

論文の内容の要旨

論文題目 拡張トピックモデルを用いた
金融テキストに対する指標構築、およびマーケット分析

氏 名 余野 京登

本論文の目的は、金融経済分野において、これまでには数値データとして観測されにくく、分析することが困難であるが、金融市場へ影響を与えるマクロ経済ファクター・心理的な要因をテキストデータを用いて、分析することにより、資産運用への活用を試みることである。個々の要因に大きく関係する非構造化データであるテキストデータに対して、拡張したトピックモデルを用いて、その潜在的な意味を踏まえ、要素ごとに抽出し数値化することにより、金融市場システムに対する変動要因を解析する。具体的には分析の対象とするマクロ経済ファクター・心理的な要因は、中央銀行の経済や物価に対するセンチメント、経済の先行きに対する不確実性、金融市場における市場参加者センチメントの三つである。

金融テキストを分析する意義は、数値データに含まれていない要因を内包しているためである。また、数値データとして、公表されるまでの時間がかかるため、それより以前に公表されるテキストデータを用いてその数値を予想することで速報性を得ることができるという利点もある。しかし、テキストデータは数値データと異なり「非構造化」されたデータである。データ量が膨大であり、かつ必要な情報の抽出が難しい。数値データと異なり、テキストデータをそのまま分析の対象とすることは難しい。そのため、テキストデータを分析に使用する場合には、非構造化データから構造化データに変換する必要がある。

文章を構造化データに変換する方法は、文章を形態素解析し、単語単位に分解したのち、その単語の頻度のベクトルとすることである。これをBoW(Bag of Words)表現と呼ぶ。しかし、このBoW表現をそのまま運用モデルに用いることも難しい。なぜなら、BoWの次元数は単語の種類数に等しく、次元の呪いにより、モデルが効率よく回帰または分類ができなくなる。そのため、非構造化データであるテキストを利用するためには、次元を落とすうえで、数値化することが必要となる。かつ次元を落とす際に情報の欠落

をできるだけ防ぐ必要がある。

本論文では、金融分野において、「非構造化データ」である金融テキストを使用し、今まで投資家が定性的にしか扱えなかった日本銀行の経済や物価に対する個別のセンチメント(第2章)、マクロ経済の先行きに対する不確かさ(第3章)をトピックモデルを用いて定量的に計測し、それを指数として構築した。加えて、金融テキストデータの資産運用への活用として、金融テキストから為替の決定要因となるマクロ経済要因、およびそれに対する投資家のセンチメントを抽出し、為替市場における短期間の価格変動の方向性予想を行った(第3章)。

本論文の学術的な貢献は、投資家にとって今まで定量的な活用が難しかった「非構造化データ」である金融テキストを、必要な情報量を落とすことなく解釈しやすいトピックごとの構造化データへ変換し、指数として定量化する手法を提案し、その指数を用いて、マクロ経済の分析や資産運用への応用を行ったことにある。また、各章それぞれにおいて、使用する金融テキストの特性や分析の目的に合わせて、使用するトピックモデルを拡張し、既存の手法に対してその有意性について示している。多くの金融テキストが未だに活用されていない中、本論文で示した拡張したトピックモデルを適応し、投資家にとって新たな経済分析や資産運用手法の開発に繋げることが可能である。各章の具体的な成果は以下のようになる。

第2章) 中央銀行の経済や物価に対する個別のセンチメントの可視化}

日本銀行のマクロ経済見通しを各マクロ経済要素ごとに定量的に把握するため、日本銀行の金融経済月報に対して、深層学習モデルを用いたセンチメントの付与とトピックモデルによるトピック分類を組み合わせたモデルにより、「個人消費」、「企業収益」、「資金供給」、「物価」等の12個の各マクロ経済の要因ごとの日本銀行のセンチメント指数を構築した。構築した各マクロ経済の要因ごとのセンチメント指数の時系列推移と日本銀行の金融政策との比較から、各トピックのセンチメントの変化と金融政策への影響について論じた。また、多くのマクロ経済指標と各センチメント指数との間に正の相関がある中、日銀の物価に対するセンチメントは、金融政策の特性上、マクロ経済指標とは連動しないことが確認した。

第3章) 拡張したトピックモデルによる不確実性指数の構築

マクロ経済の先行きの不確実性をその要因ごとに定量的に把握するため、不確実性の代理変数である市場の将来変動を示すVIX指数を教師信号としてモデル内に組み込んだトピックモデルを提案し、要因ごとの不確実性指数を構築した。拡張したトピックモデルを用いたことで各要因の不確実性がどの程度市場変動に影響を及ぼすかを同時に推

定することができ、教師信号の推定値の大きさに比例して、実際の各市場指数の変動との相関が大きいことを確認した。また、先行研究の手法との比較を通じて、提案モデルの方が、より不確実性の要因を特定したノイズが少ない指数を構築することができることを示した。マクロ経済指標への影響分析において、日本の金融政策に起因する不確実性指数の変動は、日本の鉱工業生産指数を押し下げることを確認した。

第4章) 市場センチメント指数を用いた為替市場変動の方向性の予想

為替のニューステキストから拡張したトピックモデルを用いて為替の決定要因となるマクロ経済ファクターやそれに対するポジネガを示すセンチメントを抽出し、それを特徴量として数値データを合わせて、為替市場における価格変動の方向性予測を行った。潜在変数としてトピックとセンチメントの両方の変数を持ち、加えて、教師信号の変数を持つsJSTモデルを提案し、32通貨ペアの価格変動の方向性予測を行った結果、最も精度が高いという結果となった。モデルの差に起因する分析を行い、センチメントを潜在変数と加えるトピックモデル拡張のみ、予測精度の向上に寄与することが有意に認められた。使用するテキストの内容の特性から、通貨ペア間では、取引量の少ない通貨ペアより取引量の多いメジャー通貨の方が、より多く予測精度が向上した。

以上のように、金融政策、マクロ経済、金融市場など複数の領域において、拡張したトピックモデルを活用した金融テキストの指数化を通じて、定量的に把握できなかったマクロ経済要因をより細かい粒度で資産運用への応用することができたと言える。また、金融テキストが数値データと関係性のある場合、金融テキストに投資家センチメントが含まれている場合などにおいて、トピックモデルを拡張することで、指数の精緻化や予測精度の向上等のメリットを示すことができた。

本研究で構築した指数は、金融分野の様々な分野で利用することができる。例えば、経済アナリストが、中央銀行の経済や物価に対する個別のセンチメント指数を用いて経済レポートを執筆することや、ファンドマネージャーが、要因ごとの不確実性指数を用いて、ポートフォリオのヘッジ戦略を構築したりすることなどが挙げられる。また、個別株のアナリストが決算短信などの膨大な企業に関する公開テキスト情報に、本論文で論じてきた指数化の手法を適用し抽出した特徴量を用いて、業績予想に活用することも可能だ。

今後、金融テキストはさらに増大し、機械学習、深層学習などを資産運用の分野へ活用の流れが加速していく中において、テキストデータの活用の可能性はさらに増すだろう。膨大かつ非構造化データである金融テキストデータを効率よく、かつ人間の解釈可能な形に定量化、指数化することは、資産運用モデルの構築時においても、パフォー

マンスの評価時においても非常に重要になっている。本研究では、金融政策、マクロ経済、金融市場など複数の領域において、トピックモデルによる金融テキストの指数化がマクロ経済分析や資産運用モデル構築へ有用であることを示した。