

# ChatGPT による経済レポートを用いた株価推移予測の検討

## Examining Stock Market Trends Using Economic Reports Generated by ChatGPT

西澤 克彦<sup>\*1</sup>  
Katsuhiko NISHIZAWA

小屋松 裕馬<sup>\*1</sup>  
Yuma KOYAMATSU

小笠 真治<sup>\*2</sup>  
Shinji OGASA

<sup>\*1</sup> 同志社大学  
Doshisha University

<sup>\*2</sup> 香川高等専門学校  
Kagawa technical college #2

This study delves into machine learning for stock price forecasting, merging market sentiment with time-series stock data. Recognizing the impact of factors such as central bank statements and geopolitical risks on market sentiment, we supplement conventional data with economic reports generated via ChatGPT's web search function. This approach addresses the challenge of scarce historical sentiment data. We integrate these reports with stock price data, using BERT for encoding the combined dataset into latent variables. Future stock prices are then forecasted using Generative Adversarial Networks (GAN). Our methodology demonstrates the potential of including market psychology in stock price predictions, a significant advancement in financial modeling. This approach, which bypasses the need for historical economic reports, offers a novel perspective in understanding and forecasting market trends, highlighting the synergy between natural language processing and financial data analysis.

### 1. はじめに

株価予想における、機械学習の導入が近年注目されている[Mintarya 23]. 従来の株価予想は、株価の時系列データを基に将来の値動きを予測するものであった. しかし、昨今の自然言語処理の進化により、自然言語と株価の時系列データによるマルチモーダルな学習が可能となった. そこで、本研究では市場心理と株価の時系列データを統合した株価予測を検討する.

市場心理は、中央銀行の要人発言や、経済指標、地政学リスクなど様々な要因から構成され、株価に影響を与える. しかしながら、過去の市場心理を株価のヒストリカルデータのように取得し、データセットを構築するのは容易ではない.

そこで、本研究では ChatGPT の Web 検索機能(Browse with Bing)を用いて、特定の期間における経済レポートの作成を行った. その経済レポートと株価推移のデータを用いて、将来の株価予測を行う.

株価予測には、経済レポートと株価推移とを BERT によって、潜在変数に格納し、潜在変数から GAN を用いて将来の株価の生成を行う.

本論の主たる貢献は、過去の経済レポートを保有せずとも、市場心理を考慮した株価予測の学習を可能にしたことである.

### 2. 関連研究

#### 2.1 株価予測

従来、株価予測には古典的な機械学習手法である決定木や、SVM などの手法が用いられてきた. さらに近年では、深層学習を用いた手法が採用され、DQN などの深層強化学習を用いた手法も報告されている. その中でも、LSTM は時系列データの予測に強力なアルゴリズムである[Htun 23].

#### 2.2 GAN による株価の将来株価の生成

敵対的生成ネットワーク(GAN)を用いた株価予測は、カオス株価推移の予測に LSTM による単純な教師あり学習を上回るパフォーマンスを発揮するという報告がなされており、有効な手段である[Diqi 22] [Li 22] [Abraham 21].

#### 2.3 自然言語処理を用いた株価の予測

株価は市場心理によって推移することから、市場心理を株価予測に取り入れることで予測精度が向上すると報告されている. 市場心理の取得には、特定の会社のニュースなヘッドラインのセンチメント分析を自然言語処理によって、株価推移と組み合わせる方法が用いられている[Lopez-Lira 23] [Sonkiya 21] [Jadhav 21] [Zhang 21].

### 3. 提案手法

従来研究では、市場心理を株価推移に用いる方法の有効性が示されているが、過去の市場心理をデータセットとして取得することは容易ではない.

そこで、我々は ChatGPT の Web 検索機能を用いて過去の市場心理を表すレポートを作成する手法を提案する.

提案手法は、主に(1)ChatGPT の Web 検索機能による過去の市場心理を反映する経済レポートの取得と、(2)経済レポートと株価推移を統合したマルチモーダル学習によって構成される.

#### 3.1 ChatGPT による経済レポートの取得

ChatGPT の Web 検索機能(Browse with Bing)を用いて、下記の要件で経済レポートを作成した.

