

解題

証券アナリストジャーナル編集委員会

第一小委員会委員 伊藤 敬介 CMA・CIIA

ここ数年、「ビッグデータ」という単語をよく耳にする。例えばインテルは、ビッグデータの活用で2013年に約80億円の効果を上げたとされ（注1）、また米国のウォルマート・ストアーズは顧客との取引記録を基に1時間当たり2.5ペタバイト超（1ペタバイトは約100万ギガバイト）のビッグデータを収集し（注2）、顧客の購買行動パターンを分析して商品の陳列方法を調整するなどしているという。

そもそも、ビッグデータと従来型のデータとで何が違うのであろうか？ ビッグデータに厳格な定義はないが、もともとは分析対象のデータが増えすぎてコンピューターのメモリに収まり切らなくなり、分析するには複数のマシンに分散して処理するなどのツール改良が必要になったことから、従来ツールで分析できないほど大きなデータ、すなわち「ビッグデータ」と呼ぶようになった、と言われている。そしてビッグデータと従来型データの違いとして、3つのV、すなわち「Volume（データ量）」、「Variety（データの多様性）」、「Velocity（データの発生頻度）」が挙げられる（図表1）。インターネット上の情報やニュー

ス配信、ソーシャルネットワーク上のテキストや画像、ビデオ、機械に取り付けたセンサーからの情報など、膨大な情報が該当する。

このように大量のデータが与えられた場合、その分析手法もおのずと変わってくる。従来型データの世界では、比較的質の高い管理された観測サンプルに対して「どのような因果関係から発生したか」という仮説を立て、その仮説が成り立つかを検証して結論を導く、という側面が強い。これに対しビッグデータの世界では、母集団データと思えるような膨大なデータを用いて探索的なデータ解析を行い、未知の知見を発見的に見いだすような側面が強い。つまりビッグデータの世界では、因果関係に関する仮説がなくても、探索的な解析手法や機械学習により答えをあぶり出せばよい、という発想で分析する事が多いのである。ご参考までに、ビッグデータの世界で用いられることの多い技術の例を図表2に示す。

ビッグデータはマーケティングや機械制御などさまざまな分野で活用されつつあり、マネジメント・スタイルを一変させるとも言われているが、

（注1） 日経新聞電子版2014年4月1日付記事。

（注2） McAfee, A. and E. Brynjolfsson [2012], “Big Data: The Management Revolution,” Harvard Business Review

図表 1 従来型データとビッグデータ

	従来型データ	ビッグデータ
Volume (データ量)	数メガバイト～数テラバイト※	100テラバイト以上、など
Variety (データの多様性)	構造化データ (数値データなど)	構造化データに加え、非構造化データも (文章、画像、音声など)
Velocity (データの発生頻度)	静的。バッチ処理によるデータ更新など	動的。リアルタイム更新など
データの位置付け	管理された少数の観測サンプル	母集団データと思えるような膨大なデータ
分析手法	仮説検証	機械学習／探索的データ解析

(図表注) ※1 テラバイトは、1,024ギガバイト。

(出所) 筆者作成

図表 2 ビッグデータ技術の例

技術	定義
Hadoop (ハドゥープ)	複数のサーバー上で並行してビッグデータを処理することが可能な、オープンソースのソフトウェア
MapReduce (マップリデュース)	Hadoopが依拠するアーキテクチャ
機械学習	あるデータセットに最も適合するモデルを高速で発見する技術。またそのためのアルゴリズムやソフトウェア
自然言語処理	文章を解析する技術(単語の出現頻度や文章の意味など)。またそのためのソフトウェア
インメモリ分析	コンピューターのメモリ上でビッグデータを高速で分析する技術

(出所) トーマス・ダベンポート著、小林啓倫訳、[2014]「データ・アナリティクス3.0」(日経BP)をもとに筆者作成

実は証券市場でも近年ビッグデータの活用が進みつつある。例えば株式トレーディングの世界では、証券取引所の高速取引対応に伴い、ミリ秒やマイクロ秒単位で発注を繰り返すような高頻度取引(HFT; High Frequency Trading)が広まりつつあり、この出来高シェアは日本でも50%を超えたとされる。このようなHFTデータは、構造化データ(数値データ)ではあるものの、膨大なデータ量の「ビッグデータ」である。また、株式アクティブ運用の世界でも、ニューステキストやWebデータ、決算発表やアナリストレポートなどのテキストデータ(非構造化データ)を活用する事例

