

# ファイナンス研究における新しいアプローチ ——テキストマイニングを用いた投資家心理の数値化について——

岡 田 克 彦  
羽 室 行 信

## 要 旨

ファイナンス研究において、ホモ・エコノミカスの前提に疑義が挟まれるようになって久しいが、心理的バイアスが資産価格を歪めている証拠には乏しく、投資家心理が資産価格に影響を与えるが故にアノマリーが発生するという主張はあくまでも仮説に過ぎない。本稿では、テキストマイニングの手法を援用して、投資家心理を数値化し、心理的バイアスが価格を歪めているという直接的証拠を探索する試みを紹介する。筆者らは約1年半の日本経済新聞に含まれている44,485記事をテキストマイニングして極性指数を作成し、それが株価指数との間に強い相関関係を持つことを発見した。財務的データを一切使わず、テキストの評価情報からのみ作成した極性指数と株価指数の間に強い相関関係が存在するという事は、投資家の気分が資産価格に影響を与えている可能性を示唆している。今後、ファイナンス研究において投資家心理を数値化して扱う試みは、新しいアプローチとして注目される。

## I は じ め に

資産価格形成理論に関する争点について知りたければ、マーケットを熟知している個人投資家とファイナンスの理論学習をしてきた証券アナリストに次の質問をぶつけてみると良い。

「株式市場で株価を決める最も重要な要因は何ですか？」

相場に詳しいベテラン個人投資家からはこんな答えが返ってくるだろう。「株価はね、人気があるほどあがるものだよ。だから銘柄を選ぶコツは、人気の高くなりそうな銘柄を選ぶのさ」。この考え方は、Keynes（1936）の主張する人間心理市場仮説にほかならない。Keynesは株式投資を美人コンテストになぞらえて説明している。誰が本当の美人であるかには絶対的な基準はない。どの美女が審査員達の眼鏡にかなうか（好みなのか）ということがポイントなのである。株式投資はどの企業が良いかという客観的な価値推定に依拠するよりも、多くの投資家が良いと思う企業を対象にすべきであるというのだ。

同じ質問を、ファイナンス理論を学習した証券アナリストにぶつけて見れば、次のような回答になるに違いない。「株価は企業のファンダメンタルズを反映しているのです。ですから、我々証券アナリストは、企業の業績をいち早くつかみ、分析し、その規模と質を見極めることで売り買いの推奨をしているのです」。要するに株価を決めているのは企業業績だというのだ。この証券アナリストの考える株式市場は、伝統的ファイナンス研究者たちの主張である効率的市場仮説と呼ばれるものに近い。企業価値はその企業が生み出す将来キャッシュフローの現在価値として投資家に合理的に認識され、企業価値の変動（株価の変動）は投資家の合理的な将来キャッシュフローの期待値変化が生じる場合か、或いは合理的に推定されたキャッシュフローの割引率に変化が生じる場合にのみ起こるものだと考えるのである。効率的市場仮説が考える世界は美しい数式で表現され、株価は何もかも整合的に決まっている。株式市場だけではなく、Robert Merton の提唱した Intertemporal Capital Asset Pricing Model では資産評価モデルが経済全体の均衡モデルと整合的であることを示し、Breedon (1979) は株価のリスクは一人当たりの消費と資産価格の相関によって決まるべきことを示した。

人間心理市場仮説と効率的市場仮説のどちらが正しいのかという論争は、ファイナンス研究者の間で長年展開されてきた。多くの素人投資家にとって、人間心理市場仮説については納得いくものであるが、効率的市場仮説については学問上のものであるようだ。一般的な素人投資家は、自分だけは株式市場で利益を上げることができると信じ、情報収集や分析をしっかりとすれば株式市場で儲けることができると考えている。しかし現実には、市場平均よりも高いリスク調整後リターンを達成することは、プロであっても難しいことが報告されており (Malkiel (1995)), 市場の動きは概ねランダムウォークなのだということが分かっている。市場で儲けられると思ってしまうのは、まさにギャンブラーの錯覚 (Clotfelter and Cook (1993)) であり、現実には困難である (Barber et al. (2009))。市場価格には入手可能な全ての情報が反映されており、追加的な情報はランダムにしか発生しないので、株価動向もランダムになるのだ。更に、市場が効率性を保つメカニズムについても、非常に説得力のある議論が存在する。それは Friedman (1953) のいう裁定取引の存在である。効率的市場仮説では、個人は期待効用最大化のルールに基づいて資産選択し、期待効用を算出する確率計算において Bayesian のアップデートができる、いわゆるホモ・エコノミカスの前提を持つ。もちろん、すべての個人がホモ・エコノミカスではないが、仮に一部の投資家がバイアスを持って資産価格を歪める行動をとったとしても、合理的な投資家がすぐに歪んだ価格を修正する取引（裁定取引）を行うから関係ないというわけだ。価格は合理的な水準に長期的には収束するため、非合理的な投資家は報われない。非合理的な投資家はやがては市場から排除され、市場の効率性は保たれるのである。

