

大規模言語モデルを用いた ニュースデータのセンチメント判定モデルの開発および 実体経済センチメントインデックスの構成

野間 修平 *†
Shuhei Noma

2024/11/9

概要:

近年、市場や経済といった金融ドメインにおけるニュースデータの活用が盛んに研究されている。その中でも、経済指標に替わって実体経済の様子を速やかに補足する指数や、将来の経済および市場に関する予測因子の構成は、その実用性から盛んに研究されている。ニュースデータからこうした指数や予測因子を構成する際には、その記事または記事を構成するセンテンスに対してセンチメントを判定する必要がある。センチメントの判定は辞書ベースの手法が使われることが多いものの、辞書ベースの手法は構文や文脈を正しく補足することが困難であるという問題があり、これはセンチメントを誤って判定する要因となり得る。そこで本稿では構文や文脈を勘案することができる大規模言語モデルを用いてニュースを構成するセンテンスのセンチメントを判定し、経済指標の予測や市場分析に有用なセンチメントインデックスを構成する。構成されたセンチメントインデックスは市場参加者からの注目度が高い景況感指数との連動性が高いことに加えて、その予測因子として有用であるほか、債券市場におけるレジームの判定に有用である。

Keywords: ニュースデータ, マクロ経済, 大規模言語モデル, 自然言語処理, 債券市場

* Japan Digital Design 株式会社 MUFU AI Studio(shuhei.noma@japan-d2.com)

† 本稿の内容・意見は、筆者個人に属するものであり、所属組織の見解を示すものではない。

1 序論

近年、市場や経済といった金融ドメインにおけるニュースデータの活用が盛んに研究されている。

ニュースデータを金融ドメインの分析に活用する研究は3種に大別される。1つ目は政府、企業、中央銀行などが、その意思決定のために参照すべく、実体経済の様子を捉える高頻度の指数を構成することを目指す研究である (Nguyen et al., 2020) (Shapiro et al., 2022) (Seki et al., 2022)。インフレ率や雇用者数といった経済指標はその調査や算出に多大なコストを必要することに加え、多くの指標は月次で算出、公表されるため、速報性に欠けるという問題を抱える。一方、ニュースデータから構成される指数は日次で算出することが可能であるため、こうした経済指標が抱える問題に対する処方箋となる。2つ目はGDPのような重要な指標を予測するための予測因子を構成することをモチベーションとする研究である (Barbaglia et al., 2023) (Ashwin et al., 2021)。3つ目は株価や金利など市場データの予測や要因分解にニュースデータを活用するものである。例えば、(Allen et al., 2019) および (Born et al., 2023) は株式市場、(Smales, 2014) は金市場、(Perico Ortiz, 2023) は期待インフレ率、(Beetsma et al., 2013) および (Erlwein-Sayer, 2018) は2国間の国債スプレッドについて、ニュースデータを用いた分析を行っている。

ニュースデータからセンチメントインデックスを構成する際にはニュース記事、または記事を構成するセンテンスや単語に対してセンチメントを判定する手続きを踏むこととなる。(Nguyen et al., 2020) (Ashwin et al., 2021)。センチメントの判定には事前に単語とセンチメントとを紐づけた辞書を用いることが多い (Sebastiani and Esuli, 2006) (Loughran and McDonald, 2011)。しかし、こうした辞書ベースによるセンチメントの判定は否定のような構文や文脈を考慮することができないという問題がある。

そこで本稿ではこうした問題を解決すべく、ニュースデータから得られたセンテンスのセンチメントを大規模言語モデル (Devlin, 2018) (Liu, 2019) により判定し、センチメントの強さを定量化する系列であるセンチメントインデックスを構成する。センチメントを判定するトピックの選定には任意性がある。例えば、(Audrino and Offner, 2024) では利回り、インフレーション、失業率に関するセンチメントを定量化しており、(Shapiro et al., 2022) では広くマクロ経済に関するセンチメントを定量化している。本稿では実体経済を構成する様々な因子の中でも、米連邦準備制度がその調整を義務付けられている¹ために市場参加者からの注目度が特に高い、物価動向と労働市場に対するセンチメントを定量化する。

構成されたセンチメントインデックスは市場参加者からの注目度が高いISM製造業景況感指数との連動性が高いことに加え、その予測因子として有用である。さらに、センチメントインデックスは米国債利回りの変化をよく説明し、リバーサルの傾向がみられやすい相場を検出する。

1.1 本論文の構成

本節では本論文の構成を示す。第2章では本分析で使用するニュースデータについて詳説する。特に、第2.1節ではニュースの本文から適切にセンテンスを抽出する手法を提案する。第3章と第4章では抽出されたセンテンスに対してセンチメントを判定するモデルについて詳説する。センチメントの判定は2段階からなる。1段階目ではセンテンスがマクロ経済に関するトピックについて言及しているかをキーワードおよび言語モデルにより判定する。2段階目ではマクロ経済トピックに言及していると判定されたセンテンスに対して、同じく言語モデルによりそのセンチメントを判定する。第3章では1段階目のトピック判定モデルを、第4章では2

¹ Humphrey-Hawkins Full Employment Act

段階目のセンチメント判定モデルを、それぞれ詳説する。第 5 章ではセンチメントインデックスの構成方法について詳説する。第 6 章では構成されたセンチメントインデックスの有用性を 3 つの実証分析により評価する。

2 データ

本章ではセンチメントインデックスの構成に使用するニュースデータと、そこからセンチメントを付与する対象となるセンテンスを抽出する方法を詳説する。ニュースデータはロンドン証券取引所グループが提供する Reuters ニュースのデータを使用する。本研究では 1996 年から 2023 年までに公表された、英語で執筆されており、かつ本文が存在する 13,574,861 記事を分析対象とした。

各記事は関連する事象の発生、誤記、関係者からの申し出などにより、複数回にわたって改訂されることがある。Reuters ニュースではこうした改訂の履歴をストーリーと呼ばれる概念を用いて管理している。例えば、ある国で大規模なデモが発生したとき、速報的に報じられる本文がないヘッドラインのみの記事や、その後公表されるデモについて詳細に報じた本文付きの記事、そして、デモが終了した旨が追記された記事などは、いずれも同一のデモについて報じたものであるため、同一のストーリーに属するものとして管理される。

同一のストーリーに属する記事は、その改訂がたとえ僅かなものであっても別の記事として保持される。そこで、改訂が多いストーリーに属する記事に偏ってセンチメントを判定することを避けるため、本研究では各ストーリーから 1 つの記事を抽出することとする。抽出する記事は同一のストーリーに属する記事のうち、最新のものとする。ただし、本研究では日次の頻度でセンチメントインデックスを構成することを目指すため、抽出する記事は同一のストーリーに属する最古の記事が報じられた日と同じ日に報じられた記事の中で最新のものとする。

ここで、直近の米国の労働市場について論じた記事を例にとって考える。労働市場に関する指標は雇用者数、失業率、失業保険申請件数、離職率など多岐にわたるため、それら全てが労働市場の緩みまたは引き締まりを示唆することは稀であり、記事を構成するセンテンスごとに労働市場に対するセンチメントが相異なるという状況が想定される。そこで、本研究では記事全体に対してセンチメントを付与するのではなく、記事中のセンテンスを抽出し、各センテンスに対してセンチメントを付与することを考える。

2.1 センテンスの抽出

本節ではニュース記事の本文からセンテンスを抽出するための手続きを提案する。当該手続きは以下に掲げる一連の処理から成る。

- 本文の前処理
 - ヘッダーの消去
 - フッターの消去
 - タグの消去
- センテンスへの段階的な分割
 - ブロックへの分割
 - 箇条書きの分割
 - センテンスへの分割

上記の手続きによってニュース本文に前処理を施し、センテンスへ分割する様子を図 1 に示す。以下では、各

処理を施す背景およびその詳細について述べる。

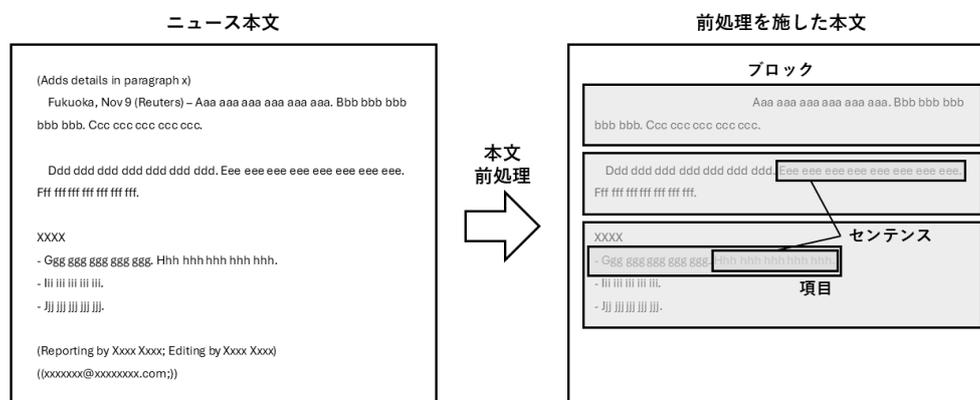


図1: ニュース本文からセンテンスへ分割する手続き. まず, 前処理としてヘッダー, フッター, タグを消去した後, 段階的にセンテンスに分割する.

