

## テキストデータを用いた株トレード支援手法の提案 A Study on a Stock Trading Support Using Technical Analysis and Text Data Analysis

伊藤 翔哉<sup>†</sup>  
Shouya Itou

樽松 理樹<sup>†</sup>  
Masaki Kurematu

### 1. はじめに

近年、ネット証券の発達により投資がしやすくなっていることや老後の資金確保などの理由から、株式投資を始める人が増えてきている。また、2022 年度から高校で金融教育が開始され、さらに株式投資を始める人が増える可能性が高い。株式投資は、収益性が高い反面、安全性が低いことから、株式投資の初心者に対し、適切な運用を支援する方法が必要となる。その 1 つとして株価予測が挙げられる。株価予測手段として、テクニカル分析とファンダメンタル分析があげられる[1]。しかし、テクニカル分析では急な変化に対応できないことや、ファンダメンタル分析では初心者には分析が難しいといった問題がある。

近年、新しいアプローチ手法として、テキストマイニングや機械学習を使う研究が増えてきている[2]。これらは、市場に対して影響を持つ人物の発言や企業の動向、事件、事故などが株価に影響を与えること、また、それらの情報を Web ニュースなどの Web から得ることが多いという考えに基づいている。このような研究の例として、J.Bollen らは、Twitter の感情解析結果から Web 上のデータの感情値と株価変動との間に相関関係があることを報告している[3]。また 2.1 節に示す Shangkun DENG らの研究では、Web ニュースの分析結果とテクニカル分析を合わせた株価予測の研究によりベースより良い結果を得ている[4]。これらの研究は英語を用いているが、増井らは、日本語も同様の傾向があることが示しており[5]、佐藤らは Twitter に加え Web ニュースを用いた研究に取り組んでいる[6]。このように現在、株価予測手段として、テキストマイニングや機械学習が用いられたものが増えてきている。

しかし、前述のような先行研究では、平均株価の予測が多い。実際の投資対象は企業単位であり、その点から企業単位の株価予測の方が実用的である。分析対象のテキストとしては SNS、Web ニュース、市場レポート、開示書類等様々あるが、テキストの更新頻度や初心者が手に入れやすいという点を考えれば、先行研究のような Twitter や Web ニュースを対象とするのが妥当である。また、初心者向けのシステムを想定する場合、予測結果を提示するだけでなく、その先として株の売買の判断についても提示することが求められる。

以上の背景から本研究では、企業の株価を対象とし、テキストマイニングとテクニカル分析を用いた株価予測及び予測結果に基づいた売買判断を行う手法を提案する。以降、2 章において、先行研究について述べる。3 章で本提案手法の概要を述べ、4 章で評価結果と考察について述べる。最後の 5 章で本論文をまとめる。

### 2. 先行研究

#### 2.1 Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction[4]

Shangkun DENG らは、Raw Data Preprocessing コンポーネント、Sentiment Analysis コンポーネント、Feature Extraction コンポーネント、Prediction & Evaluation コンポーネントからなる株価予測モデルを提案している。これらのコンポーネントを用いた処理の流れを以下に示す。なお、各コンポーネントは、RDP コンポーネント、SA コンポーネント、FE コンポーネント、P&E コンポーネントと表記する。

1. RDP コンポーネントが株価の時系列データ、ニュース、およびコメントのデータをダウンロードする。
2. SA コンポーネントは、すべてのコメントとニュースに SetiWordNet を使用したセンチメント分析を適用する。各ニュースとコメントの全体的な感情 (Objective, Subjective, Negative, Positive) を分析する。そしてターゲット企業の positive, negative, objective なニュースとコメントの数をカウントする。
3. FE コンポーネントは 3 種類のデータソースから特徴を抽出する。株価の時系列データに関してはテクニカル分析を用いて特徴を、ニュースとコメントはその頻度と SA コンポーネントで得られた感情から特徴を抽出する。
4. P&E コンポーネントは、MKL (Multiple Kernel Learning) 回帰フレームワークを使い抽出した特徴から次の取引日の株価と ROC を予測する。評価部分では、MAE、MAPE、RMSE を評価尺度として使用して、提案モデルとベースラインモデルの良さを評価する。

Shangkun DENG らは、米国株式市場における日本有名 3 社の株式を使用して検証している。この際、テクニカル分析には MACD、ROC、BIAS を使用している。これらの手法については、3.2.3 節にて説明する。Shangkun DENG らによる実験の結果を表 1 に示す。表 1 において、SVM-A は、テクニカル分析、頻度、センチメント分析の結果に SVM を用いたモデル、SVM-ta はテクニカル分析の結果のみに SVM を用いたモデル、SVM-nd は頻度の結果のみに SVM を用いたモデル、SVM-sa は、センチメント分析のみに SVM を用いたモデル、MKL-A は、テクニカル分析、頻度、センチメント分析の結果に MKL を用いたモデルである。また RMSE、MAE、MAPE は、回帰モデルの評価基準である平均二乗誤差、平均絶対誤差、平均絶対パーセント誤差をそれぞれ示す。

表 1: ある社の MAE、MAPE、RMSE の結果 [4]

