

大規模言語モデルを用いた金融テキストに対する推論ベースの極性付与

高野 海斗¹ 中川 慧¹

¹ 野村アセットマネジメント株式会社

{k-takano,k-nakagawa}@nomura-am.co.jp

概要

金融テキストマイニングでは、極性付与は重要なタスクであり、大規模言語モデル (LLM) の進歩により、適用が広がっている。一方、従来の分析対象は決算短信やニュースが中心であり、極性がテキストに直接記載されている、推論が不要なタスクである。しかし、実務上の投資判断を鑑みると、直接的に極性が記載されていないテキスト情報に対して、バックグラウンドを踏まえつつ極性を推論することが求められる。そこで本研究では、LLM の推論能力を活用し、業種に関する情報を入力し、ある特定の重大イベントが当該業種に与える影響を推論させる極性付与タスクに取り組む。そして、その出力が実務的にどの程度有用であるか検証する。

1 はじめに

金融に関連する膨大なテキストから業務効率化や投資意思決定に有益な情報を識別・抽出し、その情報と金融市場の関係性を発見・分析する技術は金融テキストマイニングと呼ばれている [1]。

金融テキストマイニングにおいて、特に重要なタスクの一つに、極性 (センチメント) の付与がある。極性を付与する方法は辞書ベースの手法と機械学習ベースの手法に大別できる (2章参照)。後者の手法に関連し、大規模言語モデル (LLM) の技術進展、特に OpenAI が開発した ChatGPT の API を介した利用が容易に可能となった。それに伴って、これまで困難であった様々なタスクに対して実務で使用できる高精度の結果をもたらし、その応用領域が拡大している [2, 3, 4]。

極性付与のタスクにおいても、Lopez ら [5] は、ChatGPT はニュースのヘッドラインの極性を適切に評価でき、その後の株式リターンに有意に正の相関関係があることを実証した。しかし、GPT-1、

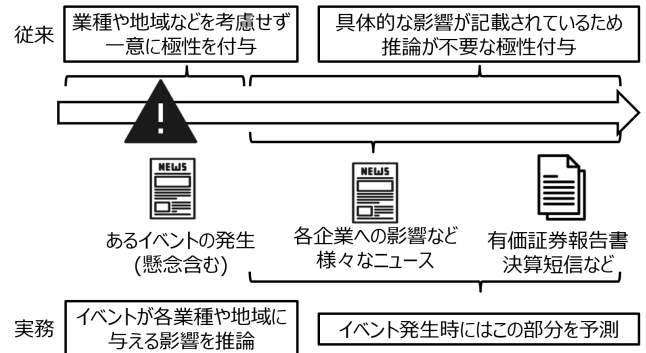


図 1 従来の極性付与の設定 (上図) と実務的な極性付与 (下図)

GPT-2, BERT などのモデルでは相関関係が見られないこともあわせて報告されている。また、Pelster ら [6] は、アウトオブサンプルの検証においても、ChatGPT の決算発表の極性評価と将来の業績発表および株式リターンとの間に正の相関関係があることを実証した。

一方で、金融テキストマイニングにおける極性付与は、日々の企業業績に関するニュースや決算短信、有価証券報告書に記載されている文を対象にしている。これらデータに対する極性分析は、事後的に過去の状況を分析するには非常に有力である。ただし、ある企業についてのテキストの情報がポジティブまたはネガティブかは直接記載されているものが大多数であるため、推論が不要な極性付与タスクであるといえる。しかしながら、実際の資産運用における投資判断を鑑みると、運用者は将来または仮定したイベントに対して推論を行い、業績あるいは株価に対してポジティブかネガティブかを判断し意思決定を行っている (図 1 参照)。

この点を考慮すると、直接的な業績への影響などが記載されていない、あるイベントに関して記載された文や文章に対して、業種特性などを踏まえてポジティブなのか、ネガティブなのかを推論することができるかという極性付与タスクに取り組んでいく

