Information Engineering Express
International Institute of Applied Informatics
2024, Vol. 10, No. 2, IEE787

Flaming Participants Detection Using Account and Stylistic Characteristics from SNS

Taisei Aoyama * , Linshuo Yang † , Daisuke Ikeda ‡

Abstract

Recent years have seen a rapid increase in social media network (SNS) users due to their swift growth. Consequently, people can easily engage in interactive communication with a vast and undefined audience. This has given rise to a recurring phenomenon called “flaming”, in which critical comments flood SNS. While various studies on flaming have been conducted, most of them have primarily focused on individuals receiving significant volumes of critical comments, rather than those who compose them, referred to as “flaming participants”. In this study, we examine the characteristics of flaming participants on Twitter (Although the name has now been changed to “X”, this paper still uses “Twitter” as its name) by using machine learn-ing to classify them into two groups: flaming participants and normal users. For the classification features, we utilize account information, i.e., statistical data for each account, and stylistic features of the postings, i.e., (1, n)-grams of the part-of-speech tags of the postings. Our experimental findings underscore the effectiveness of these features in identifying Twitter’s flaming participants. Additionally, our research re-veals that flaming participants tend to employ quote tweets more frequently than typical users, and there are distinctive word patterns observable among flaming par-ticipants.

Keywords: document classification, flaming, n-gram, Twitter

1 Introduction

In recent years, the swift growth of social networking sites (SNS) has led to a substantial surge in SNS users, enabling individuals to engage in interactive communication. Consequently, this has significantly accelerated the spread of information, granting anyone the means to effortlessly share information with a global audience. With these societal shifts, the occurrence of “flaming”, characterized by a surge of severe comments on social networking service (SNS) posts, has become increasingly common. The number of flaming has been increasing in recent years [1] [2] [3].

* Kyushu University, Fukuoka, Japan
† Kyushu University, Fukuoka, Japan
‡ Kyushu University, Fukuoka, Japan

Statements made by famous individuals often spark heated debates, triggering the flaming phenomenon. For example, harsh comments rushed to J. K. Rowling when she commented on the transgender community [4]. Although there is no universally accepted definition of flaming, Yamaguchi [5] defines it as "the phenomenon of a flood of harsh comments on social media about what a person has said or done", and this definition will be used in this paper.

In addition, we use the following terms in this paper (see Fig. 1): A *flamer* makes a statement that invites criticism and causes flaming; Then many *contributors* post comments in response to framer's statement; We call some contributors *flaming participants* if their posts about the statement are harsh.

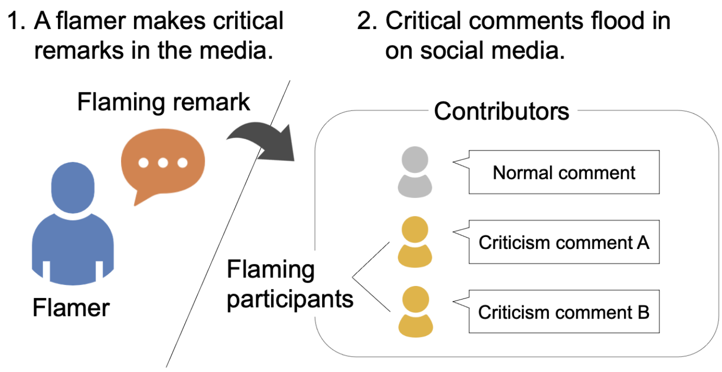

Figure 1: Overview of concepts related to flaming

Flaming's rapid spread makes it challenging to control, as harsh comments escalate quickly. Typically, flamers issue apologies or delete their accounts and often face slander and defamation, which can be mentally taxing. In fact, there have been cases of suicide linked to flaming [6]. Flaming also contributes to the decline in information sharing as it involves slander and relentless attacks, stifling productive discourse. Consequently, those unable to withstand such assaults abandon information dissemination, leading to the cyber cascade problem [7]. This problem means reduced exchange of diverse opinions on the internet, with people of extreme views only engaging with like-minded individuals, leading to even more extreme opinions.

