

日経・UTEcon 日次景気指数の作成手順

東京大学エコノミックコンサルティング株式会社

2022 年 2 月 17 日

1. 極性辞書に収録する単語の選択

本節では、極性辞書に収録する単語の選択手順について記述する。本稿では、マクロ経済分析、特に景気分析に適した極性辞書の構築を目的としていることから、経済や景気と関連のある単語を獲得するために株式会社日本経済新聞社が発行する日本経済新聞の記事データ（朝刊・夕刊・電子版）を利用する。利用する新聞記事データの期間は 1981 年 10 月から 2020 年 12 月であり、おおよそ 510 万記事が収録されている。また、新聞記事データに含まれる単語数について、異なり数が 1,631,065、延べ数が 1,621,203,196 である。はじめに、「景気」、「経済」もしくはそれらの合成語と各単語の共起回数を算出する。具体的には、各単語について、「景気」、「経済」もしくはそれらの合成語の少なくともいずれか一つと共起した記事数を共起回数とする。さらに、共起回数による単語のランキングを作成する。

ここで、考慮する単語は名詞、動詞、形容詞のみとし、またすべての単語は基本形に変換して収録する。単語分割、品詞推定および基本形の推定には、MeCab + mecab-ipadic-NEologd を利用した。次に、共起回数のランキング上位からそれぞれマクロ経済学及び金融・ファイナンスを専門分野とする筆者 2 人によって、経済や景気と関連があり、収録すべき単語をそれぞれ独立に 2,000 語を選択する。ここで、「リーマンショック」や「新型コロナウイルス」などの特定のイベント固有の単語は使用せず、景気に係る一般的な単語のみを選択した。そして、それぞれ選択した単語の重複を排除した後、全体で 3,000 語になるように調整を行い、単語リストを作成した。最終的な単語リストの品詞の内訳は形容詞：309 語、動詞：989 語、名詞：1,702 語となった。

2. アノテーションとその信頼性

1 節で得られた単語リストに収録されている 3,000 語に対して、景気動向の観点から各単語から連想される意味について、ポジティブ・ニュートラル・ネガティブの 3 つのクラスラベルの付与を行うアノテーションを行った。本稿のアノテーションには経済システムや景気動向に関する深い理解が必要であることから、アノテータとしてマクロ経済分析に従事する 3 人の専門家に依頼した。なお景気の見方については、アノテータのこれまでの経歴や置かれている立場によっても異なることが想定されるため、クラスラベルが特定の業種

の意見を反映しないように、アノテータはそれぞれ別の職種の専門家（政府・中央銀行エコ

- 景気動向の観点から、各単語から連想される意味について、ポジティブ/ネガティブ/ニュートラルの3つに分類してください。（ポジティブ: +1, ニュートラル: 0, ネガティブ: -1,）例:「悪化」→ネガティブ、「回復」→ポジティブ
- 一般的には良い意味でも、景気の文脈では悪い意味の場合はネガティブ、どちらとも言えない場合はニュートラルとしてください。悪い意味についても同様をお願いします。
- 一般的に文脈に依存する単語や係り受けによって意味が反転する単語の場合でも、可能な限り、景気が良い時期や内容に利用される場合が多い単語はポジティブ、悪い時期や内容に利用される場合が多い単語はネガティブとしてください。どちらとも言えない単語については、ニュートラルとしてください。例:「不確実」は一般的には係り受けによってどちらの意味にもなりうるが、景気の文脈だと悪い局面に使用される場合が多いのでネガティブと判断する。
- ポジティブ/ネガティブとニュートラルの境目にあると考えられる場合は、ニュートラルに分類してください。
- ひらがなについては、複数の漢字が当てはまる場合であっても、景気の文脈で最も多く使われる使用法を念頭に選んでください。

ノミスト、民間金融機関エコノミスト、報道機関エコノミスト）を選定した。アノテータへの指示書は以下のとおりである。

表1は、各アノテータによるクラス分類の結果とそれらを多数決によって集計した結果である。アノテータ間で各クラスへの分類割合は多少変わるものの、ニュートラルが一番多く、次いでネガティブ、ポジティブの順番に分類されている傾向は総じて同じである。分類割合が異なったのは各アノテータの判断基準の閾値が異なっていることが要因として想定される。

表1：アノテーションの集計結果

アノテーター	ポジティブ	ニュートラル	ネガティブ
A	372	2,203	425
B	191	2,496	313
C	383	2,110	507
多数決	282	2,317	401

また、表2はアノテータ間の評価に関する混同行列、表3は評価の一致度を測る κ 係数である。表2を見ると、アノテータ間で不一致だったクラスはポジティブとニュートラルあるいはネガティブとニュートラルの間であり、ポジティブとネガティブの間で不一致だ

ったものはわずかである。実際に、どのような単語が真反対のクラスラベルを付与されたか
 というと、「カネ余り」や「過熱感」、「バブル」といった単語であり、立場によって判断が
 分かれる単語といえる。表 3 を見ると、 κ 係数は 0.5 から 0.6 程度であり、アノテーション
 の結果は十分に信頼できる水準といえる。これらの結果を踏まえて、アノテータによる多数
 決によって 3,000 語に対するクラスラベルを決定した。

