Headline Generation for Stock Price Fluctuation Articles

Shunsuke Nishida Yuki Zenimoto Xiaotian Wang
Takuya Tamura Takehito Utsuro
Degree Programs in Systems and Information Engineering,
Graduate School of Science and Technology, University of Tsukuba
{s2320778, s2220753, s2320811, s2120744}_@_u.tsukuba.ac.jp
utsuro_@_iit.tsukuba.ac.jp

Abstract

The purpose of this paper is to construct a model for the generation of sophisticated headlines pertaining to stock price fluctuation articles, derived from the articles' content. With respect to this headline generation objective, this paper solves three distinct tasks: in addition to the task of generating article headlines, two other tasks of extracting security names, and ascertaining the trajectory of stock prices, whether they are rising or declining. Regarding the headline generation task, we also revise the task as the model utilizes the outcomes of the security name extraction and rise/decline determination tasks, thereby for the purpose of preventing the inclusion of erroneous security names. We employed state-of-the-art pre-trained models from the field of natural language processing, fine-tuning these models for each task to enhance their precision. The dataset utilized for fine-tuning comprises a collection of articles delineating the rise and decline of stock prices. Consequently, we achieved remarkably high accuracy in the dual tasks of security name extraction and stock price rise or decline determination. For the headline generation task, a significant portion of the test data yielded fitting headlines.

1 Introduction

For individuals engaged in stock trading, acquiring up-to-date information regarding stock price fluctuations is highly important. Knowledge of not only whether stock prices have risen or declined, but also the underlying causes such as product launches or sociopolitical conditions, can inform future investment strategies. While news articles on stock trading serve as a primary source of information, manually creating articles for a diverse array of securities[1] is considered challenging due to time constraints.

Hence, it is desirable to construct a system capable of automatically generating stock price fluctuation articles using quantitative stock information and related textual data. Articles typically consist of a body and a headline, with the latter expected to succinctly include, at a minimum, the security name of the fluctuating stock and a term indicating whether the stock price has risen or declined.

This paper assumes that the body of a stock price fluctuation article has already been generated automatically and aims to develop a headline generation model that produces the headline based on the article's content (Figure 1)[2].

With respect to the headline generation model, which takes the article's content as input, we solve three distinct tasks: in addition to the task of generating article headlines, two other tasks of extracting the relevant security names, and determining stock price rise or decline. Regarding the headline generation task, we also revise the task as the model utilizes the outcomes of the security name extraction and rise/decline determination tasks, thereby for the purpose of preventing the inclusion of erroneous security names.

In each task, we employ pre-trained models that have demonstrated high performance in the field of natural language processing. By fine-tuning[3] these pre-trained models for each respective task, we aim to enhance the models' accuracy. The dataset used for fine-tuning consists of the stock price fluctuation article dataset, which is com-

[1]The term "security" is used for expressing the company of the stock.

[2]Although we assume that the headline generation model developed in this paper is to be applied to the body of automatically generated stock price fluctuation articles, in the evaluation of this paper, we report the results of applying the headline generation model to the body of manually written articles. Note that automatically generated articles may have some bias which may not be the case for manually written articles, where future works include studying issues arising from this difference of automatically generated and manually written articles.

[3]Using the weights of each layer in the pre-trained model as initial values, additional training is conducted with fine-tuning datasets to make subtle adjustments to the weights.

Proceedings of the Sixth Workshop on Financial Technology and Natural Language Processing, pages 22–30
November 1, 2023. ©2023 Association for Computational Linguistics

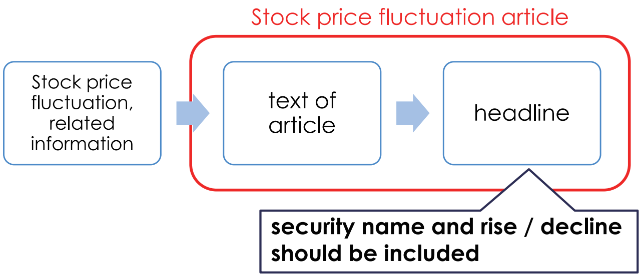

Figure 1: Overall Purpose of the Paper: Headline Generation for Stock Price Fluctuation Articles

posed of articles regarding the rise and decline of stock prices. In this paper, we report the results of the three tasks above as well as the preliminary evaluation results of the revised task of the headline generation task.

