

ニュースと株価に関する実証分析 ーディープラーニングによるニュース記事の評判分析—

五島圭一高橋大志CMA

目 次

- 1. はじめに
- 2. データ
- 3. 分析方法

- 4. 分析結果
- 5 まとめ

本研究は、人工知能分野において関心を集めているディープラーニングを用いて、ロイターニュースを指標化し、株式価格との関連性を分析したものである。分析の結果、ニュースは、マーケットに影響を与えている一方で、マーケットに対して後追いで反応していること、リターンリバーサルが見られ、センチメントに関する情報を有していること、小型株に対する影響が相対的に大きく、長続きしていること、などの結論を見いだした。

1. はじめに

近年、画像・音声・テキスト情報などの非構造 化データを資産価格分析に用いる試みが模索され ている。非構造化データは、財務データや株価チャート、格付けレーティングなどの構造化データ に比べると扱いが難しいことから、従来あまり利 用されてこなかった。しかしながら、これら非構造化データには構造化データに反映されていない情報が含まれている可能性があり、非構造化データの分析を通じ、超過収益の源泉となる有用な情報を獲得できる可能性がある。本研究では、非構造化データの中でも、ニュース記事のテキスト情報に焦点を当て、分析を試みた。



五島 圭一(ごしま けいいち)

2012年慶應義塾大学経済学部卒業。14年慶應義塾大学大学院経営管理研究科博士前期課程修了。同年より、東京工業大学大学院総合理工学研究科知能システム科学専攻博士後期課程に在籍。



高橋 大志(たかはし ひろし)

慶應義塾大学大学院経営管理研究科・慶應義塾大学ビジネススクール教授。東京大学工学部卒業。富士写真フイルム(現富士フイルム)研究員、三井信託銀行(現三井住友信託銀行)シニアリサーチャー。筑波大学大学院修士課程修了。同大学大学院博士課程修了。博士(経営学)。岡山大学大学院社会文化科学研究科准教授、キール大学経済学部客員研究員、慶應義塾大学大学院経営管理研究科准教授を経て、2014年より現職。

テキスト情報を用いた資産価格分析の先行研究 として、例えば、Antweiler and Frank [2004] では、インターネット掲示板の投稿内容には株価 の予測可能性はないとしながらも、投稿数の増加 はその後の株価変動率の上昇を予想し得ることを 報告している。Tetlock [2007] はWall Street Journal のコラムから悲観度を評価し、ダウ工業 平均株価との関連性を見いだしており、ニュース 記事には、ファンダメンタルズ情報だけでなく、 マーケットのセンチメント情報を有していると報 告している。また、Tetlock et al. [2008] では、 個別企業単位で、Wall Street Journal とDow Jones News Serviceを分析したところ、将来の 株価及び企業業績を予想できると報告しており、 この結果は、ニュース記事の内容には、ファンダ メンタルズ情報が含まれており、即座に価格に反 映されることを示唆している。

更に、資産価格との関連性だけでなく、金融市 場における投資家行動との関連性に言及する報告 が見られる。Engelberg et al. [2012] では、 Dow Jones Market Newsを分析し、空売り筋は ニュース配信後に増えるとしており、投資家はニ ュースを予測して取引するのではなく、優れたニ ュース処理能力によって超過収益を獲得している ことを示唆している。そして、ネガティブニュー スを用いることで、2年半の間に180%もの累積 リターンを獲得できることを報告している。また、 Dougal et al. [2012] では、Wall Street Journal のコラムから、執筆したジャーナリストごとに文 章の特徴を抽出し、株式市場との関連性を見いだ しており、ジャーナリストが投資家行動に大きな 影響を与えることを報告している。他にも、 Garcia 「2013] では、1905年から2005年までの 100年間のNew York Timesを分析対象とし、景 気後退時において、ニュースの内容が株価の予測

に役立つことから、不況時に株式市場に対して投 資家のセンチメントが大きな影響を持つことに言 及している。

日本証券市場を対象とした同様の研究報告がさ れている。丸山他「2008]では、Yahoo! Finance 掲示板の投稿内容を分析し、投稿数がボラティリ ティ、出来高の先行指標であることを報告してい る。上瀧他[2009]は、クレジットに焦点を当 てた分析を行っており、低格付けの社債とニュー スの間に強い関連性が見られることを見いだして いる。岡田・羽室「2011」は、ブルームバーグ配 信のニュース記事から、銘柄ごとに楽観・中立・ 悲観の分類を行い、そこから集約された市場のセ ンチメント指数が、市場ボラティリティとの間に 負の相関があり、センチメント指数が先行するこ とを報告している。また、五島・高橋 [2016] では、日経QUICKニュースをテキストマイニン グすることで、CSR活動と株価との関連性を見い だしている。最近では、会計分野においても、テ キスト分析が取り入れられている (Allee and DeAngelis [2015])