2.1.1 本文の前処理

Reuters ニュースの記事本文には報じている事象と直接的な関係がない文字列が多く含まれる. 具体的には, ヘッダー, フッター, タグが本文中に記されている. こうした文字列は記事が報じる出来事と直接的な関係がないため, センチメントインデックスを構成する上で不要な情報であると判断し, 記事本文をセンテンスへ分割するに先立ち消去する.

ヘッダーは本文の冒頭に記された一連の情報である. ヘッダーには記事を執筆した記者の名前, 記者が所属する報道機関の名称, ニュースの配信日, 事象が発生した場所などが記されている. フッターは本文の末尾に記された一連の情報である. フッターには関連するニュースへのリンク集, 本文中に表れた専門用語に関する注釈, 記事を執筆した記者の名前や連絡先, 記者が所属する企業が提供するサービスの広告などが記されている. タグは2種類存在する. 一方はニュースに関するタグであり, 当該記事と関連する出来事について報じた別の記事や, 特定の地域や資産について言及する記事をまとめたグループへのリンク機能を担っている. もう一方は資産や経済指標に関するタグであり, 当該記事中で言及された資産や経済指標の詳細やチャートを掲載するページへのリンク機能を担っている.

2.1.2 センテンスへの段階的な分割

ヘッダー, フッター, タグを消去する前処理を施した本文をセンテンスに分割する. センテンスの分割には Python の NLTK ライブラリ²に実装された `sent_tokenize()` メソッドを用いる. しかし, `sent_tokenize()` メソッドはピリオドの位置を参照するなど, 比較的シンプルなルールによってセンテンスを判定するため, 前処理を施した本文に対して同メソッドを素朴に適用するのみでは, 高品質な, すなわち, 文らしい文を得ることは叶わない. そこで, より高品質なセンテンスを得るべく, `sent_tokenize()` メソッドの適用に先立ち, さらに以下に掲げる2つ前処理を施す.

1つ目の前処理はブロックへの分割である. Reuters ニュースの記事には改段落や挿入された表の前後にお

² <https://github.com/nltk/nltk>

いて、空行もしくはハイフンなどの記号によって構成された仕切り線などによって本文が複数のブロックに分割されていることがある。本文をこうしたブロックに分割することで、ブロック中の最後のセンテンスとそれ続くブロック中の最初のセンテンスが、単一のセンテンスとして抽出されることを防ぐことができる。

2つ目の前処理は箇条書きの分割である。sent_tokenize() メソッドは Reuters ニュースの記事中に表れる箇条書きに対応することができない、すなわち、行頭文字によって区切られたセンテンスの羅列を各項目に分割することができないことが実験的に確かめられた。そこで、本文中の箇条書きについては、ルールベースの手法によって事前に項目に分割する。

本研究では以上の前処理を施したブロックまたは項目に対して sent_tokenize() メソッドを適用することでセンテメントの判定対象となるセンテンスを得る。

3 トピックの判定によるセンテンスの絞り込み

本研究ではマクロ経済の動向や個社の経営判断に関するニュースからセンチメントインデックスを構成することを目指す。センチメントを判定するモデルをファインチューニングによって獲得するために、ファインチューニングに適した訓練データを用意する必要がある。しかし、Reuters ニュースは市場や経済に関する事象のみならず、司法、政治、スポーツなど、多様な事象について報じている。こうした事象に対して言及するセンテンスのセンチメントは、正負および中立という3種の極性を付与するセンチメント判定モデルにおいては中立的とアノテートされるべきものである。従って、任意の記事から得られたセンテンスをすべて訓練データとして採用してしまうと、そのラベルは中立に偏り、訓練データの質を毀損する。

そこで、本研究では高質な訓練データを用意するため、抽出されたセンテンスがマクロ経済、市場、個社の経営判断といった、そのセンチメントを評価したいトピックに対して言及しているか否かを判定し、訓練データとして採用するセンテンスを絞り込むことを提案する。トピックの判定は2段階で行う。1段階目の判定はルールベースによる。すなわち、特定のキーワードを含むか否かという素朴なルールによりセンテンスを絞り込む。しかし、キーワードとして採用すべき単語には“job”や“cost”など、多義的なものが考えられる。そのため、こうした多義的なキーワードを広く採用すると、キーワードによるルールベースの判定のみでは関心のあるトピック以外に対して言及するセンテンスが混入してしまう。そこで、キーワードによって大まかに絞り込まれたセンテンス群から、真に関心があるトピックに対して言及するセンテンスを抽出するため、言語モデルを用いて2段階目の判定を行う。

ニュースから得られたセンテンスのトピックを判定し、センチメントを判定するまでの一連の手続きは図2のようにまとめられる。以下では、両段階によるトピック判定について詳説する。

3.1 キーワードによるトピック判定

1段階目のトピック判定は抽出されたセンテンスが特定のキーワードを含むかを判定するという素朴な方法により行う。物価動向トピックおよび労働市場トピックを判定するためのキーワードとして、本研究では以下に掲げる単語群を採用した。

物価動向トピックの判定に用いたキーワード

cost, deflation, deflator, disinflation, expense, fare, fee, freight, inflation, price, surcharge

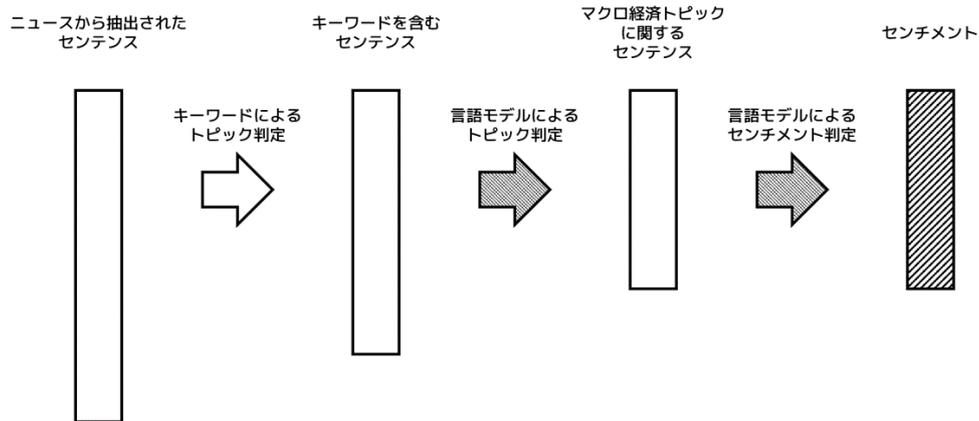


図2: センチメントを判定するまでの一連の手続き

労働市場トピックの判定に用いたキーワード

applicant, employee, employer, employment, hire, job, jobless, labor, lay off, layoff, payroll, staff, unemployment, worker

ただし、名詞についてはその複数形、動詞についてはその活用形についてもキーワードに含めてセンテンスの抽出を行った。また、大文字と小文字は区別せずにセンテンスの抽出を行った。

分析対象となる 13,574,861 記事のうち、上記キーワードのいずれかを含むものは 4,711,140 記事存在した。また、そこから第 2.1 節の方法によって抽出されたセンテンスのうち、上記キーワードのいずれかを含むものは 11,961,644 文存在した。

3.1.1 “price” を含むセンテンスについて

本研究では物価動向と労働市場に関するセンチメントの定量化を目指しているため、市場について言及しているセンテンスはセンチメントの判定対象から除かれることが望ましい。しかし、“price” という単語は物価動向について言及するセンテンスで頻出の単語であるものの、同時に株価 (stock price) やオプションの行使価格 (strike price) といった複合名詞の形で、市場について言及しているセンテンス中に頻繁に表れるという問題がある。そこで、以下に掲げる複合名詞を含むセンテンスは市場について言及しているものが多数を占めるという判断に基づき、センチメントの判定対象から除外することとする³。

センチメントの判定対象から除外するために用いたキーワード

cut-off price, issue price, share price, stock price, strike price, target price

3.2 マクロ経済トピック判定モデル

2 段階目のトピック判定は言語モデルによって行う。すなわち、第 2.1 節および第 3.1 節で提案した手続きによって抽出されたセンテンスが、マクロ経済の動向や個社の経営判断といったトピックに言及するものである

