

Economic Impact Analysis of News Articles Based on Polarity Analysis of News

Impact research of COVID-19 using market data

1st Zen Kato

Dept. of Applied Computer Engineering
Tokai University
Hiratuka, Japan
shin.lab@outlook.com

2nd Sanggyu Shin

Dept. of Applied Computer Engineering
Tokai University
Hiratuka, Japan
shin@tsc.u-tokai.ac.jp

Abstract—In this paper, we analyzed the relationship between news articles and market data (i.e., stock prices) which are expected to be affected by the contents of the articles. We analyzed news articles on the travel and tourism industry, which are thought to be strongly influenced by COVID-19, using polarity classification, and predicted their correlation. It is highly likely that the number of people infected by COVID-19, the government's countermeasures against infection, and the development of vaccines are influencing investors' investment behavior.

Index Terms—polarity analysis, news article, BERT model, COVID-19

I. INTRODUCTION

The new coronavirus (COVID-19), which has been prevalent worldwide since early 2020, has been an economic impact. In particular, in the accommodation and tourism service industries premised on interpersonal contact work, business performance deteriorated worldwide.

In Japan, the number of foreign visitors to Japan has decreased significantly due to the strengthening of border measures. In March 2020, the number of foreign visitors to Japan decreased by 93.0% from the same month of the previous year, and in April of the same year, it decreased by 99.9% from the same month of the last year, the most significant decrease in a single month since the start of statistics. As a result of the decrease in the number of foreign tourists visiting Japan, the consumption of foreign tourists visiting Japan in the January-March quarter of the same year decreased significantly to 41.6% compared to the same period of the previous year. In addition, Japanese domestic travel consumption decreased 53.1% in March of the same year compared to the same month of the previous year [1].

It is considered that general investors may refer to newspaper articles or news on the web for their investment behavior in addition to numerical information such as stock prices. Therefore, we did assume that the news about the number of people infected with COVID-19 and the government's countermeasures against infection will influence investors' investment behavior.

Already, related with COVID-19, much research occurred in the finance field [2] [3] [4] and data science field [5] [6].

In this study, we analyzed the relationship between news articles and the stock price movement of the travel and tourism industry. Anyone could expect the COVID-19 to be strongly affected the travel and tourism industries. Therefore, we expected these companies' stock prices to be affected by the content of the articles. Based on the analyzed results, we investigated the correlation between the economic impact of COVID-19 and the news articles.

The analysis targets were JAL (Japan Airlines), ANA (All Nippon Airways), H.I.S. Co., Ltd. (Travel agency company), and KNT-CT Holdings Co., Ltd. (Travel industry company), which provide tourism services listed on the Tokyo Stock Exchange. Companies with a small market capitalization may be less susceptible to news in addition to the problem of low liquidity. Still, we predicted that using stocks listed on the Tokyo Stock Exchange would solve such issues.

II. CONSTRUCTION OF POLARITY ASSESSMENT SYSTEM

A. BERT model transfer learning

D. Katayama and others developed a sentiment polarity identification model for finance using financial and economic corpus and deep learning [7]. The research by M. Costola and others, they investigated COVID-19 news, elaborated with the "Natural Language Toolkit" that uses machine learning models to extract the news' sentiment [8]. Also, N. J. Gormsen, et al. researched the economic impact of COVID-19 by about use data from aggregate stock and dividend futures markets to quantify how investors' expectations about economic growth evolved across horizons following the outbreak of the novel coronavirus (COVID-19) and subsequent policy responses until July 2020 [9].

We referenced these researches, developed a polarity model using Japanese documents on finance and the economy. The articles' polarity was classified by BERT's transfer learning (Bidirectional Encoder Representations from Transformers) [10]. The BERT is a natural language processing model that

TABLE I
DI (DIFFUSION INDEX) COMPUTATION METHOD

Assessment	Better Will get better (Good)	Slightly better Will get slightly better (Slightly good)	Unchanged Will remain unchanged (Neither good nor bad)	Slightly worse Will get slightly worse (Slightly bad)	Worse Will get worse (Bad)
Point	+1	+0.75	+0.5	+0.25	0
	⊕	○	□	▲	×

learns the relationship between sentences and a model that uses a two-way transformer. The JUMAN++ has been used for morphology the sentences [11].

