

金融テキストにおけるセンチメント分析の課題整理

高野 海斗

野村アセットマネジメント株式会社

takaito0423@gmail.com

概要

大規模言語モデル (LLM) の発展に伴い、資産運用業界においても LLM への期待が高まっているが、実際には様々な課題が存在する。これまで自然言語処理 (NLP) 技術の発展に伴い、金融テキストを対象にしたセンチメント分析も発展してきたが、その変化はセンチメント付与の手法やモデルの切り替わりが中心であり、根本的な課題や問題に対して十分な議論が不足している。そこで本研究では、従来の金融テキストにおけるセンチメント分析の課題とその解決の糸口をいくつか示す。課題を明確に示し、整理することで、金融テキストにおけるセンチメント分析が今後より発展することを期待している。

1 はじめに

LLM の急激な発展に伴い、資産運用業界においても AI ファンドマネージャー (運用者) の期待が高まっているが、実際には様々な課題が存在する。ファンドマネージャーは多種多様な情報を元に、ファンドに含まれる資産 (アセット) のポジションを変化させることで、運用収益を積み上げていくことが求められるが、その多種多様な情報のほとんどはテキスト情報である。経済指標や財務指標などに関しても、ニュースやアナリストレポートなどにコメント付きで記載されているため、数値情報に関しても間接的にテキスト情報として活用していることになる。したがって、テキストを入力することで各資産のウェイトを出力することができるモデルを作れば、ファンドマネージャーの業務を疑似的に模倣することが可能である。しかし、実際にテキストを入力してポジションを決めるハードルは非常に高く、NLP 技術活用の現状は、ファクターモデルのファクターの一つとして、テキストから算出されたセンチメントスコアを使用する程度の活用にとどまっている。さらに、人間は大量の情報の取捨選択を行いつつ、時系列変化による評価や現状の資産間の相対的

な評価を複雑に行っているため、言語モデルに同じことをさせるためには、最低限同じ量の情報の入力が必要になる。しかし、モデルに入力できるテキストの量には限界があり、大量の入力を受け入れ可能なモデルであっても、ある一定量を超えると性能が低下することが報告されている [1]。

このような背景から現在は、各テキストに対するセンチメント分析の結果を集計することでセンチメントスコアを算出し、分析や意思決定に活用する研究が盛んに行われている。しかし、実際に実務の資産運用の現場において、金融テキストから算出されたセンチメントスコアの使用は限定的である。その理由は様々だが、センチメントスコアを使用したとしても、コストやリスクに見合った十分なパフォーマンスが得られないことが理由の一つとしてあげられる。また考察にはいろいろと記述があるものの、その大半は「情報がすでに織り込まれている」ことが要因としてあげられている。しかし、十分なパフォーマンスが得られない原因には、それ以外にも意識されていない課題や、手をつけることができていない課題が存在している。

そこで本研究では、従来の金融テキストにおけるセンチメント分析の課題とその解決の糸口をいくつか示す。課題を明確に示し、整理することで、金融テキストにおけるセンチメント分析が今後より発展することを期待している。

2 センチメント分析の課題

従来の金融テキストにおけるセンチメント分析の課題として、1. 分析対象となるテキスト、2. センチメント分析手法、3. センチメントの定義付けの3つの点に焦点を当てる。

2.1 使用するテキストデータの課題

従来研究において、センチメント分析のために使用する金融テキストは、企業が公開する適時開示資料、ニュースデータ、各国の中央銀行や行政機関が

公開する資料や web ページ, web 掲示板や SNS 投稿などがあげられる。特に, 使用される頻度の高い適時開示資料とニュースデータに焦点を絞る。

まず, 企業が公開する適時開示資料に関して, 特に日本では, 決算短信や有価証券報告書がセンチメント分析に使用されることが多い。これらの資料には, 業績の要因についての言及(業績要因文), 業績結果に関する報告(業績結果文), 将来の見通しに関するコメントなどが含まれている。これらの適時開示資料をセンチメント分析で使用する際の課題として, 分析対象のテキストが作成されてから情報が公開されるまでに数か月単位のラグが発生する点があげられる。現状または過去期間の分析が目的であれば, このラグを気にする必要はほとんどないが, 資産運用における投資判断においては, なるべく最新の情報を元に判断を行いたいため, 数か月前に書かれた情報の価値は限定的である。また, 業績要因文や業績結果文は, これまでの状況に関しての言及であるため, 今後の業績を予測するという点においては, 現在の状況が今後も続くという仮定が必要である。したがって, 適時開示資料内の業績要因文や業績結果文のセンチメント分析は, 現状や過去期間の分析には適しているが, 資産運用における投資判断に使用するには工夫が必要である。また, 企業が公開する資料であることから, 主観的なテキストである点にも注意が必要である。ポジティブなことを多く記載する傾向のある企業も存在することから, クロスセクションで企業間の比較を行う場合には, 各企業の傾向を加味する工夫も必要である。

