

## 近未来金融システム創造プログラム第12回講義レポート

第12回目となる本日は、東京大学大学院工学系研究科教授の和泉潔様から「金融と技術（各論Ⅳ）オルタナティブデータと金融」という題目で講義が行われた。和泉様からはAIエージェントの発展についてと将来解決できる課題についてお話いただいた。授業後半では和泉研究室に在籍している高田様よりAIエージェントを実践していただいた、

### 1. 金融市場におけるデジタル化の三段階

金融市場における人工知能技術の応用を三段階とは、デジタルリゼーション (Digitization)、デジタルライゼーション (Digitalization)、そしてデジタルトランスフォーメーション (DX) である。第一段階のデジタルリゼーションは、紙などで記録されている情報をデジタル化する段階を指す。第二段階のデジタルライゼーションは、デジタル化されたデータを用いてビジネスの様々なプロセスを自動化する段階である。第三段階のDXは、サービス自体を変革あるいは拡張するフェーズである。

わかりやすい例として切符のデジタル化が挙げられる。まず切符がデジタル化され、次にそのデジタル切符を使って改札が自動化され、最終的にはSuicaのように交通系ICカードが駅構内や電子マネーとして多様なサービスに拡張されていった。この流れは金融市場においても同様に起きており、市場分析や資産運用におけるAI応用もこの三段階で進展している。

金融分野では特に、データの性質と分析手法のマッチングが重要である。過去には「データを大量に集めて機械学習に投入すれば何か有用な結果が得られる」という期待があったが、最近ではデータが持つ情報の性質に合わせた適切な機械学習手法を選択する視点が重要視されるようになってきた。

### 2. オルタナティブデータの可能性と限界

第一段階であるオルタナティブデータとは、従来の統計には含まれない新しいデジタルデータを市場分析に使用する試みである。具体的には、ニュースのテキスト、スマートフォンのGPSデータ、ソーシャルメディアのコメント、気象データ、衛星画像などが含まれる。特に新型コロナウイルス感染症の流行以降、現地調査が困難になったため、こうしたデータを用いた分析が注目を集めた。

実用例として、スマートフォンの位置情報や工場への出入り状況、駐車場の車両数などのデータを使えば、生産量や消費動向をリアルタイムに推定できる。衛星画像を用いて小麦畑の生育状態を判定し、コモディティ市場の先物予測に使用する事例もある。また自然言語処理によるテキストマイニングを用いて、企業の発表やニュースからセンチメントを算出し、ESGに関する方針変更を抽出して定量化する研究も大きく成長した。コロナ禍初

期の例では、クレジットカードのデータなどを使用して翌日には消費動向の急落を把握し、迅速な判断を行うことが可能になった。

オルタナティブデータにはネガティブな側面も存在する。2024年に The Journal of Finance に掲載された研究 (O. Dessaint et al.) では重要な結果が示された。1980年代から2020年までの40年間で、アナリストの短期業績予測 (1日から1年) の精度は向上している一方で、2年以上先の長期予測精度は低下しているという。著者は短期予測の精度向上について、オルタナティブデータを含む様々なデータの入手可能性が高まったことが大きな要因であると説明している。

一方で、ソーシャルメディアデータなどの利用は短期予測を改善するが、長期予測 (3年以上) の精度を低下させるという結果も示されている。著者たちの仮説は、GPSデータやソーシャルメディア、衛星画像など一般的なオルタナティブデータは短期的な情報に強いが、長期的な企業の投資や構造的変化に関してはノイズになりやすいというものである。この知見は、ダニエル・カーネマンの『ファスト&スロー』で述べられた「速い思考 (システム1)」と「遅い思考 (システム2)」の概念を想起させる。オルタナティブデータを用いた分析は、直感的・反射的に判断する「速い思考」に近い情報処理を誘発するが、企業が長期的にどのように対応して回復・成長するかといった「遅い思考」の考察は困難である。

実際の金融機関との共同研究においても、「短期的で反射的なレスポンスには様々な分析ができるが、長期的な予測は難しい」という声が聞かれる。金融はデータドリブンの試みで大きな失敗を繰り返してきた歴史があり、長期的に構造が変化する局面において、短期的な反射だけに基づいて行動すると重大な失敗につながる可能性がある。

### 3. 因果ネットワークを活用した長期予測手法

長期的な予測において必要なのは、経済や社会のデータが単独で存在しているのではなく、背後に因果関係やネットワークが存在するという認識である。例えばアメリカの関税政策の変更が他国の産業構造や輸出入に波及し、それがさらに別の国に影響を与えるといった因果のネットワークが存在する。こうした因果関係を無視してディープラーニングに投入するのは危険である。

大規模言語モデル (LLM) についても、基本的には「次の単語を予測する」学習をしているに過ぎず、現象の背後にある複雑な因果関係を自ら保持しているわけではない。したがって、LLM 単独で複雑な因果推論を行うことは難しく、因果探索などの別のツールとの併用が必要となる。

