

株価とニュースデータを用いた株価変化率予測モデル

高橋 健太(14X4055) 藤戸 四恩(18X4120) 古川 隆証(18X4121) 指導教員 五島 洋行

1. はじめに

近年、自己資産形成の必要性が謳われ、株取引を始める個人投資家の数が増加している。投資家の銘柄選択や売買の意思決定を支援するために予測手法に関する研究が様々な専門家や機関で進められている。過去の株価や株価変化率を用いて将来の株価を予測するモデルや、ニュースデータを用いて機械学習を行い将来の株価を予測する研究が発表されている。しかし、その両方を用いた研究は発表されていない。

そこで本研究では、過去の株価変化率とニュースデータの両方を用いた株価変化率予測を行い、それぞれ一つの指標から株価変化率を予測するより予測精度の向上を図る。

2. 利用するモデル

2.1 LSTM

LSTM(Long Short-Term Memory) [1]は、時系列データを扱うことができるニューラルネットワーク RNN(Recurrent Neural Network)の一種で、中間層は再帰的な構造となっている。RNNは、誤差逆伝播を用いて学習を行うが、系列が長くなると重み W と $\tanh$ の微分を何度も掛けるため勾配が徐々に小さくなり勾配消失、または徐々に大きくなり勾配爆発してしまう。対して LSTM では、データを線形和で保持するため、逆伝播での勾配が極端に大きくなったり小さくなったりせず、勾配消失または勾配爆発が起こりにくい。従来の RNN では学習できなかった長期依存を学習可能である。

2.2 BERT

BERT (Bidirectional Encoder Representations from Transformers) [2] は Transformer[3]の Encoder 部分を使用した大規模な生コーパスで Pre-Training し、各タスクで Fine-Tuning するという 2 ステップからなっている。

Pre-Training では教師なし学習の 2 つのタスクがある。1 つ目は Masked Language Modeling で、入力された単語トークンの 15% を [Mask] トークンでマスクし元のトークンを当てるタスク。2 つ目は Next

Sentence Prediction で文同士の関係を学習するタスクであり、2 文を選び連続しているかどうかを当てるタスクである。

Fine-tuning は、Pre-Training で学習済みのモデルを各タスクで教師あり学習として再学習を行う。

2.3 勾配ブースティング

勾配ブースティングはアンサンブル学習のブースティングと決定木を組み合わせた手法であり、決定木を直列に構築し、直前の決定木の結果に重みを加味し新たな決定木で予測する。アンサンブル学習をすることで 1 つの学習モデルで予測するより低バイアス、低バリエーションとなり予測精度を向上させることが可能である。決定木は目的変数の純度を最も高める説明変数を見つけ、樹木状モデルを作り分析する機械学習の手法である。

3. 実験概要

3.1 予測方法

予備実験として、1 つ目に過去の株価変化率を用いて時系列データの学習に適している LSTM モデルで予測を行う。2 つ目にニュースデータを用いて文章の学習に適している BERT モデルで予測を行う。ニュースデータはテキストデータのため、前処理をし、学習モデルに適合させる。そして、2 つのモデルをアンサンブル学習し予測を行う。予測方法の流れは図 1 のとおりである。

3.2 使用するデータ

2012 年 9 月 1 日から 2021 年 9 月 1 日までの日経平均株価を用いる。なお、株価そのものでは価格変動が大きく、予測精度が低くなってしまう可能性が高いため、対数変化率を用いる。対数変化率は以下のとおりである。

$$\log\left(\frac{S_t}{S_{t-1}}\right) = \log(S_t) - \log(S_{t-1}) \quad (1)$$

(S_t : 時刻 t における株価)

また、ニュースデータは同期間における Reuters 社のビジネスニュースを使用する。

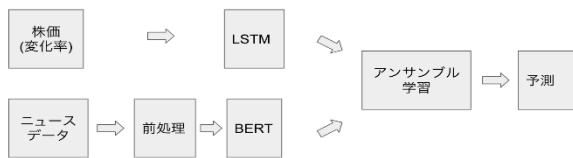


図 1 予測方法の流れ

3.3 実験環境

実験環境は以下の表 1 のとおりである。

表 1 実験環境

CPU	Intel(R) Xeon(R) CPU × 2
OS	Microsoft Windows10 Home
Memory (CPU)	24.0GB
GPU	RTX3090
Memory (GPU)	24.0GB
使用言語	Python3.8.8

4. 実験

4.1 実験設定

連続した 5 日分の株価変化率を入力として、6 日目の株価変化率の予測を行う。ニュースデータは予測する日の前日に発行されたニュースの中から 1 つをランダムに選び、予測に用いる。それぞれの実験において各データの 80%を学習データ、20%をテストデータとする。

これらの 2 つの実験から得られた予測値を特徴量とし、アンサンブル学習のスタッキングの手法で予測を行う。メタモデルには勾配ブースティングを用いて予測を行う。勾配ブースティングのフレームワークとして LightGBM を使用する。

4.2 実験結果

表 2 の実験結果において、複合モデルが RMSE, MAE 共に過去の株価変化率のみを用いて予測した場合と、ニュースデータのみを用いて予測した場合より、評価関数の値が低く、精度が高いという結果が得られた。また、表 3 より過去の株価変化率のみを用いた予測は標準偏差が小さく、小さい範囲でしか予測が行えていないことが分かる。ニュースデータのみを用いた予測は三つの予測の中で標準偏差は最も大きく、大きい範囲で予測を行うことが出来ているが、評価関数の値は最も大きく、精度が低い。複合モデルは、評価関数の値が最も小さく、平均値や中央値の両方で三つの予測モデルの中で実測値に近い値となったため、予測精度を向上させることが出来たと言える。

表 2 各予測の評価関数の値

	RMSE	MAE
過去の株価変化率のみ	0.01409	0.00972
ニュースデータのみ	0.01554	0.01153
複合モデル	0.01400	0.00970

表 3 各予測の基本統計量

	平均値	中央値	標準偏差
過去の株価変化率のみ	0.00153	0.00153	0.00026
ニュースデータのみ	-0.00381	-0.00403	0.00621
複合モデル	0.00051	0.00044	0.00140
実測値	0.00062	0.00069	0.01408

5. おわりに

本研究では、過去の株価変化率と、ニュースデータの両方を用いることで、株価に影響する複雑なマーケット環境を考慮した予測が行えると考え実験を行った。

今回の実験から、複合モデルを用いた予測が予備実験の 2 つの予測と比べ、精度が高いという結果が得られた。

本研究は複雑なマーケットに対応した分析を行えるよう、複数のデータを組み合わせた予測に対し有効な提案になると考えられる。今後、予測精度を更に高め、個人投資家の意思決定を支援する実サービスへ応用することにより、金融業界の発展の一助となれば幸いである。

参考文献

- [1] S. Hochreiter, and J. Schmidhuber : “Long Short-Term Memory”, Neural Computation, Vol.9, No.8, pp.1735-1780, 1997.
- [2] J. Devlin, M. Chang, K. Lee, and K. Toutanova : “BERT : Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint arXiv : 1810.04805, 2018.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin : ”Attention is all you need”, NIPS2017, pp.5998-6008, 2017.