³ “bond price” を含むセンテンスは物価動向や労働市場について言及していることが多いため、意図的に除外対象とするキーワードから除いている。

か否かを言語モデルによって判定する。第3章の冒頭で述べた通り、Reuters ニュースは市場や経済に関する事象のみならず、司法、政治、スポーツなど、多様な事象について報じている。そのため、本研究ではファインチューニングに際して金融など特定のドメインに特化した言語モデル (Araci, 2019) (Lee et al., 2024) をバックボーンモデルとして使用することは不適と判断し、一般的な事前学習済み言語モデルをファインチューニングすることによりトピック判定モデルを得ることとする。事前学習済みの言語モデルには RoBERTa(Liu, 2019) を用いる。

ファインチューニングには訓練データ、すなわち、マクロ経済トピックに該当するか否かのアノテーションがなされた大量のセンテンスが必要である。本研究ではコストの観点から人手によるアノテーションを避け、GPT-4 Turbo(Achiam et al., 2023) によるアノテーションを行う。

3.2.1 アノテーション

本項では各センテンスが“マクロ経済トピック”に言及するものであるか否かを GPT-4 Turbo によりアノテートする方法を詳説する。

アノテートの対象となるセンテンスは、第3.1節の方法で抽出されたキーワードを含むセンテンス群から、13,440 文をランダムサンプリングすることにより得る。ただし、抽出されるセンテンスが執筆されたタイミングが特定の時期に偏ることを防ぐべく、1996年1月から2023年12月までの336ヶ月から1月につき40文ずつ抽出する。なお、各月から抽出する40文は、物価動向に関するキーワードを含むものと労働市場に関するキーワードを含むものを20文ずつ抽出する。

本研究では GPT-4 Turbo を用いてセンテンスのアノテートを行う。プロンプトは以下のように定めた。ただし、プロンプト末尾の“xxx”にはニュースから抽出されたセンテンスが格納される。

トピックを判定するためのプロンプト

Behave like you are an outstanding economist.

Read the following sentence, and classify the sentence into “MACROECONOMICS” or “OTHER_ TOPIC” class.

Label as “MACROECONOMICS” if the sentence mainly refers to recent macroeconomics such as inflation or labor market.

Also label as “MACROECONOMICS” if the sentence mainly refers to a business decision such as job cuts or a price increase.

Label as “OTHER_ TOPIC” if the sentence mainly refers to other than macroeconomics, such as market, judiciary, politics, diplomacy, sports, environment, human rights.

Also label as “OTHER_ TOPIC” if the author’s subjective views, predictions, analyses or thoughts are described in the sentence.

Provide only the label that you classified.

The sentence: xxx

3.2.2 トピック判定の結果

第3.1節で述べた通り、事前に定めたキーワードを含むセンテンスは11,961,644文存在した。このうち、ファインチューニングされた RoBERTa によってマクロ経済トピックへの言及がなされていると判定されたセンテンスは5,198,284文存在した。

ここで、判定の精度について確認する。事前に定めたキーワードを含む 11,961,644 文のセンテンスから 112 文を無作為に抽出し⁴、これらのセンテンスに対して GPT-4 Turbo およびファインチューニングされた RoBERTa が下した判定と、当該センテンス群が実際にマクロ経済トピックへ言及しているかを調べた。ただし、抽出されるセンテンスが執筆されたタイミングが特定の時期に偏ることを防ぐべく、無作為抽出は 1996 年から 2023 年のニュースデータについて、各年から 4 文ずつセンテンスを抽出した。

図 3 は GPT-4 Turbo およびファインチューニングされた RoBERTa の精度と、両モデルによる判定の類似性を混同行列の形式でまとめたものである。なお、各センテンスが真にマクロ経済トピックへ言及しているか否かは、プロンプトの定義に従って著者が判断した。図 3-(a), (b) はそれぞれ GPT-4 Turbo およびファインチューニングされた RoBERTa の混同行列を表している。各モデルの精度はそれぞれ 84.8%, 86.6% となっており、高い精度を示していることが分かる。図 3-(c) は両モデルによる判定の類似性をまとめたものである。一致率は 87.5% となっており、ファインチューニングされた RoBERTa は GPT-4 Turbo をよく模倣することができていると分かる。

		GPT	
		Macro	Other
True	Macro	38	7
	Other	10	57

(a)

		RoBERTa	
		Macro	Other
True	Macro	39	6
	Other	9	58

(b)

		RoBERTa	
		Macro	Other
GPT	Macro	41	7
	Other	7	57

(c)

図3: GPT-4 Turbo およびファインチューニングされた RoBERTa によるニュース本文から抽出されたセンテンスがマクロ経済トピックに対して言及しているか否かを判定するタスクの精度と、両モデルによる判定の類似性を混同行列の形式でまとめたもの。(a), (b) はそれぞれ GPT-4 Turbo およびファインチューニングされた RoBERTa に対応する混同行列を表しており、(c) は両モデルによる判定の類似性をまとめている。GPT-4 Turbo およびファインチューニングされた RoBERTa の精度はそれぞれ 84.8%, 86.6% となる。

4 センチメント判定モデル

本章ではマクロ経済トピックに対して言及していると判定されたセンテンスが、物価動向および労働市場に関していずれの方向のセンチメントを有しているかを判定する手法を詳説する。センチメントを判定する手法は大きく 3 種に大別される。1 つ目は辞書ベース (Sebastiani and Esuli, 2006) (Loughran and McDonald, 2011) と呼ばれる手法であり、事前に単語とその極性のペアが定義された辞書を用意し、文章ないしセンテンスに登場する極性が定義された単語の数をカウントすることを通じてセンチメントを判定する。2 つ目はルールベース (Hutto and Gilbert, 2014) と呼ばれる手法であり、単語とその極性に関する辞書の情報に加えて、否定など言語的な構文を考慮して文章ないしセンテンスのセンチメントを判定する。3 つ目は機械学習ベースと呼ばれる手法であり、大規模言語モデルをはじめとする機械学習モデルを用いてセンチメントを判定する。

辞書ベースおよびルールベースの手法は分析目的に応じて固有の辞書を構築しなければならないことに加え、事前に定めたルールに該当しない複雑な構文に対応することは難しいという問題がある。そこで、本研究では機械学習ベースの手法を採用する。マクロ経済トピック判定モデルと同じく、汎用的な事前学習済み言語

⁴ ただし、ファインチューニングに用いたセンテンスは除いた。

モデルである RoBERTa をファインチューニングすることによりセンチメント判定モデルを獲得することを提案する。

金融ドメインの文章は語彙や表現について一般的な文章と相違点が多くみられる。そのため、金融ドメインに特化した言語モデル（以下、Fin-LLM）の開発が盛んに研究されている (Araci, 2019) (Lee et al., 2024)。こうしたモデルは金融ドメインのデータセットを大量に学習しているため、先に提案したような金融ドメインに特化していない汎用的な事前学習モデルを独自にファインチューニングすることは、一見非効率であるかに思われる。しかし、多くの Fin-LLM がその事前学習において用いているセンチメント判定のデータセットは、株価に対する影響に基づいてアノテーションされている (Malo et al., 2014) (Maia et al., 2018)。そのため、本研究の目的である実体経済に対するセンチメントを判定するために既存の Fin-LLM を用いてセンチメントを評価することは不適であると考えられる。なぜなら、実体経済に対するセンチメントとしてはネガティブと判定されるべき個別企業に関するニュースを、当該企業の株価に対してポジティブなものであると投資家がみならず事例が想定されるからである。

例えば、「ある企業が大規模な人員削減を計画している」というニュースは、人件費の観点から当該企業が発行する株式に対してポジティブな影響を及ぼすと投資家に解釈され得る。しかし、このニュースが労働市場に対して与える影響はネガティブなものとして判定されてしかるべきである。他にも、ある商品価格の下落を報じるニュースは、その商品を原材料とする製造業を営む企業にとっては増益要因、すなわち、株価に対してはポジティブで解釈され得るものである一方、物価動向に対してはネガティブなものとして判定されるべきである。

以上の考察に基づき、本研究では汎用的な言語モデルをファインチューニングすることにより、トピック判定を行うモデルを獲得することを提案する。事前学習済みの言語モデルにはマクロ経済トピック判定モデルと同じく RoBERTa(Liu, 2019) を用いる。また、ファインチューニングに用いる訓練データのアノテーションについても、マクロ経済トピック判定モデルと同じく GPT-4 Turbo によるアノテーションを行う。

4.1 アノテーション

本節ではマクロ経済トピックに対して言及していると判定されたセンテンスが、物価動向および労働市場に関していずれの方向のセンチメントを有しているかを GPT-4 Turbo によりアノテートする方法を詳説する。“物価動向モデル” および “労働市場モデル” のどちらについても、ファインチューニングに用いる訓練データは第 3.2 節で提案した手法によって “マクロ経済トピック” と判定されたセンテンス 5,198,284 文から、13,440 文をランダムサンプリングすることにより得る。ただし、抽出されるセンテンスが執筆されたタイミングが特定の時期に偏ることを防ぐべく、1996 年 1 月から 2023 年 12 月までの 336 ヶ月から 1 月につき 40 文ずつ抽出する。

