

経済ニュースによる景況感指数の足元予測

関 和広^{1,a)} 生田 祐介²

受付日 2020年9月23日, 採録日 2021年2月2日

概要: 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・景況感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）に対して景況スコアを予測する。そして、ある期間に発行された記事すべての景況スコアを統合して、その期間の景況感指数を定義する。景況スコアの予測には、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。

キーワード: ビジネスセンチメント, 深層学習, 感情分析, テキストアナリティクス

Nowcasting Business Sentiment from Economic News Articles

KAZUHIRO SEKI^{1,a)} YUSUKE IKUTA²

Received: September 23, 2020, Accepted: February 2, 2021

Abstract: Business sentiment indices released regularly by the government or the central bank play a crucial role in decision making for governmental/monetary policies, industrial production planning, and so on. However, these indices rely on traditional surveys, which are costly and time-consuming to conduct. This paper propose an approach to predicting an inexpensive and timely business sentiment index reusing daily newspaper articles. We adopt the Bidirectional Encoder Representation from Transformers (BERT) to predict a business sentiment score of a given text and aggregate the scores to define an index, named S-APIR. Also, a one-class support vector machine is applied to filter out texts irrelevant to business and economy. Moreover, we propose a simple yet useful approach to temporally analyzing how much any given factor influenced the predicted business sentiment. The validity and utility of the proposed approach are demonstrated through our experiments on 12-years-worth of newspaper articles.

Keywords: business sentiment, deep learning, sentiment analysis, text analytics

1. はじめに

政府や中央銀行が定期的に発表する景況ウォッチャー調査の現状・先行き判断 DI (diffusion index; 景況動向指

数)^{*1}や日銀短観の DI^{*2}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、景況ウォッチャー調査の現状・先行き判断 DI

¹ 甲南大学
Konan University, Kobe, Hyogo 658–8501, Japan

² 大阪産業大学
Osaka Sangyo University, Kobe, Hyogo 658–8501, Japan

a) seki@konan-u.ac.jp

^{*1} https://www5.cao.go.jp/keizai3/watcher/watcher_menu.html

^{*2} <http://www.boj.or.jp/statistics/tk/>

(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、タクシードライバーなど)2,050人を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、景気ウォッチャー指数は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。

一方、店舗の商品売り上げのような統計データ、各種ニュース記事やマイクロブログなどのテキストデータ、検索エンジンのクエリログデータ、スマートフォンなどから収集されるGPSの位置情報データ、あるいは人工衛星の衛星画像データなど、いわゆるオルタナティブデータが大量に生成・蓄積されている。このようなデータが利用可能になってきたこと、および深層学習に代表されるデータドリブンな人工知能技術の発展から、人手による調査を要さず、すでに存在する他の大規模なデータを代替的に用いた経済・金融指数の将来予測(forecast)、あるいは足元予測(nowcast)への注目が高まっている[8], [25]。具体的な取り組みとして、たとえば店舗のPOS(point of sale)データを用いた消費者物価指数の足元予測[18]、金融・経済レポートからの景況感指数の予測[26]、新聞記事からの穀物の市場価格や経済指数の予測[4], [16], [19]、ソーシャルメディアの分析による株価指数の予測[3], [7], [10]などがある。

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、2章で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、3章で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章で評価実験について報告し、語レベルでの景況感への寄与についても議論する。5章でまとめと今後の課題について述べる。

2. 関連研究

経済・金融分野では、失業率や物価指数などの統計値や個別企業の売上高など多くの数値データに加えて、新聞記事や決算短信、金融レポートなど、関連するテキストデータが豊富に存在する。従来、これらのテキスト情報は人が読み込み、他の情報・状況と考えあわせ、総合的な意思決定を行うために利用されてきた。しかしながら、公表されるすべてのテキスト情報を限られた時間で人が処理することは現実的には困難である[25]。そこで、このようなテキスト情報を用いて経済・金融指数を予測する試みが広く行われている。

表 1 景気ウォッチャー調査(現状の景気)の例

Table 1 Example responses of Economy Watchers Survey.

