

インフルエンサーが発信するツイート情報はニュース記事 に対して優位性があるのか？

Relationship between tweet information and news articles

片山慎也¹ 西良浩¹ 菅愛子¹ 高橋大志¹

Shinya Katayama¹, Yoshihiro Nishi¹, 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: SNS が世の中に与える影響が拡大しており、特にインフルエンサーが人々に与える影響が増してきている。本研究では、株のインフルエンサーを特定した上で、Google Cloud Natural Language API を用いた感情分析を行い、ツイートとニュース及び株価の関係を明らかにすることを試みる。

1. はじめに

インターネットが発達し SNS が発展する以前、情報の発信源として重要な役割を果たしていたのは、新聞やテレビに代表されるマスメディアであった。

しかし今では、レストランを探すときに食べログの星の数を参考にしたり[1]、ホテルを予約する際にトリップアドバイザーを参考にしたりする人が多い。口コミを参考にする際には、誰のどの書き込みでもいいというわけではなく、評価の高いインフルエンサーの書き込みを参考にすることが推定される。

モバイルで簡単に投資ができる環境が整い[2]、金融市場においても、SNS の役割が強くなっていることが推定される。Twitter のテキストデータから、ダウ平均の予測をする研究があるが[3]、本研究では従来の研究では限定的であったインフルエンサーのツイート情報とニュースのテキスト情報との関連性について分析を行う。

2. 関連研究

Zhang ら[4]では、twitter の特定の感情が高まる時株価（ダウ）は次の日に下がり、反対に人々が特定の感情を減らすとき、株価（ダウ）は上がるとしている。したがって、あらゆる種類の感情的な爆発を Twitter でチェックするだけで、翌日の株価の変動を予測することができるとしている。

五島・高橋・寺野 2015[5]では、機械学習によるニュース記事の評価を通して、将来の株価予測ができ

る可能性を見出している。また、ニュースのテキスト情報の極性（ポジティブ・ネガティブ）の推測について、SVR と同様に、Deep Learning が有効であることを示した。

佐藤・大原・豊田[6]では 日経平均株価の騰落予測を対象とした評価実験において無作為に抽出したユーザーのツイートを用いるよりも、経済関連ニュースに対する興味度が高いと推定したユーザーのツイートを用いる方が、正答率が向上することを確認したとしている。佐藤らは、投稿されたツイートに対する感情極性値を利用した株価騰落予測モデルを構築する際に、対象ツイートを無作為に選定するのではなく、経済関連ニュースに興味をもっていると推測されるユーザーが投稿したものに限定している。ユーザーの経済関連ニュースに対する興味度は、経済関連ニュースから事前に特徴語を抽出し、その特徴語を含むツイートの投稿頻度により定義している。

3. データ

インフルエンサーの特定に際して、ツイート（つぶやき）情報に関しては Twitter が提供する API を活用し、計 242,067 件のツイート情報を取得した。Twitter とニュースの比較において、ツイートに関しては、特定したインフルエンサー上位 10 人の中で、API 制限にかかわらず、情報が取得できる、全ツイートを分析対象とした。

ニュースに関しては、ロイターニュース 11 月分の全件を分析対象とした。ロイターニュースは、トムソンロイター社が提供するニュースであり、日本証

券市場に関する日本語のニュース記事のみを分析対象とする。データ数は14,164件である。マーケットデータに関しては株式価格ティックデータを使用する。

4. 分析手法

本研究では、2018年11月のツイート、ニュース及び株価を用いて分析を行った。まず株のインフルエンサーの特定を行った上で[7]、インフルエンサーランキングの上位10人の2018年11月のツイートを抽出した。ニュースに関しては、ロイターニュース11月分の全件を取得した。

上記のデータについて Google Cloud Natural Language API を用いた分析を行った。Google Cloud Natural Language API とは、事前トレーニングされた何千もの分類を使用し、テキストの構造と意味をすばやく明らかにできる API である。本研究では API を用いて感情分析を行う。

感情分析では、指定されたテキストを調べて、そのテキストの背景にある感情的な考え方を分析する。具体的には、投稿者の考え方がポジティブか、ネガティブか、ニュートラルかを判断する。

Google Cloud Natural Language API の感情分析では、ドキュメント内のポジティブな感情とネガティブな感情の違いを示す。ドキュメントの感情分析の score は、ドキュメントの全体的な感情を示す。Score は -1.0 (ネガティブ) ~ 1.0 (ポジティブ) のスコアで感情が表される。ドキュメントの感情分析の magnitude は、そのドキュメントに感情的な内容がどのくらい含まれているかを示す。指定したテキストの全体的な感情の強度 (ポジティブとネガティブの両方) が $0.0 \sim +\infty$ の値で示される。score と違って magnitude は正規化されていないため、テキスト内で感情が表現されるたびにテキストの magnitude の値が増加する[8]。

