

令和 6 年 4 月 23 日現在

機関番号：33107

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K01863

研究課題名（和文）集合知に基づくVI上昇予測モデルの実用化に向けたパイロットプラントの構築

研究課題名（英文）Construction of Pilot Plant of VI Rise Prediction Model Based on Collective Intelligence

研究代表者

梅原 英一（Umehara, Eiichi）

新潟国際情報大学・経営情報学部・教授

研究者番号：00645426

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は、日経平均ボラティリティ・インデックスの上昇の予測を、SNSを用いて機械学習により行った。SNSの投稿記事は速報性はあるが、文章が非常にわかりにくい。また機械学習の際に、感情（上がる/下がる）やトピックのどちらかに着目して分析を行っていた。本研究では、トピックと感情の両方を活用するモデルを開発した。この手法をSSCDVと名付けた。この結果、F-1スコアで良い結果を得られた。また学習結果を用いて、投資シミュレーションを実施した。この結果、利益が得られた。実用的な観点からも有用であることを示した。

研究成果の学術的意義や社会的意義

日経平均ボラティリティ・インデックス（以下日経平均VI）の上昇を予測することは、相場の荒れ具合を予測することである。これは機関投資家のリスク回避に役立つ。特に年金基金は株式に積極的に投資している。株式市場が今後荒れると予測できれば、現金比率を高めることでリスク回避に役立つ。本研究では、日経平均VIの予測に当たり、トピックと感情の両方を同時に解釈する手法を開発した。検証の結果、パフォーマンスも良く、我々のモデルの有効性が確認できた。

研究成果の概要（英文）：This study used machine learning to predict the rise in the Nikkei Stock Average Volatility Index using SNS. Articles posted on SNS are breaking news, but the text is very difficult to understand. In addition, during machine learning, analysis was performed by focusing on either emotion (up/down) or topic. In this study, we developed a model that utilizes both topics and emotions. We named this method SSCDV. As a result, we obtained a good result with an F-1 score. We also conducted an investment simulation using the learning results. This resulted in a profit. It was also shown to be useful from a practical point of view.

研究分野：人工知能

キーワード：株式掲示板 トピックと感情 機械学習 自然言語処理 ボラティリティインデックス ヤフー株式掲示板 日経新聞記事

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

### 1. 研究開始当初の背景

インターネット株式掲示板には一般投資家の意見や気持ちが数多く投稿されている。しかし投稿されているメッセージは集約されておらず、集合知としては活用することは実現できていない。そこで本研究では多数の一般投資家がインターネット上の投稿したメッセージを集合知化し、株式リスクの予測を試みる。株式リスクとは株式市場や株価の大幅な上昇や下落のタイミングや可能性を予測することである。具体的には株価リターンの将来の標準偏差が変化するタイミングを予測することである。この予測に関しては多くの分野でチャレンジされているが、まだ完全な予測は実現できていない。

インターネット株式掲示板を用いた株価指標予測の研究は非常に数多くある。我々もヤフー研究所にデータ提供を頂き、Yahoo! JAPAN 株式掲示板のデータを使い分析してきた。我々 (Suwa et al. (2017)) は、株式指標の中でも、恐怖指数と呼ばれる日経平均ボラティリティ・インデックス (以下日経平均 VI と呼ぶ) に着目した。日経平均 VI は大阪証券取引所に上場されている日経平均オプションの価格から導出した「日経平均株価リターンの標準偏差の予測値」であり、日経新聞にも公表されている。我々はインターネット株式掲示板による日経平均 VI 上昇日予測モデル (以下 Suwa et al. モデルと呼ぶ) を提案した。我々は潜在的ディリクレ配分法 (LDA) を用いて Yahoo! Japan 株式掲示板の投稿内容からトピックを抽出した。その上で、トピックの時系列変化から機械学習を用いて VI 上昇予測モデルを構築した。しかし我々の手法は有効だがまだ精度面で改善の余地がある。機械学習としてロジスティック回帰を用いた場合の適合率 (Precision) は 0.453、再現率 (recall) も 0.453 であり 50%未満である。しかしランダムに選択した場合の期待値 (適合率) は 0.24 であるので、機械学習による効果は確認できている。また機械学習としてランダムフォレストのテストも行ったが良い結果は得られなかった。故に我々は機械学習手法の選択も精度向上に非常に重要であることも確認している。

Suwa et al. モデルの検証として、佐々木ら (2019) は、Suwa et al. モデルの予測結果に基づくイントラデイ価格データを用いた日経平均オプションのボラティリティ・トレーディングシミュレーションを行った。その結果、大幅な収益向上が確認できた。つまり Suwa et al. モデルは実用面で有効である可能性が高いことを示した。

