

令和元年6月21日現在

機関番号：32678

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K03820

研究課題名(和文) ソーシャルセンチメントを活用した投資及びマーケティングモデルの構築

研究課題名(英文) Development of financial and marketing models using social sentiment

研究代表者

梅原 英一 (Umehara, Eiichi)

東京都市大学・メディア情報学部・教授

研究者番号：00645426

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：第一にYahoo!株式掲示板の投稿を用いて日経VI指数の上昇を予測する手法を提案した。トピックモデル及び機械学習を用いて予測モデルを構築した。モデルの妥当性を検証するためにボラティリティトレーディングシミュレーションを開発した。その結果、大幅な収益改善が確認できた。第二に政治的コミュニケーションとしてのTwitterを社会ネットワークおよび自然言語処理で分析した。その結果、アナウンスメント効果の可能性を見出すことができた。また各支持派がグループを形成しており、その間のリツイートは少ないことが分かった。つまりTwitterは意見改変の場ではなく、自分の意見を補強する場となっている可能性がある。

研究成果の学術的意義や社会的意義

ソーシャルメディアは一般市民や一般投資家の意見やセンチメントを表している。この情報は企業経営や社会に大きな影響がある。そこで本研究では、投稿が多くみられる投資と選挙に焦点を当てて2つのモデルを構築した。投資モデルにおいては、資本市場のリスク指標と投稿メッセージの関係を分析し、新たな投資意思決定モデルを開発した。今後、本研究をベースに投資リスク管理モデルが構築できれば、年金資産運用などに利用できる。選挙は事例研究を行った。その結果、支持者は同じグループ内でリツイートされていることも発見した。つまり、ソーシャルメディアは自分の意見を補強する場として利用されている可能性が高いことが分かった。

研究成果の概要(英文)：First, we proposed a method for predicting the increase of Nikkei VI using postings from Yahoo! Stock BBS. The prediction model was developed using topic model and machine learning. We developed volatility trading simulation to verify the validity of the model. As a result, we confirmed a significant improvement in profit. Second, we analyzed Twitter as a political communication using social networks and natural language processing. As a result, we might find announcement effects. Also, we found that each supporter formed a group, and there were few retweets between them. In other words, Twitter is not a place to change their own opinion, but to reinforce their opinion.

研究分野：経営情報

キーワード：ソーシャルメディア 投資意思決定 選挙行動 自然言語処理 機械学習 社会ネットワーク分析

1. 研究開始当初の背景

FaceBook、Twitter、掲示板などのソーシャルメディアは、一般市民や一般投資家の意見やセンチメントを表している。この情報の活用は企業経営や社会動向に成果を左右する可能性がある。そこで本研究では、ソーシャルメディア上の意見を企業経営にどのように活用すべきか、および選挙結果などの社会的意思決定にどのように影響するかの2点を検討する。第1に、資本市場のリスク指標とメッセージの関係を分析する。これにより、オプションなどの派生商品のプライシングや新たな投資意思決定モデルを開発できることが期待される。第2に、ソーシャルメディア上の選挙に関する投稿がある。これは一般市民からの投稿のみならず、政党や候補者の公式アカウントもある。これらの投稿メッセージが選挙とどのように関連しているかは、まだ十分に検討されていない。そこで、本研究では、選挙に関する投稿メッセージとそれを受信した市民がインターネット上でする行動を分析する。