必要がある。例えば、Covid-19 が世界的に流行する前 (2020 年 1 月) の段階で、感染力の強いウィルスが世界的に大流行した場合に、各業種に対してどのような影響が考えられるかを、理由と合わせて極性を付与することが期待される。そこで本研究においては、LLM の推論能力を活用し、業種に関する情報を入力し、ある特定の重大イベントが当該業種に与える影響を推論させる極性付与タスクに取り組む。具体的には、LLM に対して、あるイベントが特定の業種においてどのような極性を持つのかをプロンプトとして入力し、その出力が実務的にどの程度有用であるか検証する (付録の図 2 参照)。

2 金融極性付与タスクのレビュー

前章で述べた通り、文に対して極性を付与する方法は辞書ベースの手法と機械学習ベースの手法に大別できる。辞書ベースの手法は、極性語とその極性値が組となった極性辞書に基づいて、文や文章に極性を付与する [7]。辞書ベースの利点は、辞書に基づいて極性が付与されることから、解釈が容易である点があげられる。辞書ベースの問題点は、単純な方法では文脈を考慮した極性付与ができない点である。例えば、「コロナ感染者数の増加 (2.312) に伴い、売上高は減少 (-1.998) いたしました。」のような文の極性は、0.314 (ポジティブ) となるが、明らかにネガティブな文である。例に挙げたように、何が増加したかによって、「増加」のスコアを変化させるべきだが、一般的な極性辞書は、そのような問題に対処できない。解決策として、「感染者数の増加 (-1.315)」、「売上の増加 (3.312)」のように表現を拡張した辞書を作成することが挙げられる [8]。しかし、全ての表現を辞書で作成することは困難であり、類似表現に対しても、辞書に登録されていなければ極性を付与できない問題がある。また、表現を拡張できたとしても、「見込んでいた売上減少はなかったものの、」のような否定表現などに対応できないことも問題点としてあげられる。その他にも、広く利用されている一般的な極性辞書に含まれるネガティブ表現が、金融の分野において、約 4 分の 3 も該当しないことを示した研究もある [9]。したがって、極性分析を行うためには、目的に応じて極性辞書を探したり、場合によっては自作する必要がある。

一方、機械学習ベースの手法は、文や文章と極性が紐づいたデータを用いて学習させたモデルによって極性を付与する。金融に特化した機械学習ベース

の極性付与モデルとしては、FinBERT¹⁾が有名である [10]。機械学習ベースの利点は、文脈を考慮した極性を付与できる点である。また、辞書ベースの手法は、辞書に含まれている語が文に含まれていないと極性が付与できないのに対して、機械学習ベースの手法によっては、類義語が含まれていれば極性の付与が可能である。問題点として、機械学習ベースの手法は、解釈性の問題が残るため [11]、解釈性に着目したモデルの研究も盛んに行われている [12]。

従来の金融テキストマイニングにおける極性付与は、該当企業にとって、ある出来事がポジティブまたはネガティブかの情報が直接記載されているものが大多数であり、推論が不要な極性付与タスクである。したがって、LLM が持つ推論能力を十分に活用できていないタスクであるといえる。

3 実験

3.1 実験設定

LLM に対して、あるイベントの極性を業種を踏まえた上で推論させ、その出力が実務的にどの程度有用であるか確認する。入力として与えるイベントは、未知のウィルスの世界的な流行、極端な円安、ロシアとウクライナの紛争とする。また、業種は、ゲーム、飲食、海運、観光、医薬品とする。使用するプロンプトは以下の通りである。なお、< >内は、前述したイベントや業種が入る。

入力に使用するプロンプト

<イベント>は、日本の<業種>業界にどのような影響を与えますか？

使用するモデルは、OpenAI が公開している「gpt-4-1106-preview」, 「gpt-4-0613」, 「gpt-3.5-turbo-0613」を使用する。以降、それぞれのモデルを gpt-4-1106, gpt-4-0613, gpt-3.5 と本論文では表記する。gpt-4-1106 は、2023 年 4 月までのデータを使用して学習が行われており、残り二つのモデルは、2021 年 9 月までのデータを使用して学習が行われている²⁾。なお、ハイパーパラメータは再現性のため、temperature のみ 0 に設定し、後はデフォルトの値を使用する。