ニュースデータも非常によく使用される金融テキストである。ニュースデータを使用する課題としては, 時価総額の小さい(注目度の低い)企業のニュースが少ない点があげられる。数か月ニュースが出ていないような企業は, センチメントスコアに欠損が生じてしまうため, その後の分析で課題が生じることになる。また, 「コロナの流行」のようなイベントに関連するニュースデータのセンチメント分析にも課題が存在する。あるイベント X が発生してすぐのタイミングでは, 企業に対する言及は限定的であり, 実際にあるイベント X が企業に影響を与えるようなニュースが世に出る頃には, その情報はすでに市場価格に織り込まれていることが多い。例えば, 「コロナの流行」の例であれば, まず, 本日の感染者数に関するニュースが出始め, その後, 飲食チェーン店に関する言及がされたニュースなどが出

始める。そして, その後しばらくしてから, 巣籠需要でゲームの売上が好調といったニュースが出ることになる。それに対して投資家は, 「コロナの流行」というイベントが発生した時点で, 急いでイベントが影響を及ぼす企業の株式を売買する。このとき投資家は, ウイルスの流行というイベントによって将来を予想して売買行動を決定する。したがって, もし可能であるならば, 企業に対する言及がまだないタイミングで, 各企業に対するセンチメントスコアを計算する必要がある。しかし, イベントが企業に対して影響を与えるのか, そしてそれはポジティブな影響なのか, ネガティブな影響なのかは企業ごとに異なるが, 初期の頃に出るイベント発生に関するニュースは, 企業とそもそも結びついていない。したがって, そういったニュースを用いた各企業のセンチメント分析をそのまま行うことはできず, 現状の研究では分析対象になっていない。これらのニュースは景気の景況感を測定するためのセンチメント分析で使用されることはあるが, 業種など問わず一意にセンチメントをニュースに対して付与している。実際のファンドマネージャーの投資の意思決定を考えれば, 初期の頃に出るイベント発生に関するニュースを活用した個別企業のセンチメント分析が求められるが, そのような研究報告はまだほとんどない [2, 3]。

資産運用における使用するテキストに対する課題は, ここまで述べたようにいろいろと存在するが, いくつか解決策も提案されている。まず, 現状の状況などに関するテキストを分析対象にしている研究が多く, 将来に対するセンチメント分析ではない点にある。この課題を解決するシンプルな解決策は, 将来の見通しに関するコメントを分析対象にすることである。業績要因文や業績結果文のように実際に起こったことに関する記述は, 実際の財務情報や株価の情報から観測が可能な事実を反映した情報である。それに対して, 将来の見通しコメントはこれから起こると推測される意見であり, その重要性に関しては先行研究でも言及されている [4]。また, それらの文を抽出する方法に関する研究も行われている [5]。将来の見通しに関するコメントは, 適時開示資料やニュースにも含まれており, その他にもアナリストレポートなどにも含まれている。

このように, 将来の見通しに関するコメントの使用は, 前述したいくつかの課題を解消することができる。しかし, 適時開示資料は主観的であり, 相対

的な評価に不向きである点や、ニュースやアナリストレポートでは時価総額の小さい企業の観測が困難である点、イベント発生から観測までラグが発生する点は未解決のままである。これらの解決方法の一つとして、高野らはセンチメント分析するためのテキストをLLMによって生成する方法を提案している [3]。本来、個別の企業に着目したテキストが得られるのは、イベントが発生してしばらく経過してからになるが、提案手法はLLMの推論能力を活用することで即座にテキストを生成することが可能であり、これらにセンチメントを付与することで、実際の投資家や運用者の投資判断に近い意思決定をすることが期待できる。高野らの手法に関しては、スペースの関係上、Appendixで簡単に紹介する。

2.2 手法の課題

センチメント分析では、文や文章に対して極性を付与する必要がある。この極性の付与を行う方法は、極性語とその極性値が組となった極性辞書を利用する辞書ベースの手法と、言語モデルを利用する機械学習ベースの手法の2つに大別できる。

辞書ベースの手法は、シンプルな手法であることから解釈が容易である。一方で、問題点として文脈を考慮せずに極性が付与される点があげられる。例えば、何が「増加」したかによって「増加」の極性を変化させるべきだが、一般的な極性辞書はこの問題に対処できない。解決策として、「**感染者数の増加**」、「**売上の増加**」のように表現を拡張した辞書を作成する方法がある [6]。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、」のような否定表現に対応できないことも問題点として挙げられる。特に、金融テキストはポジティブな内容とネガティブな内容の両方を含む文が多い傾向にあるため、文脈を考慮できない極性辞書ベースの手法は不向きである。