表 2：アノテータ間の混同行列

		B		
		ネガティブ	ニュートラル	ポジティブ
A	ネガティブ	242	183	0
	ニュートラル	70	2,087	46
	ポジティブ	1	226	145

		C		
		ネガティブ	ニュートラル	ポジティブ
A	ネガティブ	329	90	6
	ニュートラル	175	1,885	143
	ポジティブ	3	135	234

		C		
		ネガティブ	ニュートラル	ポジティブ
B	ネガティブ	248	59	6
	ニュートラル	258	2,008	230
	ポジティブ	1	43	147

表 3：アノテータ間の一致度

	A vs. B	A vs. C	B vs. C	A vs. B vs. C
κ 統計量	0.52	0.59	0.49	0.53

3. 収録単語の補完

本節では、2 節で得られた極性辞書への収録単語の補完を行う。現段階では、新聞記事データにおいて「景気」及び「経済」とそれらの合成語と共に起した単語のみが辞書に収録されており、これら以外の単語について辞書への収録に関して取りこぼしが存在する可能性があり、補完する必要がある。そこで、本稿では単語の埋め込みベクトルと教師あり学習を用いて単語の補完を行う。具体的な手順は以下のとおりである。

- (ア) はじめに、単語の埋め込みベクトルを作成する。単語ベクトルを作成する際のコーパスとしては、日本経済新聞の記事データを利用している。なお、本稿では極性辞書に基本形のみを収録するため、入力テキストを形態素解析によって基本形に変換している。単語ベクトルを作成する際の手法として、Skip-gram(word2vec)、GloVe、fastText の 3 種類を採用した。ハイパーパラメータはデフォルトを指定しているが、ベクトルの次元は 100, 200, 400, 800 の 4 種類を試行している。
- (イ) 次に、単語ベクトルを入力、クラスラベルを出力として教師あり学習を行う。教師あり学習の手法としては、ロジスティック回帰 (MLR)、サポートベクターマシン (SVM)、順伝播型ニューラルネットワーク (FFNN)、LightGBM を用いる。なお、SVM に関しては線形カーネル、ガウシアンカーネル、シグモイドカーネル、多項式カーネルの 4 つのカーネル関数を試行するとともに、One-vs-the-rest によって 3 クラス分類を行った¹。また、表 1 が示すとおりクラスラベルの割合が不均衡であることから、その逆数を掛けることで損失関数への寄与度を調節している。
- (ウ) 各組み合わせについて 5 分割の交差検証によってテストデータに対する分類性能の評価を行い、分類性能の最も高い組み合わせを利用して、収録単語以外の単語のラベルを推定する。なお、訓練データに対する 5 分割の交差検証とグリッドサーチによって、ハイパーパラメータのチューニングを行っている²。分類性能を測る指標としては、F-measure、Precision、Recall の 3 指標について、マクロ平均、マイクログ平均、加重平均の 3 つの方法で集計した計 9 つの指標である³。

¹ なお、One-versus-one による分類精度と比較したところ、総じて One-vs-the-rest による分類精度の方が高いことを確認している。

² 具体的には、訓練データを 5 分割して、そのうちの 1 つのデータセットを取り出して開発データとし、開発データに対するあてはまりが最も良い (F-measure の加重平均スコアが最も高い) パラメータをグリッドサーチによって選択している。この一連の作業を別のデータセットを開発データとして取り出して同様にグリッドサーチを行い、5 回のスコアの平均値が高いパラメータに決定する。そして、全ての訓練データを用いて、決定したパラメータによってモデルを学習している。

