

Public Perception of COVID-19 Vaccine by Tweet Sentiment Analysis

Xuanzhou Yang
Faculty of Data Science
Musashino University
Tokyo, Japan
g2150004@stu.musashino-u.ac.jp

Virach Sornlertlamvanich
Asia AI Institute, Faculty of Data Science
Musashino University
Tokyo, Japan
virach@musashino-u.ac.jp

Abstract The COVID-19 pandemic has given a tremendous impact on the economy, changed the way of living of the whole world. Many lives are lost, the labor markets are affected and the people lifestyle are changed in order to limit the impact of COVID-19 pandemic. As of now, it seems that the COVID vaccine is the only solution for the world to be safe again. In the United Kingdom and the United States, many studies have been conducted on the sentiment analysis where the emotions of participants before vaccination and after vaccination are observed. The first batch of vaccines has been launched at the end of 2020 while some developed countries started early vaccination campaigns, and others are still in the process of ordering vaccines and remained unvaccinated until early 2021. The vaccinations are prioritized on the high-risk groups, such as medical workers and the elderly population. Vaccination for people under age 18 are still not available in the initial stage. Despite the executions of vaccinations, there are various opinions on whether the COVID-19 vaccine is safe, and a number of the population remain skeptical of taking the vaccine. In this research, we analyze tweets to understand public perception on the COVID-19 vaccine by classifying the sentiments and attitudes towards the vaccination and the available types of vaccine .[1]

Social media is an appropriate source of research to analyze public attitudes towards COVID-19 vaccine and what they feel about the various brands of the vaccine in the market. For this research, tweets written in two languages, English and Japanese, have been collected. In Japan, some related surveys on public willingness for vaccination and the

sentiment analysis are already conducted.

This study randomly surveys on the users' tweets about COVID-19 vaccination and their emotions expressed in their tweets. Due to the certain vaccination accidents, people in various countries become more concerned about the side effects and safety of the vaccine due to local deaths of various circumstances and unknown causes. In an attempt to help assess and understand public sentiment towards the initial stage of the vaccination campaign, sentiment analysis tools are utilized. It can discover that there are different sentiment patterns observed in different regions and time points as well as in different vaccine brands.

It is expected that the text categorization process will be conducted using NLTK's dedicated Twitter corpus. In social media data, users enter multiple punctuation marks, acronyms and emoticons to express their emotions. TextBlob tool will be used for analysis, which computationally identifies and classifies text into three emotions: positive, negative or neutral. TextBlob is used because it processes data by including various letters, symbols, etc. in its dictionary. In this method, each word in the dictionary is based on whether it is positive or negative, while adding an emotional analysis of commonly used expressions. In this way, people's attitudes towards vaccines in the UK, the US and Japan can be analyzed. The accuracy of the method is 73.3% in English and 71.9% in Japanese. The results show that the British and Americans are more neutral and positive about vaccines, while the Japanese are more pessimistic about vaccines.[2]

Keywords: social media, sentiment analysis, text mining, COVID-19, vaccine, tweet

I. Introduction

Since December 2019, COVID-19 has rapidly spread around the world, causing severe respiratory diseases and deadly consequences. This health complication has led to 3.94 million deaths. Since March 2020, the pandemic have also spread to Japan, the United States and the United Kingdom. As of 28 June 2021, the three countries have more than 790,000 confirmed cases, 34.5 million confirmed cases and 4.75 million confirmed cases respectively. COVID-19 has already caused more than 14,000, 610,000 and 120,000 deaths respectively in Japan, the US and the UK.

To prevent further spread of the COVID-19 outbreak and alleviate the enormous medical pressure, the development and promotion of a vaccine is critical. Several pharmaceutical companies and universities are developing COVID-19 vaccines at a faster pace than usual. More than 260 possible vaccines for COVID-19 have been proposed, but only a few of them have been approved. Some vaccines are remained at the stage of being tested.

Currently, the availability of COVID-19 vaccine remains limited. The vaccination coverage is over 50 percent in the United States, over 60 percent in the United Kingdom, and just over 20 percent in Japan, as shown in Figure 1.

Because of the severity of the epidemic, the development of the COVID-19 vaccine has been unprecedented. Every country and global organization has lowered standards for COVID-19 vaccine and shortened clinical trials of the vaccine. Although the vaccines currently in use have been rigorously tested and vetted to ensure safety, many remain skeptical about their safety. A number of people even refuse to receive the COVID-19 vaccination out of fear. In order to effectively promote the vaccine, it is undoubtedly important to collect people's perceptions on the vaccine and its side effects. However, social isolation, quarantine and travel restrictions imposed by governments hamper the data collection. Alternatively, social media like Twitter makes an ideal source of data collection for real-time data.