Our contributions are as follows:

1. We fine-tuned the pre-trained models XLM-RoBERTa (Conneau et al., 2020) and mT5 (Xue et al., 2021) using the stock price fluctuation article dataset to perform three tasks: generating article headlines, extracting target security names, and determining stock price rise and decline.

2. We developed the dataset for fine-tuning those three models, where the dataset comprises a collection of articles delineating the rise and decline of stock prices.

3. We were able to achieve quite high accuracy in the tasks of security name extraction and determining price rise and decline. For the headline generation task, appropriate headlines were generated for many test data, comparable to the actual article headlines. However, some generated headlines contained incorrect information, such as the target security names.

4. We revised the headline generation task to prevent headlines containing incorrect security names from being generated. Although we were able to reduce the number of headlines generated with incorrect security names while maintaining similar ROUGE (recall) scores as before the revision, the improvement was not significant.

2 Related Work

In the realm of research related to generating headlines from news article content, there exist preceding related work (Rush et al., 2015; Chopra et al., 2016; Nallapati et al., 2016; Kikuchi et al., 2016; Takase and Okazaki, 2019; Hitomi et al., 2019). Among them, Rush et al. (2015) developed the first approach to neural abstractive summarization. After that, Chopra et al. (2016) used the encoder-decoder framework, and Nallapati et al. (2016) incorporated additional features such as parts-of-speech tags and named entities. There also exist attempts to control the output length in neural abstractive summarization (Kikuchi et al., 2016; Takase and Okazaki, 2019; Hitomi et al., 2019). For example, in the works of Hitomi et al. (2019), they propose a corpus for evaluating headline generation models that take output length into consideration. In this paper, on the other hand, we focus on stock-specific terminology characteristic of stock-related news articles, striving to create headlines more suitable for stock price fluctuation articles.

Additionally, in the context of studies on news article headlines and stock prices, there exist several prior investigation. The proposed methods for stock price prediction using news headlines vary across different approaches. In one approach, it is suggested to combine news headlines with technical indicators to predict stock prices (Kalshani et al., 2020). Another approach is also proposed, which predicts the short-term movement of stock prices after financial news events using only the headlines of the news (Chen, 2021). In a third approach, they discussed the failure of the Efficient Market Hypothesis and proposed a project on stock trend prediction using news (Kalyani et al., 2016). Two other approaches evaluate different machine learning and deep learning methods, such as Support Vector Machines (SVM) and Long Short-term Memory (LSTM), to predict stock price movement using financial news (Liu et al., 2018; Gong et al., 2021).

3 Dataset

3.1 Stock Price News of "MINKABU"

In this paper, we utilized the web-based media platform minkabu.jp[4], which delivers news articles on finance, as the source for collecting news articles on stock price fluctuations to create our dataset. Within minkabu.jp, the distribution source "MINKABU" contains a substantial number of stock price fluctuation news articles. We collected 23,989 news articles[5] with "MINKABU" as the distribution source. Approximately 290,000 articles[6] are published with minkabu.jp as the distribution source, and it is estimated that around 81,200 of these articles pertain to stock price rise and decline (Tsutsumi and Utsuro, 2022). Therefore, it can be asserted that the scale is sufficiently ample for collecting stock price fluctuation articles.