³ 加重平均では、各クラスのデータサイズの逆数で重み付けしている。F-measure の加重

(エ) 推定されたラベルをもとに、再度人手で極性辞書に収録する単語を選定する。具体的には、固有名詞を除外したり、経済関連単語のみを抽出したりする。

平均は、Precision と Recall の加重平均を用いて算出されている。

表 4：極性分類の結果

学習モデル	埋め込み手法	次元数	F1			Precision			Recall		
			weighted	micro	macro	weighted	micro	macro	weighted	micro	macro
MLR	skipgram	100	0.823	0.820	0.688	0.828	0.820	0.676	0.820	0.820	0.705
		200	0.847	0.844	0.733	0.851	0.844	0.717	0.844	0.844	0.752
		400	0.859	0.858	0.751	0.861	0.858	0.746	0.858	0.858	0.759
		800	0.857	0.857	0.746	0.858	0.857	0.745	0.857	0.857	0.748
	glove	100	0.798	0.792	0.652	0.808	0.792	0.634	0.792	0.792	0.678
		200	0.809	0.806	0.669	0.815	0.806	0.658	0.806	0.806	0.686
		400	0.816	0.813	0.686	0.823	0.813	0.672	0.813	0.813	0.706
		800	0.825	0.824	0.694	0.827	0.824	0.692	0.824	0.824	0.702
	fasttext	100	0.819	0.816	0.681	0.823	0.816	0.671	0.816	0.816	0.693
		200	0.845	0.844	0.730	0.848	0.844	0.721	0.844	0.844	0.741
		400	0.854	0.853	0.749	0.857	0.853	0.744	0.853	0.853	0.758
		800	0.863	0.863	0.759	0.863	0.863	0.762	0.863	0.863	0.757
SVM	skipgram	100	0.863	0.865	0.750	0.863	0.865	0.766	0.865	0.865	0.739
		200	0.872	0.873	0.769	0.872	0.873	0.788	0.873	0.873	0.758
		400	0.872	0.876	0.763	0.871	0.876	0.805	0.876	0.876	0.734
		800	0.878	0.881	0.777	0.880	0.881	0.815	0.881	0.881	0.754
	glove	100	0.835	0.839	0.701	0.834	0.839	0.729	0.839	0.839	0.683
		200	0.843	0.843	0.723	0.848	0.843	0.735	0.843	0.843	0.723
		400	0.845	0.851	0.718	0.846	0.851	0.764	0.851	0.851	0.691
		800	0.852	0.860	0.727	0.853	0.860	0.792	0.860	0.860	0.686
	fasttext	100	0.852	0.853	0.738	0.853	0.853	0.747	0.853	0.853	0.734
		200	0.860	0.861	0.754	0.861	0.861	0.767	0.861	0.861	0.748
		400	0.870	0.875	0.764	0.871	0.875	0.815	0.875	0.875	0.730
		800	0.876	0.878	0.781	0.877	0.878	0.801	0.878	0.878	0.769
FFNN	skipgram	100	0.813	0.805	0.684	0.835	0.805	0.668	0.805	0.805	0.729
		200	0.853	0.854	0.737	0.855	0.854	0.745	0.854	0.854	0.736
		400	0.862	0.863	0.749	0.861	0.863	0.763	0.863	0.863	0.738
		800	0.862	0.865	0.750	0.862	0.865	0.772	0.865	0.865	0.734
	glove	100	0.768	0.746	0.643	0.824	0.746	0.604	0.746	0.746	0.743
		200	0.827	0.832	0.689	0.825	0.832	0.718	0.832	0.832	0.668
		400	0.825	0.829	0.686	0.823	0.829	0.706	0.829	0.829	0.671
		800	0.822	0.825	0.689	0.824	0.825	0.702	0.825	0.825	0.685
	fasttext	100	0.816	0.808	0.689	0.832	0.808	0.668	0.808	0.808	0.726
		200	0.849	0.853	0.733	0.849	0.853	0.763	0.853	0.853	0.711
		400	0.853	0.857	0.733	0.853	0.857	0.768	0.857	0.857	0.711
		800	0.859	0.862	0.747	0.859	0.862	0.772	0.862	0.862	0.731
LightGBM	skipgram	100	0.825	0.834	0.667	0.823	0.834	0.711	0.834	0.834	0.641
		200	0.839	0.848	0.697	0.839	0.848	0.753	0.848	0.848	0.664
		400	0.836	0.845	0.693	0.835	0.845	0.756	0.845	0.845	0.656
		800	0.829	0.841	0.677	0.832	0.841	0.760	0.841	0.841	0.634
	glove	100	0.809	0.815	0.646	0.807	0.815	0.674	0.815	0.815	0.631
		200	0.806	0.818	0.637	0.804	0.818	0.702	0.818	0.818	0.602
		400	0.815	0.820	0.663	0.816	0.820	0.690	0.820	0.820	0.652
		800	0.806	0.810	0.651	0.807	0.810	0.677	0.810	0.810	0.641
	fasttext	100	0.829	0.833	0.686	0.828	0.833	0.717	0.833	0.833	0.667
		200	0.841	0.852	0.700	0.844	0.852	0.785	0.852	0.852	0.654
		400	0.830	0.837	0.690	0.829	0.837	0.733	0.837	0.837	0.663
		800	0.845	0.852	0.716	0.848	0.852	0.762	0.852	0.852	0.692

表 4 は、各組み合わせにおけるテストデータに対する分類結果をまとめたものであり、初期値を変えて各モデルの推定・予測を 5 回試行した際の平均値である。また、最も高い分類精度の数値は太字で表示している。なお、SVM については、最も分類性能の高かった多項式カーネルの結果のみを記載している。表 4 を見ると、総じて SVM と Skip-gram の組み合わせの分類性能が最も高いことが分かる。また、次元は大きいほど分類精度が上がる傾向にあるが、400 次元と 800 次元とではほぼ同程度、もしくはわずかに 800 次元の方が良いことが分かる。これらの結果から、収録単語以外の単語のクラスラベル推定における組み合わせには、サポートベクターマシンと 800 次元の Skip-gram を採用する。そして、改めて 3,000 語を用いて学習し、新聞記事データに含まれる全単語のクラスラベルを推定する。ポジティブあるいはネガティブのクラスラベルが付与された単語からアノテーション済みの単語を除外したのち、経済や景気と関連のある単語を再度筆者らによって選定した。推定された各クラスの単語と追加収録単語の内訳については、表 5 の示すとおりである。ポジティブが 39 語、ネガティブが 152 語の計 191 語を追加収録する

表 5：推定された各クラスの単語と追加収録単語の内訳

	ポジティブ	ニュートラル	ネガティブ	計
全単語数	672	637,077	2,386	640,135
内、単語リストとの重複単語数	290	2,280	430	3,000
重複排除後の単語数	382	634,797	1,956	637,135
追加収録単語	39		152	191

3 節のアノテーションの多数決結果からニュートラルを除外したものと、得られた追加収録単語と合わせることで、最終的な景気単語極性辞書を構築する。景気単語極性辞書の収録単語の内訳は表 6 に示すとおりである。

表 6：極性辞書の収録単語の内訳

	ポジティブ	ネガティブ
名詞	198	369
動詞	85	113
形容詞	38	71
計	321	553

4. 景気単語極性辞書の応用例：経済ニュース指数の構築

本節では、構築した景気単語極性辞書の使用例の一つとして、経済ニュース指数の構築を試行する。はじめに、1980年10月～2021年12月の日経新聞記事の朝刊に対して、日ごとに登場する辞書に収録されているポジティブな単語の数とネガティブな単語の数を式(1)によって数え上げることで、スコア ($Score_t$) を算出する。

$$Score_t = \frac{Positive\ Words_t - Negative\ Words_t}{Positive\ Words_t + Negative\ Words_t} \quad (1)$$

