

## 社会的要因を用いた株価予測

1021077 松本葉太  
指導教員：新美礼彦

### Predicting Stock Prices Based on Social Factors

Yota Matsumoto

**概要：** 近年、国の政策や金利減少の影響などを受けて株取引が注目を集めている。特に現役層といわれる 20 40 歳の株主が増えている。そこで本研究では、従来より会社の言動が以前よりも株価に影響するのかを調べ、それをもとに従来の株価予測とは違ったアプローチを取り入れることで予測精度の向上を目指す。例えば、何らかの不祥事を起こし炎上した企業に対する SNS のポストを分析することで、どの程度株価に影響を与えるか予測する。他にも、新商品に対する SNS での事前評判で株価に影響があるかも調べる予定である。それぞれ別の業界や規模の会社を対象とすることで、それぞれの特徴があるのかも調べる。今回は X(旧 Twitter) のポストから SNS のポストデータを収集し、極性感情分析を行う。

**キーワード：** 株, SNS, 極性感情分析

**Abstract:** In recent years, stock trading has been attracting a lot of attention due to government policies and declining interest rates. In particular, the number of shareholders aged 20 40, considered the working-age group, has been increasing. Therefore, this study aims to improve the accuracy of forecasts by examining whether a company's words and actions affect stock prices more than before and incorporating a different approach from conventional stock price forecasts based on the results. For example, by analyzing the reaction of social networking services to a company that has been under fire for some scandal, we can predict the extent to which the company's actions will affect its stock price. We also plan to examine the impact on stock prices of the advance publicity of new products on social networking services. By targeting companies in different industries and of different sizes, we will also examine the characteristics of each company. In this study, we will collect SNS reaction data from X (formerly Twitter) posts and conduct a polar sentiment analysis.

**Keywords:** Stocks, Social networking, Emotional analysis

## 1 背景と目的

株価予測するために使われているデータは、企業の財務状況がまとまっている貸借対照表や損益計算書、キャッシュフロー計算書などの数値を使うことが多い。しかし、日本証券協会 [1] によると、ここ数年で個人株主が増え、保有銘柄数も増えている。また一人当たりの平均株式保有数は減少し、20 歳から 40 歳の現役層と呼ばれる人々が株式取引に参加し、株式の分散化が進んでいる。実際に株価を下げ発行枚数を増やす、株式分割をすることで一般人を株取引を推進し、投資家層の拡大を図る企業もある。このような状況下では、従来株価に影響しないと考えられていた要素が、より大きく株価に影響を及ぼす可能性がある。現役層が株式を持つことによって、より多くの人が会社の言動に注目しているだろう。そこで、本研究では SNS のポストを調べることで、株

価の予測を補助し、予測精度を上げることや新しい株価に影響を与える要素の発見を目的とする。

## 2 複雑系コースにおける本研究の位置づけ

複雑系コースでは現実世界で起きている事象に対して数理的にモデル化し分析・適応することや法則性を解明するなどの課題に取り組むことが求められている [2]。本研究では、人間社会で実際に動いているデータである株価と SNS に焦点を当て、SNS から人々の感情を分析し、株価の予測モデルの改善や今まで考えられていなかった株価に影響を与える要素や法則性の発見を目的としている為、学部カリキュラム・ポリシー [2] に沿っている。

## 3 関連研究

従来の予測モデルに SNS データを使い、予測モデルを向上した先行研究の例としては [3] があ

る. この研究では SNS データから特徴量を収集し, 以前から使われていた POS データと組み合わせることで, 以前の POS データだけを使用していた時よりも予測精度を上げることが成功した.

機械学習による Facebook に基づく株価予測 [4] では, 調査対象企業のアカウントに対する「いいね」などのアクション数を参照し, 話題指数としていた. それに機械学習を用いることで翌日の株価の増減を予測をしていた.

ネット炎上が株式市場に与える影響についての研究 [5] では, 炎上した時の企業の対応によって, 株価にどのように影響を与えるのかが調査されている. [5] によると, SNS での炎上が株価下落の原因になる要素として, 企業の対応の有無や報道の有無, 個人株主比率, 炎上の内容, 企業の規模や元々の注目度, 存続年数などがある. 炎上が始まる以前から当期純利益がマイナスであったり, 売り上げ率が低いなどの企業にとってマイナスな要素があると炎上を触発しやすい. また, 業界ごとに炎上の起きやすさが違い [5] の調査結果では情報通信産業が最も炎上を起こしやすいとされている.