総合magnitude: 1.1

総合score: -0.2

好決算の影響で株価が大幅上昇しそう。

magnitude: 0.2, score: 0.2

米国の雇用統計の悪化により明日の相場は荒れそう。

magnitude: 0.8, score: -0.8

図1: Google Cloud Natural Language API を用いたスコア算出の例

Google Cloud Natural Language API の感情分析機能を Python で呼び出し、出力したものが図1である。入力したテキスト全体の総合 magnitude と

score、一文ごとの magnitude と score が算出されるが、本研究では一文ごとの score を用いる。その日に発信された全ツイート、全ニュース記事に対して各文章の score を足し合わせた平均をそれぞれ算出し、その日のツイート指数、ニュース指数とする。

Google Cloud Natural Language API を用いた研究はいくつかある。江島、熊谷、村上 (2019) では、Twitter の投稿履歴から Google Cloud Natural Language API の感情分析を用いて、ユーザーの過去の感情の数値化を行うことで自己理解を支援するシステムを開発したとしている[9]。

田原、池田、松本、帆足 (2018) では、ユーザーの発話と SNS 上の投稿を用いたシステムの応答の内容の極性 (Positive・Negative・Neutral) を考慮したシステムを提案し、既存手法と比較をして、より共感度の会話を実現できることを確認したとしている[10]。

5. 分析結果

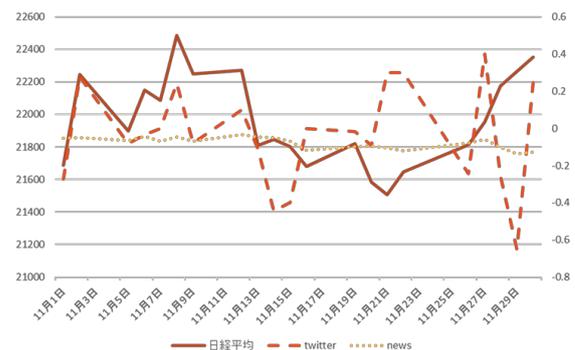


図2: Google Cloud Natural Language API によるスコアと日経平均の時系列チャート

図2のチャートを見ると、11月前半はTwitterと日経平均が近い動きをしていることが分かるが全体としては、異なった動きをしていることが見て取れる。またTwitterとニュースの動きを見ると、Twitterの方が大きな動きをしており、ニュースに関しては動きが少ない。これは当然ではあるが、Twitterの方が感情的な文章が多く、ニュースは感情を多く含まない客観的な情報が多いことが要因であるといえる。

日経平均株価と Google Cloud Natural Language API による Twitter とニュースのスコアの相関は、それぞれ 0.070, 0.193 であった。11月前半と後半に分け短期間で相関を見ると、11月前半の日経平均と

Twitter の相関係数が 0.807 と強い相関が得られたが、全体として有意な結果は得られなかった。

5. まとめ

本研究では、株のインフルエンサーを特定した上で Google Cloud Natural Language API を用いた感情分析にて、Twitter とニュースのスコアを算出し株価との関係性について分析を行った。

今回算出した指数では、ツイッター、ニュース及び株価の明確な関係は見られず、同じ動きをしている訳ではないことが分かった。

今後の課題として、日経平均と Twitter、ニュースのスコアに相関は見られなかったが、個別銘柄に絞って感情分析を行うことで新たな示唆が得られる可能性があると考えている。またスコア算出にディープラーニングを用いる等の方法を試みる。

参考文献

- [1] 広田 すみれ,高橋 聖奈:レストランクチコミサイトにおける評価の数や質と意思決定の関係, 東京都市大学横浜キャンパス情報メディアジャーナル, Vol.15, PP32-36, (2014)
- [2] 淵田 康之: フィンテックの意義と日本の課題, 資本市場, No.367, pp. 4-13, (2016)
- [3] Johan Bollen, Huina Mao, Xiao Jun Zeng: Twitter mood predicts the stock market, Journal of Computational Science, Vol.2, No.1, pp.1-8, (2011)
- [4] Zhang, X., Fuehres, H., and Gloor, P : Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”, Procedia-Social and Behavioral Sciences, Vol.26, pp.55-62 ,(2011)
- [5] 五島 圭一、高橋 大志、寺野 隆雄: ニュースのテキスト情報から株価を予測する,人工知能学会全国大会論文集,Vol.29, pp.1-3, (2015)
- [6] 佐藤 大吾, 大原 剛三, 豊田 哲也: 経済情報に関心をもつ SNS ユーザーの投稿内容に基づく株価騰落予測モデルの構築 , SIG-KBS ,Vol. B5, No.01, pp. 1-5, (2016)
- [7] 片山 慎也,菅 愛子,高橋大志: Twitter における株式のインフルエンサーの特定, 人工知能学会 第 13 回 経営課題に AI を! ビジネス・インフォマティクス研究会, (2019)

- [8] Google Cloud ホームページ「Natural Language API の基本」[<https://cloud.google.com/natural-language/docs/basics?hl=ja#sentiment-analysis-values>] (最終検索日: 2020/1/1)
- [9] 江島 直也・熊谷 直人・村上 晴美: Twitter を用いた自己理解支援システムの開発,第 81 回全国大会講演論文集,2019,Vol.1,pp397-398, (2019)
- [10] 田原 俊一・池田 和史・松本 一則・帆足 啓一郎: ウェブニュース関連ツイートの極性に着目した共感対話システムの提案と評価,第 80 回全国大会講演論文集,2018,Vol1, pp29-30,(2018)