令和3年度卒業研究論文

株価とニュースデータを用いた 株価変化率予測モデル

法政大学 理工学部 経営システム工学科

経営数理工学研究室

14X4055 高橋 健太

18X4120 藤戸 四恩

18X4121 古川 隆桢

指導教員 五島 洋行 教授

学科名	経営システム工学科	学籍番号	14X4055 18X4020 18X4121
申請者氏名		高橋健太 藤戸四恩 古川隆梃	
指導教員氏名		五島洋行	

論文要旨

論文題目	株価とニュースデータを用いた 株価変化率予測モデル
------	------------------------------

近年、日本では銀行預金の金利の低さや年金 2,000 万円問題などを受け、老後資金は自身で形成する必要があると言われており、株取引を始める人々が増加している。そうして増えた個人投資家たちの意思決定の手助けとなる分析や株価の予測システムが様々な研究機関によって進められている。過去の株価や株価変化率から将来の株価を予測する手法や、ニュースデータなどのテキストデータから株価の予測をする手法は以前から散見されている。しかし、その両方を組み合わせた予測方法は未だ見受けられない。

そこで本研究では、過去の株価変化率とニュースデータの両方を用いた株価変化率予測を行い、それぞれ一つの指標から株価変化率を予測するより予測精度の向上を図る。

比較のための予備実験として、過去の株価変化率のみ、ニュースデータのみをそれぞれ用いた株価変化率予測を行う。前者では、LSTM モデルを用いて過去の連続した 5 日分の変化率のデータから 6 日目の株価変化率の予測を行う。後者は予測する日の前日に Reuters 社から発表された Business ニュースのうちランダムで 1 つを選択し、そのニュースデータから BERT モデルを用いて 1 日分の株価変化率の予測を行う。

実験の結果、三つの予測モデルの中で 2 つのモデルを組み合わせた提案手法の精度が最も高いという結果が得られた。

目次

第1章 はじめに.....	4
1.1 研究背景と目的	4
1.2 本論文の構成	5
第2章 先行研究.....	6
第3章 関連知識.....	7
3.1 Deep Learning	7
3.2 勾配ブースティング.....	11
3.3 Adam.....	12
第4章 実験概要.....	13
4.1 提案方法.....	13
4.2 使用データ.....	13
4.3 評価関数.....	14
4.4 実験環境.....	15
第5章 過去の株価変化率のみを用いた予測	16
5.1 予測モデルの概要	16
5.2 予測結果.....	17
第6章 ニュースデータを利用した予測.....	19
6.1 予測モデルの概要.....	19
6.2 前処理	20
6.3 予測結果.....	20
第7章 2つのモデルを組み合わせた予測.....	22
7.1 予測モデルの概要	22
7.2 予測結果.....	24
第8章 実験結果の考察	25
第9章 おわりに.....	29
参考文献.....	31
謝辞.....	32

第 1 章 はじめに

1.1 研究背景と目的

近年、自己資産形成の必要性が謳われ、株取引を行う人々が増えている。自身の資産を株取引によって増やすためには、銘柄を選択し売買のタイミングを見極める事が必要となる。そこで、一般投資家の意思決定を支援するため、株価変動の指標や予測手法についての考察が様々な投資家や機関によって現在も進められている。

株価を予測するうえで用いる手法は大きく分けて2つある。1つは過去の株価などの時系列の値動きを扱うテクニカル分析である。もう1つは、企業の財務状況や業績、更には企業を取り巻く社会情勢などを基に予測を行うファンダメンタルズ分析である。

かつては株価予測において、重回帰分析など過去の株価推移や様々な経済指標を用いた手法が投資家の間でよく用いられていた。しかし、2000年頃からIT技術の発展に伴いSNSの普及や、マーケットのグローバル化が進むことで今までの計量経済学や金融工学の知識だけでは予測できない市場の変化が見られるようになってきている。

時を同じくして、機械学習の分野では2006年にジェフ・ヒントン氏が何層にもネットワークを重ねても精度を損なわない手法を提案し、深層学習(Deep Learning)[1]が注目を集めた。深層学習は、多量のデータが多層のネットワークを経由していくうちに、高次の特徴量から低次の特徴量を自動的に見つけ、学習を行うことが出来る。そのため、それまで機械学習を行う際は人間が特徴量を指定しなければならなかったという問題点が解決された。

それ以降、時系列データを扱う際に有効なRNN (Recurrent-Neural-Network) やその派生であるLSTM (Long Short-Term Memory)などで株価数値の変動を用いた株価や変化率の予測、ニュースデータを基に株価が上昇するか下落するかを予測するモデルの研究が発表されている。しかし、株価変化率の予測と、ニュースデータの両方を用いた予測を組み合わせた研究は見受けられない。

そこで本研究の目的は、過去の株価変化率とニュースデータの両方を用いた株価変化率予測を行うことで精度の向上を図り、株取引を優位に行えるようにすることである。

過去の株価変化率から予測をした場合は covid-19 などの企業を取り巻く環境の変化から起こる急激な株価変動を予測できていない。一方、ニュースデータを用いて予測をした場合、予測精度は落ちるものの急激な株価変動に対応した予測をする。そこで、両方のデータを用いて予測することでそれぞれの欠点を補完し、現在の複雑なマーケット環境を考慮した、より良い予測を行えると考える。

本研究では、予備実験として株価変化率とニュースデータそれぞれから深層学習を用い精度の評価を行う。

1.2 本論文の構成

本論文は9章で構成されている。

第2章では先行研究を述べる。

第3章では本研究で用いる学習モデルの関連知識を述べる。

第4章では本研究の実験の概要及び、使用するデータと実行環境について述べる。

第5章では過去の株価変化率のみを用いた予測モデルの概要と結果を述べる。

第6章ではニュースデータを用いた予測モデルの概要と結果を述べる。

第7章では2つのモデルを組み合わせた予測モデルの概要と結果を述べる。

第8章では第5章と第6章と第7章の結果を比較し、考察を述べる。

第9章で本研究のまとめを述べる。

第 2 章 先行研究

株価予測における手法は主に 2 つあり、それぞれテクニカル分析とファンダメンタルズ分析を用いた先行研究を紹介する。

テクニカル分析とは、過去の値動きをチャートで表し、今後の値動きや動向を予測するものである。トレンドの転換点や現在の株価水準の判断には、チャートだけでなく、独自の計算式に基づいた様々な指標が用いられる。テクニカル分析は数値やグラフの概形から予測を行うため、経済に関する知識が深くなくても予測が出来るという利点がある。

Deep Learning を株価予測に応用した研究[2]では、中間層が 1 層の簡単なフィードフォワードニューラルネットワークで 1 分足ごとの変化率を用いて学習している。予測する値は 1 分前より上がるかまたは下がるかを予測する二値分類で、67%の正解率となっている。

また、テクニカル指標であるローソク足の画像を訓練データとして深層学習を行い、直近の株価が上昇するか、下落するかを予測する研究が発表されている[3]。その研究では、CNN モデルで 3 か月以上の運用を行うと、正解率が 90%を超える。

ファンダメンタルズ分析は、企業の業績や財務状況、企業を取り巻く社会情勢を基に予測を行う手法であり、「株価は最終的に企業価値に見合った価格となる。」という前提のもとで行われ、株価が企業価値に対し、割安か割高かの判断をするために行われる。

過去には、Reuters 社が発行しているビジネスニュースの日本証券市場に関する文章をベクトル表現し、実際の株価を基にネガポジ判定をしたものを訓練データとして機械学習を行い、投資戦略を求める研究が発表されている[4]。結果は超過収益を獲得しており、機械学習によるニュース記事の評価を用いて将来の株価を予測する可能性を見出した。

第 3 章 関連知識

本章では、本研究で用いる学習モデルの関連知識について述べる。

3.1 Deep Learning

Deep Learning とは、人間の神経細胞（ニューロン）の仕組みを模した多層の人工ニューラルネットワークによる機械学習の手法である。十分なデータがあれば、多層のニューラルネットワークがデータを学習し、データ内に含まれている特徴量を見つけ出す事が可能な技術方法である。

① LSTM (Long Short-Term Memory)

LSTM[5]図 1 は、時系列データを扱うことができるニューラルネットワークである RNN の一種で、中間層は再帰的な構造となっている。RNN は、誤差逆伝播を用いて学習を行うが、系列が長くなると重み W と $\tanh$ の微分を何度も掛けるため勾配が徐々に小さくなり勾配消失、または徐々に大きくなり勾配爆発してしまう。対して LSTM では、データを線形和で保持するため、逆伝播での勾配が極端に大きくなったり小さくなったりせず、勾配消失または勾配爆発が起こりにくい。そのため従来の RNN では学習できなかった長期依存を学習可能である。

LSTM はセルの状態を更新および制御するため、忘却ゲート、入力ゲート、出力ゲートの 3 つのゲートがある。

忘却ゲートは、新しいデータが与えられた際にセルのどの情報を忘れるかを制御する。時刻 t における入力値を x_t 、セルの状態を c_t 、出力値を h_t とすると、忘却ゲート f_t は以下の式である。

$$f_t = \sigma(W_t \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

W_f は重み、 b_f は切片、 σ はシグモイド関数を表す。

入力ゲートは、新しいデータが与えられた際にどの情報を更新するかを制御する。入力ゲートの値 i_t は以下の式である。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t]) \quad (3)$$

W_i は重みを表す.