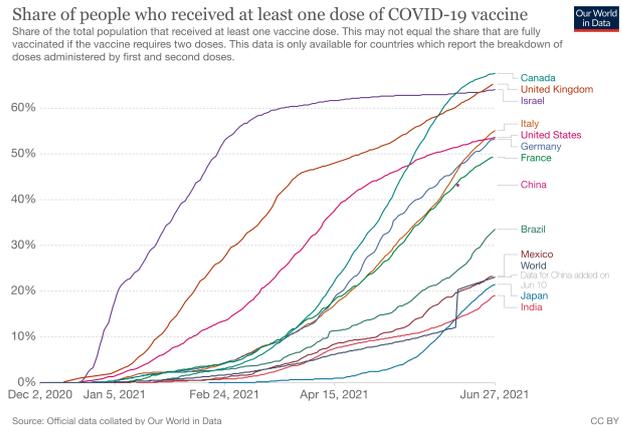


Fig. 1. Population who received at least one dose of COVID-19 vaccine [5]

To investigate the views and attitudes of citizens in the three countries, Japan, the U.S. and the U.K, we apply the Twitter API to collect tweets about the vaccine since it began. The social media analysis of those tweets is conducted. The data set is collected using the Twitter API. Based on the public data provided by Twint, the content of Tweets in English and Japanese are collected, by limiting the time and place of tweets in Japan, the United States and the United Kingdom. Since the vaccine brands vaccinated in these countries are not vetted in other countries. The vaccine brands from other regions cannot be compared in the scope of the study while collecting all the data of marketed vaccines. Thus, the vaccine brands from other regions will not be included and compared in this study. All the COVID-19 vaccine-related tweets from the Twitter stream are downloaded through the Twitter API. Information such as the Twitter ID, date and time, lang, and country location are included in the final set. The use of the data set is in full compliance with Twitter's terms of service.

II. Related Work

Sentiment analysis from Twitter is increasingly becoming a useful and powerful tool for understanding an individual's views on important things going on in society. The population's opinions on pandemics, catastrophes and vaccinations are now readily available on the internet. For the research, 10,033 English tweets and 10,079 Japanese

tweets per month are randomly collected, the English tweets are collected from the date of December 2020 to June 28, 2021 and the Japanese tweets are collected from the date of February 2021 to June 28, 2021, when vaccination started in the US and the UK, and Japan respectively. Using natural language processing and deep learning-based methods, public sentiment and trust towards different brands of COVID-19 vaccine are analyzed. Analysis of the results is expected to help promote vaccination, improve vaccination coverage, and ultimately achieve universal vaccination.

III. Methodology

A. Data collection

Twitter provides APIs to extract large amounts of data from their platform for analysis. There are two types of APIs, the stream API and the search API [3], The stream API is used to retrieve real-time data, while the search API is used to extract historical data by applying some filters. The Twitter search API called TWINT is used to collect the necessary data sets. The public dataset is downloaded, and according to the country location field in the public dataset, tweets from Japan, the US and the UK are retained. Vaccine names are searched as keywords, and the collected tweets are converted into CSV files, which are then merged for follow-up analysis.

As a result of the tweet collection, about 90,000 English tweets from December 4, 2020 to June 28, 2021 and about 100,000 Japanese tweets from February 1, 2021 to June 28, 2021 are collected by applying the keywords using the variation of the vaccine calling names as shown in Table 1. Table 2 shows the number of tweets including the calling name.

Table 1. Variation of vaccine calling name

Full Name	Tweet Name(English)	Tweet Name(Japanese)
Pfizer-BioNTech	Pfizer,pfizer,BioNTech,biontech	ファイザー
Moderna	Moderna,moderna	モデルナ
Oxford-AstraZeneca	AstraZeneca,astrazeneca	アストラゼネカ

Table 2. Frequency of the vaccine calling name

Name	Count	Name	Count	Name	Count
Pfizer	16000	AstraZeneca	16000	Moderna	11599
pfizer	16000	astrazeneca	16000	moderna	5302
BioNTech	6223				
biontech	6220				

Tweet Name(Japanese)	Count
ファイザー	40000
モデルナ	40000
アストラゼネカ	20000

For this data set, the next stage of data preprocessing is carried out in the analysing process to ensure the accuracy of the data set. During the study, the relevant changes in this data set are continuously monitored.

