

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 8 日現在

機関番号：32678

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25380481

研究課題名(和文) ソーシャルメディアと株式市場の関連性に関する研究

研究課題名(英文) Study on relations between social media and stock markets

研究代表者

梅原 英一 (Umehara, Eiichi)

東京都市大学・メディア学部・教授

研究者番号：00645426

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：本研究ではYahoo!株式掲示板と株式市場の関係を分析した。第1に株式リターンを説明する新たなファクターを発見した。自然言語処理を用いて単語出現頻度(TF/IDF)を計算し、上位700単語の主成分分析でファクターを抽出した。その結果、投資戦略・株式保有・金融経済の3ファクターを抽出した。第2に日経平均VI指数(恐怖指数)との関係を分析した。掲示板にトピック分析(LDA)を行った。トピック投稿率とVI指数に相関があることが分かった。本研究で開発した手法を用いてTwitterとTV深夜アニメ番組の次週視聴率を分析し、投稿数・実況ツイートのネガティブ単語数と有意な相関があることが分かった。

研究成果の概要(英文)：We analyzed the relationship of Yahoo! stock BBS and the stock market. The First, I found a new factor for explaining the stock returns. We calculated the term frequency (TF / IDF) using a natural language processing, and extracted factors using a principal component analysis. As a result, we could extract 3 factors (the investment strategy, the stock holdings, and the financial economy). The second, we analyzed the relationship between Nikkei Stock Average Volatility Index (VIX) and Yahoo! Stock BBS. We applied a topics analysis (Latent Dirichlet Allocations) to messages in the stock BBS. We found a correlation between the posted rate for each topic and VIX. The Third, we analyzed a relation between Twitter and Viewing rate in midnight TV anime programs using our method developed in this study. We found that there is a significant correlation between the number of messages posted on Air and the number of negative words in the tweets on Air, and next week viewing rates.

研究分野：経営情報

キーワード：ソーシャルメディア 株式市場 自然言語処理 機械学習 経営情報

1. 研究開始当初の背景

ソーシャルメディアにより一般のユーザーが自分の意見や気持ちを容易に公表することが可能になった。このうちの一つに株式投資に関心がある人々が投稿する株式掲示板がある。先行研究()により掲示板の投稿メッセージと株式リターン、ヒストリカルボラティリティ、出来高は相関関係があることが判明している。

<引用文献>

インターネット株式掲示板の投稿内容と株式指標の関係, 丸山健, 梅原英一, 諏訪博彦, 太田敏澄, 証券アナリストジャーナル 11月・12月合併号, pp.110-127, 2008.

2. 研究の目的

インターネット株式掲示板などのソーシャルメディアに投稿されたメッセージは、投資家の意見やセンチメントを表している。これの株式市場との関係进行分析することで、新たな投資意思決定モデルや新たな経営情報としての指標を開発できることが期待される。また、形態素解析・機械学習の辞書を改善することにより分析精度の向上を目指す。これにより新たなモデルや経営情報に活用できる指標や手法を開発し提言する。

3. 研究の方法

ソーシャルメディアと株式市場に関するモデルを、仮説を設定して分析を行う。また自然言語処理の分析精度の向上を行う。仮説は我々の先行研究で得た知見、ファイナンス分野の先行研究のモデルおよび経済物理学等で提案されている仮説をもとに開発する。その検証のために、クローリングを用いてインターネット上のメッセージを取得し、形態素解析および機械学習により分析し、指標を作成する。それを株価データ(株価、出来高等)との関係につき統計的に解析する。これにより、ソーシャルメディアが株式市場に与える影響に関する新たなモデルを提案する。

また、本研究で開発した自然言語処理このモデルに基づき新規に開発する手法やソーシャルメディア指標を、経営情報や社会情報として有効に活用する方策を考察する。

4. 研究成果

(1) 株式リターンやボラティリティなどの株式指標を予測することは有用である。本研究では、恐怖指数と呼ばれる投資家の不安を表す株式指標である VI 指数に着目した。VI 指数は投資家の不安心理と結びついた指標である。投資家のセンチメントを表していると考えられる Yahoo! 株式掲示板に投稿されたメッセージにより VI 指数の予測ができるか検討した。これにより、将来のリスク回避、及びボラティリティトレーディングなどの投資活動に役立てられると考えられる。