地域	回答者	景気	判断理由
北海道	タクシードライバー	×	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。

2.1 マーケットセンチメントの予測

大規模なテキスト情報を利用して金融指数を予測した初期の研究として、Bollenら[3]の研究がある。Bollenらは、マイクロブログサービスTwitterへの約10カ月分の投稿を収集し、感情分析を行った。具体的には、Twitterへの投稿のうち、書き手の感情を表している投稿だけを「I feel...」, 「I'm...」などのパターンを基に抽出し、6つの感情軸(Calm, Alert, Sure, Vital, Kind, Happy)を用いて数値化した。そして、これらの感情値がダウ平均株価の予測に有効であるかを検証した。その結果、Calm(平穏)の感情がダウ平均株価の終値と因果関係にあることが示された。Bollenらの試み以降、マイクロブログの感情分析を株価指数や個別銘柄の株価、ボラティリティ、出来高など、様々な予測に利用する研究が数多くなされている[11], [14]。

2.2 景況感指数の予測

1章で述べた内閣府の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)」までの5段階評価)とその理由(自然言語)が公開されている。いくつかの例を抜粋して表1に示す。

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、景気の実況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。

山本ら[26]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[6]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。

山本らの研究の発展として、饗場ら[20]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. 近藤ら [23], 坂地ら [24] 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. 五島ら [22] 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. また, 余野ら [27] は, 山本らの手法に加えて Latent Dirichlet Allocation (LDA) [2] によるトピック分析を行うことで, トピックごとの指数への寄与度についての考察を行った.

本研究では, 五島ら [22] と同様にニュース記事に着目し, より低コストで速報性の高い景況感指数の推定を試みる. ただし, 本研究では 3.1 節で述べるように, 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. また, 景況感を形成する要因について, その影響を語句レベルで時系列で分析する.

3. 景況感指数の予測と景況感変動要因分析

3.1 S-APIR 指数

ニュース記事を利用して景況感指数 (S-APIR 指数) の予測を行う枠組みを図 1 に示す. 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.

3.1.1 モデルの学習

先行研究 [26] にならい, 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 (表 1 参照) のペアを抽出し, モデルの学習を行う. 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.

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. 前者については, 近年, Bidirectional Encoder Representations from Transformers (BERT) [5] やその派生モデルが広く利用されており, 本研究でも BERT を採用する. BERT は, Transformer [17] 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. 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 BERT*3 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.

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) [12] 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) [13] で重み付けしたベクトル空間モデルを用いる.

3.1.2 景況感指数の予測

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.

3.2 景況感に影響を与える要因の分析

景況感, 金融政策や物価, 為替, 雇用, 賃金, 海外情

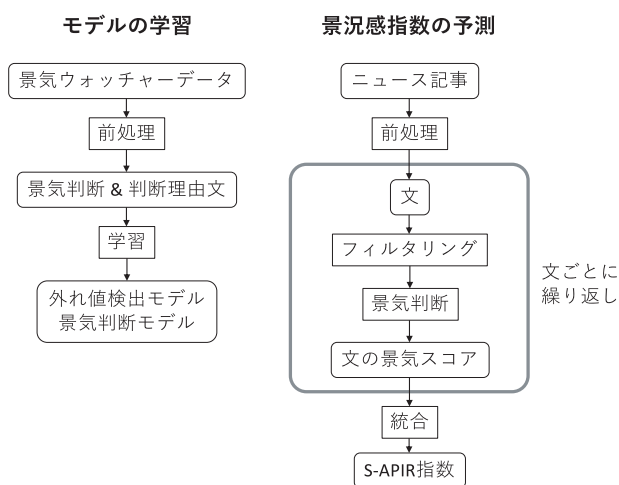


Fig. 1 Framework for computing S-APIR index.

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

勢など、様々な要因から形成されている。しかしながら、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く、経済の情勢判断を行ううえでは、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重要である。本研究では、前節の方法で推定した 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ため、以下に述べる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。

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

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。

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

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。

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

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。

4. 評価実験

4.1 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^{*4}から、2000年1月～2020年2月分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

^{*4} <https://www5.cao.go.jp>

表 2 景気ウォッチャー指数の推定結果

Table 2 Results of Economy Watchers DI prediction.

