

アジア太平洋研究所資料
23-05

「テキストデータを利用した S-APIR 指数の実用化」

研究会報告書(2022 年度)

2023年4月

一般財団法人 アジア太平洋研究所

〈はじめに〉

本報告書は、2021年度「テキストデータを利用した新しい景気指標の開発と応用」研究会と2022年度「テキストデータを利用したS-APIR指数の実用化（2022年度より研究会名称変更）」研究会の2年間の研究成果をまとめたものである。

昨今、国内外の経済を取り巻く不確実性は急速に増しており、足元の経済動向を正確に把握することは極めて困難となりつつある。一般的に、一国の経済のパフォーマンスを把握するためには、GDPや物価を始めとするマクロデータが用いられている。しかし、マクロデータには、経済の動向に対する認知ラグが存在し、経済主体の行動要因が捨象されている。このため、マクロデータを補完する形で、経済動向を迅速かつ詳細に把握する方法を新たに構築する必要がある。こうした問題意識の下で、本研究ではテキストデータを利用した新しい景気指標の開発と応用を2018年から継続的に行っている。

以下では、最初にS-APIRの予測を、代替的データセット(Web上でのNews情報)の下で行い、その頑健性を確認した。次にWeb上でのS-APIRの予測値の観察、単語ベースの足元の景況感への影響度の検索が即座に行えるデモ・システムの構築の準備を行った。こうした実践的研究を国際的な査読誌に公表するとともに、学会での報告を行った。8月には、日本経済研究センター主催の「AI・ビッグデータ経済モデル研究会」からの招待講演の要請を受け、報告を行った。

2023年4月

〈キーワード〉

テキストデータ、深層学習、景気動向調査

〈研究体制〉

研究統括	稲田 義久	アジア太平洋研究所 数量経済分析センター センター長 甲南大学 名誉教授
リサーチリーダー	関 和広	アジア太平洋研究所 主席研究員 甲南大学知能情報学部 教授
リサーチャー	松林 洋一	アジア太平洋研究所 上席研究員 神戸大学大学院経済学研究科 教授
	生田 祐介	大阪産業大学経営学部 准教授
事務局	中山 明	アジア太平洋研究所 研究推進部 総括調査役
事務局	吉田 茂一	アジア太平洋研究所 研究推進部 部員

役職名は2023年3月末時点のもの

〈執筆者〉

- 第1章 松林 洋一
- 第2章 生田 祐介, 関 和広
- 第3章 生田 祐介
- 第4章 生田 祐介, 関 和広
- 第5章 関 和広

概要

本稿は、アジア太平洋研究所のプロジェクト「テキストデータを利用した新しい景気指標の開発と応用」の2021年度研究成果と「テキストデータを利用したS-APIR指数の実用化(2022年度より名称変更)」の2022年度研究成果を、5章構成の報告書としてまとめたものである。第1章では、テキストデータを利用した新たな景況感指標を開発する意義を明確にするとともに、想定されるアウトプットと研究成果を予め紹介している。第2章では、景況感指標の構築方法を簡潔に解説している。第3章では、S-APIRの特徴を明確にするために、他の景気指数(マクロ指標、セミマクロ指標、その他指標など)との丁寧な比較を行っている。第4章では、テキストデータ(新聞記事)に登場する様々な語彙がS-APIRにどのように影響しているのかという点を、代表的語彙を用いて検証している。第5章では、新聞紙面に掲載される記事ではなく、ウェブニュースサイトで配信・公開される記事データを自動的に収集し、S-APIR指数の算出を行うシステムについて説明する。本分析は、我々の研究において画期的な試みであり、今後広範な利用が期待できる。

2023年4月

一般財団法人 アジア太平洋研究所

「テキストデータを利用したS-APIR指数の実用化」研究会

リサーチリーダー 関和広

(主席研究員／甲南大学知能情報学部 教授)

2022 年度「テキストデータを利用した S-APIR 指数の実用化」 研究会報告書

〈はじめに〉	i
〈キーワード〉	ii
〈研究体制〉	ii
概要	iii
第1章 テキストデータを利用した新しい景気指標の開発と応用	1
1. 研究の背景	1
2. 想定するアウトプット	1
3. 内容	2
4. 期待される研究成果	3
第2章 経済ニュースによる景況感指数の構築	4
要旨	4
1. はじめに	4
2. 関連研究	5
3. 景況感指数(S-APIR 指数)の予測	5
4. 評価実験	9
第3章 S-APIR 指数と他の景気指数との比較	23
1. 概要	23
2. マクロ経済指標との比較	23
3. セミマクロ経済指標との比較	25
4. 結論:S-APIR の特徴	26
第4章 S-APIR の語彙による要因分解	31
1. 概要	31
2. 指数の要因分析	31
3. 結果:景況感へのイベントの寄与度	32
4. 考察:イベントの寄与度の頑健性	34
5. 結論	34
第5章 景況感足元予測システム	38
1. 概要	38
2. システムの構成	38
3. システムの機能と使用例	39

研究会記録.....	45
研究成果.....	45

第1章 テキストデータを利用した新しい景気指標の開発と応用

松林 洋一

1. 研究の背景

従来、マクロ経済の動向を把握するには、集計データを用いることが一般的である。しかし、集計データは、リアルタイム性に欠けており、ミクロの経済要因を知るには不十分という課題がある。一方、昨今の情報技術の急速な進展により、国内外の経済活動において生成される大規模なデータ(ビッグデータ)が様々な形で利用可能になり始めている。きわめて豊富な情報を内包しているビッグデータの活用は、マクロ経済のより精緻な情勢判断と予測において、有効であると考えられる。このため、本研究ではビッグデータの一つであるテキストデータを利用して、経済の動向を把握することを試みる。

2. 想定するアウトプット

これまでの研究では、深層学習を用いることで、新聞記事上の大量のテキストデータから景気に関する指数(S-APIR 指数バージョン 1)を推定し、その頑健性を定量的に評価した。これを発展させる形で、以下に示す4つのアウトプットを予定して作業を進めてきた。

- ① 新聞記事上の少量のテキストデータ(記事の見出しとリード文程度)を利用した新たな景況感指数(S-APIR 指数バージョン 2)を開発する。これは、月次の政府統計が公表される前の時点で、景気動向を把握できる超短期予測である。
- ② S-APIR 指数を基に景気動向への単語寄与度を、オンライン検索できる「デモ・システム」を開発する。
- ③ 2021年上期までの日本経済新聞の記事を用いて、S-APIR 指数バージョン 1の分析期間を延長する。
- ④ 上記の①から③の実施で得られた成果を、査読付き学術雑誌で公表する。

3. 内容

① 分析の手法

本研究で基本となる成果物は、テキストデータから推定された景気関連指数（S-APIR 指数）である。指数を推定するため、昨年から引き続き、人工知能の一種である深層学習を用いる。深層学習のモデルとして、これまでと同様にリカレント・ニューラル・ネットワーク(Recurrent Neural Network,以下 RNN)に加え、Google 社が開発した最新の学習モデルである BERT (Bidirectional Encoder Representations from Transformers)を用いる。BERT は、RNN のように単語の順序を考慮した上では学習することはせず、文中の全ての単語同士の依存関係を学習する。その処理を基本として、S-APIR 指数の各バージョンを推定するモデルを構築する。

② 本研究のセールスポイント、具体的な工夫等

第一に、低コスト・低タイムラグである。景況感を代理する既存の指数は、CI 一致指数や消費者態度指数のように、多大な労力を投入したアンケート調査から構築されて、月次で報じられている。ところで、経済情勢の要因は一カ月単位ではなく、日々刻々と変化している。ある月の景況感指数を見て、結果的に経済情勢が前月と異なることに気付いたとしても、景気の転換局面が“いつ”であったのか、“その時”に気づくことはできない。本研究は、そのような「認知のタイムラグ」という課題を克服するため、景気に関連する「S-APIR 指数」を提案する。この指数は2つの特徴を有する。まず、既存の新聞記事から自動的に出力されるため、多くの人員を必要としない（低コスト）。さらに、日次で報じることができるため、日々起こりうる景気の転換局面を把握することが可能となる（低タイムラグ）。

第二に、単語を通じた景気動向の要因の可視化である。景気動向は、金融政策や物価、賃金、売上、来客数、あるいは他国の状況など、様々な要因により構成される。しかし、全ての要因が経済に等しく影響を与えているわけではない。経済の情勢判断を行う上で、経済を上向きあるいは下向きに動かす支配的な要因が何であるのかが知ることが重要である。そこで、S-APIR 指数への寄与度の高い単語に注目すると、その単語の表す要因が経済において支配的であると見なせる。さらに、単語の寄与度は時系列で表されるため、どの要因がどの時期に経済に影響を及ぼしているかを知ることが可能となる。

4.期待される研究成果

新聞記事のテキストデータから景況感を推定するモデルを構築し、その出力値を S-APIR 指数と称している。これを政府による既存の景況感指数と比較することで、我々のモデルが有する特徴を明らかにする。その結果を踏まえて、「S-APIR 指数」を一般に公表していく。景気動向を代理する「S-APIR 指数」を見ることで、企業の経営判断を行う際の議論に使えるようにする。そして、国や自治体に対しても、政策決定に活用して頂くことを検討する。

ビジネスや政策上の判断に役立つ手段として、5 章で説明される Web サーバを通じた「景況感足元予測システム」の利用が該当する。ユーザー自身が、システムへ興味ある単語を入力すると、その単語が S-APIR 指数にどのような影響を与えているのか知ることができる。例えば、「東京五輪」という単語を入力した場合、ミクロの波及メカニズム(例 建設需要を増やすのかどうか)まで見ることはできないが、東京五輪が最終的に景気動向へ正の影響を及ぼすのかどうかを調べるための手がかりとなる。

本研究では特にテキスト情報に着目し、日々発信されるニュース記事に基づいた新しい景況感指数(S-APIR 指数と呼ぶ)を提案する。また、予測された指数を用いて、特定の概念(語句)が景況感にどのように影響を与えたのか時系列で分析する。以降の構成は次の通りである。2 章では、テキスト情報を用いた金融・経済指数予測の研究についてまとめる。3 章では、本研究で提案する景況感指数予測の枠組みについて述べる。4 章では、評価実験について報告し、単語レベルでの景況感への寄与についても議論する。最後の 5 章では、景況感足元予測システムについて、その仕組みと実用例について解説する。

第2章 経済ニュースによる景況感指数の構築¹

生田 祐介, 関 和広

要旨

政府や中央銀行が定期的に発表する景況感指数は、金融当局の政策決定や企業の生産計画などの拠りどころとして重要な役割を担っている。しかしながら、これらの指数は大規模なアンケートに基づいており、その算出には大変な手間と時間がかかる。そこで本研究では、日々発行されるニュース記事を再利用することで、低コストで速報性の高い、新しい景況感指数を算出する手法を提案する。具体的には、外れ値検出モデルによって経済・景気に関係が薄いと判断されたニュース記事を除外したうえで、それ以外の記事（の文）に対して景気スコアを予測する。そして、ある期間に発行された記事すべての景気スコアを統合して、その期間の景況感指数を定義する。景気スコアの予測には、近年自然言語処理タスクで広く利用されている Bidirectional Encoder Representation from Transformers (BERT) をファインチューニングして用いる。さらに、算出された景況感指数を基に、任意の概念（語句）がどのように景況感に影響を与えたのかを時間軸に沿って分析する手法を提案する。約 12 年分のニュース記事を用いた実験により、提案手法の妥当性と有用性を示す。

1. はじめに

政府や中央銀行が定期的に発表する景気ウォッチャー 調査の現状・先行き判断 DI (diffusion index; 景気動向指数)²や日銀短観の DI³ などの経済指数は、金融当局の政策決定や企業の生産計画、機関投資家・個人の投資判断など、様々な経済活動のよりどころとして重要な役割を担っている。しかしながら、これらの指数算出の基礎となる調査には大変な手間とコストがかかる。

たとえば、景気ウォッチャー調査の現状・先行き判断 DI (以降「景気ウォッチャー DI」と呼ぶ) の場合、北海道、東北、北関東、南関東、甲信越、東海、北陸、近畿、中国、四国、九州、沖縄の 12 地域を対象に、地域の景気に関連の深い動きを観察できる立場にある人々 (小売店の店主、タクシー運転手など) 2,050 人を

¹ 本章の内容は、関・生田[7]と Seki et al.[10]を加筆修正したものである。

² https://www5.cao.go.jp/keizai3/watcher/watcher_menu.html

³ <http://www.boj.or.jp/statistics/tk/>

対象に地域別調査機関がアンケート調査を行い、とりまとめ調査機関において地 域ごとの調査結果を集計・分析し、指数の算出を行っている。また、このような調査は相応の時間を要するため、景気ウォッチャーDI は 1 カ月に 1 度、日銀短観は四半期に 1 度しか発表されず、速報性が低い。

2. 関連研究

山本・松尾[8] は、景気ウォッチャー調査を利用して景況感(野村AI景況感指数)の予測を行った。具体的には、Long Short-Term Memory (LSTM) ユニットから構成される両方向・多層のRecurrent Neural Network (RNN) で景気ウォッチャー調査の判断理由文から景気判断を予測するモデルを学習し、政府の月例経済報告と日銀の金融経済月報から抽出したテキストを、同モデルを用いて指数化し、これらの結果が日銀短観DIおよび景気ウォッチャーDI(現状)と高い相関(0.782~0.843)を持つことを示した。

山本・松尾[8]の研究の発展として、饗場・山本[1]はTwitterから景気に関するツイートを抽出し、同様のモデルで景況感(SNS×AI景況感指数)を予測した。近藤他[5]と坂地他[6]は、金融機関の社員が取引先企業との面談を通して作成したテキストを入力として、同様のモデルで景況感を予測した。五島他[4] は、畳み込みニューラルネットを利用し、ロイターのニュース記事から景況感を予測した。

3. 景況感指数(S-APIR 指数)の予測

本研究では、五島他[4]と同様にニュース記事を利用し、より低コストで速報性の高い景況感指数(S-APIR 指数)の推定を試みる。ただし、本研究は、記事のフィルタリングを行うことで、より正確な景況感の予測を目指すという点に特色がある。図2-1で、景況感指数の予測を行う枠組みを示す。図の左側が「モデルの学習」、右側が「景況感指数の予測」の流れである。以降の節でそれぞれについて説明する。

3.1 モデルの学習

山本・松尾[8]と同様に、内閣府の景気ウォッチャー調査のデータから、景気判断および判断理由文のペアを抽出し、モデルの学習を行う。景気ウォッチャー調査では、集計された指数(景気ウォッチャーDI)に加えて、個々のアンケート回答者が判断した現状および先行きの景気の状態(「◎(良い)」から「×(悪い)」まで5段階評価)と、その理由(自然言語文)が公開されている。いくつかの例を抜粋して表2-1に示す。

なお、景気ウォッチャー調査には現状判断と先行き判断とが含まれ、両者では理由文の特徴が異なると考えられる。本研究では現状の景況感指数の足元予測を目的とするため、現状判断のデータのみを用いる。また、景気判断は◎、○、□、△、×の記号で表現されているため、これらをそれぞれ 2, 1, 0, -1, -2 として扱う。

景気ウォッチャーDIは、表1の「景気」の5段階の判断にそれぞれ点数を与え、これらの点数を各判断区分の構成比(%)に乗じた和で算出されている。算出されたDIは、50を基準(景気に変化なし)として0から100の値をとる。なお、景気ウォッチャー調査では、景気の現状判断と先行き判断を区別して回答させるため、それぞれに指数が算出される。景気ウォッチャーDIの例を図2-2に示す。