### 2. 研究の目的

本研究の目的は集合知に基づく株式市場における実用的なリスクモデルを提案することである。株式投資は年金基金の運用の中心であり、株式市場リスクの予測は企業経営にとり重要な問題である。そこで本研究ではインターネット株式掲示板を使った株式市場リスクモデルを構築する。このリスクモデルはいわば株式市場の天気予報 (株式市場は安定するか、荒れるかを予測) として機能する。

これまでの我々の研究はヤフー研究所に Yahoo! Japan 株式掲示板のデータ提供を頂き、過去のデータを使った日経平均 VI 上昇日予測モデルの構築及び検証を行ってきた。その有効性が確認できたので、次のステップとして実証実験のフェーズに入る。本研究の目的は、機械学習のロジック見直しによるモデル精度向上である。第 1 にトピックだけで文章ベクトルを作成していた。しかし、トピック以外に感情も重要であると考え、そこで、トピックと感情の両方を抽出するモデルを開発する。第 2 に、インターネット株式掲示板だけでなく、新聞記事を加えると精度が向上するかを検証することである。

### 3. 研究の方法

第 1 の目的に対しては、既存の研究の限界を克服するために、ソーシャルメディアからのトピックと感情の両方を活用するドキュメント埋め込み技術である SSCDV を導入する。SSCDV は、感情極性辞書から参照される感情値を使用して、Sparse Composite Document Vectors (SCDV) (Mekala, Gupta, Paranjape, & Karnick, 2017) と呼ばれる埋め込みアルゴリズムを拡張することによって開発した。SCDV は、事前トレーニングされた単語埋め込みにガウス混合モデル (GMM) ソフトクラスタリング (Reynolds et al., 2009) を適用することでトピックを抽出する従来型の埋め込み手法である。

既存の SCDV はトピック情報のみに焦点を当てている。感情情報を組み込む機能が不足している。本研究で提案する手法である SSCDV は、SCDV を拡張して、ドキュメント内のトピックと感情情報の両方を使用する手法である。具体的には、円高というトピックがあったとする。このときに円高が株式相場に対して悪影響を及ぼすのか、上昇に寄与するのかはそのときの状況により異なる。この点の改善を考えた。

SSCDV は、最初の埋め込み段階でネガティブサンプリングによるスキップグラム (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) (SGNS) ベースのアプローチを採用しているため、ウィンドウサイズに依存するコンテキスト情報の使用が可能になる。その結果、トピックベースの方法と比較して、SSCDV は単語の意味表現を改善し、文書に関する意味情報を反映する文書ベクトルを生成する。さらに、SSCDV は SCDV アルゴリズムの利点の一部を継承しており、文章

レベルのコンテキストを理解し、文脈上の単語の意味変更（多義語など）を効率的に処理して文章ベクトルを生成することができる。これにより金融市場の予測にとって重要なデータが取得できる。

提案手法を評価するために、3つの課題を設定した。第1の質問「Q1: SSCDV を特徴量として用いたモデルの 日経平均 VI 予測タスクのパフォーマンスはどの程度か?」を提案手法の予測精度の評価として設定する。このために、SSCDV を使用して、日本最大の株式関連ソーシャルメディア サイトの 1 つである Yahoo! JAPAN 株式掲示板から投稿データを取得した。さまざまな埋め込みモデルをベースラインとして使用した。これらは LDA、JST、SWEM-Aver、BERT、RoBERTa、SCDV などである。

実際の投資における予測の有効性を評価するために、2 番目の質問として「Q2: 現実世界のデータセットを使用した投資シミュレーションで SSCDV 機能を使用したモデルの投資収益率はどの程度か?」を設定した。このために、日経平均 VI 上昇予測を用いた投資シミュレーションを実施した。Q1 と同様に、提案手法による投資シミュレーション結果を、ベースラインとして既存のさまざまな組み込みモデルの投資シミュレーション結果と比較した。実験結果は、予測パフォーマンスと投資シミュレーションの両方の観点から、提案手法の有効性を実証した。

提案された埋め込み手法をさらに分析するために、次の質問を作成した。「Q3: 埋め込みは機械学習モデルの学習と予測にどのような影響を与えるか?」。このために、SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017) の分析を実施した。これにより、予測に効果的なトピックと感情のペアが時系列に沿って明確に特定された。本研究の方法を図 1 に示す。