要件:

- 2018 年から 1 週間ごとのレポート  
(2019 年 25 週目の経済レポート)
- 米国, 日本, 欧州, 中国, 世界の 5 地域ごとに作成
- 1 地域あたり, 300word 以上を目安に作成
- 将来の情報を含まないように強く指示

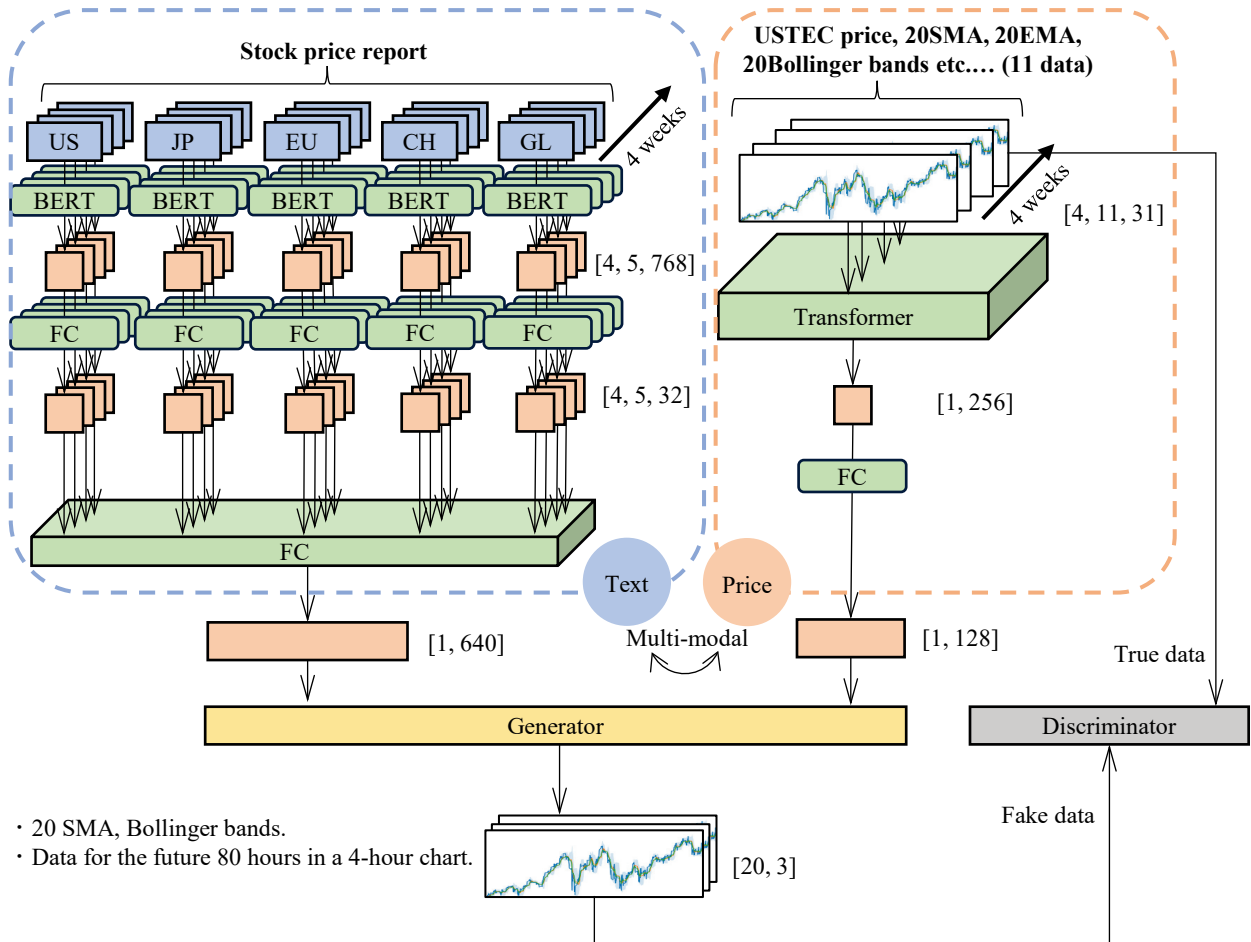


Fig. 1 Network architecture of the proposed method.

要件を踏まえたプロンプトの一例は、次のようである。

Prompt example:

You are a competent economic analyst. Please research Japanese Stock Market sentiment through the 16th week of 2017(from April 17 to April 21) and make a report. The report must include at least more than 300 words. If you include information from the following week, you will be punished.