As a pre-trained model for BERT, we used a pre-trained Japanese language model published by Kurohashi & Kawahara labs at Kyoto University [12]. This model was a transfer-learning model using data from 18 million articles on Wikipedia. The JUMAN++ divided input sentence into morpheme units and divided into sub-words by referring to the vocabulary list.

B. Data

As datasets for BERT’s transfer learning, we used the results of the Economic Watchers Survey, which is researched and published by the Cabinet Office [13], and the chABSA dataset that is TIS Corporation publishes for sentiment analysis using machine learning [14].

The Cabinet Office, the government of Japan, publishes the Economy Watchers Survey. The purpose of the survey is to gain an accurate grasp of region-by-region economic trends promptly. It is updated monthly to grasp regional economic trends and to survey economic trends. It has consisted of a five-level assessment of current and future economic conditions and the comment of assessment reasons.

On the Economic Watchers Survey, the data analyzed by DI (Diffusion Index) computation method like as Table I. Each points are assigned in accordance with the one-to-five scale for each category of response, and the DI is calculated by multiplying the points by the component ratio of each response category.

The period was from January 2000 to April 2020. The chABSA dataset is a dataset created based on the annual securities reports of listed companies. From each sentence, a negative/positive sentiment classification is made for a specific subject.

C. Data preprocessing

The current status judgments (five-level assessment) from the Economic Watchers Survey were dichotomized into positive and negative values and used as educational data for BERT. Specifically, we replaced the top two ratings (⊕ and ○) with “1” and the bottom two (▲ and ×) with “0”. We simplified the classification by excluding the middle rating (□).

The sentences, targets, and polarity classifications in the chABSA dataset were adding “1” if a sentence contained a “positive” polarity classification and added “-1” if it had a

“negative” polarity classification. The polarity of the sentence, positive or negative, is decided by completing these resulting ratings. In this study, we simplified weighting to “±1” by using polarity classification. Still, it is necessary to analyze the content of the articles in detail and subdivide the weights in future work.

As a result of these processes, we obtained 267894 samples from combined all data. Of the data obtained, 80% was used for training data, and the remaining 20% was split in half for test and validation. Table II shows the number of divisions of “training”, “test”, and “validation.”

TABLE II
NUMBER OF DATA DIVISIONS

Training	Test	Validation
214315	26790	26789

III. STOCK PRICE DATA AND NEWS ARTICLES

We obtained the four companies’ stock price data from January to June 30, 2020, using an API provided by “Investing.com.”

We collected the news data based on articles published on the Asahi Shimbun website. The Asahi Shimbun is one of the major newspapers in Japan and publishes much general news other than economic news. The analysis period was from January 1 to June 30, 2020.

This period includes the period from before the outbreak of COVID-19 to lift the declaration of a state of emergency. We have predicted that the travel and tourism industries most had received the impact of COVID-19 during this period. The Asahi Shimbun delivered 16685 news articles that related to our study during this period.

The Tokyo Stock Exchange has trading hours from 9:00 to 15:00. Therefore, we supposed the effects of news after 15:00 are reflected on the next day. So, we counted the news data as one day from 15:00 to 15:00 of the next day. Fig. 1 shows the number of articles per week.

IV. EXPERIMENT

We separated the entire article by line breaks for news data, and the polarity was evaluated by BERT each line. As a result of the polarity evaluation, we classified all tokenized sentences into either positive or negative, including the probability data. The following equation calculated the sentiment of the whole article [8].

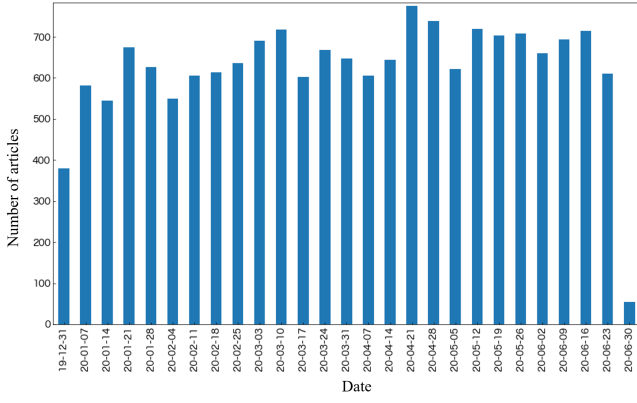


Fig. 1. Number of articles per week.

$$Sentiment_a = \frac{1}{n} \sum_{i=1}^n (P_i - N_i) \quad (1)$$

P_i and N_i indicate the probability that the sentence i contained in article a is positive or negative, including the respective. n is the number of sentences contained in the article. The overall sentiment of an article was determined by averaging the individual sentiments of each sentence contained in it.