和泉研究室では、企業の決算説明資料などのテキストから自動的に因果関係情報を抽出し、因果ネットワークを構築する技術を使用している。LLM と因果ネットワーク探索を組み合わせると、因果探索の範囲が格段に広がる。LLM に「なぜ材料価格が上昇したのか」と問うと主要な要因を8個程度挙げるが、因果ネットワークを介してさらに間接要因まで

遡ると 20 個以上の要因が明らかになることがある。

メガバンクとの共同研究において、日本の業種別の月次倒産件数をニュースデータから予測する実験を行った。本研究では、ニュース記事に含まれるキーワードを単純に全て投入するのではなく、因果関係のフィルターを途中で挟んだ。すなわち、小売業の倒産に因果的に関連しそうな要因を因果ネットワークからピックアップし、その類似度の高いキーワードのみをインプットに使用するという手法である。

実験では東京商工リサーチの月次倒産件数を目的変数とし、(1) 単純な時系列による予測、(2) ニュース記事のキーワードをそのまま全て投入した手法、(3) 因果フィルタで絞り込んだキーワードのみを使用した手法の 3 つを比較した。結果として、1 から 3 ヶ月先の短期では、過去の時系列データだけでの予測がむしろ良好であった。一方、6 ヶ月から 1 年程度の中長期予測では、因果でフィルタリングしたテキストを用いた手法が最も誤差が小さく、性能が最良であった。これは中長期になると波及効果が現れ、セクター間の連鎖的な影響が重要になるためだと考えられる。

#### 4. AI エージェントによる金融分析の自動化

第二段階であるプロセスの自動化について、現在注目を集めているのが「AI エージェント」である。金融分野でも分析や様々な作業を自律的にこなす AI エージェントが注目を集めている。AI エージェントは漠然とした指示でも、自ら必要なデータや分析手法を選択し、収集・解析してレポートを作成するような能力を目指している。

興味深い事例として「AI サイエнтиスト」がある。このシステムは、アイデア創出、先行研究の調査、実験コードの自動生成、実験の自動実行、結果の要約・論文化まで自律的に行うというものである。初期バージョンの成果物は、トピックによっては学術誌に採択される可能性もある水準という評価であった。将来的には AI が研究者の一部作業を代替する可能性があり、同様の自律化は金融アナリストやトレーダー分野にも波及するであろう。

AI エージェントの進化は単独エージェントからマルチエージェントへと進展している。ある国の経済分析を例に取れば、財務データを処理するエージェント、ニュースを分析するエージェント、衛星画像を処理するエージェントといった具合に、専門領域ごとのエージェントが協働し、最終的な統合レポートを生成する。

#### 5. マルチエージェントシステムと市場シミュレーション

証券会社との共同研究では、財務情報、株価、ニュース記事といった情報別にエージェントを作成し、それぞれのエージェントの結果を統合してレポートを生成する試みを行った。各エージェントに役割を分担させ、上位の統合エージェントが最終判断を下す構成である。結論として、単独の AI エージェントに全情報をまとめて投入するよりも、情報別に分散処理させて協調させるマルチエージェントの方が精度や安定性が高まる傾向が見ら

れた。またマルチエージェント化によって人間特有のバイアスが減少するという効果も観察された。

第三段階は、金融市場や金融サービスそのものを変革する新たなサービスを創出する段階である。国際会議では、特に LLM を用いてサービスをどう変革するか、説明可能性をどう担保してサービス化するかという議題が増加している。ただし、固定的なモデルや過去データだけに依存するアプローチは、構造変化や未知のショックに対して脆弱である。

複数の研究をまとめると、単独の LLM だけでは人間の群衆予測に劣ることが多いが、複数の LLM をアンサンブルすると人間の群衆と同等あるいは近い性能を示すという知見もある。経済や金融は固定した関係性がない領域であり、関係性は時点ごとに変化するため、様々な考え方をを持った主体を模したモデル群を用意し、集合的な挙動をシミュレートできる仕組みが重要である。これは、さらに進んでデジタルツインやエージェントベースのシミュレーションの考え方につながる。

デジタルツインの文脈では、個々の家計や企業、政府などをエージェントとして表現し、それぞれの行動規範や政策反応を模擬して経済全体をシミュレートする研究が進展している。こうした手法の利点は、実際には存在しない仮想的なショックを入力して各エージェントの反応を観察し、マクロへの影響を評価できる点である。

## 6. デジタルペルソナの生成とその実践的応用

実際の人間に対して2時間のインタビューを行い、そのインタビュー記録を基に LLM でペルソナエージェントを生成する実験がある。彼らは1000人分のペルソナを生成し、性格特性や行動実験における回答の再現性を検証した。インタビューを基に生成したペルソナは、元の人間の回答と約70%程度の再現性を示したという結果が報告されている。この技術を金融に応用すれば、様々な投資家ペルソナをデジタルツインとして作成し、市場シミュレーションに使用することで新たなサービスや分析が可能になると考えられる。