### 3.2 経済レポートと株価推移によるマルチモーダル学習

提案手法では、3.1節で作成した経済レポートと、当該週の株価推移の4週間分の情報を入力データとして、翌週の4時間足の20本分(80時間分)の20 SMA とボリンジャーバンドの上限ならびに下限を生成する。

#### 3.2.1 入力データ

- 4週間分の経済レポート  
5地域×4週間分の20個のテキストデータ
- 4週間分の株価推移  
4時間足のOHLC(Open, High, Low, Close)  
テクニカルデータ(変化率:pct\_change, 対数変化率:log\_change, 20区間単純移動平均線:20SMA, 20区間指数平滑移動平均線:20EMA, 20区間変動率指標:

20ATR, 20区間ボリンジャーバンド上限:20Upper, 20区間ボリンジャーバンド下限:20Lower)

#### 3.2.2 生成データ

- 4時間足20本分の20SMAと対応するボリンジャーバンドの上限・下限

#### 3.2.3 アーキテクチャ

アーキテクチャは図1に示すように、

- ChatGPTで作成した経済レポートを、BERTを用いて、潜在変数に変換する。
- 4週間分の株価推移(4時間足)を、transformerを用いて、潜在変数に変換する。この時、4週間分の株価推移は、休場や短縮営業などの影響で可変長であるが、Transformer-encoderによって、固定長の潜在変数に変換される。
- 経済レポートの潜在変数と、株価推移の潜在変数を、全結合し、生成器によって4時間足20本分のデータを生成する。
- 生成器によって生成されたデータ、Discriminatorによって識別し、敵対的生成ネットワークによる学習を行う。

## 4. 実験

### 4.1 データセット

株価データとして、ナスダック 100 の株価指数(USTEC)を選択した。2018 年 1 月 2 日～2023 年 11 月 1 日の期間において、1 分足の OHLC データを取得し、各種テクニカル指標を計算した後、4 時間足でリサンプリングした。経済レポートに関しては、3.1 節に従って株価データと同期間の経済レポートを取得した。以上により、当該期間における 304 週分(4 時間足で 9211 件)のデータセットを作成した。また、4.2 節で述べる手法の学習時には、2018 年～2022 年のデータを訓練データとして用い、2022 年～2024 年のデータを検証データとして用いた。

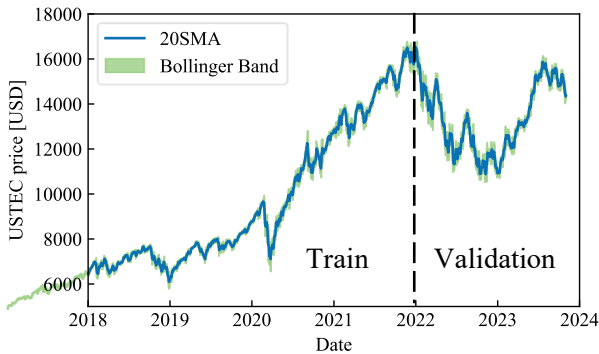


Fig. 2 Training data and validation data.

### 4.2 比較手法・評価指標

教師あり学習による手法と提案手法である GAN による学習手法とにおけるテクニカル指標の予測を比較した。評価はテクニカル指標毎に行った。評価指標として、二乗平均平方根誤差(Root mean squared error, RMSE)と平均絶対誤差(Mean absolute error, MAE)を用いた。

#### 4.2.1 教師あり学習

提案ネットワークにおいて Discriminator を排し、全結合層と Transformer encoder, Generator のパラメータを教師あり学習により更新した。損失関数として式(1)の平均二乗誤差を用いた。

$$\text{MSE} = \frac{1}{n} \sum (\hat{y}_i - y)^2 \quad (1)$$

学習はバッチサイズ 8 のミニバッチ学習を行い、エポック数は 40 とした。またパラメータの最適化関数に Adam [Kingma 14]を用い、学習率は  $1.0 \times 10^{-5}$  とした。

#### 4.2.2 GAN による学習(提案手法)

提案手法では、GAN による敵対的学習を実施するが、Discriminator 優位になる可能性を考慮して、Generator の事前学習を実施する。事前学習は 4.2.1 項と同様の教師あり学習を行う。

GAN による敵対的学習の際には、提案ネットワークの Generator 以外のパラメータを凍結させる。Generator と Discriminator は、式(2)の損失関数により最適化される。

$$\begin{aligned} \min_G \max_D V(D, G) \\ = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] \\ + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(x)))] \end{aligned} \quad (2)$$