このようにテキスト情報を用いた様々な分析がされている一方で、異なる結論を報告している研究も存在する。沖本・平澤 [2014] では、Tetlock [2007] と同じ検証モデルによって、ニュース記事を分析したところ、ファンダメンタルズ情報のみを有していると報告しており、Tetlock [2007] とは異なる結論を見いだしている。このような異なる結論に至る要因の一つとして、ニュース記事が持つポジネガ情報の推定精度に問題がある可能性がある。

この問題に対して、近年、人工知能分野においてブレイクスルーとして注目を集めているディープラーニング(深層学習)を用いることで、より精緻なニュースとマーケットとの関連性を分析で



きる可能性がある。ディープラーニングは、これまで人手で行っていた非構造化データからの特徴量抽出をコンピューターが自動で行うことを可能にし、なおかつ、より精度の高い特徴量抽出が可能であることが報告されている(Hinton et al. [2006])。そして、テキスト分析においても、ディープラーニングの適用により、従来手法に比べると高い性能が報告されている(Socher et al. [2013])。そこで本研究では、ニュース記事のテキスト情報に焦点を当て、ディープラーニングによるニュース記事の評判分析を行い、株価との関連性について分析を試みた。次章は、本分析で用いるデータに触れ、第3章では分析方法、第4章では分析結果を記す。第5章は、まとめである。

2. データ

(1) マーケットデータ

マーケットデータについては、Thomson Reuters Datastreamから、TOPIX指数、東証一部の出来高を用いた。また、㈱金融データソリューションズが提供する日本版Fama-Frenchベンチマークからサイズファクター・リターン(SMB)の日次データを使用した。

(2) ニュースデータ

ニュースデータについては、ロイターニュースを用いた。ロイターニュースは、世界で最も広く知られたニュース提供会社の一つであるトムソンロイター社が配信しているニュースである。本研究では、東京株式市場に関する英文のニュース記事を分析対象とし、ニュース記事の本文を利用した。また、タグ情報については、ニュース記事の配信日時を利用した。

テキスト分析に用いられるテキスト情報は、主に、企業が発信する情報、メディアが配信する情報、インターネットにおける投稿内容の三つに分類される(Kearney and Liu [2014])。企業が発信するニュース記事については、有価証券報告書や経営者による収支報告、アニュアルリポートなどが挙げられる。また、メディアが配信するニュース記事については、新聞記事や雑誌、ロイターや日経QUICKなどの専門ニュースが挙げられる。インターネットにおける投稿内容については、Yahoo! Finance掲示板やRanking Bullなどへの書き込みが挙げられる。

ロイターニュースは、メディアが配信する情報に分類されるものである。メディアが配信する情報は、各団体が配信する一次情報に比べると、各メディアの記者やアナリストによる情報の取捨選択が行われており、社会や市場に対して相対的に重要な情報が含まれていると考えられる。また、ロイターニュースについては日本証券市場に参加している多くの投資家が閲覧するメディアであり、新聞やテレビのニュースに比べ、イベントからニュース配信までのラグが小さいのも特徴である。分析対象期間は03年1月1日から15年5月31日とした。この間に配信された東京株式市場に関連する英文のロイターニュースを全て用いて分析を行う。

3. 分析方法

(1) ニュース指標の作成について

本研究では、ディープラーニングによるニュース指標の作成と、比較を行うために代表的な従来手法であるナイーブベイズ分類器とLoughran and McDonald Financial Sentiment Dictionaries (以下、LM辞書) によってもニュース指標の作成



を同様に行った。指標の作成には、Tetlock *et al.* [2008] 及びLoughran and McDonald [2011] を参考にした。ここでは、作成方法の詳細について、記す。

(2) 前処理について

分析を行う前のデータ整形について記述する。本分析では、まず、15時以降に配信されたニュース記事については翌営業日に編入し、市場休業日に配信されたニュース記事に関しても同様に、翌営業日に編入し、分析を進めた。これは、マーケットが閉まっている間に、配信されたニュース内容については、直後の営業日において、価格に反映されると仮定したためである。次に、文単位でポジティブな表現とネガティブな表現を集計するために、複数の文から構成されているニュース記事については、文を抽出している(注1)。一連の処理により、文の数は、990,628となった。