	SVM-A	SVM-ta	SVM-nd	SVM-sa	MKL-A
MAE	0.8803	1.0114	1.0296	1.0032	0.2368
MAPE	4.7335	5.4379	5.5167	5.3792	1.3501
RMSE	0.9002	1.0830	1.1037	1.0723	0.3025

<sup>†</sup> 岩手県立大学 Iwate Prefectural University

評価結果から、提案手法である MKL-A がベースライン手法よりも優れていることが示され、株価からの分析以外の特徴がパフォーマンスを改善したことを示している。

### 3. 提案手法

#### 3.1 手法概要

本手法の概要を図 1 に示す。本手法は、初めに、データ収集フェーズにおいて、対象企業に関する Web ニュースの記事とツイート及び株価情報を収集する。次に分析フェーズにおいて、Web ニュースとツイートに対しては、株式市場休業日に対する処理を行った後、評判分析を適用する。また株価情報に対しては、テクニカル分析を行う。次に、学習フェーズにおいて、評判分析の結果とテクニカル分析の結果、対象企業の株価に機械学習を適用し、予測モデルを構築する。次に予測フェーズにおいて、生成した予測モデルを用いて、新たに収集したデータから株価の予測を行う。最後に予測結果をもとに株売買の判断を行う。

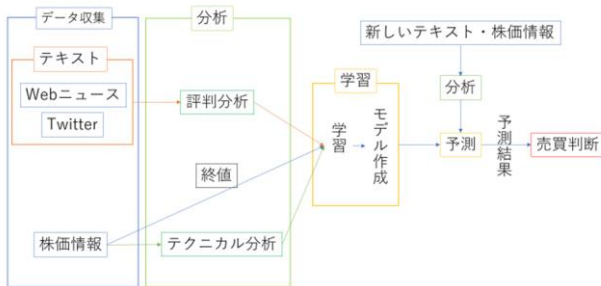


図 1 手法概要

以下、各フェーズについて説明する。

#### 3.2 各フェーズの処理

##### 3.2.1 データ収集

Web ニュースは Python で Web スクレイピングのシステムを構築し、自動的に記事を入手する。Twitter は Twitter API を用いて自動的にツイートを入手する。株価情報は手動で入手する。

また株式市場は、土曜日、日曜日、祝日、祭日（以後、土日祝日と略記）は開かれておらず、株価が変動しない。そのため、土日祝日は予測の対象外とするが、土日祝日にもテキストデータは収集されるため、土曜日・日曜日のデータは金曜日、祝日、祭日のデータはそれの前日のデータとして扱う。例えば、Web ニュースの記事が金曜日に 4 件、土曜日に 5 件、日曜日に 6 件の場合、まとめて金曜日に 15 件とする。ツイートも同様である。

##### 3.2.2 評判分析

評判分析は、テキストを数値化し、ネガティブやポジティブの感情を持っているのかを数値として算出する[7]。以後ポジティブを pos、ネガティブを neg と表記する。分析には日本語評価極性辞書[8]に株価に関連する語句を追加した辞書を使用する。日本語極性辞書には単語または用言ごとに pos あるいは neg のラベルがつけられている。本研究では、スコアを用いるため、pos のラベルには+1 のスコアを付与、neg のラベルには-1 のスコアを付与する。辞書中の語句が処理対象のテキストに出現した場合、対応するラ

ベルに付与された値を加算することでスコアを求める。そして、テキストすべての語句のスコアの合計をそのテキストのスコアとする。

本研究では、これを用いてすべての Web ニュースの記事、ツイートにスコア付けをする。例として、「A 社は 8 日、在宅勤務に関する職場からの距離制限を撤廃し、全国どこでも可能にする制度を導入したと明らかにした。出社が必要な場合は距離不問で交通費を全額支給し、従業員の負担を軽減する。育児や介護との両立支援の一環。」というニュース記事进行处理する場合、「支給」、「軽減」、「両立」、「支援」に pos、「負担」に neg がラベル付けされているため、pos が 4、neg が-1 となり、その値を加算して 3 というスコアが得られる。そして 1 日のすべてのテキストスコアを合計したものをその日のスコアとする。

##### 3.2.3 テクニカル分析

テクニカル分析には、以下の 4 つを用いる。

- MACD (Moving Average Convergence Divergence) 移動平均・収束・拡散指標と呼ばれる手法で、短期と中期の指数平滑移動平均 (EMA) がどの程度離れているか、どの程度近づいたかを分析し売買シグナルを見る手法である。n 日間の EMA は式 1 で求める。この値は、通常の移動平均と異なり、最近の株価ほど重視するが過去の株価の影響も考慮した移動平均である。式 1 において、 $k=2/(n+1)$ 、終値は当日の終値、前日 EMA(n) は前日までの EMA を意味する。この値は通常の移動平均よりも株価の動きに素早く反応する。通常、判断に用いる MACD 値は、式 2 で示す 12 日間の EMA と 26 日間の EMA の差となっている。

$$\begin{cases} 1 \text{ 日目の計算} | EMA(n) = n \text{ 日間の株価の平均} \\ 2 \text{ 日目以降} | EMA(n) = \text{終値} \times k + \text{前日} EMA(n) \times (1 - k) \end{cases} : \text{式 1}$$