機械学習ベースの手法は、BERTモデルを活用するのが主流であったが、近年ではLLMのような大規模な言語生成モデルを活用した研究が増加傾向にある。BERTモデルは、Maskした単語の予測や2つの文が連続して出現する関係にあるかなど、自動で正解データを作ることができるタスクによってパラメータの事前学習が可能である [7]。したがっ

て、事前学習済みBERTモデルをテキストと極性が紐づいたデータにより fine tuning することで、極性付与モデルを開発することが可能である。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 [8]。一方で、LLMはモデルのパラメータ数が膨大であることから、BERTモデルのように fine tuning をするにも膨大な計算リソースが必要である。しかし、LLMは様々な情報をモデルが事前に学習していることにより、モデルの入力である prompt に指示とテキストをセットで与えることで、多くのタスクを解決することが可能である。また、指示文だけでなく、具体的な例を追加して与える few-shot learning により、さらに高い精度でタスクを解くことが可能であることが報告されている。

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にある。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は、事前学習で類義語であることが学習されていれば、同様の極性の付与が可能である。一方で、機械学習ベースの手法は、解釈性が低いという課題があるため [9]、解釈性に着目したモデルの研究も盛んに行われている [10]。また、LLM活用の応用として、極性分類の結果だけでなく、その理由まで出力させることで、解釈性を高めることができないかなどの研究も行われている。

機械学習ベースの手法を採用しているセンチメント分析において、分類タスクとして学習したモデルの出力をポジティブやネガティブの強弱として使用している研究が散見される。分類タスクで学習したモデルの出力値は、あくまでそのモデルの分類に対する自信の現れであり、ポジティブやネガティブの強弱とは異なるものである。しかし、分類タスクの出力をセンチメントの強弱として利用し、分位分析によるパフォーマンスを検証している研究は意外にも多い。また、LLMなどが返ってくるセンチメントの強弱に関しては、その数値が返ってきた理由が不明慮であり、また、乱数次第で値が変化するため、ポジティブとネガティブの分類程度であれば問題ないが、その強弱などを出力させて利用するにはまだまだ課題が残る。

本来センチメントは連続値であるはずだが、連続値のラベル付けが難しいことから、多くの先行研究

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

において回帰タスクではなく分類タスクとして学習させたモデルを使用している。したがって金融テキストにおけるセンチメント分析では、連続値の出力が可能なモデルや手法の開発が重要である。連続値の出力が可能なモデルの開発には、手法だけでなくラベル付けされたデータをどのように作成するかが重要である。次節 2.3 では、金融テキストにおいて現状どのような学習データが作られているかを説明した上で、取り組むべき課題を列挙する。

2.3 センチメントの定義に関する課題

極性付与のモデルを学習させるために使用するデータは、テキストに対してポジティブ、ニュートラル、ネガティブのラベルが紐づいたデータが大半である。例えば、金融庁が公開している景気ウォッチャーデータ²⁾には、現状と先行きに対する極性が5段階評価で付けられており、さらにその根拠となるコメントも含まれている。それ以外の極性付与のモデルを学習させるデータとしては、株価のリターン情報などから被説明変数である極性をリバースエンジニアリングする方法もある。ここで問題となるのは、そのテキストがどの程度の影響を株価の動きに与えるかは不明であり、仮に影響があったとしても、それは数日以内のことかもしれないし、数か月後に現れるかもしれないという点にある。したがって、何日後のリターンに影響があるのかなどの調査も必要である。

本研究では、人手でラベル付けされるセンチメントの定義について議論する。まず、Loughran[11]らの金融極性辞書は、業績結果を紐づけることで作成している。したがって、ポジティブな単語は業績が良かったときに出現するものであり、ネガティブな単語は業績が悪かったときに出現するものである。また、Araci[8]のFinBERTは、金融の専門家16人によってラベル付けされたデータを用いて学習されている。このラベルは、次のようにラベル付けしていると記載されている。「The annotators were asked to give labels according to how they think the information in the sentence might affect the mentioned company stock price.」つまり、該当のテキストを読んだ上で、対象の企業の株価が上がる見込みがあると思えばポジティブ、下がると思えばネガティブ、変化しないと思えばニュートラルといったラベルが付与されるため、将来の株価に対する主観的なラベルが付与され

たデータである。

このようにセンチメントと一言でいっても、その定義は多種多様である。前者の研究は書き手による主観的評価に注目しており、後者の研究は読み手による客観的評価に注目している。したがって、前者は企業経営者がどのように考えているかを推測するためのセンチメントであり、後者はそれらのセンチメントによって市場がどのような行動を取るかを推測するためのセンチメントとなり得るだろう。しかし、金融におけるセンチメント分析の関連研究は、このような定義をほとんど考慮せずに使用しているものも多い。本来、これらのセンチメントがどのような定義に則っているかを踏まえ、資産のリターンとの関係性に関する仮定をおいた上で分析を進めていくべきである。