(3) ディープラーニングとナイーブベイズ分類器 によるニュース指標の作成方法

ディープラーニングとナイーブベイズ分類器に よって、それぞれ、全てのニュース記事の文に対 してポジネガ分類を行う。

本研究では、ディープラーニングのモデルとして、Recursive Neural Networkを用いた。これは、

文の句構造に従って再帰的に句(文)ベクトルを計算するモデルであり、テキストデータの評判分析において高い性能が報告されているディープラーニングモデルの一つである。文書を構文木解析したのち、ニューラルネットワークによって、構文木の形に沿って単語ベクトルの合成を行うことで、文を固定長のベクトルで表現するため、文の構造を考慮したベクトルを獲得することが可能となる。更に、Recursive Neural Networkの中でも、高い分類精度が報告されているSocher et al. [2013] が提案したRecursive Neural Tensor Network (注2)を用いた。

また、ナイーブベイズ分類器は、テキスト分類 分析において広く用いられる機械学習手法の一つ である。ベイズの定理に基づく確率的分類器であ り、自然言語処理分野ではベンチマークとされる ことが多く、本分析においても比較対象とした(奥 村・高村[2010])(注3)。

本分析の学習データについては、Socher *et al.* [2013] で用いられた11,855文から構成されている文書データを使用している (注4)。分類クラスについては、5クラス (Very Negative、Negative、Neutral、Positive、Very Positive) に分類した。

次いで、営業日ごとに分類された文をクラスご とに集計し、各クラスの文の数に対して、

⁽注1) 抽出については、一定のルールの下でプログラムにより行っている。本分析では、テキスト情報を分析 対象としたため、英文法上、センテンスになっているテキストのみを抽出した。

⁽注2) Recursive Neural Tensor Networkでは、テンソルで単語ベクトルを行列に変換してから、単語ベクトルとの積を計算することで、学習するパラメータ数を抑えながら、線形変換による句ベクトル合成を行っている(岡崎 [2015])。線形変換を行っていることによって、より表現力の高い文書ベクトルを獲得することが可能となっている。活性化関数として、中間層にはハイパボリックタンジェント関数、出力層にはソフトマックス関数が用いられている。

⁽注3) 本研究では、ナイーブベイズ分類器における、あるクラスの学習データに存在しない単語を含む文書は、そのクラスに分類されないというゼロ頻度問題を解決するために、事前に単語の出現頻度を1ずつ加算するラプラススムージングを行っている。すなわち、 $\alpha=2$ のディリクレ分布を事前確率分布としている。



Negativeには-1、Very Negativeには-2を掛け合わせた後、Neutralに分類された文以外の文の数で割ることでスコアを作成する。Tetlock et al. [2008] では、ポジティブな単語は説明力が弱いため、ネガティブな単語のみを計数して、ニュース記事の悲観度の定量化を行っており、本研究においても先行研究に倣い、ネガティブな文のみを計数し、ニュース記事の定量化を行った。具体的には、以下の数式で表される。

$$Score_t = \frac{-N - 2VN}{VP + P + N + VN}$$

VPはVery Positiveに分類された文の数、Pは Positiveに 分類 された文の数、VNはVery Negativeに分類された文の数、NはNegativeに分類された文の数を、それぞれ表している。ここで、Tetlock [2007] と同様に、各営業日のスコアを全体のスコアを用いて標準化を行う。

$$Index_t = \frac{Score_t - \mu_{Score}}{\sigma_{Score}}$$

以上の手順によってニュース指標を作成し、分 析を行った。

(4) LM辞書によるニュース指標の作成方法

ここでは、LM辞書によるニュース指標の作成について、詳細を記す。LM辞書は、Loughran and McDonald [2011] にて用いられた単語リスト (注5) であり、ファイナンス分野における

