Japan's Twitter Sphere does not Reflect Public Opinions:

Classification and Sentiment Analysis by Computational Social Science Methodology

Tanihara Tsukasa*, Odanaka Yu**

Abstract:
What is posted and what emotions accompany it in political communication on Twitter? This paper contributes to research on the polarization and representativeness of online public opinion by shedding light on this point. We collected a comprehensive set of posts mentioning the ruling Liberal Democratic Party (LDP) during the 2021 national election in Japan. Using supervised machine learning (BERT), we classified the collected posts into anti-LDP, neutral, and pro-LDP. The results showed that more than half of the posts were anti-LDP. This was significantly different from the actual election results. However, the level of retweet was higher for pro-LDP tweets than for the other classifications. We conducted a sentiment analysis of the posts using JIWC. The results revealed that anti-LDP posts were accompanied by more negative sentiment than the other classifications. These results indicate that public opinion on Twitter is polarized and may not necessarily reflect actual public opinion.

Keywords: Twitter, Political communication, Election, Natural language processing

1. Introduction

Twitter (now referred to as X, but called Twitter at the time of the study, hence we will use "Twitter" consistently in this paper) is actively used by candidates and citizens in election campaigns (See Jungherr, 2016 for a systematic review). In Japan, the ban on online election campaigning was lifted in 2013. Since then, social media-based election

* Kanda University of International Studies, Department of International Communication, Lecturer

** Kyoto University of Advanced Science, Faculty of Humanities Department of Psychology, Junior Associate Professor

campaigns have flourished（Uenohara, 2014; Yoshimi, 2017）. There are 45 million active Twitter users in Japan per month, making it a very influential platform that anyone can use for free, and it is a desirable medium for disseminating political communications.

Today, social media plays an increasingly important role in political communication. Many people consume information about politics via social media（Matsa & Shearer, 2018）and participate in politics influenced by social media（Oser & Boulianne, 2020）. Journalists refer to information on social media to set news agenda（Parmelee, 2014）. Moreover, election candidates use social media for campaigns（Jungherr, 2016）. In Japan in 2020, a total of 4.7 million posts with the hashtag "# protest the proposed amendment to the Prosecutor's Office Act" surfaced in only three days. However, it has been reported that Twitter has become a source of fake news and slander（Allcott and Gentzkow, 2017; Tanaka and Yamaguchi, 2016）. In response to such a situation, Twitter Inc. issued a statement in 2021 that read, "You may not use Twitter's services for the purpose of manipulating or interfering in elections or other civic processes" and emphasized that "The public conversation occurring on Twitter is never more important than during elections and other civic events."

As described above, Twitter is an ambivalent medium that can be a platform for active political communication and at the same time, it can be used to publish unhealthy political communication. The negative effects of Twitter in the field of political communication are the polarization of society and its bias.

2. Social media and polarization

The logic of social media-induced polarization is as follows. Unlike traditional mass media, where information has a list-like nature, in social media, it is normal to be exposed to information selectively. Thus, people select only information that fits their beliefs and are surrounded by similar information（homophily）. Users who are surrounded by similar information end up in an echo chamber. As a result, thoughts and opinions become polarized and radicalized.

Since this logic was first proposed by Sunstein（2001）, a number of empirical studies have followed; Conover（2012）collected approximately 45,000 accounts and 250,000 tweets during the 2010 congressional midterm elections and analyzed them. The retweet network was analyzed and found to be fragmented between conservatives and liberals, with limited connectivity. A similar event occurred during the 2016 U.S. presidential election between Trump and Clinton supporters (Thompson, 2016). However, other studies suggest that people are exposed to information containing opinions that differ from their own to a certain extent and that selective exposure does not occur on social media (Facebook: Beam et al, 2018; Levy, 2021, Twitter: Barberá, 2015; Tanaka and Hamaya, 2019), and some have found that only a minority of people with extreme political orientations fall into echo chambers (Eady et al, 2019).On the other hand, some empirical research suggest that exposure to differing opinions on social media can make people self-protective and reinforce the political attitudes they originally held (Bail, 2021). Furthermore, a study comparing polarization across platforms found no polarization on Facebook and WhatsApp, while polarization was observed on Twitter (Yarchi et al, 2021).

Recent research has divided the analysis on polarization into ideological and emotional level polarization (see Kubin and Sikorski, 2021 for a systematic review). The former measures an individual's ideology by asking questions about approval or disapproval of several specific policies. The latter is commonly measured using a feeling thermometer.

Kubin and Sikorski (2021), who conducted a systematic literature review, concluded that social media generally polarizes people on an ideological or emotional level. However, in a study by Levy (2021), who conducted an experiment by manipulating Facebook news feeds in the U.S., it was concluded that Facebook does not contribute to polarization at the ideological level, rather, it contributes to depolarization at the emotional level. In Japan, Tanaka and Hamaya (2019) conducted a survey study and analysis of 100,000 people using difference in difference. The study found that neither Facebook, Twitter, nor blogs contribute to polarization.

As described above, the results are inconsistent with regard to whether social media divides society. There are many possible reasons for this. These include differences between the survey study and the experimental study, differences in the wording of the questions in the survey study, differences in the platforms used in the study, and differences in the conditions of the experiment.