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

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 に示す。比較のため、以下のモデル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。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。山本ら [26] の先行研究に相当。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 BiRNN 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 [5] 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、青嶋ら [21] 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。

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、3.1.2 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

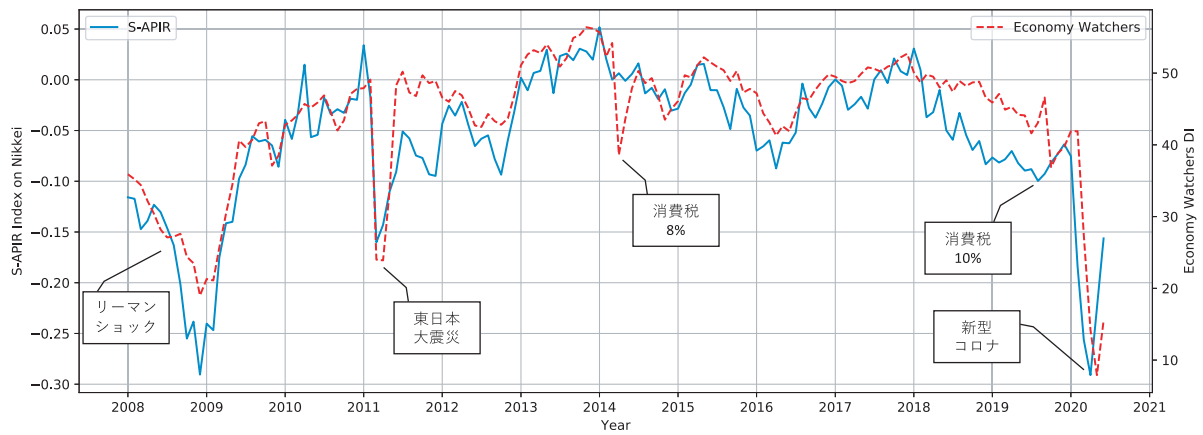


図 2 S-APIR 指数と景気ウォッチャー DI (現状) の比較 (相関係数 = 0.888)
 Fig. 2 Comparison between S-APIR index and Economy Watchers DI ($r = 0.888$).

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 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。

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

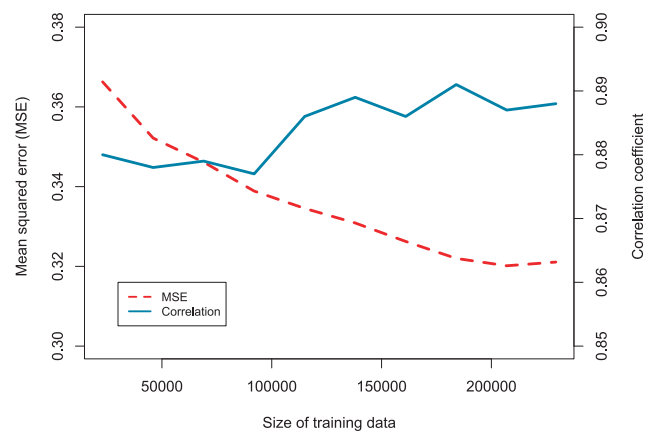


図 3 学習データ量と MSE および相関係数の関係
 Fig. 3 Relation between training data size and model performance.

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 [5]. 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 3 に示す。景気判断モデルについては MSE が学習データ 20 万件程度までは減少しているが、学習したモデルから算出された景況感指数の相関係数については、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。

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（4.2 節参照）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 [9] 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に外れ値を検出する。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 3 に示す。再現率、適合率、 F_1 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 3 には複数の閾値を試して最も F_1 値が高かったときの性能を示している。

表 3 の結果から、1 クラス SVM の方が外れ値検出の性

表 3 外れ値検出モデルの性能比較

Table 3 Performance comparison between different outlier detection models.

モデル	再現率	適合率	F_1 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

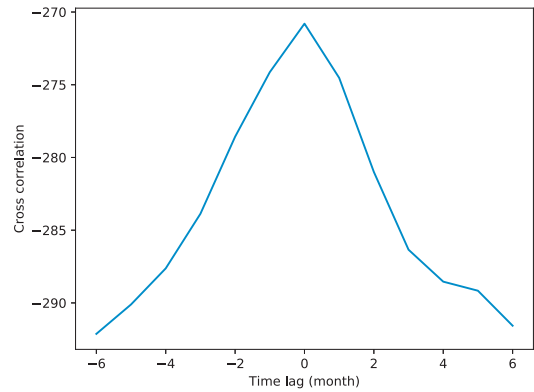


図 4 S-APIR 指数と景気ウォッチャー DI の相互相関係数
Fig. 4 Cross-correlation between S-APIR index and Economy Watchers DI.

