金融データサイエンス入門Ⅱシリーズ

# シリーズのまとめ

大　庭　昭　彦 CMA
（証券アナリストジャーナル編集委員会委員）

この数年間、金融のデジタル化と広範囲のオルタナティブデータ利用の拡大で、生成AIや深層学習をコアとする多くの定量的・工学的な技術がこれまでになかったほど大きな影響を社会に及ぼしている。データサイエンスは、従来の方法では解決できなかった問題や、そもそも従来は問題として認識できなかった問題でも活用が進んでいる。本連載「金融データサイエンス入門Ⅱシリーズ」は、会員の皆さまからのアクセスが多かった2023年の連載「金融データサイエンス入門シリーズ」の続編として企画されたものだが、引き続き、大学・民間の研究者、実務家など様々な立場の方から先端的で重要な解説をいただくことができたと考えている（注1）。

第１回の**高橋論文「データサイエンス入門・概観―機械学習、生成AI、システムサイエンス、および関連トピック―」**は、データサイエンスの学際的な定義、ビッグデータとの関連性、データサイエンティストの役割、および主要な分析技術（機械学習、生成AI、システムサイエンス）について幅広く概観している。例えば、プロンプトエンジニアリングは、生成AIへの入力（プロンプト）の与え方が、出力のクオリティに大きく影響するということを具体化している。三つの基本原則、①具体的に書く、②前後の情報を考慮する、③複雑な操作を単純なステップに分解する、などは、有効な出力を得るために非常に有効だということで参考になる。生成AIの課題としては、出力の正確性や信頼性に対する懸念（誤った意思決定や偽情報の拡散の可能性）やバイアスの問題、そしてモデルが複雑化することによる説明可能性の困難さが指摘されている。また、AIエージェントへの関心が高まっているという話も興味深い。AIエージェントとは、自然言語による指示を受け取り、ユーザーの代理で自律的に行動できるAIシステムである。最後に、データサイエンスは、対象をシステムとして捉え問題解決を図るシステムサイエンスの中で利用されることで価値を発揮するという重要な主張もされている。このために、データサイエンスでは、分析対象のドメイン知識やデータの背後にあるメカニズムの理解が、人間の役割として重要になるということだ。

第２回の**宮川論文「因果推論手法の（金融）実務利用―概念・実例と展望―」**は、データ分析手

（注１）　2023年の連載のアクセスが多かったことについては、2025年７月号の本連載「シリーズ企画の狙い」で詳しく述べた。

法の三つの類型（計測、予測・検知、因果推論）の中で、介入行為を伴うタスクに不可欠な因果推論について、その概念と金融実務での応用を紹介している。因果推論の最も重要な概念は「潜在アウトカム」であり、因果効果とは、ある処置があった場合の結果と、処置がなかった場合（反実仮想）の結果を比較することで得られる。ここで難しいのは、母集団全体での平均処置効果をATE、実際に処置を受けた集団での平均処置効果をATTとすると、一般的にATTとATEは同じではないということだ。観測データから因果効果を推定しようとすると、処置群と対照群の間にもともと存在していた違いや、処置効果が均一でないことから生じるバイアスが入り込むことがある。これらのバイアスが入らないようなランダムな割り付けを行う特別な方法をRCT（ランダム化比較試験）と呼ぶが、一般の因果推論では、これらのバイアスを排除する条件や工夫が重要だ。例えば、「自然実験」では、RCTに似た状況が自然に発生しているケースを見つけて研究する手法を指す。「回帰不連続デザイン」では、処置を受けるか否かを決める基準となる変数のわずかな変化を利用し、基準の上下で処置効果を識別する。「操作変数法」では、処置とは相関するが結果とはダイレクトに相関しない操作変数を用いることで、外生的な処置の変動を疑似的に生み出し因果効果を識別する。最後に、新しい技術として、機械学習（ML）手法を因果推論へ取り込む動きがあり、高次元の説明変数から相関の強いものを選択することで交絡要因をシステマチックに制御する手法（例：Double／Debiased Machine Learning、PDS Lasso）や、処置効果の異質性を探索する手法が紹介されており、大変興味深い。因果推論は、ドメイン知識や介入の可否に関する現実的な判断が求められるため、「マシン」に奪われにくい人の業務の典型であると考えられていることからも、分析を業務とするアナリストにとって重要な概念だと思われる。

第３回の**平野論文「AI時代の「モデル」は金融市場をどう変えるか？―数理モデルからLLMまで、モデルの活用を考える―」**は、AI時代における「モデル」に焦点を当て、注目されているモデルとして、人工市場、Deep Hedging、LLMを紹介している。

人工市場／マルチエージェントシミュレーションは、コンピュータ上に仮想市場と仮想トレーダーを構築し、市場の動きをシミュレーションする手法で、エージェント間の相互作用から創発的に発生する現象のメカニズム解明が目的である。金融危機のような非線形な挙動やレジーム変化、現実では検証困難な制度変更シナリオの検証が原理的には可能であるが、多くの技術的課題が存在する。Deep Hedgingは、デリバティブのヘッジ戦略策定、リスク管理、価格設定を行う手法である。ニューラルネットワークを用いてヘッジ量を直接計算する。従来のブラックショールズモデルでは考慮できなかった取引コストを考慮でき、不完全市場を前提としたモデル設定が可能である。学習の不安定性や解釈困難さなどの問題がある。LLM（大規模言語モデル）は、膨大な情報の中から重要な情報を取り出したり、センチメント判定をしたりする「情報抽出機」としての機能と、アナリストレポートを自動的に書くような「情報出力機」としての機能が期待されている。もっともらしい嘘を生成してしまうハルシネーション問題もあり、正確性が重視される金融分野などでの活用は進みにくいようだ。

