

大規模言語モデルを用いた新たなテキスト分析の取り組み
— 最近の賃金・物価動向に関する分析への応用 —

調査統計局 井澤公彦*、亀井郁夫、柴田菜緒、高橋悠輔、米山俊一

Bank of Japan Review

2024 年 12 月

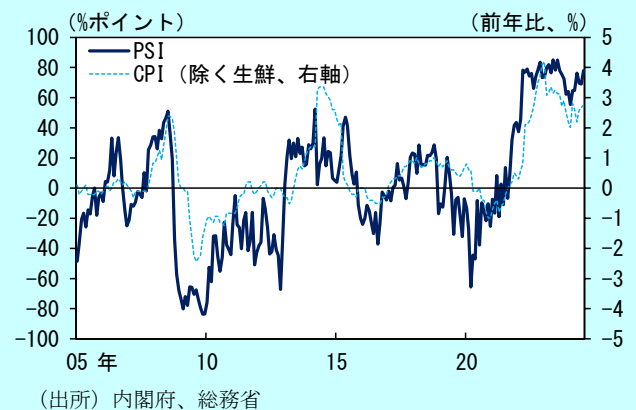
本稿では、近年の大規模言語モデル（Large Language Models：LLM）の急速な発展を踏まえ、経済情勢判断における、LLM を用いたテキストデータ分析の応用可能性を検証した。LLM は、分析者が事前にコストをかけることなく利用可能なモデル・ツールが充実していること、言語の基本的な知識を習得済のため任意の話題・文章について分析可能であること、といった利点があり、海外の中央銀行によるものを含め経済分析における活用が広がりつつある。本稿では、具体例として内閣府「景気ウォッチャー調査」のコメントを対象に、LLM を用いて最近の賃金・物価動向に関する分析を試みた。分析結果からは、販売価格上昇の要因が、原材料費から人件費に移りつつあることが示唆された。

はじめに

日本銀行では、これまで金融政策のための経済情勢判断に際し、テキストデータの活用を進めてきた。テキストデータは、高頻度データや高粒度データ等と同様にオルタナティブデータの一つであり、伝統的な統計データとは異なる特徴を持ち、経済・物価情勢の判断材料の一つとして有用である¹。例えば、内閣府「景気ウォッチャー調査」は景況感を示す指数とともにその背景に関するコメントを公表しており、日本銀行はこのコメントを用いたテキストデータ分析について、複数の研究成果を公表してきた。

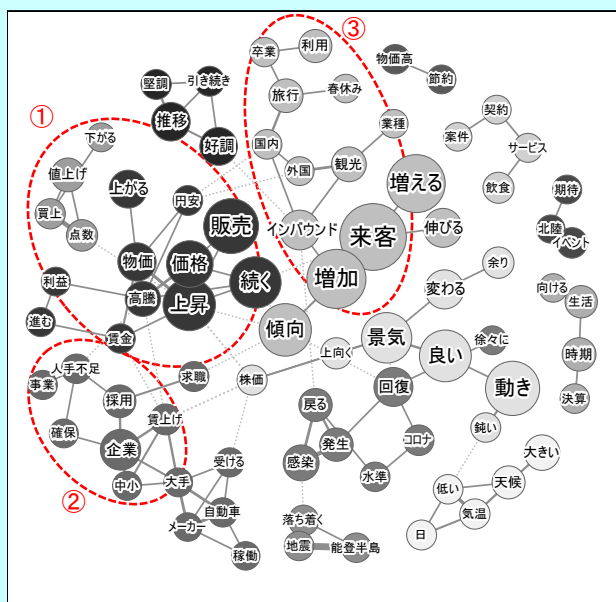
図表 1 は、同テキストデータを用いて作成された、「物価センチメント指数 (Price Sentiment Index：PSI)」である²。PSI は、機械学習の手法³を用いて物価関連コメントを抽出したうえで、それらのコメントを物価の「上昇」・「下落」・「横ばい」の категорияに分類し、「上昇」コメント数から「下落」コメント数を差し引いた値を物価関連コメント数で除すことで算出している。「景気ウォッチャー調査」は、全国の消費者物価指数より速報性があるほか、PSI は同指数に数か月間先行して相関する性質があり、先行指標として有用である。