With the increase in the number of flaming in recent years, numerous investigations into this phenomenon have been undertaken. However, most studies have focused on flamers and overlooked the flaming participants. Flaming results from the interaction of both groups, making it essential to understand flaming participants. In addition, it's worth noting that most of the attributes used in Yamaguchi's study may not be known without conducting a survey, and a few can be obtained only on platforms such as Twitter or Instagram. Therefore, drawing comparable conclusions in SNS studies is uncertain, offering potential for new insights into flaming participants.

In this study, we focus on the characteristics of flaming participants on Twitter, using machine learning to classify users into two groups: flaming participants and normal users. We employ both account information and post stylistic features. As a feature representing the stylistic characteristics, we use the (1, n)-gram of the part-of-speech tags instead of the (1, n)-gram of words. Because employing the word (1, n)-gram as the feature makes it impossible to extract consistent features for each user since the topic varies from post to post and thus various words are mixed together.

Our study offers two main contributions. First, we identified characteristics of Twitter's flaming participants and their corresponding posts, filling in gaps in existing research. In addition to the results in our previous article [8], we carried out an additional experiment to examine the combined effects of these two identified characteristics. Second, we compiled a dataset containing account information for both flaming participants and normal users. Currently, there exists no publicly accessible dataset specifically dedicated to Twitter flaming. We manually identified flaming participants and gathered their account data while also collecting data from normal users to prevent bias. We plan to release this dataset to the public soon.

The remainder of the paper is as follows: Section 2 briefly introduces related works. In Section 3, data collection and experiments are explained. Section 4 concludes the whole paper and points out the future work.

2 Related Work

In this section, we will provide an overview of previous studies on flaming and document classification. First, regarding the studies on flaming, we introduce two types of studies: flamers and flaming participants in Section 2.1. Next, regarding the studies on document classification, we introduce a study that proposed a document classification method based on content-independent stylistic features in Section 2.2.

2.1 Flaming

Iwasaki et al. [9] proposed a method for predicting SBCV type flaming on Twitter. This type of flaming arises when an opinion diverges significantly from public sentiment. They posited that flaming occurs when a user's viewpoint on a topic contradicts prevailing public opinions. To achieve this, they introduced an indicator to represent public sentiment and employed a decision tree to classify tweets as flaming or non-flaming. The method exhibited high accuracy in classification, underscoring the indicator's effectiveness. Notably, the decision tree did not rely on Twitter-specific features (number of followers, average number of retweets, etc.), suggesting its potential applicability to other media platforms.

Rajapaksha et al. [10] proposed a method introduced a method for detecting flaming through deep learning-based emotion classification. They used a deep learning model to classify the sentiment of comments on posts from three popular news media on Facebook (BBCNews, CNN, and FoxNews). They demonstrated that flaming can be detected by examining posts with a high number of negative comments. Word2Vec was used for word embedding, and their classification model was composed of three convolutional layers and one Bi-LSTM layer. The accuracy of the classification was 85%.

In contrast, Yamaguchi [5] investigated flaming and flaming participants using a questionnaire survey of Internet monitors. The study revealed an increase in flaming incidents, with only a small fraction of respondents admitting to engaging in flaming more than once. Furthermore, it found a higher likelihood of flaming involvement among individuals who spent more time on social networking services and had higher incomes. Although there have been few empirical studies on flaming, it made various findings on the reality of flaming and flaming participants. It's worth noting that while Yamaguchi's study used data obtained through questionnaires, the present study utilizes data obtained directly from Twitter.

2.2 Document Classification

In this study, we also classify flaming participants and normal users by stylistic characteristics of their posts. Consequently, we will also provide an overview of a previous study on document classification.

Baba [11] introduced a method aimed at classifying documents by stylistic features independent of the content of the text. Content words were converted to part-of-speech tags, and word (1, n)-grams were obtained from each document, which were used as a feature. SVM served as the classifier. The effectiveness of this method was demonstrated to these three tasks: citation count prediction, native language identification, and mental health prediction. The superiority of this method lies in its ability to classify documents without considering their content. In this study, we adopt this method because we want to classify each user according to stylistic features that are consistent across users, independent of the content of their posts.

As mentioned in Section 1, there are many studies on the flamers [9] [10], but there are few studies on flaming participants [5]. Yamaguchi's study [5] offered insights into the nature of flaming participants, revealing that they are a small fraction of users and have a characteristic profile. In this study, we hypothesize that some characteristics of flaming participants appear on the Twitter platform as well, and we endeavor to classify them by using stylistic characteristics of posts and account information as feature values.