英文テキスト分析の際に広く用いられる辞書である。辞書には、ポジティブな単語が354、ネガティブな単語が2,355、それぞれ定義されている。

LM辞書をもとに、営業日ごとにニュース記事内の単語を計数し、ネガティブな単語数に対してー1で掛け合わせた値を、計数した単語数で割ることでスコアを算出した。具体的には、以下の数式で表される。

$$Score_t = \frac{-N}{P+N}$$

PとNは、それぞれLM辞書で定義されている ポジティブな単語とネガティブな単語を表してい る。営業日ごとにスコアを算出した後、同様に全 体のスコアを用いて、標準化を行う。

$$Index_t = \frac{Score_t - \mu_{Score}}{\sigma_{Score}}$$

以上の手順によってニュース指標の作成を行った。

(5) ニュース指標の基本統計量

図表1は、前述までの手順で作成した各ニュース指標の統計量をまとめたものである。

IndexDLは、ディープラーニングを用いて作成したニュース指標、IndexNBは、ナイーブベイズ分類器を用いて作成したニュース指標、IndexLMは、LM辞書を用いて作成したニュース指標を表している(注6)。

⁽注4) 本研究では、学習データに金融分野の文書ではなく、一般的な文書を用いている。これは、Tetlock [2007] やTetlock et al. [2008] では、ニュース記事のポジネガ推定に心理学者が定義した一般的な辞書が用いられており、本研究はこれらに倣った。ただし、金融分野には独自の語彙が用いられる傾向があるとの報告もあるので、ニュース記事の一部を教師情報とするなどの分析は今後の課題である。文書データとラベルデータについては、http://nlp.stanford.edu/sentiment/にて公開されている。

⁽注5) 単語リストについては、http://www3.nd.edu/~mcdonald/Word_Lists.htmlにて公開されている。

	IndexDL	IndexNB	IndexLM
サンプルサイズ	3,043	3,043	3,043
平均	0	0	0
標準偏差	1	1	1
第1四分位数	-0.62	-0.63	-0.72
第2四分位数	-0.05	-0.02	-0.01
第3四分位数	0.52	0.61	0.70
自己相関係数(1)	0.18 ***	0.33 ***	0.45 ***
ニュース指標間の相関		0.28 ***	0.21 ***

図表1 ニュース指標の基本統計量

(図表注) IndexDLは、ディープラーニングを用いて作成したニュース指標、IndexNBは、ナイーブベイズ分類器を用いて作成したニュース指標、IndexLMは、LM辞書を用いて作成したニュース指標を表している。
***、**、*はそれぞれ有意水準1%、有意水準5%、有意水準10%で値が有意であることを表している。

(出所) 筆者作成。以下同じ

各ニュース指標のサンプルサイズは、03年1月から15年5月までの3,043営業日となっている。自己相関係数(1)は、1次の自己相関係数を表しており、それぞれ、有意水準1%で、IndexDLは0.18、IndexNBは0.33、IndexLMは0.45となっている。IndexDLについては、統計的に有意な値を取っているものの、その水準は相対的に低い。一方で、IndexNBとIndexLMについては、弱い自己相関が見られる。また、ニュース指標間の相関係数は、IndexDLとの相関係数を表しており、有意水準1%でIndexNBは0.28、IndexLMは0.21となっている。ニュース指標間で一定の相関関係は見られるものの、その水準は相対的に低い。

(6) VARモデルによるニュース指標の分析

前節にて作成されたニュース指標を基に、分析を行う。分析方法は、Tetlock [2007] と日本株式市場で同様の分析を行った沖本・平澤 [2014] に準ずる。具体的には、以下の4つのVARモデル(注7)によって、ニュース記事が持つ情報について、情報理論・センチメント理論・無情報理論の3点の仮説に基づき、分析を進める。

$$Tpx_{t} = \alpha_{1} + \sum_{j=1}^{5} \beta_{1j} Tpx_{t-j} + \sum_{j=1}^{5} \gamma_{1j} Index_{t-j} + \sum_{j=1}^{5} \delta_{1j} Vol_{t-j} + \varepsilon_{1t}$$
(1)

Index_t =
$$\alpha_2 + \sum_{j=1}^{5} \beta_{2j} Tpx_{t-j} + \sum_{j=1}^{5} \gamma_{2j} Index_{t-j}$$

+ $\sum_{j=1}^{5} \delta_{2j} Vol_{t-j} + \varepsilon_{2t}$ (2)