【図表 1】物価センチメント指数 (PSI)



図表 2 は、共起ネットワーク図と呼ばれる図表であり、景況感の指数だけでは分からない背景にある定性情報を定量的に可視化したものである。具体的には、分析対象回のコメントにおける、過去対比の「特徴語」を抽出したうえで、それらの「特徴語」が同じコメント内で出現している頻度、即ち「共起関係」を計測し、可視化している⁴。例えば、2024 年 3 月調査からは、①物価上昇と消費動向に関する話題、②中小企業の賃上げや人手不足に関する話題、③インバウンドや春の観光需要に関する話題の、大きく 3 つのストーリーが確認できる。

【図表 2】共起ネットワーク図（2024 年 3 月）



（注）2024 年 3 月調査の現状判断に関するコメントから特徴語を抽出し、それらの共起関係を描画したもの。単語の大きさは出現頻度、単語間の線の太さは共起関係の強さを示す。

（出所）内閣府

もっとも、これらの分析は、比較的シンプルな手法⁵を用いて単語の出現頻度や共起関係を計測したものである。他方、ここ数年で、大規模言語モデル（Large Language Models：LLM）と呼ばれる文章の文脈まで考慮することができる機械学習の手法が急速に発展し、利用が広がっている。LLM は、テキストデータや画像データ等の非構造化データ⁶を分析・解釈が容易な構造化データに、正確かつ効率的に変換することができる⁷。そこで、本稿では、そうした最新の技術を取り入れた分析を実施し、LLM の経済情勢判断への応用可能性を探った。

LLM の概要

LLM とは、大量のテキストデータを入力データとする、深層学習の手法を用いて構築された自然言語処理モデルである。コンピューターによる計算量、学習のための入力テキスト量、モデルを記述するためのパラメータ数が従来の自然言語処理モデル⁸と比較して莫大であるという特徴がある。

LLM が急速に普及するきっかけとなったモデ

ルが、2018 年に Google より公表された BERT（Bidirectional Encoder Representations from Transformers）である。BERT では、それ以前の自然言語処理モデルと比較して性能が大幅に改善し、その後の LLM のベースラインとなった。その後も 2020 年に Open-AI 社により GPT-3（Generative Pretrained Transformer）、2023 年には同社より GPT-4 が公表される等、モデルの規模、性能ともに、日々進化を遂げている^{9,10}。なお、GPT は昨今急速に利用が拡大している Chat-GPT にも搭載されている LLM である。

LLM が言語を学習する際の流れは大きく 2 つの段階に分けられる。最初の段階は「事前学習」と呼ばれるもので、Wikipedia 等から作成した大量のテキストデータを用いて教師なし学習¹¹を行い、言語の基本的な知識を習得させる段階である。この段階は計算コストが非常に大きく、桁違いに大規模な計算リソースを有していないと実行困難なプロセスである。第二段階は、「ファインチューニング」と呼ばれるプロセスで、モデルの利用目的に応じて、追加でその目的に特化した学習を行い、モデルを調整するというプロセスである。この段階は計算コストが比較的小さく、大規模な計算リソースなくとも実施可能である。

LLM を用いて分析を実施する際、上記の「事前学習」をすでに実施済みのモデルを入手し、第二段階の「ファインチューニング」のみ分析目的に応じて実施し、分析を行うことが可能である。さらに、すでに利用目的に適う「ファインチューニング」済みのモデルが公開されている場合、そうしたモデルを直接分析に利用することも可能である。こうしたことが可能なのは、LLM が「事前学習」の段階ですでに、言語の文脈を読み取る能力を獲得しているためである。

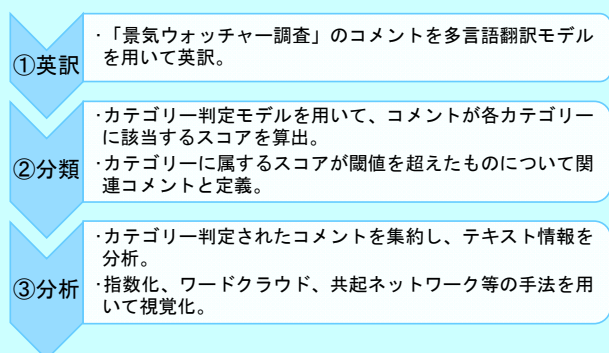
このように、LLM には、文章の文脈を考慮した精度の高いモデルを、比較的簡単に利用することができるという利点がある。こうした利点を踏まえ、海外の中央銀行では、LLM を経済情勢判断や調査研究に活用する動きがみられ始めている。例えば、オーストラリア準備銀行（Reserve Bank of Australia：RBA）では、企業の決算説明記者会見のトランスクリプトや企業との面談記録のテキストデータを対象に、LLM を用いて企業の価格設定行動の分析を行っている¹²。このほか、ニューヨー