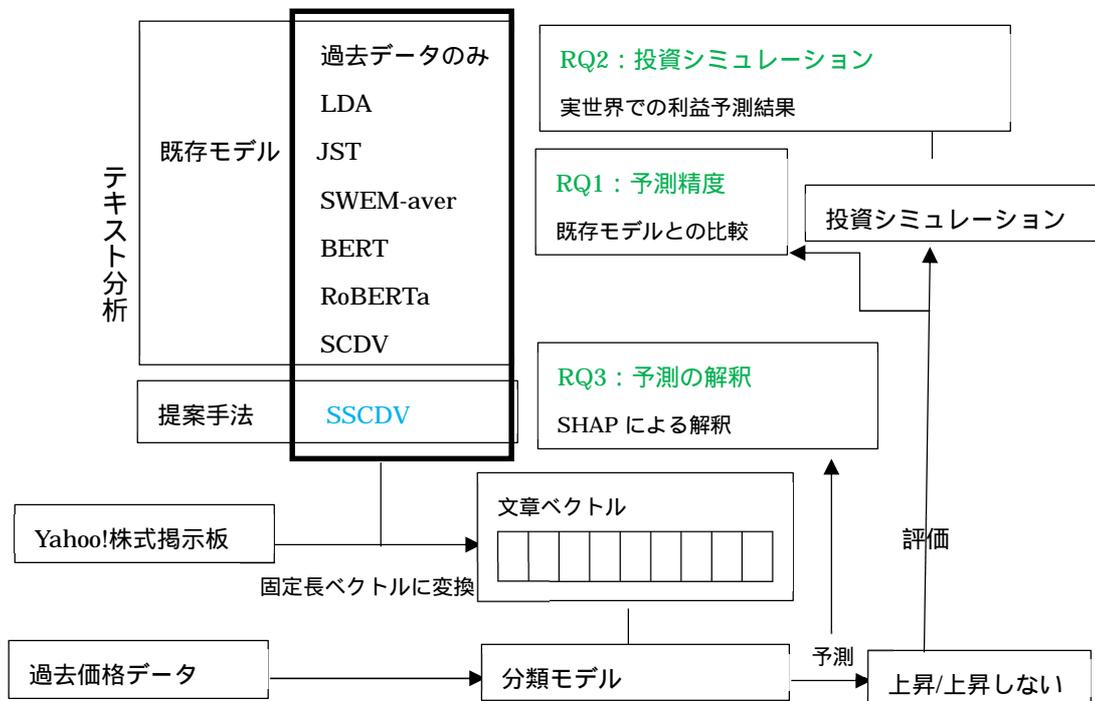


図 1 第 1 目的の研究手法

第 2 の目的に対しては、Yahoo! Japan 株式掲示板の情報に加えて、日経新聞記事を一緒に機械学習にかけます。分析手法はこれまで我々が実施してきた LDA を使う。Yahoo! Japan 株式掲示板はソーシャルメディアで速報性は高いと考えられるが、情報は不正確である可能性がある。一方、日経新聞は速報性は低い、情報はかなり正確である。株式市場に関する過去の研究では、新聞記事公表の前に株価が反応しており、新聞記事からの超過収益は見込めないという議論もある。そこで、新聞記事とソーシャルメディアを組み合わせると、どの程度の予測精度が出るかを検証する。

#### 4. 研究成果

第 1 の目的に対する研究成果を述べる。第 1 の質問「Q1: SSCDV を特徴量として用いたモデルの 日経平均 VI 予測タスクのパフォーマンスはどの程度か?」については、F-1 スコアで 0.333 となった。この値は、実施した既存のテキスト分析手法よりも高かった。ちなみに 2 位は RoBERTa の 0.332 であった。またマッシュューズ相関係数も 0.215 で一番高かった。2 位の RoBERTa は 0.213 であった。

第 2 の質問「Q2: 現実世界のデータセットを使用した投資シミュレーションで SSCDV 機能を使用したモデルの投資収益率はどの程度か?」は、2016 年 10 月 1 日から 2020 年 9 月 30 日までで +1,249,880 円の成果が得られた。毎日取引では▲7,874,570 円であった。コロナ渦で市場のボラティリティが高まっている時でも、モデルが利益を生み出す可能性があることが確認された。

現実世界の投資リスク軽減モデルとしての有効性が実証された。

第3の質問「Q3: 埋め込みは機械学習モデルの学習と予測にどのような影響を与えるか?」に関しては、SHAPを用いたモデルの解釈を行った。この結果、ソーシャルメディア上の一部のトピック - 感情ペアは一貫して市場予測に貢献した。しかし、一時的な影響しか与えないものもあることが明らかとなった。提案された埋め込み技術がトピックと感情を一緒に扱うことで効果的な予測を可能にすることを示した。この結果、人間の直感と一致しており、提案された埋め込み技術を使用したモデルが投資の意思決定をサポートするのに効果的であることを示唆している。