ここで、 $Positive\ Words_t$ と $Negative\ Words_t$ はそれぞれ、 t 日に配信された新聞記事中に出現したポジティブな単語の数とネガティブな単語の数を表す⁴。また、後述の分析における分かりやすさのため、本稿では算出されたスコアについて水準を100倍する。ただし、日次単位で計測したスコアでは不規則変動が多いため、このままではスコアの動きから景気の転換点を判定することが困難である。そこで、式(2)、(3)の指数平滑移動平均を用いて平滑化を行う。

$$News\ Index_t = Score_t$$

$$News\ Index_t = Score_t \times \alpha + News\ Index_t \times (1 - \alpha) \quad (2)$$

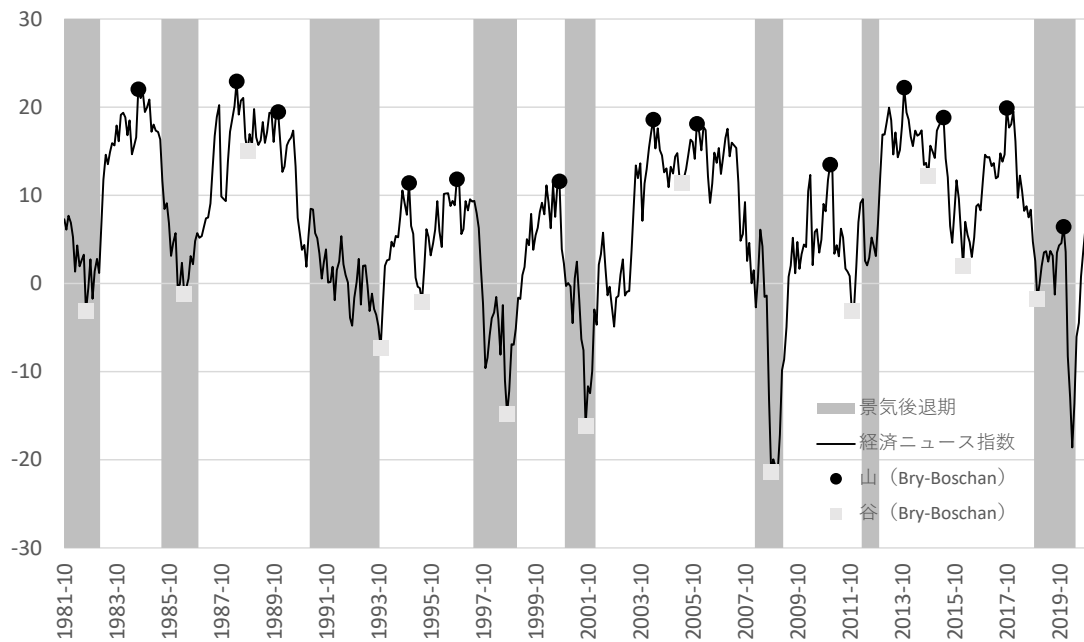
ここで、 α は平滑化係数であり、 $\alpha = 2/n + 1$ である。本稿では1か月単位での平滑化を行うため、 $n = 30$ とした。また、一般的な辞書との比較として、高村らの単語感情極性値対応表を用いて同様の手続きによってニュース指数を構築し、景気単語極性辞書に基づくニュース指数とパフォーマンスを比較する。

図1は、各辞書を用いて構築した経済ニュース指数の月末値の推移を図示したものであり、実線が構築したニュース指数、灰色の影が内閣府によって認定された景気基準日付に基づく景気後退期の期間である⁵。図1の(a)を見ると、景気単語極性辞書に基づく経済ニュース指数はバブル崩壊やITバブル崩壊、世界金融危機等の景気後退期において指数が下がっている様子が分かる。また、2020年のCovid-19の初期においても大きく指数が下がっており、景気との連動性の高い経済ニュース指数が構築できていることが見て取れる。一方で、図1の(b)を見ると、単語感情極性値対応表に基づく経済ニュース指数は景気後退期において指数が下がっている部分もあるものの、(a)と比較すると景気との連動性が総じて低いことが見て取れる。