が増えつつある。

当ジャーナルでも、これまで岡田・羽室[2011](注3)、岡田 *et al.* [2013](注4)、石田[2014](注5)など、テキストマイニングやビッグデータをテーマとした個別論文を掲載してきたが、本号では特集テーマを「ビッグデータとHFT」とし、以下の4本の論文を掲載する。

最初の諏訪部論文「データ革命と株式運用戦略」は、ビッグデータの活用が株式投資にどのような変化をもたらし、どのような課題が存在するかを考察する興味深い論文である。ビッグデータがさ

(注3) 岡田克彦・羽室行信 [2011]「相場の感情とその変動ー自然言語処理で測定するマーケットセンチメントとボラティリティー」、『証券アナリストジャーナル』、49 (8)、pp.37-48.

(注4) 岡田克彦・山崎高弘・榎原茂樹・山崎尚志 [2013]、「株価変動の季節性と投資家心理ー新聞記事に見る将来見通しとデカンショ節効果ー」、『証券アナリストジャーナル』、51 (12)、pp.96-105.

(注5) 石田隆 [2014]、「検索頻度データによる株式市場リターン予測可能性ービッグデータによる経済分析のサーベイとその実証応用例ー」、『証券アナリストジャーナル』、52 (6)、pp.83-93.

さまざまな分野で活用され、付加価値創造に寄与しつつある現状を紹介した上で、ビッグデータの特徴を3つのV、つまりVolume、Variety、Velocityの観点で整理している。特にVariety（多様性）の解説では、例えば決算発表時の経営者のスピーチの音声データを発声感情分析ソフトで分析し、声のトーンに現れた不安感から不正会計の予測を試みるなどの先行研究を紹介しており、興味深い。

ビッグデータの解析手法としては、教師信号を与えた機械学習アプローチや、決定木・ニューラルネットワークといった非線形アルゴリズムなどさまざまなものがあることを示し、例えばテキストデータから株式市場で現在どのようなテーマが注目されているかを判別し、ポートフォリオを構築した事例の紹介や、アナリストレポートに隠されたニュアンスの変化を機械学習により読み取ることで、センチメントの強弱を判別するアルゴリズムの紹介を行っている。

本格的なビッグデータの活用には運用チームの編成、データ・モデルの管理プロセス再構築等、多くの課題が存在する事も指摘。そして、株式運用でもビッグデータを用いることで、従来は人間にしかできないと思われたことをコンピュータアルゴリズムに代用させることができる時代が近づいているのかもしれない、多くの株式運用者が何らかの形でビッグデータを利用する日は近いのではないかと結んでいる。

続く宇野論文「HFTと流動性—グローバルな視点から—」は、HFT（高頻度取引）の特徴や役割、そして課題を鳥瞰する論文と言えよう。HFTの市場出来高シェアが米国で49%、日本でも50%超となっていること、HFTをマーケットメーカー（流動性供給）型とテーク（流動性需要）型に区別して理解すべきであること、などを指摘し、「ミリ秒」

（千分の1秒）や「マイクロ秒」（百万分の1秒）という単位で高速に取引を行うHFTが流動性供給の世界を劇的に変化させたことを紹介している。

HFTと類似した概念としてアルゴリズム取引（AT）があり、いずれもコンピューターによる自動執行を前提としているが、ATは機関投資家などが、VWAP取引など大口注文の取引コスト抑制のために用いることが多く、秒単位の執行スピードが前提となっているのに対し、HFTは証券会社の自己勘定部門などが薄利を積み上げる戦略として用いることが多く、ミリ秒レベルの執行スピードが必須となる点が異なる、としている。

また執行スピードを上げるには、取引所のメーソ・コンピューターのすぐ近くに発注サーバーを設置する「コロケーション・サービス」の利用と、高速コンピューターへの投資が重要となるが、そのような投資をするのは指値注文で流動性を供給する投資家（マーケットメーカー型）であることを実証分析した先行研究を紹介している。

