

経済データを用いた政府の景気基調判断文の生成 とバイアスの検証

Development of a model for generating the government's "Assessment of the current state of the economy" based on economic indicators and analysis of its tendencies

末廣徹^{1,2} 木村柚里¹ 稲垣真太郎¹

Toru Suehiro^{1,2}, Yuri Kimura¹, and Shintaro Inagaki¹

¹ みずほ証券金融市場調査部

¹ Fixed Income Research Department, Mizuho Securities, Co., Ltd.

² 法政大学大学院経済学研究科経済学専攻

² Graduate School of Economics, Hosei University

Abstract: In this paper, we examine whether it is possible to generate the government's "Assessment of the current state of the economy" by the Japanese government using several economic indicators that are believed to be consistent with the Japanese economic trend. In the Monthly Economic Report released by the Cabinet Office, the government publishes the government's "Assessment of the current state of the economy" along with expressions such as "recovering," "worsening," and "standstill", etc. In this paper, we developed a model to generate the government's "Assessment of the current state of the economy" and tested the accuracy of the predictions. In addition, we examined which economic indicators are likely to influence the government's "Assessment of the current state of the economy". It is expected that this analysis will be useful in predicting the government's policy decisions.

1. はじめに

政府や日銀は景気や物価の状況を文章で示す「基調判断」を公表している。日銀は四半期に1度の「展望レポート」で、政府は内閣府から発表される「月例経済報告」でそれぞれ示している。いずれも、政府や日銀による経済環境の評価を知る上で、金融市場において注目されることが多い。例えば、政府による景気判断が悪化すれば、経済をテコ入れするための経済対策が行われる可能性が高いという予想が増えるだろう。もっとも、政府や日銀の「基調判断」がどのような分析に基づいて決定しているのかは釈然としない面もある。そこで、本稿では様々な経済データをインプットとして、政府が「月例経済報告」で示している「基調判断文」をアウトプットするAIモデルを作成し、政府の「基調判断文」がどのように決定されているのかという実験を行った。研究分

野としては構造化データの活用をテーマとした研究分野の1つであるデータを自然言語に変換し、理解可能性を拡張する“Data-to-Text”である。

金融経済の数値データから文章を生成するという試み自体は過去にも分析された例があるものの（[1]など）、過去の研究は市場参加者に対する情報提供や業務効率化という側面が強い。一方、本稿が分析対象とする日本政府の基調判断文の「クセ」を学習することで、事前に政府の景気認識の変化やそれに伴う政策変更の予想に役立てることができるなど、フォワードルッキングな利用も大きなモチベーションとなっている。

2. 関連研究

政府や日銀のテキストデータを用いた分析は複数行われてきたが、多くの研究がテキストデータをセンチメントなどの数値データに変換することで、景気や市場の動向を予測することがテーマとなってきた。[2]は「月例経済報告」の言葉に注目し、その変化から景気指標を作成した。[3]は日銀の「金融経済月報」から金融市場の動向を分析した。[4]は、景気

連絡先: みずほ証券株式会社金融市場調査部
〒100-0004 東京都千代田区大手町 1-5-1 大手町ファーストスクエア
E-mail: toru.suehiro@mizuho-sc.com

ウォッチャー調査をもとに、月例経済報告や展望レポートをベースとした指数を作成した。これらの先行研究のように、基準を付けたり比較したりすることが困難であるテキストデータを辞書や深層学習を用いて数値データに変換することは、テキストマイニングの分野では王道であり、多くの研究が蓄積している。

他方、本稿が目している“Data-to-Text”の応用については、応用例は少ない。[1]は市場データ（日経平均）を用いて概況テキストを自動生成する研究であり、研究のプロセスは本稿と同様である。ただし、[1]は人間が行っていたタスクの自動化を目的としている一方、本稿は“Data-to-Text”のモデルを作成することで、人間が行っている文書作成の「クセ」を読み解くことが目的である。このような研究の蓄積はこれまで見られていない。

