

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

柿木研人¹ 諏訪博彦¹ 小川祐樹² 梅原英一³ 安本慶一¹

概要：ファイナンスにおいて、株式リターンや、ボラティリティだけでなく、様々な株式指標を予測することは有用である。我々は、恐怖指数と呼ばれ、投資家の不安を表す株式指標である VI 指数に着目し、ソーシャルメディアを分析することによって得られる投資家の情報と密接に関連のある指標であるという仮説を立てる。従来手法以上にソーシャルメディアの情報から投資家の感じる恐慌リスクに関する投稿を高い精度で獲得する方法として、トピックモデルによる分類を提案する。我々は、VI 指数と話題の関係を明らかにする実験として Yahoo! JAPAN の株式掲示板から取得したメッセージをトピック分類し、VI 指数と相関分析、及び重回帰分析を行った。結果として、トピックの投稿率の変動による VI 指数の予測可能性を示した。

The relationship between stock index and topic in Internet stock forum

KENTO KAKIGI¹ HIROHIKO SUWA¹ YUKI OGAWA²
EIICHI UMEHARA³ KEIICHI YASUMOTO¹

1. はじめに

ファイナンスにおいて、株式リターンだけでなく、出来高やボラティリティ、その他の金融指標などを予測することは重要であり、これまで多くの研究がなされている。例えば、ニューラルネットワークを構築して翌日の市場動向を予測した西谷ら[1]や、経済レポートをテキストマイニングすることで国債の長期的動向を予測した和泉ら[2]が挙げられる。本研究では、恐怖指数と呼ばれ、投資家の不安を表す株式指標である VI 指数に着目する。VI 指数は投資家の不安を表す株式指標で株式市場における恐慌のリスクと密接に関連のある指数である[3]。そのため投資判断の材料として重要な指標である。

我々は、ソーシャルメディアには人々の気持ちが投稿されているという前提に基づいて、ソーシャルメディアを用いて VI 指数を予測する手法を提案する。ソーシャルメディアと株式指標の関連を分析した研究として、日本の株取引掲示板を分析した丸山ら[4]、諏訪ら[5]や、Twitter を分析した Bollen ら[7]や Hong ら[8]の研究が挙げられる。これらの先行研究では、ソーシャルメディアの投稿内容の分類は強気・弱気分類、及び 6 種類の感情分類を行っている。しかし、従来手法は VI 指数と対応し

ていると考えられる市場全体の恐慌のリスクに関する投稿を獲得するためには適切でないと考えられる。

我々は、経済、金融専門の辞書を作成した和田ら[6]の研究で、単語の意味を軸とした分類の有用性が示されたことを受けて、投資家の感じる恐慌のリスクを取得するためにより適した手法として株取引に関するソーシャルメディア専用のトピックモデルを作成する。これにより、恐慌と対応して活発になるトピックを用いて VI 指数の予測ができると考えられる。

本研究では、VI 指数を予測する手法を提案することを目的とする。その中で、どのような話題が VI 指数の予測に有効であるか明らかにする。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。

実証実験のためのデータは、日本最大の株取引に関するソーシャルメディアである Yahoo! JAPAN の株式掲示板から取得した。取得したメッセージをトピックモデルにより分類し、時系列のトピックごとの投稿数を獲得した。これと、VI 指数を相関分析することにより、VI 指数と正負の相関を持つ話題を特定した。また、VI 指数を予測するために重回帰分析を行い、トピックの投稿率の変動による VI 指数の予測可能性を示した。

2. 関連研究

本章では、株式予測に関する研究のレビューを行い、本研究で分析対象とする株式指標である VI 指数のソーシャルメディアを用いた予測可能性について説明する。

¹ 奈良先端科学技術大学院大学
NAIST
² 立命館大学
RITSUMEIKAN UNIVERSITY
³ 東京都市大学
TOKYO CITY UNIVERSITY

そしてソーシャルメディア分析、及び株式指標との関係に注目している研究についてレビューを行い、本研究の目的を述べる。

2.1 株式指標予測に関する研究

ファイナンスにおいて、株式リターンや、ボラティリティだけでなく、様々な株式指標を予測することは、リターン獲得や、リスク回避につながるため有益である。そのため、様々なアプローチによってこれを実現するための研究がなされている。予測の対象とする株式指標ごとに、それに適した分析の手法は異なる。西谷ら[1]は、翌日の株式市場の動向を予測するために、前日始値、終値、高値、安値、出来高を用いたニューラルネットワークを構築し、シミュレーションによって株価変動に対応した投資行動の支援を行える可能性を示した。和泉ら[2]は、長期的な市場の動向を予測するために、金融機関の発行する経済レポートをテキストマイニングし、日本国債を対象とした運用テストにおいて運用益を上げ、市場動向の説明性を示した。