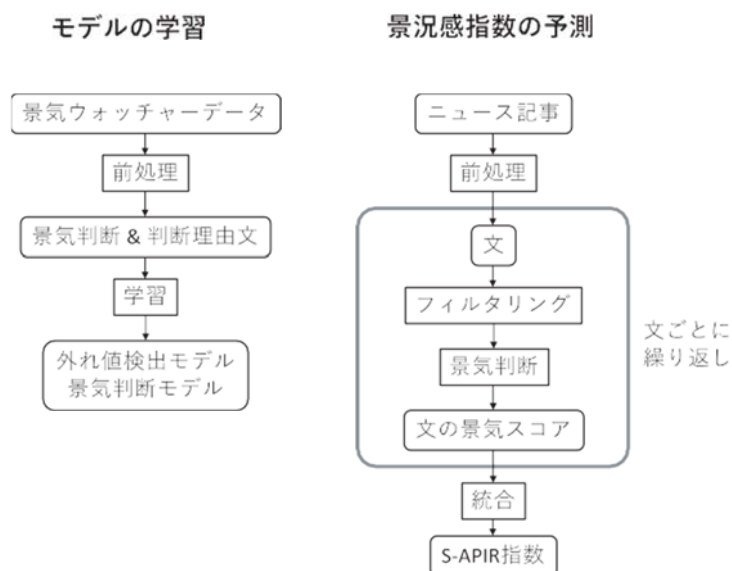


図 2-1 S-APIR 指数予測の枠組み

地域	回答者	景気	判断理由
北海道	タクシー運転手	×	売上が減少しているが、季節要因に加えて、景気が下向していることも影響している。
北関東	輸送用機械器具製造業	◎	自動車は米国への輸出が増えている。

表 2-1 景気ウォッチャー調査(現状の景気)の例

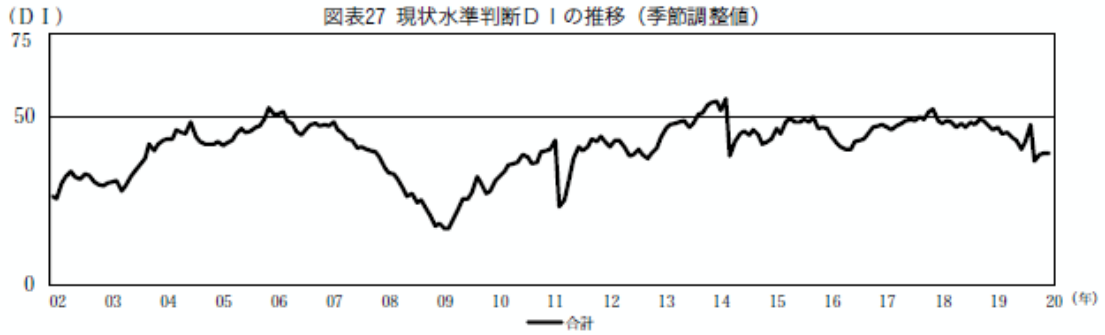
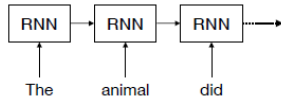


図 2-2 景気ウォッチャーDI の例

従来の言語モデルとの違い

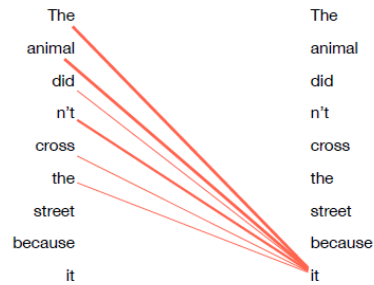
- 従来の言語モデル (RNN)

例: The animal didn't cross the street because it was too tired.



次に現れる単語を予測するモデル。RNNでは単語を順番にモデルに入力することで、前方の文脈を考慮する。

- BERT (自己注意機構)



BERTではすべての単語を一度にモデルに入力し、すべての単語同士の依存関係を推定する。

図 2-3 RNNとBERTの違い

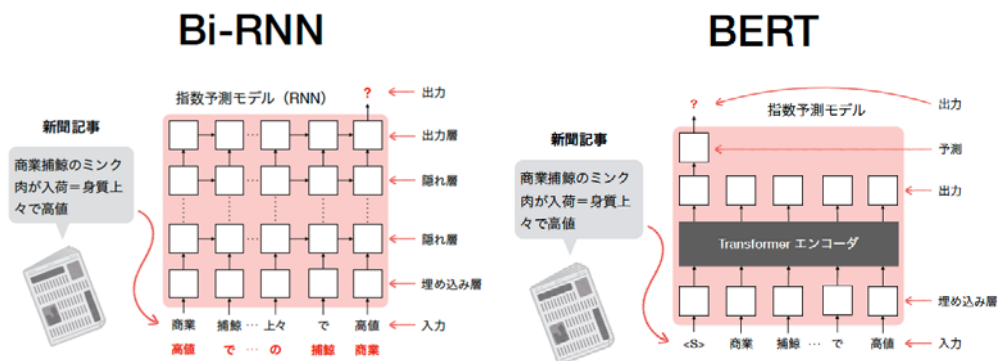


図 2-4 Bi-RNNとBERTの枠組み

抽出したデータを学習データとして、2つのモデルを学習する。1つはテキストデー

タから景気判断を行うモデルであり、もう 1 つは外れ値検出モデルである。両者を併用することにより、S-APIR 指数を算出する。

前者の景気判断を行うモデルとして、Bidirectional RNN (以降、Bi-RNN)と Bidirectional Encoder Representations from Transformers (以降、BERT)を利用した。両者の結果を比較し、予測性能の高い方を本研究の指数予測モデルに採用する。これまでの本研究プロジェクトでは、新聞記事を学習モデルに入力して景気代理指数を予測するために、RNN を基本モデルとして実験を進めてきた。他方で、BERT は、近年の機械学習において、処理の様々なタスクでスタンダードになりつつある学習モデルである。BERT は、RNN のように単語の順序を考慮して学習することはせず、文中の全ての単語同士の依存関係を学習する⁴。本研究では、日本語の大規模なコーパスで事前学習された BERT⁵に景気判断を連続値で予測する出力層を加え、景気ウォッチャー調査を学習データとしてモデル全体の重みをファインチューニングする⁶。従来の RNN と BERT との違いを図 2-3 に、RNN の応用である Bi-RNN と BERT の枠組みを図 2-4 に、それぞれ示す。

後者の外れ値検出モデルは、入力テキストのフィルタリングに用いる。本研究で景況感指数の予測に利用するデータはニュース記事であり、ニュースには多様な記事が含まれるため、外れ値検出によって経済や景気に関する文だけを選択的に

⁴ BERT は、Vaswani et al.[13]で述べられているように、Transformer とよばれる構造を利用した言語表現モデルであり、マスクされた単語を周辺の文脈から予測するというタスクを解くことで、大量のコーパスを用いて事前学習を行う。このようにして得られた初期的な言語表現モデルに、解きたいタスクに応じた出力層を加えてファインチューニングすることで、多くのタスクにおいて従来の手法を上回る性能が報告されている。ところで、BERT という名称に含まれている Encoder という用語は、機械学習で頻繁に登場する。この機会に Encoder と、それと関連する用語について、それぞれ説明しておこう。

エンコーダ(符号化:Encoder)とは、高次元のデータから共通する特徴を抽出することによって、データの次元数を減らす機能のことである。エンコーダを、人による効率的な文章読解に例えてみよう。100文字で構成される長文を効率よく読解しようとする場合、主語と動詞に該当する箇所を見つけ、そこにどのような言葉があるのか知ろうとするだろう。主語と動詞に該当する箇所にある言葉が、その文を判断する際の特徴に相当する。

これに対してデコーダ(復元化:Decoder)とは、低次元に圧縮されたデータを高次元に戻す機能のことである。エンコーダを通じて入力された大量のデータは、特徴を持った少量のデータに集約され、再度デコーダを通じて大量のデータに復元される。例えば、文を特徴づける主語と動詞に該当する部分を利用して、元の100文字の長文に戻す、という作業である。

オートエンコーダ(自己符号化:Autoencoder)とは、エンコーダとデコーダを合わせて使うことにより、出力を入力に近づけるよう学習するニューラルネットワークのことである。

⁵ <https://github.com/cl-tohoku/bert-japanese>

⁶ 事前学習とは、Wikipedia のように大量に存在しているテキストデータを用いて、文法や単語の意味等を学習させることである。この学習では人間が正解データを用意する必要がなく、データを収集してくるだけでよい。人間に例えると、幼少期の言葉を覚える段階に相当する。

利用することで、より正確な指数の予測を試みる。外れ値検出のモデルには、1 クラス Support Vector Machine (SVM) を利用する。通常の SVM では 2 クラス分類問題を扱うのに対して、1 クラス SVM では、1 クラスのデータだけを学習データとして分類器を学習し、学習データと類似していないデータを外れ値として検出することができる。そこで、景気ウォッチャー調査の景気判断理由文を与えて外れ値検出モデルを学習し、景況感指数算出の際は、景気判断理由文と類似している文だけを経済・景気に関係する文(景気関連文)と見なして、指数の算出に利用する。1 クラス SVM に用いる入力文の表現方法としては、tf-idf (term frequency-inverted document frequency) で重み付けしたベクトル空間モデルを用いる。

3.2 指数の予測

S-APIR 指数の予測にはニュース記事を用いる。ニュース記事は前処理として句点「。」をもとに文に分割し、文ごとに外れ値検出モデルに入力する。その結果外れ値と判定された文は除外し、それ以外の文(景気関連文)を景気判断モデルに入力する。その結果、文ごとに景気スコアが出力される。出力された景気スコアを日ごとや月ごとにまとめて平均値を算出し、それを日次や月次の S-APIR 指数とする。本稿では、比較の対象として月ごとに発表される景気ウォッチャーDI を用いるため、月次の S-APIR 指数を算出する。

4. 評価実験

4.1 学習データと実験設定

景況感予測モデルの学習のため、内閣府のホームページ⁷から、2000 年 1 月～2020 年 2 月までの景気ウォッチャー調査の景気判断理由文(現状)計 254,823 件を収集した。このうち、無作為に選んだ 90% を学習データ(検証データを含む)、10% をテストデータとした。モデルのファインチューニングの際のパラメタは、バッチサイズ b を 16, 32, エポック数 e を最大 5 とし、学習データの 90% をモデルの学習、10% を検証データに用い、検証データで最も平均二乗誤差(MSE)が小さかったモデル($b = 32, e = 3$)を採用した。なお、入力文の長さ(単語数)については、実験に用いた環境で実行可能な 200 単語とした。

なお、前述のように、BERT の事前学習は日本語 Wikipedia で行われたモデルを利用した。Wikipedia は本研究で景況感の予測に用いるニュース記事とは語彙や表現が異なると考えられるため、BERT の事前学習に日経新聞を加えた実験も行った。しかし、予測性能はほとんど変化しなかったため、本稿では結果の報告は割愛す

⁷ <https://www5.cao.go.jp>

る。S-APIR 指数の推定には、日経新聞 2008 年 1 月～2020 年 6 月に発行された本紙朝刊および本紙夕刊の記事見出しと記事本文を用いた。

4.2 景気ウォッチャー調査によるモデルの学習と検証

検証のため、景気判断とその理由文で学習した景気判断モデルを用いてテストデータで景気判断の予測を行い、実際の景気判断(-2~2)との MSE を算出した。結果を表 2-2 に示す。比較のため、以下で説明される二つのモデル(リッジ回帰, LSTM-BiRNN)についても同様の実験を行った。

- リッジ回帰:入力データは tf-idf で重み付けしたベクトル空間モデルで表現し、正則化パラメタ α は 1.0 とした。
- LSTM-BiRNN:LSTM を隠れ層のユニットに用いた両方向のリカレントニューラルネットワークであり、山本・松尾[8]の先行研究に相当する。予備実験の結果から、単語埋込を 300 次元、隠れ層を 2 層、隠れ層のユニット数を 512 とした。

単語の文脈を無視したベクトル空間モデルを用いたリッジ回帰、および既存研究で用いられた Bi-RNN と比較して、景気ウォッチャー調査のデータで再学習した BERT の誤差が最小となった。この結果から、他の言語処理タスクと同様に、本研究で対象とする景気判断予測についても BERT の優位性が確認できた。なお、本稿では景気判断を連続値で予測したが、青嶋・中川[2]は◎, ○, □, △, ×の 5 クラスの分類問題として同様に BERT の優位性を報告している。

モデル	MSE
リッジ回帰	0.494
LSTM-BiRNN [26]	0.355
BERT (本研究)	0.321

表 2-2 景気ウォッチャー指数の推定結果

4.3 景況感指数 S-APIR の予測

前節の比較実験により、景気ウォッチャー調査の景気判断理由文から実際の景気判断を従来のモデルより高精度に予測できることが確認できた。続いて、このモデルを用いて、3.2 項に述べた方法で S-APIR 指数を算出し、既存の景気ウォッチャー DI と比較する。

なお、S-APIR 指数は必ずしも景気ウォッチャー DI を置き換えることを目的としているわけではなく、ニュース記事を利用した新しい景況感指数という位置づけである

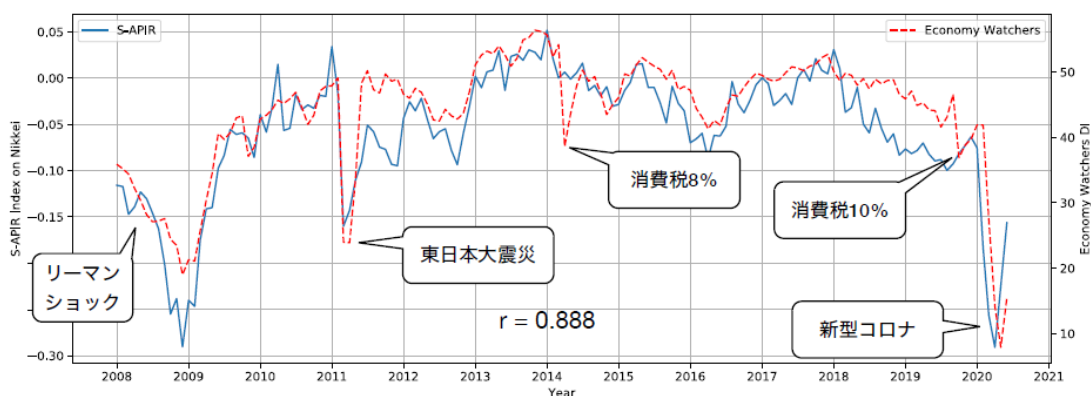
点に注意を要する。景況感に正解はなく、景気ウォッチャーDI も 2,050 名の限られた回答者の景気判断に基づく指数でしかない。本節での景気ウォッチャーDI との比較の目的は、S-APIR 指数が既存の景況感指数に類似したトレンドを持っているか確認すること、またその特性を探ることにある。

4.3.1 実験結果

4.2 節で学習した外れ値検出モデルおよび景気判断モデルを用いて、2008 年 1 月～2020 年 6 月までに発行された日経新聞の記事見出しと記事本文を入力として S-APIR 指数を算出した。比較のため、図 2-5 に、S-APIR 指数と内閣府発表の景気ウォッチャーDI (現状) を示す。リーマンブラザーズの経営破綻に端を発する金融危機や東日本大震災による景況感の低下など、S-APIR 指数はおおむね景気ウォッチャーDI のトレンドに近い動きを示しており、実際に両者の相関係数も 0.888 と高い正の相関があった。今回実験に利用したデータは一般の経済・社会情勢を伝える全国紙であり、新聞記事だけをを入力として、景況感指数の算出を主目的とする景気ウォッチャー調査に近い結果が得られたことは特筆すべきであり、S-APIR 指数の妥当性・有用性を示すものである。