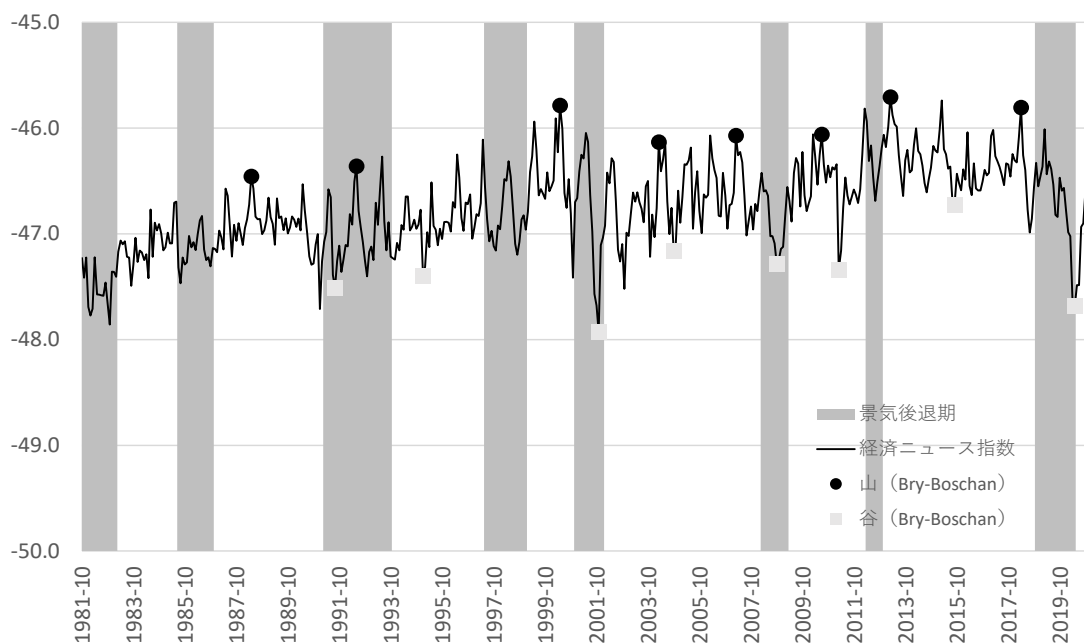
⁴ 休刊日のスコアはゼロとした。

⁵ 景気基準日付については、内閣府のウェブページ

(<https://www.esri.cao.go.jp/jp/stat/di/hiduke.html>) から取得可能である。



(a) 景気単語極性辞書に基づく経済ニュース指数



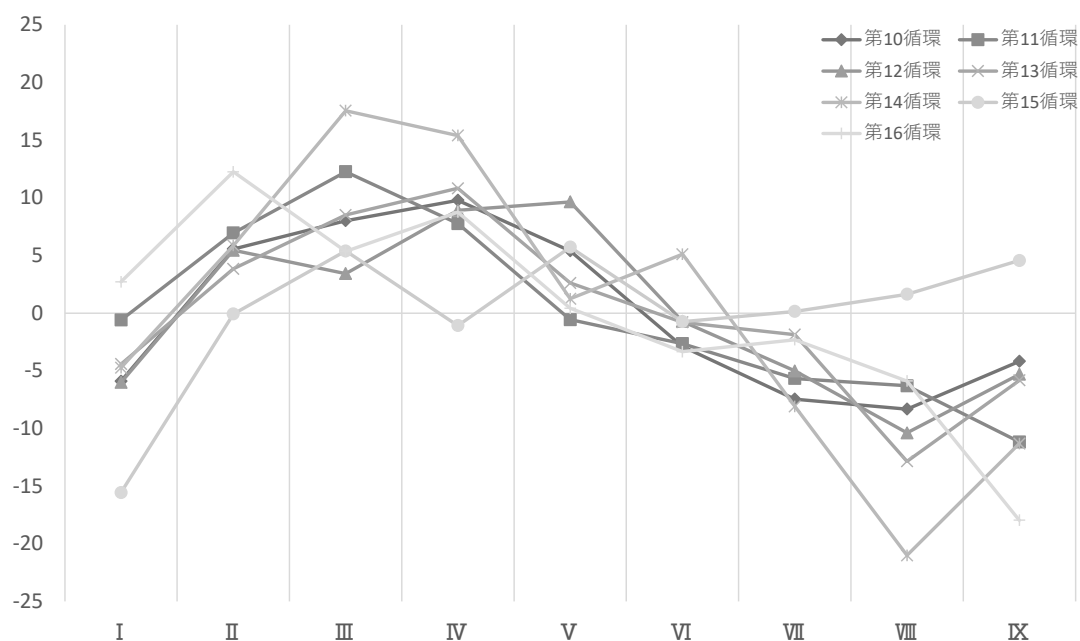
(b) 単語感情極性値対応表に基づく経済ニュース指数

図 1：経済ニュース指数の外観

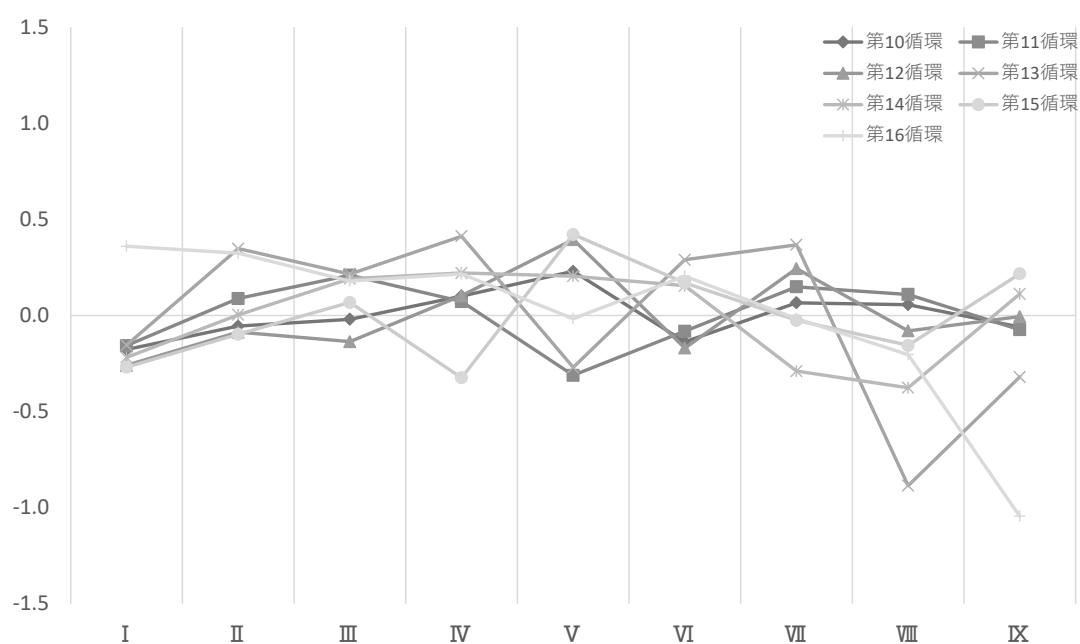
さらに、パターン分析、Bry-Boschan 法、各景気指標との時差相関の 3 つの手法を通じて、構築した経済ニュース指数と景気動向との関係について考察を行う。これらの手法は景気動向に対する個別系列のパフォーマンスを分析するための方法論として広く用いられており、わが国においても内閣府が公表する景気動向指数への採用系列を議論する際に利用されている。