我々は、リスク回避指数として注目されている VI 指数に着目する。VI 指数は、一般に恐怖指数と呼ばれ投資家の不安心理と結びついた指標であるとされている。Robert[3]は VI 指標が恐怖指数と呼ばれる理由について、VI 指数が株式市場における恐慌のリスクと密接に関連のある指数であることを 14 年分の市場分析により説明している。VI 指数を予測することは、恐慌リスクを予測することにつながり、リスク回避、及びボラティリティトレーディングなど投資活動に役立てられると考える。

なお、VI 指数は満期を T 、リスク中立確率測度による期待値を E^* 、ボラティリティを v として以下の式により定義される株式指標である。

$$VIX_T = \sqrt{E^* \left[\frac{1}{T} \int_0^T v^2 \right]}$$

この VI 指数を予測するためにはどのような手法が必要であろうか。

VI 指数は、その性質から、投資家の感情と密接に関連のある指標であることが考えられる。近年ソーシャルメディアの発展に伴い、特に株取引に関するソーシャルメディアにおいて、投資家たちは日々メッセージを発信している。これらのメッセージには投資家の感情が含まれていると考えられる。そのためメッセージを分析することによって、投資家の感情と関連の有る株式指数である VI 指数を予測できる可能性がある。

2.2 ソーシャルメディアと株式指標の関係

投資家の感情をソーシャルメディアから抽出し、利用しようとする研究はいくつか存在する。丸山ら[4]は Yahoo! 株式掲示板の銘柄毎の投稿について、自然言語処

理と機械学習による投稿内容の分析を行い、投稿の“強気”“弱気”を判定した。その結果、投稿数がボラティリティ、出来高の先行指標であり、強気投稿数と弱気投稿数の差が個別銘柄における株式のリターンやボラティリティ、出来高と関連している可能性があることを示した。

諏訪ら[5]は、[4]の結果を受け、市場全体を分析対象とすることを目的として、東証 1 部における投稿数および強気指数によるポートフォリオを構築し Fama-French の 3 ファクターモデルが成立しているかを検証している。結果として、強気指数が市場全体で株価リターンと関係している可能性があることを示した。和田ら[6]は、このテキストマイニングによる投稿分類の精度を向上させるため、経済専門の辞書の作成を行い、一般の辞書と比較して有用性を示している。

Bollen ら[7]は Twitter に着目し、自然言語処理を用いてツイートを 6 種類の感情レベルに分類し、“calm”な感情は 2 日後、及び 5 日後のダウ平均株価と正の相関があることを示した。Hong ら[8]は、[7]の結果を受け、株式に関係のあるツイートのみを収集して自然言語処理による分析を行って感情値を定義し、企業についてのツイートの感情値はその企業の株式リターンと正の相関があること、及びツイート発信者のフォロワー数が少ないほど感情値が反映する速度が遅くなることを示している。

これらの研究では、ソーシャルメディアの指標として、強気・弱気判定や 6 種類の感情分類を用いている。しかしながら、ソーシャルメディアにはさまざまな投稿がなされており、単純な強気弱気判定や 6 種類程度の大分類では整理できないさまざまな話題の投稿がなされている。より精度よく VI 指標を予測するには、より詳細な話題に投稿を分類する必要がある。そこで我々は、トピックモデルに着目する。

トピックモデルは、ソーシャルメディアなどから話題を抽出するために使われる手法の一つである。例えば、Tin ら[9]は、トピックモデルを用いて Twitter の話題を分類し、ニュースと Twitter の話題の変遷を分析している。その結果、一般的にニュースで話された話題が遅れて Twitter に波及するという結果を報告している。本研究では、Tin ら[9]の手法を参考に、株式投資に関するソーシャルメディアからさまざまな話題を抽出する。

これにより、恐慌リスクがあるときに活発に話されている話題を発見し、VI 指数の予測にどのような話題が影響しているのか明らかにする。

2.3 研究目的

株式指標に対するアプローチは数多くなされているが、リスク回避指数として有用である VI 指数について、投資家の発信する情報と関連付けて検証した研究は見当たらない。我々は、VI 指数が投資家の感情と密接に関係が