3. Social media and critical discourse

Regarding social media and elections, there is a growing body of research on whether social media posts predict election outcomes (see Chauhan et al, 2021 for a systematic review). Studies in the Western world have shown that social media posts generally predict election results (Burnap et al, 2016 for the U.K.; Wang and Gan, 2017 for France; Tsakalidis et al, 2015 for Netherlands, Greece and Germany; Heredia et al, 2018 for the United States). In Asia, research results have also shown that social media data can predict election outcomes (Jaidka et al, 2018 for Pakistan and India; Barclay et al, 2015 for India; Budiharto and Meiliana, 2018). These studies are conducted by means of volumetric, sentiment, and network analysis using big data from Twitter. In Japan, however, a study was submitted showing that opinions spread via Twitter do not match the actual election results (Toriumi, 2020). Toriumi (2020) collected tweets related to the Tokyo gubernatorial election in 2020 and analyzed the retweet network. Despite the fact that the election result was a resounding victory for Governor Koike, tweets critical of her were in the majority on Twitter, forming a large cluster. Conversely, no tweets supporting Koike were found among the high-share tweets. The results of this study indicate a significant discrepancy between the discourse on Twitter and the election results.

Some studies suggest that Twitter discourse in Japan leans toward liberals (Tanihara, 2022). Tanihara (2022) analyzed Twitter discourse in Japan during the outbreak of COVID-19 in Japan. The results revealed that information was actively disseminated by the politically liberal segment of the population, with many comments critical of the government's countermeasures. According to Toriumi (2020), the anti-Koike segment was active in Twitter discourse. This is consistent with Tanihara's (2022) findings, since Governor Koike has a conservative political viewpoint. In Japan, a conservative party (Liberal Democratic Party (LDP)) has been in power since 2012, so this result is inevitable if critical discourse tends to gather on Twitter.

The literature that studies information on social media in general suggests that social media is more likely to attract critical discourse. Fan et al. (2014) investigated how Weibo users are emotionally connected to each other. The analysis revealed that when connections are close, the emotional correlate of "anger" in particular is significantly larger than that of other emotions. In other words, people on Weibo tend to be connected based on the emotion of "anger." Wollebæk et al. (2019) also conducted a survey study in Norway to investigate the relationship between anger about the country's economic and social situation and the degree of participation in political discussions on social media. The results revealed that those with higher anger scores were more likely to participate in political discussions on social media. Lottridge and Bentley (2018) conducted an online survey of 1,000 people using 262 screenshots of news articles to explore their motivations for sharing news. The results showed that political news was the most shared (48%), and 77% of the news was viewed negatively by those who shared it. Based on these studies, it can be assumed that negative opinions that spark anger are more likely to be spread on social media.

4. Research gap

There are two questions in the previous works: "(1) Does social media divide society?" "(2) Can social media reflect public opinion during elections?" As we reviewed, the results of these issues are not consistent across survey methods and survey regions. It would be beneficial to reconsider them using new patterns. This study, however, makes a theoretical contribution to the above issues by conducting a more basic survey.

Specifically, we will identify what is posted and what emotions accompany them in the Twitter sphere at election time. This is because post volume and post sentiment have been shown to be potential predictors of election results in related studies. (Heredia et al, 2018). By clarifying these points, we can make a theoretical contribution to point (1) above by identifying the nature of discourse on Twitter as a precondition for the polarization. For issue (2), we can make a theoretical contribution by identifying the bias of the Twitter sphere during elections.

Specifically, we focus on tweets that mention the governing party. The time period for which we collected posts was the day before the day of the 2021 national election. We selected the day before the date of the election because this is the time when election-related tweets were most concentrated. The reasons for focusing only on tweets that mention the governing party are as follows. Studies conducted in Western and Asian countries that hold general elections under big two parties collect and analyze tweets that refer to both conservative and liberal parties, but this approach cannot be adopted in Japan. Because the LDP, a conservative party, has been in power since its formation in 1955, with the exception of four years from 1993 to 1994 and 2009 to 2012. The LDP was also expected to win in the most recent election. Therefore, collecting tweets about attitudes toward the LDP captures the essence of political communication during the election. One possible approach would be to collect tweets referring to the opposition parties and compare them to tweets referring to the LDP. However, since there are many opposition parties in Japan and their political beliefs significantly vary, it would be difficult to interpret the results of the analysis, so this approach was abandoned. In addition, since it is not equivalent to refer to the largest ruling party currently in power and to refer to the opposing parties, which are minority parties, it is not appropriate to compare the two. In other words, the tweets that mentioned the opposition party seemed to be accompanied by relatively strong feelings and did not fit the method of simply comparing numbers. It is not unusual to study political communication by focusing only on tweets about a particular group or person. For example, Kobayashi and Ichifuji（2015）attempted to identify the impact of political communication on voters during the 2013 national election in Japan by focusing on the tweets of the election candidate with the most followers.

Tweets that mentioned the LDP are classified as anti-LDP, neutral, or pro-LDP using a supervised machine learning method（BERT）. First, we show descriptive statistics to clarify how much of any of these posts are present. We also compare the number of retweets（RTs）for each classification. Given the findings of Tanihara（2022）and Toriumi（2020）in Japan and the findings of Fan et al（2014）and Lottridge & Bentley（2018）on tweet diffusion, we expect that in Japan, critical comments about the LDP are more likely to spread on Twitter. This is because if negative sentiment motivates posts, then those who are dissatisfied with the current political situation in Japan will be negative toward the LDP. Although there are ways to separate the tweets of the general public from those of election candidates, in this analysis, we do not distinguish between the two because our focus is to reproduce the timeline of Twitter users as accurately as possible[1)]. Based on the above, we set the following hypotheses（Study 1）.