- (注6) 手法の差異として、LM辞書(ポジネガ辞書)は事前に人手で定義した単語リストに基づいて文中の単語を計数することで文のポジネガを算出する方法であり、ナイーブベイズ分類器とディープラーニングは教師データから統計的性質に基づいて単語及び文のポジネガを算出する方法である。このうち、ディープラーニングは文の構造をも考慮したテキストマイニング手法であることから、より精緻な分析が行えることが想定される。
- (注7) 小型株への影響について、沖本・平澤 [2014] では考察されていなかったが、Tetlock [2007] では考察されているため、本研究では併せて分析を行った。



$$Vol_{t} = \alpha_{3} + \sum_{j=1}^{5} \beta_{3j} Tpx_{t-j} + \sum_{j=1}^{5} \gamma_{3j} Index_{t-j}$$

$$+ \sum_{j=1}^{5} \delta_{3j} Vol_{t-j} + \sum_{j=1}^{5} \psi_{3j} |Index_{t-j}| + \varepsilon_{3t}$$
(3)

$$SMB_{t} = \alpha_{4} + \sum_{j=1}^{5} \beta_{4j} Tpx_{t-j} + \sum_{j=1}^{5} \gamma_{4j} Index_{t-j} + \sum_{j=1}^{5} \delta_{4j} Vol_{t-j} + \varepsilon_{4t}$$
(4)

Tpxは、TOPIXの日次対数収益率 (%)、Indexは、ニュース指標、Volは、東証一部の日次出来高の対数値、SMBは、サイズファクター・リターンをそれぞれ表している。次数が5であるのは、過去5日間のTOPIXの日次リターン、出来高、ニュース指標から影響を受けると仮定しているためである。

4. 分析結果

(1) ニュース指標が株式リターンに与える影響に ついて

まず、(1)式の結果からニュース指標がTOPIX リターンに与える影響について考察する。ここで 重要なのは、(1)式における γ_{1j} である。 γ_{1j} が、い ずれかのjで、プラスになり、その後マイナスに なるのであれば、株式市場に対するニュースの影 響は一時的なものも含んでおり、ファンダメンタ ルズだけでなくセンチメントに関する情報も有し ていることになる。しかしながら、プラスの影響 がそのまま残り続けるのであれば、それはファン ダメンタルズに関する情報のみを有していること になる。また、γι,がプラスにもマイナスにもな らず、株式市場には影響を与えない場合では、ニ ュースは何も情報を持たないことになる。以上を 踏まえて、分析結果を考察する。図表2は、(1)式 のy」をまとめたものである。IndexDLは、ディ ープラーニングを用いて作成したニュース指標、 IndexNBは、ナイーブベイズ分類器によって用 いて作成したニュース指標、IndexLMは、LM辞 書を用いて作成したニュース指標を表しており、 それぞれ(1)式によって推計されたγ_{1,1}とNewey-Westの標準誤差を用いて算出したt値をまとめ たものである。

分析結果を見ると、IndexDLにおいて、有意水準 $1\% c_{\gamma_{11}}$ が0.070となり、ニュース指標は翌営業日の株式リターンにプラスの影響を与えていることが観測された。また、 γ_{14} は有意水準 10% cで -0.062となることから、ラグ 4 営業日で株価がリバウンドしており、リターンリバーサルが観測された。これは、ニュース指標は株式市場に対して影響を与えるものの、ニュースの影響は一時的

図表2 ニュース指標がTOPIXリターンに与える影響について

r_{1j}	Inde	IndexDL		IndexNB		IndexLM	
r 11	0.070 ***	(3.122)	0.003	(0.125)	0.049 *	(1.794)	
r 12	-0.047	(-1.563)	0.044	(1.639)	0.006	(0.200)	
γ ₁₃	-0.012	(-0.455)	-0.004	(-0.161)	0.000	(0.002)	
γ ₁₄	-0.062 *	(-1.771)	-0.002	(-0.080)	-0.008	(-0.282)	
γ_{15}	0.006	(0.248)	0.004	(0.163)	-0.052 *	(-1.881)	

(図表注)表は、03年 1 月から15年 5 月までの3,043営業日の各ニュース指標と TOPIXの日次対数収益率を用いて、VARモデルによって推計した値をま とめたものである。***、**、*はそれぞれ有意水準 1 %、有意水準 5 %、有意水準10%で値が有意であることを表している。また、括弧内は誤差 項について不均一分散とラグ次数 5 の系列相関に対して頑健なNewey-Westの標準誤差を用いて算出した t 値を表している。以下同じ。