“物価動向モデル” および “労働市場モデル” について、アノテーションに用いるプロンプトは以下のように定めた。ただし、プロンプト末尾の “xxx” にはニュースから抽出されたセンテンスが格納される。

物価動向トピックを判定するためのプロンプト

Behave like you are an outstanding economist.
 Read the following sentence, and classify the price trend estimated from the sentence into “HIGH”, “LOW” or “NEUTRAL” class.
 Label as “HIGH” if prices are increasing or high.
 Label as “LOW” if prices are decreasing or low.
 Label as “NEUTRAL” if prices are little changed or some price trends are conflict.
 Provide only the label that you classified.
 The sentence: xxx

労働市場トピックを判定するためのプロンプト

Behave like you are an outstanding economist.
 Read the following sentence, and classify the labor market condition estimated from the sentence into “TIGHT”, “LOOSE” or “NEUTRAL” class.
 Label as “TIGHT” if labor market is tight, even if only slightly, or if wage is increasing.
 Label as “LOOSE” if labor market is loose, even if only slightly, or if wage is decreasing.
 Label as “NEUTRAL” if labor market is unchanged or suggestions of labor market are conflict.
 Provide only the label that you classified.
 The sentence: xxx

4.2 センチメント判定の結果

マクロ経済トピックに対する言及がなされていると判定されたセンテンスに対してファインチューニングされた RoBERTa によるセンチメント判定を施した結果は表 1 にまとめられる。

表1: ファインチューニングされた RoBERTa によるセンチメントの判定結果

	物価動向			労働市場		
	pos	neu	neg	pos	neu	neg
本文を持つニュース件数	13,574,861					
キーワードを含むニュース件数	4,711,140					
キーワードを含むセンテンス数	11,961,644					
	8,885,810			3,284,631		
マクロ経済トピックへの言及があると判定されたセンテンス数	5,198,284					
	3,504,532			1,876,908		
センチメントの判定結果	1,709,127	777,615	1,017,790	595,334	487,239	794,335

ここで、精度について確認する。第 3 章の手法によりマクロ経済トピックに対して言及していると判定された 5,198,284 文のセンテンスから、物価動向および労働市場に関するキーワードを含むセンテンスをそれぞれ

112 文ずつ無作為に抽出し⁵、これらのセンテンスに対して GPT-4 Turbo およびファインチューニングされた RoBERTa が下した判定と、当該センテンスが持つ実際のセンチメントを調べた。ただし、抽出されるセンテンスが執筆されたタイミングが特定の時期に偏ることを防ぐべく、無作為抽出は 1996 年から 2023 年のニュースデータについて、各年から 4 文ずつセンテンスを抽出した。なお、各センテンスが持つ真のセンチメントは、プロンプトの定義に従って著者が判断した。

図 4 は GPT-4 Turbo およびファインチューニングされた RoBERTa の精度と、両モデルによる判定の類似性を混同行列の形式でまとめたものである。上段の 3 つの表が物価動向 (**Inflation Trend**) に対するセンチメント判定モデル、下段の 3 つの表が労働市場 (**Labor Market**) に対するセンチメント判定モデルの結果をまとめている。図 4-(IT-a), (IT-b), (LM-a), (LM-b) はそれぞれ GPT-4 Turbo およびファインチューニングされた RoBERTa の混同行列を表している。各モデルの精度は物価動向についてが 86.6%, 83.9%, 労働市場についてが 81.3%, 81.3% となっており、高い精度を示していることが分かる。(IT-c) および (LM-c) は両モデルによる判定の類似性をまとめたものである。物価動向に対するセンチメントを判定するモデルも労働市場に対するセンチメントを判定するモデルも GPT-4 Turbo およびファインチューニングされた RoBERTa の一致率はともに 78.6% となっており、ファインチューニングされた RoBERTa は GPT-4 Turbo をよく模倣することができていると分かる。

		GPT		
		High	Neu	Low
True	High	46	3	0
	Neu	2	27	1
	Low	1	8	24

(IT-a)

		RoBERTa		
		High	Neu	Low
True	High	45	2	2
	Neu	4	20	6
	Low	3	1	29

(IT-b)

		RoBERTa		
		High	Neu	Low
GPT	High	44	2	3
	Neu	6	21	11
	Low	2	0	23

(IT-c)

		GPT		
		Tight	Neu	Loose
True	Tight	25	4	3
	Neu	5	23	4
	Loose	1	4	43

(LM-a)

		RoBERTa		
		Tight	Neu	Loose
True	Tight	28	3	1
	Neu	6	21	5
	Loose	3	3	42

(LM-b)

		RoBERTa		
		Tight	Neu	Loose
GPT	Tight	26	3	2
	Neu	5	21	5
	Loose	6	3	41

(LM-c)

図4: GPT-4 Turbo およびファインチューニングされた RoBERTa によるニュース本文から抽出されたセンテンスが持つ物価動向および労働市場に対するセンチメントを判定するタスクの精度と、両モデルによる判定の類似性を混同行列の形式でまとめたもの。上段の 3 つの表が物価動向 (**Inflation Trend**) に対するセンチメント判定モデル、下段の 3 つの表が労働市場 (**Labor Market**) に対するセンチメント判定モデルの結果をまとめている。(IT-a), (LM-a) は GPT-4 Turbo に、(IT-b), (LM-b) はファインチューニングされた RoBERTa に対応する混同行列を表しており、(IT-c), (LM-c) は両モデルによる判定の類似性をまとめている。GPT-4 Turbo およびファインチューニングされた RoBERTa の精度は物価動向についてが 86.6%, 83.9%, 労働市場についてが 81.3%, 81.3% となる。

⁵ ただし、ファインチューニングに用いたセンテンスは除いた。

5 センチメントインデックスの構成

本章ではファインチューニングされた RoBERTa によるセンチメントの判定結果からセンチメントインデックスを構成する方法を詳説する。本研究では (Thorsrud, 2020) (Filippou et al., 2024) などの先行研究に倣い、ディフュージョン・インデックスと同様の定義によってインデックスを構成することを提案する。すなわち、第 k 日に報じられたニュースから抽出されたセンテンスのうち、ポジティブと判定されたセンテンスの数 P_k 、ネガティブと判定されたセンテンスの数 N_k としたとき、両者の差を両者の和で除した値を指数値とする。ただし、指数が過度に振動的になることを防ぐべく、第 t 日の指数値 I_t の計算には、第 t 日を含む直近 τ 日間に報じられたニュースから得られたセンテンスについてセンチメントの判定結果をカウントすることとする。以上の定義は、式 1 のように定式化される。

$$I_t(\tau) \stackrel{\text{def}}{=} \frac{\sum_{k \in \{t, t-1, \dots, t-\tau+1\}} (P_k - N_k)}{\sum_{k \in \{t, t-1, \dots, t-\tau+1\}} (P_k + N_k)} \quad (1)$$

続いて、休日の取り扱いについて詳説する。土日や祝日など、米国債券市場が休場となる場合は、当該日に報じられるニュースの数が少ないことを勘案し、当該休場日に報じられたニュースを翌営業日に報じられたものとしてカウントする (図 5)。

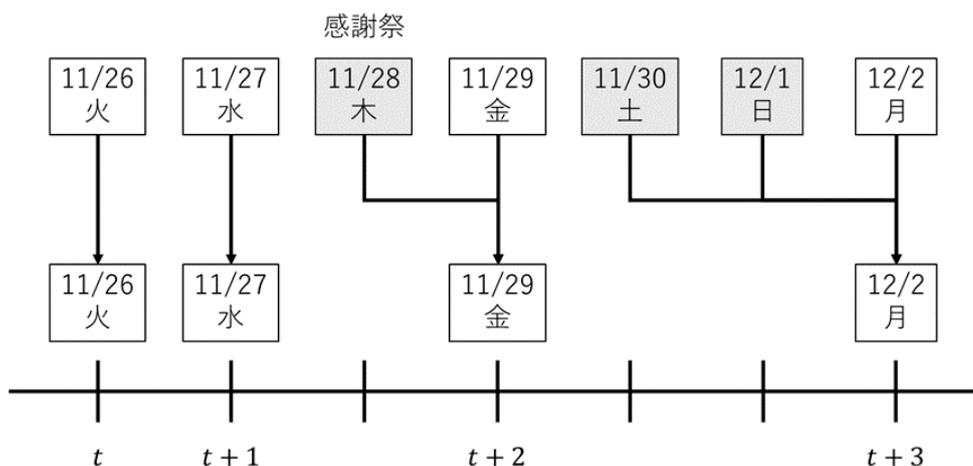


図5: センチメントインデックスを構成する際の休日の取り扱い。米国債券市場が休場となる場合は当該休場日に報じられたニュースを翌営業日に報じられたものとしてカウントする。