あるという性質から、投資家の感情が表出されるソーシャルメディアの話題から VI 指数が予測可能であると考えられる。そこで、投資家の感じる恐慌のリスクと対応した話題を抽出するために、トピックモデルを用いる。本研究では、トピックモデルで抽出した話題から VI 指数を予測することを目的とする。また、どのような話題が、VI 指数の予測に有効に働くのか明らかにする。

3. 提案手法

本章では、ソーシャルメディアの話題を用いて VI 指数を予測する手法を提案する。

3.1 概要

提案手法の概要を図 1 に示す。提案手法は、ソーシャルメディアにおける話題と VI 指数に関連がある、という仮説に基づいている。ソーシャルメディアの話題は、株取引について活発な投稿がなされている Yahoo!JAPAN の株式掲示板から取得する。まず、Yahoo!JAPAN の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する。取得したメッセージはから、投稿の内容を表す単語群を抽出するために、形態素解析を行う。形態素解析で得られた単語群から話題を抽出するために、LDA トピックモデル[10]を用いてトピック分類を行う。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、相関分析を行う。さらに、重回帰分析を用いて VI 指数の予測を行う。

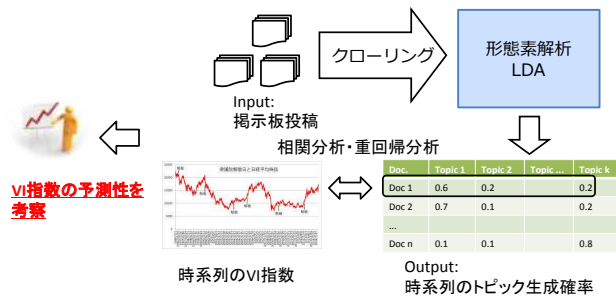


図 1 提案手法

3.2 分析対象

我々は、分析対象として、株取引に関する日本最大のソーシャルメディアとして、Yahoo!JAPAN の株式掲示板を対象とする。掲示板における話題の変遷を確かめるため、投稿数上位のアクティブなスレッドを対象とする。

3.3 データセットの作成

データセットの作成のために、まず株式掲示板からプログラムを用いて自動的にメッセージを収集する。投稿メッセージは、そのままでは処理できないため、文章の構成要素である単語（形態素）に切り分ける。このため形態素解析プログラムにより文章を形態素ごとに切り分けて、単語を抽出する。

さらに、形態素解析の結果得られた単語群から、特徴ベクトルとして適さない単語を除去するために、以下のノイズ除去処理を行う。

1. 数字、英字、記号、および日本語以外の単語の除去
2. 不要語（助詞、助動詞、接続詞、連体詞、副詞、数、代名詞、感動詞、固有名詞）の除去
3. 否定語の反映

以上のことより、トピック分類を行うために必要なデータセットを作成している。

3.4 トピックモデル

各投稿が、どのような話題を意味しているのかを判断するために、Blei ら[10]の LDA トピックモデルを用いる。これは、文書 d の生起に K 個の潜在的なトピック z_i を導入した確率モデルである。

$$p(d) = \sum_{i=1}^K p(z_i)p(d|z_i)$$

ここで、LDA では単語 w に対して $p(w|z_i)$ が得られる。つまりトピック z_i を一つのクラスターと見なすことによって、単語のクラスタリングを行っている。これによりトピック別生成確率が求められ、時系列のトピック別の投稿数を定義することができる。

3.5 相関分析・重回帰分析

VI 指数と関連を持ったトピックを明らかにするために、日別のトピックごとの投稿数と、日別の VI 指数を用いて相関分析を行う。時間間隔が日別である理由は、恐慌のリスクと関係のある指数であるため、可能な限り短い間隔での分析を行う狙いである一方で、分析が安定するために必要な投稿数を確保するためである。掲示板全体における話題の変遷を取り扱うために、その日の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、VI 指数の相関を分析する。

また、トピック間の相互作用を明らかにし、トピックモデルを用いた VI 指標の予測可能性について議論を行うために、トピック投稿率を用いた、VI 指数の重回帰分析を行う。

これらの分析において、トピック投稿率の VI 指数に対する先行性、同時性、遅行性を検証するために、投稿日の前日、同日、翌日の VI 指数を対象とする。これは、VI 指数に変化のある出来事が起きる前日、起きた当日、起きた翌日の書き込みは、それぞれ異なっており、予測に用いることができると考えられるからである。