分析手順を図 1 に示す。株式掲示板からクローリングによってメッセージを取得した。2012 年 1 月 30 日から 2013 年 3 月 31 日までの投稿数上位 50 社のメッセージ 880,321 件を取得した。これに対し形態素解析を行った。日本語以外や否定語処理を行った結果、残った投稿数は、845,400 件であった。話題を抽出するために、LDA トピックモデルを用いた。トピック数は 80 に設定した。トピック投稿数は 1,347,150 件となった。これは 1 つの投稿が 2 トピック以上に割り当てられたためである。分類された時系列の各トピックと VI 指数との関係を検証するため、日別のトピックごとの投稿数および総投稿数に占めるそのトピックの投稿数の割合(トピック投稿率)との相関分析を行った。さらに、トピック投稿率を用いた重回帰分析を用いて VI 指数の予測を行った。

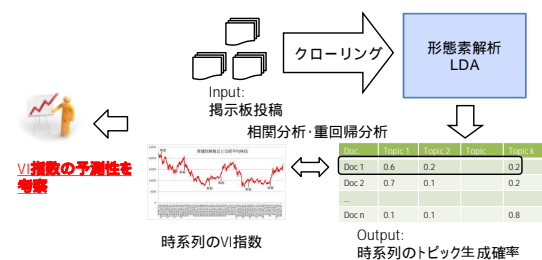


図 1 分析手法

日別のトピックごとの投稿率と前日、同日、翌日の VI 指数を相関分析した結果を表 1、表 2 に示す。これは、それぞれの時系列で VI 指数との相関係数が 1% 有意であったトピックの中で、相関係数が正、負それぞれ上位 5 件のトピックを抽出した。

VI 指数と正の相関を持つトピックを構成する単語と負の相関を持つトピックを構成する単語の傾向が異なっていることが分かった。正の相関を持つ単語群は、「日経、ドル、円安、銘柄」といった、株式投資についてのものであった。一方で、負の相関を持つ単語群は「アホ、お前、韓国、カス」といった、株式銘柄に関係のないと考えられる投稿の単語で構成されていた。

VI 指数は、相場が安定な時は低下し、相場の変動に対応して上昇する指数であるため、VI 指数との相関係数の正負は正が相場の変動、負が相場の安定と対応している。これにより、次の仮説が示唆された。

「インターネット掲示板において、株式銘柄についての投稿と株式銘柄に関係のない投稿が混在しており、VI 指数が上昇するとき、すなわち相場が変動するときには株式銘柄についての議論が活発になり、投稿率が上昇する。一方で相場が安定するときには株式銘柄についての議論は多くなされず、株式銘柄に関係のない投稿の投稿率が上昇する」

この仮説について検証するため、VI 指数との相関係数が正、負それぞれの上位 3 トピックの日別の投稿数の合計と VI 指数の関係を

調べた。VI 指数と正の相関のあるトピックは、VI 指数の上昇に応じて投稿数が上昇していた。一方で、VI 指数と負の相関のあるトピックについては、VI 指数が安定している時期に突発的なピークを迎えている。このような投稿は「煽り」投稿の可能性が考えられる。投資に関係のある投稿と、投資に関係のない、「煽り」投稿の比率が VI 指数と関連を持っていることが示唆された。煽り投稿検出技術を利用したフィルタリングを事前に行うことにより、VI 指数の予測精度を向上させられる可能性があることが分かった。

表1 VI 指数と正の相関を持つトピック

トピック単語	前日	同日	翌日
日経平均 為替	.412	.386	.333
株価	.411	.383	.333
銘柄	.403	.389	.394
ゲーム	.393	.383	
銀行 金融	.316		.325
医療		.348	.311

表2 VI 指数と負の相関を持つトピック

トピック単語	前日	同日	翌日
ユニ- アイフル シャープ	-.402	-.407	-.404
嫌い カス	-.392	-.373	-.398
日本 韓国 朝鮮	-.354	-.355	-.358
政治	-.313	-.321	-.333
通信 携帯	-.306		
中国 尖閣 沖縄		-.298	-.365