なものも含んでおり、ファンダメンタルズだけで なくセンチメントに関する情報も有していること になる。また、IndexLMについても、有意水準 10%で γ_{11} が0.049となり、ラグは異なるが γ_{15} は有 意水準10%で-0.052となることから、同様の傾 向が見られた。これらの結果は、リターンリバー サルが観測されず、ニュース指標にはセンチメン ト情報を有していないとする沖本・平澤「2014] とは対照的な結果である一方で、リターンリバー サルが見られ、センチメント情報を有するとする Tetlock [2007] とは整合的な結果となった。更 に、米国株式市場を分析対象としたTetlock [2007] では、4営業日後において有意にリター ンリバーサルが観測されることを報告しており、 日本株式市場を対象とした本研究においても同じ 4営業日後に有意にリターンリバーサルが観測さ れることは、興味深い結果の一つである。しかし ながら、IndexNBでは、どのγιについても統計 的に有意な結果が得られなかった。ナイーブベイ ズ分類器は、取り扱いが容易である一方、単語出 現の独立性など強い仮定の下、分析を行うことか ら、十分な精度が得られていない可能性があ る (注8)。

(2) 株式リターンがニュース指標に与える影響について

次に、TOPIXリターンがニュース指標に与える影響について考察する。ここでは、(2)式における β_{2j} が、重要な値となる。もし、ニュースが株式市場の動きに反応しているのみであったら、いずれかのjで β_{2j} がプラスになり、ニュースがマーケットに対して後追いで反応していることになる。一方で、 β_{2j} が有意でなかったら、それは過去の株式市場の動きとニュースとは、関係ないことになる。これらを踏まえて、分析結果を解釈する。 図表 3 は、(2)式によって推計された β_{2j} と Newey-Westの標準誤差を用いて算出した t 値をまとめたものである。

因及3 TOPIXグラーフがニュース指標に子んる影音について								
β_{2j}	IndexDL		IndexNB		IndexLM			
β_{21}	0.024 *	(1.659)	-0.003	(-0.295)	0.007	(0.674)		
β_{22}	0.005	(0.476)	-0.012	(-0.993)	-0.004	(-0.324)		
β_{23}	-0.012	(-0.600)	-0.010	(-0.745)	-0.003	(-0.279)		
${\cal B}_{24}$	0.012	(1.000)	0.004	(0.343)	0.003	(0.323)		
$oldsymbol{eta}_{25}$	0.002	(0.124)	0.000	(0.023)	-0.008	(-0.726)		

図表3 TOPIXリターンがニュース指標に与える影響について

⁽注8) 基礎研究ではディープラーニングを用いることで、従来手法より高い精度でテキストの評判分析が可能であることが報告されており、本分析でも、ディープラーニングにより相対的に高い精度の結果が得られている可能性がある。詳細な分析は今後の課題である。

⁽注9) ニュース記事には、過去のマーケット状況を記述したものも存在しており、本分析結果と整合的と捉えられる。詳細な分析は、今後の課題である。



れていない。IndexNBについても、 β_{2j} において、統計的に有意な値を獲得できていない。これらの結果は、新たな手法を通じ、従来の手法では見いだすことが困難な結果を得られる可能性を示すものである。詳細な分析は、今後の課題である。

(3) ニュース指標が出来高に与える影響について

三番目に、ニュース指標が出来高に与える影響 について考察する。ここでは、(3)式におけるy3i と ψ_3 が、重要な値となる。Coval and Shumway [2001] やAntweiler and Frank [2004] では、 メディアと取引コスト・流動性・出来高との関連 性について言及しており、(3)式でも同様にIndex。 が取引コストの代理変数となるならば、いずれか のγ₃₁においてプラスとなり、ニュース指標が小 さくなったとき、出来高を減らす影響が見られる はずである。一方で、(3)式の|*Index_{ti}*|が投資家の センチメントの代理変数となるならば、いずれか のψっがプラスとなり、ニュース指標の絶対値が 大きくなったとき、出来高を増やす影響が見られ るはずである。これは、センチメントについて平 均からの乖離が大きくなると、流動性トレーダー が株の売買を行い、それに対してマーケットメイ カーが取引を行うため、出来高が増えるとする

Campbell *et al.* [1993] やDeLong *et al.* [1990] の理論に基づくものである。以上を踏まえて、分析結果を解釈する。**図表4**は、(3)式によって推計された γ_{3j} 及び ψ_{3j} とNewey-Westの標準誤差を用いて算出した t 値をまとめたものである。