$$MACD \text{ 値} = EMA(12) - EMA(26) : \text{式 2}$$

- ROC (Rate of change) ROC は、オシレーター系のテクニカル指標の 1 つであり、相場のモメンタム (勢い) を判断するのに有効とされる。式 3 に示すように、T 日の株価  $P(T)$  を、T 日より N 日前の株価  $P(T-N)$  で割り 100 を乗じたものを用いる。一般に ROC が 100 以上の時は強気相場と判断、ROC が 100 以下の時は弱気相場と判断する。

$$ROC(T) = \frac{P(T)}{P(T-N)} \times 100 : \text{式 3}$$

- BIAS (移動平均乖離率) BIAS は、終値とその移動平均線 (SMA) の差を観察するために使用される。T 日に対する n 期間の  $BIAS_n(T)$  は、式 4 を用いて求める。

$$BIAS_n(T) = \frac{P(T) - SMA_n(T)}{SMA_n(T)} \times 100 : \text{式 4}$$

ここで、 $SMA_n(T)$  は T 日における n 期間の株価のサンプル移動平均である。この指数は現在の株価が現在の株価の移動平均からどれだけ離れているかを示す。

- RSI (Relative Strength Index) RSI は、買われすぎか、売られすぎかを判断するため

の指標として利用されている。式 5 に示すように、期間内の株価の値上がり幅の平均を、値上がり幅の平均と値下がり幅の平均の和で割った値をパーセント表示する。一般的に 80% を超える場合、買われすぎと判断し、売りシグナル、20% を下回る場合、売られすぎで判断し、買いシグナルと評価する。

$$RSI = \frac{A}{A+B} \times 100 : \text{式 5}$$

A : n 日間の値上がり幅の平均

B : n 日間の値下がり幅の平均

### 3.2.4 機械学習の適用

本手法では、株価の予測が目的であるため、機械学習手法として教師あり学習を利用する。目的変数は T 日後の終値 (T=1, 7, 30)、説明変数は N 日分の Web ニュースのスコア (N=1, 5, 10)、N 日分の Twitter のスコア、テクニカル分析のスコアとする。説明変数の組み合わせを表 2 に示す。表 2 において○は使用、－は未使用を意味する。

表 2 : 説明変数の組み合わせ

パターン	テキスト		テクニカル	日数
	Twitter	Web		
1	○	○	○	1
2	○	○	○	5
3	○	○	○	10
4	○	○	－	1
5	○	○	－	5
6	○	○	－	10
7	○	－	○	1
8	○	－	○	5
9	○	－	○	10
10	○	－	－	1
11	○	－	－	5
12	○	－	－	10
13	－	○	○	1
14	－	○	○	5
15	－	○	○	10
16	－	○	－	1
17	－	○	－	5
18	－	○	－	10
19	－	－	○	1
20	－	－	○	5
21	－	－	○	10
22	－	－	－	－

### 3.2.5 株売買の判断

株売買の判断には予測結果とデータ最終日の終値を比較し、以下の基準に従い、判断する。

$\left\{ \begin{array}{l} \text{データの最終日の終値} \leq T \text{ 日後の予測結果ならば買い} \\ \text{データの最終日の終値} > T \text{ 日後の予測結果ならば売り} \end{array} \right.$

## 4. 評価

### 4.1 評価概要

本提案手法の有用性を検証するために、作成した予測モデルの精度と売買判断に関して、以下の評価を行った。

### 4.2 評価方法

#### 4.2.1 評価データ

対象企業として 4 社の企業を用いた。Web ニュースとして、各企業名が含まれる Yahoo! ニュースの記事、ツイートとしては、企業名および株という語句が含まれる日本語のツイートをを用いた。各企業の株価情報として Yahoo! ファイナンスより企業の終値を、2022 年 6 月 11 日から同年 11 月 30 日までの約 6 か月分収集した。収集したデータ数を表 3 に示す。

表 3 : 企業ごとのデータ数 (件)

企業	Web ニュース	Twitter
A 社	274	5668
B 社	7	389
C 社	143	11946
D 社	16	596

収集したデータの前 4 か月分を訓練用データ、残り 2 か月を評価用データとし、機械学習を実施した。

#### 4.2.2 機械学習

本研究では株価の予測を行うので、回帰の教師ありの機械学習として、線形回帰と SVM を使用した。

- 線形回帰  
線形回帰は、説明変数に対して目的変数の変化を関数でモデル化する手法である。線形回帰の基本モデルは式 6 で示される。

$$Y = v + w_1 X_1 + w_2 X_2 + \dots + w_p X_p : \text{式 6}$$

式 6 において、Y は目的変数、 $X_i, i=1, 2, \dots, p$  は説明変数、 $w_i$  は係数、v は切片を示す。

- SVM  
SVM は、訓練データ内でもクラス間のマージン (点と境界線の距離) を最大化する超平面を見つけることで、データをクラス分類する。超平面とは、n 次元空間内の n-1 次元の部分空間である。超平面は、新しい観測値がクラス分類される際の決定境界となる。超平面は線形であるため、多くの場合完全にクラスを分離できない。この場合はマージンを最大化することと、分類に失敗した観測値数の最小化の間でバランスをとる必要がある。また SVM はカーネル関数を用いることで非線形な決定境界を作ることができる。