第４回の**水門・中塚論文「金融市場を対象とし**

**た自然言語処理とネットワーク分析」**は、非構造データである自然言語処理とネットワーク分析を金融市場に応用した研究を紹介している。例えば、コモディティ先物価格の週次リターンに基づいた分析では、銅が４種の中心性指標で高い中心性を示した。これは、銅がコモディティ市場における情報伝播や価格形成のハブとして機能している可能性を示している。暗号資産のリターンの独立性の分析や、ビットコイン関連のSNS投稿の共起ネットワーク分析も紹介していて、密度が濃い内容だ。なお、ここで見ているネットワークの構造は断面の分析なので、宮川論文で重視したような因果性を捉えるためには、さらなる工夫が求められることも注記されている。まだまだ分析者にとってできることは多い。

自然言語処理（NLP）は、大きなテキスト情報を定量的に分析する。金融分野では、文脈を考慮した深層学習型モデルによるセンチメント分析（極性分類）の応用が進んでいる。ここでは、企業の決算説明会テキストを対象に、NLPモデルの一つであるBERTを用いて発言内容の極性（ポジティブ／ネガティブ）を評価し、株価との関係を分析している。特筆すべきは記載内容を話者ごとに、「経営者からの説明部分」「参加者からの質問内容」「経営者による回答」の三つに分けて分析した点だ。３種ともに株価への影響が認められたが、特に質問部分の影響の推移がユニークであった点は本分析の貢献である。この結果は、経営者の説明や回答だけでなく、アナリストの質問にも株価と関係を持つユニークで重要な情報が含まれることを示唆していて興味深い。

第５回の**鈴木論文「深層学習の数理」**は、深層学習（Deep Neural Network：DNN）について、「次元の呪いの回避」や「特徴学習能力」の観点から理論を示し、現在の基盤モデルとなっているTransformerの成功の理由を専門家の立場から精緻に解説している。最近のAIに関する議論、紹介記事などにおいて、AIの成果を喧伝されることは多いが、「なぜ成果が出ているのか」についてはブラックボックスになってしまっている面もある。AI技術を分析に利用していく必要のある現代的な証券アナリストとしては、良い結果が得られる理由がブラックボックスのままでは心もとないといえる。

深層学習は、中間層においてデータの適切な表現（特徴量）を構築することで、高度な情報処理を実現している。高次元データ（画像、音声、文章など）をそのまま扱うと、次元の増加に伴い計算量が指数関数的に増加するという「次元の呪い」を受けるが、深層学習は特徴学習によってこれを回避している。このことは、深層学習が「滑らかな関数（高周波成分が早く減衰する関数）」に対し高い近似性能を持っているからだと説明される。「滑らかな関数」を“異常なギザギザ”を含まない普通の形状の関数のことだと解釈すれば、深層学習は普通の関数に強い手法だということになる。また、Transformerの主要構成要素であるAttention機構は、入力系列に応じて、その系列内のどのトークンが重要であるかを取捨選択（特徴選択）し、動的な特徴選択を実現させている。この特性により、Transformerは滑らかな関数に対して多項式オーダー（速いオーダー）で近似誤差を減らすことができ、次元の呪いを回避している（注２）。Transformerの能力の高さにつながっている、もう一つの重要なポイントが文脈内学習（In-Context Learning）である。これを事前に十分に行うことで、Transformerはデータの最適な表現を獲得できる。Attention機構の有用性と合わせて、Transformerは総合的に高い性能を持っ

ているというわけだ。この第５回は、他の回に比較して数式の多い回ではあるが、議論全体を眺め、丁寧な説明部分のおおまかな理解ができるだけでも読む価値があると思われる。

技術革新の激しいデータサイエンスの分野について、これからの証券アナリストに必要となる知見を適宜アップデートしていくことを期待して、前回の入門シリーズに続いて今回の入門Ⅱシリーズを企画した。結果として、全体像の拡大や、予測・検知にとどまらない介入を伴う問題へのアプローチ（因果推論）の理解、モデルや応用の広がりについて紹介することができた。また、「なぜ従来の技術で難しかったことが、深層学習でできるようになったのか」についての理論的整理を、第一線の研究者の方から説明していただくこともできた。本シリーズで紹介されてきたことを踏まえ、金融とデータサイエンスのかかわり、さらには人とAIとの新しい関係について、証券アナリストとしての見方がますます深まっていくことが期待される。

（注２）　鈴木教授は、このことを示した論文の共著者である。Takakura, S. and T. Suzuki “Approximation and Estimation Ability of Transformers for Sequence-to-Sequence Functions with Infinite Dimensional Input”（ICML23、2023年）。