評価に関しては、定性的な評価を行う。方法論が確立されていないため、定量的な評価は今後の課題

1) <https://huggingface.co/ProsusAI/finbert>

2) <https://platform.openai.com/docs/models>

表1 gpt-4-1106 の出力結果の集計

イベント	業種	posi	nega	neutral
未知のウィルスの世界的な流行	ゲーム	1	5	3
	飲食	0	5	3
	海運	0	7	2
	観光	0	6	1
	医薬	6	1	1
ロシアとウクライナの紛争	ゲーム	0	6	1
	飲食	0	5	1
	海運	0	7	0
	観光	0	6	0
	医薬	1	3	2
極端な円安	ゲーム	3	3	1
	飲食	2	3	1
	海運	2	4	0
	観光	3	2	2
	医薬	2	6	0

表2 gpt-4-0613 の出力結果の集計

イベント	業種	posi	nega	neutral
未知のウィルスの世界的な流行	ゲーム	2	4	0
	飲食	0	5	1
	海運	0	5	0
	観光	0	5	0
	医薬	3	2	0
ロシアとウクライナの紛争	ゲーム	0	4	1
	飲食	0	3	1
	海運	0	5	0
	観光	0	4	0
	医薬	1	3	0
極端な円安	ゲーム	3	2	0
	飲食	1	3	0
	海運	1	4	0
	観光	2	2	0
	医薬	1	3	0

とする。

3.2 実験結果

参考として、ロシアとウクライナの紛争が飲食業界に与える影響について、得られた出力をモデルごとに付録 A に示す。また、表 1, 表 2, 表 3 は、各モデルの出力結果に記載されていた影響を極性別に集計したものである。

4 考察

4.1 集計結果からの考察

表 1 から表 3 より、数の変動はあるものの、モデルによってポジティブな影響とネガティブな影響の大小関係が逆転することはなかった。

まず注目したいのは、ウィルスの流行に対して、多くの業界でネガティブな影響が優勢の中、医薬

表3 gpt-3.5 の出力結果の集計

イベント	業種	posi	nega	neutral
未知のウィルスの世界的な流行	ゲーム	1	4	0
	飲食	0	5	0
	海運	0	5	0
	観光	0	5	0
	医薬	2	1	2
ロシアとウクライナの紛争	ゲーム	0	4	0
	飲食	0	4	0
	海運	0	4	0
	観光	0	4	0
	医薬	0	4	0
極端な円安	ゲーム	4	1	0
	飲食	2	3	0
	海運	1	2	2
	観光	2	2	1
	医薬	1	3	1

業界はポジティブな影響の方が多く出力されている点である。例えば、FinBERT に「Global outbreak of unknown virus.」と入力すると、「**negative: 0.752, neutral: 0.230, positive: 0.018**」と出力される。しかしながら、本研究の主題通り、業界という背景によって極性は異なるべきであり、今回の結果はそれを顕著に示している。

次に、ウィルスの流行に対するゲーム業界の結果からわかる通り、ポジティブ、ネガティブの個数を用いた分析では不十分である。ゲーム業界は、コロナ流行による外出自粛の影響を受けて需要が急増し、多くの企業で収益が増加した。ゲーム需要の増加は各モデルの出力に含まれているものの、その他に懸念されるリスクや変化が多く記載されており、どの項目がどれだけのインパクトがあるのかを記載内容から考慮する必要がある。

4.2 モデルに着目した考察

表 1 から表 3 より、gpt-4-1106 は他のモデルが平均 5 つを下回る影響を出力するのに対して、平均 7 つを超える影響を出力した。これは、モデルのアップデートに伴い全体的に精度が低下するという黒木ら [3] や Chen ら [13] の指摘とは異なり、モデルのアップデートに伴い、より多様な出力が可能になっていることを示唆している。

次に、モデルを学習させているデータの学習期間に着目すると、付録 A の出力結果に示した通り、gpt-4-1106 は、「価格が上昇しました。」のように実際に起こった事実として出力していることがわかる。このような出力になった要因は、2022 年 2 月からロシアはウクライナへ本格的な軍事侵攻を開始