カウントの窓幅を $\tau = 20$ と定めた場合の物価動向トピックおよび労働市場トピックに関するセンチメントインデックスの推移をそれぞれ図 6, 7 に示す。ただし、背景色がグレーに塗りつぶされている期間はそれぞれ CPI⁶の前年同月比が 1% 未満となる期間と、米国経済がリセッション⁷と判定されている期間を表す。物価動向トピックに関するセンチメントインデックスはインフレ率が 1% 未満に減速する局面においてその水準を

⁶ 米労働統計局 (Bureau of Labor Statistics) が算出、公表する米国の消費者物価指数。
<https://www.bls.gov/news.release/cpi.toc.htm>

⁷ 全米経済研究所 (National Bureau of Economic Research) が定める景気後退期。
<https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions>

切り下げしており、物価動向を補足していることが伺える。また、労働市場トピックに関するセンチメントインデックスについてもリセッション期間においてその水準を急激に切り下げしており、景気後退に伴う労働市場の緩みを補足していることが伺える。

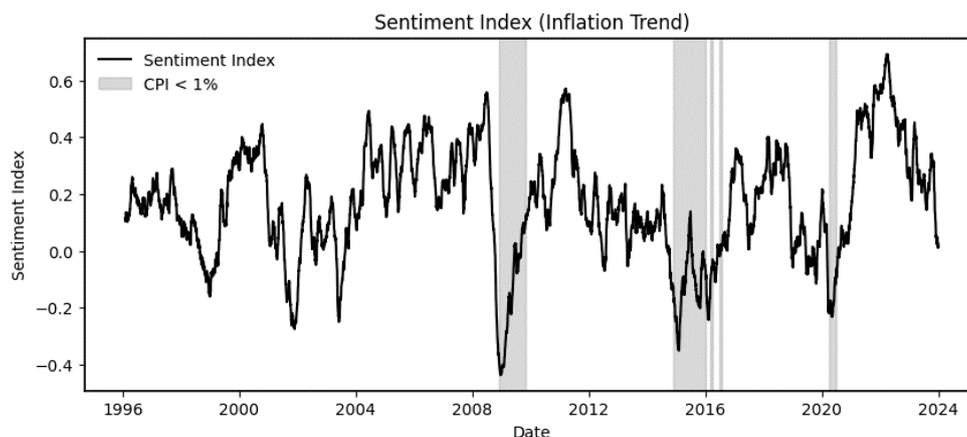


図6: 物価動向トピックに関するセンチメントインデックスの推移。背景色がグレーに塗りつぶされている期間は CPI の前年同月比が 1% 未満となる期間を表す。

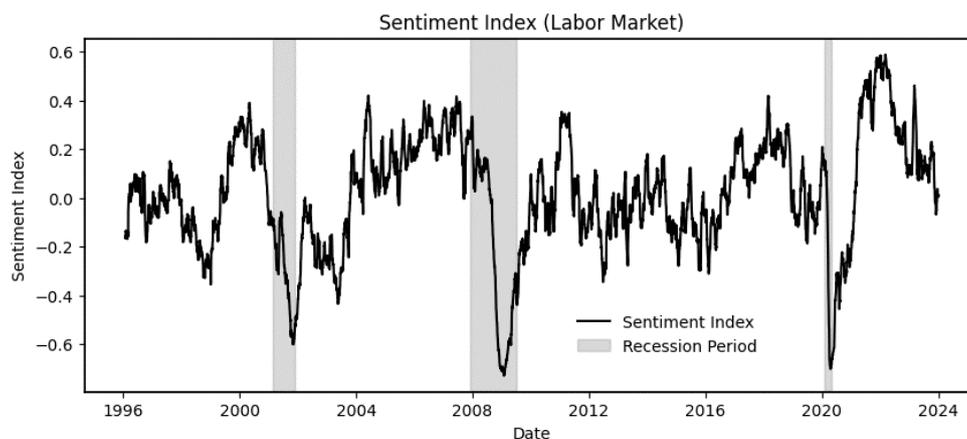


図7: 労働市場トピックに関するセンチメントインデックスの推移。背景色がグレーに塗りつぶされている期間は NBER により米国経済がリセッションと判定されている期間を表す。

6 センチメントインデックスの評価

本章では構成されたセンチメントインデックスの有用性を評価すべく執り行った 3 つの実証分析について詳説する。

6.1 景況感指数との連動性

1つ目の分析では構成したセンチメントインデックスと、米供給管理協会(Institute for Supply Management)が算出、公表する ISM 製造業景況感指数⁸との連動性を確認する。

ISM 製造業景況感指数（以下、ISM 指数）は米国で製造業を営む 300 以上の企業に対するアンケート調査の結果を要約した指数である。アンケートは生産、新規受注、在庫など 10 の項目から構成され、各アンケートの結果を要約したサブ指数が存在する。ISM 指数は経済指標の中でも市場参加者からの注目度が高く、GDP や株価に対して先行性があるとする見方もある (Banerjee and Marcellino, 2006) (Schmeling, 2009)。加えて、ISM 指数は米産業審議会 (The Conference Board) が算出、公表する景気先行指数 (The Conference Board Leading Economic Index)⁹の入力変数として採用されるなど、その有用性が広く認められている。

図 8は物価動向トピックに関するセンチメントインデックスおよび原材料価格の動向に関するアンケートから構成された ISM 支払価格指数（以下、ISM-price）の推移を示したものである。2003 年、2023 年に一時的に乖離する場面があるものの、多くの期間において両指数の連動性は極めて高いことが見受けられる。また、図 9は両指数を散布図に示した¹⁰ものである。両指数の相関係数は 76.9% となり、高い連動性を示していることが分かる。

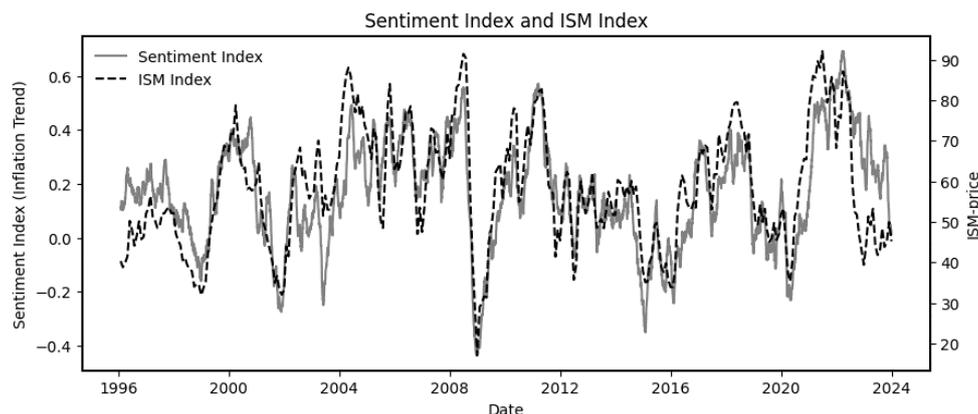


図8: 物価動向トピックに関するセンチメントインデックスおよび原材料価格の動向に関するアンケートから構成された ISM 指数の推移

図 10は労働市場トピックに関するセンチメントインデックスおよび各企業における雇用者数の動向に関するアンケートから構成された ISM 雇用者数指数（以下、ISM-labor）の推移を示したものである。2009 年から 2016 年にかけて一時的にセンチメントインデックスが ISM-labor の上方に位置する期間が存在するものの、多くの期間において両指数の連動性は極めて高いことが見受けられる。また、図 11は両指数を散布図に示したものである。両指数の相関係数は 63.7% となり、高い連動性を示していることが分かる。

⁸ <https://www.ismworld.org/supply-management-news-and-reports/reports/ism-report-on-business/>

⁹ <https://www.conference-board.org/topics/us-leading-indicators>

¹⁰ センチメントインデックスの値には ISM 指数が発表される日の前営業日の値を採用した。

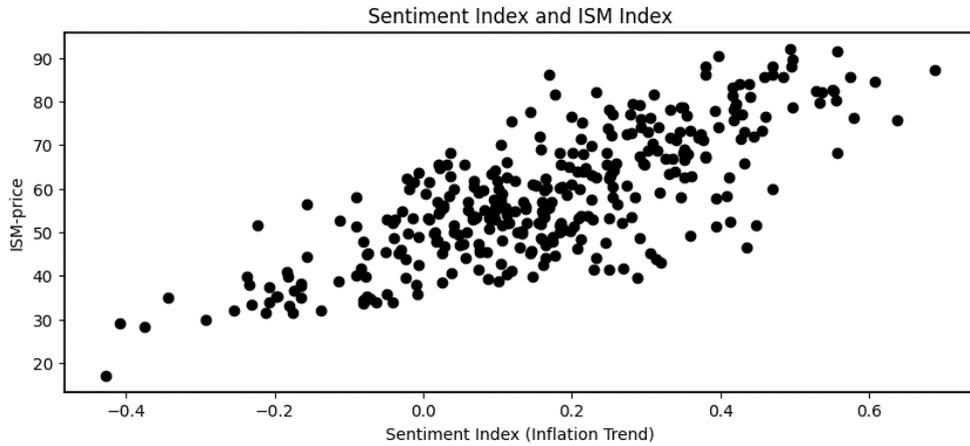


図9: 物価動向トピックに関するセンチメントインデックスおよび原材料価格の動向に関するアンケートから構成された ISM 指数. 両指数の相関係数は 76.9% となり, 高い連動性を示していることが分かる.