2. 研究の目的

第1の投資意思決定モデルに関しては、本研究では株式市場のリスク指標に焦点を当てる。株式市場とSNS投稿メッセージに関する関係については、Antiweiler et al.(2004)、丸山ら(2009)など数多くの研究がある。先行研究から分かっていることは、SNSの投稿メッセージと出来高、ボラティリティは有意な相関が見られる、リターンに関しては統計的に有意である可能性はあるがその差は大きくはないということであった。そこで本研究では、リターンではなくリスク指標であるボラティリティの予測に着目する。ボラティリティの指標としては、大阪証券取引所に上場されている日経平均オプションから計算される日経ボラティリティインデックス(以下VIと呼ぶ)を用いる。VIは現在のオプション市場で見込まれている日経平均株価の1か月先までの変動率を示している。VIは恐怖指数とも呼ばれ、投資家の市場変動に対する恐怖を示していると言われる。Whaley(2000)は恐怖指数と呼ばれる理由について、VIが株価の恐怖のリスクと密接に関連している指数であることを14年間の市場分析により示している。つまりVIは投資家の株式市場に対する気持ち(センチメント)を表している可能性がある。このVIを予測することができれば、株式市場の大幅な変動等(暴落・暴騰)を予測できる可能性がある。つまり、VI予測モデルを構築できれば、有益な投資意思決定モデルを構築できる。

ソーシャルメディアとして日本の株式掲示板では日本最大のヤフー株式掲示板のデータを取得して用いる。株式掲示板のメッセージは大多数の一般のインターネット投資家の気持ちが投稿されている可能性がある。そこでこれらのメッセージを自然言語処理により分析し、現在投稿されている内容(トピック)を特定する。これを入力変数とし、目的変数にVIの変動を入力し、機械学習を用いてVI指上昇予測モデルを構築する。

第2のソーシャルメディアと選挙の分析に関しては、まだ分析事例の報告件数が少ないと考えられる。日本では、2013年より選挙へのソーシャルメディアが解禁された。7月に行われた第23回参議院選挙で佐藤ら(2014)はTwitterやFacebookといったSNSや、YouTubeといった動画サイト等の様々なメディアを用いたネット選挙が立候補者の選挙活動、また有権者にどのような影響を与えたのかという分析を行った。しかし、ネット選挙が有権者の行動に大きく影響しているという確証は得ることができなかったと述べている。そこで本研究では、調査対象とするSNSはTwitterに限定し、そのツイート・リツイートの動向と選挙結果の関係を分析し、アナウンスメント効果等の世論形成理論との比較を行う。

3. 研究の方法

3.1. VI 上昇予測モデル

本研究が提案する手法は、5つのステップから構成される。

(1) Yahoo!株式掲示板に投稿されたメッセージを形態素解析により形態素に分割する。掲示板の投稿メッセージのデータはヤフー株式会社より提供頂いた。形態素解析エンジンはIG0を用いた。

(2) 形態素を入力データとして、LDAを用いて各メッセージをトピックに分類する。各メッセージがどのようなトピックを意味しているのかを判断するために、Blei et al. (2003)のLDAトピックモデルを用いる。これは、「潜在的ディリクレ配分法」と呼ばれ、文書の確率的生成モデルである。各文書には潜在トピックがあると仮定し、統計的に共起しやすい単語の集合が生成される要因を、この潜在トピックという観測できない確率変数で定式化する。この手法を用いてメッセージを100種類のトピックに分類した。

(3) 分類された各トピックのメッセージ数の割合をトピック別投稿率として定義し算出する。分類したトピックと過去の日経平均株価データを用いてクラス分類を行った。説明変数として、トピック投稿数及び時系列トピックとVIを12種類を標準特徴量を設定した。これは総投稿数、VI、総投稿数の前日差、VIの前日差、総投稿数の前日比、VIの前日比、7取引日平均の総投稿数、7取引日平均のVI、当日の総投稿数と過去7取引日平均の総投稿数との差、当日のVIと過去7取引日平均のVIとの差、当日の総投稿数と過去7取引日平均の総投稿数との比、当日のVI指数と過去7取引日平均のVI指数との比である。

(4) トピック別投稿率を特徴量として、機械学習によるVIの上昇予測モデルを構築する。トピック投稿数及び時系列トピックと標準特徴量を組み合わせて表1のモデルを作成し、機械学習の特徴量として用いた。