さらに、センチメントに関しては強弱が存在するが、強弱を反映したデータセットを作成することは非常に難しい。例えば、ポジティブ、ネガティブ、ニュートラルの3値分類でも、ラベル付けに迷うものも存在する。そして、強いポジティブ、弱いポジティブのような分類の細分化は、その境界がさらに曖昧であり、アナテーターごとにその境界が異なるため、多くの人の協力が必要となる。

ここまで述べてきたことに加えて、根柢のセンチメントの軸も重要である。例えば、資産運用におけるセンチメントには、割安、割高などの軸も考えられる。また、対象が株価のリターンなのか、社債の利回りなのかによって、異なるセンチメントになることもあるだろう。

3 まとめと今後の展望

金融テキストにおけるセンチメント分析には、対応しなければならない課題が山積みである。2.1節で述べたように、同じテキストでも業種ごとにセンチメントが異なることも考えられる。したがって、何のためにセンチメントモデルを開発するのか、そして、そのためにどのような定義に則ったデータを作成する必要があるのかを検討することにもっと時間を割く必要がある。このような課題が山積みであるという背景もあり、先行研究のいくつかはこのような議論を避けている。そして、そのような研究を参考に行われている多くの研究では、こういった課題があることすら認識されていない現状がある。急激に技術が発展した今こそ、これらの課題に対して改めて取り組むべきだと考える。

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

参考文献

- [1] Mosh Levy, Alon Jacoby, and Yoav Goldberg. Same task, more tokens: the impact of input length on the reasoning performance of large language models. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 15339–15353, 2024.
- [2] 高野海斗, 中川慧. 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. 言語処理学会第30回年次大会, pp. 1295–1300, 2024.
- [3] 高野海斗, 中川慧. 大規模言語モデルによる事業概要を考慮した金融テキストの推論ベース極性分析. 第38回人工知能学会全国大会, pp. 4Xin211–4Xin211, 2024.
- [4] 小寺俊哉, 田中良典, 佐藤史仁, 佐久間洋明, 坂地泰紀, 和泉潔. 有価証券報告書からの未来志向文の抽出. 第33回人工知能学会全国大会, pp. 2O1J1305–2O1J1305, 2019.
- [5] 北森詩織, 酒井浩之, 坂地泰紀. 決算短信 pdf からの業績予測文の抽出. 電子情報通信学会論文誌 D, Vol. J100-D, No. 2, pp. 150–161, 2017.
- [6] 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.
- [7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4171–4186, 2019.
- [8] Dogu Araci. FinBERT: Financial sentiment analysis with pre-trained language models. **arXiv preprint arXiv:1908.10063**, 2019.
- [9] David Gunning and David William Aha. DARPA’s explainable artificial intelligence program. **AI Magazine**, Vol. 40, No. 2, pp. 44–58, 2019.
- [10] 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.
- [11] 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.

Appendix

高野らの研究では、LLM の推論能力を活用し、(1) ある企業の事業に関する情報を入力として与え、(2) ある特定の重大イベントが当該企業に与える影響を推論させることで、センチメント分析を行うためのテキストを生成する方法を提案している。提案手法によるセンチメント分析のイメージを図 1 に示す。

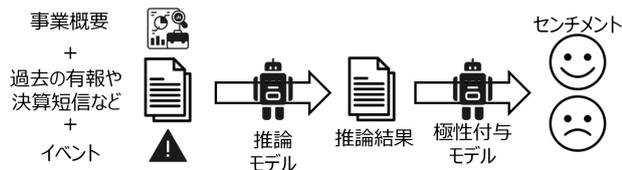


図 1 個別企業の情報を加味した推論ベース極性分析

具体的には、以下のような prompt を用いる。

あるイベントの影響を推論させるための prompt

<ある会社 C の事業概要 x_C >
上記の内容を踏まえて、<イベント E> が起こった場合、この企業にどのような影響があると考えられますか？

提案手法によって、ロシアとウクライナの戦争がブリヂストンの業績にどのような影響を与える可能性があるか推論した結果を以下に示す。

ブリヂストンの業績への影響の推論結果

1. 原材料調達の困難：ロシアは天然ガスや石油などのエネルギー資源を豊富に持つ国であり、これらはタイヤ製造の重要な原材料となります。戦争が起こると、これらの資源の供給が困難になり、原材料の価格が上昇する可能性があります。これにより、製造コストが上昇し、利益率が低下する可能性があります。
2. 販売市場の縮小：ロシアやウクライナは、ブリヂストンの製品を購入する消費者市場でもあります。戦争が起こると、これらの市場での販売が困難になり、売上が減少する可能性があります。