しており、gpt-4-1106 が 2023 年 4 月までのデータを使用して学習されているためである。したがって、2021 年 9 月までのデータを使用して学習された残り 2 つのモデルには、「しました。」のような表現が含まれず、「可能性があります。」という表現が多様されている。厳密なアウトオブサンプルの分析をするのであれば、未知のウィルスの世界的な流行に対する推論は、2019 年までのデータを用いて学習された LLM を用いるべきである。なぜなら、2020 年以降、コロナウィルスが世の中に与えた影響に言及するテキスト情報は多数存在しており、本論文で使用したモデルは、それらの情報を学習済みである。仮に 2019 年までのデータを用いて学習されたモデルを使用する場合には、出力の悪化が想定される。実際の運用を考えると、将来のデータを使用することは不可能であり、LLM を用いた分析は十分に注意が必要である。

より詳細にモデルごとの出力結果を見ると、gpt-3.5 の出力は「紛争」によって生じる一般論となっており、各国の国名を別の国名に置換しても、内容にほとんど矛盾が生じない出力である。それに対して、gpt-4-0613 は付録 A で太字で示した通り、「ウクライナは世界最大級の穀物輸出国」という記載があり、gpt-3.5 よりも優れた出力となっている。さらに、学習データの leak はあるものの、gpt-4-1106 は、「ロシアとウクライナは小麦やひまわり油などの農産物の大きな輸出国」、「ロシアは天然ガスや石油の大手輸出国」などの情報も出力に追加されている。本論文の検証結果から想定される現段階の LLM におけるシンプルな推論の限界は、実際に起こっていないことに関しては、「穀物価格の上昇」から食材のコストが増加することまで推論が可能であり、実際に起こったことがあることに関しては、「パンや麺類、揚げ物など」のように、その食材によって作られる食品の推論までが可能であった。実際に、ウクライナからの穀物輸出が減ったことで、間接的に大きな影響を与えている食材の一つに卵があるが³⁾、そこまでの推論は、実際に起こった出来事でもまだ推論することは難しい。しかし、gpt-3.5 に「卵の価格上昇要因を教えてください。」というプロンプトを与えると、出力される要因の一つに「飼料価格の上昇：鶏の飼料（主に穀物や大豆）の価格が上昇すると、養鶏業者の生産コストが増加します。」のような記述が含まれるため、1step ではな

3) 餌代の高騰が卵の価格上昇の大きな要因となっている。

く、複数 step で推論を行う方法に拡張することで、現状のモデルでも、より深い推論が実現できるかもしれない。

また、LLM において Hallucination[14]⁴⁾が問題となるが、gpt-3.5 は特に多く発生した。その中でも**極端な円安**においては、「円安により、日本のゲーム会社は海外からの人材を採用しやすくなる可能性があります。⁵⁾」や「円安により、海外企業の製品が日本市場で競争力を持つ可能性もあります。⁶⁾」のような、事実とは正反対の出力が顕著に生じていた。極性分析において、正反対の内容の出力は、極性が反転するため、Hallucination の発生しにくいモデルを使用することや、その発生を抑制する工夫が重要となる。

5 おわりに

本研究では、LLM の推論能力を活用し、業種に関する情報を入力し、ある特定の重大イベントが当該業種に与える影響を推論させる極性付与タスクに取り組んだ。出力結果からは、一部期待するような結果を得ることができたが、課題も山積みである。付録 A に示した結果の通り、過去に起こった出来事に対しての推論は、詳細な要素の出力も多く含まれるため、個人の投資家が思考の壁打ちをするツールとして有力であるが、ベテランの運用者やエコノミストからすると周知の事実がほとんどであるため、もう少しプロンプトや方法に工夫が必要である。推論能力を向上させるために、因果関係データセットの構築も重要な課題の 1 つである [15]。

また、4.1 節の中で医薬業界の結果を取り上げたが、実際に医薬業界の中でも、コロナウィルスに関連する取り組みを行い業績を伸ばした企業も存在すれば、在宅自粛などによって従来の風邪やインフルエンザの患者数が激減したことで、業績が落ち込んだ企業も存在する。したがって、より具体的な事業概要や過去の業績結果などをプロンプトに組み込むことで、個々の企業にどのような影響が推論がどの程度可能であるか検証を行いたい（付録の図 3 参照）。

