

デジタルトランスフォーメーション（DX）時代の 資産価格評価

—データおよび情報技術の利活用と取り巻く課題—

高橋 大志 CMA

(証券アナリストジャーナル編集委員会委員)

1. はじめに

利用可能なデータの拡大・多様化や情報技術の急速な進展は、経済活動をはじめ様々な形で大きな影響をもたらしている。近年、デジタルトランスフォーメーション(DX: Digital Transformation)との言葉を耳にする機会が増えてきた。デジタルトランスフォーメーションを推進するためのガイドライン(経済産業省[2018])は、DXの定義として、企業がビジネス環境の激しい変化に対応し、データとデジタル技術を活用して、顧客や社会のニーズを基に、製品やサービス、ビジネスモデルを変革するとともに、業務そのものや、組織、プロセス、企業文化・風土を変革し、競争上の優位性を確立することと記述しており、その重要性を推し測ることができる。2018年5月時点のS&P500構成銘柄にて時価総額が最も大きい5社(Apple、Alphabet、Microsoft、Amazon、Facebook)はいずれもデジタル関連企業であるといった指摘(Verhoef *et al.* [2021])などからも、その影響の大きさがうかがえる。

デジタル化への対応といってもいくつかの選択肢がありそうである。Verhoef *et al.* [2021] は、デジタル化への取り組みの程度について分類を行っており、情報の電子化を指すデジタイゼーシ

ン(Digitization)、デジタル化がビジネスに用いられるデジタイゼーション(Digitalization)、さらに、既存のビジネスプロセスに変化をもたらすデジタルトランスフォーメーション(DX)といった三つの分類を示している。また、DXの推進に関し、Matt *et al.* [2015] は、DXにはトップマネジメントの関与が不可欠であると指摘している。データおよび情報技術の利活用の果たす役割は高まっており、企業や各組織をはじめ社会全体における効率的な取り組みが期待される(高橋[2018])。

2. 統計調査におけるデータの検証

統計調査なども情報技術などを通じたデータ利活用への取り組みが行われている分野の一つに挙げられる。例えば、総務省[2020]は、経済統計に関する調査においても、社会におけるDXの取り組みが進められていること、また、ビッグデータの利用が広がっていることから、ビッグデータの経済統計への活用促進に期待を寄せている。さらに、その活用法やメリットに関して、既存統計の補完、報告者負担や業務負担の軽減などが挙げられており、例えば、既存統計の補完に関しては、(1)速報性、(2)詳細化、(3)カバレッジの拡大、

(4)新たな指標の作成などの観点を示している。

近年、新型コロナウイルス感染拡大を背景に、メディアを通じ繁華街の人出の増減などの情報を目にする機会が多くなった。そのような流動人口データは、ビジネス領域を含めこれまで以上に関心を集めているが、経済統計への利活用を検討する際、データの特性は、把握しておくべき項目の一つに挙げられる。菅ほか [2019] は、流動人口データ (GPSデータ) を国勢調査などと比較を行い、GPSデータは、国勢調査などと相関があり、一定レベルの信頼性が確保されているとともに、利活用に当たり国勢調査を補完し得るものとして有効であると指摘している。さらに近年、地球環境や気象などの観測データを対象とした検証なども意欲的に行われている (石田 [2021]、佐藤 [2021])。このような地道な取り組みは、今後の利活用の促進に大きく貢献するものと期待される。

3. 資産価格評価に関する取り組み

資産価格評価に情報技術を応用する事例も、数多く報告されている。Mullainathan and Spiess [2017] は、計量経済学における機械学習について議論を行っており、その中で、機械学習の各モデルに触れ、Lasso、Ridge、Elastic net、深層学習、アンサンブル学習などといった手法を概観している。

機械学習の手法は多岐にわたるが、Yang *et al.* [2021] は、深層学習の役割が高まってきていると指摘している。自然言語処理は、深層学習の主要な応用領域の一つであり、資産価格評価に関する分析事例も数多い (Devlin *et al.* [2018])。