#### 4.2.3 株価の予測精度評価

各手法、各企業の株価の予測精度は、式 7 で求める MAE を用いて評価する。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| : \text{式 7}$$

(n = データ数、 $\hat{y}$  = 予測値、y = 正解値)

4.2.4 売買判断の評価

売買判断の評価は次の方法を用いる。

テスト期間中の各日の対数収益率を、式 8 を用いて算出する。対数収益率がプラス、つまり T 日の実際の終値がデータ最終日の実際の終値より大きい場合、買い判断ならそのまま資産がプラス、売り判断ならマイナスになるため、買い判断が適切と評価する。逆に対数収益率がマイナスの場合は買い判断だとマイナス、売り判断ならプラスで売り判断が適切である。この点を加味し、式 9 に示すように、その日の売買判断が買いならば 1 を、売りならば-1 をかけた値を評価収益率とする。

$$\text{対数収益率} = \log \frac{T \text{ 日の実際の終値}}{\text{データ最終日の実際の終値}} \quad \text{式 8}$$

評価収益率

$$= \text{対数収益率} \times \alpha \begin{cases} \alpha = 1 & (\text{判断が買いの場合}) \\ \alpha = -1 & (\text{判断が売りの場合}) \end{cases} \quad \text{式 9}$$

上記で求めたテスト期間の評価収益率の総和を算出する。評価収益率が大きいほど適切な売買判断ができていると評価する。

4.3 評価結果

予測精度の結果を表 4 に、売買判断の評価結果を表 5 に示す。表 4 において、MAE は低いほど良い結果となっており、行で一番良い結果の色を変更し、企業内で一番良い結果を太字にしている。表 5 に表す売買判断の評価は期間の対数収益率の総和であり、多いほどよい結果である。色や太字の指定は、表 3 と同様である。

結果から、どの手法も株価のみで予測したものよりも良い結果を得ている。予測精度に関しては 1 日後が一番高く、遠くなるほど精度が悪くなっており、企業ごとに良いパターンも悪いパターンもどれも異なっている。売買判断に関

しては予測日での差は見られないが、企業ごとに良いパターンも悪いパターンもどれも異なっている。

4.4 考察

表 4 から予測精度はどの企業で線形回帰、SVM でも 1 日後の予測精度が一番高く、1 週間後、1 ヶ月後になると精度が悪くなるが、A 社、C 社は 1 日後、1 週間後、1 ヶ月後の差が小さいが、B 社、D 社は大きくなっている。これは学習データに偏りがあるためと考えられる。

図 2 に各企業の株価情報の動きを示す。縦線より左側が学習用データ、右側が評価用データである。A 社と C 社は学習用データと評価用データにおいて、株価の動きに変化が少ないことと、1 日後、1 週間後、1 ヶ月後でも動きに変化が少ないため予測精度の差が出なかった。一方、B 社、D 社は変化が大きいので予測精度に差が出ている。

売買判断の評価においては、表 6 に示すように、すべての企業、手法も一番良い結果は、すべて買い判断した場合よりも高い値であることから、適切な売買判断ができたとして評価する。しかし、パターンによっては悪くなっている場合もある。これは、判断のミスが続いたことと一度の判断ミスで大きな損失をしたことが考えられる。これは予測精度と売買判断の評価においても良い結果と悪い結果がリンクしていない要因と考えられる。

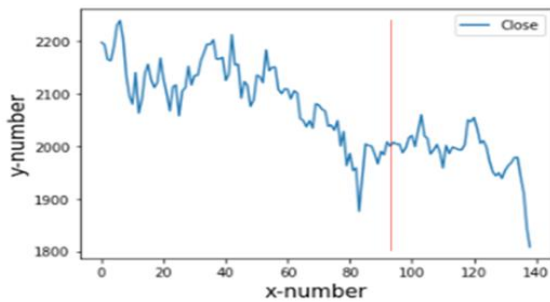
今回、テクニカル分析とテキストマイニングを掛け合わせて複数のパターンで検証した。テキストデータは企業によってデータ数に差があり、線形回帰では、テキストデータ数が多いほうがテキストデータを使用した際に悪い結果になりやすく、逆にテキストデータ数が少ない場合はよい結果になりやすい傾向が見受けられた。これは、今回集めたテキストデータは Twitter と Web ニュースであるためテキスト数が多い知名度がある、話題になっている企業ほど、関連が低いノイズとなるテキストデータが多くなりやすい傾向があると考えられる。逆にテキストデータ数が少ない企業のテキストデータは、関連度が高いものになりやすいと考えられる。

表 4 : 予測精度結果 (MAE)