B. Data preprocessing

Since Twitter text data has some typical but non-semantic characteristics, these characteristics are addressed in the data preprocessing step.

First of all, almost every tweet contains a short link, which can be "https://t.". This link has no actual semantics and will cause ambiguity after markup, which will bring unnecessary trouble to future analysis. Therefore, this link needs to be deleted. Second, other punctuation marks are deleted. Third, unnecessary newlines are removed, since this is the most common nonsensical identifier. Fourth, all text content are converted to lowercase to ensure that each word appeared in a consistent format [2]. TextBlob are used for English tweets, and according to Textblob documentation, Textblob takes advantage of Naïve Bayes (NB) model for classification as shown in Figure 2. NB classifier has been trained on NLTK (Natural Language ToolKit) to detect valence of aggregated tweets.

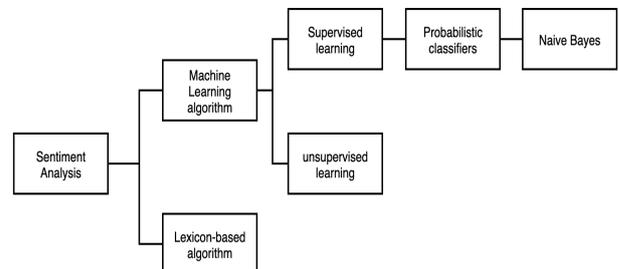


Fig. 2. Naïve Bayes Hierarchy

NB is a probabilistic algorithm that uses Bayes theorem to compute sentiment distribution over the data. However, NB dissects any text to a bag of words which

means the positions of the words are completely disregarded. The Bayes equation to predict the sentiment probability is:

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(\text{features}|\text{label})}{P(\text{features})}$$

Where P(label) is the prior probability of a label, P(features|label) is the prior probability that a given feature set is being classified as a label, and P(features) is the prior probability that a given feature set is occurring.

In the processing of Japanese tweets, classified emotion Japanese Text is used, and each emotion word is summarized into a Text, which is finally combined into an emotion word Text. By summarizing and classifying all the words in the dictionary, a more convenient emotion analysis tool is obtained compared with machine learning. Accuracy still needs to be improved, but when dealing with a large number of data sets, the processing speed is very fast and rapid analysis.

C. Sentiment analysis

Sentiment analysis is used to understand citizens' attitudes towards COVID-19 vaccine from the perspective of emotional expression. At the same time, how people accept different vaccine brands and how people's perceptions of vaccines change over time are explained from multiple dimensions. To remove the stop word, by using TextBlob, the sentiment of all the English tweets are divided into optimistic and pessimistic, as shown in Figure 3.

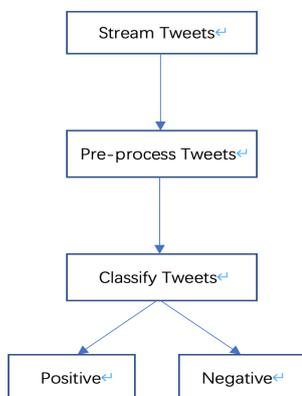


Fig. 3. English Tweets divided into optimistic and pessimistic

Text classification of Japanese tweets is based on the use of emotional word classification. By analyzing and querying whether the words or phrases existing in the tweets also exist in the text, the corresponding analysis results are output after finding the corresponding words. Emotional results are examined and classified according to emotions such as joy, surprise, sadness, disgust, fear and anger, and the results of these emotions are summarized and displayed respectively.

IV. Results and analysis

In the US and UK, the majority of tweets are optimistic about the vaccine, which indicates that the majority of people in the US and UK have a positive attitude towards the different brands of COVID-19 vaccine and a more realistic view as shown in Figure 4.

