

# 過去のニュース発信傾向及びニュースのセンチメントによる日本株式市場の反応分析

Analysis of the reaction of the Japanese stock market by  
the tendency of past news transmission and news sentiments

ユン ソンジェ  
尹 聖在<sup>1</sup> 高橋大志<sup>1</sup>

Sungjae Yoon<sup>1</sup>, Hiroshi Takahashi<sup>1</sup>

<sup>1</sup>慶應義塾大学 大学院経営管理研究科

<sup>1</sup> Graduate School of Business Administration, Keio University

**Abstract:** In this research, we studied the relationship between news sentiment polarities (positive, negative, neutral) and volume using news data and high-frequency data for the Japanese stock market. For the classification of news sentiment, the FinBERT model, which is a machine-learning method proposed for analysis in the field of finance, was used. The tendency of past news transmission was taken into consideration, and the reaction of the volume to the sentiment of each alert news was evaluated. As a result of the analysis, it was found that the trading volume tends to increase when the sentiment polarities (positive, negative, neutral) of past news in the same company are mixed. These are consistent with the previous studies using daily data and are interesting consequences that contribute to the clarification of the price change mechanism in the stock market.

## 1 はじめに

近年、ニュース、有価証券報告書などのテキスト情報に焦点を当てた先行研究が数多く報告されている。アメリカ市場を対象とした研究[1][2][3][4]をはじめ、日本市場を対象とした研究[5][6]も含め数多くの分析が報告されている。例えば、Antweiler & Frank(2004)[1]は、Yahoo! Finance Raging Bull に掲示された米国株式市場の上場銘柄45社の150万件以上のメッセージを分析し、日次データを用い、ニュースと市場の変化（出来高、ボラティリティ）に密接に関連しているとの結果を報告しており、さらに投稿されているメッセージ間の不一致は出来高の増加に関連しているとの主張を行っている。

本研究では、ニュース及び高頻度データを用いて過去のニュース発信傾向及びニュースのセンチメントによる日本株式市場の反応に関する分析を行う。

## 2 先行研究

テキストと市場の関係を分析した先行研究も数多く報告されている。例えば、アメリカ市場の研究に関して、Tetlock(2007)[2]、Tetlock et al.(2008)[3]ではメディアの悲観度による属性や影響を取り入れた分析を報告している。また、Dougal et al.(2012)[4]は、Dow Jones Industrial AverageとWall Street Journalのコラムから文章の特徴を取り出し、コラムを執筆したジャーナリストが投資家の行動に大きな影響をもた

らすと報告している。

近年、日本株式市場を対象とした研究報告も見られる。沖本・平澤(2014)[5]は、QUICK社の端末により配信される日経ニュースからニュース指標を作成し、株式市場への影響を分析し、翌営業日のリターンや出来高と有意に正の影響をもつこと、先行研究で報告されたリバウンドが観測されないことを報告している。また、五島・高橋[6]は、Thomson Reutersのニュースを指標化し、ニュース指標とマーケットの反応との関係を分析している。

近年、Googleによって開発された自然言語処理の手法のBERT (Bidirectional Encoder Representations from Transformers) [7]が注目を浴びているが、BERTにファイナンス分野の辞書を事前学習させ、ニュース分類 (Positive, Neutral, Negative) をできるFinBERTモデルが提案されている。本研究では、FinBERTモデル (Araci(2019)[8]) を通じたニュース分類からニュースのセンチメント及び過去ニュースの発信傾向を算出し、高頻度データを同時に用いる分析を行うことが特徴の一つである。

## 3 目的

本研究は、ニュース発表が株式市場の出来高に与える影響を明らかにすることを目的とする。本分析では、ファイナンス分野の分析に特化した機械学習

手法の FinBERT モデル<sup>1</sup>を用い、日本市場の上場企業のニュースを分類し、過去のニュース発信の傾向及び市場全体の価格変動の傾向を考慮し、ニュースのセンチメントによる反応分析を行う。

## 4 分析対象及びデータ

分析対象は、日本市場（東証一部）の 2018 年末基準の時価総額上位 50 位以内の銘柄の内、アラートニュース（以下アラート）の分析可能な銘柄（32 社）を対象とした。表 1 では対象銘柄の時価総額を確認できる。分析期間は 2007 年 1 月-2019 年 2 月である。

表 1 対象銘柄の時価総額分布

項目	時価総額（億円）
東証一部の全体	5,621,213(100%)
対象銘柄（32 社）	1,489,789(26.5%)

本研究では、Thomson Reuters 社の英語ニュースデータを用いた。また、株式価格および取引データについては Thomson Reuters 社、Datastream を用いた。

