

# ティックデータを用いたニュースと株価との関連性分析

## Analyzing The Relationship between News and The Tokyo Stock Exchange Using Intraday Price Data

五島圭一<sup>1\*</sup> 高橋大志<sup>2</sup> 寺野隆雄<sup>1</sup>  
Keiichi Goshima<sup>1</sup> Hiroshi Takahashi<sup>2</sup> Takao Terano<sup>1</sup>

<sup>1</sup> 東京工業大学 大学院総合理工学研究科

<sup>1</sup> Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

<sup>2</sup> 慶應義塾大学 大学院経営管理研究科

<sup>2</sup> Graduate School of Business Administration, Keio University

**Abstract:** This paper studies the relationship between news and the Tokyo Stock Exchange using intraday price data. Our main finding is that news information almost reflects on stock prices within one second.

## 1 はじめに

増益・増配・M&A・増配・株式分割・自社株買いなどのニュースに対してマーケットがどのように反応するかについて、数多くの実証研究がなされてきた [1]。また、日々発信されるニュースがどのくらい株価変動（リターンの絶対値・ボラティリティ）を説明できるかについて、数多く検証されてきた。近年では、これらニュースの分析に、テキストマイニングが取り入れられている。ニュースをテキストマイニングすることでその内容を評価し、マーケットの反応を観測するアプローチがなされている。これまで観測が難しかったイベントを分析したり、ニュースの頻度だけでなく、テキストのポジネガ度合いを定量化することで、投資家のセンチメントと関連付けたり、テキスト情報を用いた様々な研究報告がある [2]。これら先行研究は日次単位での分析が多いことから、本研究では秒単位での分析を行う。

ニュースが資産価格に影響を与える新しい情報を有していれば、それらのニュースの配信を受けて、価格が変化する可能性がある。一方、既に織り込まれた情報であれば、ニュースは価格に影響を与えない可能性がある。また、ニュースが何も情報をもたないのであれば、ニュースと価格との関連性が見られない可能性がある。本研究では、2009年から2010年の2年間に焦点をあて、ニュースデータをテキストマイニングすることで、ニュースのポジネガを定量化し、ティック

データを用いることで、秒単位でのニュース記事と日本株式市場との関連性の分析を試みる。次章は、本分析で用いるデータに触れ、3章では分析方法、4章では分析結果を記す。5章は、まとめである。

## 2 データ

### 2.1 マーケットデータ・ニュースデータ

本研究では、個別銘柄の株価データについて、Thomson Reuter Datastream から、ティックデータを用いた。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、日本証券市場に関する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、ニュース記事の発信日時とニュースと関連する企業の証券コードを利用した。

## 3 分析方法

### 3.1 分析手順について

ここでは、本研究の分析手順を記述する。はじめに、過去の株価データとニュース記事を基に、ニュース記事にポジネガスコア（ポジティブあるいはネガティブの度合いを表す値）の付与を行った [3]。スコアが付与されたニュース記事を対象として、スコアの値が  $\geq 0.975$  を超えたニュース記事を Positive なニュースとして、

\*連絡先：  
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  
〒226-8502 神奈川県横浜市緑区長津田町 4259-J2-1705  
E-mail: goshima.k.aa@m.titech.ac.jp

$-z_{0.975}$ を下回ったニュース記事を Negative なニュースとして分類した。どちらにも分類されないニュースは、Neutral に分類した。本分析の前処理として、ロイターニュースについて、第一報のニュースのみを分析対象とするため、再送記事と訂正記事について分析対象外とし、さらに、ニュース記事のテキスト情報に注目したため、決算情報のみのニュース記事についても、分析対象外としている。また、ティックデータによる分析を行うため、営業日において9時から15時までに配信されたニュース記事のみを分析対象としている。ただし、学習データには全てのニュースを用いている。選別後の各クラスのニュース記事数については、表1に示す。のべニュース数とは、一つのニュース記事に複数の証券コードが付随している場合、重複して計数した値である。

表 1: のべニュース数

	ニュース数
Positive	859
Neutral	8,113
Negative	634
Total	9,606

次いで、各ニュース配信時刻付近において関連する銘柄の株価リターンを算出し、三分割したクラスごとに集計する。ニュース配信付近のリターンとして、1時間前から1時間後まで10分間隔で計算する。加えて、ニュース配信時刻1秒前から1秒後までのリターンも計算する。

## 4 分析結果

表2と図1は、ニュース配信時刻付近の各クラスの累積平均株価リターンをまとめたものである。分析結果を見ると、全体の傾向として、ニュース配信時刻1秒前から1秒後までに大きく変動していることがわかる。また、ニュースが配信される前から、ポジティブなニュース記事はプラスに、ネガティブなニュース記事はマイナスに累積平均株価リターンが徐々に、変動していることが見て取れる。

表 2: ニュース配信時刻前後の累積株価リターン

分	-50	-40	-30	-20	-10	-1/60	1/60	10	20	30	40	50	60
Positive	0.2 %	0.4 %	0.6 %	0.9 %	1.4 %	1.7 %	4.8 %	4.9 %	4.8 %	4.9 %	4.8 %	4.8 %	4.8 %
Neutral	0.1 %	0.1 %	0.2 %	0.3 %	0.5 %	0.6 %	0.9 %	0.9 %	1.0 %	0.9 %	0.9 %	0.9 %	0.9 %
Negative	-0.2 %	-0.3 %	-0.5 %	-0.7 %	-1.2 %	-1.6 %	-4.4 %	-4.5 %	-4.4 %	-4.4 %	-4.4 %	-4.3 %	-4.3 %

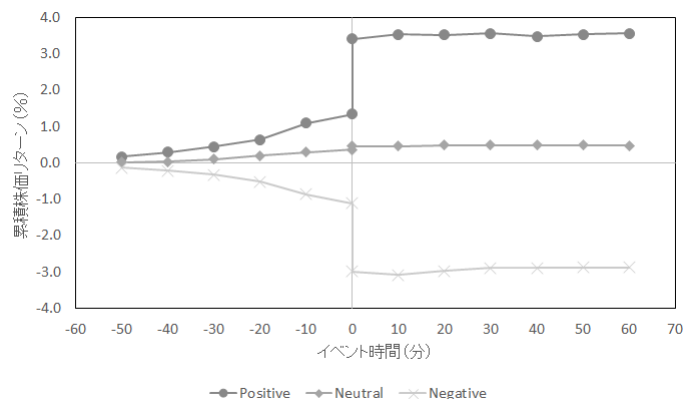


図 1: ニュース配信時刻前後の累積株価リターンの推移

## 5 まとめ

本研究では、ニュースデータをテキストマイニングすることで、ニュースのポジネガを定量化し、ティックデータを用いることで、秒単位でのニュース記事と日本株式市場との関連性の分析を試みた。分析の結果、ポジティブ及びネガティブに分類されたニュースについては、ニュース配信時から1秒以内において株価が大きく変動していることが示された。

## 参考文献

- [1] Campbell, J. Y., A. W. Lo, and A. C. MacKinlay.: *The Econometrics of Financial Markets*, Princeton University Press (1997). 祝迫・大橋・中村・本多・和田訳: ファイナンスのための計量分析, 共立出版 (2003).
- [2] Kearney, C. and Liu, S.: *Textual Sentiment in Finance: A Survey of Methods and Models*, *International Review of Financial Analysis*, Vol. 33, pp. 171-185 (2014).
- [3] 五島圭一・高橋大志: 株価を用いたニュース記事評価に関する研究, 日本ファイナンス学会予稿集 (2015).