## 4 提案する理論

SNS を用いた株価予測について説明する. 今回は X(旧 Twitter) のポストからデータを集める. 1 年分を集める予定であり, 半年を実験用にしよう半分を検証用にする. 対象企業は業界や規模が別々になるように集め, それぞれの特徴や法則性がないか調べる. データには投稿日時, テキスト (ポスト内容), 閲覧数, リプライ数, RT 数, 引用数, アカウント名フォロワー数が含まれている. 収集データのテキストを極性感情分析し株価予測に使う. 極性感情分析には, 事前学習済感情分析モデルである bert-japanese-finetuned-sentiment とトークナイザー事前学習済モデル bert-base-japanese-whole-word-masking を用いる. トークナイザーとは日本語を分析する際に, 文章を意味ごとに分解するもののことである. このトークナイザーは形態素分析という手法で文章を分解している. これらのモデルを使い, ポスト内容をネガティブ, ポジティブ, ニュートラルの三つの感情に分ける. 1 日ごとに極性感情分析の結果を集計, スコアする. また, [4] のようにいいね数, RT 数, リプライ数, 閲覧数などを使い同

じ一つのポストでも, 話題指数による重さ付けをする. これらの情報をもとに株価予測モデルを作る. 株価予測モデルを作る際は [3] のように従来のモデルに補助する形で今回のデータを使い, 精度の向上を目指す. 予測モデルを作る際は実際の出来事が株価に影響を与えるのに時間差があることを考慮する.

## 5 実験と評価

企業はカドカワ (9468) を対象に X(旧 Twitter-twitter) にて検索ワードで「kadokawa」で検索し, 45785 件のポスト情報を集めた. 期間は 2024 年 4 月 23 日 2024 年 10 月 25 日であり, 返信やリンクがついているポストは広告や告知である可能性が多いため除外した. 集めたデータを事前学習モデル bert-japanese-finetuned-sentiment とトークナイザー事前学習済モデル bert-base-japanese-whole-word-masking で極性感情分析し, POSITIVE, NEGATIVE, NEUTRAL の 3 つの感情に分類した. POSITIVE, NEGATIVE, NEUTRAL をそれぞれ 1, -1, 0 とした. それらを 1 日ごとに和をとって集計し, 株価と図にした (図 1). 青のグラ

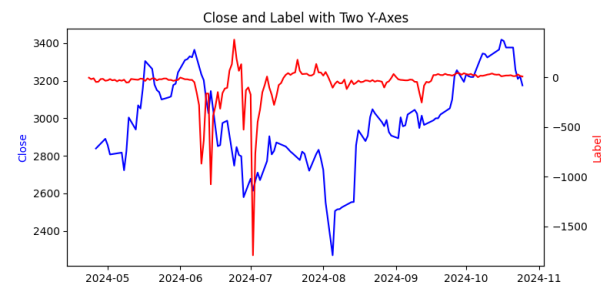


図 1 X のポストと株価

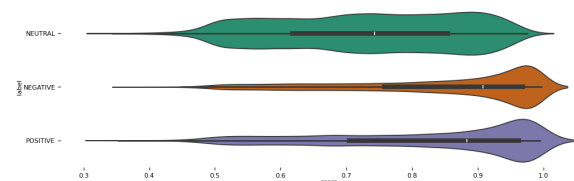


図 2 X のポストと株価

フがカドカワの株価, 赤のグラフが X でのポストである. また, 2 つの株価はその日の最後の株価である終値を採用した. 感情の分類の信頼度は, 学習モデルに備え付けられていたものを使用. 2. 数値が高いほど信頼度が高いことを示す. POSITIVE, NEUTRAL は, 1.0 の方にグラフが偏って

おり問題はない。NEGATIVE だけが 0.5 あたりの信頼度のものが多く、本来 POSITIVE に分類されるべきものが NEGATIVE に分類されていないか、少し不安が残るものとなった。実際は NEGATIVE の判定が特に多かった 6 月 14 日と 7 月 2 日周辺はネットニュースなどになるほど話題になっており、明らかところは外していないことが分かった。

## 6 考察

図 1 より、実験を踏まえてわかったことは、炎上が起こると NEGATIVE の判定のポストが多くなり、株価下落に大きな影響を与えることだ。6 月 12 日の炎上から少し遅れて株価が下落している。これは想定通りの動きとなった。7 月 2 日の炎上では、株価は炎上が起きるより先に株価が暴落している。このことから、この炎上は株価に関係ない、もしくは暴落していることについてのポストと予想できる。その後 1 か月後に暴落があるがあまりにも期間が開いている為関連性が薄い。そこで、4 を見ると日経平均株価の影響だとわかる。この図は見やすくするために平均 0 分散 1 の標準化を施した。