ここからは和泉研究室の高田様によるデモを紹介する。

デモでは、AI インタビュアーが登場し、被験者に対して質問を行う。インタビューは約2時間を要する半構造化面接で、人生経験や様々な社会問題に対する意見を聞く形式である。マイクを通じた音声リアルタイムで文字認識され、その回答を踏まえてAIが深掘りする質問をしていく仕組みである。インタビューは主要な質問が約100問あり、それぞれに所要時間が設定されている。

高田氏は実際に2時間のインタビューを行い、自身のエージェントを作成した。このエージェントに「あなたは誰ですか」と質問すると、自身の名前や年齢を回答した。ここで話している声は本人の声を5秒ほど録音し、その音声をもとに合成した音声を用いているため、本人の声で喋っているように聞こえる。研究内容について尋ねると、エージェントは「最近ではジェネレーティブエージェントの開発に取り組んでいます。エージェントがどのように学習し、集団内での行動を形成するかを研究しています」といった適切な説明を

返した。

インタビューの内容はエージェントの記憶として保持され、後の質問で深掘りに用いられる。また「あなたの今の上司についてどう思っていますか」とエージェントに尋ねると、「私の上司については非常に尊敬しています。先生は研究に対する情熱が強く、常に新しい試練を与えてくださいます」といった具体的な応答が得られた。

さらに高田氏以外に2時間のインタビューを行ったエージェントとの高田様のエージェント同士で会話させることも可能である。その会話例では、「近未来金融システム創生の講義受講生に伝えたいこと」というテーマで議論が進行し、テクノロジーの進化が金融システムに与える影響などについてエージェント同士で自然な議論が成立した。

このようにエージェントを作成すれば、エージェント同士の会話が成立し、自分の考えを整理するための相手としても利用できる。この技術は金融市場における投資家行動のシミュレーションや、様々なステークホルダーの意思決定プロセスをモデル化する上で重要な役割を果たすことが期待される。本講義で紹介した一連の技術は、金融市場におけるデータ活用の三段階を体現するものであり、今後の金融技術の発展において中心的な役割を果たすと考えられる。

## Q&A

Q1: オルタナティブデータが多数ある中で、どのようにして使いそうなデータを選別すればよいか。

A1: データ同士の関係性や因果性を基準に選別することが重要である。業種や目的に関連する因果関係の抜けがないかを確認し、多様性（バラエティ）を持ったデータソースを確保する一方で、冗長でノイズになり得る情報は排除することが推奨される。

Q2: AI エージェントやペルソナの応答生成は計算量が大きいのか。どの程度の計算資源が必要か。

A2: 実装によって大きく異なるが、今回示されたエージェントは膨大なトランスクリプトを記憶しているものの、毎回全体を総当たりで検索するのではなく、質問との関連度が高い記憶だけを取り出して回答を生成する仕組みであるため、必ずしも極端な計算資源を要するものではない。

Q3: 金融市場のシミュレーションが理想的に機能して全員が最適な行動をとると市場の流動性は失われるのではないのか。市場の役割はどうか。

A3: 公開されたシミュレーション結果があればそれが市場参加者の行動に影響を与えるが、情報を非公開にして利用する者は優位に立ち得る。また、すべてが同一行動になると見えた場合でも、マイノリティや裏をかく戦略をとる者が報酬を得る可能性があるため、

市場が完全に停止するとは考えにくい。

Q4: ペルソナ実験で人間の回答の約 70%が再現できたとあるが、再現できなかった残り 30%はどのような質問や領域か。インタビュー時間を延ばすと再現性は改善するか。

A4: 再現性が低かったのは主に経済ゲームや他者とのインタラクションに関する応答である。通常の半構造化インタビューは個人の人生史や背景に重点を置くため、対人相互作用に依存する行動は再現しにくい。MBTI のような性格指標の一部は部分的に再現されるが完全ではない。

Q5: ペルソナ創作時に個人の第一印象や感情の多様性、場面に応じた表現を織り込むことは可能か。

A5: 可能であるとされる。発言者は場面や相手に応じて異なるペルソナを演じ分けることは実装されつつあり、即時の応答や文脈適応は比較的容易に実現できると説明した。ただし、間を置いた反応や深い個性に基づく振る舞いといった遅延的な特性の表現は課題として残っている。

Q6: 株式トレードのような心理戦やブラフを伴う状況（ポーカーのような場面）は LLM で学習・実行可能か。

A6: 原理的には可能である。近年の LLM は他者の心的状態を推測する性能が向上しており、対戦相手の行動を予測し、戦略的に振る舞うためのモデル化が実用化されつつある。