H1 Tweets critical of the LDP are more likely to be posted than tweets that are not.
H2 Tweets critical of the LDP are more likely to spread than those that are not.

Next, we analyze the sentiment of each tweet. We use a sentiment dictionary to identify the sentiment of each tweet. Specifically, we use the JIWC dictionary, a Japanese language emotion dictionary created by the social computing lab at Nara Institute of Science and Technology (Shibata et al, 2017). The JIWC dictionary is a dictionary that divides frequently occurring Japanese expressions into seven emotional categories: "sadness," "anxiety," "anger," "disgust," "trust," "surprise," and "joy." In this paper, we focus specifically on negative sentiment. In light of related studies (Tanihara, 2022; Toriumi, 2020; Fan et al, 2014; Lottridge and Bentley, 2018), tweets classified as anti-LDP should be accompanied by negative emotions. Thus, the following hypothesis is set (Study 2).

H3 Tweets classified as anti-LDP are accompanied by negative sentiment.

5. Study1

5.1. Methodology

In Study 1, we classify the collected tweets into anti-LDP, neutral, and pro-LDP using a supervised machine learning method to clarify the distribution of discourse in the Twitter sphere. We also compare the number of RTs of the tweets classified into each of these categories to reveal differences in the ease of diffusion based on political attitudes.

Specifically, we processed the data in the following way: We used the Twitter API to comprehensively collect tweets mentioning "LDP." The target time point was October 30, 2021, the day before the opening day of the House of Representatives election (335,820

---

1) Some may argue that tweets from election candidates should be excluded from the analysis because of the high number of retweets, which may be outliers in terms of statistical analysis (however, in this study, we used a nonparametric test, so the results are not affected by the mean.). However, doing so would exclude tweets that are frequently retweeted and easily seen by users, which would undermine the purpose of this study.

tweets). Election campaigns are prohibited by law before the day of public announcement and on the voting day. Therefore, the day before polling day was the day when the most election-related posts were tweeted. The first objective of this study is to classify this data into anti-LDP, neutral, and pro-LDP. To do so we employed supervised machine learning methods. We employed a natural language processing model called BERT.

Bidirectional Encoder Representations from Transformers (BERT) is a natural language processing model using deep learning (Devlin et al, 2019). BERT has higher prediction accuracy and has been used for text classification. For example, a model specific to tweets related to the COVID-19 can classify tweets in terms of attitudes toward the corona vaccine (positive, negative, or neutral) (Müller et al, 2020).

BERT and other natural language processing models are based on the distributional hypothesis, which holds that the meaning of a word is based on its relationship to its surrounding words. According to this hypothesis, many natural language processing models learn the meanings of words and sentences and represent them as vectors. BERT is based on a model called Transformer, which enables a more flexible learning of word relations. The main idea of Transformer is that it employs only the Attention mechanism in learning the relationships between words. Specifically, attention is a mechanism that can consider the relationships among all words that appear in a given sentence regardless of word order. It overcomes the problem of conventional models, which only consider context in a unidirectional manner, and calculates on which other words to focus when understanding the meaning of a word in a sentence in a bidirectional manner.

In this paper, BERT is used to classify whether the collected tweets are anti-LDP, neutral, or pro-LDP. The topic model is well-known for machine learning-based document classification. Topic model extracts the multiple topics that make up a given set of texts and allows classification in terms of which topics embody each text (Blei et al, 2003). However, it has been noted that this is not a suitable learning model when the texts to be learned are short, such as tweets (Curiskis et al, 2019). Conversely, BERT has proven to be effective in classifying tweets, with a model created specifically for COVID-19-related tweets (Müller et al, 2020).

BERT learning is created through two stages, namely, pre-training and fine tuning. Pre-training consists of two tasks, namely, Masked Language Modeling and Next Sentence Prediction. This study uses a pre-trained Japanese language model developed by the Inui Lab at Tohoku University (https://github.com/cl-tohoku/bert-japanese/blob/main/README.md). This model was learned from the August 2020 Japanese Wikipedia article on general trends of word usage in Japanese. Several pre-trained models of Japanese in BERT are available. However, we select this model, because scholars have reported high levels of accuracy of its finetuned version in classifying Japanese tweets, which is suitable for the study objectives (Ikegami et al. 2021; Hatanaka & Toriumi, 2022).

Fine tuning is the learning of features of a particular type of sentence to be analyzed, in addition to the pre-trained understanding of a given language in general. For the classification task that this paper performs, sentences with classification labels are further trained into a pre-trained model. For example, the target is a collection of tweets that have been assigned the labels anti, neutral and pro for each of the tweets. By having BERT learn not only the sequence of words in the sentences, but also the relationship between them and the labels, it will be able to guess the corresponding labels given a sequence of words. Note that while pre-training requires a large corpus and research resources, fine tuning can be done by training a relatively small number of texts.

The authors manually coded the training data for fine tuning. We randomly selected 1,500 tweets for analysis; a university faculty member specializing in sociology and the author classified each tweet as anti-LDP, neutral, or pro-LDP. Cronbach's alpha coefficient between the two was 0.838, which was sufficiently reliable. Fine tuning was performed using this training data. A total of 70% of the training data was used for training, and the remaining 30% was used for testing. The purpose of this step was to measure accuracy by having the trained model predict the labels of the 30% of the data and matching them with the manually coded ones. Accuracy was tested and found to be 0.9933, which was deemed sufficient for the purpose of identifying trends in big data as a whole. The study then used the fine-tuned model to classify the data by estimating under which labels the remaining tweets belong. The program and parameters for running the analysis by BERT are based on the commentary and published code by Omi et al. (2021) (https://github.com/stockmarkteam/bert-book).