3.2 The Procedure of Dataset Development

The method for creating the stock price fluctuation article dataset utilized in this paper is outlined below. In the creation of the stock price rise article dataset, we selected "individual words representing rise in stock prices", which are vocabulary indicating a stock price increase. We also selected "individual words representing cause of stock price rise", which are words used when explaining the reasons for a stock price increase. Total number of those words is 76 which include more than ten stock price domain specific words (Tsutsumi and Utsuro, 2022) such as "反発 (correction)", "続伸 (continued to rise)", "高値 (high price)", "カイ気配 (bid price)", "大幅高 (large rise)", "上昇 (rise)", "ストップ高 (hit limit high)", "急伸 (rise rapidly)", "連騰 (winning streak)", "堅調 (increase steadily)", "急騰 (sharp rise)", and "好感 (favorable)". Next, from 3,300 articles[7] by the distribution source "MINKABU" we extracted 2,734 articles containing either "individual words representing rise in stock prices" or "individual words representing cause of stock price rise" in the article body text. Furthermore, from these 2,734 articles, we excluded articles in which the company code did not appear in the text[8] and articles whose headlines had a low probability of being stock price rise news[9], resulting in the extraction of 1,185 articles with a high likelihood of being stock price rise news. After filtering, we manually created a dataset of 617 stock price rise articles out of the 1,185 articles.

For stock price decline articles, we selected "individual words representing decline in stock prices", which are vocabulary indicating a stock price decrease, and "individual words representing cause of stock price decline", which are words used when explaining the reasons for a stock price decrease. Total number of those words is 103 which also include more than ten stock price domain specific words (Tsutsumi and Utsuro, 2022) such as "嫌気 (unfavorable)", "反落 (reactionary fall)", "続落 (continued to decline)", "赤字 (deficit)", "急落 (fall rapidly)", "減益 (decrease in profit)", "出尽くし感 (material exhaustion)", "転落 (fall)", "下落 (decline)", "下振れ (downside)", "大幅安 (large decline)", and "引き下げ (reduction)". Next, from the 23,989 articles by the distribution source "MINKABU", we extracted 12,887 articles containing either "individual words representing decline in stock prices" or "individual words representing cause of stock price decline" in the article body text[10]. Furthermore, from these 12,887 articles, we excluded articles in which the company code did not appear in the text and articles whose headlines had a low probability of being stock price decline news, resulting in the extraction of 7,986 articles with a high likelihood of being stock price decline news. After filtering, we manually created a dataset of 777 stock price rise articles out of the 7,986 articles.

Those 76 stock price rise related words and 103 stock price decline related words have a certain overlap such as "発表 (announcement)", "影響 (influence)", "要因 (cause)", and "見通し (estimation)". There could be cases where articles may include both stock price rise and decline related words, or may include those overlapping words. Even in such cases, however, most articles usually report only either stock price rise or decline, but do not discuss the trend in change of the stock price

[4] https://minkabu.jp/

[5] 15,300 articles distributed from June 30, 2020, to December 3, 2020, and 8,689 articles distributed from March 5, 2021, to June 1, 2021.

[6] As of November 2021.

[7] Articles distributed from October 30, 2020, to December 3, 2020.

[8] Articles where ".T>" does not appear in the text.

[9] Articles whose title begins with "<". This is used in headlines for articles on foreign exchange, bonds, and individual investor trends.

[10] For both stock price rise and decline articles, those excluded articles mostly do not report stock price fluctuation.

over longer duration. We also manually exclude those case of exceptional article types.

4 Headline Generation Model

4.1 Overall Procedure

In this paper, with respect to the headline generation model, using the article's main text as input, we solve three distinct tasks: in addition to the task of generating article headlines, two other tasks of extracting the relevant security names, and determining stock price rise or decline.

For the stock price rise and decline judgment task and security name extraction task, we used the XLM-RoBERTa (base-sized model) (Conneau et al., 2020)[11] as the pre-trained model[12]. XLM-RoBERTa is a multilingual model pre-trained on Common-Crawl data[13] containing 100 languages[14].

For the headline generation task, we used the mT5 (small-sized model) (Xue et al., 2021)[15] as the pre-trained model[16]. mT5 is a multilingual model pre-trained on the mC4 corpus[17], which includes 101 languages[18].

For fine-tuning each pre-trained model, we used the stock price fluctuation article dataset mentioned in section 3. As a preprocessing step for this dataset, all security name codes[19] appearing in the dataset's context were removed to prevent easy identification of security name positions. After ensuring that the training, validation and test data did not contain articles about the same security name, we randomly divided the whole dataset into three parts of 200 validation examples, 200 test examples and the remaining 994 examples for training. Those divided datasets were used for all tasks (Table 1).