次に、日別のトピックごとの投稿数と前日、同日、翌日の VI 指数で重回帰分析を行った。時系列による大きな差は見られなかった。これは、株式掲示板で投資家が話題にするトピックは、一日単位で変化するのではなく数日間わたる継続性がある可能性が考えられる。また VI 指数と 0.606 の有意な重相関係数を持つことが分かった。同一データを対象に先行研究[1]と同様の手法で機械学習を行って強気/弱気投稿数と VI 指数の相関係数は 0.416 であった。この結果、従来手法と比較して、本研究の提案手法が VI 指数予測手法として有用であると考えられる。

(2) 本研究は、資本資産評価モデル (CAPM) において株式掲示板によるアノマリー (超過リターン) が存在するかどうかを検証する。CAPM に従うとポートフォリオの超過リターンはポートフォリオの市場感応度 () とマーケットリターンにより表すことができる。CAPM で説明できないリターンがアノマリーと呼ばれる。投資家は数値情報だけでなくニュースなどの文字情報も参考にして投資

判断を下す。近年はインターネット株式掲示板が発達しており他の投資家の声を蓄積している。我々の先行研究では投稿内容の強気・弱気に基づく強気指数によるファクターを開発した。しかし、強気指数だけではまだアノマリーを十分に説明できなかった。そこで本研究では強気指数以外に掲示板指標としてアノマリーを説明できるファクターが存在するか調査した。

分析対象は Yahoo! 株式掲示板を用いた。対象銘柄は東証 1 部上場株式である。Yahoo! 株式掲示板からクローラーにより投稿メッセージを取得した。取得したメッセージを形態素解析にかけて単語を抽出し、単語ごとの TF/IDF 値を算出した。単語出現頻度が 1000 回以上に絞り、TF/IDF 値の上位 700 単語を抽出した。この 700 単語に対して主成分分析を行った。主成分負荷量が大きい単語を調べることで、その主成分がどのような意味を持っているかを決定した。

アノマリーの存在を見るために東証 1 部全銘柄を各銘柄の主成分得点に従い表 3 の 6 つのポートフォリオに分けた。

表3 ポートフォリオ構築

ポートフォリオ	主成分得点
PF1	上位 20%
PF2	20%-40%
PF3	40%-60%
PF4	60%-80%
PF5	下位 20%
PFx	投稿なし

PF1 から PF5 の 5 ポートフォリオに対して CAPM に従い超過リターンが存在するか統計的な検定を行った。主成分分析で寄与率の高い上位 10 成分の主成分得点と我々の先行研究に基づく強気指数との相関を表 4 示す。この結果、主成分 1, 2, 7, 8 とリターンは相関があり、主成分 4, 5, 6, 7, 10 は強気指数と相関があることが分かった。

表4 主成分得点とリターン・強気指数

	Return	Bullishness
PCS1	0.0223**	0.0025
PCS2	0.0175**	-0.0036
PCS3	0.0065	0.0075
PCS4	-0.0032	-0.0362**
PCS5	0.0061	-0.0374**
PCS6	-0.0035	-0.0315**
PCS7	0.0126*	0.0208**
PCS8	0.0194**	-0.0009
PCS9	-0.0097	0.0025
PCS10	-0.0020	-0.0223**

* p<0.05, ** p<0.01

次に主成分1を利用した5つのポートフォリオを作成し、CAPMで超過リターンが観測できるか調査した。この結果、PF1,2,5において超過リターンが検出された。また主成分得点が高いポートフォリオほど超過リターンが高くなり、低いポートフォリオはマイナスであることが分かった。

さらにPF1とPF5のリターンの差を新たなファクターと定義し、CAPMにファクターを変数として追加したマルチファクターモデルを構築した。その結果、統計的に有意な超過リターンは観測されなかった。つまり、この新たなファクターはリターンを説明する可能性があることが分かった。

表5 マルチファクターモデルの有効性

	CAPM	Multi-Factor Model
PCS1		×
PCS2		
PCS3		
PCS4	×	-
PCS5	×	-
PCS6	×	-
PCS7	×	-
PCS8		×
PCS9		×
PCS10	×	-

○: Portfolio with the abnormal return exists.
 ×: Portfolio with the abnormal return does not exist.
 -: It was not analyzed, because abnormal returns did not exist in CAPM.