分析結果を見ると、IndexDLでは、y₃₄は有意 水準10%で0.008となるものの、γ,,は有意な値は 取らないため、翌営業日に影響を与えている結果 とはならなかった。そのため、取引コストの代理 変数になっていないことを示している。一方で、 ψ_{21} は有意水準10%で0.009となることから、ニュ ース指標の絶対値は出来高に影響を与えることに なり、センチメントの代理変数となっている可能 性を示している。ここでも、Tetlock [2007] と 整合的な結果となった。しかしながら、他のニュ ース指標について、IndexNBでは、γ₃₁が有意な 値を取ることから、取引コストの代理変数となる 可能性があるが、 ψ_{31} は有意な値を取らず、セン チメントの代理変数とはならない可能性がある。 より適切な代理変数による分析は今後の課題であ る。

(4) ニュース指標が小型株に与える影響について 最後に、ニュース指標が小型株に与える影響に

r_{3}	, Inde	IndexDL		IndexNB		IndexLM	
r_{3}	-0.005	(-1.158)	0.011 ***	(2.705)	-0.003	(-0.528)	
γ_{33}	0.006	(1.275)	0.000	(-0.107)	-0.004	(-0.966)	
γ_3	0.003	(0.687)	0.006	(1.232)	0.004	(0.843)	
γ_3	0.008 *	(1.957)	-0.007 *	(-1.834)	-0.007	(-1.525)	
γ_{3}	-0.001	(-0.338)	0.000	(0.028)	0.011 **	(2.314)	

図表4 ニュース指標が出来高に与える影響について

$\psi_{_{3j}}$	IndexDL		IndexNB		IndexLM	
₩ ₃₁	0.009 *	(1.710)	-0.002	(-0.336)	-0.010	(-1.375)
$\psi_{_{32}}$	0.001	(0.215)	-0.008	(-1.138)	0.011	(1.596)
$\psi_{_{33}}$	0.004	(0.688)	0.003	(0.425)	0.008	(1.312)
$\psi_{_{34}}$	0.001	(0.274)	0.013 **	(2.362)	0.002	(0.300)
$\psi_{_{35}}$	0.011 *	(1.914)	0.007	(1.123)	-0.001	(-0.135)

ついて考察する。ここでは、(4)式における γ_{4j} が 重要な値となる。時価総額の小さい企業は、個人 投資家の影響が相対的に大きい、情報が入手しに くいなどの特徴があり、ニュースの影響について 大型株とは異なる可能性がある。 γ_{4j} が統計的に 有意な値となるならば、TOPIXリターンに対す る予測力を別にして、ニュース指標は小型株に影 響を与えていることになる。これらを踏まえて、 分析結果を解釈する。**図表5**は、(4)式によって推 計された γ_{4j} とNewey-Westの標準誤差を用いて算 出したt値をまとめたものである。

分析結果を見ると、IndexDLでは、742は有意 水準 5%で0.025、γ₄₃は有意水準 1%で0.032と なり、TOPIXリターンに対する予測力を別にし て、ニュース指標は小型株に影響を与えている結 果となった。 γ_{42} と γ_{43} がプラスとなるのは、小型 株に対するニュース指標の影響が相対的に大き く、また、長続きしていることを意味している。 特に、ニュース指標が2営業日後及び3営業日後 のSMBに影響を与えていることは、小型株に対 してニュースの情報が徐々に反映されることを示 している。これは、Tetlock [2007] においても 4営業日後のSMBへ影響を与えているとの報告 をしていることから、同様の傾向が観測された。 一方で、IndexNBとIndexLMのγ₄は統計的に有 意な値は得られていない。新たな手法(IndexDL) を通じ、従来の手法 (IndexNB、IndexLM) で

は見いだすことが困難な結果を得られている点は 興味深い。詳細な分析は今後の課題である。

5. まとめ

本研究では、ニュース記事のテキスト情報に焦点を当て、従来手法との比較を通じて、ディープラーニングによるニュース記事の評判分析を行い、株価との関連性について分析を行った。分析の結果、ニュース指標はマーケットに対して後追いで反応している一方で、マーケットに対して後追いで反応している可能性があること、リターンリバーサルが見られることからニュース指標にはマーケットのセンチメントに関する情報を有している可能性があること、小型株に対するニュース指標の影響が相対的に大きく、長続きしていること、などの結論を見いだした(注10)。新たな手法を通じ、国内株式市場において、ニュース指標によるリターンリバーサルや株価の後追いが生じている可能性を示した点は本論文の特徴の一つとして挙げられる。