ただし、Generator の入力である  $z$  は株価データと経済レポートの潜在表現である。学習のパラメータは 4.2.1 項と同様である。

### 4.3 結果と考察

表 1 に各手法における定量的な結果を示す。太字は良好な結果であることを示す。どのテクニカル指標においても、教師あり学習手法で誤差が小さい。提案手法は教師あり学習手法よりも劣っていると推察される。また、図 3 に教師あり学習手法における推定曲線、図 4 に提案手法における推定曲線を示す。明らかに学習に失敗していることが見て取れる。さらに、提案手法における 10 エポック毎の生成曲線を図 4 に示す。敵対的学習初期は推定対象の波形の特徴を捉えているが、終盤に至っては特徴を全く捉えていないことがわかる。この原因として、Generator と Discriminator の学習の様々なパラメータの調整が上手くいっていないことが考えられる。Generator と Discriminator のネットワーク構造を詳細に検討していくことや学習率の調整が望まれる。

なお、教師あり学習手法においては、年単位で大域的に見れば比較的高精度に株価が推定できている。しかし、80 時間単位での精度評価は未だにできていないため、推定の良し悪しは不明である。実際にトレードを行うために、短期間での精度評価を実施していく。

本研究は、ChatGPT の Web 検索機能と株価の時系列データを統合して株価を予測する新しいシステムの初期検討を行った点で大変有用であると考えられる。

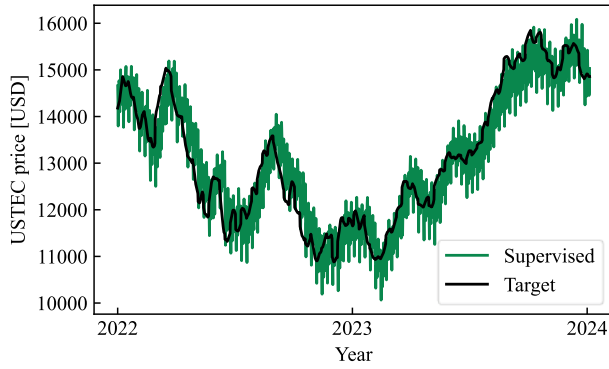
## 5. まとめ

本研究では市場心理と株価の時系列データを統合した新しい株価予測手法を提案した。この手法では、ChatGPT の Web 検索機能を活用して経済レポートを生成し、BERT と GAN を用いて株価予測を行う。実験結果は、教師あり学習手法に比べて提案手法が劣ることを示している。特に、提案手法は学習の終盤で株価の特徴を捉えることができなかった。これは、Generator と Discriminator のパラメータ調整に課題があることを示唆している。

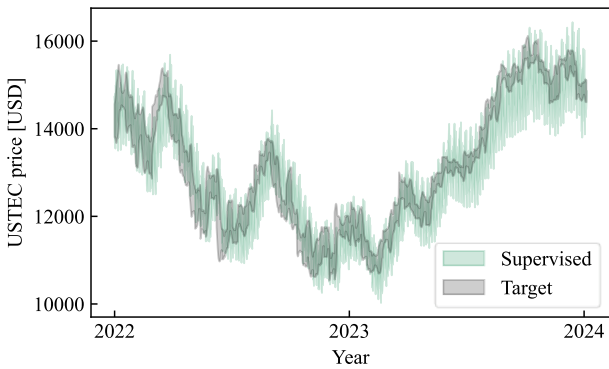
今後の研究では、これらのネットワーク構造の改善と学習パラメータの最適化が必要である。また、教師あり学習手法は長期的な視点では有効であるが、短期的な精度は未だ不明であるため、短期的な予測精度の向上にも注力すべきである。最終的には、市場心理を考慮した株価予測の実現に向けて、さらなる研究と改善が求められる。

Table 1 Quantitative results of supervised and proposed methods.

Method	Technical index	RMSE [USD]	MAE [USD]
Supervised	20SMA	<b>489.98</b>	<b>393.68</b>
	Upper	<b>533.28</b>	<b>428.93</b>
	Lower	<b>567.82</b>	<b>462.96</b>
Proposed	20SMA	60151.51	48314.23
	Upper	60374.13	50317.00
	Lower	43962.34	38499.93



(a) 20SMA.



(b) Bollinger bands.