2008~2020/06 のデータ

景気ウォッチャーDI (全体) との比較



2008~2020/06 のデータ

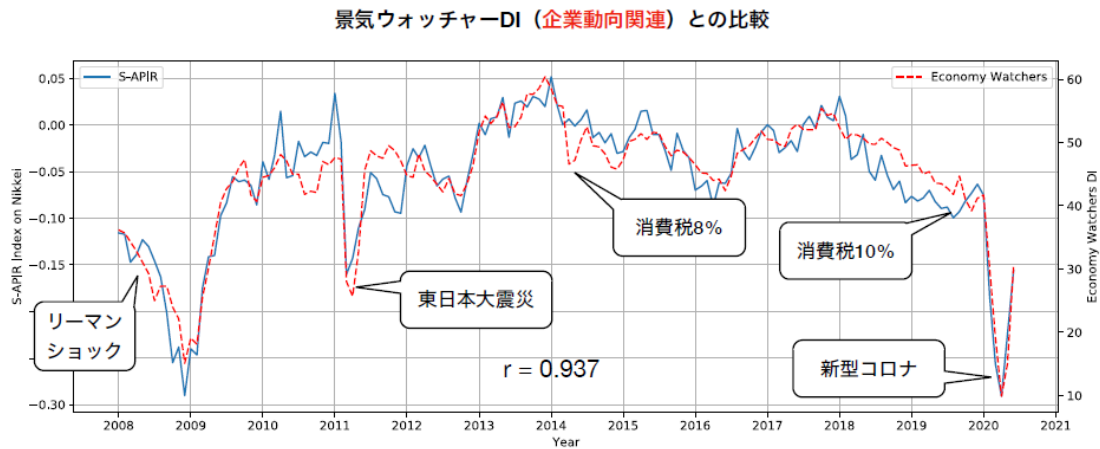


図 2-5

S-APIR 指数と景気ウォッチャー DI (現状,全体) の比較 (相関係数 = 0.888) (上)

S-APIR 指数と景気ウォッチャー DI (現状,企業) の比較 (相関係数 = 0.937) (下)

ところで、景気ウォッチャー調査の回答者は、約 7 割が家計動向関連、約 2 割が企業動向関連、約 1 割が雇用関連の業種に就いており、家計動向の影響が相対的に大きい。よって、景気ウォッチャーDI も家計動向の影響をより強く受けた指数であるといえる。一方、本研究で指数の推定に用いた日経新聞は経済紙であるため、ビジネス関連の記事が多いと考えられる。そこで、企業動向関連業従事者のみから算出された景気ウォッチャーDI と S-APIR 指数を比較したところ、相関係数は 0.888 から 0.937 に上昇した。この結果は、日経新聞から算出した S-APIR 指数は、より企業動向を反映した景況感指数であることを示唆している。

4.4 追加実験と考察

4.4.1 学習データと予測性能の関係

BERT のエンコーダは大規模なデータで事前学習が行われているため、タスクごとのファインチューニングでは、少量の学習データで比較的高精度のモデルが学習できる。景況感の予測において、学習データが十分であるのか、あるいは、学習データを増やすことで性能向上の余地があるのかを調べるため、学習データ量と MSE および相関係数の関係を調べた (前者は景気ウォッチャー調査のテストデータにおける MSE, 後者は景気ウォッチャーDI との相関係数である)。結果を図 2-6 に示す。景

気判断モデルについては MSE(図中点線)が学習データ 20 万件程度までは減少している。他方, 学習したモデルから算出された景況感指数の相関係数(図中実線)については, 14 万件程度で学習したモデル以降は頭打ちになっている。このことから学習データ量は十分であり, さらなる学習データの増加による効果は見込めないと考えられる。

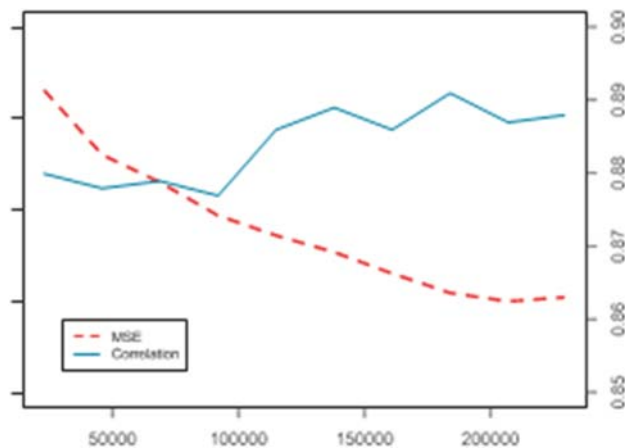


図 2-6 学習データ量と MSE および相関係数の関係

4.4.2 外れ値検出モデル

経済・景気に関係する文のみを景況感の予測に利用するため, 本研究では外れ値検出モデルとして 1 クラス SVM を用い, モデル入力文のフィルタリングを行った。その有効性を検証するため, フィルタリングを行わず, すべての記事を利用して S-APIR 指数を算出した。この結果を景気ウォッチャーDI と比較したところ, 相関係数は 0.878 であった(フィルタリングを行った場合は 0.888)。また, 企業動向関連のみの景気ウォッチャーDI との相関係数は 0.919 であった(同 0.937)。フィルタリングを行わない場合, 全体, 企業動向関連のいずれの場合も相関係数が低下し, ニュース記事には景況感の予測には適さない情報が含まれているという予想を裏付ける結果となった。

なお, 先行研究で用いられた LSTM-BiRNN で景況感指数を算出した場合, 景気ウォッチャーDI との相関係数は, フィルタリングなしで 0.765, フィルタリングありで 0.875 であった。また, 企業動向関連のみの景気ウォッチャーDI との相関係数は, フィルタリングなしで 0.805, フィルタリングありで 0.922 であり, いずれの場合も BERT を利用したときよりもフィルタリングの効果が大きかった。この結果から, 景気・経済に無関係な記事が混在していても, BERT は LSTM-BiRNN よりも頑健な景気判断予測ができていることが分かる。

さらに、本研究で採用した 1 クラス SVM の妥当性を評価するため、LSTM-RNN オートエンコーダを用いた外れ値検出モデルとの比較実験を行った。このモデルは、長短期記憶 (LSTM) を隠れ層のユニットに用いた回帰型ニューラルネットワークで、入力データ復元時の誤差の大きさで外れ値を検出する。学習時のパラメタとして、単語埋め込みベクトルの次元数 d を 100, 200, 300, 隠れ層のユニット数 u を 32, 64, バッチサイズ b を 8, 16 とし、各パラメタ値の組合せを試行した。学習データには景気ウォッチャー調査の 90% (4.1 節参照) を利用し、残り 10% のテストデータで最も損失が少ないモデル ($d = 300, u = 64, b = 16$) を採用した。

なお、外れ値検出の定量的な評価実験を行うためには、正常値を持つ事例と外れ値を持つ事例からなるデータセットが必要である。しかしながら、そのようなデータセットを手で作成するのはコストが高い。そこで、ここでは正常値の事例として外れ値検出モデルの学習に用いていない景気ウォッチャー調査のデータ (2020 年 3~8 月分) 7,962 件を利用した。外れ値については、毎日新聞 2019 年版の中で「芸能」というカテゴリが付与されている記事 14,912 文を簡易的に外れ値と見なして利用した。もちろん、芸能カテゴリの記事にも景気に関する記事はありうるものの、他のカテゴリと比較すれば関係しない記事 (外れ値) が多いと考えられる。結果を表 2-3 に示す。再現率、適合率、F1 値は、正常値と外れ値を 2 つのクラスと考えたときの分類性能 (マクロ平均) である。なお、オートエンコーダについては、出力値 (データ復元時の誤差の大きさ) に閾値を設けて正常値か外れ値か判断する必要があるため、表 2-3 には複数の閾値を試して最も F1 値が高かったときの性能を示している。

モデル	再現率	適合率	F1値
オートエンコーダ	0.928	0.858	0.892
1 クラス SVM	0.929	0.916	0.923

表 2-3 外れ値検出モデルの性能比較

表 2-3 の結果から、1 クラス SVM の方が外れ値検出の性能が高いことが分かる。直感的には、文脈をある程度考慮できる RNN ベースのオートエンコーダが有利であると考えられるが、ある文が経済・景気に関するものかそうでないかを区別するという単純な本タスクにおいては、単語レベルの特徴で十分であることが示唆される。なお、ここで学習したオートエンコーダを外れ値検出に用いて 2008 年 1 月~2020 年 6 月の景況感指数を算出したところ、景気ウォッチャー DI との相関係数は 0.875 であり、外れ値検出を行わない場合 (0.878) と比べてむしろわずかに相関係数が悪化する結果となった。

4.4.3 S-APIR の先行性・速報性

景気ウォッチャーなどの景況感指数は、景気の動向を迅速に把握することを一つの目的として作成されている。そのため、景気ウォッチャーDI に対して S-APIR 指数の動きが先行していれば、景気動向の変化をいち早くとらえる先行指数として価値がある。そこで、相互相関係数 (cross-correlation) によって S-APIR の先行性・遅行性を評価した。図 2-7 に結果を示す。なお、相互相関は正規化されていないため、 $-1 \sim 1$ の範囲には収まらない。

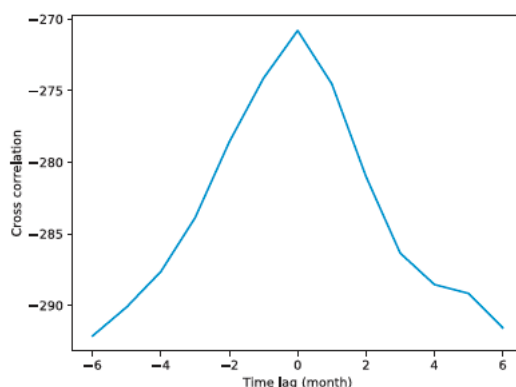


図 2-7 S-APIR 指数と景気ウォッチャーDI の相互相関関係

この結果から、時差なし (time lag が 0) の場合の相関が最大であり、少なくとも月単位では S-APIR 指数は景気ウォッチャーDI の (先行も遅行もない) 一致指数であることが分かった。ただし、ある月の景気ウォッチャーDI の公表は翌月の 10 日前後であるのに対して、S-APIR 指数はその月の末日に算出可能であり速報性が高い。そのため、たとえば景気ウォッチャーDI が発表されるまでの暫定指数としての価値もある。また、本稿では月次で算出したが、S-APIR 指数は日次・週次などでの算出も容易であるため、より粒度の高い景況感予測も可能となる。

情報源	フィルタリング	相関係数
日経新聞	なし	0.873
	あり	0.892 (+2.1%)
毎日新聞	なし	0.738
	あり	0.817(+10.7%)

表 2-4 情報源として利用する新聞別の景況感と景気ウォッチャーDI との相関係数の関係。
(カッコ内の数字はフィルタリングの有無による相関係数の増加率)

4.4.4 一般紙による景況感予測

これまでの実験では、景況感を予測するという目的に鑑み、経済や産業関係の記事が多い専門誌として日経新聞の記事を基に S-APIR 指数を算出した。その結果、S-APIR 指数は既存の景気ウォッチャーDIと高い相関を示し、特に企業動向関連業の DI とより高い相関を持つことが明らかになった。

本節では、情報源の違いによる景況感指数の違いを調査するため、一般紙を入力として景況感の予測を行い、日経新聞を用いた場合の結果と比較検討した。具体的には、毎日新聞本社版 2008 年 1 月～2019 年 12 月（毎日新聞記事データ集 08 版～19 版）の記事を利用し、フィルタリングの有無も条件に加えて S-APIR 指数を算出した。結果を表 2-4 にまとめる。なお、比較のために日経新聞を用いた場合の結果も表に含めている。この結果が 4.3 節とわずかに異なるのは、毎日新聞に合わせて 2020 年 1～6 月の記事を分析対象から除外したためである。

この結果から、いくつかの重要な観察ができる。1 つは、経済紙である日経新聞の方が景況感の予測に有用なことである。もう 1 つは、フィルタリングの効果が毎日新聞の方が大きいことである。具体的には、日経新聞ではフィルタリングによって相関係数が 2.1% 向上しているのに対して、毎日新聞では 10.7% の向上を示している。この結果は、一般紙である毎日新聞の方が経済・景気に無関係な記事を多く含んでいるという直感と一致する。

なお、試みとして日経新聞と毎日新聞の両方を使って S-APIR 指数を算出したところ、いずれかの新聞を単独で用いるよりわずかながら相関係数が向上した ($R = 0.899$)。毎日新聞のように、個々では必ずしも最適な情報源ではなくても、組み合わせることで他のより有用な情報源（ここでは日経新聞）に補完的に働く可能性が示されたことはビッグデータ利用の観点からも興味深い。

第2章の付録：RNN

BERT との比較実験に用いた Bi-RNN は、テキストの両方向から入力して予測する学習モデルであり、先行研究で指数の予測に用いられている⁸。名称から予想できるように、Bi-RNN は RNN (Recurrent Neural Network) を基本としている。ここで、Bi-RNN への理解を進めるために、RNN の構造について簡単に復習しておこう⁹。

⁸ 山本・松尾[8]の研究に相当するモデルである。

⁹ RNN の詳細に関心のある方は、APIR から発行されたディスカッションペーパーである生田他[3]を参照のこと。

付図 2-1 の上の図は、RNN を構成する基本的なニューラルネットワークを表しており、複数の層に分けて処理することによって入力から出力を予測させる。埋め込み層では、コンピュータに入力された「来客数が増えてこない」という文を、単語に分解して、それを数字に変換することで解析できるようにする。単語をベクトル空間に埋め込むことから、埋め込み層という名が付いている。隠れ層では、大量の情報を圧縮して重要な特徴を抽出する。そして、出力層では、景気を表す指数が Δ ($=-1$) に近づくように、隠れ層から受け取った特徴間で重みを柔軟に設定する。この例では、学習したモデルに当初の文を入力すると、予測される指数は -0.73 である。なお、実際の RNN の学習は、文を一度に読み込むのではなく、下の図のように文字列の順に読み込む¹⁰。

付図 2-2 の上の図は、RNN の詳細であり、隠れ層に GRU(Gated Recurrent Unit)という計算コストを抑えるユニットが含まれている。この図では、埋め込み層が 300 次元、各隠れ層のユニット数が 512 であるとして、「来客数が増えてこない」という文を学習する。ところで、GRU の歴史を見ると、古典的な RNN と、LSTM(Long Short-Term Memory)というユニットを取り入れた RNN の延長線上にある。下の図は、古典的な RNN と LSTM を比較したものであり、LSTM は古典的な RNN よりも、様々な長い文を文脈に沿って学習できる、つまり可変長の時系列データを解析できる、という特徴を持つ。ただし、ご覧の通り、LSTM ユニットには多くの矢印が描かれている。これは、LSTM は処理の種類が多く、計算コストが高いことを意味する。こうした課題への対処として、LSTM の特徴を持ちながらコントロール作業を減らすことで計算コストを抑えることができる GRU が開発された。