この手順を他の主成分に対しても行った。この結果、主成分1、8、9は強気指数とは別の、株式リターンを説明できる新たな掲示板ファクターである可能性があることが分かった。主成分1に含まれる単語は「ぴん、利益狙い、買値」などの株式売買に関する単語、主成分8は「経済、金融、投資」などマクロ経済や金融関係の単語、主成分9は「保有割合、証券取引、株券」などの株式保有に関する単語であった。

(3) 最近では、ツイートの内容に着目し、テレビ番組の視聴率の指数にツイッターを取り入れる動きまでもみられている。ジャンルごとの特性も分かっており、1分あたりのツイート件数平均で最も多いジャンルはアニメである。そこで本研究ではアニメ番組に関するツイートについて分析を行う。ツイート数、ツイートから得られる社会ネットワーク、ツイート内容の3つと視聴率の関係性を明らかにする。

深夜アニメのクールの切り替わりである2014年7月3日から、10話が終わる2014年9月14日までとした。収集対象テレビ番組は11タイトルである。なお情報の正確性を高めるため公式推奨ハッシュタグを含まないツイートは取得しないものとした。取得した総ツイート数は1,598,593件であった。その内、リツイートとリプライは1,108,615件、ツイ

ートは489,978件であった。

ツイート数と視聴率の関係は表6である。この結果、視聴者によるツイッターへの投稿が多い深夜アニメは、視聴率も高いと言える。さらに各話のツイート数は、翌週の視聴率とも正の相関がみられた。これにより、アニメ番組放送中に投稿されたツイート数は、翌週の番組視聴率の予測指数となる可能性が得られた。

表6 ツイート数と視聴率の相関分析

	相関係数	p値	結果
(1) 同時指標	.4226**	.0000	1%有意
(2) 遅行指標	.3825**	.0001	1%有意
(3) 先行指標	.4396**	.0000	1%有意

** : 1%有意

第2にリツイート・リプライを無方向性のグラフと考えて社会ネットワークを構築し、各番組の平均視聴率と各番組の次数中心性・媒介中心性のクラスター係数で相関をとった。次数中心性は1%有意な正の相関が見られた。つまり、リツイートやリプライに関する次数中心性のクラスター係数の値が大きいアニメ番組ほど、平均視聴率が高いという関係が成立する可能性がある。この結果、「視聴率の高い番組は、番組公式アカウントへの興味が深い」という仮説が成立する。しかし媒介中心性に関しては視聴率と有意な関係はなかった。つまり、深夜アニメにおいては、非公式アカウントへのリツイート数と視聴率には関係がない可能性がある。つまり平均視聴率は、番組公式アカウントへの注目度から得られる次数中心性と強い関係があることが明らかとなった。

第3に投稿内容を分析するために番組放送中のツイートに対して形態素解析をおこなった。「かわいい」や「おもしろい」などのプラス感情の単語には1、「死ぬ」「嫉妬」などのマイナス感情の単語は-1、キャラクター名などのプラス感情とマイナス感情どちらにも分類されないものは中立とし0を割り付けた。最後にプラス感情とマイナス感情を算出した結果と、それぞれの次週の視聴率との相関を見た。

この結果、プラス感情と評価した単語の個数と視聴率は有意な相関は見られなかった。「かわいい」「良い」といったプラス感情の単語は、ほぼ全ての番組で共通して常に形態素解析の上位に食い込んでいたため、次週の視聴率の変動に影響しづらかった可能性がある。しかしながら、マイナス感情と評価した単語の個数と視聴率で相関分析をした結果は、5%有意で正の相関が見られた。「死ぬ」「酷い」といったマイナス感情の単語は、形態素解析の上位に入ることが少なかった。そのため、常に上位にあるプラス感情の単語をマイナス感情の単語が追い抜くような回では、マイナス感情で煽ることによって一種のスリルを味わわせ、次回への期待を高めてい