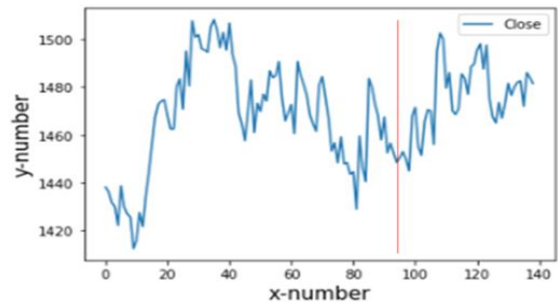
	パターン	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
線形回帰	A社	1日後	25.831	21.448	23.47	24.917	23.863	22.192	26.66	<b>21.063</b>	22.343	24.175	25.299	21.758	31.804	22.332	26.758	23.875	22.443	21.337	21.828	65.218	57.855	24.089
		1週間後	85.279	76.679	88.315	24.112	44.688	23.837	69.664	78.335	53.427	25.012	24.105	23.461	83.581	54.062	47.5	<b>22.727</b>	31.684	23.69	65.391	43.614	48.7	23.808
		1ヵ月後	170.05	160.97	242.65	42.187	46.944	39.859	196.29	176.93	259.32	41.503	43.492	47.425	168.83	153.5	241.36	41.996	<b>36.775</b>	39.532	195.35	105.652	93.66	40.611
	B社	1日後	8.041	8.186	8.153	9.89	10.355	10.293	8.002	8.246	7.85	9.889	10.127	10.218	7.829	7.757	8.671	9.914	9.85	10.116	<b>7.82</b>	28.521	28.72	9.979
		1週間後	40.187	40.005	46.189	33.098	37.482	37.373	40.19	39.431	39.408	33.019	38.664	34.623	40.273	40.688	49.001	<b>32.794</b>	35.533	35.95	40.279	42.348	43.894	32.796
		1ヵ月後	71.174	66.894	70.058	62.447	63.578	64.604	71.232	67.18	69.218	<b>62.243</b>	64.046	63.687	72.085	72.274	72.403	63.571	63.546	66.684	72.323	71.789	74.143	63.39
	C社	1日後	10.868	10.222	9.247	10.123	9.533	8.836	10.727	10.456	9.257	9.977	9.747	9.298	10.432	10.006	9.711	9.991	9.653	<b>9.088</b>	10.195	11.286	12.173	9.837
		1週間後	16.113	15.199	13.273	14.83	14.469	11.959	16.155	15.259	14.115	14.783	15.567	13.516	15.59	14.909	13.593	14.967	13.96	11.859	15.483	15.48	<b>12.74</b>	14.863
		1ヵ月後	13.234	16.876	17.107	11.749	19.799	<b>10.534</b>	13.25	16.447	14.204	11.895	20.179	10.809	13.639	14.047	15.658	11.368	11.508	10.769	13.663	13.007	14.638	11.481
	D社	1日後	80.111	82.779	100.73	63.998	<b>51.064</b>	58.63	67.986	90.53	73.747	62.042	52.563	60.045	84.262	66.057	85.507	57.094	52.145	53.626	68.522	443.08	399.73	51.447
		1週間後	244.45	273.9	152.72	102.5	155.742	106.16	236.14	281.25	188.1	100.95	161.39	119.282	258.39	240.6	132.42	102.31	99.827	<b>98.258</b>	244.3	252.044	330.16	99.485
		1ヵ月後	105.08	157.24	247.5	241.87	252.033	227.18	97.947	127.45	97.172	235.152	252.51	211.083	94.392	104.14	163.02	260.85	263.241	276.76	92.036	<b>88.485</b>	99.347	257.16
SVM	A社	1日後	47.277	48.722	43.306	58.842	22.115	19.761	51.737	49.529	46.278	61.664	23.297	21.459	44.853	48.271	42.869	60.463	24.952	<b>19.52</b>	48.475	107.512	93.016	61.938
		1週間後	46.781	39.767	37.402	68.344	<b>34.786</b>	60.328	44.917	44.618	37.972	69.528	41.299	59.403	50.489	42.307	39.227	65.273	37.612	58.248	50.876	81.617	74.006	61.655
		1ヵ月後	48.837	47.756	61.415	51.539	34.807	52.402	50.709	53.544	73.204	54.722	41.792	58.919	47.741	44.789	60.11	50.789	<b>32.646</b>	46.147	49.951	64.498	69.604	52.128
	B社	1日後	11.45	10.108	12.32	10.69	10.402	10.697	11.474	11.302	12.183	10.732	10.776	10.537	<b>9.887</b>	10.147	12.302	11.058	9.715	10.496	10.931	29.59	30.841	11.058
		1週間後	39.899	37.376	41.275	38.994	39.961	35.902	39.899	40.82	45.124	38.994	40.611	41.704	39.962	37.09	40.299	38.955	39.964	<b>35.327</b>	39.962	43.914	45.32	38.955
		1ヵ月後	68.498	67.902	67.226	63.574	66.448	69.465	68.783	69.179	63.905	<b>63.111</b>	68.534	63.285	67.804	67.676	69.758	64.532	65.721	69.704	67.734	67.644	71.468	64.781
	C社	1日後	9.805	9.787	9.096	10.178	10.087	<b>8.96</b>	9.704	9.927	9.109	10.015	9.909	9.391	9.833	9.605	9.228	10.22	10.096	9.449	9.853	11.547	11.965	10.094
		1週間後	14.634	15.237	12.88	14.441	13.367	13.195	14.681	14.767	13.651	14.106	14.152	13.574	15.035	14.867	13.349	14.385	13.845	13.377	14.932	13.538	<b>12.551</b>	14.049
		1ヵ月後	11.258	13.204	11.05	11.845	13.592	11.364	11.101	13.433	10.76	11.129	13.874	11.018	11.757	11.496	11.146	11.726	<b>10.189</b>	11.52	11.692	11.519	14.33	10.898
	D社	1日後	95.81	92.974	77.043	104.51	99.388	84.656	93.979	97.414	81.109	102.754	99.155	86.327	85.851	84.085	<b>75.7</b>	100.4	90.969	90.436	84.611	119.832	113.46	98.642
		1週間後	93.109	90.547	<b>83.203</b>	105.21	98.221	89.437	93.829	98.062	90.959	105.175	104.43	94.557	97.215	94.221	90.867	128.43	103.299	109.7	96.364	96.159	91.997	108.01