ここからは、RNN を用いて新聞記事から景気指数を予測する手順について解説する。付図 2-3 は、基本的な予測方法を表している。まず、景気ウォッチャー調査(景気判断とその理由文のペア)で学習したモデルに、検証のために再度同じ理由文を入力すると、 -0.73 という景気判断が予測として出力されるとする(上の図)。そして、学習済みモデルに、新聞記事を入力して景気指数を予測したい(中の図)。下の図の通り、学習した RNN はそのまま予測に使われる。これら一連の学習と予測を、手法 A と称することにしよう。

手法 A により指数を出力することができるが、精度は高くないかもしれない。その原因は、予測に用いる新聞記事にある。新聞には景気に関する記事以外にも、様々なジャンルの記事が含まれる。よって、景気に関係しない記事を含めたすべての記事を予測作業に用いることが、指数予測の精度を落としていると考えられる。この問題に対処するため、新聞記事の中から景気動向に関する記事のみを取り出して予測する、という方針を考える。

景気動向に関する記事のみで予測することは、技術的に、景気に関する記事以外

¹⁰ データの順序が分析上重要であることから、RNN は時系列データ分析に近い。

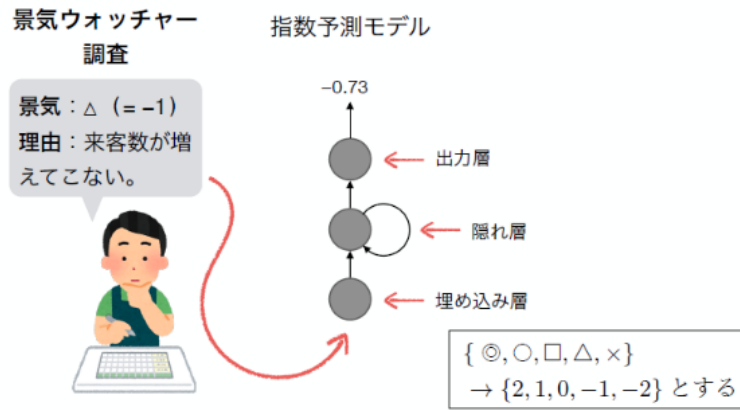
の記事を異常値として除外できるような、異常値検出問題として扱うことができる。具体的に、様々な文体や表現のテキストの中から、景気に関する記事以外の記事をふるいにかけて、景気動向に関する記事のみを抽出できるような、1クラスSVMと呼ばれるモデルを、別途学習する。この場合の1クラスSVMの学習では、景気動向に関する文の特徴さえ学べば良いので、学習データには景気ウォッチャー調査の景気判断理由文を使う。学習した1クラスSVMで新聞記事をフィルタリングさせると、異常値を除いた文を抽出することができる。フィルタリングされた文のことを、以降では景気関連文と呼ぶことにする。そして、学習したRNNに景気関連文を入力することで、景気指数を予測する。これら一連の学習と予測のまとまりを手法Bと称し、それは付図2-4で表されている。

手法Bでは、新聞記事の景気関連文のみを利用することで、指数の予測精度を向上させることを試みたわけだが、課題は残る。それは、モデルの学習に用いた景気ウォッチャー調査の景気判断理由文と新聞記事では、用いられている表現や文体が異なるということである。そこで、新聞記事によりモデルを再学習することで、新聞記事に表れる表現に適切な重みを与えて、予測できるようにしたい。これは、新聞記事によるモデルのファインチューニングといい、モデルを新聞記事に慣れさせることである。

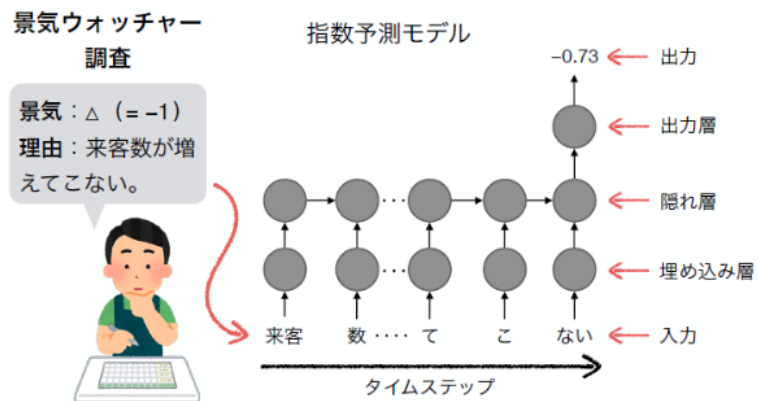
ファインチューニングを工程に含めた予測は、三段階で構成される。まず、手法Bと同様に、景気関連文から初期モデルで指数を予測する。次に、予測値の絶対値が大きい文は、景気動向をよく示していると考えられるので、そのような文だけを訓練データに用いてモデルを再学習する。最後に、再学習したモデルで指数を予測する。なお、予測に用いるニュース記事は、事前の再学習で用いられたニュース記事とは異なるものである。上記の三段階で構成された方法を手法Cと称することにし、付図2-5で表されている。

この付録での説明は、景気代理指数を予測するための学習モデルとして、まずは通常版のRNNに注目した。そして、RNNに含まれる情報処理ユニットをLSTMやGRUなどに置き換えることで、出力結果を評価した。そうした通常版のRNNでは、テキストを文頭から文末にかけて一方向から入力して予測する。本研究では、RNNの応用版であり、テキストを双方向から入力して予測するBidirectional RNN(以下、Bi-RNN)でも、同様に実験を行った。

RNN

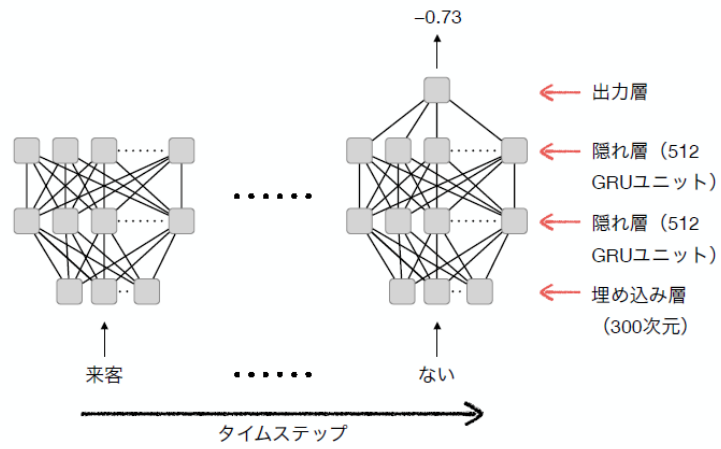


RNN

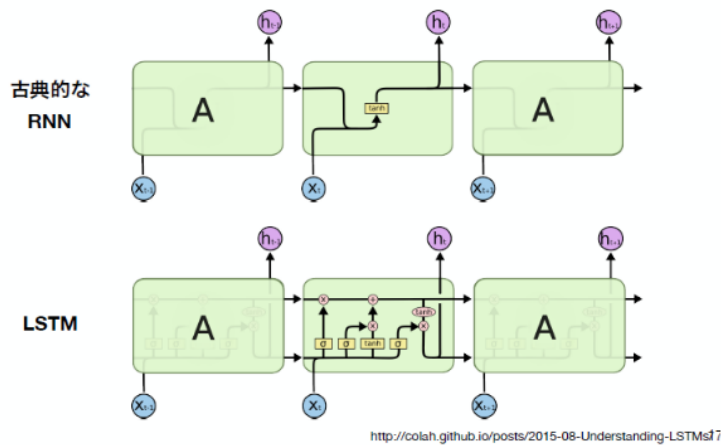


付図 2-1 RNN の基本となるニューラルネットワーク(上)
RNN による学習(下)

モデルの詳細



Gated Recurrent Unit (GRU)



付図 2-2 RNN の詳細(上)
LSTM ユニットを取り入れた RNN (下)

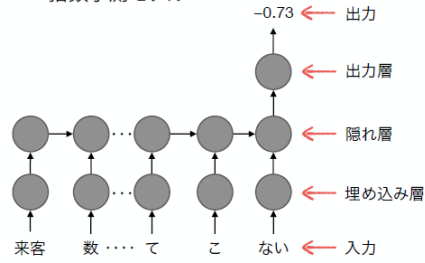
A. 新聞記事で指数を予測

学習

景気ウォッチャー調査

指数予測モデル

景気：△ (= -1)
理由：来客数が増えてこない。



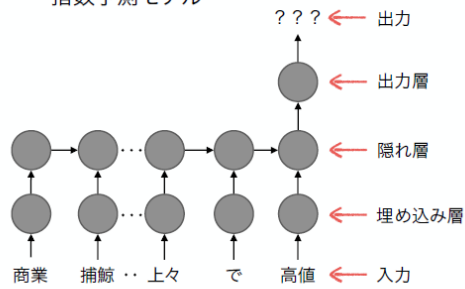
A. 新聞記事で指数を予測

予測

新聞記事

指数予測モデル

商業捕鯨のミンク
肉が入荷=身質上
々で高値



A. 新聞記事で指数を予測

学習

景気ウォッチャー
調査

RNN



予測

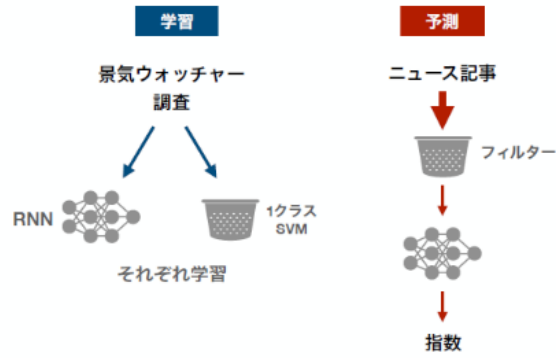
ニュース記事

指数



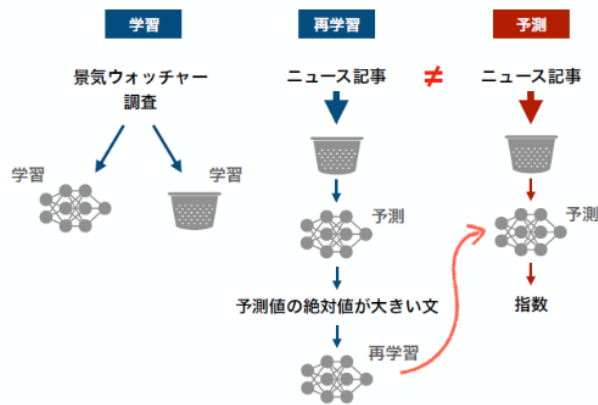
付図 2-3 景気ウォッチャー調査における学習 (上)
新聞記事による予測 (中)
学習済みモデルに全ての記事を用いた予測 (下)

B. 景気動向に関する記事で予測



付図 2-4 フィルタリングにより景気動向に関する記事を用いた予測

C. モデルを再学習して予測



付図 2-5 再学習されたモデルによる予測

第3章 S-APIR 指数と他の景気指数との比較

生田 祐介

1. 概要

本章では、S-APIR と他の景気指標との関係について説明する。第 2 章で説明したとおり、S-APIR は、内閣府の『景気ウォッチャー調査』（景気判断とその理由文のペアで構成されるアンケート回答）を学習データに使い、ファイン・チューニングされた BERT モデルに基づいて算出されたものである。その結果、S-APIR は、景気ウォッチャー調査ではなく新聞記事から算出されたものであるが、景気ウォッチャー調査に基づく景気ウォッチャーDI と強い相関があることがわかった。景気ウォッチャーDI との強い相関関係それ自体は有益であるが、景気ウォッチャーDI はアンケート調査に基づいた数ある景気指標の 1 つに過ぎない。したがって、他の代表的な景気指標と比較することで、一般的な景気指標としての S-APIR の特徴を調べるのが重要である。以下では、2008 年 1 月～2020 年 6 月において、月次の S-APIR と他の景気指数を比較する。

2. マクロ経済指標との比較

ここでは、S-APIR を二つのマクロ経済指標と比較する。第一に GDP と比較し、第二に複数の経済変数の共通因子を景気と見なした人工的指標と比較する。

2.1 GDP

まず、図 3-1 で S-APIR をマクロ経済指標の代表である GDP と比較した。ただし、月次の S-APIR と比較するために、観察できる四半期 GDP から仮想的に作り出された月次 GDP を利用している¹¹。これら二つの変数の動きを細かく見ていくと、大きなショックがあったときに S-APIR は GDP よりも若干先行している。例えば、2008 年のリーマンショック後の世界金融危機では、GDP は年末頃から減少し始め、2009 年初めに底を打った。一方、S-APIR は、すでに 2008 年 9 月頃から急激に減少し始め、2009 年初めに底を打っている。ほぼ半年にわたって、S-APIR は景気動向を先取りしていたことになる。また、S-APIR は、2020 年の COVID-19 パンデミックによる景気後退を若干先取りしている。

¹¹ 月次 GDP は、日本経済研究センター (JCER) によって、四半期 GDP と景気を反映する月次データ (例 鉱工業生産指数) を使い、混合頻度アプローチ (ブリッジ方程式) によって推定されたものである。

2.2. DFM による人工的指標との比較

比較対象となる第二のマクロ経済指標は、ダイナミック・ファクター・モデル（以下、DFM）によって人為的に作り出された数値である¹²。DFMとは、いくつかの観測できる変数から、観測できないが共通する要因を、状態空間モデルを用いて推定する方法である。図 3-2 は DFM における観測変数と非観測変数との関係を示したものであり、鉱工業生産指数 (IIP) のような観測可能な個別経済データ $y_{i,t}$ は、観測できない共通のマクロ経済指標 $x_{i,t}$ から生成されると仮定している。なお、 $i \in \{1, \dots, N\}$ は個別経済指標、 $t \in \{1, \dots, T\}$ は時系列指標を、それぞれ表す。

変数 $y_{i,t}$ と x_t の関係は、以下の(3-1)式、(3-2)式、そして(3-3)式から構成された DFM で定式化される。

$$y_{i,t} = \beta_{i,0} + \gamma_i x_t + u_{i,t} \quad (3-1)$$

$$x_t = \phi_1 x_{t-1} + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \eta_{i,t} \quad (3-2)$$

$$u_{i,t} = d_{i,1} u_{i,t-1} + d_{i,2} u_{i,t-2} + \dots + d_{i,q} u_{i,t-q} + \epsilon_{i,t} \quad (3-3)$$

ただし、 $u_{i,t}$ は x_i と相関しない固有のショックであり、 $\eta_{i,t}$ と $\epsilon_{i,t}$ は誤差項である。

式 (3-1) は測定方程式 (Measurement Equation) であり、個々の経済変数 $y_{i,t}$ は $u_{i,t}$ に依存する。式(3-2) は遷移方程式 (Transition Equation) であり、 x_t は p 次の自己回帰過程に従っている。固有のショック $u_{i,t}$ は式 (3-3) のように q 次の自己回帰過程に従う。変数 x_t は観測可能ではないので、各方程式は状態空間表現に変換される。すると、変数 $y_{i,t}$ と x_t の関係を決定するパラメータは、カルマン・フィルタによって推定することができる。

本研究では、 $y_{i,t}$ に当てはめるデータを生産、消費、雇用、輸出に関する 4 種類の統計から利用した(つまり、 $N = 4$ である)。これらの 4 つの観測可能な経済変数は、共通の観測不能な経済指標 x_t を推定するための基本系列であると仮定している。なお、推定された x_t のことを、S-APIR と比較する際に分かりやすさを優先するため、以降では単に DFM とよぶ。

図 3-3 に示すように、S-APIR と DFM はほぼ同じ動きをしている ($r = 0.749$)。また、GDP の場合と同様に、S-APIR では 主要なショックに対して、DFM よりも若干早い動きとなっている。