4. 分析概要

本章では、具体的な分析の方法について述べる。

4.1 対象データ

我々は、分析対象として、2012 年 1 月 30 日から 2013 年 3 月 31 日までの Yahoo!JAPAN の株式掲示板の投稿数

上位 50 社のメッセージ 880321 件を取得した。2012 年のデータの取得については 2015 年 6 月に、2013 年のデータの取得については 2014 年 3 月に行った。表 1 にメッセージ取得企業の一部を示す。

表 1 メッセージ取得企業例（投稿数上位 10 社）

銘柄コード	企業名	投稿数
8411	みずほフィナンシャルグループ	100940
3632	グリーン	76392
9984	ソフトバンク	65929
6753	シャープ	63502
8306	三菱 UFJ フィナンシャルグループ	49479
2432	DeNA	40249
5912	OSJB ホールディングス	36591
6758	ソニー	34155
8515	アイフル	32102
7261	マツダ	25836

また、2012 年 1 月から 2013 年 3 月の市場開催日の日別日経平均 VI 指数を図 2、日別日経平均終値を図 3 に示す。対象期間において、日経平均終値の動きは、2012 年初頭から 2012 年 3 月にかけて上昇したのち急落、2012 年 6 月初頭に年初以来最安値を記録したのち 11 月までは小さな上昇下降を繰り返す、2012 年 11 月以降は上昇している。この期間における VI 指数の動きは、平時は 20 前後を推移し、二度大きく立ち上がっている。一度目の立ち上がりは日経平均が年初来安値を記録した 2012 年 6 月、二度目の立ち上がりは自民党政権が誕生し、アベノミクスが宣言された 2012 年 12 月に起きている。これを分析の対象とする。

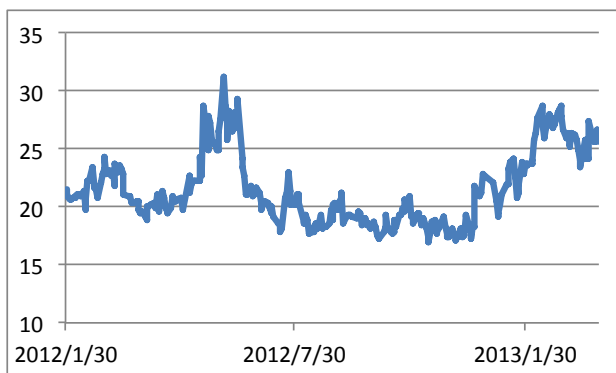


図 2 対象期間の VI 指数

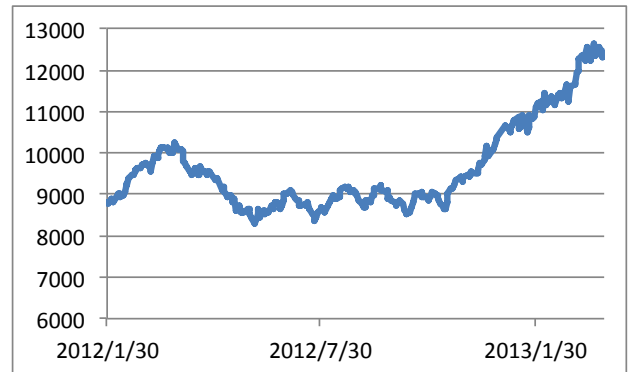


図 3 日経平均終値

4.2 形態素解析

前章で述べた形態素解析によって、テキストデータとして取得した投稿を、投稿ごとの単語群の形にする。ノイズ除去の結果、1 単語以上が残った投稿数は、845,400 件であった。

4.3 トピック分類

形態素解析により抽出した単語群から、時系列のトピック投稿数の定義について説明する。

前章で述べたトピックモデルにより、単語群から、それぞれのトピックの生成確率を求める。ここで、トピックの生成確率を求めるための[10]パラメータ α は 50 に固定した上で、予備実験としてより少ない生成トピック数で分類を実施した際に、文章量が極めて多い投稿しかトピック生成確率に明確な差が現れなかった。中程度の文章量の投稿もトピックに分類することを狙って、生成トピック数は分析環境における最大数である 80 とする。生成トピック全体の生成確率を p 、トピックの生成確率全体の標準偏差を σ としたとき、 $p+2\sigma$ を閾値とし、トピック T の生成確率がそれ以上である投稿について、T について話されている投稿であると定義し、その個数をトピック T の投稿数とする。また、トピック T の投稿数をその日の掲示板投稿数で割ったものをトピック T の投稿率とする。この投稿率の推移が、そのトピックの盛り上がり、盛り下がりを表すと考えられる。時系列のトピック別投稿率を、VI 指数と分析する対象とする。