表 5 : 売買判断の評価結果

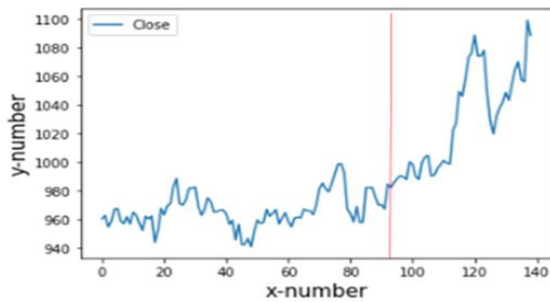
		パターン	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
線形 復帰	A社	1日後	0.21	0.17	0.1	-0.08	-0.001	-0.1	0.1	0.3	0.14	-0.03	-0.03	-0.06	0.22	0.23	0.15	-0.08	0.01	-0.11	<b>0.28</b>	0.03	-0.02	-0.03
		1週間後	0.03	0.04	-0.01	-0.03	0.06	0.03	0.03	0.03	0.01	-0.03	-0.03	-0.1	0.03	<b>0.05</b>	-0.01	-0.03	-0.002	-0.02	0.03	0.09	0.06	-0.03
		1ヵ月後	-0.03	0.03	-0.03	0.06	0.02	0.08	-0.03	0.03	-0.03	0.05	0.06	0.03	-0.03	-0.03	-0.03	<b>0.07</b>	0.03	-0.05	-0.03	-0.03	0.08	<b>0.14</b>
	B社	1日後	0.01	0.09	0.1	0.14	0.14	0.14	0.01	0.06	0.09	0.14	0.14	<b>0.15</b>	0.08	0.08	0.09	0.14	0.14	0.14	0.08	0.13	0.12	0.14
		1週間後	0.13	0.12	0.14	0.14	0.14	0.14	0.13	0.13	<b>0.17</b>	0.14	0.14	0.15	0.13	0.13	0.12	0.14	0.14	0.14	0.13	0.13	0.14	0.14
		1ヵ月後	0.14	0.14	0.16	0.13	0.13	0.15	0.14	0.16	<b>0.16</b>	0.13	0.13	0.15	0.14	0.14	0.15	0.14	0.13	0.15	0.14	0.11	0.14	0.13
	C社	1日後	0.23	<b>0.25</b>	0.23	0.08	0.1	0.14	0.24	0.24	0.23	0.09	0.07	0.11	<b>0.25</b>	0.24	0.18	0.05	0.02	0.17	<b>0.25</b>	0.23	0.2	0.08
		1週間後	0.19	0.2	0.19	0.07	0.08	0.12	0.22	0.21	0.2	0.09	0.11	0.03	0.18	<b>0.21</b>	0.11	0.07	0.07	0.11	0.19	0.2	0.19	0.08
		1ヵ月後	0.07	0.12	0.13	0.06	0.08	0.08	0.08	0.08	0.08	0.06	0.08	0.08	0.08	0.08	0.11	0.08	0.08	0.08	0.08	0.08	<b>0.15</b>	0.12
	D社	1日後	0.14	0.08	0.21	-0.08	0.03	0.21	-0.01	0.12	0.21	-0.02	0.16	0.21	0.14	0.14	0.21	<b>0.32</b>	0.09	0.12	0.14	-0.13	-0.13	0.15
		1週間後	0.07	0.13	0.21	0.29	-0.04	0.25	0.07	0.08	0.21	0.2	0.14	0.27	0.14	0.14	-0.05	0.29	0.13	<b>0.31</b>	0.14	-0.13	-0.13	0.29
		1ヵ月後	-0.14	-0.13	-0.14	0.11	0.22	<b>0.43</b>	-0.14	0.002	-0.03	0.11	0.23	0.44	-0.14	-0.14	-0.13	0.15	0.15	<b>0.38</b>	-0.14	-0.003	0.09	0.15
S V M	A社	1日後	-0.03	-0.03	-0.06	-0.03	<b>0.1</b>	0.06	-0.03	-0.03	-0.06	-0.03	<b>0.1</b>	0.01	-0.07	-0.03	-0.06	-0.03	0.04	0.06	-0.03	-0.03	-0.03	-0.03
		1週間後	-0.03	-0.03	-0.03	-0.03	<b>-0.002</b>	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.003	-0.03	-0.03	-0.03	-0.03	-0.03
		1ヵ月後	-0.03	0.12	-0.07	-0.04	0.07	-0.07	0.12	0.05	-0.03	<b>-0.002</b>	<b>0.19</b>	-0.003	0.02	0.12	-0.11	-0.05	0.11	-0.02	0.1	0.03	0.05	0.04
	B社	1日後	0.12	0.14	0.15	0.15	0.14	<b>0.19</b>	0.12	0.12	0.15	0.15	0.14	0.16	0.14	0.14	0.15	0.15	0.14	0.19	0.13	0.13	0.12	0.15
		1週間後	0.13	0.13	0.14	0.14	0.14	0.14	0.13	0.13	0.14	0.14	0.14	0.14	0.13	0.13	<b>0.15</b>	0.14	0.14	0.14	0.13	0.11	0.14	0.14
		1ヵ月後	0.14	0.14	0.14	<b>0.15</b>	0.14	0.14	0.14	0.14	0.14	<b>0.15</b>	0.14	<b>0.15</b>	0.14	0.14	0.14	<b>0.15</b>	0.14	0.14	0.14	0.13	<b>0.15</b>	<b>0.15</b>
	C社	1日後	0.2	0.2	0.19	0.05	0.19	0.18	<b>0.27</b>	0.23	0.19	0.07	0.21	0.14	0.18	0.19	0.22	0.07	0.19	0.16	0.22	0.23	0.2	0.09
		1週間後	0.18	0.16	0.14	0.11	0.07	0.04	0.15	0.18	0.14	0.13	0.08	0.04	0.16	0.17	0.05	0.12	0.11	0.05	0.16	<b>0.21</b>	0.17	0.13
		1ヵ月後	0.05	0.06	0.05	0.06	0.05	0.08	0.05	0.06	0.06	0.06	0.05	0.08	0.07	0.07	0.08	0.08	0.09	0.08	0.07	0.13	<b>0.15</b>	0.09
	D社	1日後	0.23	0.12	0.27	0.2	0.28	0.35	0.23	0.12	0.27	0.2	0.25	<b>0.35</b>	0.09	0.12	0.26	0.16	0.27	0.32	0.12	0.12	0.12	0.18
		1週間後	0.19	0.2	0.23	0.17	0.18	0.24	0.17	0.25	<b>0.38</b>	0.17	0.25	<b>0.24</b>	0.25	0.18	0.35	0.25	0.27	0.3	0.25	0.12	0.12	0.28
		1ヵ月後	0.17	0.22	0.29	0.29	0.29	<b>0.43</b>	0.19	0.22	0.24	0.29	0.32	<b>0.43</b>	0.16	0.19	0.26	0.29	0.29	0.37	0.19	0.22	0.08	0.29