3 Experiment

In this study, we examine the characteristics of flaming participants on Twitter from two perspectives: their account information and stylistic features of their posts. To facilitate our analysis, we constructed a dataset comprising account information and posts of both flaming participants and normal users. With the dataset, we conducted two experiments: classification by account information and stylistic features of posts. This chapter describes the data collection methods used in the experiments and the details of the experiments.

3.1 Data Collection

The primary objective of this study is to identify the characteristics of flaming participants and to gain new knowledge about flaming. Achieving this goal necessitates the initial step of identifying and gathering data on flaming and flaming participants because there are no publicly available datasets specific to flaming. This section describes the methods employed for data collection.

Twitter API [12] was used to collect account information. Account information encompasses information about the profile and settings of a Twitter account, including the username, number of followers, and account creation date. These data points can be obtained using the Twitter API. The following is the specific methodology employed for collecting data of flaming participants and normal users.

In order to identify the flaming participants, it is necessary to identify the cases of the flaming at first. We identified flaming cases by referring to Togetter[1]. Many articles focused on flaming incidents have been published on Togetter, often introducing the accounts of flamers along with critical comments in response. Subsequently, from the accounts posting such replies, we collected account information on users who sent replies that met one of the following conditions as flaming participants.

- Users write outbursts for the purpose of offending the flamer, with little reference to the content of the flaming.
- Users who interact with other users without listening to other opinions, but sticking to their own opinions, ranting and raving.

We obtained account information of 100 flaming participants in this way. Table 1 presents examples of user classification in a flaming case based on these conditions. For example, the user posted "Eat your lipstick, ugly bitch." was categorized as a flaming participant because the contributor did not express his/her own opinion on the flaming but instead used the word "ugly bitch" to make the flamer feel uncomfortable. While, the post "You should consider the feelings of the restaurant." was judged not to be a contributor to the flames because it mentioned the contents of the flaming and expressed its own opinion without using abusive words.

Table 1: Examples of user classification in a flaming case

Flaming case	
A female beautician made a statement on Twitter in favor of cancellations at restaurants without notice and received a large number of critical comments.	
Flaming Participant	**Other**
"Eat your lipstick, ugly bitch." "You idiot" "Your brain is beautiful! So slippery!"	"You should consider the feelings of the restaurant." "If you're going to expose your insanity, you might as well not be on Twitter. It's just embarrassing."

These are the author's translations of Japanese tweets.

Next, we explain how we collected data of normal users. First, 500 Japanese language tweets were randomly obtained every hour for 24 hours to prevent potential bias based on the time of the tweet. Next, we obtained the account information of the user who made each tweet. From the pool of 12,000 accounts obtained in this way, 100 were finally utilized in the experiment Accounts meeting the following conditions were excluded as unsuitable for the experiment:

- Official corporate accounts,
- BOT accounts,
- Accounts whose tweets are mostly due to the auto-tweet.

[1] `https://togetter.com/`

3.2 Classification by Account Information

In this part, we will elucidate the process of classifying Twitter users into flaming participants and normal users using the user's account information. Table 2 shows the features employed for classification.

Follows, followers, and follower_per_follow are adopted to represent the size of the community of accounts. Likes and tweets are adopted to represent the account's activity level, while retweet_ratio, quote_ratio, and reply_ratio are used to represent the form of the posts. N_liked_avg, n_retweeted_avg are adopted to represent the influence to the post.

Table 2: Features used for classification by account information

Features	Explain
follows	Number of followings
followers	Number of followers
likes	Number of likes
tweets	Number of tweets
retweet_ratio	Percentage of retweets in the total of tweets
quote_ratio	Percentage of quotes in the total of tweets
reply_ratio	Percentage of replies in the total of tweets
n_liked_avg	Average number of likes earned per tweet divided by the number of followers
n_retweeted_avg	Average number of retweets earned per tweet divided by the number of followers
follower_per_follow	Number of followers per following

These features were used for classification using machine learning models, specifically Linear SVM and random forest. Each hyperparameter was determined by grid search. The search range for grid search is as follows: n_estimators of random forest (10, 11, ..., 19), max_depth of random forest (5, 6, ..., 14), and C of SVM (10, 20, ..., 70). Stratified 5-fold cross validation was performed using 100 users for both flaming participants and normal users.