表 2 は、対象銘柄の英語ニュースデータにおける統計量を示したものである。ニュースは 2 種類が存在する。具体的には、(1) ヘッドライン（タイトルを示す）と本文（詳細内容）を含んで発信されるケース（以下、本文）と (2) 緊急や速報として発信され、即時性が高い一方、ヘッドラインのみ発信されるケース（以下、アラート）がある。ニュースの分析においては、アラートと本文いずれも対象としている。本分析では、いずれもヘッドラインのみを用いて分析した。

表 2 対象銘柄のニュースの件数（開場時間のみ）<sup>2</sup>

種類	対象銘柄（32 社）の ニュース数
アラート	1,809
本文	13,226
合計	15,035

なお、株式取引データ（高頻度データ）は対象銘柄の 1 分単位の株価と出来高のデータ、TOPIX の日単位の終値データを利用した。

## 5 分析方法

以下の手順で分析を行う。

1) 分析を行う英語ニュースのヘッドラインの前

処理を行う。ここで、ヘッドラインに記載されている重複ニュースの削除、証券コードの削除、小文字化等を行う。

2) 1)の前処理後のニュースのヘッドラインに FinBERT モデル(Araci(2019)[ 8 ])を適用し、ニュースのセンチメントを 3 パターン（ポジティブ、ニュートラル、ネガティブ）で分類する。

3) 株式市場の開場時間内に発信された各ニュースに対し、発信時刻の分から+1 分~+10 分の間の合計出来高を算出し、以下の式で定義された出来高比率を算出する。

$$\text{出来高比率} = \frac{\text{発信後+1分~10分の間の合計出来高}}{\text{取引日の合計出来高}}$$

4) 株式市場の開場時間内に発信された各ニュースに対し、以下の式で定義されたリターンを算出する。

$$\text{リターン} = \frac{\text{発信後+10分の株価} - \text{発信時の株価}}{\text{発信時の株価}}$$

5) 該当銘柄ニュース発信日の以前の 10 日間における各センチメント分類のニュース（アラート及び本文）の数を集計する。

6) 5)で集計したポジティブ及びネガティブのニュース数を用い、ニュースの発信傾向を以下の式で評価する。

$$\text{ニュースの発信傾向} = \text{ポジティブの数} - \text{ネガティブの数}$$

なお、ニュースの発信傾向は、以下の 3 パターンで分類する。

- ① ニュースの発信傾向 > 0 : ポジティブ寄り状態
- ② ニュースの発信傾向 = 0 : 混在状態  
(ポジティブとネガティブが同数)
- ③ ニュースの発信傾向 < 0 : ネガティブ寄り状態

7) 株式市場の変化を調べるため、TOPIX の日単位の終値データを用いて以下の式で TOPIX の変化率を評価する。

$$\text{TOPIXの変化率} = \frac{\text{取引日の指数} - \text{取引日の10営業日前の指数}}{\text{取引日の10営業日前の指数}}$$

また、TOPIX の変化率は以下の 2 パターンで分類する。

- ① TOPIX の変化率 > 0 : 上昇
- ② TOPIX の変化率 < 0 : 下落

なお、今回の分析では TOPIX の変化率=0 のケースは表れていない。

8) 3)4)5)6)で導出した結果に基づき、以下のケース間の対象企業の緊急ニュース（アラート）<sup>3</sup>の出来高

<sup>1</sup> 本研究で用いた FinBERT モデル(Araci(2019)[ 8 ])では Reuters の TRC2-financial のコーパスを用いる。

<sup>2</sup> 表中の値は開場時間内のニュース及び重複するニュースを除外したものである。なお、1 件のニュースが複数の銘柄の内容を含む場合も存在する。

<sup>3</sup> アラートは緊急性が高い性質を持っており、先行研究[ 9 ][ 1 0 ]では日本株式市場でアラートは本文

比率とリターンの検定 (t 検定) を行う。

① 分析対象

- 混在状態とポジティブ寄り状態
- 混在状態とネガティブ寄り状態

② t 検定の種類：

- 出来高比率：片側、分散が等しくないと仮定
- リターン：両側、分散が等しくないと仮定

9) 3)4)5)7)で導出した結果に基づき、以下のケース間の対象企業の緊急ニュース (アラート) の出来高比率とリターンの検定 (t 検定) を行う。

① 分析対象

- TOPIX の変化率が上昇状態と下落状態 (FinBERT による同分類間の比較)

② t 検定の種類：8)②と同一

10) 3)4)5)6)7)で導出した結果に基づき、以下のケース間の緊急ニュース (アラート) の出来高比率とリターンの検定 (t 検定) を行う。

① 分析対象

- 過去のニュース傾向が 6)②「混在状態」における TOPIX が上昇傾向と下落傾向の状態 (FinBERT による同分類間の比較)