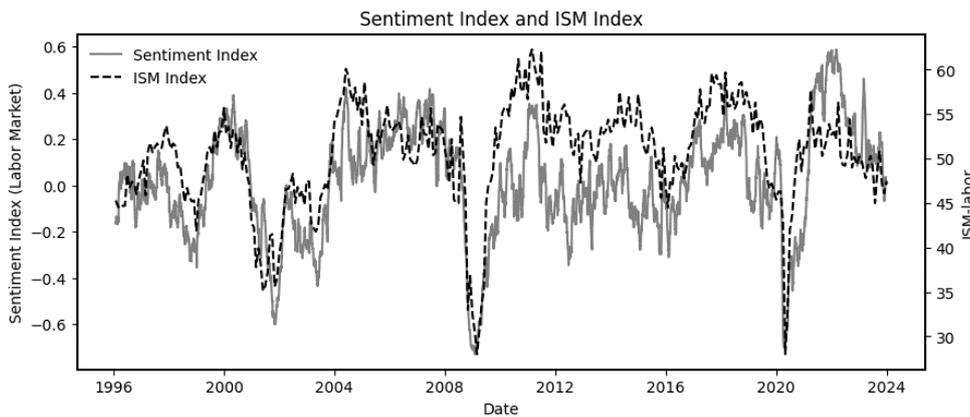


図10: 労働市場トピックに関するセンチメントインデックスおよび各企業における雇用者数の動向に関するアンケートから構成された ISM 指数の推移

6.2 ISM 製造業景況感指数に対する予測力の確認

2つ目の分析では構成したセンチメントインデックスが ISM 指数の予測因子となり得るかを検証する. 特に, 物価動向に関するセンチメントインデックスが ISM-price に対して予測力を有するか否かを検証する.

予測力の有無は特徴量選択の手続きを介して判断する. すなわち, ISM-price に関する L1 正則化付きの線形予測モデル (Zou, 2006) を学習するとき, 正則化を強めてもセンチメントインデックスにかかる回帰係数が他の説明変数と比して 0 に収束しにくいことを確認する.

ISM-price は米国経済がリセッションに陥るタイミングで著しく低下する性質を持つため, 損失関数として MSE (Mean Squared Error) を採用すると, その鋭敏性からたとえ平常時に予測力を持たなくとも, ISM-price が著しく低下するタイミングで大きく変動する説明変数が特徴量として選択されてしまう恐れがある. そこ

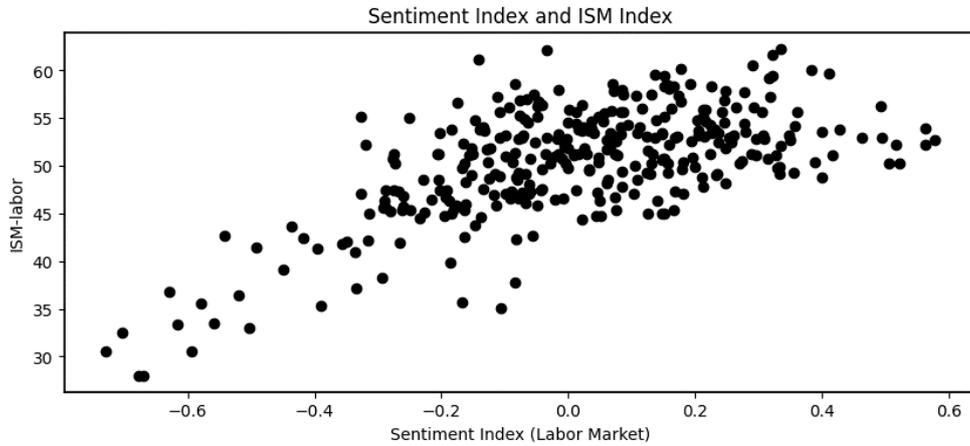


図11: 労働市場トピックに関するセンチメントインデックスおよび各企業における雇用者数の動向に関するアンケートから構成された ISM 指数. 両指数の相関係数は 63.7% となり, 高い連動性を示していることが分かる.

で, こうした鋭敏性の問題を緩和すべく, 損失関数には MAE (Mean Absolute Error) を採用する¹¹. 従って, L1 正則化項と併せて, 本予測モデルの損失関数は式 2 のようになる. ただし, $\lambda \in \mathbb{R}_{\geq 0}$ は正則化の強さをコントロールするチューニングパラメータである.

$$\sum_t |y_t - \alpha - \beta^\top x_t| + \lambda \cdot \|\beta\|_1 \quad (2)$$

次に, 本分析で採用した説明変数を記す. 本分析で採用した説明変数は表 2 にまとめられる. 本分析で採用した説明変数は 4 種に大別される. 1 種目の説明変数はセンチメントインデックスである. 物価動向トピックに関するセンチメントインデックスについて, その水準と 20 営業日変化を採用する. ただし, ISM 指数の調査期間と平仄を合わせるべく, 毎月 20 日におけるセンチメントインデックスの値をその水準として採用した. 2 種目の説明変数は ISM-price 自身である. ただし, 予測時点で知りえない情報を用いないよう, 第 t 月の ISM-price を予測する際には第 $t-1$ 月までのデータを用いる. ISM-price はその水準のみならず, 第 $t-1$ 月までに至る 1 ヶ月変化および 3 ヶ月変化も採用する. 3 種目の説明変数はニューヨーク連邦準備銀行およびフィラデルフィア連邦準備銀行が算出, 公表する景況感指数である. これらの景況感指数はそれぞれ Empire State Manufacturing Survey¹², Manufacturing Business Outlook Survey¹³ なる名称であり, ISM 指数と同じく製造業を営む企業に対するアンケート調査の結果を要約した指数である. ISM 指数との相違点としては, 調査対象となる企業の所在地が全米ではなくそれぞれの地区連銀が管轄する地区内に限られるという点が挙げられる. 加えて, 両景況感指数は第 t 月の調査結果に基づき構成される指数が第 t 月中に公表されるという特徴を持つ. 一方, ISM 指数は第 t 月の調査結果に基づき構成される指数が第 $t+1$ 月の第 1 営業日に公表されるため, 第 t 月の ISM 指数を予測する説明変数として両連銀が公表する当該景況感指数における第 t 月の指数値を採用することができる. 両地区連銀の景況感指数のうち, 支払い価格に関するアンケート結果から

¹¹ 中央値回帰モデルを採用することと同義となる.

¹² https://www.newyorkfed.org/survey/empire/empiresurvey_overview

¹³ <https://www.philadelphiafed.org/surveys-and-data/regional-economic-analysis/manufacturing-business-outlook-survey>

構成される指数について、その水準、1ヶ月変化、3ヶ月変化を採用する。また、両地区連銀の景況感指数は前月から当月に至る変化のみならず、将来6ヶ月における見通しについてもアンケートを執り行っており、本予測モデルにはこれらのアンケートから構成される指数も説明変数として採用する。4種類の説明変数は日次の価格指数であり、Baltic Dry Index と Goldman Sachs Commodity Index を採用する。Baltic Dry Index はばら積み船の運賃を、Goldman Sachs Commodity Index は原油や銅といった様々な商品の価格を測る指数である。Goldman Sachs Commodity Index は多様な商品の総合的な振る舞いを測る Composite 指数の他にも、エネルギーや産業用金属といったセクター別の指数が用意されており、本予測モデルではこれらを全て採用する。各価格指数について、その20営業日変化率および65営業日変化率を説明変数に採用する。以上より、採用した説明変数は表2に示す31系列となる。

表2: ISM-price の予測モデルにおいて採用した説明変数のリスト

指数	本稿で用いる略称	採用指数	前処理
センチメントインデックス	Index	物価動向トピック	水準 20 営業日変化
ISM 製造業景況感指数の前月値	ISM	支払価格指数	水準 1 ヶ月変化 3 ヶ月変化
Empire State Manufacturing Survey	NY	支払価格指数 (実績) 支払価格指数 (6 ヶ月先予測)	水準 1 ヶ月変化 3 ヶ月変化
Manufacturing Business Outlook Survey	Philly	支払価格指数 (実績) 支払価格指数 (6 ヶ月先予測)	水準 1 ヶ月変化 3 ヶ月変化
Baltic Dry Index	Baltic	指数値	20 営業日変化率 65 営業日変化率
Goldman Sachs Commodity Index	Composite	GSCI	指数値 20 営業日変化率 65 営業日変化率
	Energy	EN	指数値 20 営業日変化率 65 営業日変化率
	Industrial Metal	IM	指数値 20 営業日変化率 65 営業日変化率
	Precious Metal	PM	指数値 20 営業日変化率 65 営業日変化率
	Agriculture	AG	指数値 20 営業日変化率 65 営業日変化率
	Livestock	LV	指数値 20 営業日変化率 65 営業日変化率