Table 3 provides the mean values for each indicator from the 5-fold cross validation. The hyperparameters that yielded the best classification accuracy are n_estimators (17), max_depth (13), and C (60). The results demonstrate the effectiveness of account information in classifying flaming participants and normal users.

Table 3: Classification results by account information

Model	Precision	Recall	F1
Linear SVM	0.64	0.60	0.62
Random Forest	0.67	0.66	0.67

Mann-Whitney U-test was performed for each feature to validate the features that were effective in classification. Mann-Whitney U-test is a nonparametric test used for two uncorrelated groups. In this context, it was employed to ascertain whether there were significant differences in the representative values of the 10 features used in the experiment between the two user groups: flaming participants and normal users. Table 4 shows the results of the test. Among the 10 features, significant differences were observed in five features: quote_ratio, reply_ratio, n_liked_avg, n_retweeted_avg, and follower_perfect_follow. The median value of the "follower_per_follow" was higher for normal users, and the other four features were higher for flaming participants.

Table 4: Results of Mann-Whitney U-test

feature	median (flaming) (participants)	median (normal users)	p-value	
follows	281.000	282.000	0.9659	
followers	218.500	163.000	0.4172	
likes	10400.000	9315.000	0.7259	
tweets	6354.000	7739.500	0.4208	
retweet_ratio	0.208	0.257	0.5124	
quote_ratio	0.059	0.025	0.0024	*
reply_rato	0.274	0.153	0.0130	*
n_liked_avg	0.005	0.002	0.0031	*
n_retweeted_avg	0.000	0.000	0.0000	*
follower_per_follow	0.513	0.808	0.0442	*

* $p < 0.05$

Based on these results, we checked the tweets of the flaming participants. The tweets of the flaming participants showed that they usually responded to the comments of others who had not been flamed. Furthermore, we observed that they often quote and retweet to express their own opinions. Many of them comment in a negative tone that make the listener feel uncomfortable, such as "-しろ (imperative form)," "-だろ (high-pressure tone)," "4 ね (Japanese slang for the imperative form of "die")," and "きもい (meaning gross)." Flaming participants frequently directed their quotations and replies towards tweets from politicians or news media. This aligns with the findings of with the results of previous studies [5]: people who believe it is acceptable to strongly criticize others on the Internet are more likely to participate in flaming than those who do not think so.

Rather than using it primarily for sharing personal daily life events, flaming participants leverage Twitter as a powerful tool to assert their opinions on contentious and controversial topics such as politics and discrimination against women. However, we found that the content of these posts was often emotional and harsh in tone, giving the impression that the perpetrators wanted to persist in their opinions at all costs.

3.3 Classification by Stylistic Features of Posts

This part describes the classification using stylistic features of posts. This experiment consists of two parts: ngram-based classification of non-polarized/polarized part-of-speech tags in posted sentences.

In this experiment, we gathered 1,000 posted sentences from each user and converted them into part-of-speech tag sequences. Then, the TF-IDF of those (1, n)-grams were used as features. In other words, 1,000 tweets for one person corresponds to one document. The pre-processing steps are detailed below. First, URLs, emoticons, and emojis are replaced with special tags during the text cleaning process. Then, morphological analysis were performed on the cleaned text using MeCab[2] as the engine with mecab-ipadic-NEologd as the dictionary, and each morpheme was replaced with a part-of-speech tag such as [noun], [verb], etc.

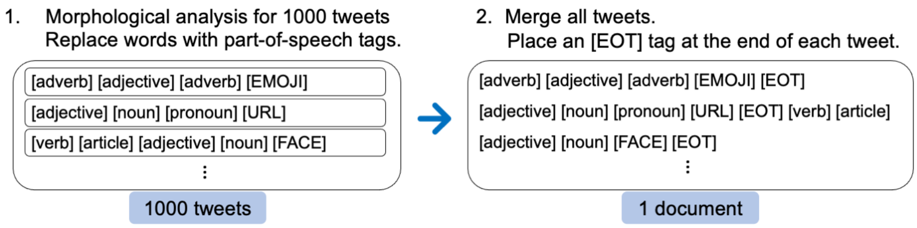

Figure 2: Overview of pre-processing for one user

In cases where polarity was assigned, referring to the polarity dictionary [13] [14], we assigned p to positive words and n to negative ones, e.g. [n-noun]. Subsequently, all of the processed sentences were combined into a single document for each user, and part-of-speech tag sequences were created for the number of users. And an [EOT] tag was appended at the end of each tweet to indicate tweet boundaries. An overview of the preprocessing for one user is presented in Fig. 2.

