

審査の結果の要旨

氏名 余野 京登

本研究では、自然言語処理におけるトピックモデルを拡張させ、金融経済テキストに適応させることにより、非構造化データである膨大なテキスト情報に埋もれた潜在的な要因・イベントを抽出する手法を新たに提案する。提案手法を用いて、数値データが時々刻々変動する市場において、どの要因やイベントがどの程度影響したのかを分析し考察を行った。本研究では、為替概況ニュースやマクロ経済カテゴリーのニュースを対象に分析を行った。一般的なトピックモデルを拡張し、マーケットの数値情報を教師信号とするモデル、内部にセンチメントを代表する隠れ変数を有するモデル等を用いて、ニュースから潜在的な要因・イベントの抽出、および可視化を行った。本研究で提案する手法によって、投資家が投資判断を行う際の現状判断に有用な情報抽出を可能にした。経済・金融分野への工学的アプローチを試みることにより、金融市場システムにおける変動要因の解明に貢献する。

第1章では本研究の背景・モチベーションを議論し、第2章から第4章まで金融経済テキストからの情報抽出に関して、抽出される数値データの利用目的や頻度が異なる条件に適した手法を各章で提案した。

第2章では日本銀行が発行しているテキストを対象に中央銀行の経済や物価に対する個別のセンチメントをモデルから計算し、時系列推移として可視化する手法を提案した。日本銀行の景気に対するセンチメントを日本銀行の発行するテキストに対して、トピックモデルおよびニューラルネットワークを用いて、トピック事に分解した指数化を行った。モデルは LSTM RNN(Long Short Term Memory Recurrent Neural Networks) による各文章に対するスコア付与と LDA(Latent Dirichlet Allocation)によるトピック分類を統合した手法を開発した。単語の共起性をもとにした潜在変数(トピック)を導入し、潜在変数のセンチメントを計測することで、2014-2015年にかけて足元では全体的なセンチメントはプラスで推移する一方、物価トピックのセンチメントはマイナスで推移していたというトピック毎の詳細な分析を非構造化データで可能となった。

第3章では、テキスト作成の背景情報に着目し、状態空間モデルをベースとしたモデルを提案し、物価トピックに対する更に詳細なセンチメントの構築を行った。潜在変数の推移を可視化した数値、テキストの両方の影響を受けた新たな指標を提案することができた。第5章では、ニューステキストから不確実性指数の構築を行う。その際にトピックモデルを用いたニュースの分類を行うが、米国株式市場の株価変動指数を用いた教師ありトピックモデルである supervised LDA モデルを適用した。その結果、2009年から2019年までの10年間で時期毎に市場の不確実性を支配する世界中の各市場に対応するトピックを抽出することができた。さらに、インパルス応答分析により各国の金融政策の動きと各トピックの関連性の高さを明らかにした。

第4章では、教師信号を考慮したセンチメントとトピックを同時に抽出可能な supervised joint sentiment topic (sJST)モデルを新たに提案した。本モデルによって抽出されたトピック得点を用いて、32組の通貨ペアの為替の価格変動を複数の機械学習モデルにより予測する実験を行った。既存手法と比較した結果、本モデルにより抽出した特徴量を用いた市場予測結果が、他のモデルを用いた場合より高い精度を示した。

第5章において結論を述べている。本論文の成果および今後の課題を記している。付録Aの章において、本論文で開発した手法をシステム実装し、テキストデータから分析した不確実性指数を可視化するツールの紹介を行った。

以上を要するに本研究は、金融経済分野において、これまで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や市場心理に関わる要因を、テキストデータを用いて分析することにより、資産運用への活用する新たな手法を提案した。本手法により、経済金融分野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テキストの指数化を通じて、定量的に把握できなかった要因をより細かい粒度で社会分析に応用することができると期待される。本研究に関する課題設定・モデル構築・結果分析について工学的研究として十分な水準にあると評価された。本研究の有用性や新規性についても評価され、研究の完成度や発表実績についても十分であると認められた。

よって本論文は博士（工学）の学位請求論文として合格と認められる。