5.2. Result

Table 1 presents the results of the classification using BERT: 52.4% were anti-LDP, 28.6% were neutral, and 18.9% were pro-LDP. As expected, anti-discourse is predominant, which accounted for more than half of the total. Thus, the result support H1.

Table 1 Number of tweets per political attitude

	tweets (excluding RTs)	tweets (including RTs)	ratio (%)
anti-LDP	33,226	176,097	52.4%
neutral	20,310	96,138	28.6%
pro-LDP	6,040	63,585	18.9%

5.3. Comparison of the number of RTs in each category

We elucidate the relationship between political attitudes and information diffusion by comparing the number of RTs per category. First, we compared the average number of RTs. Contrary to expectations, pro-LDP exhibited the highest average number of diffusion (i.e., 9.5 times). Anti-LDP is 4.3 times, and neutral is 3.7 times. However, what is the degree of representativeness of the average number of RTs? Out of the 59,576 original tweets analyzed, 41,337 had zero RTs, which indicate a skewed distribution. In addition, a few tweets received an extremely large number of RTs. Therefore, considering the distribution of RTs as normal is difficult. The Kolmogorov–Smirnov test revealed that the null hypothesis of a normal distribution was rejected at the $p < 0.001$ level.

Table 2 Comparison of retweet counts

	RT	SE
anti-LDP	4.3000	0.3573
neutral	3.7335	0.4529
pro-LDP	9.5273	0.6758

Therefore, the study performed a nonparametric test to compare the levels per classification of RT counts. We first performed the Kruskal–Wallis test, because we compared three groups. The result indicated that the null hypothesis that the three groups have equal rank sums was rejected at the $p < 0.001$ level. In addition, the Steel–Dwass test was conducted as a post hoc test. Table 2 presents the results along with the effect sizes. Although a statistically significant difference is noted between the two groups, the effect size among the anti-neutral groups is extremely small, such that it cannot provide a substantial implication. In contrast, small effect sizes were detected between pro-neutral and pro-anti. This means that tweets classified as pro are more likely to be diffused than those classified as nonpro. Therefore, the results reject H2. Effect size γ was calculated using the following formula（Mizumoto & Takeuchi, 2011）.

$$\gamma = \frac{Z}{\sqrt{N}}$$

where Z is the standardized z-value in the nonparametric test, and N is the sample size.

Table 3 Nonparametric test results（RT）

	Effect Size	p-value
anti-neutral	0.06	0.001
anti-pro	-0.19	0.001
neutral-pro	-0.27	0.001

6. Study2

6.1. Methodology

Study 2 is an analysis of the emotions associated with tweets. Related studies have repeatedly pointed out that negative emotions are easily diffused. We will examine whether this is the case in Japan.

With regard to sentiment analysis, we conducted a dictionary-based analysis. We used the JIWC dictionary（Shibata et al, 2017, https://github.com/sociocom/JIWC-Dictionary）. The LIWC（Pennebaker et al, 2015）is the de facto standard for emotion dictionaries in the social sciences. JIWC is a Japanese emotional dictionary created by the social computing lab at Nara Institute of Science and Technology, following in the footsteps of LIWC. The JIWC dictionary is created from the episode bank. The episode bank consists of "sadness," "anxiety," "anger," "disgust," "trust," and "surprise." The project asked people to write about their feelings and their corresponding episodes, and collected approximately 4,000 sentences for each emotion. The relationship between each emotion and the words used to express it was analyzed to show the connection between a word and various emotions. For example, the word "treatment" corresponds to "anxiety" and "trust," respectively.

The analysis in this paper was conducted in the following steps. First, a morphological analysis of each tweet was performed. The parts of speech extracted were "noun," "verb," "adjective," and "adjectival verb." MeCab was used as the morphological analyzer. Second, for the list of morphologically analyzed words, we counted words that fit the sentiment words registered in the JIWC. Thus, the sentiment value of each tweet was listed. The list shown in Table 4 was created for the number of tweets. Third, the four negative emotions of "sadness," "anxiety," "anger," "disgust" were considered as negative emotions, and these values for each tweet were added and averaged to create the negative value for that tweet. The statistical analyses and natural language processing analyses in this paper were conducted using Stata 16.1 and Python 3.8.3.

Table 4 Examples of emotional value assignment

text	label	retweet	Sadness	Anxiety	Anger	Disgust	Trust	Surprise	Joy
バリバリ自民臭の漂うセージ屋だから、力で捩じ伏せるんが好きだいね。本当次は退いてもらうんべ。	0	0	0	1	0	0	0	0	0
現金給付だけで景気回復するわけもなく経済政策とセットでしょうーコロナ当初から若い人のために経済まわさなきゃと言ってきたNHK党の給付の仕方、れいわの先を見据えた給付公約はブレてない自民党は「検討」打った対策は的外れ与党なのに検討してる時間長すぎでしょしてる詐欺	0	0	1	4	2	5	0	2	1
この調子で明日の投票日も過去最多の投票率にしましょう🎉衆院選の期日前投票１６６２万人、過去最多だった前回を９７・９万人上回る🏃#選挙に行こう#報道特集#Yahooニュース#こんなひどい政治に鉄ついを#自民党 #維新は自民党の補完勢力 https://t.co/lkBRbR8gzK	0	24	0	2	1	3	0	0	1

.
.
.
.
.
.