② t 検定の種類：8)②及び9)②と同一

## 6 分析結果

### 6.1. 過去に発信された情報の影響

表 3 及び表 4 は、対象企業の全アラートに対し、過去のニュースの発信傾向が「混在状態」と「ポジティブ寄り状態」、「混在状態」と「ネガティブ寄り状態」の出来高比率とリターンに対し、両グループ間の検定 (t 検定) (分散が等しくないと仮定) を行った結果である。

分析の結果、出来高比率の場合は「混在状態」が「ポジティブ寄り状態」又は「ネガティブ寄り状態」のいずれよりも平均出来高比率が、統計的有意 (\*\*\*) :  $p < 0.01$ ) に高いことを確認できる。一方、リターンについては、統計的有意な差はみられなかった。

### 6.2. 過去発信傾向、TOPIX の変化率の影響

表 5 から表 8 までは、対象企業の全アラートに対し、TOPIX の変化率を考慮した過去のニュースの発信傾向が「混在状態」と「ポジティブ寄り状態」、「混在状態」と「ネガティブ寄り状態」の出来高比率とリターンに対し、両グループ間の検定 (t 検定) (分散が等しくないと仮定) を行った結果である。表 5

より出来高比率が高いことが確認されている。

及び表 6 は TOPIX 変化率が「上昇のみ」のケースであり、表 7 及び表 8 は TOPIX 変化率が「下落のみ」のケースである。

分析の結果、TOPIX 変化率が「上昇のみ」と「下落のみ」のいずれも出来高比率の場合は「混在状態」が「ポジティブ寄り状態」又は「ネガティブ寄り状態」のいずれよりも平均出来高比率が、統計的有意 (\*\*\*) :  $p < 0.01$ ) に高いことを確認できる。一方、リターンについては、統計的有意な差はみられなかった。

### 6.3. 混在状態の TOPIX の変化率による影響

表 9 及び表 10 は過去ニュースの発信傾向が「混在状態」に該当する対象企業の全アラートに対し、TOPIX の変化率が「上昇」と「下落」の出来高比率とリターンに対し、両グループ間の検定 (t 検定) (分散が等しくないと仮定) を行った結果である。分析の結果、出来高比率の場合は「ネガティブ」に分類されたアラートは、「下落」が「上昇」より統計的有意に高い傾向にあることを見出した。一方、リターンについては、統計的有意な差はみられなかった。

## 7 まとめ

本研究では、日本株式市場を対象とし、ニュースと高頻度データを用い、ニュース記事の極性 (ポジティブ、ネガティブ、ニュートラル) と出来高の関連性について分析を実施した。分析の結果、同一企業における過去のニュース記事の極性 (ポジティブ、ネガティブ、ニュートラル) が混在している場合に、出来高が大きくなる傾向を確認した。これらの結果は、日次データを用いた報告と整合的な結果であり、株式市場における価格変動メカニズムの解明に寄与する興味深い結果である。

## 参考文献

- [1] Antweiler, W., Frank, M. Z. : Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *Journal of Finance*, Vol. 59, No. 3, pp. 1259-1293 (2004)
- [2] Tetlock, P. C. : Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance*, Vol. 62, No. 3, pp. 1139-1168 (2007)
- [3] Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S. : More than Words: Quantifying Language to Measure Firms' Fundamentals, *Journal of Finance*, Vol. 63, No. 3, pp. 1437-1467 (2008)
- [4] Dougal, C., Engelberg, J., Garcia, D., and Parsons C. A. : Journalists and The Stock Market, *Review of Financial Studies*, Vol. 25, No. 3, pp. 639-679 (2012)

- [ 5 ] 沖本竜義, 平澤英司 : 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, Vol. 52, No. 4, pp. 67-75 (2014)
- [ 6 ] 五島圭一, 高橋大志 : ニュースと株価に関する実証分析-ディープラーニングによるニュース記事の評判分析-,証券アナリストジャーナル, Vol. 54, No. 3, pp. 76-86 (2016)
- [ 7 ] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805 (2018)
- [ 8 ] Araci, D.: FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models, arXiv:1908.10063 (2019)
- [ 9 ] 尹聖在, 高橋大志 : FinBERT を用いたニュースと株式市場の関連性分析, 2021 年度経営情報学会全国大会予稿集 (2021) (to appear)
- [ 1 0 ] Yoon, S.J., Takahashi, H. : A Study of News and Stock Markets Using FinBERT Models: Evidence from Japan and Korea, Ai-Biz 2021 (2021) (to appear)
- [ 1 1 ] Yoon, S.J., Takahashi, H. : A Study of News and Stock Markets Using FinBERT Models: Evidence from Japan and Korea, Ai-Biz 2021 (2021) (to appear)