一方、人間心理が価格形成に影響を与えるという研究者による議論のポイントは三点ある。一つは、Friedman (1953) の裁定取引には現実的な限界があるという問題である。合理的な投資家が裁定取引を実施しようと思っても、市場の構造から裁定が効かないのだ。この顕著な例は、NTT 株と NTT ドコモ株に見られるような親子上場に見られる。相当の期間 NTT 株の時価総額は自らが保有するドコモ株の時価総額以下で取引されていた(岡田 (2007))。NTT 株を全株取得し、保有の NTT ドコモ株を売却するだけで巨額の利益が得られる計算になる。現実には NTT の大株主である財務大臣から株式を買い付けることは不可能であるため、実現性には乏しい。このような裁定の限界 (limits of arbitrage) が市場の随所に見られるため、効率性が保たれる保証はない。第二点として、合理的な投資家でさえ、短期的には非合理的な価格形成に加担するという問題がある。非合理的な投資家の中には positive feedback trader と呼ばれる投資家が存在し、彼らは価格が上がれば買いに走り、価格が下がれば売りにまわる。合理的な投資家は、このような非合理的な投資家の行動を予想し、本来価値から乖離した価格であっても positive feedback trader が買いに走る前に買い、売り始める前に売るという行動を取るというのである (De Long et al. (2000))。そうであれば、ホモ・エコノミカスなトレーダーが存在したとしても、彼らの取引がファンダメンタル価値に収束させることはなく、逆にファンダメンタル価値から乖離させる力として働くことになる。第三点目として、どうしても説明できないアノマリーが長期的に残存している事実可依拠して、市場価格は長期的にファンダメンタル価値に収束していないという指摘である。最もよく知られているものに 1 月効果、週末効果 (Rogalski (1984)), ハロウィーン効果 (Bouman and Jacobsen (2002)), デカンショ節効果 (Sakakibara, Yamasaki and Okada (2009)), 天候効果 (Hirshleifer and Shumway (2003)) 等が挙げられるが、このようなアノマリーについては長期間消滅していない。<sup>i</sup>

しかしながら、これらの三点を列挙したとしても、効率的市場仮説が間違いであるとは言えない。基本的には市場は効率的であるが、リスクをうまく記述できていないだけだという主張ができるからだ。効率的市場仮説の基本的メッセージは、リスクに見合ったリターンしか期待できない、即ちフリーランチは存在しないということである。制度的な問題を解決し裁定の限界を最小化し、リスクファクターを特定すれば、一時的な価格の乖離はあったとしても、概ね市場は効率的である事がわかり、将来はアノマリーの問題は消滅するだろうと、効率的市場仮説派は主張する。

この論争は今も継続しているが、人間心理市場仮説陣営の弱みは、心理的バイアスと、投資行動についての直接的な連関を示す証拠が少ないことにある。例えば、日本や世界の主な株式市場で報告されている天候効果についても、明らかになっているのは、天気の良い日の株式市場のリターンが統計的に有意に高いということだけである。天気が良いとい

う状態が、人間に明るい心理状態をもたらすところまでは心理学の知見で明らかとなっている。しかし、明るい投資家心理が積極的に株式を買うという投資行動に直結するかどうかは、あくまでも推測の域を出ないのだ。投資家心理が直接投資家行動に影響しているという証拠が観察できない限り、天気というファクターが投資行動に影響を与えているとは結論づけられない。これからのファイナンス研究者の重要な役割の一つは、心理と行動と結果を結びつける実証的証拠を見出すことであると言えよう。

本稿では、この問題に取り組むための新しいアプローチとして、情報工学分野のテキストマイニングという手法を取り入れたものを紹介する。第二節では、テキストマイニングのファイナンスへの応用について述べ、第三節では筆者らの日本経済新聞を用いたテキストマイニングの実験例について触れる。第四節では実証結果の提示と解釈について述べ、第五節ではこれからの課題とこの研究分野の展望について述べる。