ク連銀でも、LLM を用いて連邦公開市場委員会（FOMC）の議事録や Tealbook を対象にテキスト分析を行っている¹³。こうした海外での潮流を踏まえ、本稿では、日本経済の情勢判断における、LLM を用いたテキストデータ分析の応用可能性を検証する。

分析の流れ

ここでは、分析の流れを説明したうえで、今回利用した LLM を紹介する。これを踏まえて、次節では日本における賃金・物価動向の分析を行う。なお、本稿における分析では、「景気ウォッチャー調査」のコメントをテキストデータとして用いる。分析の流れは、図表 3 の通りである。まず、①機械翻訳用 LLM を用いて、「景気ウォッチャー調査」のコメント集を英訳する。続いて、②カテゴリ判定用 LLM を用いて、英訳した各コメントが述べているテーマのカテゴリを判定する。最後に、③判定されたカテゴリをもとに分析や指数化・視覚化を行う。

【図表 3】分析の流れ



（コメントの機械的な日英翻訳）

今回の分析は、前処理の段階でコメントの英訳を実施しているという特徴がある。分析に使用するカテゴリ判定モデルは、カテゴリ判定というタスクに特化して「ファインチューニング」が行われている点で優れているものの、対応言語が英語のみである。そこで、LLM を用いた機械翻訳を活用することで、日本語で書かれた「景気ウォッチャー調査」のコメントを分析可能な英語という形に効率的に変換した。ここでは、Meta 社が公表している、NLLB（No Language Left Behind）という機械翻訳モデル¹⁴を用いた。

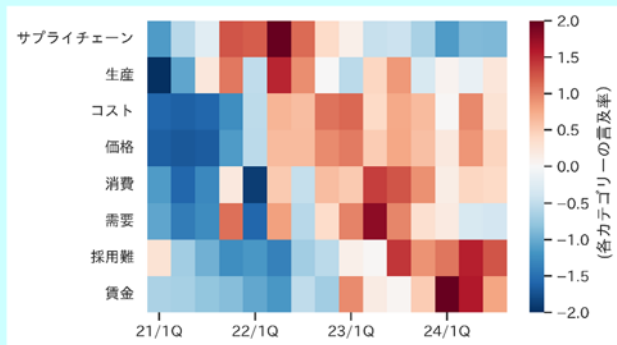
ちなみに、多くの LLM が英語テキストをもとに開発されている中、非英語圏の国々では自国語に対応し、かつ分析目的に適った LLM が必ずしも利用可能ではないという問題があった。それに対し、本稿で活用した機械翻訳モデルは、200 もの言語に対応しており、一度英訳というプロセスを経ることで、非英語圏の多くの国々に対し LLM を用いた分析の門戸を広げる可能性を秘めていると考えられる。

（コメントのカテゴリ判定）

続いて、カテゴリ判定用の LLM を用い、英訳された各コメントについて、それらが言及している話題を判定する。今回使用するモデル¹⁵は、コメントを入力すると、コメントが指定したカテゴリを含意している確率を 0~1 のスコアとして出力する（1 に近いほど当該カテゴリである確率が高い）。同様のカテゴリ判定モデルは、先述の RBA の研究でも用いられており、本稿では同研究に倣い、スコアが一定の閾値（0.7）を上回る場合に当該カテゴリに属していると判定した。