表1 機械学習に用いた特徴量データ

データ1	トピック投稿数+標準特徴量
データ2	時系列トピック+標準特徴量
データ3	トピック投稿数の前日差+標準特徴量
データ4	時系列トピックの前日差+標準特徴量
データ5	トピック投稿数の前日比+標準特徴量
データ6	時系列トピックの前日比+標準特徴量
データ7	7取引日平均のトピック投稿数+標準特徴量
データ8	7取引日平均の時系列トピック+標準特徴量
データ9	当日のトピック投稿数と過去7取引日平均のトピック投稿数との差+標準特徴量
データ10	当日の時系列トピックと過去7取引日平均の時系列トピックとの差+標準特徴量
データ11	当日のトピック投稿数と過去7取引日平均のトピック投稿数との比+標準特徴量
データ12	当日の時系列トピックと過去7取引日平均の時系列トピックとの比+標準特徴量
データ13	全ての特徴量

「今後の7取引日間で、当日のVI 始値から2ポイント増加した日が含まれる」を目的変数として、ランダムフォレスト及びロジスティック回帰を用いたクラス推定を行なった。

(5) VI 上昇予測モデルの妥当性検証

機械学習により得られたモデルの妥当性を検証するために、日経平均オプションをバタフライスプレッドによる売買シミュレーションを行った。オプションを用いることにより、ボラティリティトレーディングが可能になる。これは、日経平均のボラティリティを予測して、その変動に合うような売買戦略を構築するものである。本研究のモデルはVI が大きく動くタイミングを予測するので、4つのオプションを組み合わせるショートバタフライ戦略を利用する。なお、売買は当日の始値を用いた。

3.2. ソーシャルメディアと選挙

分析手法は3ステップで行う。なお、Twitter メッセージはTwitter-API を用いて開発した。メッセージ収集に用いるキーワードは新聞記事より取得した。

(1) 時系列分析

Twitter の投稿量をツイートとリツイートそれぞれで測定した。支持者間でのツイート数の数・比率を時系列で分析し、世論調査の結果と比較することで、アナウンスメント効果などの投票者の行動を分析する。

(2) ツイートの投稿内容分析

ツイートを形態素解析及びLDAを用いてトピックに分類した。選挙の序盤、中盤、終盤でどのような内容が吐かれていたのかを分析する。

(3) リツイートの社会ネットワーク分析

リツイートを無方向性グラフと考えて、社会ネットワーク分析を行った。これにより、投票者の集団を区別し、どの集団がどのような影響を持っているかを分析する。

4. 研究成果

4.1. VI 上昇予測モデルの結果

機械学習の結果を表2に示す。この結果、ロジスティック回帰による全データを入力データとしたデータ13が適合率0.45及び、再現率0.45となり、実際に利用可能なモデルであると判断した。なお本研究では適合率ばかりでなく再現率も重視した。なお学習期間は2012年11月22日から2014年11月14日である。

表2 機械学習の結果

	ランダムフォレスト			ロジスティック回帰		
	適合率	再現率	データ数	適合率	再現率	データ数
データ1	0.444	0.042	4/9	0.000	0.000	0/0
データ2	0.000	0.000	0/2	0.235	0.768	73/311
データ3	0.500	0.032	3/6	0.250	0.537	51/204
データ4	0.000	0.000	0/2	0.236	0.779	74/314
データ5	0.667	0.021	2/3	0.439	0.263	25/57
データ6	0.143	0.011	1/7	0.393	0.232	22/56
データ7	0.444	0.042	4/9	0.000	0.000	0/0
データ8	0.000	0.000	0/2	0.235	0.768	73/311
データ9	0.667	0.063	6/9	0.233	0.505	48/206
データ10	0.000	0.000	0/2	0.236	0.779	74/314
データ11	0.500	0.011	1/2	0.421	0.253	24/57
データ12	0.286	0.021	2/7	0.414	0.253	24/58
データ13	0.237	0.095	9/38	0.453	0.453	43/95

このロジスティック回帰の結果をもとに、このVI 上昇予測モデルの有効性を売買シミュレーションにより検証する。ベンチマークとしての機械学習を用いない毎日取引と機械学習を用い