## II テキストマイニングの手法のファイナンス研究への応用

テキストマイニングとはテキスト（文章）の中から必要な情報を取り出す方法論のひとつである。テキストはそもそも情報を伝えるものであるが、これまで研究対象とされてきたテキスト情報は、新聞記事、学術論文に代表されるように、事実を記述するものがほとんどであった。それに対し、チャット、Web 掲示板、Weblog 等の普及、利用者の増大に示されるように、インターネット上では、一般の個人が手軽に情報発信できる環境が整うとともに、個人の発信する情報に、ある対象に関するその人の評価等、個人の意見が多数記述されるようになってきている。この個人の評価に関する情報（評価情報）をテキスト中から抽出し、整理し、提示することは、対象の提供者である企業やサイト運営者、また、対象を利用する立場の一般の人々双方にとってメリットとなる。このため、自然言語処理の分野では、近年急速に評価情報を扱う研究が活発化している。筆者らは近年発展してきたテキストマイニングの手法を、ファイナンス分野において活用することを考えた。

効率的市場仮説が言うように、株価が将来キャッシュフローの現在価値として決まっているのであれば、事実のみが株価に反映されているはずである。しかし、ホモ・エコノミカスでない投資家が株式評価をすると考えれば、株価を評価する時点の心理状態が評価額に影響を及ぼすと考えられる。この心理状態に影響を与えるものが、投資家一般に言われる「相場のムード」、「場味」のようなものであると言えよう。この「場味」を抽出する手法として、テキストの評価情報を利用する。テキストが伝える事実そのものではなく、事実がどのように伝えられているかという点を客観的分析に耐える数値情報に変換することにより、投資家心理の株価に及ぼす影響について、経済学の方法論を用いて調査すること

ができるのである。

例えば、「ソニーの今期一株あたりの利益は200円となりそうだ。この水準では現在のソニーの株価収益倍率（PER）は20倍以下となり、値ごろ感が出ている。」という記述と、「ソニーの今期一株あたりの利益は200円となりそうだ。この水準では現在のソニーの株価収益倍率（PER）は20倍以下となるが、米国アップル社とほぼ同等の水準である。」という記述であれば、明らかに書き手のソニー株に対する評価は前者の方が高いことがわかる。効率的市場仮説が考える様に、事実情報だけが株価を決めるとすれば、将来期待キャッシュフローは一株当たり200円という情報から推定され、現在の株価はソニーの持つ事業リスクに相応しい割引率を用いた割引現在価値と解釈される。しかし、Bayesian update ができない人間が、バイアスを持った将来見通しを形成し、バイアスを持ったリスク評価をしているのであれば、事実情報以外にも、評価情報が株価に反映されているはずである。人間が心理的バイアスを持ち、それによって投資の意思決定が影響を受けていることは既に報告されている（例えば Sutan and Willinger (2009)）。しかし、それが市場の場において形成される株価に影響を与えるかどうかという点については、十分な証拠が提出されていない。テキストマイニングによる評価情報の抽出は、心理と株価の重要な連関について説得力のある証拠を提出してくれると期待される。

### Ⅲ 新聞記事から評価情報の抽出

#### 1 新聞記事データについて

筆者らは、新聞記事に含まれるテキストデータの中の評価情報が場味の形成に関与していると考え、新聞記事情報の中の評価情報を数値化することにした。株式相場全体の雰囲気を形成するものの中には、株式市場に関するニュースに限らず、経済全体、政治、社会と幅広いと考えられる。そこで、今回は1面から10面までの日本経済新聞記事の全記事を取得することにした。新聞各紙の中で日本経済新聞を選択したのは、経済関連の記事の充実度、300万部を超える発行部数等を考慮し、相場のセンチメントに係る分析を行うに最も適したメディアと判断したからである。検証対象期間は2008年12月1日から2010年6月16日であり、この期間に取得した記事数は44,485記事、文章数は535,851文である（表1参照）。記事のテキストファイルは日経テレコン21からダウンロードした。

表1 解析対象記事の統計データ

記録収集期間	2008年12月1日 2010年6月16日
記事数	44,485
文章数	535,851
文節数	4,783,931
形態素数	14,052,180
用言句文節数	1,405,313
疑問文数	6,916
逆接接続文節数	48,910
接続文節数	117,403
引用文節数	89,969
願望文節数	4,945
否定文節数	81,224
妥当評価表現文節数	1,380,993
同定用言句文節数	610,153