5. 結果

本章では、分析の結果について述べる。

5.1 トピック分類結果

トピック分類の結果について述べる。今回のデータセットにおけるトピックの生成確率の平均値は 0.0125、標準偏差は 0.0138 であった。これを元に先に述べた手法でトピック分類を行った結果、トピック投稿数は 1,347,150 件となった。表 2 に、投稿と、分類トピックの例を示す。文章量が極めて少ない投稿について、トピックが割り当てられなかった一方で、文章量の多い投稿が 2 トピック以上に割り当てられたため、本文のない投稿などを除い

た総投稿数 845,400 件より多くなった。

5.2 トピックと VI 指数の相関分析の結果

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 3、表 4 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、負であるものの中からそれぞれ上位 5 件に位置するトピックをピックアップし、そのトピックに所属する 10 単語を表している。正の相関、負の相関どちらの場合も時系列による相関係数の傾向に大きな差は見られず、各時系列における相関係数上位 5 件は表に示すトピックから成っている。

5.3 重回帰分析の結果

日別のトピックごとの投稿数と前日、同日、翌日の VI 指数を重回帰分析した結果を表 5、表 6 に示す。表 5 は、全 80 トピックを説明変数として重回帰分析を行い、そののちに、係数が 1%有意であるトピックのみを説明変数として分析を行ったものである。表 6 は、各時系列における、各トピックの標準化係数を表したものである。標準化係数についても、前項で示した VI 指数と正の相関を持つトピックが正の係数を持ち、負の相関を持つトピックが負の係数を持った。時系列による大きな差は見られなかった。

表 2 投稿と分類トピックの例

	タイトル	本文	分類トピック 1	分類トピック 2
分類トピックなし	春が来た	いい勢いですね	なし	
1 トピックに分類	Re: 罰当りの親	お前はアホカ？ そんな作り話はもう古い。	俺 お前 奴 男 麻雀 爺 相手 掲示板 みたい アバキ 誰	なし
2 トピックに分類	鹿島は、原発除染処理で伸びる	鹿島建設は被災地地元企業であるし、復興予算の中でも、除染処理などの費用が大きいので、これから長期にわたり、事業が伸びるだろう。	建設 大 震災 市 県 復興 東京 地震 日本橋 東日本	電力 原発 稼働 東電 関西電気 関電 事故 エネルギー 福島

表 3 VI 指数と正の相関係数を持つトピック(前日との相関係数順、空欄は 5 位圏外)

トピック番号	トピック単語	前日	同日	翌日
39	日経 ドル 円安 為替 円高 平均 ダウ 上昇 安 高	.412	.386	.333
16	倍 銘柄 関連 不動産 期待 高 株価 割安 注目 大	.411	.383	.333
10	ここ 銘柄 他 ホールド 推奨 方 参考 参戦 ケネ 動き	.403	.389	.394
14	ゲーム 任天堂 提供 ソフト グリー サービス ユーザー サイト ネット 版	.393	.383	
66	銀行 金 資金 アイフル 倒産 借金 利益 資産 額 融資	.316		.325
32	医療 細胞 大学 研究 技術 開発 教授 的 ため がん		.348	.311

表 4 VI 指数と負の相関係数を持つトピック(前日との相関係数順、空欄は 5 位圏外)

トピック番号	トピック単語	前日	同日	翌日
53	アホ ソニー アイフル シャープ りんこ ストップ グリー チャート ♪(-402	-407	-404
38	俺 お前 奴 男 麻雀 爺 虫 嫌い カス	-392	-373	-398
30	日本 韓国 日本人 花 王国 朝鮮 在日 朝鮮人 韓国人 歴史	-354	-355	-358
22	民主党 国民 政治 政権 自民党 選挙 首相 氏 野田	-313	-321	-333
33	ソフトバンク ドコモ 買収 孫 通信 数 携帯 件 契約 利用	-306		
62	中国 日本 政府 日 島 尖閣 問題 沖縄 尖閣諸島 北朝鮮 国		-298	-365