カテゴリ判定 LLM を用いたシンプルな分析例として、ヒートマップを活用したトピックの変遷の視覚化を試みた。図表 4 は、「景気ウォッチャー調査」のコメントにつき、カテゴリとして、「生産」、「価格」、「賃金」等を指定したうえで、各カテゴリに属すると判定されたコメント数の割合の推移を示している（過去平均からの乖離として基準化）。プラス（赤色）は、そのカテゴリに属するコメント数の割合が過去平均対比高いことを意味し、マイナス（青色）は、その割合が過去平均対比低いことを意味する。結果を確認すると、半導体等の供給制約が話題となっていた 2021 年下期から 2022 年下期にかけては、「サプライチェーン」や「生産」といった話題の言及率が過去平均対比高い結果となっていた。他方、2022 年下期以降は、「コスト」や「価格」といった物価に関するコメントが過去平均対比増加し、さらにごく足もとについてみると、「賃金」や「採用難」といった雇用関係に言及するコメントが過去に比べて増加している。このようにカテゴリ判定を活用することで、回答者が景況感を答えるうえで、何が論点であったかを理解することができる。

【図表 4】 カテゴリー別ヒートマップ



(注) 各カテゴリーの言及率は、各カテゴリーに属するコメントの全コメントに対する割合を標準化したもの。

(出所) 内閣府

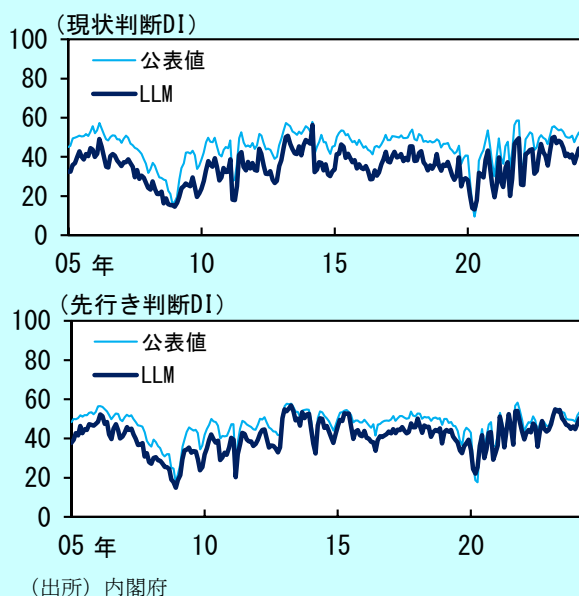
なお、カテゴリーについては、分析者が任意に設定することが可能である。例えば、「生産」、「物価」、「賃金」等のように単語そのものを指定しても良いし、「生産増加」、「物価上昇」、「賃金上昇」等のように、方向感まで含めてもよい。実際、先述の RBA の研究でも、方向感まで含めたカテゴリーを用いて分析を行っている。さらに、同手法では、カテゴリー毎にスコアを算出するため、一部のコメントは複数のカテゴリーに属していると判定されうる。そのため、あるカテゴリーAに属するコメントがカテゴリーBにも属している割合を計算することで、カテゴリー間の連関性についても検証が可能である。次節では、こうした特徴を生かしつつ、分析を深めていく。

(コメントのセンチメント分析)

ここでは、参考として、次節の分析では用いないものの、テキストのセンチメントを判定できる LLM を用いて、日本経済関連のテキストデータに対する LLM の性能を確認する。「景気ウォッチャー調査」では、回答者ごとに、景気の現状および先行きに関する 5 段階の判断とその判断理由となるコメントが同時に公表されている。それゆえ、LLM を用いて回答者のコメントのセンチメントを計測し、そのセンチメントを集計した DI が、実際に公表されている景気の現状判断 DI および先行き判断 DI を再現できていた場合、LLM の性能は高いと評価できる。ここでは、Prosus AI 社より公表されている、金融経済のセンチメント計測用にファインチューニングされた FinBERT というモデルを利用する。同モデルは英語にのみ利用可能な LLM であるため、先述のとおり機械翻訳

用の LLM を用いて日本語のコメントを機械的に翻訳したうえでセンチメントを計測する。

【図表 5】 DI



(出所) 内閣府

図表 5 は、LLM を用いて、各回答者によるコメントのセンチメント（ポジティブ、ネガティブ、中立）を判定したうえで、その結果から作成された DI を示している。ここで DI は、(ポジティブコメント比率×100+中立コメント比率×50+ネガティブコメント比率×0) / 100 として作成した。LLM を用いて算出した DI を現状判断 DI および先行き判断 DI の公表値と比較すると極めて似た動きをしており、LLM によるセンチメント計測は、回答者のコメントの文脈やセンチメントをうまく捉えていると評価できる。すなわち、今回使用した LLM は、機械翻訳、センチメント計測ともに、相応の精度を有していると考えられる¹⁶。