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 1 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 4 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。

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

表 4 情報源として利用する新聞ごと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

Table 4 Correlation coefficients for different information sources with/without filtering.

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%)

るため、より粒度の高い景況感予測も可能となる。

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 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。

4.5 景況感に影響を与える要因の分析

最後に、3.2 節に述べた方法で景況感に影響を与える要因について分析した。この方法では任意の語句について分析が可能であるが、ここでは例として「貿易」、「アベノミ

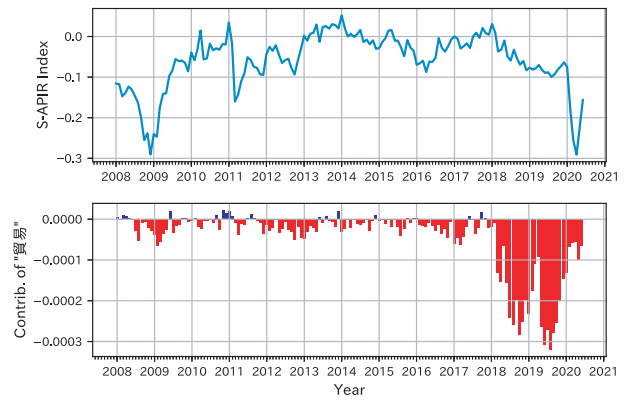


図 5 景況感への「貿易」の寄与

Fig. 5 Contribution of “trade” to S-APIR index.

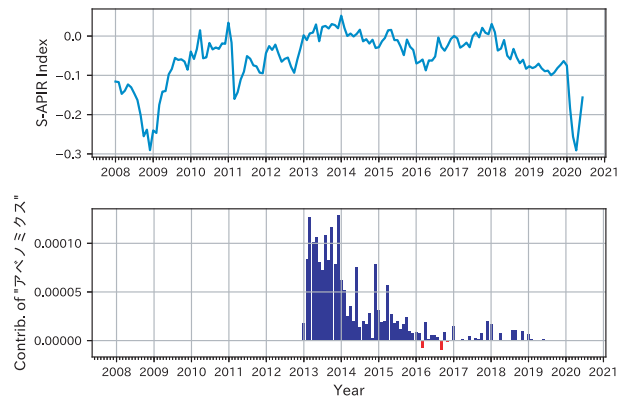


図 6 景況感への「アベノミクス」の寄与

Fig. 6 Contribution of “Abenomics” to S-APIR index.

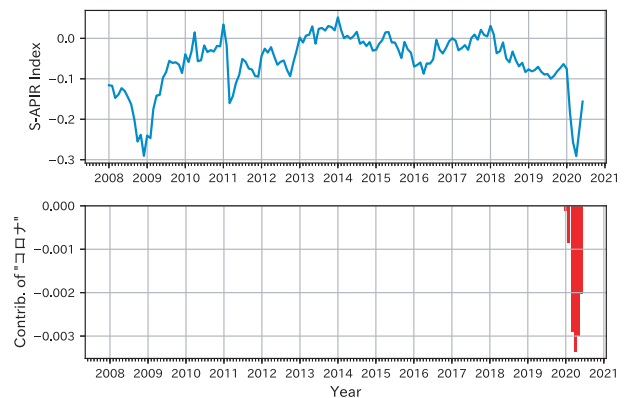


図 7 景況感への「コロナ」の寄与

Fig. 7 Contribution of “Coronavirus” to S-APIR index.