¹² DFM は、Stock and Watson [11, 12]で開発された景気推定手法である。

以上のことからわかるように、S-APIR 指数は、全体の傾向としては GDP や DFM と似ており、マクロ経済全体を捉えることができるという特徴があることがわかる。また、S-APIR はマクロ経済指標にある程度先行しており、経済の見通しや経済主体（消費者や生産者など）のセンチメントに関する情報がある程度包含しているという別の特性も示唆される。

3. セミマクロ経済指標との比較

経済センチメントに関する特徴をより詳細に分析するために、いくつかのセミマクロ指標と S-APIR の関係を検証する。センチメント指標としては、消費者と売り手の側面をそれぞれ表す消費者態度指数 (CCI) と購買担当者指数 (PMI) に注目する¹³。また、実態指標としては、需要側と供給側の両方の統計から算出される消費総合指数 (SCI) と、供給側の統計のみから算出される消費活動指数 (CAI) を取り上げた。

3.1. センチメント指標との比較

図 3-4 は CCI と S-APIR を比較したものであるが、両者の動きは似ており、強い正の相関（相関係数=0.848）となっている。実際、2008 年の金融危機や 2020 年の COVID-19 の大流行といった大きなショックが発生した際の低下と回復のタイミングは概ね一致しており、S-APIR が将来に対する消費者心理を反映していることが伺える。

次に、販売面からの見通し指標として、製造業 PMI (以降, MPMI と表記) とサービス業 PMI (以降, SPMI と表記) に注目する。PMI は、企業の購買担当者から毎月アンケートを取り、その結果に基づいて算出されたものである。調査内容は、生産、新規受注、在庫水準、雇用状況、価格などである。通常、企業の購買担当者は、需要に応じて商品を購入し、販売に備える。そのため、PMI を通じて、消費に先行する販売の立場から経済活動を見ることになる。したがって、PMI は、市場の動きに敏感な購買担当者による景気の先行指標と言える。MPMI は製造業を対象としているため、企業間取引 (B2B) に関連する情報が含まれている。一方、SPMI はサービス業を対象としており、企業対消費者 (B2C) 取引に関する情報を含んでいる。

図 3-5 は、PMI と S-APIR を比較したものである。2008 年の金融危機の際、S-APIR は SPMI よりも MPMI に近い動きをしていることが分かる。一方、2020 年の COVID-19 パンデミック時には、S-APIR は MPMI よりも SPMI に近い動きをしている。金融危機時に MPMI が急激に低下したのは、自動車需要の大幅な減

¹³ 消費者態度指数は、消費者の今後 6 ヶ月間の消費動向の見通しを表している。「雇用環境」「収入の増え方」「暮らし向き」「耐久消費財の買い時判断」を、消費者がそれぞれ 5 段階に評価する調査結果から、指数を算出している。

少が一因である。COVID-19パンデミック時にSPMIが低下したのは、食料、小売、宿泊の需要が大きく低下したためである。当該の期間において、S-APIRは、B2BとB2Cという異なるタイプの売り手の見通しを正確に捉えている。このことから、S-APIRの情報源である日経新聞は、全ての期間に当てはまらないものの、製造業とサービス業のうち大きな負の影響を受けた業種について、より頻繁に報道していることが示唆される。

3.2. 実態指数との比較

以上のように、S-APIRとセンチメント指標(CCIやPMI)との比較から、S-APIRは消費活動についての予測を上手く捉えていることがわかった。しかし、この結果は、S-APIRが実際の消費活動に先行しているかどうかを示すものではない。そこで、S-APIRと実際の消費活動に関する指標(SCI、CAI)を比較することにより、S-APIRの消費活動の先行指標としての適否を検討したい。図3-6では、2008年の金融危機、2014年と2019年の増税、COVID-19の大流行など、特に大きな経済イベントにおいて、S-APIRが実際の消費に先行していることが分かる。例外は東日本大震災であり、S-APIRはSCIやCAIと同調して動いているように見える。

4. 結論：S-APIRの特徴

以上の一連の比較から、S-APIR指数の大きな特徴を2つ確認した。第一に、S-APIRはマクロ経済の全体的な動向を捉えている。第二に、S-APIRは消費と販売の見通しに関する情報を多く含んでいる。これは、モデルの学習データである「景気ウォッチャー調査」に、家計消費と企業の販売活動の見通しを示す様々な文章が含まれていることに起因していると思われる。

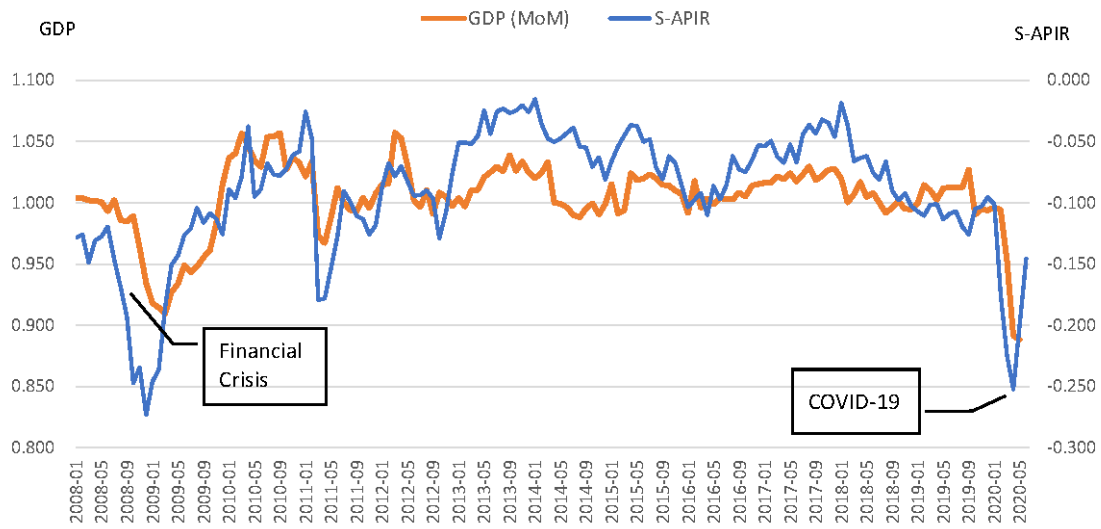


図 3-1 S-APIR と月次 GDP との比較 (相関係数 = 0.698)

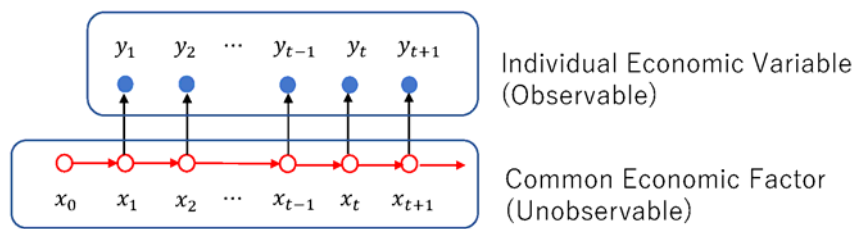


図 3-2 ダイナミック・ファクター・モデル (DFM) の概念図

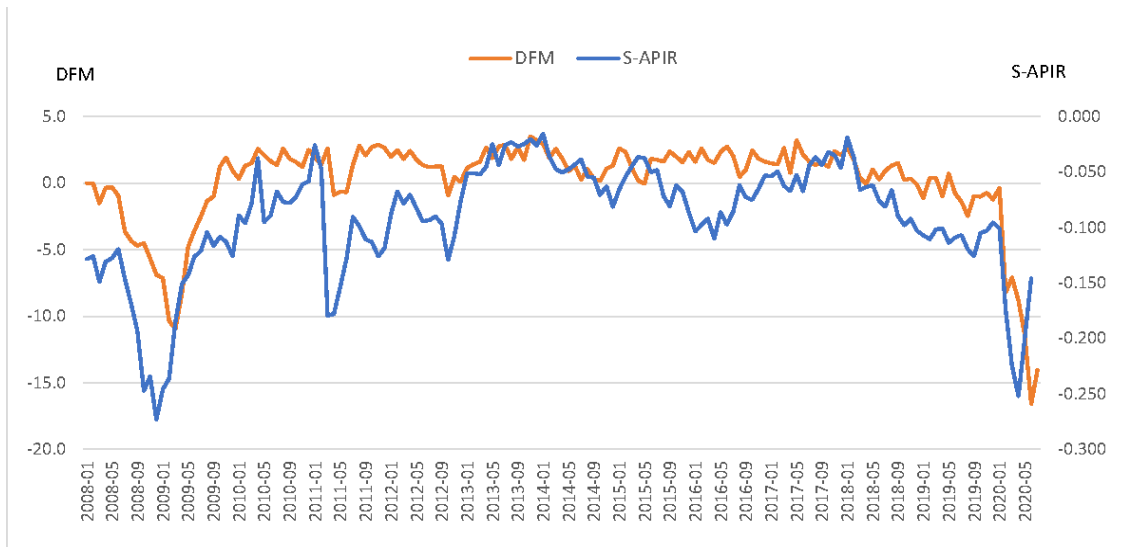


図 3-3 S-APIR と DFM との比較 (相関係数 = 0.749)

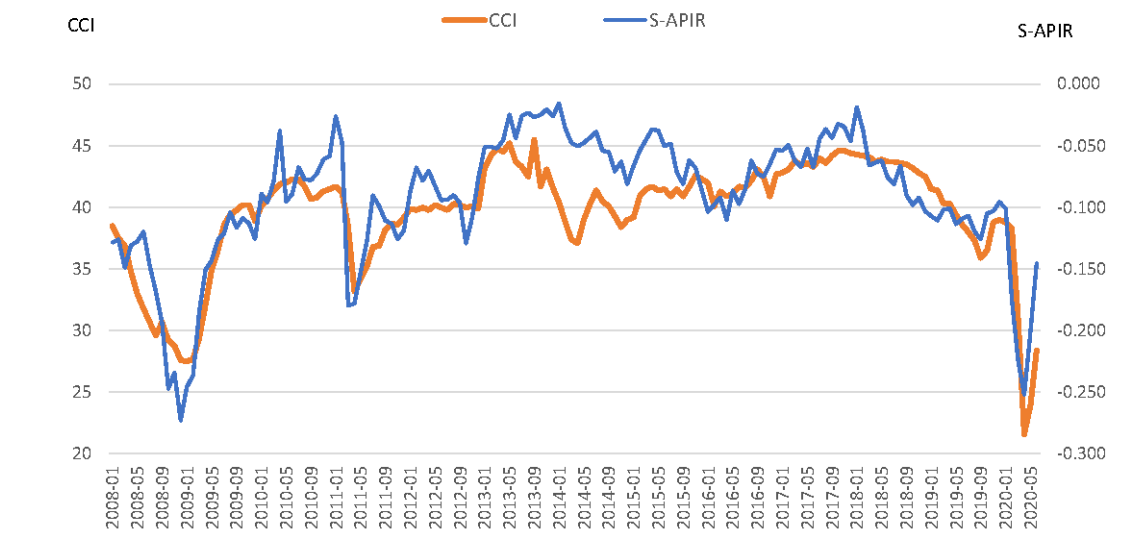


図 3-4 S-APIR と CCI との比較 (相関係数 = 0.848)

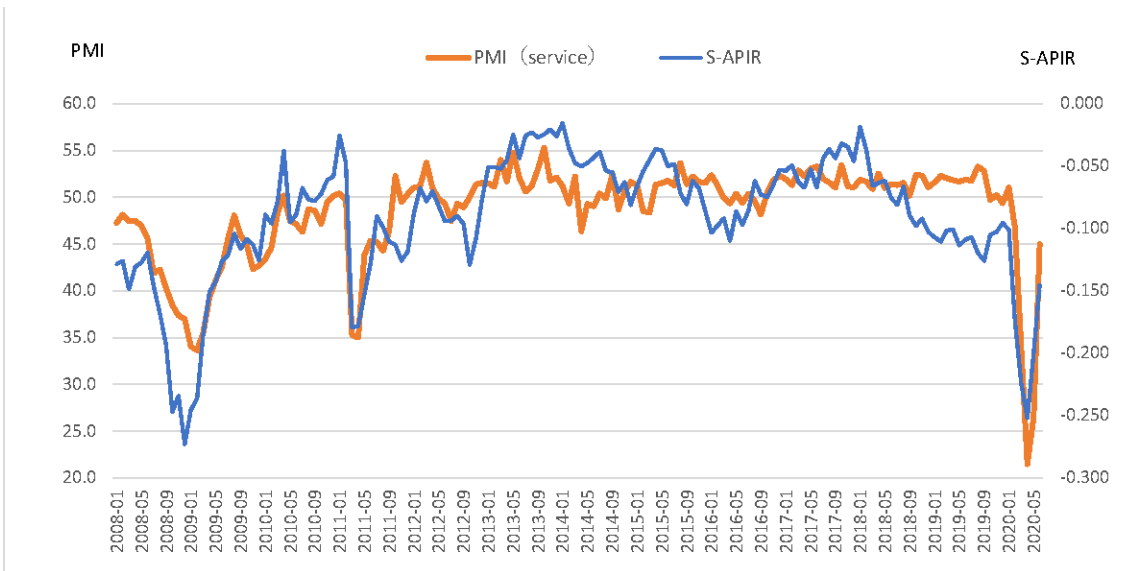
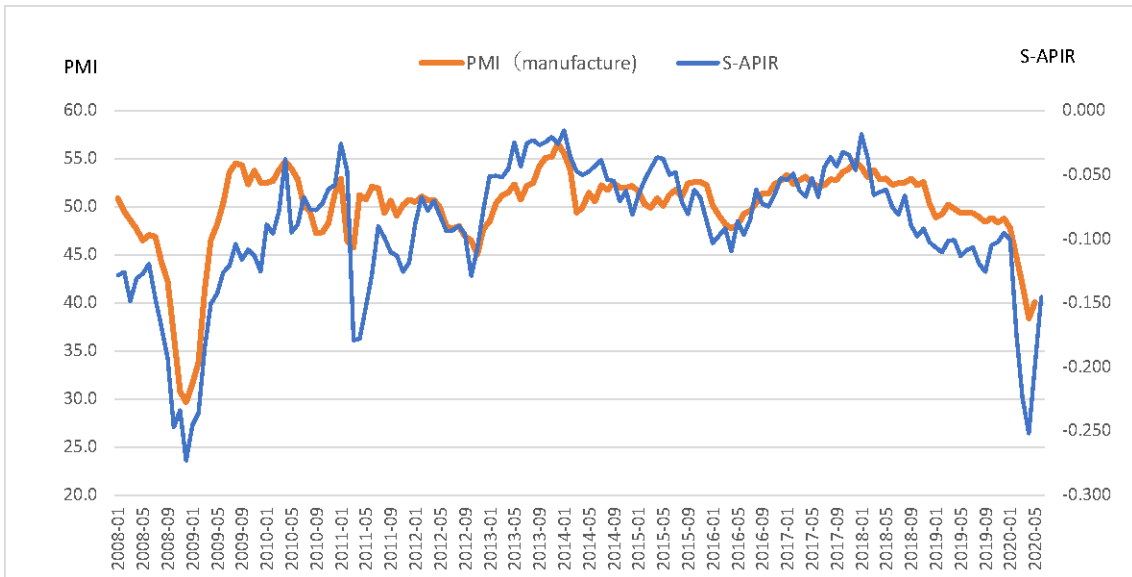