表5 重回帰分析結果

	R2	補正 R2	F 値	Df	p 値	AIC	BIC	CAIC
前日	.605	.591	42.572	10, 278	.000	1232.679	1276.676	1233.809
同日	.606	.589	35.430	12, 276	.000	1235.646	1286.976	1237.179
翌日	.606	.584	27.936	15, 273	.000	1242.271	1304.600	1244.529

表6 トピックごとの標準化係数 (1%有意のトピック)

前日			同日			翌日		
トピック 番号	標準化 係数	VIF	トピック 番号	標準化 係数	VIF	トピック 番号	標準化 係数	VIF
66	0.289	2.101	40	0.299	1.461	40	0.287	1.492
40	0.288	1.238	14	0.239	1.373	14	0.165	1.269
16	0.26	1.736	16	0.229	1.918	32	0.16	1.344
14	0.224	1.334	66	0.184	1.524	60	0.153	1.086
64	0.195	1.686	32	0.163	1.602	13	-0.109	1.05
32	0.141	1.425	29	0.121	1.259	48	-0.118	1.277
23	0.116	1.127	46	0.106	1.123	38	-0.134	1.378
30	-0.193	1.845	33	-0.117	1.185	11	-0.142	1.71
11	-0.193	1.92	30	-0.184	1.749	73	-0.157	1.352
38	-0.208	1.378	38	-0.229	1.426	22	-0.165	1.579
			11	-0.23	1.498	35	-0.165	1.891
						1	-0.179	1.161
						59	-0.192	1.652
						30	-0.224	1.674
						33	-0.229	1.173

6. 考察

本章では、分析結果について考察する。

6.1 相関係数の正負

表2, 表3で確認できるように、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。正の相関を持つ単語群は、「日経、ドル、円安、銘柄」といった、株式銘柄について議論を交わしていると考えられる投稿の単語で構成されている。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る。

ここで、VI指数の性質について考える。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と考えられる。VI指数と正、負の相関を持つ単語群の傾向、VI指数の持つ意味を考察すると、次の仮説が示唆される。

- インターネット掲示板においては、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、同日のVI指数との相関係数が正、負である投稿の、それぞれ相関係数上位3トピックの日別の投稿数の合計、及びVI指数を図4に示す。VI指数と正の相関のあるトピックは、VI指数の立ち上がりに応じて投稿数が増加している様子が観測される。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。対象期間の終盤においては、急激に投稿数が減少している。この投稿数減少の原因については、データ取得時期の差が考えられる。4章で述べたように、掲示板投稿のクロール時期について、2012年の書き込み2012年のデータの取得については2015年6月に、2013年のデータの取得については2014年3月に行っている。相関係数負のトピックについては、5章表2中段に示されるような、投資に関係のない攻撃的な投稿が主となる。このような投稿について、2度目のデータ取得時期には「荒らし」として削除されてしまっていることにより、取得できた投稿数が減少している可能性が考えられる。