3. 使用データ

内閣府月例経済報告の1998年1月から2019年8月で示される景気の基調判断のテキストデータを用いた。月例経済報告は景気に対する日本政府の公式見解を示す資料であり、「月例経済報告等に関する関係閣僚会議」において経済財政政策担当大臣を中心に議論されて内容が決まる。表1に「景気の基調判断」の例を示す。「景気の基調判断」では景気動向指数も参考にされるものの、CI一致指数などによって機械的に基調判断が決定するわけではない。従って、日本政府の基調判断はどのような基準で作成されているのか不透明な点が多いことで知られている。

基調判断の文章の構成についても決まりがあるわけではなく、「景気は、回復している」などシンプルな構成の場合もあれば、「景気は、輸出や生産に弱さがみられるが、持ち直している」など様々な付帯条件が記されることもある。とはいえ、ほとんどのケースにおいて基調判断は「景気は」という語で始まる文章であり、文の最後は「弱含んでいる」「改善に足踏みが見られる」「持ち直しの動きが見られる」などのように結論が示されていることが特徴となっており、一定の精度で文章の生成が可能であると期待できる。なお、分析に用いたデータセットでは、「東日本大震災の影響により」や「復興需要等を背景として」といった経済データだけでは判断しにくいとみられる文言はセンテンスから取り除いた。他にも、輸出や個人消費、生産と言った細かい表現については景気全体の水準や方向性とは直接的な関係がないことから、これもデータから除いた。さらに、2センテンス以上から構成されているコメントについては、初めのセンテンスのみを取り出してターゲットデータとした。

表1 月例経済報告の基調判断の例文

景気は、おおむね横ばいとなっているが、イラク情勢等から不透明感が増している。
景気は、おおむね横ばいとなっている。株価やアメリカ経済の動向など、我が国の景気を巡る環境に変化の兆しがみられる。
景気は、依然として厳しい状況にあるものの、復興需要等を背景として、緩やかに回復しつつある。

テキストデータについては、助詞、助動詞、非自立的動詞、接続詞は区切らず、手前で出現したトークンに接続した（例えば下記）。

[“景気は” , “,” , “緩やかに” , “回復しつつ” , “ある”]

作成したトークンをリストに格納し、ID化することで埋め込み処理を可能にした。

一方、文章生成の元となる経済データは、景気動向指数のCI一致指数の個別系列(9系列)を中心に、CI先行系列(11系列)、CI遅行系列(9系列)を用いた。CI(Composite Index)は生産、消費、雇用など経済活動での重要かつ景気に敏感に反応する様々な指標を統合して1つの指標にすることによって、景気の現状把握を行うために作成されている。

景気動向指数には、景気に対して先行して動く先行指数、ほぼ一致して動く一致指数、遅れて動く遅行指数の3つの指数がある。そのうち一致指数のデータを中心に内閣府の景気動向指数研究会での議論を踏まえ、景気循環の転換点である景気基準日付(景気の山・谷)を設定している。政府の景気に対する基調判断と、景気動向指数研究会の判断は必ずしも一致しないものの、一般的に景気の基調判断を決めるデータセットとしては過不足がないものと言える。

指標データに関して、本稿では当月を除く過去12ヶ月分を入力データとした。当月分については、月例経済報告が公表された段階で経済指標が発表されていないものが多いことから、基調判断を決める際の参考にすることはできないからである。データセットの合計は260で、ここから80%を訓練データ、20%を検証用データとして学習を行ったすなわちモデル入力データのサイズは208である。

なお、指標データを用いる際には、月次の差分を標準化してから用いた。

4. モデル

本稿では文書生成タスクについて encoder-decoder モデルを使用した。Encoder, Decoder 内部のアーキテクチャにはどちらも GRU を使用した。GRU とはゲート機構を持つ回帰型ニューラルネットワークであり、LSTM と同等の精度が期待されるが、出力ゲートが特に小規模なデータセットに対しては GRU が有効とされている。

モデル構築における条件について、パラメータの最適化手法には Adam を使用し、ミニバッチのサイズを 12, encoder の隠れ状態の次元を 256, epoch 数を 100 とした。形態素解析器には Mecab, モデルの実装には Pytorch を使用した。