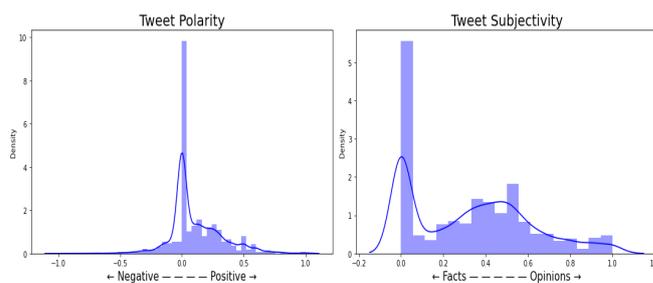


Fig. 4. Tweet Polarity and Sensitivity

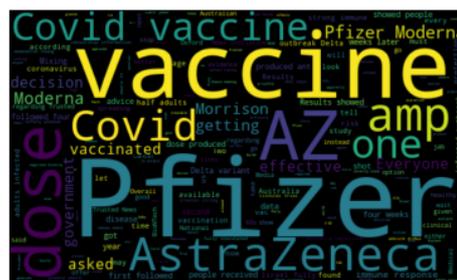


Fig. 5. Pfizer appearing the most times

Meanwhile, among the keywords concerning Vaccine, according to the analysis of WordCloud, among all the keywords, the keyword Pfizer appears the most times, followed by AstraZenca as shown in Figure 5.

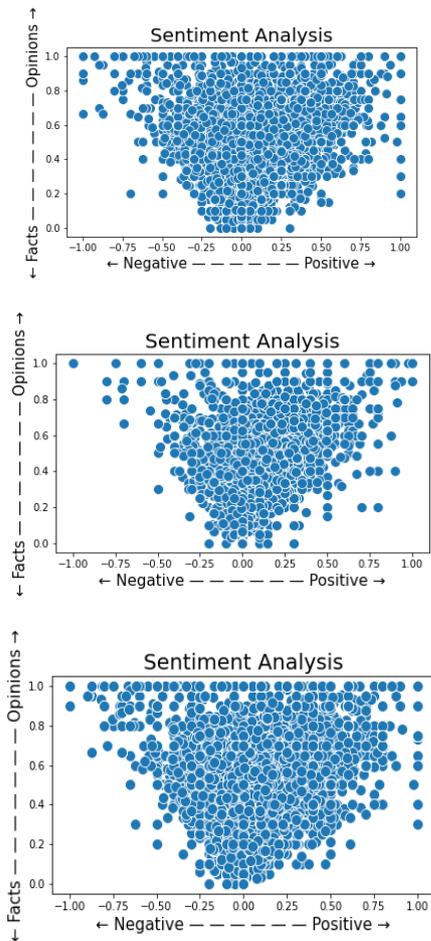


Fig. 6. Sentiment Analysis for Pfizer, AstraZenca, Moderna

TextBlob is used to conduct a sentiment analysis of all the keywords related to Pfizer, and the conclusion is that more people are optimistic and neutral than pessimistic. Among all the analysis results shown in Figure 6, more people commented on Pfizer and Moderna vaccines, indicating that people are more concerned about these two vaccines.

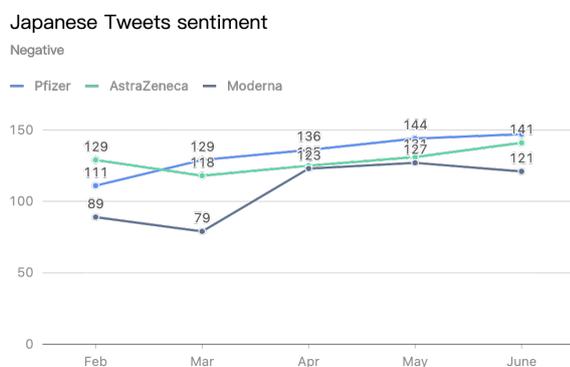


Fig. 7. Negative emotions illustrated by Japanese Tweets

Compared with the United States and the United Kingdom, Japan is completely on the contrary. Although many Japanese people are happy about vaccines, they also have a large aversion to them. Moderna vaccine is the least negative, followed by AstraZenca vaccine. Out of every 10,000 tweets, March's Moderna vaccine had the lowest negative sentiment (79), while April's Pfizer vaccine had the highest (144). As shown in Figure 7, although people's aversion to vaccines remains high, they also have a happy attitude towards vaccination. Out of every 10,000 tweets, Positive emotion has been rising every month.