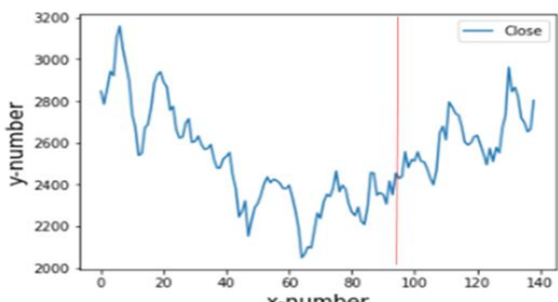
A社



C社



B社



D社

図 2 各企業の株価情報

表 6 : 対数収益率の総和の比較

	A社	B社	C社	D社
すべて買い	0.03	0.11	0.03	0.14
テクニカル分析	0.28	0.15	0.25	0.22
テクニカル分析+テキスト	0.23	0.15	0.27	0.29
テキスト	0.10	0.19	0.21	0.43

しかし、全体的にみると、どのパターンが良いのか、悪いのかの結果を見だせていない。そのため、一定の方向性を見出すためにサンプル数を増やす必要がある、またサンプル数を増やすことによって株価の動きに合うパターンなど新しい発見を得られる可能性がある。

#### 4.5 今後の課題

今後の課題としては、以下のことが挙げられる。

1 つとして学習、テストデータの検討が挙げられる。A社、D社は株価が下がる場合の学習が多く、上がる場合の

学習が少ない。評価の際に A 社のテストデータは下がりになっており、D 社は上がりになっているため、D 社の予測精度が悪くなっている。このような差をなくすためにデータを整理する必要がある。

2 つ目は、機械学習手法の検討である。本実験では機械学習に線形回帰と SVM を使用しているが、パラメータチューニングをしていない。このことからパラメータチューニングをすることでより良い結果が得られると考えられる。また、株価の全体的な動きとして線形回帰や SVM よりも良い機械学習手法を検討していく必要がある。

3 つ目は評価方法の検討である。本研究では売買判断の評価として、対数収益率を基本として用いたが、対数収益率で評価した場合、一度の判断ミスで大きく損をした場合に、他の判断が適切でも最終的な評価が悪くなる場合がある。ここは判断がどれだけ正しかったかを評価するのか、大きな損失をしない判断を重視するかといった判断の評価基準とそれに適した方法を検討していく必要がある。

4 つ目としてサンプルの追加が挙げられる。検証では 4 つの企業を用いたが、良い結果、悪い結果のパターンが違ったため、本手法の有用性を見出すのに至っていない。サンプルを追加し、検証することで、本手法が有用となる条件があるかを検証する必要がある。またそれらを踏まえることで、株価の動き方などに適したパターンの発見などができると考えられる。