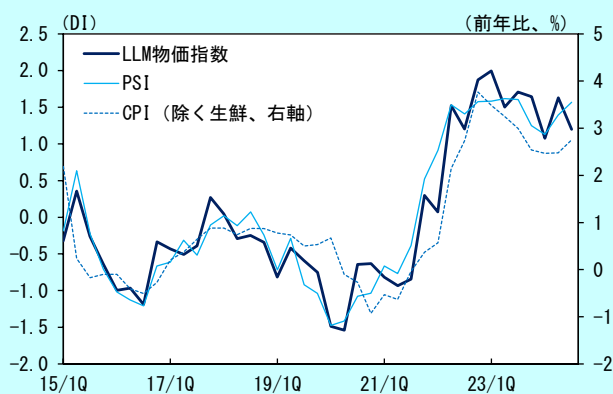
最近の賃金・物価動向の分析

前節で説明した LLM を用い、本節では経済情勢分析の一例として、「景気ウォッチャー調査」を対象とした昨今の賃金・物価動向に関する分析結果を示す¹⁷。

最初に、冒頭で紹介した PSI と同様の発想で、LLM を用いて物価指数を作成した。PSI は、「景気ウォッチャー調査」の各コメントが「価格上昇／下落を意味しているかどうか」を機械学習の手法で判定し、指数化したものであるが、ここでは各コメントが、文脈上「価格上昇／下落を含意す

るか」をカテゴリー判定 LLM で判定したうえで、指数を作成した¹⁸。

【図表 6】LLM 物価指数

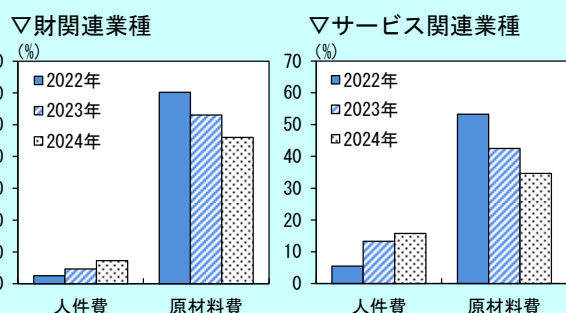


(注) LLM 物価指数および PSI は、グラフ表示期間内のデータで標準化した指数。

(出所) 内閣府、総務省

図表 6 を確認すると、LLM を用いて計算した指数は、価格上昇／下落に特化して学習させた PSI と同様の推移を示している。近年の物価上昇局面を概ねうまく捕捉できており、PSI と同様、物価動向を測るうえで有用と考えられる。なお、特定の単語を分析者が選択し、その出現状況をもとに算出された PSI と比較して、LLM は分析者が単語を選択することなくコメントの文脈を読み取ったうえで指数を作成しているため、恣意性が少ないというメリットがある。

【図表 7】物価上昇の背景



(注) 各年 1～9 月の景気の現状及び先行き判断に対するコメントについて、カテゴリー判定モデルを用いて「販売価格上昇」を含意すると判定されたコメントのうち、「人件費」ないし「原材料費」を含意すると判定されたコメントの割合を示している。財関連業種は、製造業と小売関連。サービス関連業種は、非製造業（除く建設業）、飲食関連、サービス関連、住宅関連、雇用関連。

(出所) 内閣府

続いて、昨今の値上げの背景を分析した。カテゴリー判定 LLM を用いて、「販売価格上昇を含意

する」と判定されたコメントのうち、「人件費を含意する」ないし「原材料費を含意する」と判定されたコメントの割合をそれぞれ計算した。財関連業種、サービス関連業種それぞれにおいて同割合を計算し、価格上昇の要因変化を確認した。

図表 7 をみると、いずれの業種においても、2022 年以降、原材料費に言及したコメントの割合が減少する一方、人件費に言及したコメントの割合は、低水準ながらも増加してきている。このことから、コストとして人件費が次第に意識されるようになってきたことが分かる。なお、業種間で比較すると、サービス関連業種では、財関連業種と比べて、人件費の言及が多い一方で原材料費の言及が少ない姿となっている。