6.2. Result

The focus of this paper is on whether tweets classified as anti-LDP are accompanied by negative sentiment. The average negative value for each classification is shown in Table 5. Indeed, anti-LDP tweets appear to have a higher negative value than the other classifications. However, as with the number of RTs, doubt arises as to whether the negative values are normally distributed. Therefore, we conducted the Kolmogorov–Smirnov test, and the null hypothesis that the distribution is normal was rejected at the $p < 0.001$ level.

Table 5 Comparison of negative value

	Mean of negative value	SE
anti-LDP	1.043	0.004
neutral	0.819	0.005
pro-LDP	0.799	0.009

Therefore, a nonparametric test was performed to compare the levels of negative values in each category. Since we compared three groups, we first performed the Kruskal–Wallis test. As a result, the null hypothesis that all three groups have equal rank sums was rejected at the $p < 0.001$ level. In addition, the Steel–Dwass test was conducted as a post hoc test. The results are shown in Table 6, along with the effect sizes. According to the results of the

Steel–Dwass test, the null hypothesis that there is no difference between neutral and pro cannot be rejected. On the other hand, although the effect sizes were small, there was a statistically significant difference between anti-neutral and anti-pro. Thus, H3 is supported.

Table 6 Nonparametric test results（emotion）

	Effect Size	p-value
anti-neutral	0.15	0.001
anti-pro	0.12	0.001
neutral-pro	-	0.317

7. General discussion

Based on the above, we give comprehensive discussions. Study 1 reveals that anti-LDP tweets account for more than half of all tweets and that pro-LDP tweets are more diffused than other tweets. While the former is as hypothesized, the latter is an unexpected result.

This indicates that anti-LDP can be a stronger motivation for posting tweets, but pro-LDP can be a stronger motivation for RTs. Although few empirical studies have analyzed posts and RTs separately, Tanihara（2022）provides a few hints; Tanihara（2022）provides a few hints by analyzing the difference in the nature of posts and RTs as follows:

> The psychological cost of retweeting is considered relatively lower than that of tweeting because retweeting is a reaction to original tweets. Unlike tweeting, which requires the creation of original text, retweeting can be done in one click. Therefore, the decision-making process to retweet is considered to be less burdensome than the decision to tweet（p. 78）.

What does it mean that anti-LDP is active only in posts with higher psychological costs? To understand this, we must understand the position of LDP in Japan. First, the actual election results were a resounding victory for the LDP（Table 7）. Second, the LDP has been in power for most of Japan's constitutional history. Thus, every election is conducted with the expectation that the LDP will win. Therefore, this is not an election in which voters choose between a conservative or liberal party, as in the United States or the United Kingdom. In such a situation, anti-LDP is inevitably more active. This is because if negative sentiment is the motivation for tweeting, then those who are dissatisfied with the current political situation in Japan will be negative toward the LDP. Given the related research on the accounts of election candidates, in which those who lag in their prior predictions are more active on Twitter (Rossini et al., 2018), this study suggests that the accounts of ordinary citizens display a similar trend.

Table 7 Result of the election

		Number of parliamentary seats
Ruling party	Liberal Democratic Party (LDP)	261
	New Komeito (NK)	32
Opposition party	The Constitutional Democratic Party of Japan (CDP)	96
	Japan Restoration Party (JRP)	41
	The Democratic Party for the People (DPP)	11
	Japanese Communist Party (JCP)	10
	Others	14

On the other hand, what does it mean when the level of RTs is higher for pro-LDP? Based on Tanihara's (2022) analysis of the difference between posts and RTs, this study infers that RTs represent a loose endorsement that does not require one to post by oneself. This point is interesting when considered in conjunction with the fact that LDP supporters are in the majority in actual public opinion. Table 2 shows that pro-LDP posts are very small at 18.9%, less than neutral. However, there is a high number of diffusions.

When examined in light of related studies, it is possible to understand that the majority is not motivated post anything. In other words, since the LDP is expected to win the election even if they do not contribute anything, they refrain from contributing anything that is psychologically burdensome. However, compared to neutral voters, they have a more positive political attitude, which can be understood as a loose endorsement in the form of RTs.

Next, we interpret the relationship with sentiment. The results show that anti-LDP posts are accompanied by more negative emotions than neutral and pro-LDP posts. Although related studies suggest that anti-LDP posts with negative sentiment are more likely to be diffused than other posts, the results of nonparametric tests suggest that pro-LDP posts have a larger number of RTs. In other words, it can be inferred that the diffusion effect of emotion is smaller than that of political attitude. For reference, we tried to find the correlation coefficient between negative sentiment values and RT counts, but the coefficient, while statistically significant, was too small and the effect on RTs was not supported in the present analysis（$r = 0.0177$, $p < 0.001$）.

However, it is significant that the nonparametric test revealed in a scientific manner that anti-LDP tweets are accompanied by negative sentiment. While we did not find evidence that negative sentiment motivated RTs, it did motivate posts. Based on the aforementioned difference in the nature of posting and RT, negative emotions motivate posting behavior, which is a stronger statement of intention, rather than RT, which is a loose endorsement. This is also interesting when considered in the context of Japan's unique situation of a long-term LDP administration. Since the LDP was expected to win the election, the anti-LDP hoped to create an anti-LDP atmosphere on Twitter to change people's voting behavior. In other words, they had a stronger motivation to post their opinions than those who were neutral and pro-LDP. This motivation is thought to be expressed as negative emotions.