(注) 疑問文、願望文、引用文は評価の対象外としている

## 2 評価情報の抽出

テキストの評価情報には大きくわけて二つある。書き手が対象に対して良いと感じている場合と悪いと感じている場合である。この評価情報の良い・悪いに関する軸を評価極性と呼ぶ。ある評価情報が良い評価をもつことを「肯定極性をもつ」と呼び、逆に悪い評価をもつことを「否定極性をもつ」と呼ぶ。また、肯定極性か否定極性をもつ評価情報がテキスト内で記述された表現を評価表現と呼ぶ。テキスト評価分析を支える要素技術に関する研究には、注目するテキスト断片の粒度によって、①評価表現辞書の構築に関する諸研究、②評価情報を観点とした文書分類に関する諸研究、③評価情報を含む文の抽出に関する諸研究、④評価情報の要素組の抽出に関する諸研究に分けられるが、紙幅の関係上本稿での詳細な記載は避ける。<sup>ii</sup> ただ本稿の目的が、一般的な商品・サービス等への評価ではなく、株式市場の場味という特殊な評価を数値化することであるから、①の評価表現辞書の構築方法については慎重に検討した。

評価表現辞書の構築方法には主に3つの流れが存在する。一つは語彙ネットワークを利用した構築方法である。語彙ネットワークを利用した評価表現辞書構築の手法は、既存の語彙知識の情報を利用して、評価表現の評価極性を求める手法である。まず、シソーラスや国語辞書の情報を基にして、語彙をノードとする語彙ネットワークを用意する。そして、評価極性情報をネットワーク上で伝播させることによって、語彙ネットワーク中のすべてのノード（語彙）の評価極性を求めるのである。この手法は「類義関係にある語彙の評価



極性は一致しやすい」という仮定に基づいてアルゴリズムが構成されている (Kamps et al. (2004))。

もう一つは共起情報を利用した評価表現辞書構築の手法である。肯定極性をもつ典型的な表現 (「good」や「excellent」) と否定極性をもつ典型的な表現 (「bad」や「poor」) を種表現として始めに用意しておき、種表現と共起する比率に従って語句の評価極性を判定する。この手法は「肯定極性をもつ語句の周辺文脈には肯定極性をもつ語句が現れやすく、否定極性をもつ語句の周辺文脈には否定極性をもつ語句が現れやすい」という考え方に基づく。

第三に周辺文脈の情報を利用した評価表現辞書構築の手法がある。まず、評価極性が既知である少数の種表現を幾つか用意する。そして、辞書構築アルゴリズムでは、繰り返し過程の中で、種表現から評価表現を順次増やしていくブートストラップ的な戦略をとる。これは、テキスト中に評価表現が存在すると、その周囲にも同様の評価表現が連続する文脈が形成されることが多いという、文脈の一貫性の考え方に基づく。

株式相場を語る場合、「上値が重い」「底堅い」等の特有の言い回しが存在することから、本稿では予め用意した辞書で評価極性を判定するのではなく、周辺文脈の情報を利用した手法を採用することにした。

### 3 那須川・金山 (2004) の手法による辞書構築

那須川・金山 (2004) は評価表現の周辺文脈について、「文書中に評価表現が存在すると、その周囲に評価表現の連続する文脈が形成されることが多く、その中では、明示されない限り、肯定／否定の極性が一致する傾向にある」と述べている。筆者らは株式相場の評価についての特殊性を考慮し、評価表現辞書作成の為に、ブートストラップ的に評価表現を収集するこの手法を採用した。ある評価表現  $t$  の周辺文脈に注目した場合、「しかし」や「～だが」などの逆接関係を導く表現が存在しなければ、文脈中にある評価表現の評価極性は  $t$  の評価極性と等しくなり、逆に、逆接関係を導く表現が存在すれば、それ以降の極性は  $t$  の極性から反転すると考える。まず、評価極性が既知である少数の評価表現を種表現として用意し、上記の仮定に従って、文書内からブートストラップ的に評価表現候補を収集する。複合的な評価表現に対応するため、用言と同時に、用言に最も近い格助詞句を加えた複合表現を評価表現候補として選んでいる。考慮する周辺文脈は、評価表現  $t$  と同一文内の節、前文の主節、後文の主節としている。

例えば、評価極性が既知である種表現として、「回復する」を予め肯定極性を持つとして指定しているとしよう。今、「東京株式市場は 日経平均株価が 四日ぶりに 反発した。終値は 前日比 二一三円四三銭高の 八〇三八円九四銭で、 四営業日ぶりに 八