さらに、カテゴリー判定モデルで「価格上昇」と判定されたコメントのワードクラウドを作成することでも、物価上昇の変化要因を確認した¹⁹。ここでは、2022 年および 2024 年において、LLM によって「販売価格上昇を含意する」と判定されたコメントを対象とした²⁰。

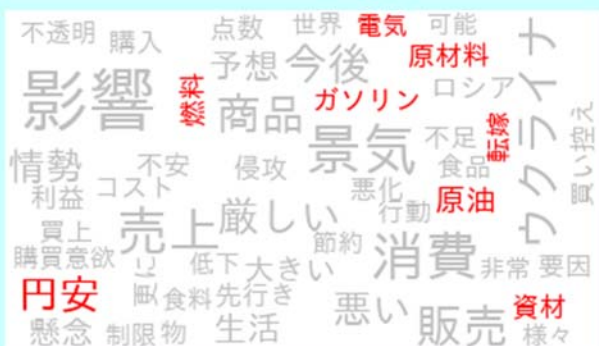
図表 8 を確認すると、2022 年には「原油」や「資材」といった原材料費関連の単語が目立った一方、2024 年には「賃上げ」や「人件」といった人件費関連の単語が目立っており、「景気ウォッチャー調査」の回答者において、価格転嫁の対象となるコスト要因の認識が原材料費から人件費へと変化してきたことが示唆される。なお、原材料費に関わる「円安」は、2022、2024 年共に多くの言及があり、物価の押し上げ要因として継続的に認識されてきたことがうかがわれる。

LLM を用いると、以上の様なテキスト分析を行う際に、分析者の事前準備の負担を抑えることができる。例えば PSI では、教師データを作成し、モデルを推計するというプロセスが必要であった。一方、LLM は、一般的な言語知識を学習済みであるため、例えばカテゴリー判定 LLM であれば、ユーザーサイドで学習の手間をかけることなく、任意の単語（今回であれば「価格上昇／下落」等）を指定するだけでカテゴリーの分類が可能である。また、今回用いた機械翻訳 LLM やカテゴリー判定 LLM を含め、ファインチューニング済みのモデル・ツールが多数公開されていることから、ユーザーは目的に応じた分析にすぐに取り掛

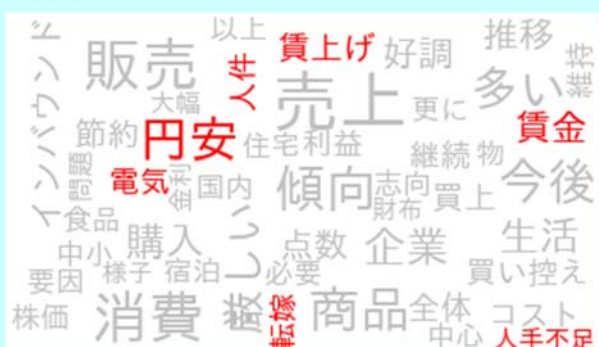
ることができる。

【図表 8】 ワードクラウド

▽2022年



▽2024年



(注) 各年1～9月の景気の現状及び先行き判断に対するコメントについて、カテゴリ判定モデルを用いて「販売価格上昇」を含意するコメントを抽出。抽出されたコメント全体と比較して、分析対象年のコメントにおいて相対的に出現頻度が高い単語をワードクラウドで描画。単語の大きさは出現頻度、赤字は値上げの背景と考えられる単語(筆者が判断)を示す。なお、ここでの対象は名詞・形容詞・副詞のみ。ただし、価格上昇関連単語(物価、物価高、上昇、値上げ、値上がり、上がる、単価、料金、高、高い、インフレ、価格、高騰)は除外。

(出所) 内閣府

おわりに

本稿では、近年急速に発展している LLM について、経済情勢判断における応用可能性について検証した。LLM は、目的に応じた多様なモデルが入手可能であるほか、言語の一般的な知識を学習済であることによる汎用性の高さ、といった利点がある。今回、本稿における LLM を用いた賃金・物価動向の分析結果からは、足もと販売価格上昇の要因として人件費が大きくなりつつあることが示唆された。