Emotionally charged tweets are likely to influence people. During the 2013 national election, a field experiment was conducted on the impact of the exposure of the Twitter accounts of prominent politicians on voters（Kobayashi & Ichifuji, 2015）. The results showed that exposure to Twitter did not affect voters' political knowledge or voting behavior, but it did affect their emotions. Twitter has a 140-character limit. It is difficult to influence approval or disapproval of an issue that requires knowledge to make a decision in a short text. However, it is possible to appeal to emotions even with short sentences. Thus, the fact that over half of the tweets mentioning the LDP are anti-LDP and are accompanied by negative emotions means that the Twitter sphere can induce people to dislike the LDP. Despite this, the actual election result was a resounding victory for the LDP. Here, we can see the separation between public opinion in the Twitter sphere and actual public opinion. As far as the discourse on the LDP is concerned, the opinion of Twitter users and offline public opinion are separated in Japan.

What contribution can the above discussion make to previous research? First, what contribution can this study make to the question, "Does social media divide society?" If those who are anti-LDP make more posts and those who are pro-LDP retweet more, then the information we see on Twitter is either anti-LDP or pro-LDP, and the discourse on Twitter appears to be divided into two. This is why many people respond that the Internet is a scary place and why those who use social media are more likely to believe that the world is polarized (Gollwitzer, 2018 for USA; Yamaguchi, 2015 for Japan). While continued research is needed to determine whether social media divides society, it is, at least, clear from this study that social media discourse is divided. These conclusions are consistent with related research demonstrating that holding extreme views predicts social media postings in Japan, the U.S., and South Korea (Yamaguchi, 2022). In the U.S., politically moderate individuals tend not to post on social media (Bail, 2021). This is because they try to avoid being attacked by those with strong political attitudes.

Second, what contribution can this study make to the question, "Can social media reflect public opinion during elections?" The actual election results showed a resounding victory for the LDP, as shown in Table 7. In other words, tweets that mention the LDP did not predict the outcome of the election. This can be considered to represent a characteristic of Internet public opinion. Traditional public opinion polls are conducted by researchers or governments that distribute questionnaires at random. People answer questions because they are asked. This is passive. On the other hand, online public opinion is an active form of transmission. Therefore, it is theoretically possible that it is accompanied by relatively strong feelings. In fact, anti-LDP tweets are accompanied by more negative feelings than other tweets. And this leads to the act of tweeting. However, it is not representative of the actual election results.

Furthermore, this study has methodological contributions. Most previous studies on elections and social media use lexicon-based techniques and few use deep learning techniques（Chauhan et al, 2021）. This study is one of the few attempts to reveal aspects of political communication on Twitter using state-of-the-art technology called BERT. Many studies on election forecasting by social media have tried to discover how to make forecasts more accurate. However, it is a new discovery that in some cases, such as the Japanese case revealed in this study, there is a discrepancy between Twitter public opinion and actual public opinion in the first place. In addition, several studies on social media and political communication focus on the U.S., and there is a need for research in other countries（Kubin and Sikors, 2021）. Japan is in a unique situation where the LDP has been the ruling party for a long time. Examining this situation in conjunction with research on political communication in the U.S. and the U.K. will reveal both commonalities and differences across cultures and situations.

8. Conclusion

In this paper, we used state-of-the-art computational social science methods to analyze the pre-election discourse in the Twitter sphere. This work reveals the distributions of people's opinions toward the ruling LDP. While the anti-LDP tweets were in the majority, the pro-LDP tweets had the highest level of RTs. Furthermore, anti-LDP posts were found to be accompanied by negative sentiment.

It has been suggested that political communication in the Japan's Twitter sphere may be a place where dissatisfaction is reinforced（Tanihara, 2022）. The results of this study support this idea. Even pre-election predictions predicted an LDP victory, and indeed the LDP easily won. In the Twitter sphere, it is possible that those dissatisfied with the long-running LDP government may form clusters. This could have resulted in a large number of anti-LDP posts and accompanying negative sentiment in the Twitter sphere.

On the other hand, this paper has its limitations. As previously noted, the paper focuses only on tweets that mention the LDP. However, there are a number of minority parties in Japan that are very active online（NHK Party, Reiwa Shinsengumi）. By clarifying the characteristics of the social media discourse surrounding them, the overall picture of political communication during elections will become clearer. Such a study would reveal the characteristics of bias in the Japan's Twitter sphere. That is, if there is a liberal bias, there will be more pro-tweets than anti-tweets when it comes to parties with liberal policies. If there is a negative bias, then there will be more anti-tweets for all parties, regardless of their policies.

As mentioned above, the political situation in Japan is unique. Unlike the U.S. or the U.K., Japan does not have two major political parties, and there is no situation where one has to choose between liberal and conservative. Under these circumstances, political communication in the Twitter sphere may be evolving in its own unique way. This paper is novel in that it reveals that the discourse in the Twitter sphere, although limited in its attitude toward the LDP, it showed the opposite aspect of the actual election results.

Disclosure Statement

There are no potential conflicts of interest with respect to the research and authorship of this article.

Data

The replication data for this article can be found at https://doi.org/10.7910/DVN/0RDSHI

References

Allcott, H. & G. Matthew（2017）Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31（2）, pp.211-236. DOI: 10.1257/jep.31.2.211

Bail, C.（2021）*Breaking the social media prism: How to make our platform less polarizing*, Princeton: Princeton University Press.