図 3-5 S-APIR と MPMI の比較 (相関係数 = 0.773) (上)
S-APIR と SPMI の比較 (相関係数 = 0.802) (下)

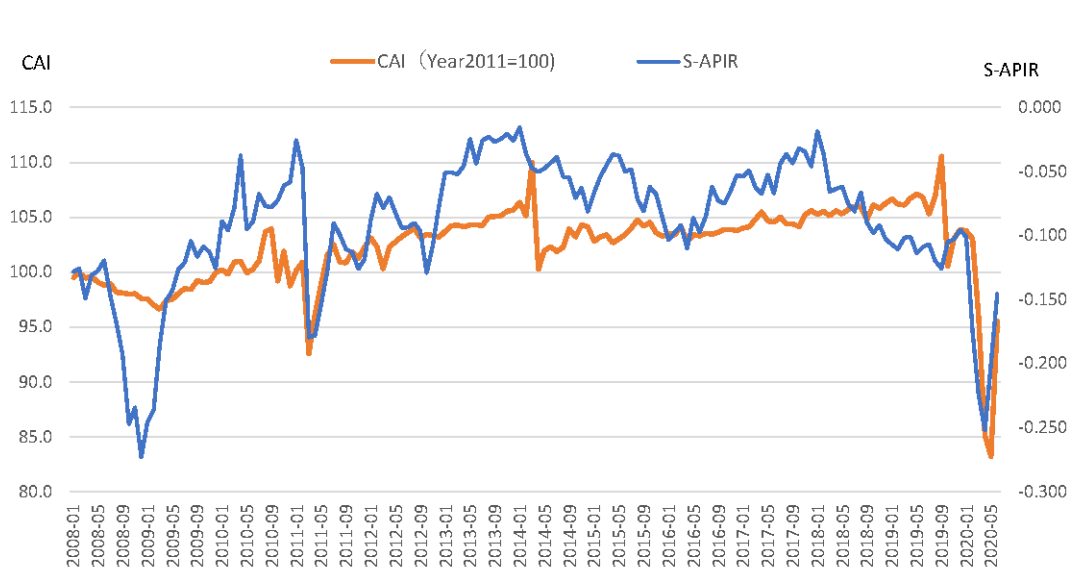
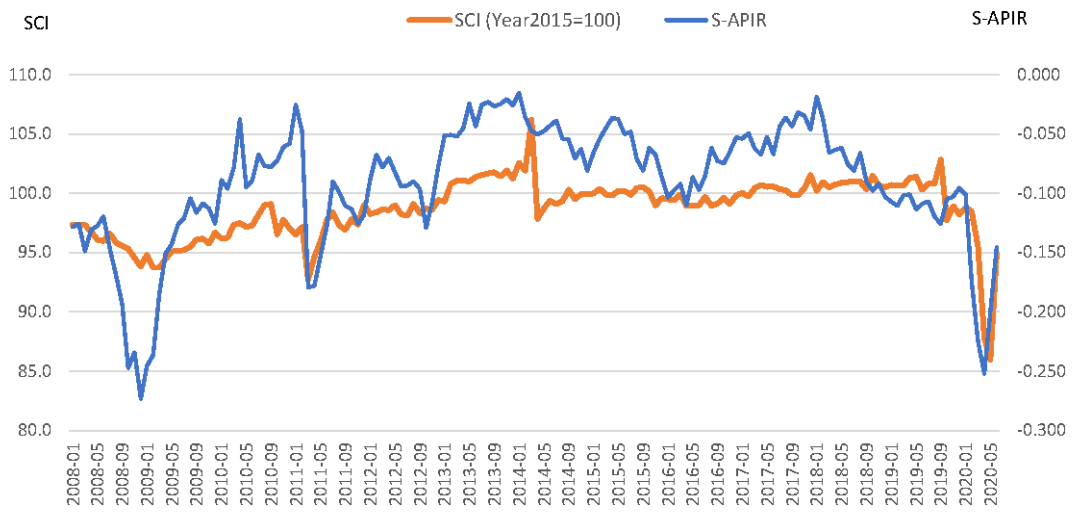


図 3-6 S-APIRとSCIの比較(相関係数 = 0.727) (上)
S-APIRとCAIの比較(相関係数 = 0.650) (下)

第4章 S-APIR の語彙による要因分解¹⁴

生田 祐介, 関 和広

1. 概要

景況感は、金融政策や物価、為替、雇用、賃金、海外情勢など、様々な要因から形成されている。しかし、すべての要因が経済に等しく影響を与えているわけではないことから、経済を上向きあるいは下向きに動かす要因が何であるのかを知ることが、経済の情勢判断を行う上で重要である。本章では、第2章で推定した S-APIR 指数に対する語の寄与を考えることで、どのような要因（語句）が、いつ、どの程度、景況感に影響を与えたのかを分析する。

従来 of 定量分析では、ある事象や要因が経済動向にどの程度影響を与えるかを分析する場合、その要因の時系列（例えば、原油価格）と経済動向を表す指標（例えば、GDP）を回帰分析で検証している。これに対して、我々は景況感の測定に使用した入力ニュース文章に着目し、「原油価格」という言葉が登場する文脈を、ニュース文章の予測景況感スコアによって利用する。つまり、事後的に明らかになる定量データだけでは捉えきれない経済主体の行動パターンや将来予測から、景気変動の要因を考察しようとするものである。

2. 指数の要因分析

S-APIR 指数は文のスコアを統合したものであることを踏まえ、S-APIR 指数を単語へ要因分解するために、以下のような簡便な方法を採用する。まず、文 s の景気スコア p_s は、文 s を構成する語（形態素） $w \in S$ の景気スコア $p_{s,w}$ の総和であると仮定する。

$$p_s = \sum_{w \in S} tf_{s,w} \cdot p_{s,w} \quad (4-1)$$

なお、 $tf_{s,w}$ は文 s における語 w の出現頻度である。さらに、すべての単語 $w \in S$ が等しく文 s の景気スコアに貢献すると仮定すると、 $|S|$ を文 s を構成する単語数として $p_{s,w}$ は次式で表せる。

¹⁴ 本章は、関・生田[7]と Seki et al.[10]の内容を、加筆修正したものである。

$$p_{s,w} = \frac{p_s}{|S|} \quad (4-2)$$

ここで、期間 t 内の文の集合を S_t とする。そして、単語 w についてすべての文 $s \in S_t$ で $p_{s,w}$ の総和を取り、総文数 $|S_t|$ で除することで、期間 t における語 w (のみ) の景況感指数 $p_{t,w}$ を次式で定義する。

$$p_{t,w} = \frac{1}{|S_t|} \sum_{s \in S_t} t f_{s,w} \cdot \frac{p_s}{|S|} \quad (4-3)$$

なお、 w が複合語である場合は、それを構成する語の語数を式 (2) に乗ずる。直感的には、期間 t の記事に現れたすべての語 w の景況感指数 $p_{t,w}$ の総和が、期間 t における S-APIR 指数であるというように解釈できる。

3. 結果：景況感へのイベントの寄与度

以下では、近年の日本経済で注目されている 4 つの事象に着目し、S-APIR 指数との関係を検証してみることにする。まず、「インバウンド」(おおまかに言えば、「外国人観光客」) について見てみよう。インバウンドという言葉は、英語の“inbound”の借用語であり、2010 年代以降に急増した日本への外国人旅行者を指す。図 4-1 からわかるように、この言葉は 2015 年頃から経済に好影響を与え始め、2017 年末までは順調に上昇していた。2010 年代半ば、日本では外国人観光客による“爆買い”という現象が見られ、この現象は 3 年間続いたこと確認できる。

前述したように、我々のアプローチでは、様々な事象がいつ、どの程度、経済心理に影響を与えるかをタイムリーに定量化することが可能である。COVID-19 の流行が始まってから、訪日外国人の数が途絶え、図 4-1 に示すように、経済に大きな影響を与えるようになった。しかし同時に、2020 年初頭が最も影響が大きく、それ以降はある程度状況が改善されていることも確認できる。このことはインバウンド需要の減少による経済への影響がそれほど持続しない可能性を示唆しており、政策立案者が経済情勢を把握する上で非常に有用である。すなわち、外国人観光客の減少による経済の減速が比較的一過性のものであれば、政策立案者は観光産業に対する大規模な支援策を実施する必要はないのかもしれない。その代わりに、限られた資源を、緊急の支援を必要とする他の産業や個人に割り当てることができるだろう。

次に、「増税」との関係を見てみよう。2010 年代以降、日本では 2014 年 4 月に

5%から8%へ、2019年10月に8%から10%へ、二度の消費税増税が実施された。消費税増税の実施は事前に発表されていたため、増税直前の駆け込み需要とその後の落ち込みという形で経済に影響を与えた。図4-2は、2014年の最初の増税の影響を明確に示している。しかし、二度目の2019年の増税では、駆け込み需要の影響はほとんど見られず、消費の反動減は微妙なものとなっている。消費者のセンチメントは弱まり始めており、増税のショックにほとんど反応していないことを示唆している。また、増税が景気に与える影響は、訪日外国人観光客のように従来の定量分析で容易に確認できるものではなく、我々のアプローチの有用性を立証している。

三つ目に注目した単語は、「コロナ(COVID-19)」である。図4-3では、2020年初頭からコロナウィルスの寄与がマイナス側に大きく現れており、景況感を押し下げる大きな要因になっていることが分かる。なお、「コロナ」という語は景気ウォッチャー調査には2020年1月まで1度も出現していない。今回は2020年2月までのデータを学習データとしたが、学習データに含まれない語(未知語)に対して本手法が頑健か確認するため、試みとして2019年12月までのデータを用いて景気判断モデルを学習し、同じ実験を行った。その結果、図4-3とほぼ同じ結果が得られた(結果は割愛する)。本手法は文の景気スコアを単語に均等に分配するという単純な方法であるため、文全体として景気判断ができれば未知語であっても適切にその貢献を定量化できることが確認できる。

最後に、2020年に開催が予定されていた一大イベントである「東京五輪」を見てみよう。図4-4からわかるように、東京での開催が決まった2013年9月から約7年間、オリンピックは概ね経済にプラスの効果を与えていた。しかし、2020年に向けては、COVID-19のパンデミックにより開催が懸念されるようになり、経済への影響は一転して悪化している。もし、パンデミックが発生しなければ、オリンピックが日本経済に与えるプラス効果は一定して安定していたのではないかと推測する。東京五輪の経済への影響は、開催国決定から開催日まで8年近くに渡って発生していると考えられるため、従来の定量分析では、特定期間の経済心理への影響を評価することは困難であった。これに対し、我々のアプローチでは、新聞で取り上げられる記事である限り、任意の時期の任意のイベントを対象とした時間的な分析が可能である。

4. 考察: イベントの寄与度の頑健性

このように、ある事象が企業の景況感に与える影響を分析することが可能であることを示した。一方、文を構成する単語がその文の景気スコアに等しく貢献するという仮定は過度な単純化とも考えられ、その結果の妥当性の検証が必要である。本研究で用いた BERT は自己注意機構というシステムに基づくため、景気判断予測の際、入力系列のどの要素に注意が払われていたのかをアテンション値 (attention weight) で内部的に定量化している。しかし、その値は後ろの層に近づくほど均等になる傾向があり、必ずしも入力文の単語そのものの重要性を示すわけではない。そこで、単語の重要性をより適切に表現する指標として、層間のアテンションを有向グラフと見なして最大フローを求める attention flow やアテンション値の積和で定義される attention rollout が提案されている (Abnar and Zuidema, [9])。ここでは、より計算量が少ない attention rollout を単語の重要度として利用し、前述の仮定の妥当性を検証した。具体的には、BERT の最終層の [CLS] トークン (文全体を表す特殊なシンボル) から各単語 w への attention rollout 値 r_w の大きさに応じて文 s の景気スコア p_s を各単語に分配した。すなわち、式(4-2)を次式で置き換えた。

$$p_{s,w} = p_s \cdot \frac{r_w}{\sum_{w' \in S} r_{w'}} \quad (4-4)$$

この定義に基づいて「東京五輪」の貢献度を算出し直した結果を図 4-5 に示す。図 4-5 の上下のグラフを比較すると、(a) attention rollout 値を利用した下のグラフでは全体的に値が大きくなっていること、(b) わずかな違いはあるものの傾向としてはほとんど変わらないことが分かる。実際に、両者の相関係数は 0.995 であり、非常に高い正の相関があった (他の三つの単語についても同様であった)。以上の結果から、アテンション値に基づいて語の重要性を考慮した場合も (絶対的な値は変化しても) 全体の傾向はほとんど変化しないため、語の貢献の傾向を分析するうえでは式(4-2)の仮定は問題にはならないといえる。