4.2 Rise and Decline Detection

For fine-tuning the pre-trained model for the stock price rise and decline judgment task, we used the Huggingface Transformers Text classification library[20]. For model fine-tuning[21], we used 994 training examples from the stock price fluctuation article dataset.

After fine-tuning, we inputted the 200 test examples from the stock price fluctuation article dataset to the model and tested its performance. We used accuracy as the performance evaluation metric[22]. The accuracy was 1.0, indicating that the model correctly determined the rise and decline for all 200 test examples.

4.3 Extraction of Security Names

For fine-tuning the pre-trained model for the security name extraction task, we used the Huggingface Transformers Question Answering library[23]. For model fine-tuning[24], we used 994 training examples from the stock price fluctuation article dataset. We set the question text for each example to be blank and the answer to be the security name covered in the article, and inputted the pair of answers and article body text to the model.

After fine-tuning, we inputted the 200 test examples from the stock price fluctuation article dataset to the model and tested its performance. We used exact match rate as the performance evaluation metric[25]. The exact match rate was 99.5%, indicating that the model could correctly extract security names for almost all of the 200 test examples.

4.4 Headline Generation

For fine-tuning the pre-trained model for the headline generation task, we used the Huggingface Transformers Summarization library[26]. For fine-tuning[27], we used 994 training examples from the stock price fluctuation article dataset. We treated

[11] https://huggingface.co/xlm-roberta-base

[12] XLM-RoBERTa is designed to be easy to apply to classification tasks and sequence labeling tasks.

[13] https://commoncrawl.org

[14] Including the Japanese language.

[15] https://huggingface.co/google/mt5-small

[16] mT5 is designed to be easy to apply to text to text conversion tasks including summarization and generation tasks.

[17] https://www.tensorflow.org/datasets/catalog/c4

[18] Including the Japanese language.

[19] Assigned to all listed companies, composed of "unique name code with 4 digits + 1 reserve code digit."

[20] https://github.com/huggingface/transformers/tree/v4.21.1/examples/pytorch/text-classification

[21] The learning rate was set to 0.00002, the batch size was set to 8, and the number of epochs is 10.

[22] Accuracy is given by (number of correctly classified data)/(number of data).

[23] https://github.com/huggingface/transformers/tree/v4.21.1/examples/pytorch/question-answering

[24] The learning rate was set to 0.00003, the batch size was set to 8, and the number of epochs is 2.

[25] The proportion of cases where the model's output string exactly matched the reference security name string.

[26] https://github.com/huggingface/transformers/tree/v4.21.1/examples/pytorch/summarization

[27] The learning rate was set to 0.00003, and the batch size was set to 8. The model for the minimum validation loss is selected.

	stock price rising articles	stock price declining articles	Total
training data	432	562	994
validation data	94	106	200
test data	91	109	200
total	617	777	1,394

Table 1: Number of Articles of Each Type

the article headlines as summaries for each data and inputted the pair of summary and article body text to the model.

After fine-tuning, we inputted the 200 test examples from the stock price fluctuation article dataset to the model and tested its performance. We used ROUGE (recall) as the performance evaluation metric. ROUGE (recall) measures the degree of match between the model-generated summary and the reference summary, where it is measured as the rate of the intersection of the model-generated summary and the reference summary over the reference summary. ROUGE-1 (recall) measures the match at the 1-gram (word) level, ROUGE-2 (recall) at the 2-gram (bi-gram) level, and ROUGE-L (recall) measures the match of the longest common sequence. As shown in the row of "before task revision" in Table 4, ROUGE-1 (recall) was 53.03, ROUGE-2 (recall) was 35.73, and ROUGE-L (recall) was 52.77[28].

Next, we conducted a manual evaluation of 100 out of the 200 headline generation results for the test data. The column of "before task revision" in Table 2 shows examples of model-generated headlines that were manually evaluated. We examined whether the information in the model-generated headlines corresponded to the information in the input context and found that 80 out of the 100 test examples contained relevant information. Of these 80 examples, 38 (47.5%) had a perfect string match, and 65 (81.3%) were appropriate as headlines.

Additionally, while all the 100 test examples contained correct information about stock price rise or decline, there were six cases where the security name was incorrect.