たVI上昇予測モデルによる売買指示の2つを比較した。その結果を表3に示す。なお、検証期間は2014年11月18日～2016年6月29日である。この結果、コールオプション及びプットオプションを利用したバタフライスプレッドの両方とも、機械学習の売買指示を用いれば大きな収益の改善が実現できたことが分かった。これにより本研究で提案したVI上昇予測モデルが有効である可能性を確認することができた。今後の課題としては、機械学習の手法を見直すことによる予測モデルの精度の一層の向上が必要である。

表3 シミュレーション結果

ケース	毎日取引		VI 上昇予測モデル	
	損益 (円)	売買回数	損益 (円)	売買回数
コールオプション	-3,926	279	536	71
プットオプション	-4,818	270	-779	75

4.2. ソーシャルメディアと選挙

2014年12月の衆議院選挙、2015年6月の大阪都構想の住民投票、2016年7月の平成28年度東京都知事選挙を分析した。分析結果は3つの事例で同様の結果が得られたので、ここでは2015年の大阪都構想の結果について述べる。

(1) 時系列分析

図1の賛成アカウント数と反対アカウント数の割合を示す。賛成アカウントは5月5日から5月10日まで少数派となっている。5月11日からは賛成派アカウントの割合が増加し、最終的には賛成59.6%、反対40.4%となり賛成多数という結果になった。これは住民投票とは異なる結果となった。アカウント数が反対優勢から賛成優勢に転じた5月11日は、新聞やニュースなどによる反対派優勢の世論調査結果のツイートが多数あり、前日の5月10日も同様のツイートが多くあった。賛成派不利を知ったユーザが5月11日にツイートをを行った可能性がある。このことから反対有利という報道が賛成派ユーザに対してアンダードッグ効果に似た効果を与えた可能性がある。また、反対アカウントは5月11日に減少してから減少傾向であり、反対アカウントにもアナウンスメント効果が影響している可能性がある。その後も反対アカウントが減少傾向であるのは、離脱効果が起きたためと考えることもできる。反対派有利という報道に安心し、Twitter上での活動を鈍らせた可能性がある。

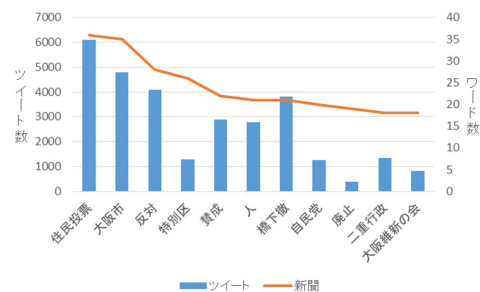
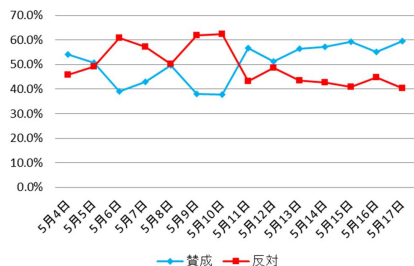


図1 賛成・反対アカウント数の比較

図2 トピック別ツイート数と新聞記事

(2) トピック分析

図2に新聞大手4社の記事の頻出単語とTwitterに投稿されたトピック数を示す。新聞記事で多く扱われている単語はツイートでも上でも多く出現することがわかる。ただ、「廃止」や「大阪維新の会」は他の単語と比較してみるとTwitterにおける出現数が少ない。さらに、トピック分析で賛成派の主張として存在した「既得権益」という単語は新聞記事には存在しなかった。