続いて、採用した31系列の説明変数および、被説明変数である ISM-price に対して施した前処理について詳説する。採用した説明変数のうち、センチメントインデックス、ISM-price、Empire State Manufacturing Survey、Manufacturing Business Outlook Survey はそれぞれ -1 から 1, 0 から 100, -100 から 100, -100 から 100 の間の値をとる。一方、本稿では予測モデルとして式2に示したような線形回帰モデルを採用しているため、これらの系列はロジット関数や逆双曲線正接関数を用いて実数全体へ変換する。これらの系列は水準のみならず、その20営業日変化や1ヶ月変化なども説明変数として採用しているため、こうした前処理につい

でも実数全体へ変換した系列から変化を計算する。さらに、本予測モデルは L1 正則化が施されているため、各説明変数が正則化項に及ぼす寄与を均一化すべく、各説明変数はその分散が 1 となるように正規化を施す (Zou, 2006)。

解パス、すなわち、正則化を強めていった際の各説明変数にかかる回帰係数の推移は図 12 のようになる。ただし、正則化を強めるにつれ、各説明変数にかかる回帰係数は 0 へ収束するか否かによらず、その絶対値は縮小されるため、式 2 を最小化した結果得られた回帰係数 β^* をそのまま解パスとして描画すると、正則化を強めた際にどの説明変数が非 0 の回帰係数を持っているかを視認するのが難しくなる。そこで、図 12 では、変数選択の様子を確認しやすくするため、説明力が最大となるよう最適な回帰係数 β^* を一律に定数倍した値を描画している。具体的には、次式の c に関する最小解 $c^* \in \mathbb{R}_{>0}$ を用いて、 $c^* \cdot \beta^*$ を描画する。

$$\sum_t \left| y_t - \alpha^* - (c \cdot \beta^*)^\top x_t \right| \quad (3)$$

また、ISM の前月値にかかる回帰係数は正則化の強さによらず大きな値となるため、その他の説明変数にかかる回帰係数の振る舞いを確認しやすくすべく、描画していない。

$\lambda = 10$ と設定したモデルにおいては 31 系列の説明変数のうち、本稿で提案したセンチメントインデックスを含む 5 つが採用されており、センチメントインデックスの予測因子としての有用性を示すものとなっている。

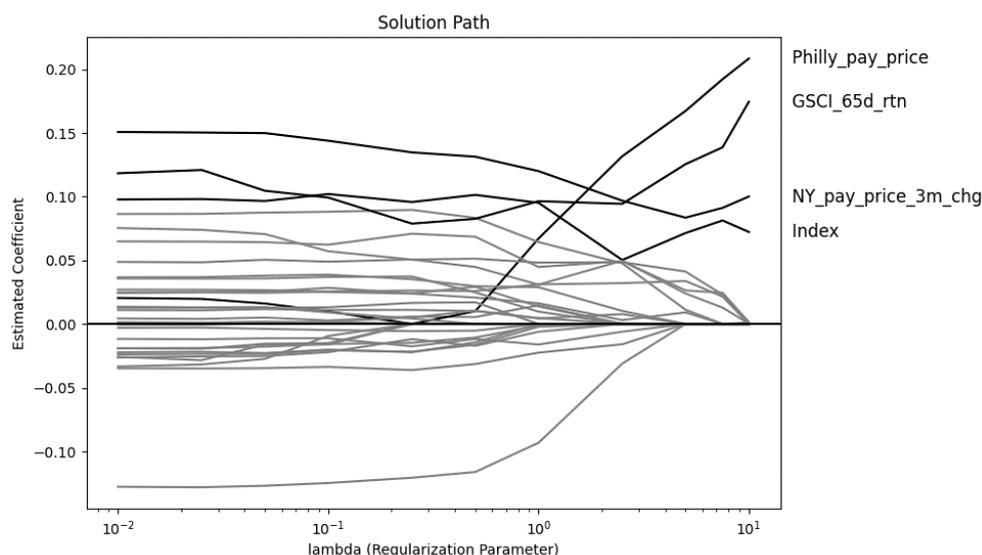


図12: ISM 製造業景況感指数のサブ指数である ISM 支払価格指数の予測モデルの解パス。ただし、ISM 支払価格指数の前月値にかかる回帰係数については描画していない。正則化を強めた結果、特徴量としては ISM 支払価格指数の前月値、Manufacturing Business Outlook Survey の当月値、Goldman Sachs Commodity Index の 65 営業日変化、Empire State Manufacturing Survey の 3 ヶ月変化、そして本研究で構成したセンチメントインデックスが採用された。これはセンチメントインデックスが ISM 支払価格指数の予測因子として有用であることを示す結果となっている。

6.3 リバーサル相場の検出

3つ目の分析では構成したセンチメントインデックスが債券市場における相場の検出に有用であることを示す。

ニュースデータのセンチメントと国債市場との関係を調べた文献としては (Erlwein-Sayer, 2018), (Beetsma et al., 2013), (Perico Ortiz, 2023) などが存在する。中でも (Erlwein-Sayer, 2018) は国債市場の相場、すなわちレジームを推定している点が他の2文献にはない特徴となっている。具体的には、ソブリン債スプレッドと、発行体である2国のマクロ経済に関して報じているニュースの件数とのローリング相関を用いて相場の変化を検出している。本稿では (Erlwein-Sayer, 2018) に倣い、構成したセンチメントインデックスと米国債利回りとのローリング相関によってレジームを定義し、将来の金利変化に関する示唆を探る。分析対象とする米国債利回りには、米連邦準備制度が算出する Treasury Nominal and Inflation-Indexed Constant Maturity Series¹⁴ なる系列群のうち、米国債10年利回り系列を用いる。

第6.1節で確認したように、センチメントインデックスはISM製造業景況感指数との連動性が高いため、実体経済の強さを定量化すると考えられる。他方、流動性プレミアム仮説によれば、国債の利回りは将来にわたる政策金利の期待値の平均と、リスク・プレミアムの和として定式化される (Ang, 2014)。さらに、テイラー・ルール (Taylor, 1993) の議論に従えば、政策金利はインフレや生産といった実体経済に依存するため、前述の流動性プレミアム仮説とあわせて考えれば、実体経済の強さやその予測を国債利回りの変動要因としてみなすことができよう (Smith and Taylor, 2009)。以上の議論を踏まえると、実体経済の強さを表すと考えられるセンチメントインデックスの変化は国債利回りの変化を説明することが期待される。実際、図13に示すように、センチメントインデックスの5営業日変化と米国債利回りの5営業日変化の、窓幅を125営業日としたローリング相関は最大で62.7%、平均で23.4%となり、センチメントインデックスが米国債利回り変化をよく説明していることが確認される。逆に、このローリング相関が低位となるレジームは、足元の実体経済ではなく、その将来予測やタム・プレミアムといった、直接観測することができないファクターが主たる利回りの変動要因となっていると考えられる。

以下では、当該レジームを“低相関レジーム”と呼称し、低相関レジーム下における利回り変化の傾向を探る。ただし、ローリング相関が低位であるとは、物価動向および労働市場に関するセンチメントインデックスと利回りのローリング相関がともに10%未満であるものとして定義した。前述のような直接観測することができないファクターによって利回りが急激に変化した際には、その不確実性ゆえ、その後平均回帰的な揺り戻しが発生しやすい、すなわち、リバーサルが発生しやすいと考えられる。この仮説を検証するため、ローリング相関が低位であるか否かという観点での局面分割に加え、以下に掲げる2つの局面分割を考える。

1つ目の分割は直近20営業日で利回りが15bp以上変化したか否かという観点に基づく。これは、前述の仮説における「利回りが急激に変化する」という条件に対応する。2つ目の分割は直近の利回り変化とセンチメントインデックス変化が同符号であるか異符号であるかという観点に基づく。これは、前述の仮説における「実体経済を代理するセンチメントインデックス以外のファクターが利回り変化の主たる変動要因となっているか」を判定している。

表3は前述の3つ観点で局面を分割したとき、各局面における将来20営業日の“リバーサル傾向”の平均値を示したものである。ただし、リバーサル傾向は将来20営業日の利回り変化に、直近20営業日利回り変化の