ると考えられる。それにより、必然的に次週の視聴率への影響が大きくなった可能性がある。

最後に、実況ツイートに着目して、投稿されたツイート数、マイナス感情を持つ単語の出現数の2変数を独立変数、次回視聴率を従属変数とし、ステップワイズ重回帰分析を行った。その結果、この2変数が抽出され、決定係数は.227で有意となった。故に、この実況ツイートに関する2つの変数は視聴率の予測に利用できる可能性がある。

5. 主な発表論文等

[雑誌論文](計 15 件)

井上聖吾, 梅原英一: 2014年12月の衆議院選挙のTwitter分析, 査読無, 情報メディアジャーナル, Vol.17, No.1, pp.7-17, 2016.

原田彩加, 梅原英一: プレーヤの発話量及び行動量による協力ゲーム(禁断の島)の達成度の相違分析, 査読無, 電子情報通信学会技術報告書, Vol.115, No.486, pp.187-192, 2016.

相川実穂, 仙海かおり, 安田智亜紀, 酒井啓江, 小川祐樹, 梅原英一: 深夜アニメ番組に関するつぶやきに着目したTwitter分析, 査読有, 第14回情報科学技術フォーラム, Vol.4, pp.85-92, 2015.

K. Miyata, H. Yamamoto, Y. Ogawa: What Affects the Spiral of Silence and the Hard Core on Twitter? An Analysis of the Nuclear Power Issue in Japan, 査読有, American Behavioral Scientist, Vol.59, No.9, pp.1129-1141, 2015.

真継りまこ, 梅原英一: 感性工学による好まれるイラストのデザイン分析, 査読無, 情報メディアジャーナル, Vol.16, No.1, pp.8-18, 2015.

H. Suwa, E. Umehara, T. Ohta: Stock BBS Factor Model using Principal Component Score, 査読有, Advances in Computational Social Science, Agent based Social System, Springer, Vol.11, pp.3-15, 2014.
DOI:10.1007/978-4-431-54847-8_1.

K. Watabe, K. IWASAKI, E. Umehara: Factors and Models for Promoting Consumer Use of Electronic Money, 査読有, International Journal of Japan Association for Management Systems, Vol.6, No.1, pp.7-14, 2014.

渡部和雄, 岩崎邦彦, 梅原英一: 電子マ

ネー利用の規定要因と利用頻度・利用意向のモデル化による示唆, 査読有, 経営情報学会誌, Vol.23, No.2, pp.105-119, 2014.

堺里美, 佐藤香, 梅原英一: 人生ゲームで投資シミュレーション学習, 査読無, 情報メディアジャーナル, Vol.15, No.1, pp.52-59, 2014.

加藤菜美絵, 諏訪博彦, 太田敏澄: 企業内SNS導入に関する利用者調査, 査読有, 情報処理学会論文誌, Vol.55, No.1, pp.221-229, 2014.

小川祐樹, 山本仁志, 宮田加久子: Twitterにおける意見の多数派認知とパーソナルネットワークの同質性が発言に与える影響 - 原子力発電を争点としTwitter上での沈黙の螺旋理論の検証 -, 査読有, 人工知能学会論文誌, Vol.29, No.5, pp.483-492, 2014.

梅原英一, 渡部和雄: 電子マネー普及に関するシミュレーションモデルの構築, 査読無, 情報メディアジャーナル, Vol.14, No.1, pp.84-92, 2013.

三浦大樹, 諏訪博彦, 鳥海不二夫, 鬼塚真: ソーシャルサーチのための効率的な検索アルゴリズムの提案, 査読有, 情報処理学会論文誌データベース, Vol.6, No.3, pp.29-39, 2013.

大野光太郎, 小川祐樹, 諏訪博彦, 太田敏澄: 東京消防庁における消防活動経験の伝承を支援するSNSの提案, 査読有, 情報処理学会論文誌, Vol.54, No1, pp.284-294, 2013.

平澤真大, 小川祐樹, 諏訪博彦, 太田敏澄: ニコニコ動画のログデータに基づくソーシャルノベルティのある動画の発見手法の提案, 査読有, 情報処理学会論文誌, Vol.54, No1, pp.214-222, 2013.