セルの更新値 $\tilde{c}_t$ は以下の式である.

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t]) \quad (4)$$

W_c は重みを表す.

出力ゲートは、どの情報を出力するかを制御する. 出力ゲートの値 o_t と出力値 h_t は次の式で与えられる.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

W_o は重み, b_o は切片を表す.

セルの状態は以下の線形和の式となる.

$$C_t = C_{t-1} \otimes f_t \oplus \tilde{c}_t \otimes i_t \quad (7)$$

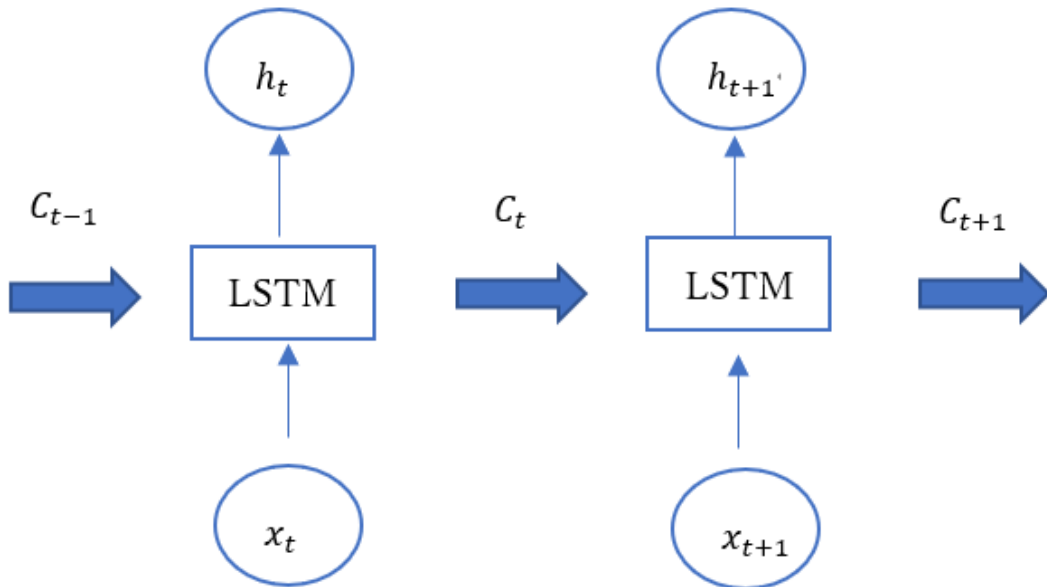


図1 LSTM モデル

② Transformer

Transformer[6]は、Attentionのみを使用したEncoder-Decoderモデルであり、長距離の依存関係を捉えることができる。

また並列計算をすることにより学習時間の短縮をすることが可能である。Transformerの全体像は図2に示す。左側がEncoder、右側がDecoderである。

AttentionはQueryとKeyとValueの3つのベクトルで計算される。EncoderのQuery, Key, Valueは入力単語Embedding X にそれぞれの重み(W^Q , W^K , W^V)を掛け合わせた(XW^Q , XW^K , XW^V)である。QueryとKeyは次元 d_k を持つ。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\left(\frac{QK^T}{\sqrt{d_k}}\right)V\right) \quad (8)$$

MultiHead Attentionは、各単語ベクトルをheadの数で分割し、それぞれのheadを重み W^O でconcatする。

$$\text{MultiHead}(Q, K, V) = \text{concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (9)$$

$$\text{Where } \text{head}_h = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

$$\text{Where the projections are parameter matrices} \quad (10)$$

$$W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}, \quad W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}, \quad W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$$

$$\text{and } W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}.$$

Positional Encodingは文章上の位置な依存関係を学習する。

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (11)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right) \quad (12)$$

FeedForwardはMultiHead Attention後に2層のニューラルネットワークで1層目が線形和で2層目がReLUを持つ。

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (13)$$

b_1, b_2 は切片を表す。

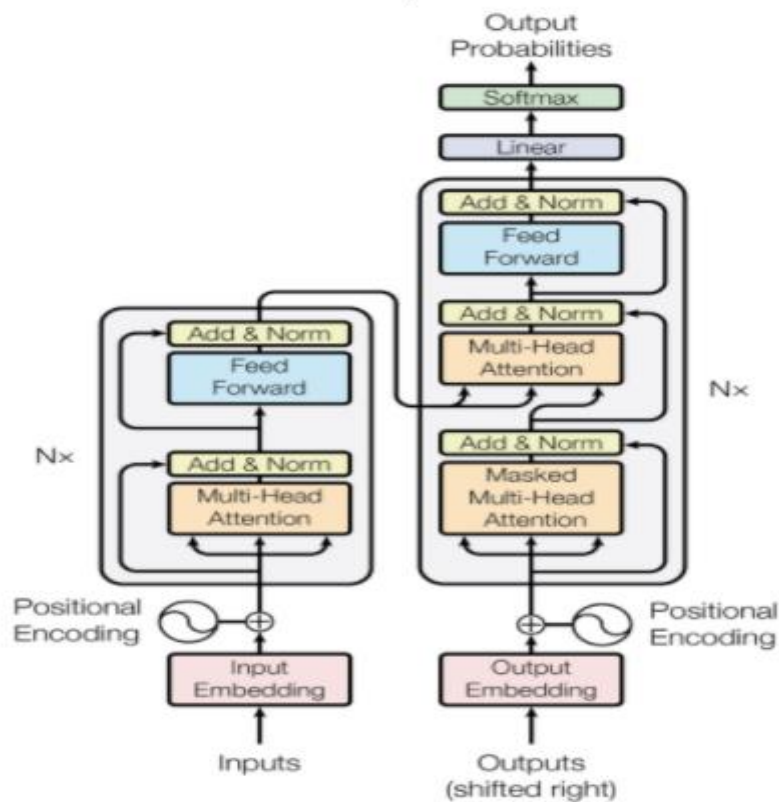


図 2 Transformer モデルの概要[4]

③ BERT

BERT(Bidirectional Encoder Representations from Transformers)[7] 図 3 は Transformer の Encoder 部分を使用した大規模な生コーパスで Pre-Training し、各タスクで Fine-Tuning するという 2 ステップからなっている。

Pre-training では教師なし学習の 2 つのタスクがある。1 つ目は Masked Language Modeling で、入力された単語トークンの 15% を [Mask] トークンでマスクし元のトークンを当てるタスク。2 つ目は Next Sentence Prediction で、文同士の関係を学習するタスクであり、2 文選びそれらが連続しているかどうかを当てるタスクである。

