

カテゴリーの異なるニュースを用いた日本株式市場の分析 —高頻度データおよび LSTM によるニュース分類—

LSTM Model for Explaining the Association between News Data and Stock Price Fluctuations

高山理璃子¹ 菅愛子¹ 高橋大志¹

Liliko Takayama¹, Aiko Suge¹, and Hiroshi Takahashi¹

¹慶應義塾大学大学院経営管理研究科

¹ Graduate School of Business Administration, Keio University

Abstract: Research on the association between stock price fluctuations and news data is popular and numerous studies have been conducted. This paper aims to improve the accuracy of text classification into three categories (negative, neutral, or positive) by employing high frequency data and LSTM model. One of the novelties of our paper is to use several types of news articles in the analysis.

1 はじめに

近年、情報蓄積の技術が進展し、金融市場分野ではビックデータを用いて深層学習を行い、株式変動を説明する新たな手法が検討されている。テキストマイニング手法を用いたニュースと株式市場の関連性を説明する試みは、このような流れの一つである。これまで日時データや月次データを用いた、ニュースと株式市場に関する分析が報告されているが、ニュース配信による株式市場への影響を観察するには、短期変動に着目し分析を行う必要がある。また、個々のニュース属性を考慮することで、より詳細な分析結果が得られる可能性がある。本論文はニュース配信による株式市場の短期変動及び、ニュース属性の分類に焦点を当て分析を行った。

テキストマイニングを用いた先行研究としては、例えば、Gidofalvi et al.[1]はニュース発信前後 20 分間 (計 40 分) の株価をナイーブベイズ分類器により分類し、ニュースと株価変動との関連性を報告している。また、Tetlock et al.[2][3] は Wall Street Journal を用いて株価と日時ニュースの関連性についてテキストマイニングを行い、過度に悲観的なニュースは株価を押し下げ、取引量を増大させる傾向があることを報告している。Chan[4]は月次の株価変動とヘッドラインニュースの関連性について分析を行い、悲観的なニュース配信の後、株価の下落傾

向が継続している関連性について報告している。より詳細な単位での研究としては Brogaard et al.[5]のミリ秒単位での研究がある。当研究では、高頻度取引が長期的な価格変動の方向と同様に取引することで、市場の効率性に寄与することを見出している。

本研究では、2 種類の異なるニュース：市場全体に関する記事 (マクロニュース) と個別企業に関する記事 (銘柄ニュース) を用いて、株式市場の短期間の変動を観察する。高頻度データを用いたニュースと株式市場の分析では、五島・高橋[6]が挙げられるが、ニュースの特徴ごとの分類は行っていなかった。また Schumaker and Chen[7]は、ニュースをカテゴリー化し、分析をおこなっているが、カテゴリーごとそれぞれ単体で分析をおこなっており、カテゴリー化した上で同時にモデル構築をする研究は限定的である。

2 使用データ

分析には株式データとニュースデータを用い、分析のサンプル対象を日経 225 社とする。いずれも Thomson 社より取得した。

2-1. 株式データ

日経 225 に属する企業を対象に株式価格、取引時間、取引量を取得した。(表 1)

表 1 : 使用する株式データ

	2013 年	2014 年
日経 225	955,215,608 行	1,137,155,162 行
日経平均先物	2,899,243 行	2,538,704 行

2-2. ニュースデータ

本研究ではカテゴリーの異なるニュースとしてマクロニュース及び銘柄ニュースの 2 種類を用いる。分析対象は日本株式市場の開場時間内に英語で発信されたヘッドラインニュースで、ニュース配信 2.5 分前に取引が行われていたニュースに限る。日経 225 に属する企業のコード番号が付随したニュースを銘柄ニュース、日経 225 のコード番号が入ったニュースをマクロニュースとした。(表 2)

表 2 : 使用するニュースデータ

	2013 年	2014 年
ニュース数	2155	1201

3 分析方法及び分析結果

本研究では LSTM (Long Short Term Memory) [8] を用いて分析を行った。以下の定義式よりニュース配信前後の株式変動率を求め、教師データとする。クラス分類に用いる閾値は今回 0.1% とした。

$$\text{株式変動率 (\%)} = \frac{(\text{ニュース配信 2.5 分後の価格}) - (\text{ニュース配信 2.5 分前の価格})}{(\text{ニュース配信 2.5 分前の価格})} \times 100$$

ネガティブ : $-0.1\% > \alpha$
 ニュートラル : $-0.1\% \leq \alpha \leq 0.1\%$
 ポジティブ : $0.1\% < \alpha$

以上の分類を教師データとして、2013 年のデータでニュースと株式変動率の分類学習を行い、2014 年のデータで検証を行った。分析には銘柄ニュース、マクロニュースをそれぞれ単体で用いた場合と両方用いて分析した場合の 3 通りで行った。分析の結果、マクロニュース及び銘柄ニュースの両方を用いた場合が最も高い正答率となった。(表 3)

表 3 : 分類分析の正答率

	学習	検証
銘柄ニュース	76.05%	56.76%
マクロニュース	62.53%	59.53%
銘柄及びマクロ ニュース	76.39%	72.22%

4 まとめ

本論文では、日経 225 に属する企業を対象にニュースと株式変動の関連性について LSTM を用いて分析を行った。ニュースデータには、カテゴリーの異なるニュースとしてマクロニュース及び銘柄ニュースを使用し、株式変動率を教師データとしてニュース分類を行った。

分類分析では銘柄ニュースのみならず、マクロニュースを加えることで分類精度の向上が確認された。

本研究を通じ、高頻度データを教師データとして、株式短期変動の分類モデル構築ができる可能性があること、またニュースを特徴ごとに分類し、分析する意義が確認できた。分析期間の拡張および企業属性を考慮した分析など詳細な分析は今後の課題である。

参考文献

- [1] G. Gidófalvi and C. Elkan, "Using news articles to predict stock price movements," Department of Computer Science and Engineering, University of California. (2001)
- [2] Tetlock, P. C. "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance* 62 (3), pp. 1139-1168. (2007)
- [3] Tetlock, P. C., M. Saar - Tsechansky and S. Macskassy "More than words: Quantifying language to measure firms' fundamentals," *The Journal of Finance* 63 (3), pp. 1437-1467. (2008)
- [4] W.S.Chan, "Stock price reaction to news and no-news : drift and reversal after headlines," *Journal of Financial Economics*, vol.70, no.2, pp.233-260. (2003)
- [5] Brogaard, J., T. Hendershott and R. Riordan "High-frequency trading and price discovery," *The Review of Financial Studies* 27 (8), pp. 2267-2306. (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』 (2016)

[7] R.P. Schumaker and H.Chen, "Textual Analysis
of Stock Market Prediction Using Financial News
Articles," in Proceedings of the 12th Americas
Conference on Information Systems (AMCIS) ,
pp.1-20. (2006)

[8] S. Hochreiter, J. Schmidhuber "Long Short-Term
Memory," Neural Computation 9(8) pp.1735-1780
(1997)