

SNSのデータを用いた株価変動と新型コロナ関連トピックの 関係性の分析についての研究

Research on relationship analysis of Stock price fluctuations using SNS data and topics related to the covid-19

小駒佑輝[†] 齋藤泰寛[†] 佐竹夏輝[†] 大石海斗[†]
迫田巧大[†] 西田龍史[†] 池辺正典[†]
Yuki Kogoma[†] Yasuhiro Saito[†] Natsuki Satake[†] Kaito Oishi[†]
Takuto Sakota[†] Tatsufumi Nishida[†] Masanori Ikebe[†]

[†] 文教大学 情報学部

[†] Faculty of Information and Communications, Bunkyo University.

要旨

将来の株価を予測する研究は古くから行われており、様々な方法が提案されている。従来の方式では過去の株価のような定量的なデータから予測する方法が多かったが、本研究ではそれに加えて Twitter のデータを参考にすることでネガポジ判定等による数値化した情報を利用する。そして、Twitter 上の話題などを反映した株価予測の手法を検討する。

1. はじめに

1.1. 背景

現在、新型コロナが世界的に流行により経済に大きな影響が出ている。ニュースや新聞でも連日新型コロナの影響で景気が悪化しているといった報道がされている。しかし、実際にはどの程度の影響が出ているかといったことは明確ではない。そこで、SNS からのデータを利用して景気の動向について予測することで、コロナ禍の経済における株価の変動という視点から研究を進める。

1.2. 株価について

本研究では東証1部の上場企業である楽天、KDDI、ソフトバンク、ドコモを対象とする。これらの4社を対象とした理由は、ある程度の取引規模があり、Twitter 等でも情報の収集が可能であること、主要キャリア4社による比較が可能であることから今回の研究対象とした。これらの企業の株価について、社会情勢などの情報を Twitter から収集し、株価との相関を分析する。

1.3. ツイッターについて

Twitter とはユーザーがツイートという短文のテキスト投稿をできる SNS である。ツイートには文章だけではなく画像や動画なども併せて投稿することができる。また、キーワードでユーザーやツイートを検索することも可能である。そして、株価についてのツイートも多く、「株価」で検索することで、株価に関するニュースや株価の変動を予測するようなツイートも投稿されている。

2. 研究の内容

2.1. 既存研究

Twitter からの収集情報を分析した研究[1]やそれを株価に応用した研究[2]の他に、株価の情報を可視化するための研究[3]も行われている。また、SNS からの感情分析によって傾向を見出す研究[4][5]も行われていることから、本研究では株価の変動を社会情勢やそれを受けた感情分析により傾向を見出す。

2.2. 研究の目的

本研究の目的としては、SNS（主に Twitter）から得られる情報から株価を予測し、新型コロナ感染症関連のトピックとの繋がりから、景気分析を行う。分析結果は時系列等を考慮した形で可視化することで分析のための可視化も研究する。元となるデータは Twitter を対象としており、収集したデータの可視化等によって株価と景気、コロナ関連の情報の把握を助けることを目的と。

2.3. 研究の概要

2.4. 処理方法

Twitter API を用いて収集したツイートの感情分析を行う。分析対象は楽天, KDDI, ソフトバンク, NTT ドコモの4社とし, ツイートと株価はそれぞれ TwitterAPI, Yahoo Finance API を利用して取得する。感情分析には, janome の形態素解析と日本語評価極性辞書を使用する。この辞書は評価極性を持つ名詞, 約 8,500 表現に対して評価極性情報を付与した辞書である。感情分析の手順については取得したツイートを janome で単語を形態素に分割して単語の「基本形」「読み」「品詞」を取得する。取得単語を極性辞書にかけて, ネガポジの極性判定を行う。ツイートはテキストデータであるため, 日本語評価極性辞書を使用して文章がネガティブかポジティブかを判定することで数値化する。設定したキーワードを含み文章の内容がポジティブ, またはネガティブなツイート数と変動した株価の相関係数を算出する。

2.5. ツイートの収集処理

ツイートの収集方法については Twitter API を用いて行った。この API を用いることで, Twitter に投稿されているツイートの情報を取得することができる。今回は Twitter の検索機能にあたる Search API を利用し, これにより任意のキーワードを含むツイートを検索し, Python で作成したプログラムで取得する。今回用いた検索ワードは「楽天」, 「Softbank」, 「KDDI」, 「NTT ドコモ」の4種類である。キーワードを用いた収集方法では過去一週間のツイートしか収集できないため, キーワードごとに1週間で1万件程度のツイートを収集した。

2.6. 可視化処理

前述したツイートの感情分析データと Yahoo Finance API で取得した過去の株価の変動データを使用する。また, 可視化には, D3.js を使用した。D3.js とは可視化ツールとしてウェブブラウザ上で動的コンテンツを描画する JavaScript ライブラリである。可視化の手順は, ツイートの感情分析データを1日ごとにネガティブ単語とポジティブ単語で集計を行い個別に縦軸は単語数, 横軸は日付の折れ線グラフの形にして出力を行う。Yahoo Finance API で取得した過去の株価の変動データのうち, 終値のデータを使用して縦軸は価格, 横軸は日付の折れ線グラフの形にして出力を行う。株価の予測には終値とツイートのネガティブ単語とポジティブ単語それぞれの数との相関係数を求め, その情報を可視化する。

3. まとめ

今回の研究では, 新型コロナによって不安定な経済の中で, Twitter から収集したツイートデータと Yahoo Finance API で取得した株価のデータを用いて株価の変動を予測し, それを可視化した。分析内容を見ると, 個々のデータは個人の発言であるが, SNS 全体の傾向として見た場合には傾向を示す情報が含まれていることもあった。既存事例ではツイートの感情から株価を予想するものなどもあり, 一つの些細な発言でも, 視点を変えることで大きな情報を入手できるものとなる。今回は株価という視点で見えていたが, 流行や政治など, 様々なコンテンツについて調査するのにも SNS のデータは活用できると考える。

参考文献

- [1] 鳥海不二夫, Twitter 上のビッグデータ収集と分析, 組織科学, 組織学会, Vol48, No.4, pp.47-59, 2015.
- [2] 長尾将宏, 長尾智晴, Twitter を用いた株式市場の変動予測, 第76回全国大会講演論文集, 情報処理学会, Vol.2014, No.1, pp.377-378, 2014.
- [3] 石川大輝, 斉藤 隆文, 多数の株価の比較のための可視化手法, 第77回全国大会講演論文集, 情報処理学会, Vol.2015, No.1, pp.101-102, 2015.
- [4] 桃井達明, 須鎗弘樹: Twitter から生成した感情モデルと社会経済的現象との相関(数理モデル(2),A 分野:モデル・アルゴリズム・プログラミング), 情報科学技術フォーラム講演論文集, 電子情報通信学会・情報処理学会, Vol.11, No.1, pp.127-130, 2012.
- [5] 三井健史, 伊藤智也, 中西勇人, 濱川礼: SNS の投稿を用いた感情記録ライフログシステム~Emote~, 研究報告エンタテインメントコンピューティング (EC), Vol.2014-EC-32, No.1, pp.1-6, 2014.