Fine-Tuning は、Pre-Training で学習済みのモデルを各タスクで教師あり学習として再学習を行う。

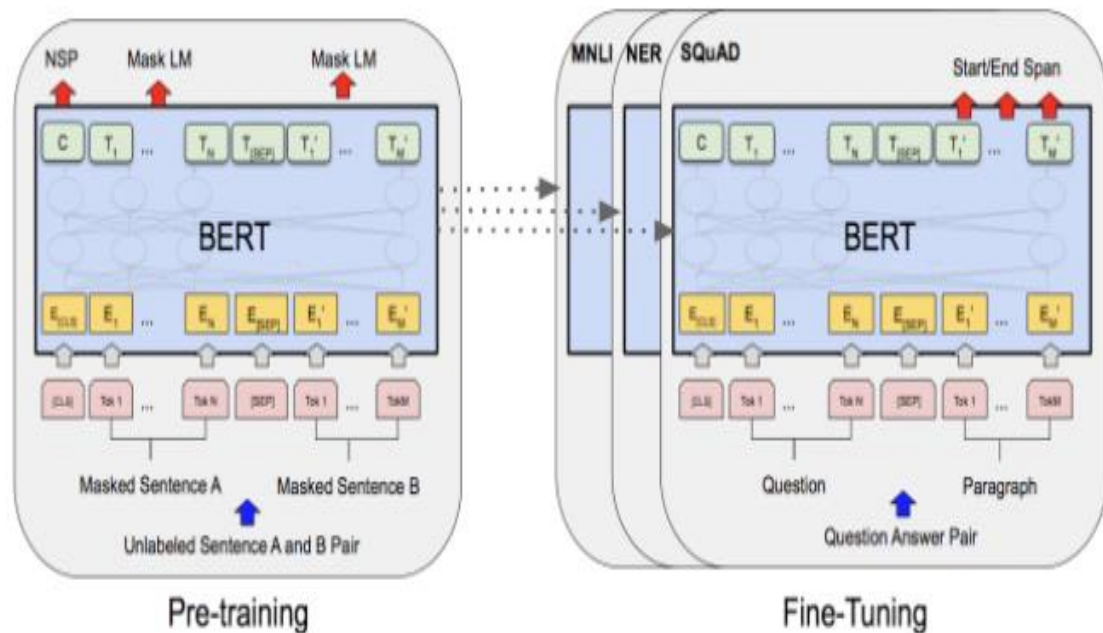


図3 BERTモデルの概要 [7]

3.2 勾配ブースティング

アンサンブル学習のブースティングと決定木を組み合わせた手法である。アンサンブル学習は複数のモデルの分類器からそれぞれ予測させ、その結果を統合し全体の予測を行う。

1つの学習モデルではバイアスとバリエーションはトレードオフ関係にあるため、低バイアスの状態では、高バリエーションとなり過学習となることが多い。バイアスとは実際値と予測値との誤差の平均、バリエーションは予測値の分散である。アンサンブル学習をすることで、1つの学習モデルで予測するより低バイアス、低バリエーションとなり予測精度を向上させることが可能である。

アンサンブル学習はバギング、ブースティング、スタッキングの3つの手法がある。バギングは複数の分類器の結果の平均や多数決で予測を行う手法で、それぞれの分類器を並列に学習することが可能である。ブースティングは弱学習器を直列に構築し、直前の弱学

習器の結果に重みを加味し新たな弱学習器で学習させることで予測を行う手法。スタッキングは複数の異なる学習モデルの結果を特徴量の入力としてメタモデルで学習し、予測を行う手法である。

決定木は目的変数の純度を最も高める説明変数を見つけ、樹木状モデルを作り分析する機械学習の手法の1つである。対象を分類する分類木と対象の数値を推定する回帰木の総称を決定木という。

勾配ブースティングは決定木を直列に構築し、直前の決定木の結果に重みを加味し新たな決定木で予測する。その際に、勾配降下法を用いて損失関数を最小化するように学習させる。

3.3 Adam

機械学習では学習を行う際、損失関数の値と勾配からパラメータを更新する。予測精度が最大になるよう重みパラメータの最適化を行うアルゴリズムのことを Optimizer という。

Adam[8]は Optimizer の中でも移動平均で振動を抑制する Momentum と学習率を調整して振動を抑制する RMSProp を組み合わせ、改良したものである。

以下の計算式で重みパラメータの更新を行う。

$$g_t = \nabla_{\theta} f_t(\theta_{t-1})$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (14)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \widehat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad \theta_t = \theta_{t-1} - \alpha \frac{\widehat{m}_t}{\sqrt{\widehat{v}_t + \epsilon}} \quad (15)$$

(θ :重み, f_t :誤差関数, $\alpha=0.001$, $\beta_1=0.9$, $\beta_2=0.999$, $\epsilon = 10^{-8}$, $m_0, v_0 = 0$)

第 4 章 実験概要

本章では本研究の実験の概要及び、使用するデータ、評価関数、実行環境について述べる。

4.1 提案方法

予備実験として、第 5 章では過去の株価変化率のみを用いて時系列データの学習に適している LSTM モデルで株価変化率の予測を行う。第 6 章ではニュースデータを文章の学習に適している BERT モデルで株価変化率の予測を行う。ニュースデータはテキストデータのため、前処理をして学習モデルに適合させる。第 7 章では 2 つのモデルをアンサンブル学習し、予測を行う。複合モデルの予測の流れは図 4 のとおりである。

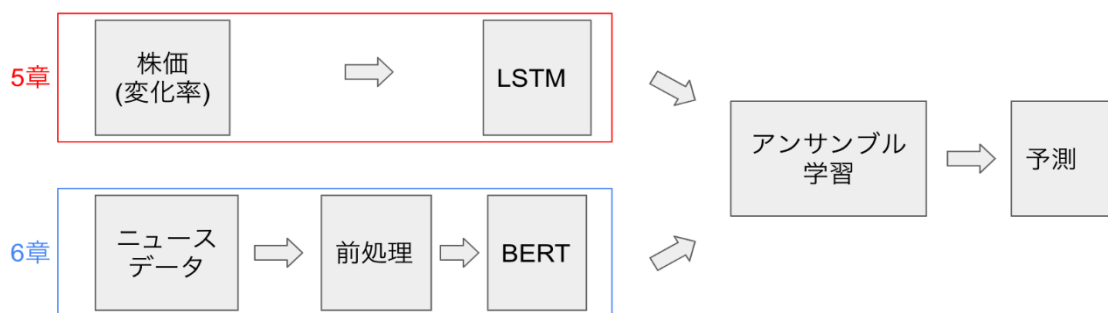


図 4 複合モデルの予測の流れ

4.2 使用データ

① 株価データ

2012 年 9 月 1 日から 2021 年 9 月 1 日までの日経平均株価終値 2,227 個のデータを用いる。日経平均株価は日本経済新聞社が東京証券取引所一部に上場している企業の中から選んだ 225 銘柄で構成されており、終値は取引時間終了時の取引価格である。

また、株価そのものでは今までにない変化が多く精度が落ちてしまう可能性が高いため、対数変化率を用いる。

対数変化率を用いる利点は大きく 2 点ある。1 点目は数値の変動幅が小さくなり、分散が小さく抑えられる点であり、2 点目は上昇時と下落時の変化率を等しく扱うことが出来

る点である.

時刻 t における対数変化率は以下のとおりである.

$$\log\left(\frac{S_t}{S_{t-1}}\right) = \log(S_t) - \log(S_{t-1}) \quad (16)$$

ただし, S_t は時刻 t における株価を表す.

② ニュースデータ

Reuters社のビジネスニュースを使用する. 対象期間は株価データ同様, 2012年9月1日から2021年9月1日までとする.

4.3 評価関数

本研究では予測精度を比較する際に以下の評価関数を用いる. 数値が低いほど誤差が小さく, 精度が良いといえる.

N はデータ数, y_i は実測値, $\hat{y}_i$ は予測値を表す.