クス」、「コロナ」について景況感指数への寄与を算出した。結果を図 5、図 6、図 7 に示す。いずれも上側のグラフは S-APIR 指数の推移であり、比較のために示している。

図 5 を見ると、貿易に関する記事は 2018 年始めまで長らく景況感にほとんど影響を与えていなかったことが分かる。しかしながら、その後急速に寄与がマイナス方向に増加し、景況感に負の影響を与えている。これは、2018 年前半に始まった米中貿易摩擦を反映しているものと考えられる。

次に図 6 (アベノミクス) を見ると、2012 年 12 月に第二次安倍内閣が発足した直後から景況感にプラスに寄与していたものの、その期待・効果は年を追うごとに徐々に低下していったことが見てとれる。最後に図 7 (コロナ) 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 7 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。

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 [17] 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。しかしながら、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 [15]。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 [1]。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 (2) を次式で置き換えた。

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

この定義に基づいて「アベノミクス」の貢献度を算出し直した結果を図 8 に示す。図 8 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が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 (「貿易」と「コロナ」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 (2) の仮定は問題にはならないといえる。

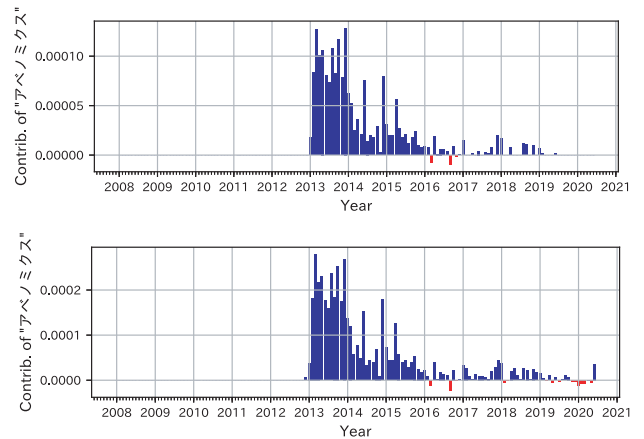


図 8 景況感への「アベノミクス」の寄与 (上は景気スコアを均等に分配 (図 6 の再掲)、下は景気スコアを attention rollout 値に応じて分配)

Fig. 8 Contribution of “Abenomics” to S-APIR index, where a sentence sentiment score was evenly distributed to its constituent words for the upper figure whereas it was divided proportionally to attention rollouts for the bottom figure.

5. おわりに

本研究では、低コストで速報性の高い景況感指数を算出すること、および景況感に影響を与える要因を時間軸に沿って分析することを目的として、日々発信されるニュース記事を再利用した指数予測の枠組みについて検討した。具体的には、1 クラス SVM による経済・景気関連文の同定、BERT による入力文の景気判断を行い、それを基に景況感指数 S-APIR を算出した。そして、2008 年以降の日経新聞を利用してその妥当性を検証した。その結果、S-APIR 指数は既存の景気ウォッチャー DI に対して高い相関 ($R = 0.888$) が得られることを確認した。また、比較の対象を企業動向に限定した場合、相関係数がさらに向上した ($R = 0.937$)。この結果は、日経新聞を利用して算出した S-APIR 指数は、企業動向をより強く反映した指数であることを示唆している。速報性については、実験によって S-APIR 指数は景気ウォッチャー調査の一致指数であることが分かった。ただし、S-APIR 指数の集計は即時に終了するため、景気ウォッチャー調査よりも 10 日程度早く公表が可能である。最後に、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。

なお、本研究では景気判断モデルに BERT を用いたが、深層学習を利用した言語処理の進展は著しく、他のより高性能なモデルが続々と発表されている。それらのモデルの利用や、景況感指数の速報値としての効果の分析は今後の課題である。

謝辞 本研究は、一般財団法人アジア太平洋研究所 (APIR) の自主研究プロジェクト「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」、JSPS 科研費 18K11558, 20H05633, 平成 30 年度私立大学等経常費補助金特別補助「大学間連携等による共同研究」により行った。APIR の宮原秀夫氏, 岩野宏氏, 本多佑三氏, 稲田義久氏, 松林洋一氏, 中山明氏の御支援に感謝する。