自然言語処理を通じたテキストの可視化などは、最も分かりやすい取り組み事例として挙げら

れるかもしれない。例えば、特許文書を対象とした分析として、Yonemura *et al.* [2021] は、自然言語処理を通じ特許文書のベクトル表現を獲得した上で可視化を行い、各企業の保有する特許の時系列推移を比較している。Chen *et al.* [2021] は、日本および米国両国の特許を対象とし、ソフトクラスタリングを通じ特許文書から各国の産業構成の時系列推移を推定し、両国の相違点を指摘している。松本ほか [2020] は、特許文書の分析を通じ企業の多角化に焦点を当てた研究を行っており、企業の多角化により企業価値の減少が生じるコングロマリットディスカウントと整合的な結果を見いだしている。

自然言語処理を通じた文章分類なども広く行われているが、近年、関心を集めている手法の一つに文章生成、文章要約などが挙げられる。例えば、Nishi *et al.* [2021a] は、ロイターニュースを対象に分析を実施し、近年注目を集めている文章生成モデルGPT-2 (Radford *et al.* [2019]) による文章生成を通じ利用可能なデータを拡張することで、分類モデル (RoBERTa、BiLSTMなど) の精度向上に貢献できることを示している。また、野矢・高橋 [2021] は、企業に関する過去のニュース (例：過去1カ月) を基に当該企業に関するニュースの要約文章を作成し、新たなニュースと要約文章との違いに着目し、株式市場におけるニュースの反応との関連性について分析を試みている。

情報技術の進展は多岐にわたり、自然言語処理を含め数多くのモデルの提案および応用が行われている。これらの成果を資産価格評価に取り込むことで、新たな視点を含んだ取り組みを実施できる可能性がある。

大きく三つのアプローチ（ナレッジベース、統計的機械学習、深層学習）が存在すると指摘している。

(3) モデルのメカニズム

モデルが機能するメカニズムの解明も課題の一つに挙げられる。深層学習などにおいては、モデル構築に利用可能なデータ数と比較して推定すべきパラメータの数が多く、そのようなモデルが優れたパフォーマンスを示す理由については、必ずしも十分な理解が得られていない。モデルのメカニズムの解明を試みる基礎的研究も試みられている（Amari [2020]）。

(4) 個人情報を含むデータ

また、個人情報を含むデータ（パーソナルデータ）の取り扱いも課題の一つである。内閣官房IT総合戦略室 [2017] は、データに関して、個人情報を含むデータ、匿名加工されたデータ、個人にかかわらないデータの三つに分類した上で、個人情報を含むデータについては、利活用が十分に進んでいるとは言い難いと指摘している。パーソナルデータを含めたデータの利活用などについても、産官学交えた検討が進められている（総務省 [2018]、システムイノベーションセンター）。

(5) 人材育成、組織体制の構築

データおよび情報技術の利活用を進めるための人材育成や組織体制の構築も重要な課題として挙げられる。例えば、総務省 [2020] は、ビッグデータの活用を社会的に推進していくために、データを扱うスキルや知識を有する人材の育成を進めるとともに、具体的にデータを活用し業務を遂行する組織では、そういった人材を確保し組織的

にノウハウの蓄積を進めていく必要があると指摘している。また、各領域の慣習の違いの理解などといったものも関連した課題に挙げられるかもしれない。

本章に記した以外にも数多くの課題が存在している。課題解決に向け、産官学連携への期待も高まっているように思われる。

5. 終わりに

本稿では、データおよび情報技術の利活用への取り組みに関し、われわれの取り組み事例を含め、いくつかのトピックに触れた後、取り巻く課題および解決に向けた取り組みなどについて概観した。金融分野におけるデジタルデータおよび情報技術の利活用の影響が、アナログデータのデジタル化のみにとどまるのか、大きなプロセスの変化をもたらすものなのか、今後の動向が興味深い。

本研究の一部は、科学研究費補助金（JP20K01751）の助成を受けた。また、日本証券アナリスト協会、青山学院大学大学院国際マネジメント研究科共催セミナー「非財務情報とオルタナティブデータを用いた企業評価の潮流」参加者から有益なコメントをいただいた。ここに記して感謝したい。

〔参考文献〕

- 石田中 [2021] 「地球観測衛星データを用いたSDG15.4.2（山地植生被覆指数）の試算について」、総務省第3回観測データ利活用検証WG資料、経済産業省 [2018] 「デジタルトランスフォーメーションを推進するためのガイドライン」、佐藤彰洋 [2021] 「Kapos分類ごとのMGCIの区間推定値」、総務省第3回観測データ利活用検証WG資料、システムイノベーションセンター（SIC）「ウェブサイト」、<https://sysic.org>。