① 二乗平均平方根誤差(RMSE : Root Mean Squared Error)

RMSEは以下の式で求められる.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (17)$$

② 平均絶対誤差(MAE : Mean Absolute Error)

MAEは以下の式で求められる.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (18)$$

4.4 実験環境

本研究における実行環境は表 1 のとおりである。

表 1 実験環境

CPU	Intel(R) Xeon(R) CPU × 2
OS	Microsoft Windows10 Home
Memory (CPU)	24.0GB
GPU	RTX3090
Memory (GPU)	24.0GB
使用言語	Python3.8.8

第 5 章 過去の株価変化率のみを用いた予測

本章では，過去の株価変化率のみを用いた予測モデルの概要と予測結果について述べる．

5.1 予測モデルの概要

LSTM モデルを用いて株価変化率の予測を行う．日経平均株価のデータを時系列順に 80%を学習データ(train)，20%をテストデータ(test)として使用する．

連続した 5 日分の株価変化率を入力値として与え，6 日目の株価変化率を予測値として出力する．Optimizer は Adam を利用し学習している．

学習に用いた各設定パラメータは表 2 のとおりであり，予測の流れは図 5 で示す．

表 2 LSTM モデルの各設定パラメータ

隠れ層	120
学習率 α	0.05
バッチサイズ	32
epoch 数	25



図 5 過去の株価変化率のみを用いた予測の流れ

5.2 予測結果

予測値に基づいた評価関数の値は表3に記す。

図6、図7のとおり RMSE, MAE に関しては train, test 共に epoch 数が増加し、学習が進むにつれ評価関数値の変動幅が小さくなり、収束に向かっている。これは学習の回数を重ねるにつれ予測値と実測値の誤差が小さくなっていることを示している。つまり、正しく学習が進んだと言える。また、RMSE の値が最小となった epoch 数は 17, MAE の値が最小となった epoch 数は 14 である。これ以上 epoch 数を増やしても最小値の更新が見込めなかったため、epoch25 まで予測を行った結果を記す。

表3 株価変化率を用いた予測の評価値

評価関数	最小値	最小値を記録した epoch 数
RMSE	0.01409	epoch 17
MAE	0.00982	epoch 14

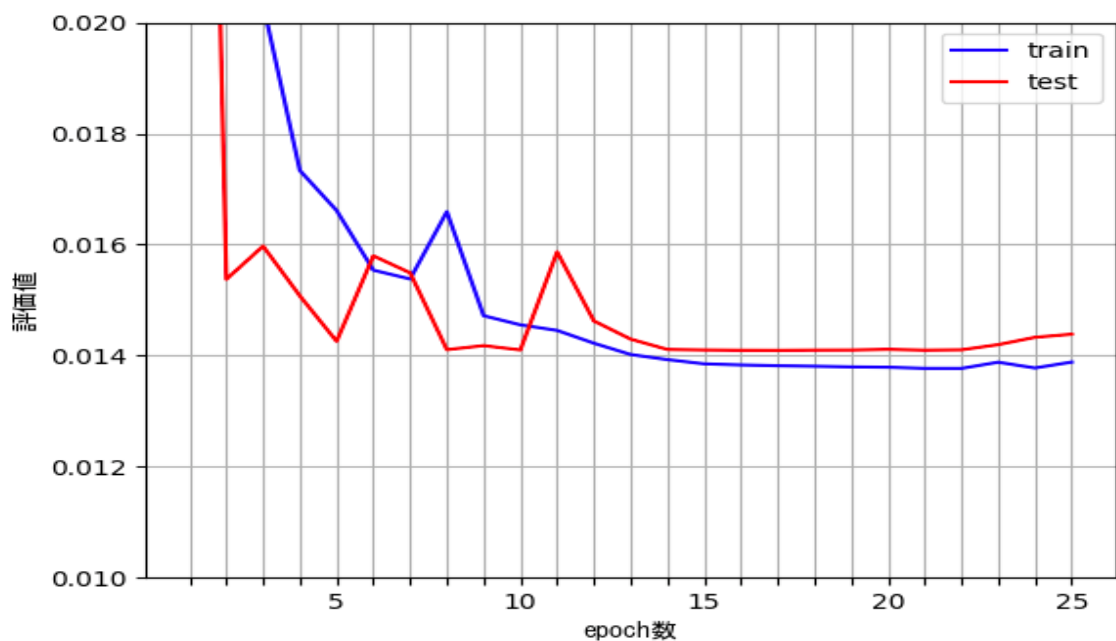


図6 株価変化率を用いた予測の RMSE

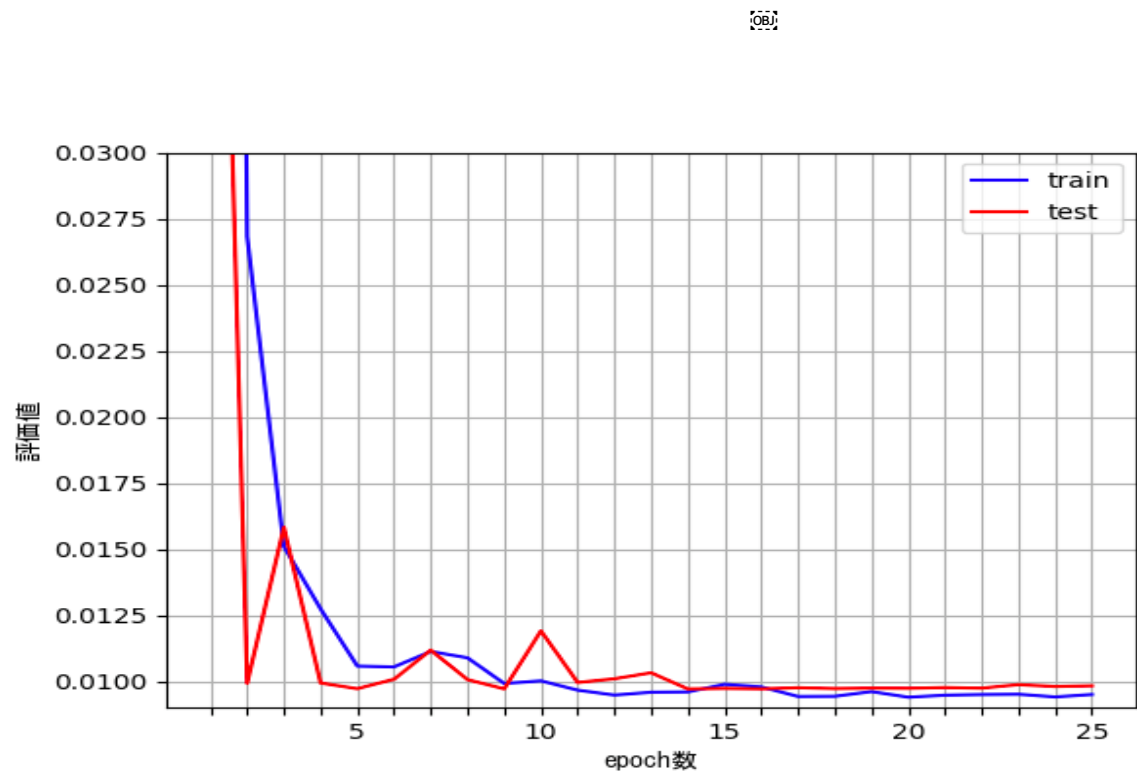


図 7 株価変化率を用いた予測の MAE

第 6 章 ニュースデータを利用した予測

本章では，ニュースデータを用いた予測モデルの概要と予測結果について述べる．

6.1 予測モデルの概要

BERT モデルを用いて株価変化率の予測を行う．ニュースデータの 80%を学習データ，20%をテストデータとして利用する．予測する日の前日のニュースデータのうち，ランダムで 1 つを選び，株価変化率の予測を行う．

ニュースデータでは日本語を利用するため，日本版 Wikipedia で事前学習している東北大学乾研究室の日本語 BERT モデル[9]を使用する．Optimizer は Adam を利用し学習している．