最後に、LLM を用いた分析における留意点について言及する。1 点目は、LLM を用いた分析に限

らず、テキストデータを用いた分析全般にあてはまる留意点であるが、分析対象となるテキストデータの偏り（バイアス）の問題である。テキストデータが実態を偏りなく反映していることは現実的には難しい。回答者の属性分布による偏り、同一回答者が継続して回答することによる偏り、また質問自体が誘導となることによる回答の偏りや、回答自体も例えば不満を持っている回答者のコメントは現実よりも強いといった偏りも考えられる。分析を行う際にはテキストデータが持つ偏りを絶えず確認し、偏りを認識したうえで結果を解釈する必要がある。

2 点目は、LLM を用いた分析にはノイズが多少なりとも存在するという点である。例えば、カテゴリ判定モデルにおいては、販売価格上昇と判断されたコメントの一部に、中間財価格上昇に関するコメント等も含まれていた。現時点ではこのようなノイズを完全に除去することは難しく、全てのカテゴリ判定結果が分析者の意図した結果とは限らない点に留意が必要である。また、判定プロセスがブラックボックス化しており、判定に至った事由の理解は困難である。そのため、分類されたコメントの内容をチェックし、結果が妥当性を有するかも確認していく必要がある。

3 点目は、LLM の計算コストの大きさである。LLM を用いた分析に当たっては、学習済みのモデルが容易に入手可能であるために、分析用のモデルを用意する負担は小さい。他方、「事前学習」を行う際に必要な桁違いに大きい計算リソースは不要ではあるものの、LLM を用いての分析は、従来のテキスト分析手法対比で計算負荷は大きい。数行程度のテキストデータの処理時間は短くとも、大量のデータやモデルの規模次第では、従来の手法対比で相応の計算リソースや時間が必要となる点は留意する必要がある。なお、LLM の分野は技術進歩が極めて速く、新たな高性能モデルが次々とリリースされている。LLM を用いた分析を行う際は、最新の動向をフォローし、分析の目的に適うモデルを選択することも重要である。

以上、本稿では LLM のメリットやデメリットを挙げつつ、経済情勢判断への LLM の利用可能性について検証した。情勢判断においては統計データが基本となるものの、本稿の分析の様に LLM によるテキスト分析等、非構造化データを含めた

* 現・大阪支店

¹ オルタナティブデータを用いた日本銀行のリサーチについては、以下のページにまとめられている。

<https://www.boj.or.jp/research/bigdata/index.htm>

また、こうしたリサーチの概観については、以下の資料を参照。

亀田制作（2021）「オルタナティブデータを用いた日銀リサーチの紹介」、日銀レビュー・シリーズ、No.2021-J-16

² PSI の詳細については、以下の資料を参照。

中島上智・山縣広晃・奥田達志・香月信之輔・篠原武史「景気ウォッチャー調査のテキスト分析からみた企業の短期インフレ予想」、日本銀行ワーキングペーパー・シリーズ、No.21-J-12

³ PSI の作成には、ナイーブベイズ分類器と呼ばれる機械学習手法を用いている。

⁴ 詳細は、以下の資料を参照。

大高一樹・菅和聖（2018）「機械学習による景気分析—『景気ウォッチャー調査』のテキストマイニング—」、日本銀行ワーキングペーパー・シリーズ、No.18-J-8

⁵ 例えば、PSI の算出は、あくまで言及されている「単語」のみに着目して行われており、コメントの「文脈」までは考慮していない。

⁶ Excel ファイルや CSV ファイル等にきれいに整形されていない状態のデータ。そのため、非構造化データそのものを直接分析に用いることは難しい。

⁷ LLM をはじめとする、深層学習技術の経済学への応用については、以下の資料を参照。

Dell, M., "Deep Learning for Economists," *Journal of Economic Literature*, forthcoming.