参考文献

- [1] Abnar, S. and Zuidema, W.: Quantifying attention flow in transformers, *Proc. 58th Annual Meeting of the Association for Computational Linguistics*, pp.4190–4197 (2020).
- [2] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent dirichlet allocation. *The Journal of Machine Learning Research*, Vol.3, pp.993–1022 (2003).
- [3] Bollen, J., Mao, H. and Zeng, X.-J.: Twitter mood predicts the stock market, *Journal of Computer Science*, Vol.2, No.1, pp.1–8 (2011).
- [4] Chakraborty, S., Venkataraman, A., Jagabathula, S. and Subramanian, L.: Predicting socio-economic indicators using news events, *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1455–1464 (2016).
- [5] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.4171–4186 (2019).
- [6] Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural Computation*, Vol.9, No.8, pp.1735–1780 (1997).
- [7] Huang, T.-C., Zaeem, R.N. and Barber, K.S.: It is an equal failing to trust everybody and to trust nobody: Stock price prediction using trust filters and enhanced user sentiment on Twitter, *ACM Trans. Internet Technology*, Vol.19, No.4 (Sep. 2019).
- [8] Kapetanios, G. and Papailias, F.: *Big data & macroeconomic nowcasting: Methodological review*, Economic Statistics Centre of Excellence (ESCoE) Discussion Papers ESCoE DP-2018-12, Economic Statistics Centre of Excellence (ESCoE) (2018).
- [9] Kieu, T., Yang, B. and Jensen, C.S.: Outlier detection for multidimensional time series using deep neural networks, *Proc. 19th IEEE International Conference on Mobile Data Management (MDM)*, pp.125–134 (2018).
- [10] Levenberg, A., Pulman, S., Moilanen, K., Simpson, E. and Roberts, S.: Predicting economic indicators from web text using sentiment composition, *International Journal of Computer and Communication Engineering*, Vol.3, No.2, pp.109–115 (2014).
- [11] Li, B., Chan, K., C.C. Ou, C. and Ruifeng, S.: Discovering public sentiment in social media for predicting stock movement of publicly listed companies, *Information Systems*, Vol.69, pp.81–92 (2017).
- [12] Manevitz, L.M. and Yousef, M.: One-class SVMs for document classification, *The Journal of Machine Learning Research*, Vol.2, pp.139–154 (2002).
- [13] Manning, C.D., Raghavan, P. and Schütze, H.: *Introduction to information retrieval*, Cambridge University Press, New York (2008).
- [14] Oliveira, N., Cortez, P. and Areal, N.: The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices, *Expert Systems with Applications*, Vol.73, pp.125–144 (2017).
- [15] Serrano, S. and Smith, N.A.: Is attention interpretable?, *Proc. 57th Annual Meeting of the Association for Computational Linguistics*, pp.2931–2951 (2019).
- [16] Shapiro, A.H., Sudhof, M. and Wilson, D.: *Measuring news sentiment*, Federal Reserve Bank of San Francisco Working Paper 2017-01, Federal Reserve Bank of San Francisco (2017).
- [17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I.: Attention is all you need, *Proc. 31st International Conference on Neural Information Processing Systems*, pp.6000–6010 (2017).
- [18] Watanabe, K. and Watanabe, T.: Estimating daily inflation using scanner data: A progress report, *CARF Working Paper Series*, No.CARF-F-342 (2014).
- [19] Yoshihara, A., Seki, K. and Uehara, K.: Leveraging temporal properties of news events for stock market prediction, *Artificial Intelligence Research*, Vol.5, No.1, pp.103–110 (2016).
- [20] 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).
- [21] 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).
- [22] 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).
- [23] 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).
- [24] 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).
- [25] 関 和広, 水田孝信, 八木 勲, 落合友四郎, 酒井浩之, 和泉 潔: 金融情報学: ファイナンスにおける人工知能応用, 人工知能, Vol.32, No.6, pp.905–910 (2017).
- [26] 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).
- [27] 余野京登, 和泉 潔: 金融レポート, およびマクロ経済指数によるリアルタイム日銀センチメントの予測, 第 31 回人工知能学会全国大会論文集, 2D1-3 (2017).



関 和広 (正会員)

2002 年図書館情報大学情報メディア研究科修士課程修了。2006 年インディアナ大学図書館情報学研究科博士課程修了。神戸大学助教, 准教授等を経て現在, 甲南大学知能情報学部教授。Ph.D. 人工知能学会会員。



生田 祐介

2016年神戸大学経済学研究科博士後期課程修了。アジア太平洋研究所研究員等を経て現在大阪産業大学経営学部講師。博士（経済学）。日本経済政策学会，日本経済学会，人工知能学会各会員。