

解題

証券アナリストジャーナル編集委員会

第二小委員会委員長 菅原周一 CMA

株式時価総額にどれほどの意味があるのかは議論の分かれるところであるが、2018年8月2日にアップルの株式時価総額が初めて1兆ドルを超え、アマゾンも9月4日に1兆ドルを超えたことは、世界的に大きな変化が起きていることを表す象徴的な出来事であったと言ってよいであろう。過去の株式時価総額ランキングを見ると、例えば07年末時点では、エクソンモービル、ペトロチャイナ、GEがトップスリーで石油関連会社が2社入っていたが、17年末になると、アップル、アルファベット（グーグルの親会社）、マイクロソフトがトップスリーとなっている（第4位がアマゾン、第5位がフェイスブックであった）。世界トップ上位5社を全てIT関連企業で独占し、これらのうち、マイクロソフトを除いたグーグル、アップル、フェイスブック、アマゾン・ドット・コムはその頭文字を並べてGAFAと呼ばれ、SNS、検索サイト、通販などの新しい事業を通じて、世界中の利用者の大量データを収集、蓄積、分析し、これまでにないサービスや製品を生み出し続けている。IT分野に積極的な中国の代表的IT企業であるバaidu、アリババ集団、テンセントの3社（その頭文字を並べてBATと呼ばれている）も、SNS、検索サイト、通販などの新しい事業を通じて巨額の利益を上げている。ITを駆使してネ

ット上でビッグデータを活用してビジネス展開する企業は「プラットフォーマー」と呼ばれ、GAFAやBATはその代表的企業であり、投資家からも成長への大きな期待がかけられている。

「データは21世紀の石油である」という指摘がどこまで将来を予言する適切な表現であるかは分からないが、ビッグデータをハンドリングする環境が整った今、世界は新たなステージに進んだことは間違いないであろう。情報は宝の山であり、経済的な成長だけでなく社会全体に大きな変化をもたらすものとなる可能性がある。これを新たな資源と呼ぶ人がいるが、顧客情報に代表されるデータは時間の経過とともに劣化する一方で、データ量が急激に増え続けているという点で資源とは異なる側面を持っている。あらゆるものがインターネットにつながり、ハードウェアも進化し、大量のデータ（ビッグデータ）を蓄積することができるようになることで、医療、物流の世界が大きく変わり、マーケティングの方法も大きく変わることが予測される。日本のように地方の隅々まで情報通信インフラが整っていれば（超高速ブロードバンドが利用できる世帯は15年時点で99.98%、光ファイバーの契約数の割合はOECD加盟国中1位）、少子高齢化の対策として、地方活性化に生かすことも考えられる。

金融の世界でも、ビッグデータを活用することで大きな変化が生まれつつある。例えば、センチメントやメディアの影響がどのように資産価格に影響を与えるかという問題意識を持った研究が数多く行われており、新たな研究分野も生まれている。例えば、18年1月5日に開催されたAmerican Economic Associationアニュアルミーティングにおいて、“Big Data and the Cross-section of Stock Returns”というペーパーセッションが設けられたほどである。

そこで、今月号の特集では、ビッグデータと金融に関する最新の研究成果を取り上げることとした。注目されている分野だけに、最近では興味深い研究成果が報告され、今後も期待されている。今回はそれらの中でも、最近の研究成果として4本の論文を紹介することとした。

1本目は、**佐々木論文「企業間データ連携が生み出す価値の可能性」**である。この論文では、ヤフー(株)におけるデータの活用方法の現状を解説し、垣根を越えた企業間のデータ連携により多面的にユーザーを理解し価値を提供できることを明らかにするとともに、(森の実りを動物や植物が分け合って森が成長していくエコシステムになぞらえた)データフォレスト構想を推進していることを報告している。

具体的には、ヤフー(株)におけるデータの活用方法について三本柱である「ビッグデータ」、「処理基盤」、「サイエンス」をヤフー(株)及び関係会社がどのように活用しているかについて解説し、次にヤフー(株)が抱えるデータ活用時の障害となっている「サイロ化」問題の解決策を紹介し、最後にヤフーグループの垣根を越えた企業、自治体とのデータ共有ならびに活用による新たな試みであるデータフォレスト構想について、自動車メーカー、