- 菅愛子・飯島信也・兵頭大史・藤原直哉・水野貴之・松本裕介・武藤杏里・瞿雪吟・伊藤武真・松井伸司・五十嵐盛仁・上田聖 [2019] 「東京都における流動人口データの有効性の検証」、総務省統計委員会ワーキングペーパー 2019-WP03.
- 総務省 [2018] 『情報通信白書 (平成30年版)』.
- [2020] 「公的統計へのビッグデータの更なる活用に向けて」.
- 高橋大志 [2018] 「情報技術が産業・ビジネスに与える影響—技術革新と資本配分—」、『統計』69 (1).
- 内閣官房IT総合戦略室 [2017] 「AI、IoT時代におけるデータ活用ワーキンググループ 中間とりまとめの概要」.
- 野矢淳、高橋大志 [2021] 「要約文章と機械学習による株価変動の分類モデルの構築」、人工知能学会第18回ビジネス・インフォマティクス研究会 (SIG-BI)、2021年.
- 松本裕介・菅愛子・高橋大志 [2020] 「テキスト情報を用いた企業の技術的多角化と企業価値との関連性分析—特許データを用いた分析—」、日本ファイナンス学会第28回大会.
- Adadi, A. and M. Berrada [2018] “Peeking inside the black-box: a survey on explainable artificial intelligence (XAI),” *IEEE access* 6.
- Amari, S. I. [2020] “Any target function exists in a neighborhood of any sufficiently wide random network: A geometrical perspective,” *Neural Computation* 32(8).
- Arrieta, A. B., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, ... and F. Herrera [2020] “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion* 58.
- Athey, S. [2017] “Beyond prediction: Using big data for policy problems,” *Science* 355 (6324).
- Chen, Z., Y. Matsumoto, A. Suge and H. Takahashi [2021] “The Visualization of Innovation Pathway Based on Patent Data—Comparison Between Japan and America,” *Smart Innovation, Systems and Technologies* 241, Springer.
- Devlin, J., M. W. Chang, K. Lee and K. Toutanova [2018] “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805.
- Goudet, O., D. Kalainathan, P. Caillou, I. Guyon, D. Lopez-Paz and M. Sebag [2017] “Causal generative neural networks,” arXiv preprint arXiv:1711.08936.
- Lundberg, S., and S. I. Lee [2017] “A unified approach to interpreting model predictions,” arXiv preprint arXiv:1705.07874.
- Matt, C., T. Hess and A. Benlian [2015] “Digital transformation strategies,” *Business & Information Systems Engineering* 57(5).
- Miller, T. [2019] “Explanation in artificial intelligence: Insights from the social sciences,” *Artificial intelligence* 267.
- Mullainathan, S., and J. Spiess [2017] “Machine Learning: An Applied Econometric Approach,” *Journal of Economic Perspectives* 31(2).
- Nishi, Y., A. Suge and H. Takahashi [2021a] “Construction of a News Article Evaluation Model Utilizing High-Frequency Data and a Large-Scale Language Generation Model,” *SN Business & Economics*, Springer.
- [2021b] “Elucidation of Factors Affecting Price Formation through High-Frequency Data and News Evaluation,” the SCE 27th International Conference on Computing in Economics and Finance (CEF 2021).
- Raddant, M. and H. Takahashi [2021] “Corporate boards, interorganizational ties and profitability: The case of Japan,” *Empirical Economics*.
- Radford, A., J. Wu, R. Child D. Luan, D. Amodei and I. Sutskever [2019] “Language models are unsupervised multitask learners,” *OpenAI blog* 1 (8).
- Ribeiro, M. T., S. Singh and C. Guestrin [2016] “Why should I trust you?: Explaining the predictions of any classifier,” Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.
- Spirtes, P., and K. Zhang [2016] “Causal discovery and inference: concepts and recent methodological advances,” *Applied informatics* 3 (1).
- Verhoef, P. C., T. Broekhuizen, Y. Bart, A. Bhattacharya, J. Q. Dong, N. Fabian and M.

