI'16), 2016年05月25日, 京都府京都市 京都テルサ

(13) 中島義裕, 森直樹, 住田知也「証券市場におけるアルゴリズム取引の影響」, 第20回進化経済学会 東京大会, 2016年3月26日, 東京大学 駒場キャンパス

(14) Ivan Romic, Yoshihiro Nakajima, "Ecosystem Engineering in Natural and Social Systems", JAFEE 20th Annual Meeting , International Conference on Socio-economic systems with ICT and Networks, 2016年3月27日, 東京大学 駒場キャンパス

(15) 住田和也, 松本啓之亮, 森直樹, 「人工株式市場分析のための標準エージェントセットに関する一考察」, 第59回システム制御情報学会研究発表講演会, 2015年05月20日, 中央電気倶楽部(大阪府・大阪市)

(16) 三浦秀之, 森直樹, 松本啓之亮, 「進化的な株式戦略獲得法へのサポートベクターマシンの導入」, 第59回システム制御情報学会研究発表講演会, 2015年05月20日, 中央電気倶楽部(大阪府・大阪市)

〔図書〕(計1件)

(1) Ed. Hajime Kita, Kazuhisa Taniguchi, Yoshihiro Nakajima, "Realistic Simulation of Financial Market", Springer, 2016年, 202ページ. (Chap. 4 Naoki Mori, "Evolution of Day Trade Agent Strategy by Means of Genetic Programming with Machine Learning" pp. 97-115, Chap. 5 Yoshihiro Nakajima, "How to Estimate Market Maker Models in an Artificial Market", pp. 117-135.)

〔産業財産権〕

出願状況(計0件)

取得状況(計0件)

〔その他〕  
ホームページ等

6. 研究組織

(1)研究代表者

中島 義裕 (NAKAJIMA YOSHIHIRO)  
大阪市立大学・大学院経済学研究科・教授  
研究者番号: 40336798

(2)研究分担者

森 直樹 (MORI NAOKI)  
大阪府立大学・工学研究科・准教授  
研究者番号： 90295717

(3)連携研究者

(4)研究協力者