¹⁴ <https://www.federalreserve.gov/releases/h15/>

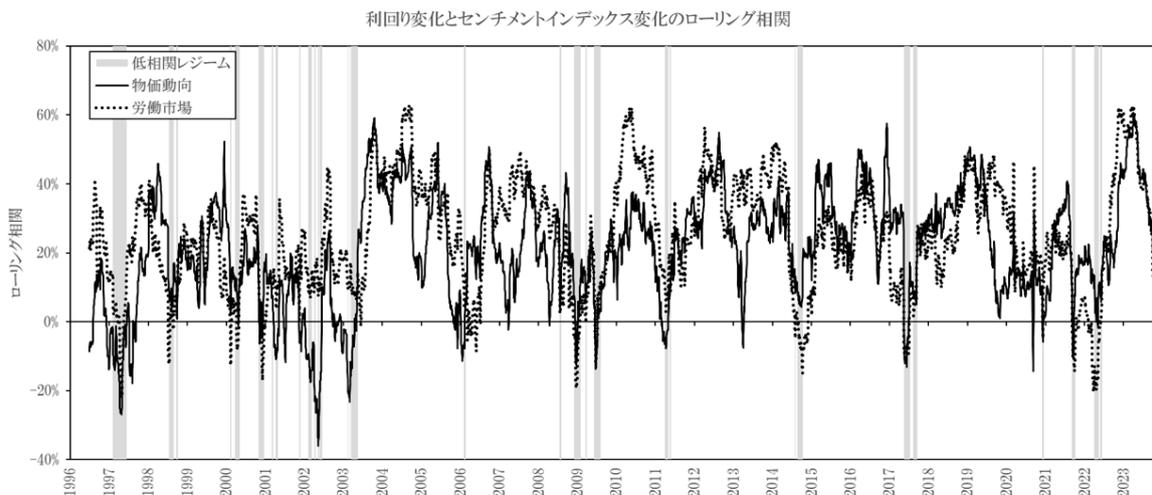


図13: センチメントインデックス変化と米国債利回り変化のローリング相関の推移。ローリング相関は最大で62.7%を示すなど、ニュースデータから構成されたセンチメントインデックスは米国債利回りの変化をよく説明する。

逆符号を乗じたものとして定義した¹⁵。先に掲げた仮説の通り、センチメントインデックスが代理すると考えられる実体経済以外のファクターによって利回りが急激に変化した局面においては、リバーサルの傾向がみられる。

表3: 各局面におけるリバーサル傾向。リバーサル傾向は将来 20 営業日の利回り変化に直近 20 営業日利回り変化の逆符号を乗じたものとして定義される。直接観測することができないファクターによって利回りが急激に変化した際には、その後にリバーサルが観測される傾向がみられる。

直近の利回り変化 直近のセンチメント変化と利回り変化	大きい		
	異符号	同符号	小さい
通常レジームにおけるリバーサル傾向 (bp)	-0.1	-1.8	0.1
低相関レジームにおけるリバーサル傾向 (bp)	9.9	2.8	0.8

7 結論

本稿では Reuters ニュースから物価動向および労働市場に対するセンチメントを捉えるセンチメントインデックスを構成した。センチメントの判定においては大規模言語モデルを用いることで、構文や文脈を勘案したセンチメント判定を行った。構成されたセンチメントインデックスはリセッションやインフレ率の減速に呼応した振る舞いを示すほか、重要な経済指標である ISM 製造業景況感指数と高い連動性を示した。また、物価動向トピックに関するセンチメントインデックスは ISM 製造業景況感指数のサブ指数である ISM 支払価格

¹⁵ 定義より、将来 20 営業日の利回り変化が直近 20 営業日の利回り変化と逆方向であるとき、リバーサル傾向は正となる。

指数の予測因子として有用であることを示した。さらに、センチメントインデックスは米国債利回りの変化をよく説明し、米国債利回りがリバーサル傾向を示しやすいレジームを検出することを示した。

参考文献

- Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat et al. (2023) “Gpt-4 technical report,” *arXiv preprint arXiv:2303.08774*.
- Allen, David E, Michael McAleer, and Abhay K Singh (2019) “Daily market news sentiment and stock prices,” *Applied Economics*, Vol. 51, No. 30, pp. 3212–3235.
- Ang, Andrew (2014) *Asset management: A systematic approach to factor investing*: Oxford University Press.
- Araci, D (2019) “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models,” *arXiv preprint arXiv:1908.10063*.
- Ashwin, Julian, Eleni Kalamara, and Lorena Saiz (2021) “Nowcasting euro area GDP with news sentiment: a tale of two crises,” *Journal of Applied Econometrics*.
- Audrino, Francesco and Eric A Offner (2024) “The impact of macroeconomic news sentiment on interest rates,” *International Review of Financial Analysis*, Vol. 94, p. 103293.
- Banerjee, Anindya and Massimiliano Marcellino (2006) “Are there any reliable leading indicators for US inflation and GDP growth?” *International Journal of Forecasting*, Vol. 22, No. 1, pp. 137–151.
- Barbaglia, Luca, Sergio Consoli, and Sebastiano Manzan (2023) “Forecasting with economic news,” *Journal of Business & Economic Statistics*, Vol. 41, No. 3, pp. 708–719.
- Beetsma, Roel, Massimo Giuliodori, Frank De Jong, and Daniel Widijanto (2013) “Spread the news: The impact of news on the European sovereign bond markets during the crisis,” *Journal of International Money and Finance*, Vol. 34, pp. 83–101.
- Born, Benjamin, Jonas Dovern, and Zeno Enders (2023) “Expectation dispersion, uncertainty, and the reaction to news,” *European Economic Review*, Vol. 154, p. 104440.
- Brown, Tom B (2020) “Language models are few-shot learners,” *arXiv preprint arXiv:2005.14165*.
- Devlin, Jacob (2018) “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*.
- Erlwein-Sayer, Christina (2018) “Macroeconomic news sentiment: enhanced risk assessment for sovereign bonds,” *Risks*, Vol. 6, No. 4, p. 141.
- Filippou, Ilias, Christian Garciga, James Mitchell, and My T Nguyen (2024) “Regional Economic Sentiment: Constructing Quantitative Estimates from the Beige Book and Testing Their Ability to Forecast Recessions,” *Economic Commentary*, No. 2024-08.
- Hutto, Clayton and Eric Gilbert (2014) “Vader: A parsimonious rule-based model for sentiment analysis of social media text,” in *Proceedings of the international AAAI conference on web and social media*, Vol. 8, pp. 216–225.
- Lee, Jean, Nicholas Stevens, Soyeon Caren Han, and Minseok Song (2024) “A survey of large language models in finance (finllms),” *arXiv preprint arXiv:2402.02315*.
- Liu, Yinhan (2019) “Roberta: A robustly optimized bert pretraining approach,” *arXiv preprint arXiv:1907.11692*.

- Loughran, Tim and Bill McDonald (2011) “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks,” *The Journal of finance*, Vol. 66, No. 1, pp. 35–65.
- Maia, Macedo, Siegfried Handschuh, André Freitas, Brian Davis, Ross McDermott, Manel Zarrouk, and Alexandra Balahur (2018) “Www’18 open challenge: financial opinion mining and question answering,” in *Companion proceedings of the the web conference 2018*, pp. 1941–1942.
- Malo, Pekka, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala (2014) “Good debt or bad debt: Detecting semantic orientations in economic texts,” *Journal of the Association for Information Science and Technology*, Vol. 65, No. 4, pp. 782–796.
- Nguyen, Kim, Gianni La Cava et al. (2020) “News Sentiment and the Economy| Bulletin–June 2020.”
- Perico Ortiz, Daniel (2023) “Inflation news coverage, expectations and risk premium,” Technical report, FAU Discussion Papers in Economics.
- Schmeling, Maik (2009) “Investor sentiment and stock returns: Some international evidence,” *Journal of empirical finance*, Vol. 16, No. 3, pp. 394–408.
- Sebastiani, Fabrizio and Andrea Esuli (2006) “Sentiwordnet: A publicly available lexical resource for opinion mining,” in *Proceedings of the 5th international conference on language resources and evaluation*, pp. 417–422, European Language Resources Association (ELRA) Genoa, Italy.
- Seki, Kazuhiro, Yusuke Ikuta, and Yoichi Matsubayashi (2022) “News-based business sentiment and its properties as an economic index,” *Information Processing & Management*, Vol. 59, No. 2, p. 102795.
- Shapiro, Adam Hale, Moritz Sudhof, and Daniel J Wilson (2022) “Measuring news sentiment,” *Journal of econometrics*, Vol. 228, No. 2, pp. 221–243.
- Smales, Lee A (2014) “News sentiment in the gold futures market,” *Journal of Banking & Finance*, Vol. 49, pp. 275–286.
- Smith, Josephine M and John B Taylor (2009) “The term structure of policy rules,” *Journal of Monetary Economics*, Vol. 56, No. 7, pp. 907–917.
- Taylor, John B (1993) “Discretion versus policy rules in practice,” in *Carnegie-Rochester conference series on public policy*, Vol. 39, pp. 195–214, Elsevier.
- Thorsrud, Leif Anders (2020) “Words are the new numbers: A newsy coincident index of the business cycle,” *Journal of Business & Economic Statistics*, Vol. 38, No. 2, pp. 393–409.
- Zou, Hui (2006) “The adaptive lasso and its oracle properties,” *Journal of the American statistical association*, Vol. 101, No. 476, pp. 1418–1429.