食品メーカー、スポーツ産業等の実例を紹介している。

この構想は一企業の目先の利益優先を超えたチャレンジングな試みであり、人々の生活を更に豊かにする可能性を秘めている。

2本目は菅・高橋論文「高頻度データを通じたニュースと株式市場の関連性の分析」である。この論文では、東京証券取引所一部上場企業を対象とした大規模データの一つである高頻度データを使い、近年進展してきた情報技術の一つである深層学習を利用して、ニュース記事の内容と株価変動の関連性について分析している。

具体的には、単語(文章)のベクトル表現の獲得及び文章分類にSkip-gram及びLSTM(Long short-term memory)を採用し、分析対象データを二つに分類し、前半(13年1~6月)のデータで分類モデルの構築を行い(学習)、後半(13年7~12月)のデータでモデル検証を行っている。ニュースデータについては、Thomson Reuters社が提供しているニュース記事のヘッドラインから、13年の一年間で抽出されたヘッドラインニュース、1,864報を基に分析している。マーケットデータについては、Thomson Reuter社より取得した高頻度取引データ(各株式銘柄の取引成立価格・取引量)2億6,700万行以上という大規模データを対象としている。分析に際しては、自然言語処理分野においてその有用性から近年関心を集めているLSTMを用い、深層学習を用いた自然言語処理手法(ニューラル言語モデル)を通じニュース記事内容(非構造化データ)を解析し、短期的な価格変動との関連性をモデル化している。分析の結果、短期間の価格変動を教師データとした場合においてもニュースの分類モデル構築をできる可能性があること、ニュース発信後

の短期間で取引量が大幅に増加する傾向にあること等の結論を見いだしたと報告している。

大規模高頻度データを教師データとして深層学習によりモデル化を行った先端的な試みであり、更なる研究の発展が期待される分野である。

3本目は村宮・竹原論文「ビッグデータと会計研究」である。この論文では、ティックデータを利用して、業績予想修正の開示というイベントに対する市場反応を5分単位で分析した結果を報告し、幾つかの有益な示唆が得られたとしている。

具体的には、TDnetを通じて公表された業績予想修正に関するデータを分析対象として、開示が行われた時刻を分単位で収録したデータとティックデータを5分刻みで集計したデータを結合し、業績予想修正に関する情報がどのような出来高反応を生み出し、その情報がどのように株価に織り込まれていくかを分析している。その結果、出来高反応の分析結果として、立会時間中の開示の方が、時間外の開示に比べて出来高反応が大きく、また、立会時間中の方が開示時刻周辺の時間帯で大きく反応し、その後、30分から1時間という短時間の間で急激に反応が弱まっているという事実や、株価反応の分析結果として立会時間中に業績予想修正が開示されれば極めて迅速にその情報が株価に織り込まれること、arrowhead（東京証券取引所が提供する株式売買システム）稼働後は稼働前に比べて、相対的に早く株価への織り込みが進むことを明らかにしたとしている。

ティックデータを用いた会計研究はこれからであり、有望な研究課題が多く残されていて、今後の研究の拡大が期待される分野である。

4本目は高橋・岡田論文「投資家の注目が新規

公開株式の価格形成に与える影響」である。この論文では、ビッグデータを用いることで投資家行動を定量的に捉え、それがIPOの価格形成に与える影響について分析している。すなわち、公開価格と初値が大きく乖離する原因として、情報の非対称性や行動ファイナンスの視点から、幾つもの説明がなされているが、本論文では後者の立場からIPOパズルの説明を試み、これを支持する結果を示している。

具体的には、投資家の注目度の指標として発行企業ホームページの閲覧数（ページビューあるいはPVという）を用いて、PVと初期収益率の関係を分析し、公開価格アナウンスメント後のPVと初期収益率の間に正の統計的に有意な関係があること、ブックビルディング期間のPVは情報を持つ投資家の需要申告というよりむしろ個人投資家の注目を捉えていると考えられることを実証し、個人投資家の注目が初値の上昇圧力となっている可能性が高いことを示している。

IPO研究だけでなく、ファイナンスのいろいろな分野で、ビッグデータを使った分析が進み、新たな知見が得られることが期待されている。ビッグデータを使った実証研究の成果が、理論の進化に寄与することが期待される。

ビッグデータ（特に個人情報関連）に関しては、欧州で規制の動きがあり、日本でも同様の動きが見られ、今後、何らかの規制が課せられる可能性がある。金融の世界もこの動きの影響を今後受ける可能性がある。しかし、ビッグデータと進歩を続ける情報技術による分析は多くの新たな知見をわれわれにもたらす可能性がある。今回はその一端を紹介することができればと考えての特集である。今後のこの分野での研究成果に期待したい。