(3) 社会ネットワーク分析

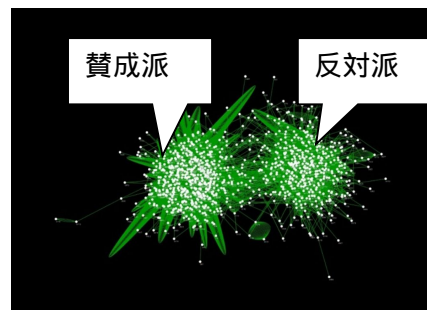
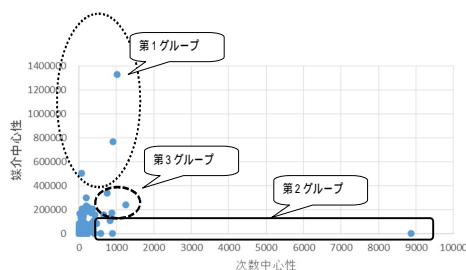


図3 リツイートの有力アカウントの抽出

図4 社会ネットワーク分析

図3にリツイートの社会ネットワーク分析における有力アカウントを示す。ごく少数のアカウントがリツイートを集めていることがわかる。媒介中心性が高い第1グループのツイート内容では、第一に単純な大阪都構想に関する情報を提供するのみのツイートである。第二に情報の提供に加え自分のコメントを加えたツイートである。また、自身の立場(賛成,または反対)を先に表明しているものも多く見られた。このようなタイプのユーザは自身の意見の賛同者を増やすことも目的の一つとしているのかもしれない。回数中心性が高い第2グループはユーザのツイートは「これは大阪市解体構想です。」といったような個人の反対意見が多かった。また賛成のアカウントとして維新の会等の団体が含まれていた。回数中心性,媒介中心性がともに高い第3グループには橋下徹氏のアカウントが含まれていた。

図4に社会ネットワーク分析を示す。明らかに2つのクラスターが存在し、賛成派・反対派に分かれている。ユーザは自分の意見に近いツイートしか見ないという選択的接触が行われている可能性が高い。回数中心性,媒介中心性が高く,かつツイート数が多かったアカウントも同じ意見を持つアカウントとしか接触していない可能性が高かった。故にTwitter上では意見の改変はなく,補強のみが行われている可能性がある。

5. 主な発表論文等

[雑誌論文](計 13件)

渡部和雄,梅原英一,岩崎邦彦,紙出版物利用者と紙・電子出版物利用者の意識や行動の定量分析、日本印刷学会誌、査読有、56巻、2019、印刷中

梅木寿人,中村優吾,藤本まなと,水本旭洋,諏訪博彦,荒川豊,安本慶一、混雑度の偏りを考慮した避難所決定手法、情報処理学会論文誌、査読有、60巻、2019、608-616

渡部和雄,岩崎邦彦,梅原英一、紙出版物利用者と紙・電子出版物利用者の意識や行動の定量分析、経営情報学会論文誌、査読有、27巻、2019、301-315

Maenaka,S., Suwa,H., Arakawa,Y., Yasumoto, K., Heart Rate Prediction for Easy Warking Route Planning、SICE Journal of Control, Measurement, and System Integration、査読有、11巻、2018、284-291

DOI: 10.9746/jcmsi.11.284

Kobayashi,T., Ogawa,Y., Suzuki,T., Yamamoto, H., News audience fragmentation in the Japanese Twittersphere、Asian Journal of Communication、査読有、29巻、2018、274-290

DOI: 0.1080/01292986.2018.1458326

Maenaka, S., Suwa, H., Arakawa, Y., Yasumot, K., Heart Rate Prediction for Easy Waking Route Planning、SICE Journal of Control, Measurement, and System Integration、査読有、11巻4号、2018、284-291

神辺圭一,諏訪博彦,篠田孝祐,栗原聡、共同利用施設における実験終了後の研究成果数予測、情報処理学会論文誌数理モデル化と応用、査読有、11巻、2018、1-11

Luhanga,E.T., Elder Akpa A.H, Suwa,H., Arakawa, Y., Yasumoto, K., Identifying and Evaluating User Requirements for Smartphone Group Fitness Application、IEEE Access、査読有、6巻、2018、3256-3269

DOI: 0.1109/ACCESS.2018.2793844

荒川周造,諏訪博彦,小川祐樹,荒川豊,安本慶一,太田敏澄、暗黙知に基づく飲食店向け不動産賃料推定モデルの提案、情報処理学会論文誌、査読有、59巻、2018、33-42