[28]Without fine-tuning, the accuracy of rise and decline detection was 0, and the exact match rate of extraction of security names was 45.5%, which was pretty low compared with after fine-tuning. ROUGE (recall) of headline generation was also pretty low, where ROUGE-1 (recall) was 2.56, ROUGE-2 (recall) was 0.29, and ROUGE-L (recall) was 2.60.

5 Revised Task of Headline Generation and its Preliminary Evaluation Results

In the headline generation model described in the previous section, high accuracy was achieved for the tasks of determining whether the stock price rose or declined and extracting the security name. However, for the headline generation task, the model generated titles containing incorrect information not found in the input context for 20 out of 100 test examples. Here, we focus on the fact that six of these examples had incorrect security names and aim to improve the headline generation task by inserting the security name and a tag ("r" or "d") representing whether the stock price rose or declined in the input context[29][30] to prevent headlines with incorrect security names from being generated.

For fine-tuning the pre-trained model for the revised headline generation task, we used the Huggingface Transformers Summarization library, as in the case of the headline generation task of section 4.4. The training data of the stock price fluctuation article dataset and the various parameters during training were also set to the same values as the previous headline generation task. We inputted the pair of the article headline and the body text with the security name and rise/decline tags added to the model.

After fine-tuning, we inputted the 200 test examples from the stock price fluctuation article dataset to the model and performed the same performance test as in the previous headline generation task. As shown in the row of "after task revision" in Table 4, ROUGE-1 (recall) was 57.22%,

[29]For example, "company-name-A is surging ..." → "r company-name-A company-name-A is surging ..."

[30]Both in the training and in the test, we insert the reference tag representing the correct information on whether rise or decline, as well as the reference security name. This is simply because the test performance presented in section 4.2 and section 4.3 is almost perfect. We also confirmed that, when we inserted a randomly chosen security name in this revised task, the generated headline was with an incorrect security name for 13% cases.