そしてHFTが流動性供給の側面を持つ一方で、市場のボラティリティを高めるという批判もあり、これを背景にHFT規制がさまざまな国で検討されていること、HFTの存在感が高まる中で流動性供給のメカニズムを深く理解し、市場運営のあり方を考える必要があることを指摘している。

林論文「高頻度注文板データによる2014年東証ティックサイズ変更の国内株式市場への影響分析」は、東京証券取引所（東証）が14年1月と7月に呼値の刻み（ティックサイズ）の縮小をしたことに伴い、東証や私設取引システム（PTS；Proprietary Trading System）の間でどのような市場構造変化が起こったかを、ミリ秒刻みの高頻度データを用いて実証分析する論文である。

東証では10年に新株式取引システム「アロー

ヘッド」が稼働して以降、HFTの本格的な利用が始まっており、また当論文で東証と比較するために用いたPTSは、チャイエックス・ジャパン（ChiX Japan）とSBIジャパンネクスト証券（JNX）の2つである。

分析の結果、14年の東証のティックサイズ縮小後に、東証、ChiX Japan、そしてJNXの3市場間でさまざまな変化が確認されたとしている。まず、銘柄のビッドアスク・スプレッドであるが、従前は3市場とも同程度の水準で推移していたのが、東証のティックサイズ縮小後は、東証のスプレッドが他の2市場と比べ大幅に縮小した、としている。また取引シェアも東証が大きく伸ばしたことを確認している。

さらに3市場間の価格形成（気配値更新）の先行運行関係を分析すると、東証のティックサイズ縮小後はJNXに対し東証は相対スピードを速め、ChiX Japanはその東証に対してさらに相対スピードを速めたことを観測している。このようなビッグデータの分析を常時行うことで、執行コストの削減や市場センチメントの把握など、さまざまな応用が考えられよう。

最後の大崎論文「HFTの規制をめぐる米国の動向と日本市場への示唆」は、単に規制の動向を論じるだけでなく、そもそもHFTが生み出される背景となった米国の株式市場構造の変化やHFTに対する誤解など、HFTの実態とその規制の変遷を多面的に解説し、これらがわが国のHFT市場にどのような示唆を与えるかを考察する論文である。

1990年代初めまで、NYSE上場銘柄の取引金額の90%以上がNYSEで取引されていたのに対し、現在は20%程度しかNYSEで取引されていない。この背景には、注文取扱ルールの制定やトレードスルー規制の見直し、市場間の裁定取引の活発化、

呼値の10進法への移行などさまざまな要因が重なり、取引市場の多様化が進み、その結果マーケットメイクもHFTを駆使するスタイルへと移行したこと、などを解説しており、興味深い。

また10年5月に、ダウ平均が5分間で573ドル暴落し、その後1分半で543ドル暴騰した、いわゆる「フラッシュ・クラッシュ」では、HFTが混乱の原因ではないかという一部の批判があったが、実はHFTは最初の大口売り注文に買い向かい、マーケット・インパクトを緩和する機能すら果たした面があることを紹介している。

日本では、10年のアローヘッド稼働以降、HFTの利用が進みつつあるが、米国のような市場の分裂は見られていないため、米国におけるHFTの規制をめぐる議論をそのまま日本に当てはめるのはミスリーディングであること、また市場間競争を促すことで市場全体の機能を高めようという米国規制当局の姿勢やそれに支えられた米国市場のダイナミズムからわが国が学ぶべき点も多いこと、などを指摘している。

今後も利用可能なデータやその分析技術が日進月歩で進化し、「ビッグデータ」や「HFT」の利用により市場構造や投資家行動の解明、そしてそれらを踏まえた投資戦略の策定や効率的な売買執行が着実に進むことが予想される。これはすなわち、より透明でコストが安く、完備な資本市場の形成につながり得る話であり、社会的な意義は大きい。同時に、金融・証券市場に携わるわれわれ自身、このような技術や情報の活用が、われわれ自身の付加価値を大きく左右する時代になりつつあることを示唆する。本号の特集が、幅広い投資家、および金融・資産運用業界に携わるわれわれにどのようなインプリケーションをもたらすかを改めて考えるきっかけとなれば幸いである。