[学会発表](計 14 件)

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, 協調とモバイルシンポジウム, 鳥羽シーサイドホテル(三重), 2016年7月6日 7日

原田彩加, 梅原英一: 協力ボードゲームにおけるチーム成功度と参加者の行動の関連性分析 スマホを用いた社会物理学実験, 第22回社会情報システム学シンポジウム, 電気通信大学(東京), 2016年1月21日

大和田拓也、梅原英一：首都直下地震の
徒歩帰宅者トイレに関するエージェ
ン・シミュレーション，第22回社会情
報システム学シンポジウム，電気通信大
学（東京），2016年1月21日

梅原英一：マルチエージェントシミュレ
ーションによる地域電子マネーの普及
条件分析，経営情報学会 2015 年秋季全
国研究発表大会，沖縄コンベンションセ
ンター（沖縄），2015年11月28日 - 29
日

Y. Ogawa, T. Kobayashi, H. Yamamoto,
T. Suzuki: Estimating Media
Partisanship from Twitter data: A case
in Japan, World Association for Public
Opinion Research (WAPOR) 68th Annual
Conference, Buenos
Aires(Argentina) ,June 16-19 2015.

Y. Ogawa, T. Kobayashi, H. Yamamoto,
T. Suzuki: News Audience
Fragmentation in Japanese
Twittersphere, International
Communication Association (ICA) 65th
Annual Conference, Preconference: New
Media and Citizenship in Asia: Civic
Engagement for Sustainable
Development Across the Life Span, San
Juan (PuertoRico), May 21-25 2015.

A. Ito, Y. Arakawa, H. Suwa, A. Fukuda:
Region-to-Region Similarity Analysis
based on Foursquare Venue Database,
ACM HotMobile 2015, Santa Fe(USA),
Feb. 12-13 2015.

Y.Ogawa, H.Yamamoto, K.Miyata, and
K.Ikeda: Association between
Selective Exposure and Attitude on
Twitter, The 2014 IEEE/WIC/ACM
International Conference on
Intelligent Agent Technology (IAT'14),
Warsaw (Poland), Aug. 11-14 2014.

K. Ikeda, Y. Okada, F. Toriumi, T.
Sakaki, K. Kazama, I. Noda, K. Shinoda,
H. Suwa, S. Kurihara: Multi-Agent
Information Diffusion Model for
Twitter, The 2014 IEEE/WIC/ACM
International Conference on Web
Intelligence, pp.21-26, Warsaw
(Poland), Aug. 11-14 2014.

Y.Ogawa, H.Yamamoto, K.Miyata, and
K.Ikeda: Association between
Selective Exposure and Attitude on
Twitter, The 2014 IEEE/WIC/ACM

International Conference on
Intelligent Agent Technology (IAT'14),
Warsaw (Poland), Aug. 11-14 2014.

渡部和雄,岩崎邦彦,梅原英一：電子マネ
ー利用意向の規定要因および規定要因
間の関係の実証分析,第12回情報科学技
術フォーラム,鳥取大学（鳥取）,2013
年9月4日 6日.

梅原英一、諏訪博彦：日中足株価を用い
た株式掲示板と株式市場の関係：
SoftBank の事例,第12回科学技術フォ
ーラム,鳥取大学（鳥取）,2013年9月4
日 6日.

梅原英一,渡部和雄：電子マネー普及の
差異：マルチ・エージェント・シミュレ
ーションによるアプローチ,経営情報
学会 2013 年春季全国大会,慶応大学（東
京）,2013年6月29日 30日.

E. Umehara, H. Suwa: Relationship
between Stock BBS and Stock Market
using Intra-day Stock Prices: Case of
SoftBank, Econophysics Colloquium
2013 & Asia Pacific Econophysics
Conference (APEC) 2013, POSTEC
(Korea), July. 29-31 2013.

6. 研究組織

(1)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

(2)研究分担者

諏訪 博彦 (SUWA, Hirohiko)
奈良先端科学技術大学院大学・情報科学研究
科・助教
研究者番号：70447580

(3) 研究分担者

小川 裕樹 (OGAWA, Yuki)
立命館大学・情報理工学部・助教
研究者番号：40655985