BERT モデルの Fine-Tuning における各種値は表 4 に記し，予測の流れは図 8 で示す．

表 4 BERT モデルの設定パラメータ

隠れ層の最終層の学習率 α	0.00005
出力層の学習率 β	0.00001
バッチサイズ	32
epoch 数	25

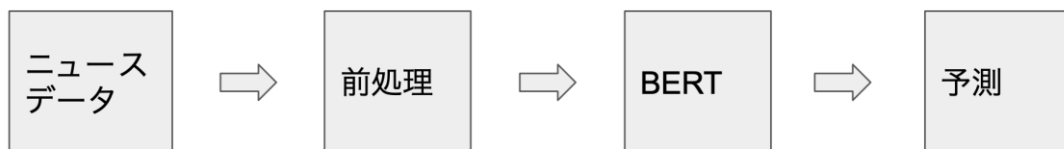


図 8 ニュースデータのみを用いた予測の流れ

6.2 前処理

ニュースデータは同日中に複数の記事があるため、1日1記事をランダムで抽出する。
前処理は以下の3つを行う。

①テキストの正規化では、半角カタカナを全角に、全角の記号や英単語を半角に変換した。

②テキストのクリーニングでは、記号、htmlタグ、ハイパーリンクなどを除く。

③ニュースデータの文章の形態素解析を行い、頻出でありかつ品詞が名詞の単語を除く。

名詞の頻出単語上位4つは表5で記す。

表5 名詞の頻出単語上位4つ

単語	出現頻度 (回)
日	1,760
月	1,597
年	1,402
ロイター	1,069

6.3 予測結果

予測値に基づく評価関数の値は表6で記す。

表6 ニュースデータを用いた予測の評価値

評価関数	最小値	最小値を記録した epoch 数
RMSE	0.01455	epoch 22
MAE	0.01153	epoch 21

図9、図10のとおりRMSE、MAE共に学習が進むにつれ、評価関数の値が小さくなり、変動の幅も小さくなっているため、学習がうまく進んでいると判断される。また、epoch22以降でRMSE、MAE共にtrainデータの値が小さくなっていることに対し、test

データの評価関数の値が増加していることが読み取れる。それは、これ以上学習を進めても精度が上がることはないということを示している。

epoch 数は 第 5 章の実験と同じ 25 としている。

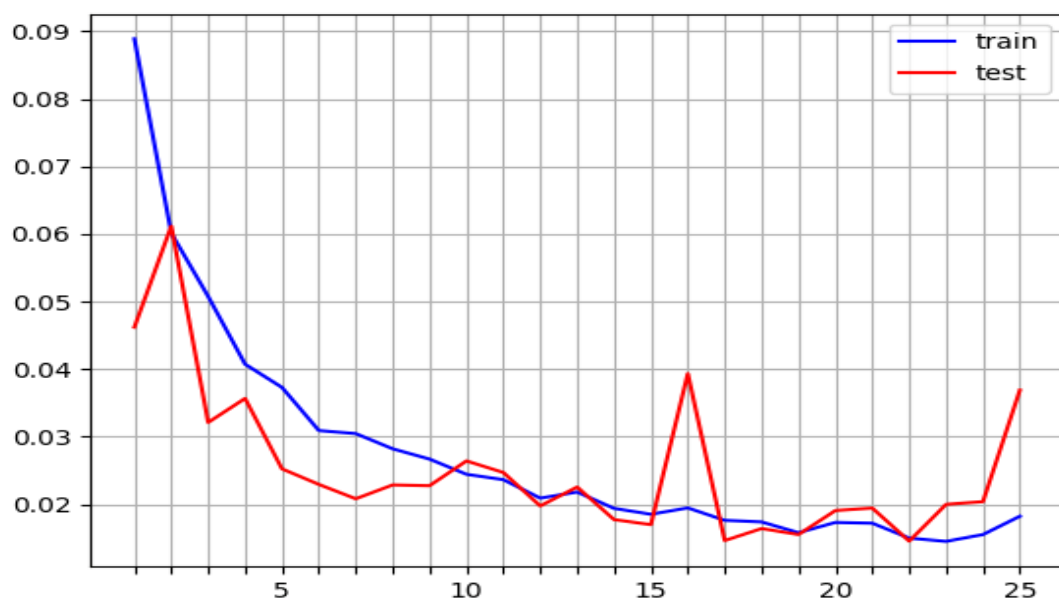


図 9 ニュースデータを用いた予測の RMSE

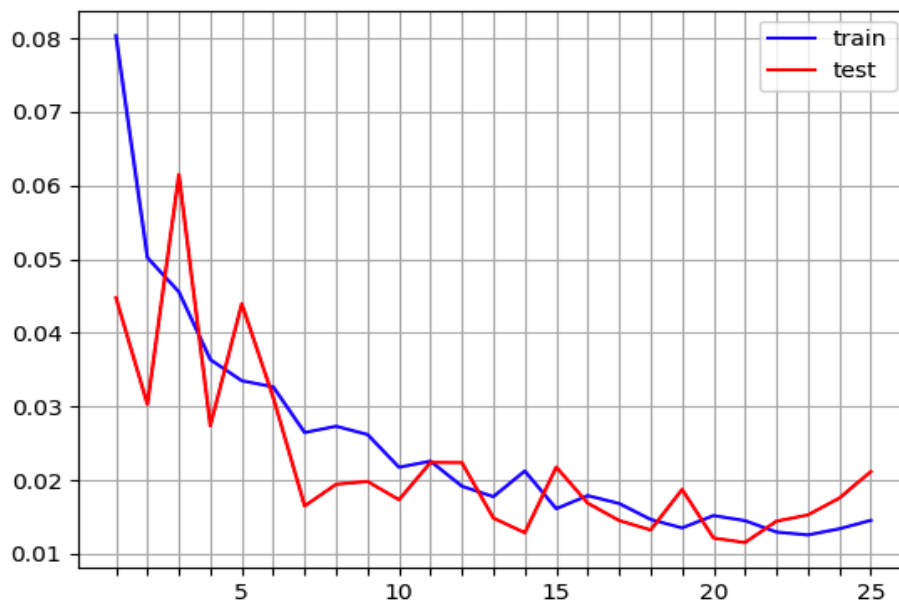


図 10 ニュースデータを用いた予測の MAE

第 7 章 2 つのモデルを組み合わせた予測

本章では、2 つのモデルを組み合わせた予測モデルの概要と予測結果について述べる。

7.1 予測モデルの概要

株価の値のみを LSTM モデルで予測した値とニュースデータを BERT モデルで予測をした値のデータを特徴量とし、アンサンブル学習のスタッキングの手法で予測を行う。メタモデルには勾配ブースティングを用いて予測を行う。勾配ブースティングのフレームワークとして LightGBM を使用する。以前のフレームワークと比較し LightGBM の特徴は計算時間が短くできる。工夫点は 2 点あり、1 点目は決定木の分岐が階層ごとに分岐させる Level-wise から葉ごとに分岐させる Leaf-wise にしている。そのため、全体の葉の数が少なくなる。2 点目は変数をヒストグラム化しているため、数値ごとに読み込むより時間を短縮できる。

LightGBM では、学習率は 0.0005 で学習させる。株価変化率とニュースデータを組み合わせた予測の流れは図 11 で示す。

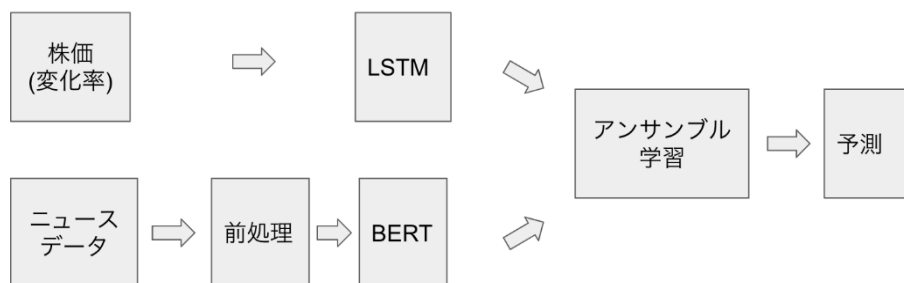


図 11 株価変化率とニュースデータを組み合わせた予測の流れ

7.2 予測結果

RMSE と MAE 共に表 7 の結果が得られた.

表 7 複合モデルで予測の評価値

評価関数	値
RMSE	0.01400
MAE	0.00970

評価関数の値は予備実験の 2 つの結果と比べ, 2 つモデルを組み合わせて予測を行う複合モデルはどちらも最も低い値となった.