SSCDVは、インターネット株式掲示板のノイズに対して特に堅牢であった。従来のトピックベースの手法よりも豊富な情報を文章から取得できた。さらに、金融市場の予測に不可欠なトピックと感情の間の複雑な相互作用をベクトルに明示的に埋め込むことができた。

第2の目的に対する研究成果について述べる。日経平均VIの予測結果では、提案手法である「Yahoo!Japan 株式掲示板 + 日経新聞 + 金融時系列」のF-1値が0.30となり一番高かった。また、日経平均VIの上昇が不均衡データであるために、PR-AUCも検証した。この結果でも、提案手法でPR-AUCが0.19となり一番高かった。ちなみに2位は「Yahoo!Japan 株式掲示板 + 金融時系列」でF-1値が0.26、PR-AUCが0.18であった。この結果から、「Yahoo!Japan 株式掲示板 + 日経新聞 + 金融時系列」の予測精度が確認できた。

## 5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 4件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 梅原英一, 加藤奈美恵, 諏訪博彦, 小川祐樹, 杉浦昌	4. 巻 8
2. 論文標題 組織における個人情報保護行動モデルの構築 - 従業員の個人情報保護行動を促進するためには -	5. 発行年 2020年
3. 雑誌名 社会情報学	6. 最初と最後の頁 81-95
掲載論文のDOI (デジタルオブジェクト識別子) 10.14836/ssi.8.3_81	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 林浩輝, 梅原英一, 小川祐樹	4. 巻 8
2. 論文標題 否決された大阪都構想のTwitter投稿における世論形成理論成立の考察	5. 発行年 2020年
3. 雑誌名 社会情報学	6. 最初と最後の頁 165-175
掲載論文のDOI (デジタルオブジェクト識別子) 10.14836/ssi.8.3_165	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 梅原英一	4. 巻 58
2. 論文標題 大学図書館のデジタル・トランスフォーメーションに対する組織のIT活用能力	5. 発行年 2021年
3. 雑誌名 日本印刷学会誌	6. 最初と最後の頁 18-22
掲載論文のDOI (デジタルオブジェクト識別子) 10.11413/nig.58.18	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Ueda Kentaro, Suwa Hirohiko, Yamada Masaki, Ogawa Yuki, Umehara Eiichi, Yamashita Tatsuo, Tsubouchi Kota, Yasumoto Keiichi	4. 巻 245
2. 論文標題 SSCDV: Social media document embedding with sentiment and topics for financial market forecasting	5. 発行年 2024年
3. 雑誌名 Expert Systems with Applications	6. 最初と最後の頁 in printing
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.eswa.2023.122988	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 梅原英一, 渡部和雄, 岩崎邦彦	4. 巻 61
2. 論文標題 読者調査に見る紙マンガと電子マンガの嗜好の違い	5. 発行年 2024年
3. 雑誌名 日本印刷学会誌	6. 最初と最後の頁 8-14
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 渡部和雄, 梅原英一, 岩崎邦彦	4. 巻 60
2. 論文標題 消費者の紙の出版物, 電子出版物への意識と行動の分析と利用促進策 テキストでの回答の分析を中心として	5. 発行年 2023年
3. 雑誌名 日本印刷学会誌	6. 最初と最後の頁 23-31
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計18件 (うち招待講演 1件 / うち国際学会 1件)

1. 発表者名 細川 蓮, 小川 祐樹, 上田 健太郎, 諏訪 博彦, 梅原 英一, 山下 達雄, 坪内 孝太
2. 発表標題 新聞メディアと株式掲示板を用いた日経VI指数予測
3. 学会等名 2022年度 人工知能学会全国大会
4. 発表年 2022年

1. 発表者名 細川蓮, 上田健太郎, 諏訪博彦, 梅原英一, 山下達雄, 坪内孝太, 小川祐樹
2. 発表標題 新聞とインターネット株式掲示板を用いた金融指標の予測
3. 学会等名 第30回 人工知能学会 金融情報学研究会
4. 発表年 2023年

1. 発表者名 Ueda Kentaro, Sasaki Kodai, Suwa Hirohiko, Ogawa Yuki, Umehara Eiichi, Yamashita Tatsuo, Tsubouchi Kota, Yasumoto Keiichi
2. 発表標題 Prediction of Nikkei VI increase for reducing investment risk using Yahoo! JAPAN stock BBS
3. 学会等名 The 6th International Workshop on Application of Big Data for Computational Social Science in WI-IAT 2021 (国際学会)
4. 発表年 2021年