図 2 は、各辞書を用いて構築したニュース指数に対して、パターン分析を適用した結果であり、各景気循環・各ステージにおけるニュース指数を図示したものである。なお、本稿の経済ニュース指数は、戦後の景気循環のうち第 10 循環から第 16 循環までの期間（1983 年 2 月～2020 年 5 月）を分析対象とする⁶。

⁶ 各景気循環の山谷については、内閣府のウェブページ
(<https://www.esri.cao.go.jp/jp/stat/di/hiduke.html>) を参照されたい。



(a) 景気単語極性辞書に基づく経済ニュース指数



(b) 単語感情極性値対応表に基づく経済ニュース指数

図2：パターン分析

パターン分析においては、第5ステージが景気のピークに対応するが、図2の(a)に示すように、経済ニュース指数のピークの多くは第3ステージ～第4ステージに位置している。また、第9ステージが景気の底に対応するが、ニュース指数の底は概ね第8ステージに位置している。このように、ニュース指数の変動は、現実の景気に対して概ね第1ステージ～2ステージ程度前に変動が現れる。換言すれば、景気の回復局面、後退局面が、先行してニ

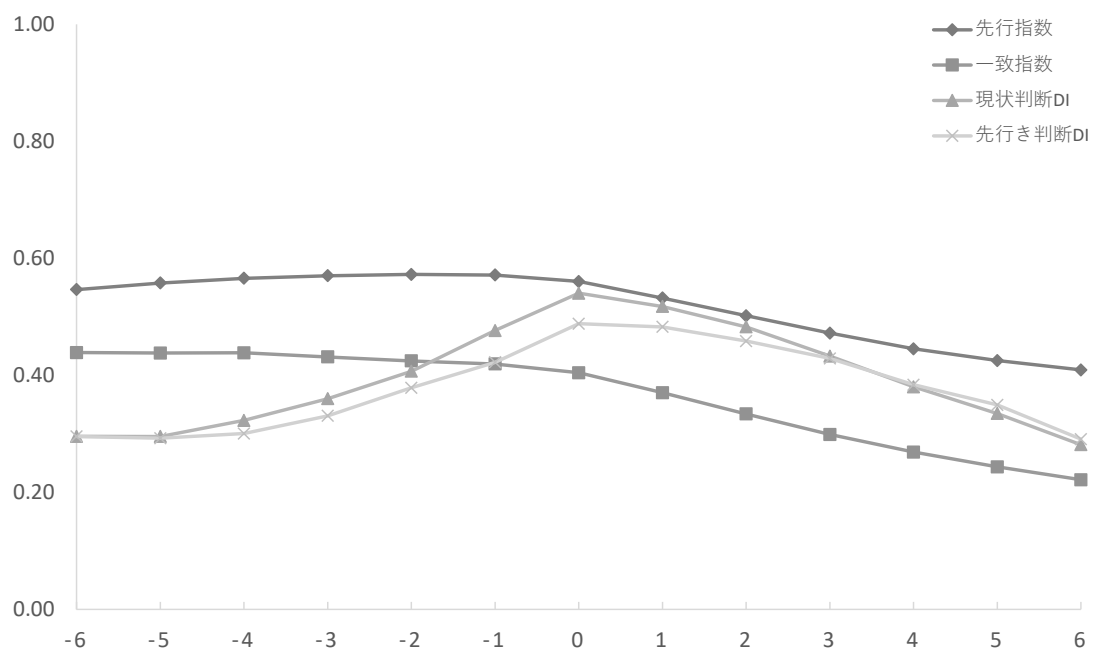
ニュース指数に現れる。つまり、ニュース指数には、現実には景気動向に先行してその変動が現れるので、ニュース指数を参照することで、将来の景気動向の変動を予測することが可能となる。一方で、図2の(b)を見ると、単語感情極性値対応表による経済ニュース指数は、ステージごとに上下が相対的に少ないことから景気動向の関連性が低く、また先行性も観察されない。

Bry-Boschan 法とは、山と谷との間隔が5か月以上必要である、1循環の長さが15か月以上必要であるなどのルールを条件として与え、各種移動平均を施すなどして系列の山谷を確定するアルゴリズムである。系列の山谷をみることにより、景気基準日付とのリード・ラグの関係や系列の振幅など、当該系列の特徴をチェックできる。図1に示されている山・谷は Bry-Boschan 法によって認定された山・谷である。景気単語極性辞書に基づく経済ニュース指数に関して、系列の振幅の回数は景気循環より多いものの、概ねタイミングは景気基準日付と一致している。また、山・谷ともに政府が認定した景気循環より早くに転換点が来ていることが分かる。一方で、単語感情極性値対応表に基づく経済ニュース指数に関しては、系列の山・谷のタイミングは景気基準日付と一致していない場合も多い。