5 つ目にテキストデータの選別が挙げられる。テキストデータが多い場合に関連度の低いノイズが入っていると考えられるため、フィルタリングなどをして関連度の高いものだけにする。また現在は、Twitter は株という単語と企業名が入っているツイートを、Web ニュースは企業名が入っている記事だけを集めている。しかし、株価に関連するニュースとしては企業名が出ていなくても経済全体、業界全体に影響を与えるものもあるため、それらのツイートや Web ニュースについても集める必要がある。

6 つ目にテクニカル分析の検討である。本研究では MACD、ROC、BIAS、RSI の 4 つのテクニカル分析手法を用いた。しかし、これらの手法以外のテクニカル分析手法もあるため、他の分析手法を用いたパターンや、4 つすべてを用いないパターンパターンの検討も課題である。

## 5. おわりに

株式投資を始める人が増えてきており、そのような初心者の株式投資を支援する方法の 1 つとして株価予測がある。株価予測の手段として、近年、テキストマイニングや機械学習を使う研究が増えてきている。

以上の背景から、本研究では、先行研究で多い平均株価ではなく企業の株価を予測することに着目し、テキストデータとテクニカル分析を用いた株価予測手法を提案した。テキストデータとしては、正確な情報であると考えられる Web ニュースと大量のデータとニュースにならないような企業の評判を取り入れられる Twitter の両方を使用する。テキストデータに評判分析を適用し得たスコアを用いて機械学習で構築したモデルにより企業株価を予測する。さらに予測結果を用いて株の売買判断を行う手法について提案した。

提案手法では、Web ニュース、Twitter から集めたテキストデータに対して評判分析を適用しスコアを算出、株価にデータに対してテクニカル分析を適用し、その 2 つの分析

結果と株価情報を用いて機械学習により予測モデルを作成する。このモデルを用いた予測結果をもとに株の売買判断を行う。

本手法の有効性を検証するために、データとして、4 つの企業をサンプル企業とし、Web ニュース、Twitter、株式市場から 2022 年 6 月 11 日から同年 11 月 30 日までの約 6 か月分の株式情報を集め、前半 4 か月を用いて予測モデル構築、後半 2 か月のデータを用いてを検証した。機械学習としては、線形回帰と SVM を使用し、目的変数は T (T=1, 7, 30) 日後の終値、評判分析したツイート、Web ニュースとテクニカル分析結果の組み合わせで複数のパターンで評価を行った。株価の予測精度と評価収益率を用いた評価の結果、予測精度として最も良い結果は株価のみを用いたものより良い結果を得たが、悪くなった場合も見受けられた。このような結果を得た理由としては、予測精度には使用したデータの偏りが結果に影響していると考えられる。売買の評価は、買い判断をし続けた場合より、良い結果を得ることが出来たことから、適切な判断ができていると評価する。しかしパターンによっては提案手法の方が悪くなっている。これは、判断のミスが続いたことや一度の判断ミスで大きな損失をしたことが考えられる。このことは予測精度と売買判断の評価においても良い結果と悪い結果がリンクしていない要因にもなっている。また予測精度、売買の評価どちらも良い結果のパターン、悪い結果のパターンが企業ごとに違っており、汎用的なパターンの発見には至っていない。

今後の課題として、データの整理、機械学習手法の検討、評価方法の検討、サンプル企業の追加、テキストデータの選別、テクニカル分析手法の検討などが挙げられる。

## 参考文献

- [1] 谷畑侑昭, 初心者でも分かる株式投資新しい「定石」と「読み」一簡単・確実!基礎知識から応用・テクニカル分析まで一, 東京, 現代書林 (1996)
- [2] 鈴木智也, 中川慧, 伊藤友貴, 坂地泰紀:金融におけるテキストマイニングと機械学習応用, 人工知能学会, 36 号 3 号, 270-278(2021)
- [3] Johan Bollen, Huina Mao, Xiao-Jun Zeng : Twitter mood predicts the stock market, Journal of Computation Science, 2, 1-8 (2010)
- [4] Shangkun Deng, Takashi Mitsubuchi, Kei Shioda, Tatsuro Shimada, Akito Sakurai : Combining technical analysis with sentiment analysis for stock price prediction, IEEE 9th International Conference on Dependable, Autonomic and Secure Computing, DASC 2011, 800-807(2011)
- [5] 増井佑亮, 藤野巖 : Twitter の多軸的感情情報を利用した株価の予測, 言語処理学会第 24 回年次大会発表論文集, 69-72(2018)
- [6] 佐藤謙太, 高知宏, 黒岩文介, 白井治彦 : ネガポジ解析による Web データと株価変動の相関関係評価, 福井大学大学院工学研究科研究報告, 63, 75-86 (2015)
- [7] 大塚裕子, 乾孝司, 奥村学, 意見分析エンジン—計算言語学と社会学の接点, コロナ社 (2007)
- [8] 日本語評価極性辞書:  
[http://www.cl.ecei.tohoku.ac.jp/Open\\_Resources-Japanese\\_Sentiment\\_Polarity\\_Dictionary.html](http://www.cl.ecei.tohoku.ac.jp/Open_Resources-Japanese_Sentiment_Polarity_Dictionary.html)  
(最終アクセス: 2022 年 6 月 10 日)