⁸ 従来の自然言語処理に使用される手法としては、PSI の作成に用いているナイーブベイズ分類器のほか、深層学習を使用した手法としては、RNN（Recurrent Neural Network）やその派生形である LSTM（Long Short-Term Memory）を応用したモデル等が挙げられる。

⁹ BERT のパラメータ数が 3 億 4000 万であるのに対し、GPT-3 では 1750 億、GPT-4 では非公認ながらも GPT-3 をはるかに上回るパラメータ数といわれている。

¹⁰ 自然言語処理の性能を測る指標として、GLUE（General Language Understanding Evaluation）Benchmark というものがある。文法の正しさの判定、センチメントの判定、文章の含意関係の判定等のタスクの総合評価で自然言語処理モデルの性能をスコア化しており、人間のスコアを上回る LLM が登場している。なお、GLUE は英語表現を評価するためのベンチマークであるが、日本語表現を評価するベンチマークとしては、JGLUE がある。

¹¹ データそのものからパターンを学習する手法。これに対し、教師あり学習は文章とラベルの対となっているデータをもとに、文章が属するラベルを予測できるよう学習を行う手法。例えば、先述の PSI は、物価上昇／下落といったラベル付きテキストを用いてモデルを推計しており、教師あり学習を用いたモデルである。

¹² Windsor, C., and M. Zang (2023), "Firms' Price-setting Behaviour: Insights from Earnings Calls," Reserve Bank of Australia Research Discussion Paper, 2023-06.

¹³ Fischer, E., R. McCaughrin, S. Prazad, and M. Vandergon (2023),

"Fed Transparency and Policy Expectation Errors: A Text Analysis Approach," Federal Reserve Bank of New York Staff Reports, No.1081.

¹⁴ 具体的には、Hugging Face 社のウェブサイト上にて公表されている"facebook/nllb-200-distilled-600M"を用いた。

¹⁵ 具体的には、Hugging Face 社のウェブサイト上にて公表されている"MoritzLaurer/DeBERTa-v3-base-mnli-fever-anli"を用いた。同モデルは、Microsoft 社が開発した DeBERTa-v3-base というモデルをベースとしている。本稿では、経済に関する幅広い話題に対して汎用的に分析を行う観点からシンプルな汎用モデルを用いたが、より専門的な話題を扱う際には、学習データを丁寧に確認しつつ、分析用のモデルを選択することが肝心である。また、より規模の大きい最先端のモデルを用いれば、更に精度が改善すると考えられる。

¹⁶ 一般論として、センチメントの計測結果については、同じ事象に対し経済主体によって受け止め方が異なるために解釈が難しい場合がある。例えば「賃金上昇」は、消費者にとってはポジティブな事象でも、企業にとってはコスト上昇というネガティブな事象としても捉えられうるため、コメントする立場によってセンチメントが逆となりうる。こうしたコメント主体の属性を確りと見極めつつ、センチメントの計測結果を利用する必要がある。

¹⁷ 日本における最近の賃金・物価動向の分析としては、たとえば、以下の資料を参照。

尾崎達哉・神保真宏・八木智之・吉井彬人（2024）「賃金・物価の相互連関を巡る最近の状況について」、日銀レビュー・シリーズ、No.2024-J-2

¹⁸ 文章が「価格上昇」、「価格下落」を含意するかをそれぞれ判定させ、スコアが 0.7 を上回った場合に価格上昇／下落を含意すると判定されたとした。指数は（価格上昇コメント数－価格下落コメント数）／総コメント数として算出した。

¹⁹ LLM によるテキストの判定では、単語の出現頻度をベースとした機械学習の手法と比較して、判定メカニズムがブラックボックス化しており分かりにくいという課題がある。この課題への対処としては、判定後のテキストを対象にワードクラウドや共起ネットワーク図等を用いて単語の出現頻度を調べ、正確に分類できているかを検証することが有効である。

²⁰ ワードクラウドを描画する際には、「販売価格上昇」に関連する単語を除外している。これは、「販売価格上昇」と分類されるコメントを対象に描画しているため、関連単語がワードクラウドに大きく表示されてしまい、分類されたコメントの中でのそれ以外の傾向の変化が見にくくなってしまうためである。

日銀レビュー・シリーズは、最近の金融経済の話題を、金融経済に関心を有する幅広い読者層を対象として、平易かつ簡潔に解説するために、日本銀行が編集・発行しているものです。ただし、レポートで示された意見は執筆者に属し、必ずしも日本銀行の見解を示すものではありません。

内容に関するご質問等に関しましては、日本銀行調査統計局（代表 03-3279-1111）までお知らせ下さい。なお、日銀レビュー・シリーズおよび日本銀行ワーキングペーパー・シリーズは、<https://www.boj.or.jp> で入手できます。