テキストデータの分析は、数値データの分析と 比較し、相対的に誤差が大きいと考えられ、より 適切な手法を用いた分析は今後の課題である。ま た、他国の証券市場や他のメディアの分析、証券 投資への応用、同営業日のニュース指標の影響の 考慮などについても今後の課題である。

凶表 5	ニュース指標かSMBに与える影響について

$r_{_{4j}}$	IndexDL		Inde	IndexNB		IndexLM	
r 41	-0.020	(-1.644)	0.009	(0.731)	0.003	(0.202)	
r_{42}	0.025 **	(2.352)	-0.006	(-0.479)	-0.013	(-1.066)	
γ_{43}	0.032 ***	(2.693)	0.002	(0.184)	0.006	(0.454)	
r 44	0.015	(1.249)	-0.008	(-0.700)	0.008	(0.603)	
\mathcal{Y}_{45}	0.010	(0.693)	-0.007	(-0.513)	0.005	(0.403)	

(注10) これらの結果は、米国株式市場を分析対象とした先行研究と整合的な結果である。



本稿の作成に当たり、匿名のレフェリーから貴重なコメントをいただいた。また、本研究は、JSPS科研費23310106の助成を受けている。記して感謝したい。

(参考文献)

- 岡崎直観 [2015]「単語の分散表現と構成性の計算モデルの発展」2015年度人工知能学会全国大会(第29回)OS-1 意味と理解のコンピューティング(2).
- 岡田克彦・羽室行信 [2011] 「相場の感情とその変動 一自然言語処理で測定するマーケットセンチメン トとボラティリティー」『証券アナリストジャーナ ル』49(8)、37-48頁.
- 沖本竜義・平澤英司 [2014] 「ニュース指標による株式市場の予測可能性」『証券アナリストジャーナル』 52 (4)、67-75頁.
- 奥村学・高村大地 [2010] 『言語処理のための機械学習入門』コロナ社.
- 五島圭一・高橋大志 [2016]「ニュースを用いたCSR 活動が株価に与える影響の分析」『ジャフィー・ジャーナル』、8-35頁.
- 上瀧弘晃・高橋悟・高橋大志 [2009]「クレジット市場におけるヘッドラインニュースの効果」『日本ファイナンス学会第17回大会予稿集』、113-122頁.
- 丸山健・梅原英一・諏訪博彦・太田敏澄 [2008]「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』46 (11・12)、110-127頁
- Allee, K. and M. D. DeAngelis [2015] "The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion," *Journal of Accounting Research* 53(2), pp.241-274.
- Antweiler, W., and M. Z. Frank [2004] "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *Journal of Finance* 59(3), pp.1259-1294.
- Campbell, J. Y., S. J. Grossman, and J. Wang [1993] "Trading Volume and Serial Correlation in Stock Returns," *Quarterly Journal of Economics* 108(4), pp.905-939.

- Coval, J. D., and T. Shumway [2001] "Is Sound Just Noise?," *Journal of Finance* 56(5), pp.1887-1910.
- DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann [1990] "Noise Trader Risk in Financial Markets," *Journal of Political Economy* 98(4), pp.703-738.
- Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons [2012] "Journalists and The Stock Market," *Review of Financial Studies* 25(3), pp.639-679.
- Engelberg, J., A. V. Reed, and M. C. Ringgenberg [2012] "How Are Shorts Informed? Short Sellers, News, and Information Processing," *Journal of Financial Economics* 105(2), pp.260-278.
- Garcia, D. [2013] "Sentiment during Recessions," *Journal of Finance* 68(3), pp.1267-1300.
- Hinton, G. E., S. Osindero and Y. Teh [2006] "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation* 18(7), pp.1527-1554.
- Kearney, C. and S. Liu [2014] "Textual Sentiment in Finance: A Survey of Methods and Models," *International Review of Financial Analysis* 33, pp.171-185.
- Loughran, T. and B. McDonald [2011] "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance* 66(1), pp.35-65.
- Socher, R., A. Perelygin, J. Wu, J. Chuang, C. Manning, A. Ng and C. Potts [2013] "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," *Conference on Empirical Methods in Natural Language Processing*.
- Tetlock, P. C. [2007] "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance* 62(3), pp.1139-1168.
- Tetlock, P. C., M. Saar-Tsechansky and S. Macskassy [2008] "More Than Words: Quantifying Language to Measure Firms' Fundamentals," *Journal of Finance* 63(3), pp.1437-1467.

(この論文は投稿論稿を採用したものです。)