一方、POSITIVE のポストが多いとき株価が上がるかはわからない。図 3 は、1 日ごとの POSITIVE のポスト数の合計である。

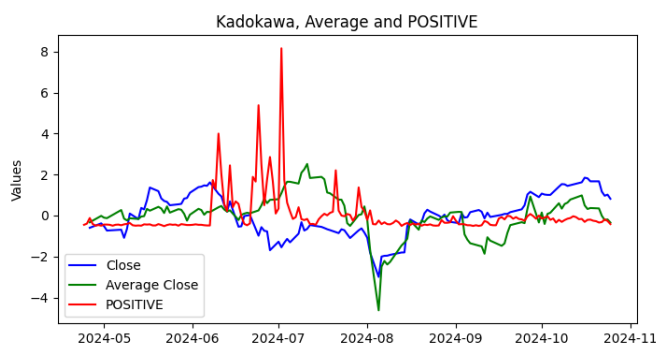


図 3 POSITIVE ポスト数

NEGATIVE のポスト数が多い日であっても、少なからず POSITIVE のポストがあるが、今回は NEGATIVE の方が影響度も数も大きく、株価暴落に至ったことがわかる。しかし、いいね数などの指数をつかえばポスト数が少なくとも影響度をより詳しく調べられる可能性がある。もう一度、図 4 を見ると、6 月 14 日の炎上の後、POSITIVE のポストが多くなっているが、日経平均株価も上がっている為、どちらの影響でカドカワの株価が

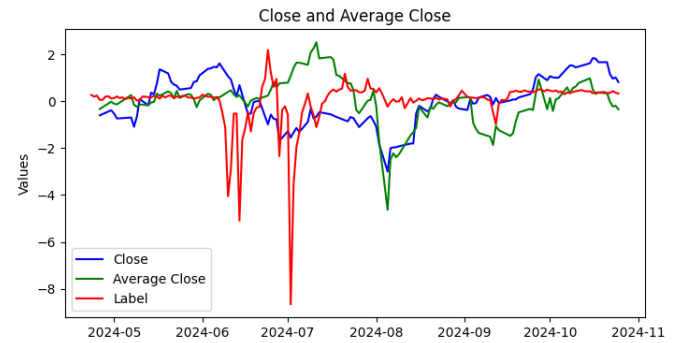


図 4 カドカワの株価と日経平均株価と X のポスト

上がっているか不明である。その他 8 月入ってすぐのところでは POSITIVE のポストが多いにもかかわらず日経平均株価に影響されカドカワの株価は下落しているところや、NEUTRAL のポストが多い 9 月以降も基本的に日経平均株価に沿って変動しているように見える。このことから日経平均株価の影響を踏まえたうえで、X のポストがどの程度株価に影響しているか見定め株価予測モデルを構築しなくてはならない。また、企業によって日経平均株価の影響を受ける度合いが変わっている可能性や為替などの他の考えうる影響が大きい要素にも注意して他の企業についても調査していく必要がある。

## 7 結言

実験を通して、SNS のポストから株価の予測ができるのではないかとという仮説の一部を検証することができた。今後の展望 1 つ目はいいね数やれプライ数などからそのポストの影響度を計り、重み付けすることで精度の向上を目指す。またこれにより、今回のようにポスト数では負けている POSITIVE のポストの影響度を調べられるようになる可能性がある。2 つ目はより多くの企業、業界を分析しそれぞれの特徴を調査することや今回わからなかった SNS の POSITIVE なポストの株価に対する影響度を調べることである。3 つ目は実験から得たデータを使い、株価予測モデルの精度向上をすることである。

## 参考文献

- [1] 日本証券業協会, 「個人株主の動向について」, 9, 21, 2022.
- [2] 公立はこだて未来大学, 学部カリキュラム・ポリシー, 31, 10, 2024.

- [3] 西口 真央, 鳥海 不二夫, 河尻 耕太郎, 吉田 光男, 小売店 POS データおよび SNS データを用いた缶ビールの需要予測, 2024 年度人工知能学会全国大会 4 ページ, 2024.
- [4] 及川 健一郎, 堀田 政二, 機械学習による Facebook に基づく株価予測 (若葉研究者の集い 3, サマーセミナー 2014～未来を拓くビジョン技術～), 映像情報メディア学会技術報告, 39-40, 2014.
- [5] 武田 史子, 森 継哉, ネット炎上が株式市場に与える影響についての研究, 30, 7, 2020.