〇〇〇円台を 回復した。米株高が 好感され、輸出関連株を 中心に 買われた。」というテキストから評価情報を抽出するための辞書を作りたい。評価極性辞書の構築は、「回復した」という肯定の種表現から「終値は 前日比 二一三円四三銭高の 八〇三八円九四銭で、四営業日ぶりに 八〇〇〇円台を 回復した。」という文に現れる全ての用言を肯定極性と判断する。更に、前文の「東京株式市場は 日経平均株価が 四日ぶりに反発した。」と後文の「米株高が 好感され、輸出関連株を 中心に 買われた。」も逆転を示す接続詞がないために両方とも肯定的な用言から構成されていると判断する。こうして、次々に肯定用言、否定用言を収集し、候補辞書を作成する。候補辞書の中から、出現頻度、肯定極性、否定極性の偏りを総合的に判断し最終的な辞書を作成する。今回の実験では、「回復する」「改善する」「上方修正する」「改善」「底堅い」「好調」を肯定極性の種表現として、「悪化する」「悪化」「下方修正する」「上値重い」「不調」を否定極性の種表現として辞書構築した。最終的に構築された辞書は表2に示す。

表2 評価表現辞書

肯定極性 (76エントリ)	否定極性 (70エントリ)
SE 続伸、回復する、改善する、上方修正、底堅い、回復、下げ止まる、改善、好調、上方修正する、続伸する	不調、悪化、上値重い、悪化する、下方修正する
CE なる〈黒字：に〉 示す〈動向：を〉 いう〈回復：と〉 かかる〈悪化：に〉 縮小する〈赤字幅：が〉 持ち直す〈効果：で〉 上昇する〈鉱工業生産指数：が〉 比べる〈13月期：と〉 更新する〈年初出来高値：を〉 終える〈上海総合指数：が〉 反発する〈日経平均株価：が〉 なる〈黒字：と〉 なる〈一致指数：が〉 なる〈プラス成長：と〉 反発する〈東京株式市場：で〉 上回る〈前年実績：を〉 縮小する〈赤字：が〉 進展する〈在庫調整：が〉など	よる〈雇用統計：に〉 発表する〈米労働省：が〉 する〈悪化：と〉

(注) SE = Single entry CE = Combined entry. 太字は評価表現辞書構築に用いた種表現であることを示す。



#### 4 極性スコアの時系列観察

辞書完成後、収集された535,851文について極性を判定していく。文章中に存在する肯定用言が単独、あるいは複合で含まれていれば、+1と判定し、否定用言が単独、あるいは複合で含まれていれば-1と判定する。この集計は日次に行い、t日の肯定極性のスコア（ $pos_t$ ）はt日の日本経済新聞1面から10面全文章に現れる肯定極性の合計値である。同様に否定極性のスコア（ $neg_t$ ）はt日の日本経済新聞1面から10面全文章に現れる否定極性の合計値である。

$$pos_t = \sum_{i=1}^m ppol_t$$

$$neg_t = \sum_{i=1}^m npol_t$$

但し、 $neg_t$ はt日の1面から10面までの全文章 m 個の中で負の極性値（ $npol$ ）を持つものの合計値、 $pos_t$ は正の極性値（ $ppol$ ）を持つものの合計値である。極性値の推移はそのまま観察すればかなりのばらつきを示すことから、以下の方法で線形加重移動平均を算出し観察することにする。

$$WMA_t(n) = \frac{n}{S} p_t + \frac{n-1}{S} p_{t-1} + \dots + \frac{2}{S} p_{t-n+2} + \frac{1}{S} p_{t-n+1}$$

$$S = \sum_{i=1}^n i$$

但し、 $WMA_t(n)$ はt期における n 日間の線形加重移動平均値、 $p_t$ はt期における極性値スコアであり、正の極性値スコア（ $pos_t$ ）か、負の極性値スコア（ $neg_t$ ）のどちらかである。

## IV 評価情報と株式市場

### 1 実証結果

記事データの収集期間は2008年の12月から2010年6月までである。この期間は金融危機直後からの回復期に相当するが先行きに対する不透明感が非常に強い時期であった。リーマンショックが発生したのが2008年9月であり、米国の大統領選挙が行われたのが11月である。リーマンショック後7000円台まで下落していた日経平均株価指数は、2008年12月から回復し始め9200円まで上昇したものの、3月初旬にかけて急落し7054円まで下落した。その後世界経済、とりわけアジア経済の回復と歩を合わせるように上昇し、2010年には10000円を挟んだ動きとなっている。検証期間内の株式市場は、事後的に考えれば概ね上昇基調にあるものの、金融危機の余波をどの程度読みこめば良いのか意見のわかれた時期