4) 生成モデルにおいて、もっともらしいウソ (=事実とは異なる内容や文脈と無関係な内容) の出力が生成される現象。

5) 円安により、賃金が下がるため採用は難しくなる傾向にある。

6) 円安により、海外企業の製品は価格が上昇するため、海外企業の製品の競争力は低下する傾向にある。

参考文献

- [1] 中川慧, 伊藤友貴. 機械が読む英文開示. 企業会計, Vol. 75, No. 3, pp. 339–348, 2023.
- [2] Irene S Gabashvili. The impact and applications of chatgpt: a systematic review of literature reviews. **arXiv preprint arXiv:2305.18086**, 2023.
- [3] Yutaka Kuroki, Tomonori Manabe, and Kei Nakagawa. Fact or opinion?—essential value for financial results briefing. In **2023 14th International Congress on Advanced Applied Informatics (IIAI-AAI)**. IEEE, 2023.
- [4] 高野海斗, 中川慧, 藤本悠吾. Chatgpt を活用した運用報告書の市況コメントの自動生成. 人工知能学会第二種研究会資料, No. FIN-031, pp. 61–67, 2023.
- [5] Alejandro Lopez-Lira and Yuehua Tang. Can chatgpt forecast stock price movements? return predictability and large language models. **arXiv preprint arXiv:2304.07619**, 2023.
- [6] Matthias Pelster and Joel Val. Can chatgpt assist in picking stocks? **Finance Research Letters**, p. 104786, 2023.
- [7] Tomoki Ito, Hiroki Sakaji, Kota Tsubouchi, Kiyoshi Izumi, and Tatsuo Yamashita. Text-visualizing neural network model: understanding online financial textual data. In **Pacific-Asia Conference on Knowledge Discovery and Data Mining**, pp. 247–259, 2018.
- [8] 今井康太, 酒井浩之, 永並健吾, 稲垣真太郎. 債券市場における金融極性辞書の自動構築. 知能と情報, Vol. 34, No. 4, pp. 673–682, 2022.
- [9] Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. **The Journal of Finance**, Vol. 66, No. 1, pp. 35–65, 2011.
- [10] Dogu Araci. FinBERT: Financial sentiment analysis with pre-trained language models. **arXiv preprint arXiv:1908.10063**, 2019.
- [11] David Gunning and David William Aha. DARPA’s explainable artificial intelligence program. **AI Magazine**, Vol. 40, No. 2, pp. 44–58, 2019.
- [12] Tomoki Ito, Kota Tsubouchi, Hiroki Sakaji, Tatsuo Yamashita, and Kiyoshi Izumi. Word-level contextual sentiment analysis with interpretability. In **AAAI**, Vol. 34, pp. 4231–4238, 2020.
- [13] Lingjiao Chen, Matei Zaharia, and James Zou. How is chatgpt’s behavior changing over time? **arXiv preprint arXiv:2307.09009**, 2023.
- [14] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. **ACM Computing Surveys**, Vol. 55, No. 12, pp. 1–38, 2023.
- [15] 張辰聖子, 持橋大地, 小林一郎. 手がかり表現に基づく非論理的な言語推論の学習. 言語処理学会 第 29 回年次大会, pp. 494–499, 2023.

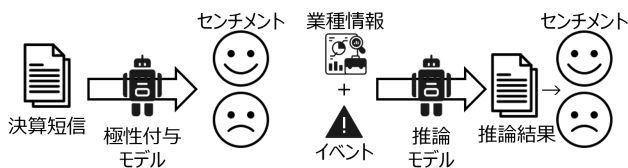


図2 従来の極性付与(左図)と本研究の極性付与(右図)