content of the article (partially omitted)	reference headline	model-generated headline
Sグループは続伸している。25日の取引終了後、株主優待制度の内容を変更すると発表しており、これが好感されているようだ。創業30周年記念優待制度の内容を継続し、21年12月末以降も100株以上保有者を対象に、保有株数と継続保有期間に応じて1000円から1万円分のオリジナルクオカードを贈呈するという。(= S Group continues to rise. After the close of trading on the 25th, they announced a change in their shareholder benefits program, which seems to be well-received. They will continue the 30th anniversary commemorative benefits program, and after December 2021, they will give original QUO cards worth 1,000 to 10,000 yen to shareholders who hold 100 or more shares, depending on the number of shares held and the continuous holding period.)	SGは続伸、創業30周年記念優待制度の内容を継続へ (= SG Continues to Rise, Continues 30th Anniversary Commemorative Benefits Program)	SGは続伸、株主優待制度の変更を好感 (= SG Continues to Rise, Favorably Received Changes to Shareholder Benefits Program) (Information corresponding to the content of the article is included in the model-generated headline, which is comparable to the reference headline.)
M社が大幅続伸している。きょう付の日本経済新聞朝刊で、「1千キロメートル離れた場所から複数のドローンをまとめて操作できるシステムを2021年度にも実用化する」と報じられており、これが好材料視されているようだ。. . . (= M Company has made significant gains. According to today's morning edition of the Nikkei Shimbun, it was reported that they "will commercialize a system that can control multiple drones from a distance of 1,000 kilometers by the end of fiscal 2021," which seems to be viewed as good news. . . .)	M社が大幅続伸、1000キロ先のドローン操作する技術を21年度にも実用化と報じられる (= M Company Surges, Reported to Commercialize 1000km Distant Drone Control Technology by FY 2021)	M社が大幅続伸、1千キロメートル離れた場所から複数のドローンをまとめて操作 (= M Company Surges, Control Multiple Drones from a Distance of 1,000 Kilometers) (Information corresponding to the content of the article is included in the model-generated headline, while the model-generated headline is not sufficient)
K社が一時ストップ高まで買われた。同社はきょう、画像処理検査エンジンの販売強化などを目的に12月1日から社長直轄のプロジェクトチームでの活動を開始すると発表しており、今後の展開などが期待されているようだ。同社はこれまでに培った技術やノウハウを生かし、さまざまな顧客ニーズに応えられる画像処理外観検査用のエンジンを開発し、従来の液晶向け以外の分野にも進出･拡販することで、収益の拡大と安定によって、収益性の高い事業体制を確立するとしている。(= K Company's stock temporarily hit the daily limit high. The company announced today that it will start a project team under the direct control of the president from December 1st, aiming to strengthen the sales of image processing inspection engines, among other goals. It seems that future developments are expected. The company plans to establish a highly profitable business structure by expanding and stabilizing revenues, by developing image processing exterior inspection engines that can respond to various customer needs using the technology and know-how they have cultivated so far, and expanding into fields other than traditional liquid crystal displays.)	K社は一時S高、画像処理検査エンジンの販売強化に向けた活動開始へ (= K Company Temporarily Hits Daily Limit High, Activities to Strengthen Sales of Image Processing Inspection Engines Begin)	K社が一時S高、画像処理検査エンジンの販売強化などを目的に事業展開 (= K Company Temporarily Hits S High, Business Expansion Aimed at Strengthening Sales of Image Processing Inspection Engines) (Information corresponding to the content of the article is included in the model-generated headline, while the model-generated headline is not sufficient. Especially, the financial terminology "daily limit high" is abbreviated as "S high".)
J社が急反落。23日付の日本経済新聞朝刊で、「世界の飛行機の3分の1、8600機は休暇シーズンの8月に入っても地上に留め置かれている。需要拡大を前提に投資してきたリース産業や金融商品もリスクにさらされ、新造機はキャンセルが相次ぐ」と報じられており、航空機リースを主力とする同社への警戒感が強まっているようだ。. . . (= J Company's stock is plummeting. According to the morning edition of the Nihon Keizai Shimbun on the 23rd, "One-third of the world's airplanes, or 8,600 aircraft, are still grounded even in August, the vacation season. The leasing industry and financial products that have invested based on demand expansion are exposed to risks, and new aircraft are being canceled one after another." It seems that vigilance is increasing towards the company, which mainly deals with aircraft leasing. . . .)	J社が急反落、眠る飛行機8600機と報じられ業績への影響を警戒 (= J Company's Stock Plummets, 8,600 Grounded Aircraft Reported, Increasing Caution on Earnings Impact)	J社が急反落、国内有力証券が急反落、国内有力機はキャンセルと報じられる (= J Company's Stock Plummets, Domestic Leading Securities Plunge, Domestic Leading Aircraft Reported to be Canceled) (Information irrelevant to the content of the article is included in the model-generated headline, such as "domestic leading securities".)

Table 2: Examples of Stock Price Fluctuation Article Content, Headlines, and Model-generated Headlines in the Test Data (Japanese text with English translation)