であり、株式市場に対する期待が高まった後に裏切られるような価格動向も見られた（図1参照）。

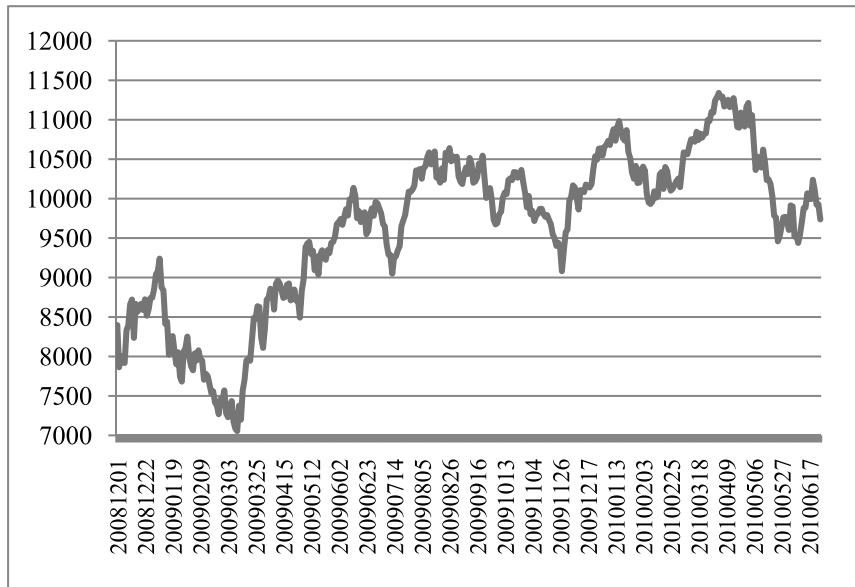


図1 対象期間中の日経平均株価指数の日次推移

検証期間において、新聞記事の肯定極性スコアと否定極性スコアの45日間の線形荷重移動平均値を計算し、その時系列推移を図2と図3に示している。肯定表現を日次で集約し移動平均を計算したスコアについての特徴は、上下のばらつきが大きいということであろう。また肯定極性スコアは上昇する時期と下落する時期にある程度のもメンタムが存在するようだ。肯定極性スコアの推移に比較し、否定極性スコアの推移の変動率は小さく、安定している。

興味深いのは、新聞記事の極性値の算出の過程に株価は全く関与していないにもかかわらず、算出された肯定極性スコアと日経平均株価の動向には正の相関（0.71）が見られることである。株価の上昇が新聞記事の中に肯定的なものをもたらすという因果関係を想像するのは容易であるが、取得記事が1面から10面までの政治、経済、文化等の全記事であることに鑑みれば、両者の相関が高いのは驚きである。更に、筆者らが注目したのは否定極性スコアと株価指数の相関係数（0.90）である。否定極性スコアであるから数値自体は負の値を取っているが、否定評価表現の減少が株価指数の上昇に強い関係を持っているようだ。