Barberá, P.（2015）*How social media reduces mass political polarization: evidence from Germany, Spain, and the U.S.*, Working Paper, New York University, http://rubenson.org/wp-content/uploads/2015/10/barbera-tpbw.pdf.

Barclay, F. P., C. Pichandy, A. Venkat & S. Sudhakaran（2015）India 2014: Facebook 'Like' as a Predictor of Election Outcomes. *Asian Journal of Political Science* 23（2）, pp.134-160. https://doi.org/10.1080/02185377.2015.1020319

Beam, M. A., M. J. Hutchens & J. D. Hmielowski（2018）Facebook news and（de） polarization:

reinforcing spirals in the 2016 US election. *Information, Communication & Society* 21（7）, pp.940-958. https://doi.org/10.1080/1369118X.2018.1444783

Blei, D. M., A. Ng & M. I. Jordan（2003）Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3, 993–1022, https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Budiharto, W. & M. Meiliana（2018）Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big Data* 5, pp.1-10. https://doi.org/10.1186/s40537-018-0164-1

Burnap, T., R. Gibson, L. Sloan, R. Southern & M. Williams（2016）140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies* 41, pp.230-233. https://doi.org/10.1016/j.electstud.2015.11.017

Chauhan, P., N. Sharma & G. Sikka（2021）The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing* 12（2）, pp.2601-2627. https://doi.org/10.1007/s12652-020-02423-y

Conover, M. D., B. Gonçalves, A. Flammini & F. Menczer（2012）Partisan asymmetries in online political activity. *EPJ Data Science* 1（6）, pp.1-19. https://doi.org/10.1140/epjds6

Curiskis S. A., B. Drake, T. R. Osborn & P. J. Kennedy（2019）An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management* 57（2）, 102034. https://doi.org/10.1016/j.ipm.2019.04.002

Devlin, J., M. W. Chang, K. Lee & K. Toutanova（2019）*BERT: pre-training of deep bidirectional transformers for language understanding*. In Proceedings. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1, pp.4171-4186.

Eady, G., J. Nagler, A. Guess, J. Zilinsky & J. A. Tucker（2019）How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *SAGE Open*. pp.1-21, https://doi.org/10.1177/2158244019832705

Fan, R., J. Zhao, Y. Chen & K. Xu（2014）Anger is more influential than joy: sentiment correlation in Weibo. *PLOS ONE* 9（10）, e110184.
https://doi.org/10.1371/journal.pone.0110184

Gollwitzer, A.（2018）Social media use relates to perceiving the United States as more politically polarized, Available at *SSRN*, https://ssrn.com/abstract=3137325 or http://dx.doi.org/10.2139/ssrn.3137325

Hatanaka, R., & F. Toriumi（2022）. Corona ka ni okeru twitter campaign no user bunseki［Analysis of giveaway tweets users in Covid-19］. *Proceedings of the 36th annual conference of the Japanese Society for Artificial Intelligence*, 2B6GS1001-2B6GS1001. https://doi.org/10.11517/pjsai.JSAI2022.0_2B6GS1001

Heredia B., J. D. Prusa & T. M. Khoshgoftaar（2018）Social media for polling and predicting United States election outcome. *Social Network Analysis and Mining* 8, pp.1-16. https://doi.

org/10.1007/s13278-018-0525-y

Ikegami, T., F. Nin, S. Nishide & S. Kou (2021). Tweet nihonngo kanjyo corpus ni taisuru BERT no kouka no kensyo [Testing the effect of BERT on a corpus of emotions of Japanese in tweet]. *Proceedings of the 83th annual conference of the Information Processing Society of Japan*, (1), pp.527-528.

Jaidka K., S. Ahmed, M. Skoric & M. Hilbert (2018) Predicting elections from social media: a three-country, three-method comparative study. *Asian Journal of Communication* 29(3), pp.1-22. https://doi.org/10.1080/01292986.2018.1453849

Jungherr, A. (2016) Twitter use in election campaigns: a systematic literature review. *Journal of Information Technology & Politics* 13(1), pp.72-91, https://doi.org/10.1080/19331681.2015.1132401

Kubin, E. & C. von Sikorski (2021) The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), pp.188-206. https://doi.org/10.1080/23808985.2021.1976070

Kobayashi, T. & Y. Ichifuji (2015) Tweets that matter: evidence from a randomized field experiment in Japan. *Political Communication* 32(4), pp.574-593. https://doi.org/10.1080/10584609.2014.986696

Levy, R. (2021) Social media, news consumption, and polarization: evidence from a field experiment. *American Economic Review* 111(3), pp.831-870, DOI: 10.1257/aer.20191777

Lottridge, D. & F. R. Bentley (2018) Let's hate together: How people share news in messaging, social, and public networks, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp.1-13, https://doi.org/10.1145/3173574.3173634

Matsa, K. E. & E. Shearer (2018, September 10). *News use across social media platforms 2018*. Pew Research Center. https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/

Mizumoto, A. & O. Takeuchi (2011) Koukaryo to kenteiryoku bunseki nyumon: toukeiteki kentei wo tadasiku tsukautameni (Introduction to effect size and statistical power analysis: For proper use of statistical tests) *Methods for better foreign language education: The Japan association for language education & technology Kansai department Methodology research division 2010 annual report*, pp.48-74, https://www.mizumot.com/method/mizumoto-takeuchi.pdf

Müller, M., M. Salathé & P. E. Kummervold (2020) COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter, *arXiv preprint arXiv:2005.07503*

Omi, T., K. Kaneda, M. Morinaga & A. Emami (2021) *BERT ni yoru sizen gengo syori nyumon: Transformers wo tsukatta jissen programming*. (Introduction to Natural Language Processing with BERT: Practical Programming with Transformers.) Tokyo: Ohm-sha.