Fig. 2 shows a 7-day moving average of sentiment over time. The sentiment is negative in all ranges, but it stopped declining as of May and has been on an upward trend since then. This result may be due to the announcement of an extension of a state of emergency in early May and the lift declaration in late May.

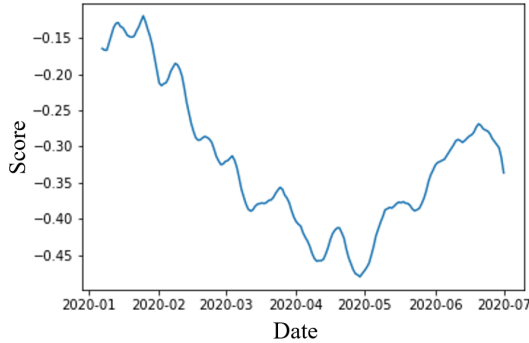


Fig. 2. 7-day moving average of sentiment.

In order to investigate the correlation between the 7-day moving average of sentiment and the closing price of the stock under investigation, we calculated Spearman's rank correlation coefficient by the following equation.

$$\rho = 1 - \frac{6 \sum D^2}{N^3 - N} \quad (2)$$

Spearman's rank correlation coefficient is a technique that can be used to summarize the strength and direction (negative or positive) of a relationship between two variables. We

evaluated the monotonic relationship between two variables based on the rank value of each correlated variable. We calculated the sentiment for two types of data: one using all news data and the other filtering out news containing "corona" and "covid."

V. RESULT

Fig. 3 shows a plot of the sentiment score and closing stock price. As shown in Table III, the correlation coefficient and P-value between the two values showed a significantly strong positive correlation between the sentiment calculated for the entire news and the closing stock price. This result is consistent with the hypothesis that an increase in positive news positively impacts stock prices.

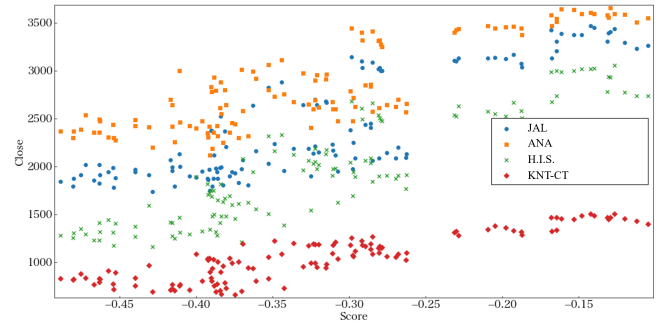


Fig. 3. Overall news and stock prices.

TABLE III
CORRELATION BETWEEN SENTIMENT AND STOCK PRICES
(ALL NEWS DATA)

Company	Correlation coefficient	ρ -value
JAL	0.832	5.473e-32
ANA	0.802	3.365e-28
H.L.S.	0.890	5.679e-42
KNT-CT	0.847	4.081e-34

As shown in Fig. 4, there was a negative correlation between the sentiment score calculated only from news related to corona and the stock price. The correlation coefficient and P-value are shown in Table IV.

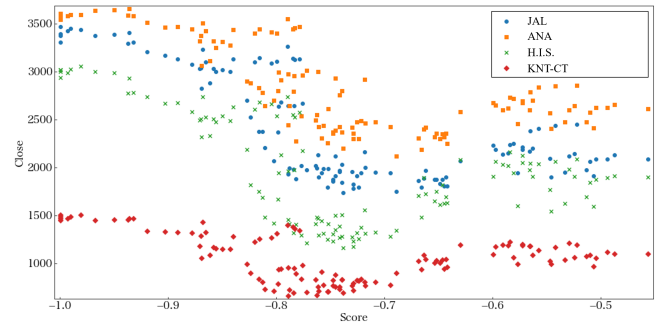


Fig. 4. News and stock prices related to COVID-19.