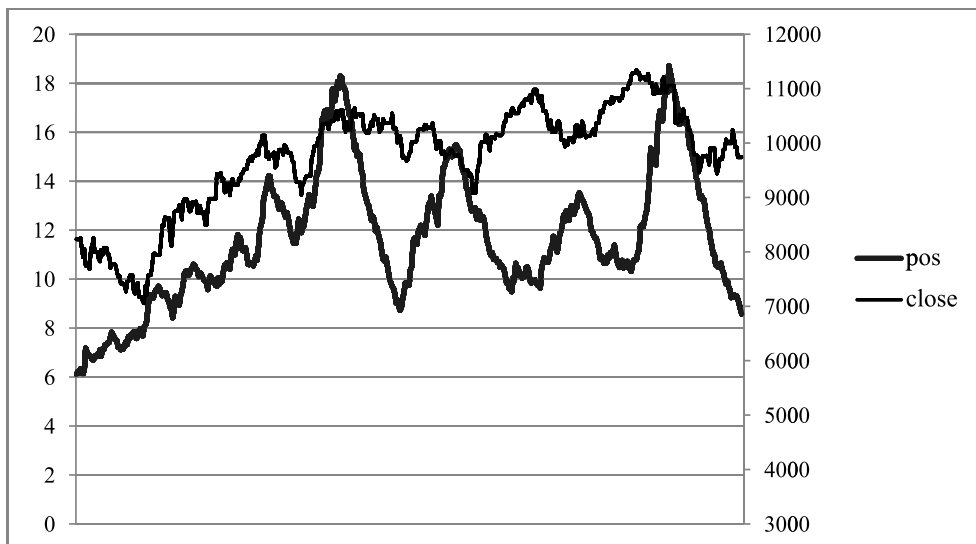


図2 45日線形加重移動平均の日次肯定評価スコア (pos) と日経平均株価指数 (close) の終値推移

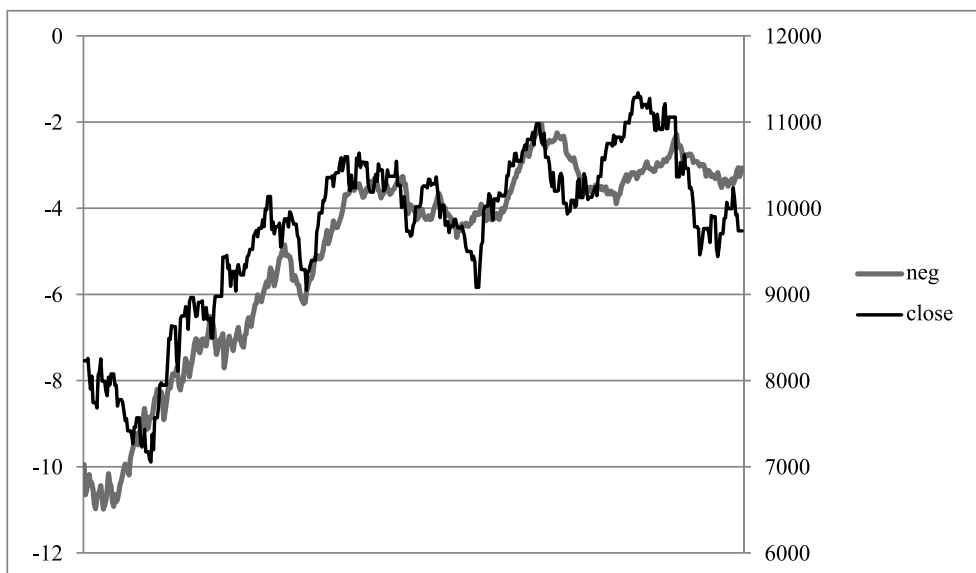


図3 45日線形加重移動平均の日次否定評価スコア (neg) と日経平均株価指数 (close) の終値推移

## 2 考察

評価情報の抽出には株式相場に関するインプットは皆無である。単に種表現から評価表現辞書を構築し記事の極性スコアを判断しただけである。しかしながら、そこから抽出さ

れた数値と株価指数とは非常に高い相関を示す。そもそも取得した記事の中には、経済や相場に関するもの以外の記事の方が多い。従って、株価が上昇しているから肯定的な記事が多いのだという因果関係で全てを説明するには無理であろう。また、否定極性スコアが肯定極性スコアよりも強い相関を示すのも大変興味深い。

肯定極性スコアと否定極性スコアの動向から言えば、否定極性スコアの方が株式相場の動向と近く、場味を斟酌する場合は否定評価表現に焦点を当てる方が良いとの示唆が得られる。相場が金融危機から回復しようとしているとき、一般的な評者であればどう感じるだろうか。金融危機が最悪期を脱して今後は経済を含めた全体が回復、改善に向かうだろうという表現を使いにくいのではないか。何故なら、そのような積極的な表現を用いたとした場合、経済の二番底が発生した場合、評者は明らかに間違えたことになるからだ。評者の先行きの不安感の減少は、肯定評価表現の増加というよりも、否定評価表現の減少という形で現れる可能性がある。マイニングの対象となった記事が経済関連だけに限定されていないため、あくまで推測の域を出ないが、人間心理を反映するものは、否定評価表現であると考えられるのではないだろうか。