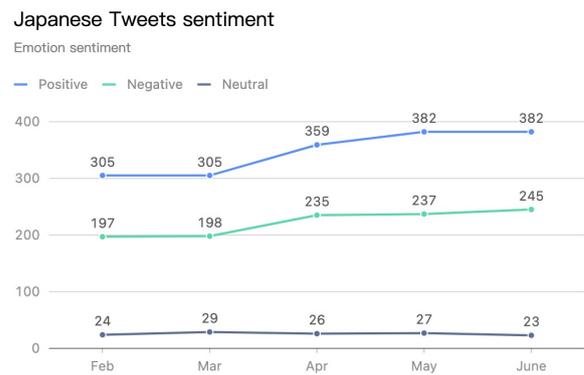


Fig. 8. Emotion Sentiment for Japanese tweets

That is to say, for these two vaccines, people's negative emotions are relatively small, and their willingness to get vaccinated is relatively high. Through analysis shown in Figure 8, it is concluded that the negative emotions towards the Biontech vaccine is partly due to the death caused after local vaccination. It is likely that the local vaccine incidents in the past have led to a skeptical attitude toward vaccines, and confidence upon vaccination will increase once vaccination coverage increases and safety is more widely certified.

According to previous studies on the acceptability and hesitation of COVID-19 vaccine in Japan, nearly 62.1% of adults in the country are willing to receive COVID-19 vaccine. Certain sociodemographic factors, such as gender, age and income level, as well as some psychological factors, such as perceptions of the effectiveness of the COVID-19 vaccine and willingness to protect others through vaccination, are associated with vaccine acceptance [4]. In the study, it is noted that acceptance of the vaccine was low among some

sociodemographic groups, such as women, adults aged 20 to 49, and low-income people. Efforts to increase confidence in the effectiveness of the COVID-19 vaccine and to increase an individual's willingness to protect others through vaccination may be a key to promoting vaccine acceptance.

V. Conclusion

Pfizer, AstraZeneca and Moderna are the leading vaccines given in the UK, the US and Japan, with Japan suspending use of AstraZeneca after a number of medical incidents. For the United Kingdom and the United States, TextBlob is used to process the sentences, and TextBlob used Naive Bayes classification to greatly improve the accuracy. It can be seen from the analysis results that people's attitude towards the three vaccines is generally neutral and positive, with relatively small negative emotions. The future of vaccinations seemed glorious, with Pfizer leading the list of positive emotions, followed by Moderna Vaccines. Both vaccines could be further promoted with improved safety and sufficient actual vaccine production. For Japanese tweets, the use of sentiment analysis text in Japanese is more convenient and time-saving than other machine learning methods. The emotions and attitudes expressed in the sentences can be compared and analyzed through edited and summarized emotion words and corresponding emotions. As shown in the formula, the accuracy of English is 73.3% and that of Japanese is 71.9% according to the formula. True Positive is TP, True Negative is TN, False Positive is FP, and False Negative is FN.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

As the analysis showed, Japan's aversion to the vaccine still remained. However, it also maintained a relatively high level of happiness. In particular, the third ranking is confidence, which indicates that Japanese people's willingness to get vaccinated does exist.

With the improvement of vaccination coverage and the increased number of vaccination, safety is more widely accepted and people could gradually reduce disgusting attitude. Through the analysis of the data results on negative emotions, Moderna, relative to other vaccines,

scored the lowest on the negative sentiment value, meaning it has the best impression. At least two vaccines, which can be show in Japan, Moderna vaccine acceptance may be the best. The second best is Pfizer. On the premise of improving safety and sufficient actual vaccine production, the two vaccines can increase their usage promotion in Japan.

REFERENCES

- [1] T. Na, W. Cheng, D. Li, W. Lu, H. Li, "Insight from nlp analysis: covid-19 vaccines sentiments on social media", Department of Computer Science, University of Manchester, 2021, pp. 1-2.
- [2] A. Shelar and C. Y. Huang, "Sentiment Analysis of Twitter Data", 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018, pp. 1301-1302, doi: 10.1109/CSCI46756.2018.00252.
- [3] K. H. Manguri, R. N. Ramadhan, P. R. M. Amin, "Twitter sentiment analysis on worldwide covid-19 outbreaks", KJAR, 2020, pp. 5.
- [4] R. Batra, A. Sharia Imran, Z. Kastrati, A. Ghafoor, S. M. Daudpota, S. Shaikh, "Evaluating polarity trend amidst the coronavirus crisis in people's attitudes toward the vaccination drive", MDPI, 2021, pp. 6-7.
- [5] M. Machida, I. Nakamura, T. Kojima, R. Saito, T. Nakaya, T. Hanibuchi, T. Takamiya, Y. Odagiri, N. Fukushima, H. Kikuchi, S. Amagasa, H. Watanabe, S. Inoue, "Acceptance of a covid-19 vaccine in japan during the covid-19 pandemic", MDPI, 2021, pp. 1.
- [6] Official data collated by Our World in Data, 2021.