TABLE IV
CORRELATION BETWEEN SENTIMENT AND STOCK PRICES
(RELATED TO COVID-19)

Company	Correlation coefficient	p -value
JAL	-0.6352	6.596e-15
ANA	-0.6737	3.404e-17
H.I.S.	-0.4749	4.259e-08
KNT-CT	-0.2839	1.676e-03

VI. CONCLUSION

In this paper, we investigated the economic impact of COVID-19 on the stock market, especially on the tourism service industry. We have developed a model for identifying news sentiment using a machine learning model for 16685 online articles published in the Asahi Shimbun between January and June 2020.

As a result, it was found that there is a significant and strong positive relationship between news and movement of stock prices. It was confirmed that an increase in the sentiment score meant an increase in positive news, and an increase in positive news had a positive effect on the stock price.

However, there was a negative correlation between the sentiment score of articles related to COVID-19 and the stock price of tourism service companies. This result may be because the model wrong interpreted negative factors such as “increasing trend” and “exceeding” in the articles as favorable as if they were financial terms like rising the stock price.

As a future task, we will consider developing a model that understands the context of each industry and calculates the sentiment score, such as positive factors for one sector such as “weak yen” and “strong yen,” which can be damaging for other industries.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP 19K03006 and Transcosmos foundation.

REFERENCES

- [1] JNTO (Japan National Tourism Organization), Statistics of foreign visitors to Japan (press release). https://www.jnto.go.jp/jpn/statistics/data_info_listing/index.html (Japanese).
- [2] S. R. Baker et al., “The Unprecedented Stock Market Impact of COVID-19,” NBER Working Paper, no. 26945, June 2020.
- [3] A. M. Al-Awadhi, K. Alsaifi, A. Al-Awadhi and S. Alhammadi, “Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns,” *Journal of Behavioral and Experimental Finance*, vol. 27, September 2020.
- [4] M. Mazur, M. Dang and M. Vega, “COVID-19 and the march 2020 stock market crash. Evidence from S&P1500,” *Finance Research Letters*, vol. 38, January 2021.
- [5] T. M. Akase, S. R. Sheyigari and A. A. Auta, “EVALUATION OF SOCIAL MEDIA INFLUENCE ON THE OUTBREAK OF COVID-19 IN NIGERIA,” *MKAR JOURNAL OF MEDIA AND CULTURE*, vol. 5, no. 1, March 2021.
- [6] B. A. S. Machado, K. V. S. Hodel, V. G. Barbosa-Júnior, M. B. P. Soares and R. Badaró, “The Main Molecular and Serological Methods for Diagnosing COVID-19: An Overview Based on the Literature,” *Viruses* 2021, vol. 13, no. 1, December 2020.

- [7] D. Katayama, Y. Kino and K. Tsuda, “A Method of Sentiment Polarity Identification in Financial News using Deep Learning,” *Procedia Computer Science* 159, pp.1287–1294, 2019.
- [8] M. Costola, M. Nofer, O. Hinz and L. Pelizzon, “Machine Learning Sentiment Analysis, COVID-19 News and Stock Market Reactions,” SAFE Working Paper No. 288, September 2020.
- [9] N. J. Gormsen and R. S. J. Koijen, “Coronavirus: Impact on Stock Prices and Growth Expectations,” *The Review of Asset Pricing Studies*, vol. 10, issue 4, pp.574–597, December 2020.
- [10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of NAACL-HLT 2019*, pp.4171–4186.
- [11] Kurohashi-Chu-Murawaki Lab., JUMAN++ - A Japanese Morphological Analyzer. <https://nlp.ist.i.kyoto-u.ac.jp/EN/?JUMAN%2B%2B>
- [12] Kurohashi-Chu-Murawaki Lab., BERT Japanese Pretrained model. https://nlp.ist.i.kyoto-u.ac.jp/index.php?ku_bert_japanese (Japanese).
- [13] Cabinet Office (Japan), Economy Watchers Survey. <https://www5.cao.go.jp/keizai3/watcher-e/index-e.html>
- [14] TIS Inc., chABSA dataset - Aspect-Based Sentiment Analysis dataset. <https://github.com/chakki-works/chABSA-dataset>