モデルの評価には BLEU-4 スコアを用いた。データサイズが限られているため評価データを用いない評価を試みた。具体的には validation data に対する BLEU-4 スコアの最大値を取得するという試行をそれぞれ 10 回行い、その平均値を求めた。結果は 60.3 となり、良好とされる目安である 40 を上回った²。

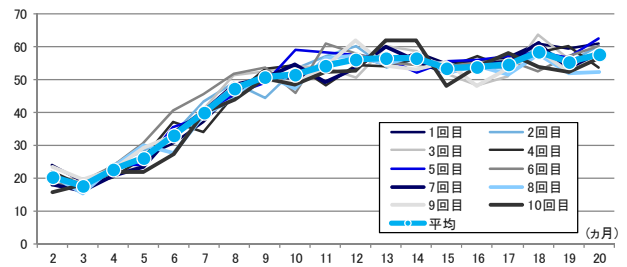
5. 分析

5-1. 経済データの範囲

モデルの作成条件と BLEU-4 のスコアを比較することで、月例経済報告の基調判断の特徴を分析した。まずは、学習データである経済指標 (CI 一致指数の個別 9 系列) の過去データは何ヵ月分用いるのが最も再現性が高いかを調べた。最も基調判断分の再現性の高いデータ入力する数を調べることで、政府が過去何ヵ月分のデータセットを見て基調判断文を考えているのかを予想することができる。

訓練データを 2~20 ヶ月分まで変化させて BLEU-4 スコアを比較すると (図 1), 2~12 ヶ月までは過去データの数を増やした分だけ BLEU-4 スコアが上昇する一方、それ以降はデータ数を増やしてもスコアの改善が限定的である。つまり、過去 12 ヶ月より前のデータを追加しても基調判断の生成にはほとんど影響を与えていないことが分かった。政府が基調判断文を作成する際には、過去 12 ヶ月程度のデータが参照されている可能性が高いと言えよう。

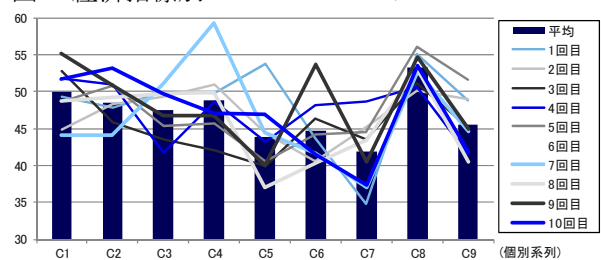
図 1 訓練データの過去データ数と BLEU-4 スコア



5-2. 基調判断に影響を与える指標

モデル作成に用いた経済指標 (CI 一致指数の個別 9 系列)のうち、1 つだけを訓練データとして BLEU-4 スコアを比較すると、図表 2 のようになった。CI 一致指標の 9 系列は「C1: 生産指数 (鉱工業)」「C2: 鉱工業用生産財出荷指数」「C3: 耐久消費財出荷指数」「C4: 所定外労働時間指数 (調査産業計)」「C5: 投資財出荷指数 (除く輸送機械)」「C6: 商業販売額 (小売業) (前年同月比)」「C7: 商業販売額 (卸売業) (前年同月比)」「C8: 営業利益 (全産業)」「C9: 有効求人倍率 (除く学卒)」だが、C7 (商業販売額) の BLEU-4 スコアが低く (再現性が低い)、C8 (営業利益) が高い (再現性が高い) ことが分かった。他方、その他の指標間ではそれほどスコアの差がなかった。この結果から、政府が基調判断文を作成する際には、CI 一致指数の 9 系列の中では「営業利益」が重視されている可能性が高いと言えよう。

図 2 経済指標別の BLEU-4 スコア

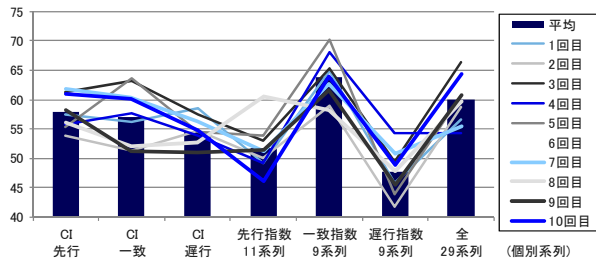


さらに、学習データの範囲を CI 一致指数以外にも広げた結果を検証した (図 3)。CI 先行指数を用いた場合の BLEU-4 スコアは CI 一致指数と同様に良好な結果 (再現性が高い) となることが分かった。一方、CI 遅行指数のスコアは他に比べやや低い (基調判断に対する影響は限定的) ことが分かった。