山縣史哉,梅原英一、平成28年度東京都知事選挙のTwitter分析、電子情報通信学会技術報告書、査読無、117巻464号、2018、61-66

岩本茂子,小川祐樹,諏訪博彦,太田敏澄、つぶやきトピックによる残業意欲の分析、電子情報通信学会論文誌D、査読有、J100-D巻、2017、760-772

井上聖吾,梅原英一、2014年12月の衆議院選挙のTwitter分析、東京都市大学情報メディアジャーナル、査読無、17巻、2016、7-17

DOI: 10.9746/jcmsi.11.284

渡部和雄,梅原英一,岩崎邦彦、消費者のO2O(Online to Offline)行動の差異に基づいた消費者特性の分析と実店舗への誘導への示唆、2016、情報処理学会論文誌、査読有、57巻、1887-1897

[学会発表](計 9件)

佐々木皓大,梅原英一,諏訪博彦,小川祐樹,山下達雄,坪内孝太、LDAを用いた株式掲示板の投稿メッセージによる恐怖指数上昇予測の提案、第22回人工知能学 金融情報学研究会、2019

Sasaki,K., Hirose,Y., Umehara,E., Suwa,H., Ogawa,Y., Yamashita,T., Tsubouchi,K., Simulation of Volatility Trading using Nikkei Stock Index Option based on Stock Bulletin Board、The 3rd International Workshop on Application of Big Data for Computational Social Science in IEEE BigData2018、2018

高野雅典,高史明,森下壮一郎,西朋里,小川祐樹、ニュースを起点とするレイシズム表出におけるニュース番組の性質と個人の性質の関連:インターネットテレビに投稿される差別的コメントの分析、第3回計算社会科学ワークショップ、2018

廣瀬由衣, 梅原英一, 株式掲示板の分析結果に基づく売買指示に従う日経平均オプション取引のシミュレータの開発、第 24 回社会情報システム学シンポジウム、2018

高野雅典, 森下壮一郎, 小川祐樹, インターネットテレビ局 AbemaTV におけるニュースの「チラ見」がニュース視聴の動機に与える影響、日本行動計量学会 第 46 回大会、2018

Suwa, H., Ogawa, Y., Umehara, E., Kakigi, K., Yasumoto, K., Yamashita, T., Tsubouchi, K., Develop Method to Predict the Increase in the Nikkei VI index, the 2nd International Workshop on Application of Big Data for Computational Social Science、2017

Hayashi, k., Umehara, E., Ogawa, Y., Analysis of Twitter Messages about the Osaka Metropolis Plan in Japan, the 2nd International Workshop on Application of Big Data for Computational Social Science、2017

小川祐樹, 高史明, 鳥海不二夫, Twitter ユーザのパーソナリティ推定に基づく炎上現象の分析、第 2 回計算社会科学ワークショップ、2017

柿木研人, 諏訪博彦, 小川祐樹, 梅原英一, 安本慶一, インターネット株式掲示板における話題と株式指標の関係、マルチメディア、分散、協調とモバイル(DICOM02016)シンポジウム、2016

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

6. 研究組織

(1)研究分担者

研究分担者氏名：諏訪 博彦

ローマ字氏名：(SUWA, hirohiko)

所属研究機関名：奈良先端科学技術大学院大学

部局名：先端科学技術研究科

職名：特任准教授

研究者番号(8桁): 70447580

(2)研究分担者

研究分担者氏名：小川 祐樹

ローマ字氏名：(OGAWA, yuuki)

所属研究機関名：立命館大学

部局名：情報理工学部

職名：助教

研究者番号(8桁): 40625985

(3)研究分担者

研究分担者氏名：渡部 和雄

ローマ字氏名：(WATABE, kazuo)

所属研究機関名：東京都市大学

部局名：知識工学部

職名：教授

研究者番号(8桁): 90244532

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、研究者個人に帰属されます。