<table>
<tr><th rowspan="2">content of the article (partially omitted)</th><th rowspan="2">reference headline</th><th colspan="2">model-generated headline</th></tr>
<tr><th>before task revision</th><th>after task revision</th></tr>
<tr><td>Uホールディングスが続急伸している。12日の取引終了後、第2四半期累計(4 9月)連結決算を発表しており、売上高20億9300万円(前年同期比8.8%増)、営業利益3億9300万円(同22.0%増)、純利益2億5400万円(同7.7%増)となり、営業利益、純利益とも第2四半期における過去最高益更新したことが好感されている。. . . (= U Holdings continues to surge rapidly. After the close of trading on the 12th, the company announced its second quarter consolidated financial results (from April to September), with net sales of 2.093 billion yen (an increase of 8.8% year-on-year), operating profit of 393 million yen (an increase of 22.0% year-on-year), and net profit of 254 million yen (an increase of 7.7% year-on-year). The record-breaking operating profit and net profit in the second quarter are well-received.. . .)</td><td>U が続急伸、高収益サブスクモデルの早期確立で第2四半期は過去最高益更新 (= U continues to surge, setting record-high earnings in the second quarter with early establishment of high-profit subscription model)</td><td>BBHD が続急伸、第2四半期営業利益 22%増 (= BBHD continues to surge, 22% increase in second quarter operating profit)</td><td>U が続急伸、第2四半期営業利益は過去最高益更新 (= U continues to surge, updating record-high operating profit in the second quarter)
(The security name was incorrect before task modification, while it is generated correctly after task modification.)</td></tr>
<tr><td>T社が6連騰と上げ足を強め、25日移動平均線を大きく上に放れてきた。同社は樹脂封鎖装置を主力とする半導体製造装置メーカーで、コンプレッション型を中心に収益を伸ばしている。9日取引終了後に21年3月期業績予想の修正を発表、トップラインは計画ラインを減額したものの、営業利益は従来予想の20億円から23億1000万円(前期比2.8倍)に大幅増額しており、これを手掛かり材料に上値を見込んだ買いが継続している。. . . (= T Corp has strengthened its upward momentum with a six-day winning streak, breaking significantly above its 25-day moving average. The company, a semiconductor manufacturing equipment manufacturer specializing in resin sealing devices, has been expanding its revenues mainly through compression-type devices. After the close of trading on the 9th, the company announced a revision to its earnings forecast for the fiscal year ending March 2021. Although the top line was reduced from the initial plan, operating profit was significantly increased from the previous forecast of 2 billion yen to 2.31 billion yen (2.8 times the previous fiscal year), and this has served as a catalyst for continued buying with an upward outlook.. . .)</td><td>T社が6連騰と上げ足加速、21年3月期営業大幅増額で前期比2.8倍に (= T Corp's 6-day winning streak accelerates, with operating profit for the fiscal year ending March 2021 significantly increased to 2.8 times the previous fiscal year)</td><td>T社が6連騰、半導体製造装置の受注拡大で21年3月期営業利益予想を大幅増額 (= T Corp on a 6-day winning streak, significantly raising its operating profit forecast for the fiscal year ending March 2021 due to expanded orders for semiconductor manufacturing equipment)</td><td>T社T社が25日線を大きく上放れ、半導体製造装置の受注拡大で21年3月期営業利益予想を大幅増額 (= T Corp T Corp surges significantly above the 25-day line, significantly raising its operating profit forecast for the fiscal year ending March 2021 due to expanded orders for semiconductor manufacturing equipment)
(Before task modification, the security name was correctly identified, while after inserting the tag through revision, extra security name words “T Corp” were incorrectly generated.)</td></tr>
</table>

Table 3: Examples of Stock Price Fluctuation Article Content, Headlines, and Model-generated Headlines in the Test Data (before and after task revision, Japanese text with English translation)

ROUGE-2 (recall) was 36.94%, and ROUGE-L (recall) was 52.27%. Table 4 also compares the ROUGE values before and after task revision, where, although the ROUGE-1 (recall) value improved slightly after the revision, the ROUGE-2 (recall) and ROUGE-L (recall) values hardly changed.

Next, we manually evaluated the 100 headlines generated for the test data, which were previously evaluated in the headline generation task of section 4.4. Table 3 shows examples of differences in the model-generated headlines before and after the task revision based on the manual evaluation.

Upon examining whether the information in the generated headlines corresponded to the information in the input context, we found that 82 out of 100 examples contained relevant information. Out of these 82 examples, 30 (36.6%) had exact matches in terms of character strings, and 62 (75.6%) were considered appropriate headlines. In addition, while all the 100 examples contained correct information about stock price rise and decline, 5 cases had errors in the security names. Table 5 shows the results of the manual evaluation before and after the revision.

From the above, it can be seen that, by the task revision, the number of headlines generated with incorrect security names is reduced, while main-

	ROUGE-1	ROUGE-2	ROUGE-L
before task revision	53.03	35.73	52.77
after task revision	57.22	36.94	52.27

Table 4: ROUGE (recall) Scores Before and After Task Revision

evaluation point		before task revision	after task revision
(i)	Relevant information exists in the model-generated headline.	80	82
(ii)	(Out of (i)) The model-generated headline is appropriate as a headline.	65	62
(iii)	(Out of (ii)) The model-generated and reference headlines matched exactly.	38	30
(iv)	The model-generated headline contains correct information on stock price rise/decline.	100	100
(v)	The model-generated security name is correct (including appropriate abbreviations).	94	95

Table 5: Results of Manual Evaluation for 100 Test Examples (before and after task revision)

taining the same level of accuracy as before. However, it can be said that the improvement was not significant.