1. 発表者名 上田 健太郎, 諏訪 博彦, 小川 祐樹, 梅原 英一, 山下 達雄, 坪内 孝太, 安本 慶一
2. 発表標題 金融指標予測のためのソーシャルメディアに適した分散表現獲得手法の検討
3. 学会等名 第28回社会情報システム学シンポジウム
4. 発表年 2022年

1. 発表者名 上田 健太郎, 諏訪 博彦, 小川 祐樹, 梅原 英一, 山下 達雄, 坪内 孝太, 安本 慶一
2. 発表標題 日経VI予測のためのソーシャルメディアの感情とトピックを用いた文書分散表現獲得手法の提案
3. 学会等名 社会システムと情報技術研究ウィーク2022 (WSSIT2022)
4. 発表年 2022年

1. 発表者名 梅原英一, 富岡雄大
2. 発表標題 電子図書館のビジネスモデルとTCO予測シミュレーション
3. 学会等名 日本印刷学会2021年研究発表大会
4. 発表年 2021年

1. 発表者名 梅原英一, 富岡雄大
2. 発表標題 東京都市大学の電子図書館の TCO
3. 学会等名 2021年経営情報学会全国研究発表大会
4. 発表年 2021年

1. 発表者名 佐々木皓大, 諏訪博彦, 小川祐樹, 梅原英一, 山下達雄, 坪内孝太, 安本慶一
2. 発表標題 ヤフーファイナンス掲示板を用いたコロナ・ショックにおける投資家の話題変化と株式指標の関係
3. 学会等名 第27回社会情報システム学シンポジウム
4. 発表年 2021年

1. 発表者名 佐々木皓大, 諏訪博彦, 小川祐樹, 梅原英一, 山下達雄, 坪内孝太, 安本慶一
2. 発表標題 ヤフーファイナンス掲示板の投稿を用いた株投資リスク低減のための日経VI上昇予測
3. 学会等名 社会システムと情報技術研究ウィーク
4. 発表年 2021年

1. 発表者名 梅原英一, 小川祐樹, 平川敦貴
2. 発表標題 2019年大阪ダブル選挙におけるTwitter分析
3. 学会等名 2020年人工知能学会全国大会
4. 発表年 2020年

1. 発表者名 梅原英一
2. 発表標題 大学のデジタル化に対する組織のIT活用能力と従業員の能力開発
3. 学会等名 経営情報学会2020年秋季全国研究発表大会
4. 発表年 2020年

1. 発表者名 梅原英一
2. 発表標題 大学のデジタル化と大学図書館の電子化の課題
3. 学会等名 日本印刷学会2020年全国大会（招待講演）
4. 発表年 2020年

1. 発表者名 杉本明隆，梅原英一
2. 発表標題 マルチエージェント・シミュレーションによる従業員を新型コロナ感染から守る対策の提案
3. 学会等名 第27回社会情報システム学シンポジウム
4. 発表年 2021年

1. 発表者名 有谷 公輔，梅原 英一
2. 発表標題 2019 年参議院議員選挙におけるTwitter分析
3. 学会等名 信学技報 L01S2020-47
4. 発表年 2021年

1. 発表者名 何 双兵, 梅原 英一
2. 発表標題 遠隔グループワーク支援システムの開発
3. 学会等名 情報処理学会第83回全国大会
4. 発表年 2021年

1. 発表者名 梅原英一, 渡部和雄, 岩崎邦彦
2. 発表標題 マンガの紙利用者と電子利用者の調査
3. 学会等名 2023年度社会情報学会全国大会
4. 発表年 2023年

1. 発表者名 梅原英一, 渡部和雄, 岩崎邦彦
2. 発表標題 紙マンガと電子マンガの読者同行
3. 学会等名 経営情報学会2023年全国研究発表大会
4. 発表年 2023年

1. 発表者名 梅原英一, 渡部和雄, 岩崎邦彦
2. 発表標題 紙と電子マンガのインタビュー調査
3. 学会等名 日本印刷学会第150回研究発表会
4. 発表年 2023年

〔図書〕 計1件

1. 著者名 Eiichi Umehara	4. 発行年 2022年
2. 出版社 Springer	5. 総ページ数 14
3. 書名 " A Game Theory Investigation of Contract Between IT Vendor and User in Problems of Information System " in Traditional Systems Science Vol.27: Systems Research II	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	諏訪 博彦 (Suwa Hirohiko)  (70447580)	奈良先端科学技術大学院大学・先端科学技術研究科・准教授  (14603)	
研究分担者	小川 祐樹 (Ogawa Yuki)  (40625985)	東京都市大学・デザイン・データ科学部・准教授  (34315)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------