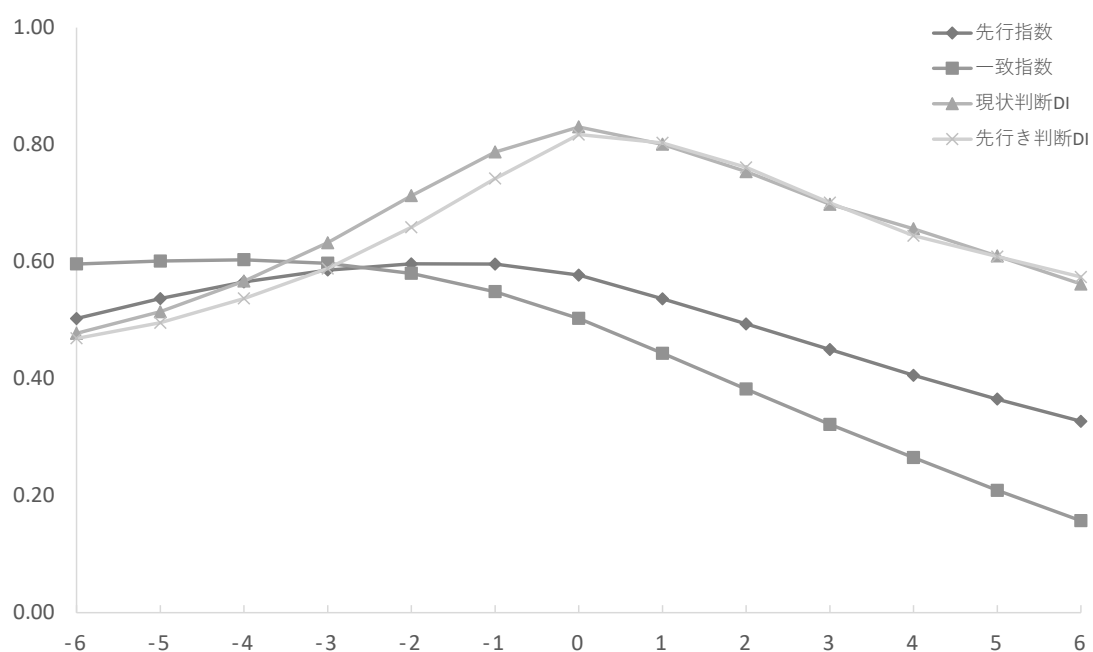
最後に、政府や中央銀行が公表している各景気指標に対する経済ニュース指数の先行遅行関係について、時差相関係数を通じて考察する。時差相関係数は2つの変数のうち、一方にタイムラグをとり、これともう一つの変数との相関係数を計算することによって求められる。そして、いくつかのラグ期間を試行し、相関係数の推移を観察することで、2つの時系列データの先行遅行関係を考察する。本稿では、内閣府及び日本銀行が公表する景気指標との先行遅行関係を分析する。表7は利用する各景気指標のまとめたものである。

表7：各景気指標のまとめ

景気指標	公表者	公表頻度	利用系列	内容
景気動向指数	内閣府	月次	先行指数 一致指数	生産、雇用など様々な経済活動での重要かつ景気に敏感に反応する指標の動きを統合することによって、景気の現状把握及び将来予測に資するために作成された指標
景気ウォッチャー調査			現状判断DI 先行き判断DI	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の協力を得て、地域ごとの景気動向を的確かつ迅速に把握し、景気動向判断の基礎資料
全国企業短期経済観測調査 (日銀短観)	日本銀行	四半期	大企業-製造業-実績 大企業-製造業-予測 大企業-非製造業-実績 大企業-非製造業-予測	全国の企業動向を的確に把握し、金融政策の適切な運営に資することを目的に作成された指標
生活意識に関する アンケート調査			景況感-前年比 景況感-予想	生活者が現状において抱いている生活実感や、金融・経済環境の変化がもたらす生活者の意識や行動への影響を把握するために作成される指標