表3 過去ニュースの発信傾向における出来高比率のt検定結果  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

t検定の比較対象	統計量項目	Aの値	Bの値	自由度	t	p-value (片側)
混在状態(A)と ポジティブ寄り状態(B)	平均	7.85%	4.98%	1,120	8.814	<b>2.27E-18***</b>
	標準偏差	6.60%	4.85%			
	サンプル数	852	432			
混在状態(A)と ネガティブ寄り状態(B)	平均	7.85%	4.77%	1,471	11.238	<b>1.79E-18***</b>
	標準偏差	6.60%	4.26%			
	サンプル数	852	762			

表4 過去ニュースの発信傾向におけるリターンのt検定結果  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

t検定の比較対象	統計量項目	Aの値	Bの値	自由度	t	p-value (両側)
混在状態(A)と ポジティブ寄り状態(B)	平均	0.022%	-0.033%	1,188	1.084	0.279
	標準偏差	1.076%	0.720%			
	サンプル数	852	432			
混在状態(A)と ネガティブ寄り状態(B)	平均	0.022%	0.005%	1,552	0.348	0.728
	標準偏差	1.076%	0.787%			
	サンプル数	852	762			

表5 過去ニュースの発信傾向における出来高比率のt検定結果 (TOPIXの変化率：上昇のみ)  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

t検定の比較対象	統計量項目	Aの値	Bの値	自由度	t	p-value (片側)
混在状態(A)と ポジティブ寄り状態(B)	平均	7.61%	4.65%	734	7.131	<b>1.25E-12***</b>
	標準偏差	6.63%	4.59%			
	サンプル数	476	262			
混在状態(A)と ネガティブ寄り状態(B)	平均	7.61%	4.59%	771	7.807	<b>9.08E-15***</b>
	標準偏差	6.63%	4.41%			
	サンプル数	476	337			

表6 過去ニュースの発信傾向におけるリターンのt検定結果 (TOPIXの変化率：上昇のみ)  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

t検定の比較対象	統計量項目	Aの値	Bの値	自由度	t	p-value (両側)
混在状態(A)と ポジティブ寄り状態(B)	平均	0.039%	0.019%	734	0.318	0.751
	標準偏差	1.116%	0.579%			
	サンプル数	476	262			
混在状態(A)と ネガティブ寄り状態(B)	平均	0.039%	0.048%	771	0.155	0.877
	標準偏差	1.246%	0.381%			
	サンプル数	476	337			

表7 過去ニュースの発信傾向における出来高比率の t 検定結果 (TOPIX の変化率：下落のみ)  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

t 検定の比較対象	統計量項目	A の値	B の値	自由度	t	p-value (片側)
混在状態(A)と ポジティブ寄り状態(B)	平均	8.14%	5.50%	405	5.061	<b>3.17E-07***</b>
	標準偏差	6.57%	5.20%			
	サンプル数	376	170			
混在状態(A)と ネガティブ寄り状態(B)	平均	8.14%	4.91%	618	8.207	<b>6.61E-16***</b>
	標準偏差	6.57%	4.14%			
	サンプル数	376	425			

表8 過去ニュースの発信傾向におけるリターンの t 検定結果 (TOPIX の変化率：下落のみ)  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

t 検定の比較対象	統計量項目	A の値	B の値	自由度	t	p-value (両側)
混在状態(A)と ポジティブ寄り状態(B)	平均	-0.000%	-0.114%	371	1.314	0.190
	標準偏差	1.024%	0.891%			
	サンプル数	376	170			
混在状態(A)と ネガティブ寄り状態(B)	平均	-0.000%	-0.029%	751	0.416	0.677
	標準偏差	1.024%	0.898%			
	サンプル数	376	425			

表9 「混在状態」の TOPIX 傾向における出来高比率の t 検定結果  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

FinBERT の分類結果	統計量項目	上昇	下落	自由度	t	p-value (片側)
ポジティブ	平均	6.99%	7.62%	238	0.746	0.229
	標準偏差	6.75%	6.31%			
	サンプル数	131	111			
ニュートラル	平均	7.61%	7.24%	318	0.518	0.302
	標準偏差	6.77%	6.55%			
	サンプル数	215	146			
ネガティブ	平均	8.25%	9.74%	242	1.823	<b>0.035**</b>
	標準偏差	6.26%	6.59%			
	サンプル数	130	119			

表10 「混在状態」の TOPIX の変化率におけるリターンの t 検定結果  
(\* : p<0.1, \*\* : p<0.05, \*\*\* : p<0.01)

FinBERT の分類結果	統計量項目	上昇	下落	自由度	t	p-value (両側)
ポジティブ	平均	0.066%	0.075%	239	0.082	0.935
	標準偏差	0.854%	0.768%			
	サンプル数	131	111			
ニュートラル	平均	0.054%	-0.040%	359	0.808	0.605
	標準偏差	1.338%	0.880%			
	サンプル数	215	146			
ネガティブ	平均	-0.014%	-0.022%	208	0.048	0.961
	標準偏差	0.938%	1.350%			
	サンプル数	130	119			