Fig. 3 Prediction results of the supervised method.

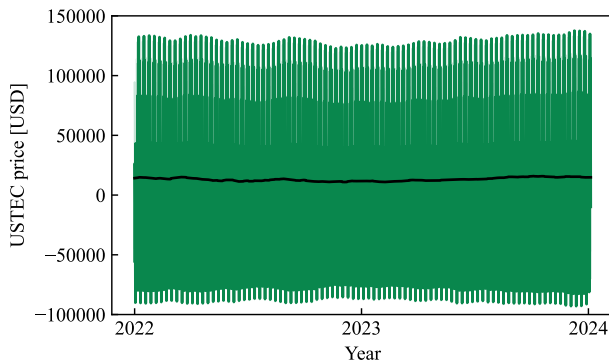
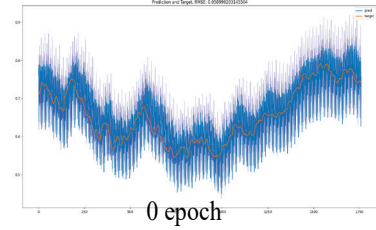
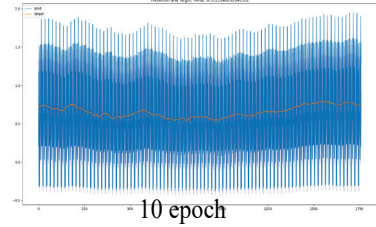


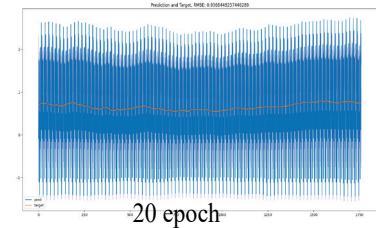
Fig. 4 Prediction results of the proposed method.



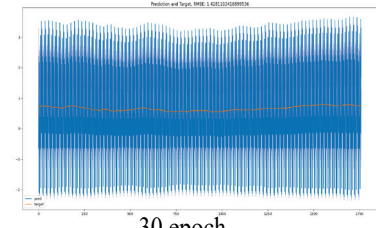
0 epoch



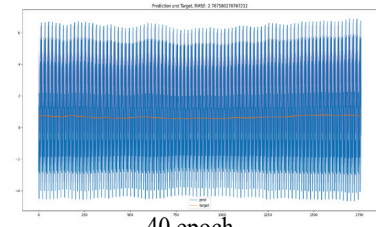
10 epoch



20 epoch



30 epoch



40 epoch

Fig. 5 Generated data result per epoch in the proposed method

## 参考文献

- [Kingma 14] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- [Mintarya 23] Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: a systematic literature review. *Procedia Computer Science*, 216, 96-102.
- [Htun 23] Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation*, 9(1), 26.
- [Diqu 22] Diqu, M., Hiswati, M. E., & Nur, A. S. (2022). StockGAN: robust stock price prediction using GAN

algorithm. *International Journal of Information Technology*, 14(5), 2309-2315.

- [Li 22] Li, Y., Cheng, D., Huang, X., & Li, C. (2022, January). Stock price prediction Based on Generative Adversarial Network. In *2022 international conference on big data, information and computer network (BDICN)* (pp. 637-641). IEEE.
- [Abraham 21] Abraham, J. B. (2021). Improving stock price prediction with GAN-based data augmentation. *Indonesian Journal of Artificial Intelligence and Data Mining*, 4(1), 1-10.
- [Lopez-Lira 23] Lopez-Lira, A., & Tang, Y. (2023). Can chatgpt forecast stock price movements? return predictability and large language models. arXiv preprint arXiv:2304.07619.

- [Sonkiya 21] Sonkiya, P., Bajpai, V., & Bansal, A. (2021). Stock price prediction using BERT and GAN. arXiv preprint arXiv:2107.09055.
- [Jadhav 21] Jadhav, R., Sinha, S., Wattamwar, S., & Kosamkar, P. (2021, October). Leveraging Market Sentiment for Stock Price Prediction using GAN. In 2021 2nd Global Conference for Advancement in Technology (GCAT) (pp. 1-6). IEEE.
- [Zhang 21] Zhang, Y., Li, J., Wang, H., & Choi, S. C. T. (2021). Sentiment-guided adversarial learning for stock price prediction. *Frontiers in Applied Mathematics and Statistics*, 7, 601105.