また、個別系列をそれぞれ別のデータとして学習した場合は、CI 一致系列の 9 系列を用いた場合の BLEU-4 スコアが最も高かった。僅差ながらも全 29 系列すべてを用いた場合よりも高くなったことから、基調判断を予想する場合は CI 一致指数の個別系列の変化を確認することが特に重要と言えよう。

² 本稿におけるモデルの詳細は[5].

図3 景気動向指数（CI 先行指数，CI 一致指数，CI 遅行指数）別の BLEU-4 スコア



5-3. 訓練データの期間による比較

最後に、政府による基調判断文の作成において、構造変化があったかどうかの検証を行った。政府の基調判断文は常に同じ判断基準で公平に行われていることが望ましいことは間違いないが、基調判断文作成のルールが明文化されているわけでもなく、そのときの政権によって基調判断文の傾向が変わる可能性もあるだろう。そこで、直近で政権交代が生じた（民主党政権から自民党の安倍政権に変わった）タイミングの前で「基調判断の判断基準が変わったかどうか」を検証した。

具体的には、12年末までのデータセットを学習データとしたモデルを作成した上で、13年1月以降のデータセット（過去12ヵ月分のCI一致指数の個別9系列）から想定される基調判断の生成を行った。安倍政権のアベノミクス以降のデータを学習データから外すことにより、アベノミクス前の判断基準を反映したモデルを作成することができる。仮に判断基準が変わっていないのであれば、実際の基調判断を再現することが可能であるはずである。逆に、再現ができなければ政府の基調判断の判断基準に変化があった可能性が疑われる。

19年以降の基調判断について、実際の基調判断とアベノミクス前のモデルを使って生成した文を比較すると（表2）、明らかにモデルの生成文の方が弱気

な表現が並んだ。つまり、アベノミクス以降（13年1月以降）は基調判断の判断基準が変化し、より強気（楽観的）になっている可能性が高いと言える。

19年12月の基調判断については、実際は「弱さが一段と増しているものの」という警戒文がついてはいるが、「緩やかに回復している」という表現は前月から維持された。しかし、アベノミクス前のモデルを使った生成文は「一層厳しさを増している」という非常に弱い表現となった。

また、実際の基調判断を用いたスコア（スコアの作成方法は5段階評価）と、アベノミクス前のモデルが生成した文のスコアを比較すると（図4）、後者の方がCI一致指数と連動していることが分かった。つまり、アベノミクス前まではCI一致指数に連動した基調判断が発表されていたとみられる一方、アベノミクス以降は基調判断の判断基準が強気化したことで連動性がなくなってしまったと言える。

少なくとも、アベノミクス前後において、基調判断文の作成のルールが政府内で変化した可能性が高い。以上から、モデルの作成期間を変更することにより、構造変化が生じている可能性を示すことができた。