第 8 章 実験結果の考察

本章では第 5 章，第 6 章，第 7 章それぞれの実験結果をもとに考察を述べる．

表 8 において各評価関数の値を比較すると，RMSE，MAE 共にごく僅かな差ではあるが過去の株価変化率とニュースデータの 2 つを用いた予測が最も精度が高い結果となった．

表 8 各予測における評価関数の値

	RMSE	MAE
過去の株価変化率のみ	0.01409	0.00982
ニュースデータのみ	0.01455	0.01153
複合モデル	0.01400	0.00970

表 9 および図 12 の過去の株価変化率のみを用いた予測は，予測値がすべて 0 以上 0.05 以下という小さい範囲で求められたということが分かる．つまり，下落に関しては一切予測をすることが出来なかった．しかし，評価関数が最も精度の高かった複合モデルの値に近いのは，表 9 と図 13 から読み取れるとおり，実測値における株価変化率の期待値が正かつ 0 に近い数値であるためと推測される．

表 9 各予測と実測値の統計量

	平均値	中央値	最大値	最小値	標準偏差
過去の株価変化率のみ	0.00153	0.00153	0.00259	0.00071	0.00026
ニュースデータのみ	-0.00381	-0.00403	0.01973	-0.01918	0.00621
複合モデル	0.00051	0.00044	0.00467	-0.0073	0.00140
実測値	0.00062	0.00069	0.07731	-0.06274	0.01408

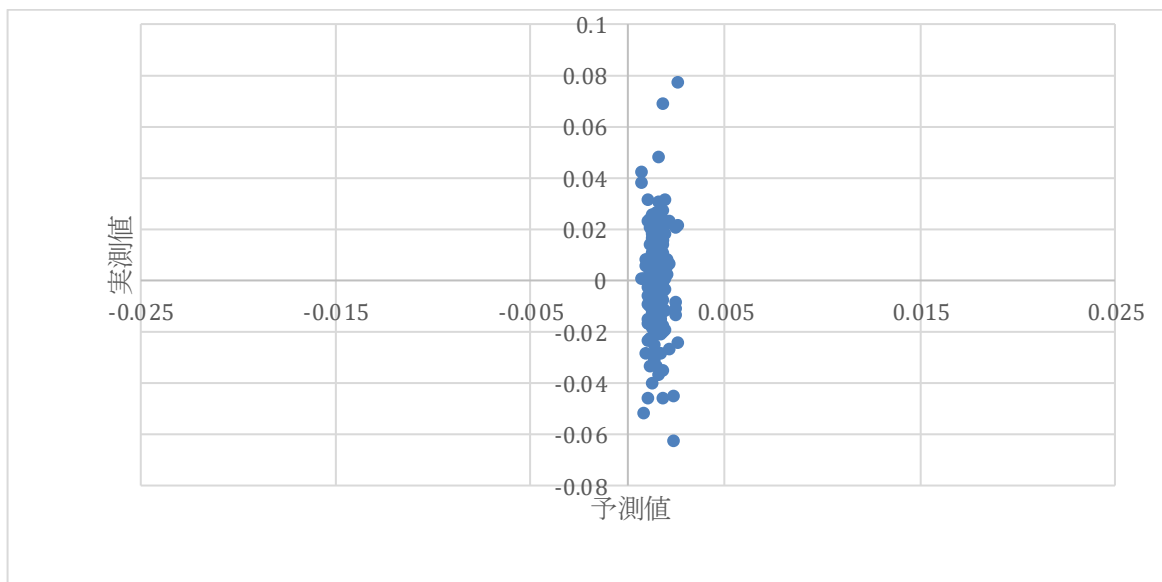


図 12 過去の株価変化率のみを用いた予測値

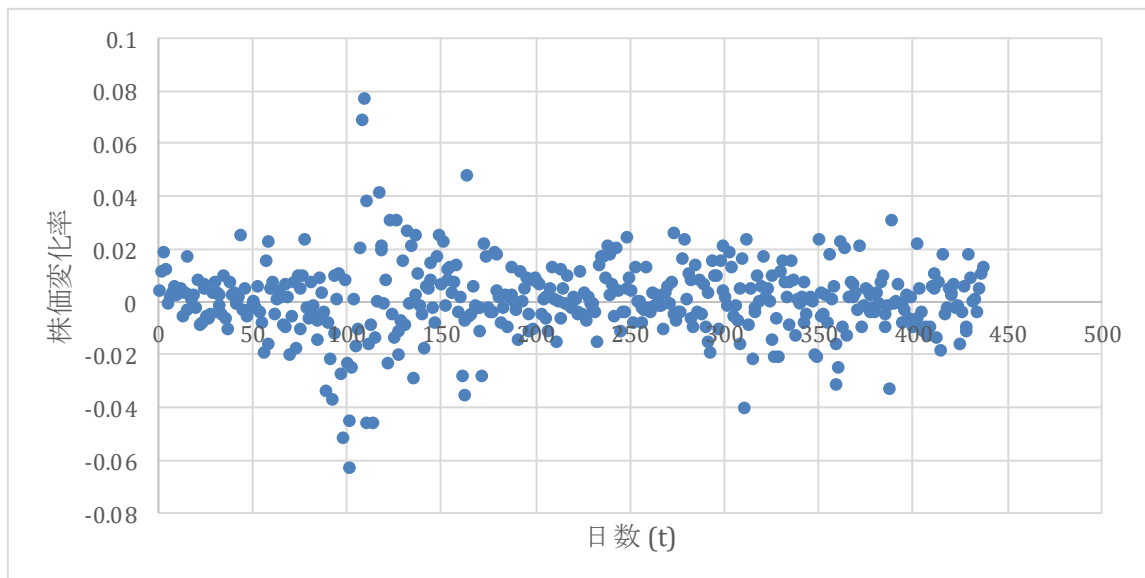


図 13 実測値の散布図

ニュースデータを用いた予測は、標準偏差が三つの予測モデルの中で最も大きく、図 14 からも正負問わず幅広く予測を行っていることが分かる。しかし、統計量を見ると平均値中央値共に負の値をとっている。これは、実測値において絶対値が 0.03 以上の値が負の値に多く、その影響を受けて予測値も負の値が多くなったと考えられる。

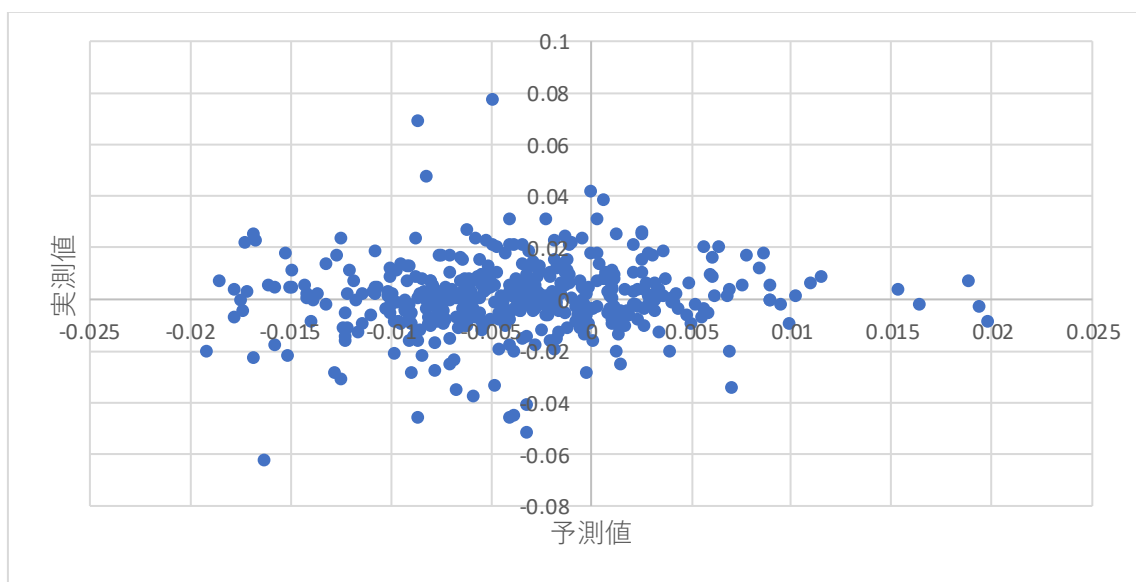


図 14 ニュースデータのみを用いた予測値

複合モデルは各評価関数の値が最も小さく、平均値と中央値は最も実測値に近い値を取った。図 15 から株価変化率のみを用いた予測と似た散布図の概形ながら下落に対しても予測が出来ているため、過去の株価変化率のみの予測とニュースデータのみの予測それぞれの特徴を捉え、予測精度を向上させることが出来たと言える。しかし、実測値の最大値は 0.07731、最小値は -0.06274 であるが、予測値の最大値は 0.00467、最小値は -0.00730 であり、標準偏差もかなりの差があるため、まだまだ実用的な精度であるとは言えない結果となった。