Oser, J. & S. Boulianne (2020). Reinforcement effects between digital media use and political participation: A meta-analysis of repeated-wave panel data. *Public Opinion Quarterly*, 84(S1), pp.355-365, https://doi.org/10.1093/poq/nfaa017.

Parmelee, J. H.（2014）. The agenda-building function of political tweets. *New Media & Society*, 16（3）, pp.434-450. https://doi.org/10.1177/1461444813487955

Pennebaker, J. W., R. L. Boyd, K. Jordan & K. Blackburn（2015）*The Development and psychometric properties of LIWC2015*. The University of Texas at Austin.

Rossini P., J. Stromer-Galley, K. Kenski, J. Hemsley, F. Zhang & B. Dobreski（2018）. The relationship between race competitiveness, standing in the polls, and social media communication strategies during the 2014 U.S. gubernatorial campaigns. *Journal of Information Technology & Politics*, 15（3）, pp.245-261, https://doi.org/10.1080/19331681.2018.1485606

Shibata, D., S. Wakamiya, K. Ito & E. Aramaki（2017）*JIWC: Cloud sourcing ni yoru nihongo kanjyou hyougen jisyo no koutciku*.（JIWC: Crowdsourcing to build a Japanese emotional expression dictionary）, paper presented at the annual meeting of the association for natural language processing（23th）pp.13-17 Mar, Tsukuba, Japan, available at: https://www.anlp.jp/proceedings/annual_meeting/2017/pdf_dir/B5-3.pdf

Sunstein, C. R.（2001）*Republic.com*. Princeton: Princeton University Press.

Tanaka, T. & S. Hamaya（2019）*Netto ha syakai wo bundan shinai*（The Internet does not polarize society）.Tokyo: KADOKAWA.

Tanaka, T. & S. Yamaguchi（2016）*Netto enjoy no kenkyu*（Research on online flaming）. Tokyo: Keisoshobo.

Tanihara, T.（2022）The bias of Twitter as an agenda-setter on COVID-19: An empirical research using log data and survey data in Japan. *Communication and the Public* 7（2）, pp.67-83, https://doi.org/10.1177/20570473221075846

Thompson, A.（2016, December 8）Journalists and Trump voters live in separate online bubbles, MIT analysis shows. *VICE NEWS*, https://www.vice.com/en_us/article/d3xamx/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows

Toriumi, F.（2020, July 8）. 2020 tochijisen de koike tochiji heno ouen message ga Twitter jyou ni hotondo nkatta ken（There were almost no messages of support for Governor Koike on Twitter in the 2020 governorship race.）*Yahoo! News*, https://news.yahoo.co.jp/byline/toriumifujio/20200708-00186958

Twitter, Inc.（2021）*Civic integrity policy*, https://help.twitter.com/en/rules-and-policies/election-integrity-policy

Tsakalidis A., S. Papadopoulos, A. I. Cristea & Y. Kompatsiaris（2015）Predicting elections for multiple countries using Twitter and polls. *IEEE Intelligent System* 30（2）,pp.10-17. DOI: 10.1109/MIS.2015.17

Uenohara, H.（2014）2013nen sangiinsenkyo ni okeru social media［Social media in the 2013 House of Councillors election］, *Senkyo kenkyu*, 30（2）: pp.116-128, https://doi.org/10.14854/jaes.30.2_116

Wollebæk, D., R. Karlsen, K. Steen-Johnsen & B. Enjolras (2019) Anger, fear, and echo chambers: the emotional basis for online behavior. *Social Media + Society* 5(2). https://doi.org/10.1177/2056305119829859

Wang H., D. Can, A. Kazemzadeh et al. (2012) *A system for real-time Twitter sentiment analysis of 2012 US Presidential election cycle*. In: Proceedings of the 50th annual meeting of the association for computational linguistics, pp.115-120, https://aclanthology.org/P12-3020.pdf

Yamaguchi, S. (2015) Net enjoy no jittai to seisakutekitaiou no kentou: Jissyoubunnseki kara miru syakaiteki eikyou to meiyo kisonzai, seigenteki honnin kakunin seido, Internet literacy kyouiku no arikata (Consideration of the actual situation and policy to online flaming: An empirical analysis of the social impact, and defamation, restrictive identification systems, and Internet literacy education) *Jyouhou tsushin seisaku review, 6*, 52-74

Yamaguchi, S. (2022) Why are there so many extreme opinions online?: An empirical, comparative analysis of Japan, Korea and the USA, *Online Information Review, ahead-of-print*. https://doi.org/10.1108/OIR-07-2020-0310

Yarchi, M., C. Baden & N. K. Vilenchik (2021) Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media, *Political Communication*, 38(1-2), pp. 98-139, https://doi.org/10.1080/10584609.2020.1785067

Yoshimi, K. (2017) 2014nen syuugiingiinsousenkyo kikan tyu no kouhosya ni okeru Twitter no riyoukeikou (Twitter usage trends among candidates during the 2014 general election for the House of Representatives), *Jyouhou tsushin gakkaishi* 34(4), pp.155-165, https://doi.org/10.11430/jsicr.34.4_155