図4 実際の基調判断スコアと、アベノミクス前のモデルが生成した基調判断スコア

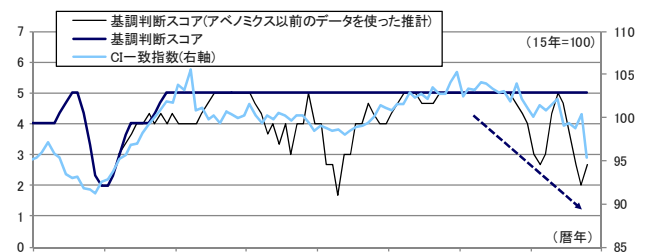


表2 19年以降の基調判断と、アベノミクス以前の訓練データを用いたモデル推計値

	実際の基調判断の文	アベノミクス前のモデルを使った生成文
19年1月	景気は、緩やかに回復している	景気は、依然として厳しい状況にある
19年2月	景気は、緩やかに回復している	景気は、このところ停滞している
19年3月	景気は、このところ輸出や生産の一部に弱さもみられるが、緩やかに回復している	景気は、一部に弱い動きが続いており、回復が緩やかになっている
19年4月	景気は、このところ輸出や生産の一部に弱さもみられるが、緩やかに回復している	景気は、一部に弱い動きが続くものの、緩やかに回復している
19年5月	景気は、輸出や生産の弱さが続いているものの、緩やかに回復している	景気は、一部に弱い動きがみられ、このところ回復が緩やかになっている
19年6月	景気は、輸出や生産の弱さが続いているものの、緩やかに回復している	景気は、依然として厳しい状況にあるが、一部に持ち直しの動きがみられる
19年7月	景気は、輸出を中心に弱さが続いているものの、緩やかに回復している	景気は、このところ弱い動きとなっている
19年8月	景気は、輸出を中心に弱さが続いているものの、緩やかに回復している	景気は、このところ弱い動きとなっている
19年9月	景気は、輸出を中心に弱さが続いているものの、緩やかに回復している	景気は、このところ弱い動きとなっている
19年10月	景気は、輸出を中心に弱さが長引いているものの、緩やかに回復している	景気は、依然として厳しい状況にあるものの、このところ上向きの動きがみられる
19年11月	景気は、輸出を中心に弱さが長引いているものの、緩やかに回復している	景気は、引き続き低迷状態が長引き、はなはだ厳しい状況にある
19年12月	景気は、輸出が引き続き弱含むなかで、製造業を中心に弱さが一段と増しているものの、緩やかに回復している	景気は、引き続き停滞し、一層厳しさを増している

6. 結論

本稿では、日本の経済指標によって日本政府の基調判断文のテキストデータを生成することができるのか、という検証を行った。数値データである経済指標から基調判断文のテキストデータを作成するという“Data-to-Text”のタスクにおいて、深層学習を使うことにより、一定の精度で文章生成が可能であることが分かった。

また、作成したモデルを用いて政府の基調判断文作成の特徴やバイアス（いわゆる「クセ」）を検証した。具体的には、(1)入力する経済指標のデータの時系列方向でのデータ数を変更することにより、政府は概ね過去 12 ヶ月程度のデータを参考に基調判断文を作成している可能性が高いことを示し、(2)入力する経済指標の種類を変化させることにより、政府は「営業利益」の項目を重視している可能性が高いことを示した。さらに民主党政権から自民党の安倍政権に政権交代したタイミングの前後（アベノミクス前後）でモデルを分割したところ、アベノミクス以降の基調判断文はそれ以前と比較して景気に対して強気な基調判断文を選択していたことを示した。

本稿が分析対象とした政府の基調判断文に限らず、データを文章で説明するケースは少なくない（各国中央銀行の声明文など）。深層学習を用いたモデルを作ることにより、各種基調判断文を予測することは市場参加者にとって有益だろう。また、本稿のようにモデルを検証することによって基調判断文の妥当性を議論することも可能である。数値データと異なり、文章の正しさに疑義がある場合は、解釈の仕方などによって水掛け論になってしまうことが多いが、モデルを作成することによってその一貫性を問うことも可能になる。本稿の研究がこのような応用例の一例となるだろう。

参考文献

- [1] 村上聡一郎, 渡邊亮彦, 宮澤彬, 五島圭一, 柳瀬利彦, 高村大也, 宮尾祐介: 時系列数値データからの概況テキストの自動生成, 言語処理学会 第 23 回年次大会 発表論文集, 2017
- [2] 山澤成康: 景気指標としての月例経済報告, JCER Discussion Paper 124, 日本経済研究センター, 2009
- [3] 和泉潔, 後藤卓, 松井藤五郎: 経済キスト情報を用いた長期的な市場動向推定, 情報処理学会論文誌 Vol.52 No.12, 3309-3315 頁, 2011
- [4] 山本裕樹, 松尾豊: 景気ウォッチャー調査の深層学習を用いた金融レポートの指数化, 第 30 回人工知能学会全国大会, 2016

- [5] 木村柚里, 末廣徹, 稲垣真太郎: 小規模データセットを用いた景気の基調判断文自動生成, 言語処理学会 第 26 回年次大会, 2020