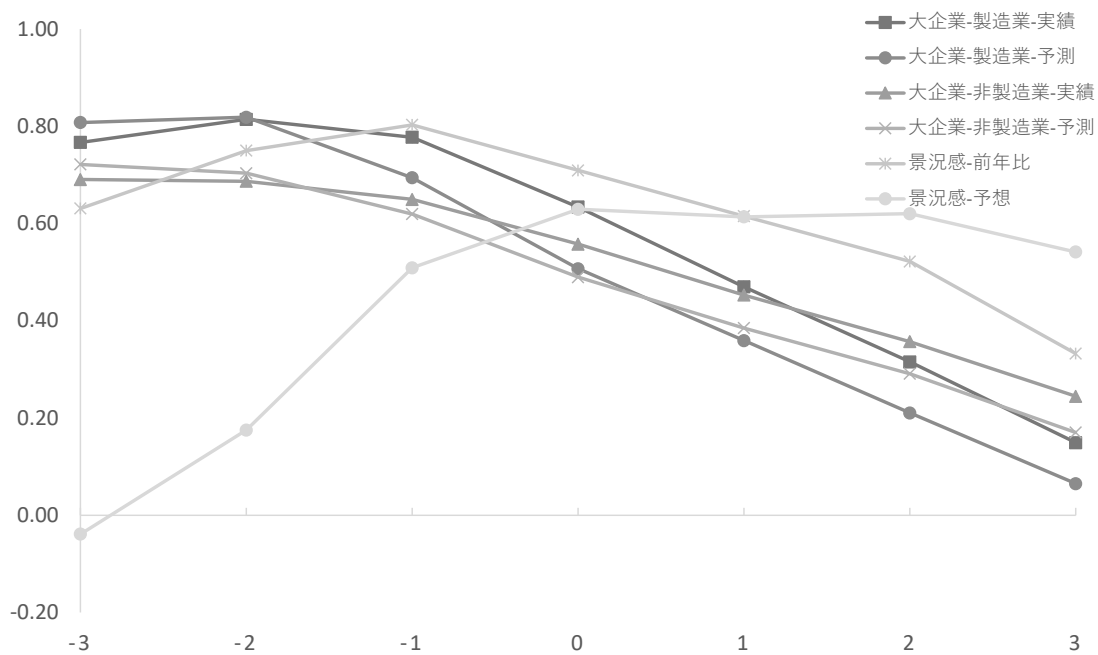


(a) 景気単語極性辞書に基づく経済ニュース指数

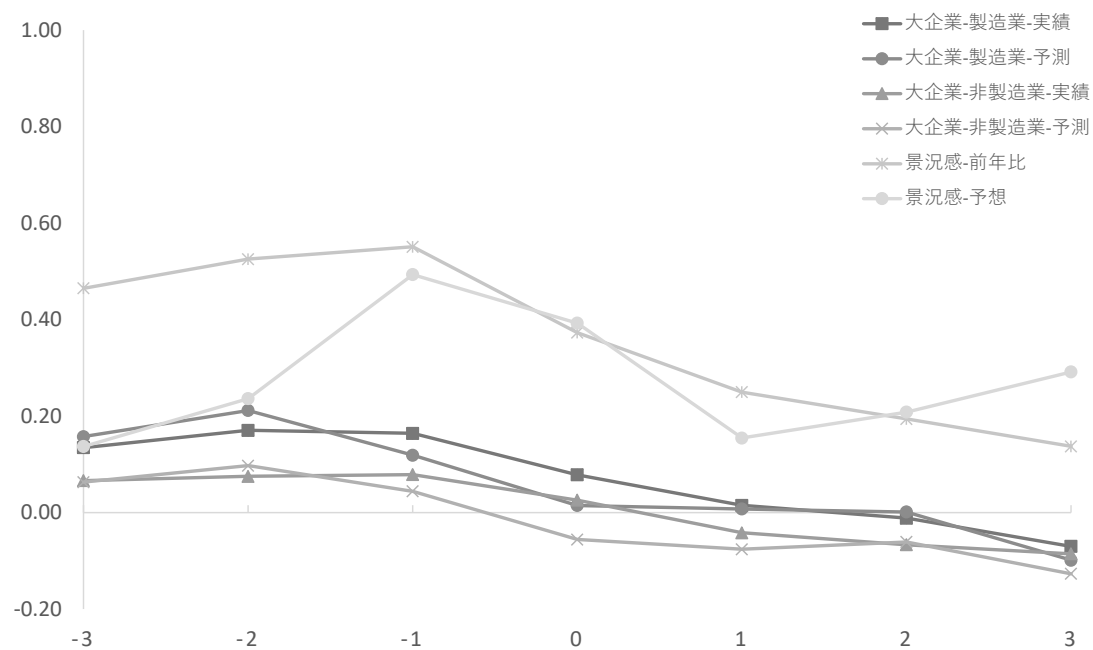


(b) 単語感情極性値対応表に基づく経済ニュース指数

図 3：内閣府が公表する景気指標との時差相関



(a) 景気単語極性辞書に基づく経済ニュース指数



(b) 単語感情極性値対応表に基づく経済ニュース指数

図4：日本銀行が公表する景気指標との時差相関

図3と図4は、経済ニュース指数と各景気指標との時差相関係数をまとめたものであり、縦軸は相関係数、横軸は時差（月/四半期）を示している。そして、左方向は経済ニュース指数が先行する方向、右方向は経済ニュース指数が遅行する方向を示している。図3の(a)を見ると、内閣府が公表している各景気指標と高い相関があり、構築したニュース指数は景

気動向を捉えていることが分かる。景気動向指数に対しては先行性が見られ、先行指数とは-2 カ月、一致指数とは-4 カ月で相関係数が最も大きい値を取ることから、景気動向のいち早い捕捉ができる可能性を示唆している。景気ウォッチャー調査に対しては、先行性は見られないものの同時点で相関係数が 0.8 と非常に高い値を取っており、景気動向との連動性が高いことが分かる。また、図 4 の(a)を見ると、日本銀行が公表している各景気指標と高い相関があり、構築したニュース指数は景気動向を捉えていることが分かる。日銀短観に対しては、-3~-2 四半期で相関係数が最も大きい値を取ることから、同様に先行性が見られる。生活意識に関するアンケート調査に対しては、景況感-前年比とは-1 四半期、景況感-予想とは同時点で相関係数が最大となっている。さらに、これらの指標は集計から発表までのラグがあるため、経済ニュース指数は実務上の足許予測（ナウキャスト）の面でも優れている。一方で、図 3 と図 4 の(b)を見ると、単語感情極性値対応表に基づく経済ニュース指数は総じて各景気指標との相関係数が相対的に低いことが分かる。これらの結果から、景気動向との関連性という観点では、本稿で構築した景気単語極性辞書に基づく経済ニュース指数のパフォーマンスが良いことが分かる。そして、景気単語極性辞書を用いれば、景気動向をいち早く捕捉できる経済ニュース指数を構築することができる可能性を示している。

5. まとめ

本稿では、経済分析におけるテキストデータ分析の発展に寄与することを目的として、経済ドメインに適した極性辞書の構築とその応用例を示した。具体的には、日本経済新聞の新聞記事コーパスを用いるとともに、複数人の専門家による単語の選定によって、経済ドメインに適した極性辞書の構築を行った。また、応用例の一つとして、構築した極性辞書を用いて景気動向を捕捉することが可能な経済ニュース指数を構築した。