6 Conclusion

In this paper, we fine-tuned the pre-trained models XLM-RoBERTa and mT5 using the stock price fluctuation article dataset to perform three tasks: generating article headlines, extracting target security names, and determining stock price rise and decline. We constructed a stock price fluctuation article headline generation model.

We were able to achieve quite high accuracy in the tasks of security name extraction and determining stock price rise and decline. For the headline generation task, appropriate headlines were generated for many test data, comparable to the actual article headlines. However, some generated headlines contained incorrect information, such as the target security names. We revised the headline generation task to prevent headlines containing incorrect security names from being generated. Although we were able to reduce the number of headlines generated with incorrect security names while maintaining similar ROUGE (recall) scores as before the revision, the improvement was not significant.

As future work, we would like to investigate methods such as setting loss functions for each of the tasks of security name extraction, stock price rise and decline determination, and headline generation, and performing multi-task learning to ensure that the security name information in the headlines is correct. This is to clarify whether the multi-task learning approach can improve the performance of the method employed in this paper, which is quite specific to the financial domain compared with the general language models studied in Radford et al. (2019). Another future work is to apply Large Language Models (LLMs) such as ChatGPT (OpenAI, 2023) and LLaMA (Touvron et al., 2023) to the task studied in this paper.

References

Qinkai Chen. 2021. Stock movement prediction with financial news using contextualized embedding from BERT. http://arxiv.org/abs/2107.08721.

Sumit Chopra, Michael Auli, and Alexander M. Rush. 2016. Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 93–98, San Diego, California. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Asso-*

ciation for Computational Linguistics, pages 8440–8451, Online. Association for Computational Linguistics.

Jiaying Gong, Bradley Paye, Gregory Kadlec, and Hoda Eldardiry. 2021. Predicting stock price movement using financial news sentiment. In *Proceedings of the 22nd Engineering Applications of Neural Networks Conference*, pages 503–517, Cham. Springer International Publishing.

Yuta Hitomi, Yuya Taguchi, Hideaki Tamori, Ko Kikuta, Jiro Nishitoba, Naoaki Okazaki, Kentaro Inui, and Manabu Okumura. 2019. A large-scale multi-length headline corpus for analyzing length-constrained headline generation model evaluation. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 333–343, Tokyo, Japan. Association for Computational Linguistics.

Ali Hassanzadeh Kalshani, Ahmad Razavi, and Reza Asadi. 2020. Stock market prediction using daily news headlines. https://ssrn.com/abstract=3685530.

Joshi Kalyani, H. N. Bharathi, and Rao Jyothi. 2016. Stock trend prediction using news sentiment analysis. http://arxiv.org/abs/1607.01958.

Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. 2016. Controlling output length in neural encoder-decoders. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1328–1338, Austin, Texas. Association for Computational Linguistics.

Yang Liu, Qingguo Zeng, Huanrui Yang, and Adrian Carrio. 2018. Stock price movement prediction from financial news with deep learning and knowledge graph embedding. In *Knowledge Management and Acquisition for Intelligent Systems*, pages 102–113, Cham. Springer International Publishing.

Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Çağlar Gu?lçehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence RNNs and beyond. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pages 280–290, Berlin, Germany. Association for Computational Linguistics.

OpenAI. 2023. GPT-4 technical report. http://arxiv.org/abs/2303.08774.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.

Sho Takase and Naoaki Okazaki. 2019. Positional encoding to control output sequence length. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3999–4004, Minneapolis, Minnesota. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothee Lacroix, Baptiste Roziere, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and efficient foundation language models. http://arxiv.org/abs/2302.13971.

Gakuto Tsutsumi and Takehito Utsuro. 2022. Detecting causes of stock price rise and decline by machine reading comprehension with BERT. In *Proceedings of the 4th Financial Narrative Processing Workshop @LREC2022*, pages 27–35, Marseille, France. European Language Resources Association.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.