5. 結論

本章では、文単位で予測した景気スコアを単語単位に分割して再集計することで、任意の要因が景況感に与える影響を時間軸に沿って分析できることをいくつかの例

とともに示した。また、景気スコアが均等に分布しているか、attention rollout 値に比例しているかに関わらず、結果は頑健であることが示された。

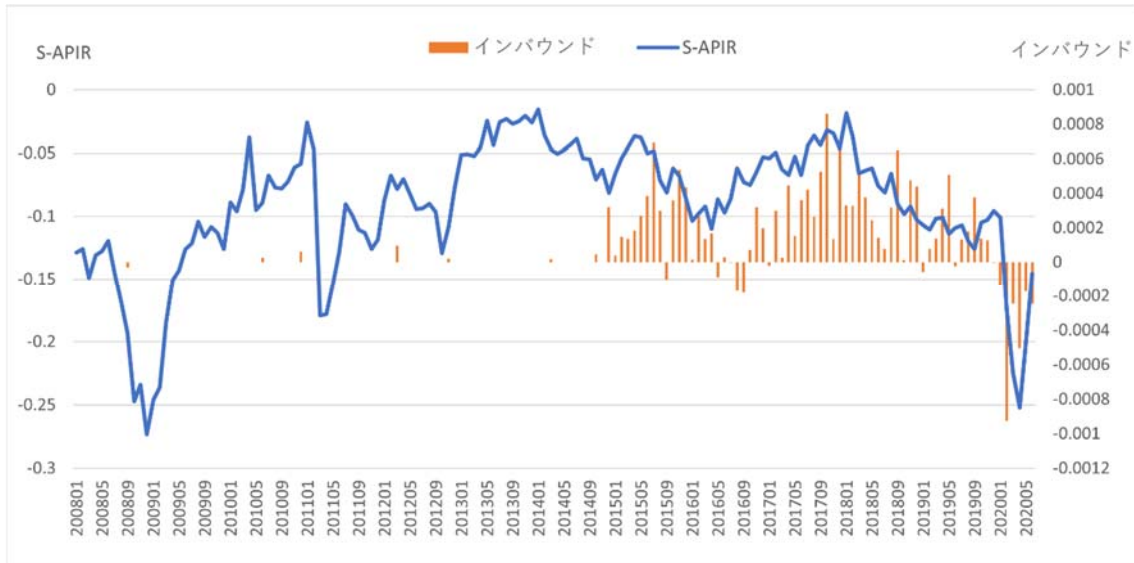


図 4-1 S-APIR 指数への「インバウンド」の寄与

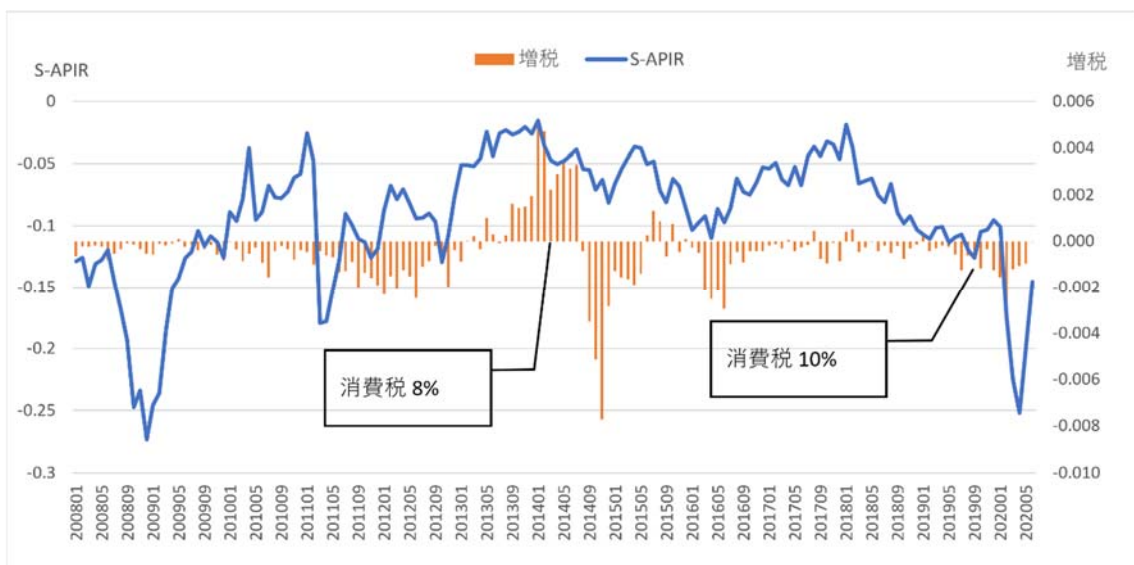


図 4-2 S-APIR 指数への「増税」の寄与

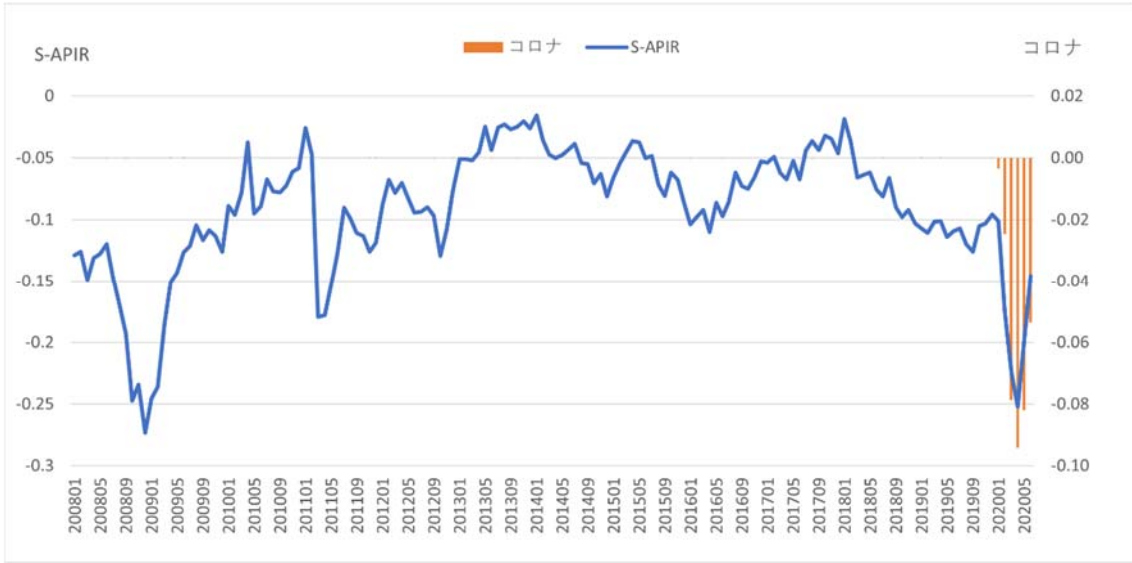


図 4-3 S-APIR 指数への「コロナ」の寄与

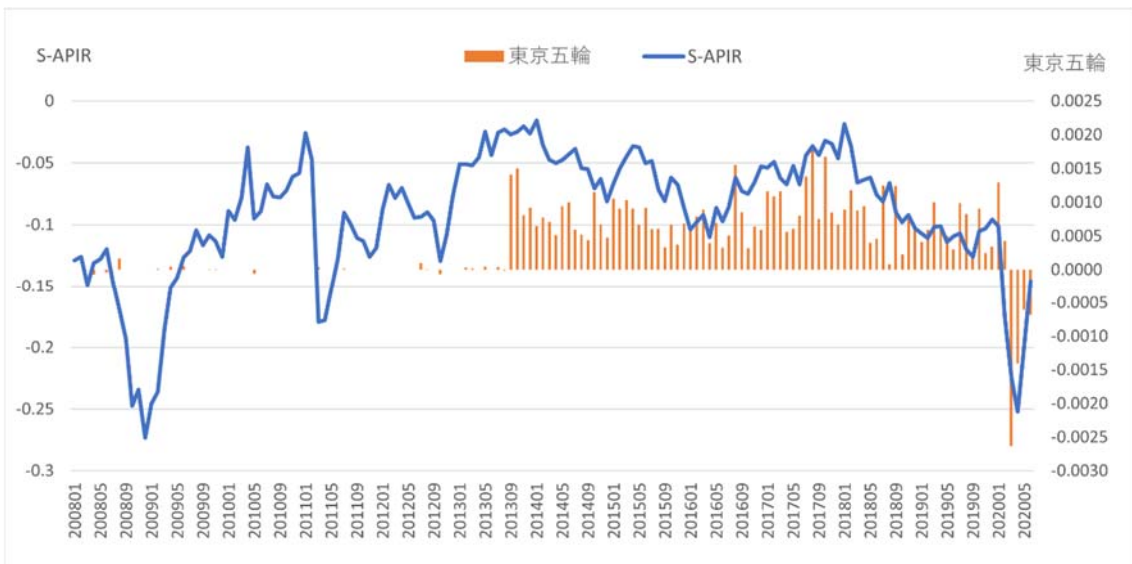


図 4-4 S-APIR 指数への「東京五輪」の寄与

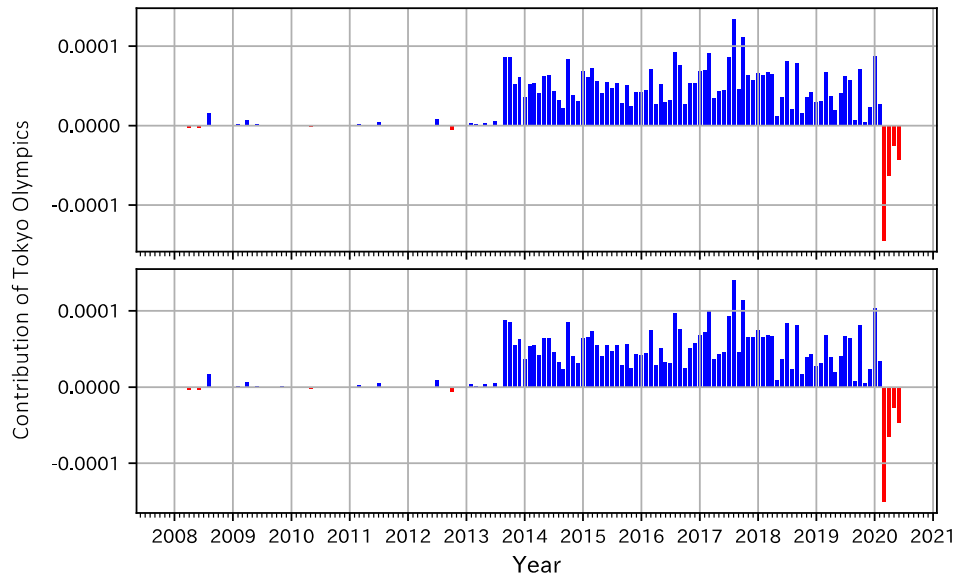


図 4-5 S-APIR 指数への「東京五輪」の寄与 (上は景気スコアを均等に分配 (図 4-4 と同様), 下は景気スコアを attention rollout 値に応じて分配)

第5章 景況感足元予測システム

関 和広

1. 概要

S-APIR 指数は、学習済み景況感予測モデルに大量の新聞記事データを入力することにより算出される。第 2 章では、このモデルに 2008 年 1 月から 2020 年 6 月までに発行された日経新聞の記事を与え、その結果得られた S-APIR 指数を同期間の過去の景気ウォッチャーDIと比較することで、S-APIR 指数の妥当性を示した。しかしながら、速報性の高い景気判断、すなわち景況感の足元予測を可能にするという S-APIR 指数開発のそもそもの趣旨に立ち返れば、過去のデータではなく、今まさに報道されている記事を用いて現在の景況感を予測する枠組みを構築する必要がある。

本章では、新聞紙面に掲載される記事ではなく、ウェブニュースサイトで配信・公開される記事データを自動的に収集し、S-APIR 指数の算出を行う「景況感足元予測システム」の構成について述べ、実例を交えてシステムの使用方法を説明する。

2. システムの構成

景況感足元予測システムの構成の概要を図 5-1 に示す。

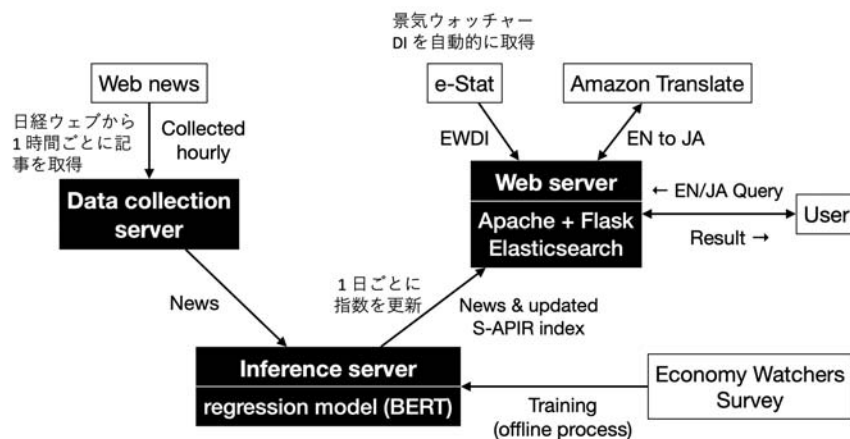


図 5-1. 景況感足元予測システムの構成

システムは 3 つのサーバ、すなわちデータ収集サーバ (data collection server)、推論サーバ (inference server)、ウェブサーバ (web server) からなる。

データ収集サーバは 1 時間ごとに対象のニュースサイトにアクセスし、新しい記事の HTML を取得・保存する。取得された HTML ファイルは、目的とする記事以外の雑多な情報を含んでいるため、スクレーピングによって記事のタイトルと記事本文（一部）のみを抽出する。

抽出された記事データは、毎日一回定刻に推論サーバに送信され、文分割などの前処理やフィルタリングが行われたのち、各文について景気スコアが予測される。予測に用いるモデルは、これまでと同様、あらかじめ景気ウォッチャー調査データでファインチューニングされた BERT である。予測した景気スコアは月次で集計され、ウェブサーバに送信される。

ウェブサーバは、S-APIR 指数とともに景気ウォッチャーDI をユーザに提供する。なお、Web サーバでは、月次の景況感指数を表示するだけではなく、特定のキーワードが景況感にどの程度影響したのか、特定のキーワードに着目した分析を行うことができる（3 章参照）。キーワードはユーザが検索クエリとして自由に入力し、そのクエリが検索エンジン Elasticsearch で検索される。

Elasticsearch は、収集されたウェブ記事を文単位でインデックスしており、入力されたクエリが含まれる文とその景気スコアを即座に検索・取得し、キーワード単位に分解して再集計することで、そのキーワードの景況感指数への寄与を時系列で出力する。なお、一度検索したキーワードに関する要因分解の結果は、次回以降の検索時に速やかに応答できるよう関係データベース SQLAlchemy に保存される。

また、入力されたキーワードはウォッチリストとしてブラウザのキャッシュに保存される。この機構によって、同一ユーザが同一ブラウザで本ウェブシステムに再訪問したときには、そのキーワードに関する結果が自動的に再度表示されるため、ユーザが興味を持っている特定の要因による影響を継続的にモニタできる。

なお、本システムは日本語のニュースサイトのみを対象に記事データの収集を行っているため、要因分解は日本語のキーワードを想定している。しかし、国内外で広くデモンストレーションを行えるよう、英語にも対応している。英語のキーワード（アルファベットのみからなるクエリ）が入力された場合は、それを翻訳ウェブサービス Amazon Translate に送信し、日本語に翻訳したのちに分析を行う。

3. システムの機能と使用例

景況感足元予測システムのスクリーンショットをエラー！参照元が見つかりません。に示す。ランディングページには、検索クエリを入力するためのテキストボックスと景況感を示すグラフのみが表示されている。表示されている景況感は、月次の S-APIR 指数と景気ウォッチャーDI である。なお、当月の S-APIR 指数は、月の始めから前日までに取得された記事を基に算出された暫定値であり、毎日最新の記事を反映し

て更新される。景気ウォッチャーDI は政府のポータルサイト e-Stat から自動的に取得している。なお、ある月の景気ウォッチャーDI が内閣府のホームページで公開されるのは翌月の 8 日頃（正確には月の始めから 6 営業日）、この統計値が e-Stat で公開されるのは 25 日ごろであり、前者でも月末から 1 週間以上の遅延がある。これに対して、S-APIR 指数は月次の確定値であれば翌月 1 日に、暫定値であれば当月内のいつでも知ることができるため、きわめて速報性が高い。

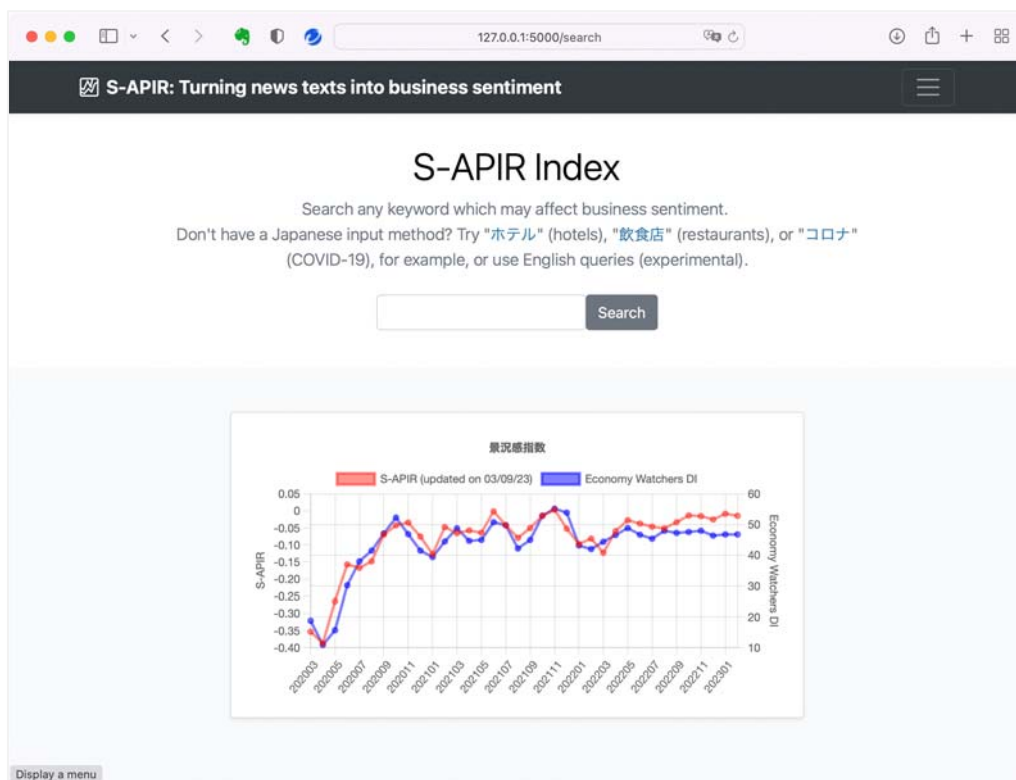


図 5-2. 景況感足元予測システムのランディングページ

次に、特定のキーワードに着目した分析について具体的な例をあげて紹介する。近年の日本経済に大きな影響を与えている要因の一つとして、「中国」がある。そこでエコノミストや政策立案者が、「中国」が景況感に与える影響について時系列で定量的な分析を行いたいという状況を想定する。このような場合、ユーザはキーワード「中国」で検索を行い、図 5-3 のような結果を得る。なお、ここでは 2023 年 2 月までのデータを使用した。このグラフには、ポジティブな影響（青い棒グラフ）とネガティブな影響（赤い棒グラフ）がそれぞれ月次で集計され、さらに両者を合計した総合的な影響が折れ線グラフで示されている。