## V 今後の展望

ファイナンスと社会学、心理学の学際的研究分野は総称して「行動ファイナンス」と呼ばれている。行動ファイナンスの学際的で自由なアプローチから生まれてくる研究によって、金融市場の真実の理解を深めるための新たな知見が次々と得られている。心理学実験とファイナンス実験の組み合わせや、ミクロ経済学では扱われない心理学的な変数の資産価格への影響を見るなど、従来のミクロ経済学の延長線上にあるファイナンスでは考えられなかった研究アプローチが、多様な発想で実施されている。第1節で述べた天候効果にしても、当初はとても荒唐無稽なテーマだと考えられたが、前日の雲の量で翌日の株価がある程度予測できることが報告され学界に衝撃を与えた。これは天気の良い日の人間に与える心理的効果が投資家を楽観的行動に駆り立てるからだと推測されているが、はっきりしたメカニズムは解っていない。人間の心理がマーケットに与える影響は、それを解読して一儲けができるというレベルのものだけではない。市場の根本的な問明できないばかりか、合理的モデルに依拠することによって、かえって価格の正当性を異常に高い成長率の予測や割引率の低さに求めることになり、バブルを助長してしまうおそれがある。この意味で、心理的要因をできるだけ客観的にとらえる方法の開発はこれからのファイナンス研究の広がりのためには必要不可欠であろう。

本稿ではファイナンス研究の新しいアプローチとして、テキストマイニングの手法を提

案した。筆者らが知る限りにおいて、海外の研究においても、テキストマイニングの手法を応用した研究は散見されるがまだまだ稀少である。本稿では、特に幅広い分野の日本経済新聞の記事に焦点を当て、相場全体、日本全体を覆う空気感のようなものの数値化を試みた。その結果、株価の構成要因とは無関係なテキスト記事から計算される極性スコアが、株価指数と高い相関を示すことが明らかになった。今回の実験はまだパイロットテスト的な色彩が濃く、改良の余地は大きい。しかし、現段階で得られた結果だけからみても、テキストマイニングのアプローチが今後のファイナンス研究の発展に大きく寄与するのは間違いないと思われる。

### 謝 辞

本研究は JST ERATO 湊離構造処理系プロジェクトの協力のもとに可能となりました。ここに記して心より感謝申し上げます。

### 注

- i Shiller (1990) はこのようなアノマリーの存在に加えて、株式市場の変動率が高すぎるものが効率的市場仮説と非整合的だと指摘する。
- ii テキストマイニングの最近の分析動向については乾・奥村 (2006) を参照されたい。

### 参 考 文 献

- Barber, B., Y. Lee, Y. Liu and T. Odean (2009), "Just How Much Do Investors Lose by Trading?," *Review of Financial Studies*, 22, pp. 151-186
- Bouman, S. and B. Jacobsen (2002), "The Halloween Indicator, 'Sell in May and Go Away': Another Puzzle," *American Economic Review*, 92, pp. 1618-1635
- Breeden, D. (1979), "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, pp. 265-296
- Clotfelter, T. and P. Cook (1993), "The 'Gambler's Fallacy' in Lottery Play," *Management Science* 39, pp. 93-95
- De Long, J., A. Shleifer, H. L. Summers and J. Waldman (1990), "Positive Feedback Investment Strategies and Destabilizing Rational Speculation", *Journal of Finance* 45(3), pp. 379-395
- Friedman, M., "The Case for Flexible Exchange Rates," in *Essays in Positive Economics* (Chicago: The University of Chicago Press 1953)
- Hirshleifer, D. and T. Shumway (2003), "Good Day Sunshine: Stock Returns and the Weather," *Journal of Finance* 58, pp. 1009-1032
- Kamps, J., Marx, M., Mokken, R. J., and de Rijke, M. (2004). "Using WordNet to Measure Semantic Orientations of Adjectives." In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*.

- Keynes, J.M. (1936), "*The General Theory of Employment, Interest and Money*", Macmillan, London
- Malkiel, B. (1995), "Returns from Investing in Equity Mutual Funds 1971-1991," *Journal of Finance* 50, pp. 549-589
- Merton, R. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41 pp. 1151-1168
- Rogalski, R. (1984), "New Findings Regarding Day of the week Returns over Trading and Non Trading Periods: A Note," *Journal of Finance* 39, pp. 1603-1614
- Sakakibara, S., T. Yamasaki and K. Okada (2009), "The Calendar Structure of the Japanese Stock Market," *Discussion Paper Series 2009-31 Kobe University*
- Shiller, R. (1990), "Market Volatility and Investor Behavior," *American Economic Review*, 80 pp. 58-62
- Sutan, A. and M. Willinger (2009), "Guessing with Negative Feedback: An experiment," *Journal of Economic Dynamic & Control*, 33, pp. 1123-1133
- 乾孝司・奥村学 (2006), 「テキストを対象とした評価情報の分析に関する研究動向」自然言語処理, 13, pp. 201-241
- 岡田克彦 (2007), 「裁定取引が機能不全に陥る諸要因について」ビジネスアンドアカウンティングレビュー 第2号, pp. 43-56