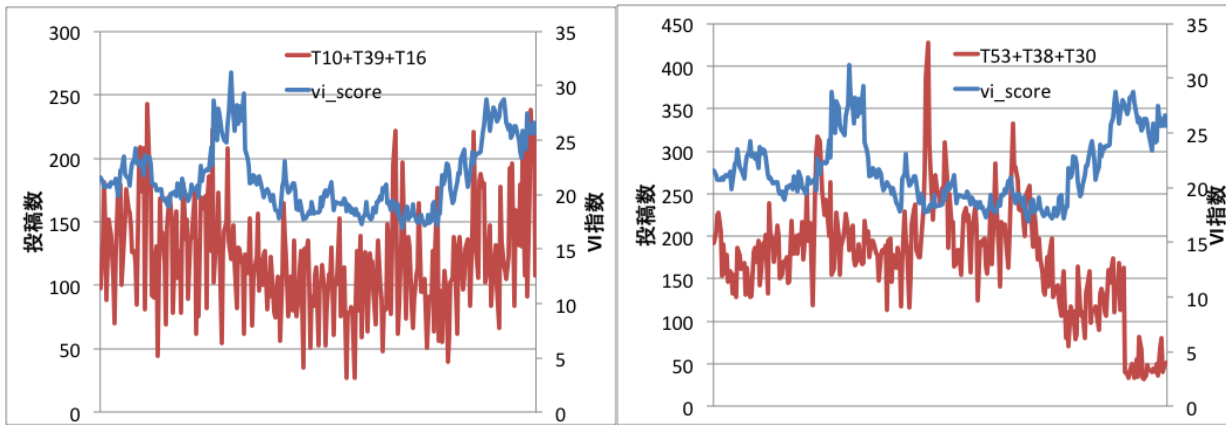


図4 VI指数と正の相関のあるトピック（左）と負の相関のあるトピック（右）の投稿数

6.2 重回帰分析の結果

重回帰分析によって、投資に関係のあるトピックの投稿率を正、投資に関係のないトピックの投稿率を負とする重回帰式を作成し、VI指数と有意な重相関係数を持つ重回帰式を得ることができた。ここで、我々は、予備実験として同一データを対象に先行研究[3]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0.416を導出していた。本研究における重相関係数は、これよりも優れており、従来手法と比較して、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。

6.3 分析時系列についての変化

相関分析、重回帰分析、どちらの場合においても、投稿日時の日、及び前後一日のVI指数との相関結果の傾向に大きな差は見られなかった。これに関しては、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にわたる継続性がある可能性が考えられる。この点に関する分析は今後の検討課題である。

6.4 煽り投稿検出技術との併用可能性

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。ここで、ソーシャルメディアにおける”煽り”投稿の検出は課題として多くの研究がなされており、山下ら[11]はYahoo!JAPANの株式掲示板における投稿を、その投稿に対する返信の内容を利用して機械学習で”煽り”投稿の検出可能性を検証している。このような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ことが考えられる。

7. まとめ

株式指標の一つであるVI指数について、ソーシャルメディアと密接に関連している株式指標であるという仮説に基づき、トピックモデルを用いた実際のインターネット掲示板投稿による分析を行った。結果として、VI指数と正の相関のあるトピック、負の相関のあるトピッ

クを発見し、VI指数と有意な相関を持つ重回帰式を定義することができた。今後の課題としては、VI指数予測モデルの構築、検証、及びインターネット掲示板における煽り投稿検出技術の利用などが挙げられる。また例えば、トピック内における投稿内容の強気・弱気の程度が株式指標の上昇・下降の方向性に影響を与えている可能性もある。この点に関する検証も今後の検討課題である。

参考文献

- [1] 西谷遼,最上義夫,小野典彦,池田健司,”ニューラルネットワークを用いた株売買支援システムの構築と検証”,自動制御連合講演会,2010.
- [2] 和泉潔,後藤卓,松井藤五郎,”経済テキスト情報を用いた長期的な市場動向推定”,情報処理学会論文誌,Vol.52,No.12,pp.3309-3315,2011.
- [3] Robert E. Whaley,”The Investor Fear Gauge”,The Journal of Portfolio Management,Volume 26,No.3,pp.12-17,2000.
- [4] 丸山健,梅原英一,諏訪博彦,太田敏澄,”インターネット株式掲示板の投稿内容と株式市場の関係”,証券アナリストジャーナル,Vol.46, No.11-12, pp.110-127,2008.
- [5] 諏訪博彦,梅原英一,太田敏澄,”インターネット株式掲示板の投稿内容分析に基づくファクターモデル構築の可能性”,人工知能学会論文誌,Vol.27,No.6,pp.376-383,2012.
- [6] 和田英一,諏訪博彦,小川祐樹,太田敏澄,”専門辞書を用いたテキストマイニングによるインターネット株式掲示板の投稿分析に関する研究”,経営情報学会秋季全国大会,pp.115-118,2012.
- [7] Johan Bollena,Huina Maoa,Xiaojun Zengb,”Twitter mood predicts the stock market”,Journal of Computational Science,Volume 2, Issue 1, pp. 1-8,2011.
- [8] Hong Keel Sul,Alan R. Dennis,Lingyao Yuan,”Trading on Twitter:The Financial Information Content of Emotion in Social Media”,Hawaii International Conference on System Science,pp.806-815,2014.
- [9] Ting Hua, Yue Ning, Feng Cheny, Chang-Tien Lu, Naren Ramakrishnan,”Topical Analysis of Interactions Between News and Social Media”,Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016.
- [10] D. Blei, A. Ng, and M. Jordan,”Latent Dirichlet Allocation”,Journal of Machine Learning Research, Volume3,pp. 993-1022,2003.
- [11] 山下達雄,坪内孝太,”株価掲示板情報における煽り情報の検出”,人工知能学会,2015.