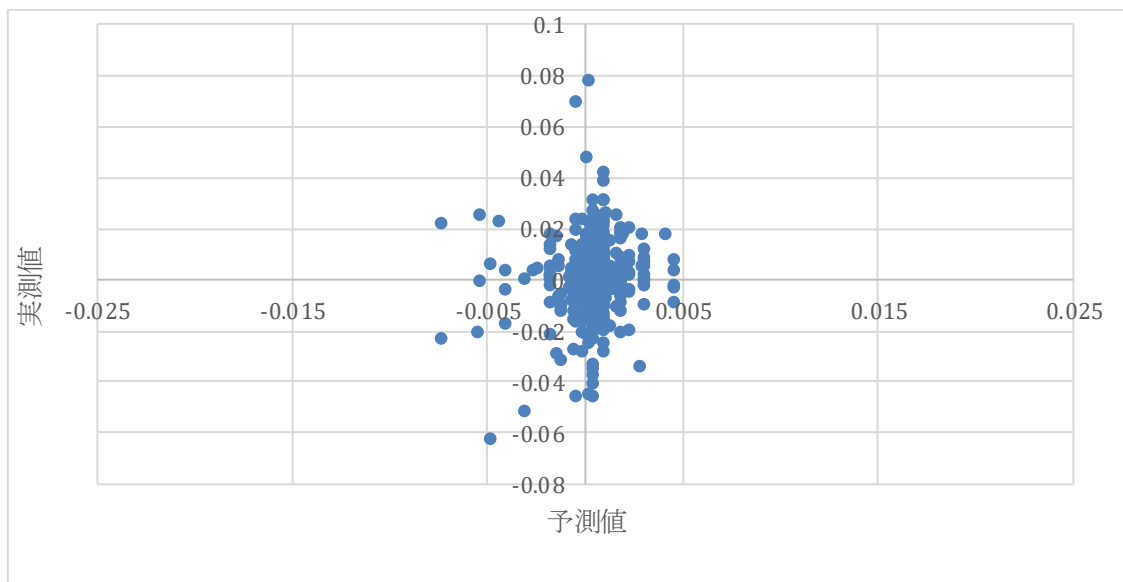


図 15 複合モデルの散布図

第9章 おわりに

本研究では、テクニカル分析だけを用いた株価、ファンダメンタルズ分析の両方の手法を用いることで、株価に影響する複雑なマーケット環境を考慮したよりよい予測が行える株価予測モデルを提案した。

第1章では研究背景と目的として以下の事を述べた。老後資金は自身で形成する必要があるといわれており、資産形成のための株取引数が増えている、このような一般投資家の意思決定を支援するため、さまざまな株価変動の予測手法が発表されている。

過去の研究では、株価数値の変動を用いた株価や変化率の予測、ニュースデータを基に株価が上昇するか下落するかの予測をするモデルの研究が発表されている。しかし、過去に株価変化率の予測と、ニュースデータの両方を用いた予測を組み合わせた研究はあまり発表されていない。過去の株価変化率から予測を行った場合は企業を取り巻く環境の変化から起こる急激な株価変動を予測できず、ニュースデータを用いて予測した場合、予測精度は落ちるものの急激な株価変動に対応した予測をするという特徴がある。そこで私たちは、両方のデータを用いて予測を行うことでそれぞれの欠点を補完し、現在の複雑なマーケット環境を考慮した、より精度の高い予測を行えると考えた。

上記から、それぞれ一つの指標から株価変化率の予測をするより予測精度の向上を図ることを目標として本研究に取り組んだ。第2章では、テクニカル分析とファンダメンタルズ分析それぞれの手法の先行研究を述べた。テクニカル分析を用いた先行研究と、ファンダメンタルズ分析の機械学習を用いた先行研究はそれぞれ散見されるが、テクニカル分析とファンダメンタルズ分析の両方を用いた先行研究は見受けられない。第3章では、本研究をするにあたり関連知識について述べた。第4章では実験概要として本研究の提案方法と使用データ、実験結果に対する評価関数、実験環境について述べた。予備実験として第5章では、予備実験として私たちは、株価変化率のみから LSTM モデルを用いて予測を行い、第6章では、ニュースデータのみから BERT モデルを用いて予測を行った。データ

は、2012年9月1日から2021年9月1日までの日経平均株価とReuters社のビジネスニュースを使用した。予測精度を比較するためRMSEとMAEを評価関数として用いた。それらの評価関数の値が0に近いほど予測値と実測値の誤差が小さく、精度が良いと言える。第7章では、これら2つ学習モデルの特徴量を学習データとしてアンサンブル学習による予測を行った。第8章では、実験の考察を述べた。予測結果はやはり、複合モデルが評価関数の値が最も0に近く、精度が良いという結果が得られた。過去の株価変化率のみを用いた予測値は範囲が小さく、ニュースデータのみを用いた予測値は範囲が大きいものの精度が芳しくなかった。複合モデルでは、その両方の特徴を捉え予測精度を向上させることが出来たと言える。

今回の実験結果から、過去の株価変化率とニュースデータの両方を用いた予測の有用性を示すことが出来た。様々な指標が考案されているため、今後は株価変化率だけでなく、他の指標同士を掛け合わせた予測の研究に本研究は役立てられると考えられる。

参考文献

- [1] G. Hinton, S. Osindero, and Y. W. Teh, : “A fast learning algorithm for deep belief nets, Neural Computation”, Vol.18, No.7, pp.1527–1554, 2006
- [2] 池戸鉄一, 林田実 : “ディープラーニングの株価予測への応用”, 北九州市立大学「商経論集」 vol.52, pp.13-26, March, 2017.
- [3] R. M. I. Kusuma, T. Ho, W. Kao Y. Ou, K. Hua : ”Using Deep Learning Neural Networks and Candlestick Chart Representation to Predict Stock Market” , arXiv:1903.12258, 13pages, 2019.
- [4] 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”, 全国知能学会全国大会, 2G4-OS-25a-4, vol.29, pp.1-3, June, 2015.
- [5] S. Hochreiter, and J. Schmidhuber : “Long Short-Term Memory”, Neural Computation, Vol.9, No.8, pp.1735-1780, 1997.
- [6] A. Vaswani, N. Shazeer , N. Parmar, J. Uszkoreit, L. Jones , A. N. Gomez, L. Kaiser, and I. Polosukhin : “Attention is all you need”, NIPS2017, pp.5998-6008, 2017.
- [7] J.Devlin, M.Chang, K.Lee, and K.Toutanova, : “BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding”, Proceedings of NAACL-HLT 2019, pp.4171-4186, June, 2019.
- [8] D. P. Kingma, and L. J. Ba : ”Adam : A method for Stochastic Optimization”, International Conference on Learning Representations , ICLR2015 , 13pages, 2015.
- [9] Pretrained Japanese BERT Models / 日本語 BERT 訓練済みモデル,
<https://github.com/cl-tohoku/bert-japanese>, アクセス日 : 2021-01-08.