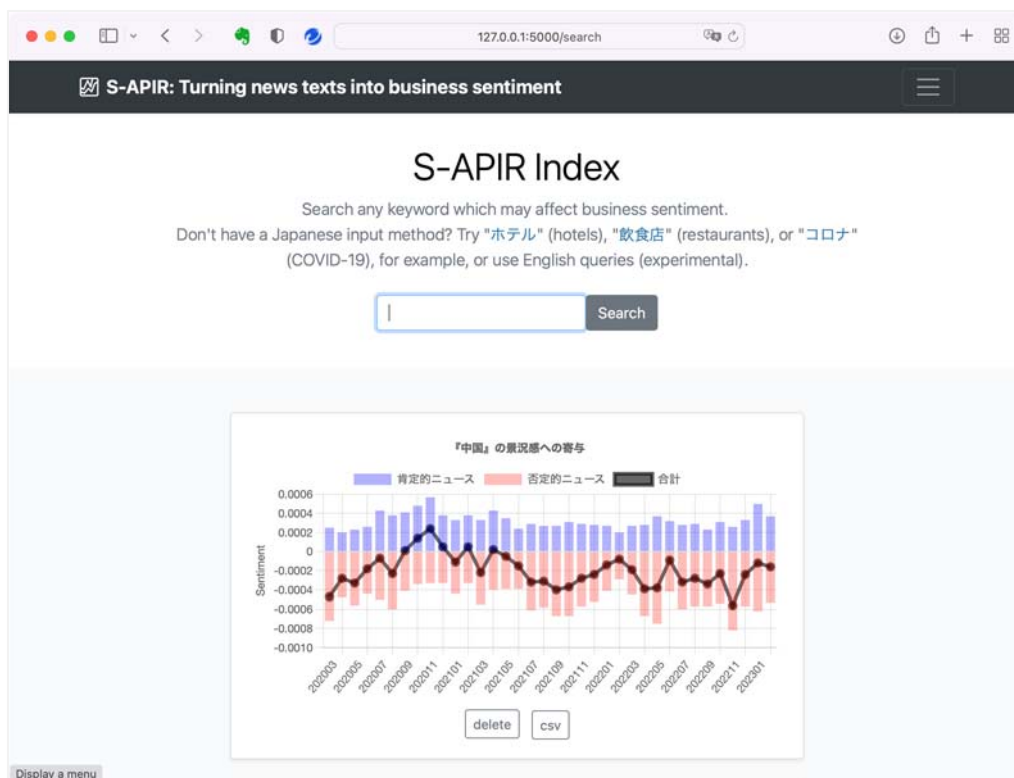


図 5-3. キーワード「中国」で要因分解したときの結果（2023 年 1 月までのデータを使用）

このグラフから、ポジティブな影響も多くあるものの、2020 年 3 月以降、中国は一時期（2020 年秋）を除いて総合的にはほぼ一貫してネガティブな影響を与えていることがうかがわれる。では、中国について実際にどのような報道がなされ、景況感にネガティブな影響を与えているのだろうか。具体的な報道内容について知りたい場合は、マウスカーソルを所望の月の棒グラフに重ねることで、その月に報道された特徴的な文をいくつか閲覧することができる。たとえば、赤い棒グラフが最も長い 2022 年 11 月にマウスカーソルを重ねれば、図 5-4 のように新型コロナウイルスの感染が再拡大していることを報道する記事の一部が表示される（図では著作権の関係で文にモザイク処理をかけてある）。なお、2020 年 11 月に景況感が向上している際の記事を同様に閲覧すると、中国の景気が回復基調にあり、輸出や販売が大幅に増加していることが理由であることが分かる。この機能により、ユーザが興味を持っている任意の要因（例：中国）が景況感に与える影響が特定の時期に変化した背景が知りたいときなど、過去の新聞の縮刷版や記事データベースにあたることなく、即座に当時の状況を垣間見ることができる。

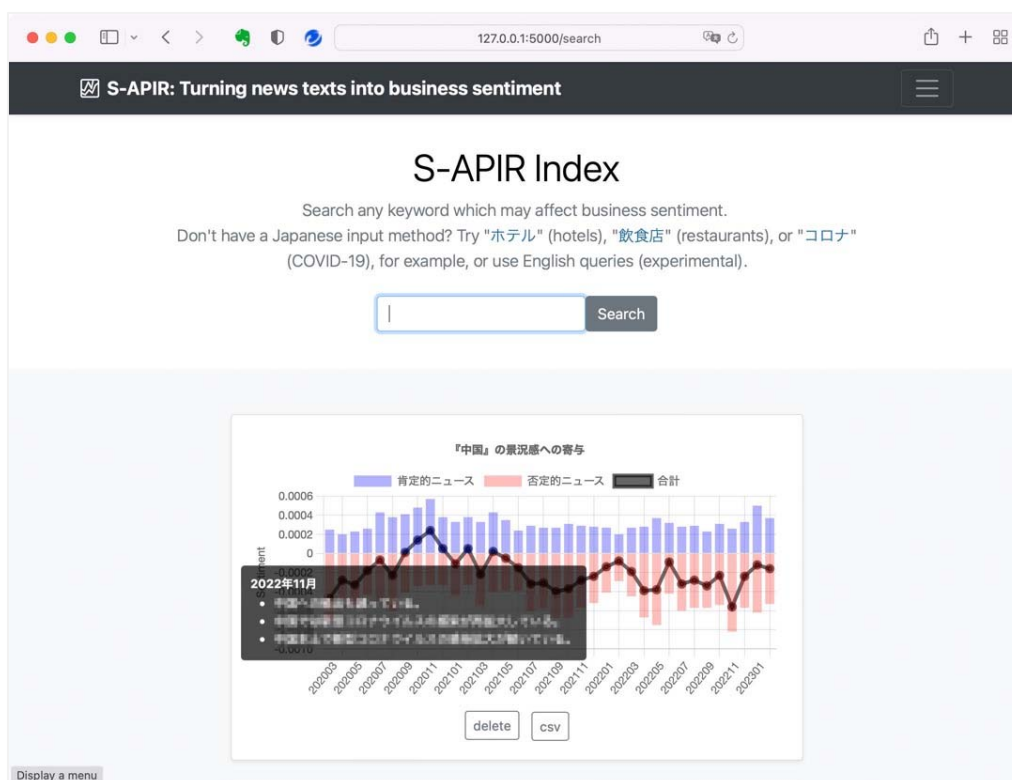


図 5-4. 景況感が下落した 2022 年 11 月の特徴的な文

さらに、複数のキーワードで要因分解を行った場合、各要因の景況感への影響の大きさを視覚的に容易に比較できるようにするため、図 5 のように積み上げ棒グラフが表示される。図 5 は、近年、日本経済に大きなショックを与えた(与えている)「コロナ」、「ウクライナ」、「インフレ」について要因分解を行ったときの例であり、このグラフから例えば次のような観測が可能である。

- いずれも景況感に大きな影響を与えていると考えられる「ウクライナ」、「インフレ」と比較しても、特に 2020 年 3～5 月の「コロナ」の影響ははるかに大きい。
- コロナの影響が 2022 年 3 月あたり(コロナ第 6 波の時期)をピークに徐々に落ち着いてきている。
- ロシアがウクライナに侵攻した 2022 年 2 月から数か月は、「ウクライナ」が「コロナ」以上の影響を持っている。
- 「コロナ」、「ウクライナ」ほどの影響は無いものの、2021 年秋頃から「インフレ」が継続して景況感を下げる一定の要因になっている。

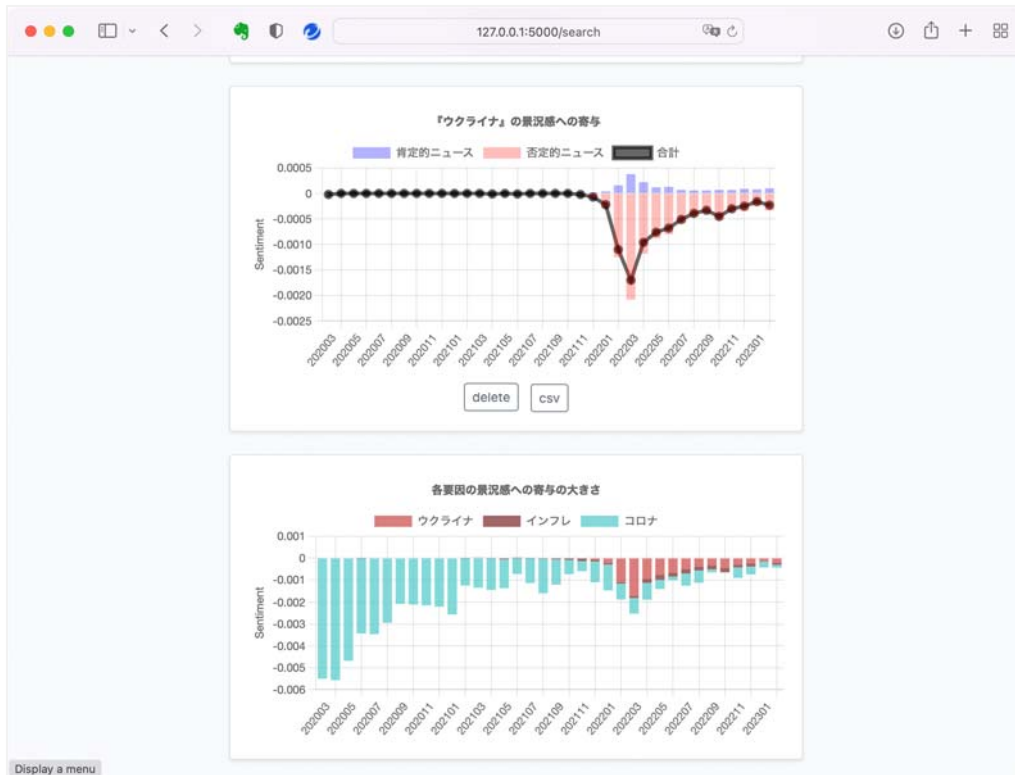


図 5-5. 「ウクライナ」、「インフレ」、「コロナ」の景況感への影響の大きさを比較した例（下のグラフ）

参考文献

- [1] 饗場行洋, 山本裕樹: データサイエンスと新しい金融工学, 財界観測, Vol.81, No.2, pp.30-41 (2018).
- [2] 青嶋智久, 中川慧: 日本語BERT モデルを用いた経済テキストデータのセンチメント分析, 第33 回人工知能学会全国大会論文集, 4Rin127 (2019).
- [3] 生田祐介, 関和広, 松林洋一: テキストデータを利用した新しい景況感指標の開発と応用(下)— 応用編: 深層学習を利用したテキスト分析 —, APIR Discussion Paper Series No. 48 (2021).
- [4] 五島圭一, 高橋大志, 山田哲也: 自然言語処理による景況感ニュース指数の構築とボラティリティ予測への応用, 金融研究, Vol.38, No.3 (2019).
- [5] 近藤浩史, 與五澤守, 成瀬道紀, 森正和: 金融機関のテキストデータを活用した景気センチメントの計測, 人工知能学会全国大会論文集(2019).
- [6] 坂地泰紀, 和泉潔, 松島裕康, 川瀬和哉, 林寛: 接触履歴を用いた地域景況インデックスの自動生成, 知能と情報, Vol.31, No.2, pp.626-635 (2019).
- [7] 関和広, 生田祐介: 経済ニュースによる景況感指数の足元予測, 情報処理学会論文誌, Vol.62, No.5, pp1288-1297 (May 2021).
- [8] 山本裕樹, 松尾豊: 景気ウォッチャー調査を学習データに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論文集, 3L3-OS-16a-2 (2016).
- [9] Abnar, S. and Zuidema, W.: Quantifying attention flow in transformers, Proc. 58th Annual Meeting of the Association for Computational Linguistics, pp.4190-4197 (2020).
- [10] Seki, K., Ikuta, Y & Matsubayashi, Y. (2022). News-based business sentiment and its properties as an economic index. *Information Processing and Management*, 59, Article 102795.
- [11] Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, 4, 351-394.
- [12] Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicators. In G. Moore, & K. Lahiri (Eds.), *The leading economic indicators: new approaches and forecasting records* (pp. 63-90). Cambridge University Press.
- [13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. & Polosukhin, I.: Attention is all you need, Proc. 31st *International Conference on Neural Information Processing Systems*, pp.6000-6010 (2017).

研究会記録

2022年 2月22日 2021年度所内研究会

2022年 10月14日 2022年度第1回所内研究会

2023年 3月16日 2022年度第2回所内研究会

研究成果

(論文)

- ・ 生田祐介, 関和広, 松林洋一. テキストデータを利用した新しい景況感指標の開発と応用 (下) —応用編: 深層学習を利用したテキスト分析—. APIR Discussion Paper Series, No. 48, <https://www.apir.or.jp/research/9261>, 49 pages, 2021.
- ・ 関和広, 生田祐介. 経済ニュースによる景況感指数の足元予測. 情報処理学会論文誌, Vol. 62, No. 5, pp. 1288-1297, 2021.
- ・ K. Seki, Y. Ikuta, and Y. Matsubayashi. News-based Business Sentiment and its Properties as an Economic Index. Information Processing & Management, Vol. 59, No. 2, 102795, 2022.

(国際会議)

- ・ K. Seki. Turning News Texts into Business Sentiment. In Proc. of the 44th European Conference on Information Retrieval, pp. 311-315, 2022.

(口頭発表)

- ・ 関和広, 生田祐介, 松林洋一. テキストデータを利用した新しい景況感指標の開発と応用. 「第13回 AI・ビッグデータ経済モデル研究会」, 日本経済研究センター, 2022年8月5日

「テキストデータを利用した S-APIR 指数の実用化」
研究会報告書

発行日 2023(令和5)年4月
発行所 〒530-0011
大阪市北区大深町3番1号
グランフロント大阪 ナレッジキャピタル
タワーC 7階
一般財団法人 アジア太平洋研究所
Asia Pacific Institute of Research (APIR)
TEL(06)6485-7690(代表)
FAX(06)6485-7689
発行者 小浪 明

ISBN978-4-87769-138-7