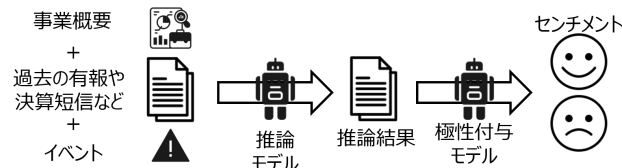


図3 個別企業の情報を加味した推論ベースの極性付与

A ロシアとウクライナの紛争が飲食業界に与える影響の推論結果

gpt-3.5

ロシアとウクライナの紛争は、日本の飲食業界に以下のような影響を与える可能性があります。

1. 食材の供給不足：ロシアやウクライナからの食材の輸入が制限される可能性があります。これにより、日本の飲食業界は特にロシアやウクライナからの食材に依存しているレストランや食品メーカーに影響を与える可能性があります。代替の供給源を見つける必要があるかもしれません。
2. 輸出市場の減少：ロシアやウクライナは日本の飲食業界にとって重要な輸出市場です。紛争が激化すると、両国の経済が停滞し、需要が減少する可能性があります。これにより、日本の飲食業界は輸出市場の減少に直面することになります。
3. 国際関係の影響：ロシアとウクライナの紛争は国際的な関係にも影響を与えます。日本は国際社会での立場を考慮しながら外交政策を運営しており、紛争による国際的な緊張が日本の飲食業界にも波及する可能性があります。
4. 消費者の意識の変化：紛争が報道されると、消費者の意識が変化する可能性があります。例えば、ロシアやウクライナと関連する食品やレストランへのボイコットが起こるかもしれません。これにより、日本の飲食業界においてロシアやウクライナとの関係が影響を受ける可能性があります。

以上のような要素が組み合わさり、日本の飲食業界にはロシアとウクライナの紛争が様々な影響を与える可能性があります。ただし、具体的な影響は紛争の進展や国際情勢の変化によって異なる可能性があります。

gpt-4-0613

ロシアとウクライナの紛争が日本の飲食業界に与える影響は、主に以下のようなものが考えられます。

1. 食材の輸入制限や価格変動：ロシアやウクライナから輸入している食材がある場合、紛争により輸出が制限されたり、価格が変動する可能性があります。特に**ウクライナは世界最大級の穀物輸出国**であり、紛争が長引くと穀物価格の上昇や供給不安が生じる可能性があります。
 2. 為替レートの変動：紛争によりロシアルーブルやウクライナ通貨の為替レートが変動すると、それらの国からの輸入コストが変わる可能性があります。これは飲食業界だけでなく、他の業界にも影響を及ぼす可能性があります。
 3. 消費者の意識の変化：紛争により消費者の意識が変わる可能性もあります。例えば、ロシアやウクライナの食材を使った料理を提供している店舗は、消費者の反ロシア感情や反ウクライナ感情により売上が下がる可能性があります。
 4. ロシアやウクライナの観光客の減少：紛争によりロシアやウクライナからの観光客が減少すると、特に彼らをターゲットにした飲食店の売上が下がる可能性があります。
- 以上のような影響が考えられますが、具体的な影響は紛争の進行状況やその後の国際的な対応によります。

gpt-4-1106

ロシアとウクライナの紛争は、世界的な経済に多くの影響を与えており、日本の飲食業界にも様々な形で影響を及ぼしています。以下はその影響のいくつかです：

1. 原材料の価格上昇：**ロシアとウクライナは小麦やひまわり油などの農産物の大きな輸出国**です。紛争によりこれらの商品の供給が不安定になり、国際市場での価格が上昇しました。これは、パンや麺類、揚げ物などを提供する飲食店にとってコスト増加の要因となります。
 2. エネルギー価格の上昇：**ロシアは天然ガスや石油の大手輸出国**であり、紛争によってエネルギー価格が上昇しました。これは、飲食店の運営に必要な電気やガスのコストを押し上げ、経営に影響を与える可能性があります。
 3. 物流コストの増加：紛争により、特定の航路や輸送ルートが使用できなくなる可能性があります。これにより、食材の輸入コストが上昇し、飲食店の仕入れ価格が高騰する可能性があります。
 4. 消費者の心理的影響：紛争による経済的不安定さや物価の上昇は、消費者の支出意欲に影響を与える可能性があります。不安定な経済状況の中で、消費者は外食に対して慎重になり、飲食業界の売上に影響を及ぼすかもしれません。
 5. メニューの変更や価格の見直し：原材料の価格上昇に対応するため、飲食店はメニューの内容を変更したり、価格を見直したりする必要が出てくるかもしれません。これにより、顧客の選択肢や価格感に影響を与えることとなります。
 6. サプライチェーンの再構築：長期的には、紛争によるサプライチェーンの不安定さを避けるために、飲食業界はより多様な供給源を確保する必要があるかもしれません。これには、国内産の食材へのシフトや、他の国からの輸入に依存するなどの戦略が含まれます。
- これらの影響は、紛争の進展や国際的な対応によって変化する可能性があります。飲食業界は、これらの変動に柔軟に対応し、経営戦略を適宜調整することが求められます。