The TF-IDF of (1, n)-grams of part-of-speech tags was computed from the part-of-speech tag sequence obtained by preprocessing, and was used as a feature. Linear SVM was used for the classification model, and the model was evaluated by the 5-fold cross validation. The hyperparameters were determined by grid search. The search range for hyperparameter C is 10, 20, ..., 70. The number of data used is 100 for both flaming participants and normal users.

Table 5 presents the outcomes when non-polarized part-of-speech tags are used, and Table 6 shows those when polarized part-of-speech tags are used. The value of each indicator is the mean value in the 5-fold cross validation process. These also show the hyperparameter C corresponding to the highest F1-score. The results show that classification using polarized part-of-speech tags for all n has higher F1 values than non-polarized.

In an additional experiment, the classification was conducted using both account information and polarized part-of-speech tags as features. The experimental setup is the same as in the previous experiment, except for the features. Table 7 shows the results of this classification and also the hyperparameter C corresponding to the highest F1-score. Although the results were expected to be more accurate than using polarized part-of-speech tags, the F1-score was lower for each n.

To gain a deeper understanding of these results, we examined the features that significantly impacted the classification. SVM computes weights for each feature during training. Larger weights indicate features that are more characteristic of

[2] https://taku910.github.io/mecab/

flaming participants, while smaller weights indicate features more characteristic of normal users.

Table 5: Classification results using (1, n)-gram of non-polarized part-of-speech tags

	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
Precision	0.67	0.68	0.68	0.70	0.70	0.68
Recall	0.78	0.78	0.76	0.74	0.74	0.73
F1	0.72	0.72	0.71	0.72	0.72	0.70
C	60	30	40	50	40	40

Table 6: Classification results using (1, n)-gram of polarized part-of-speech tags

	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
Precision	0.78	0.81	0.82	0.85	0.80	0.82
Recall	0.70	0.69	0.69	0.72	0.70	0.67
F1	0.73	0.74	0.75	0.78	0.74	0.73
C	60	30	30	20	20	30

Table 7: Classification results using the account information and (1, n)-gram of polarized part-of-speech tags

	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
Precision	0.77	0.79	0.84	0.85	0.86	0.85
Recall	0.64	0.65	0.65	0.64	0.63	0.64
F1	0.70	0.72	0.73	0.73	0.72	0.73
C	50	60	70	60	70	60

First, we introduced features that significantly affected the classification using non-polarized part-of-speech tags. Table 8 shows a description of tag name abbreviations. Table 9 shows the features with the highest weights for the non-polarized models generated during cross validation with ¥ngram4. Notably, it was observed that the weights of pattern [verb] [a.v.] [noun] are generally greater in the second, third, and fourth validation. The phrases that follow this pattern include "-したこと (used to ask about experience)", "-したわけ (used to deny)", and "-した方 (used to recommend)". Also, [nouns] [a.v.] [p.p.] appear in common in the third, fourth, and fifth validation. Examples of this pattern include "-だから (meaning 'because')", "-なのに," "んだけど (meaning like 'but')", "-ですか (interrogative form)", "ですよね", "んだ (used at the end of a sentence)", etc. We confirmed that tweets containing such phrases express that they are questioning the other person or are dismayed about something.

Next, we introduced the features that significantly affected classification by polarized part-of-speech tags. Table 10 shows the features with the highest weights for the polarized models generated during cross validation with (1, 4)-gram. The

Table 8: Description of tag name abbreviations

Tag name	Explain
p.p.	Postpositional Particle
a.v.	Auxiliary Verb
pre-n.a	Pre-Noun Adjectival

Table 9: Features with high weights for each training of the 5-fold cross validation with (1, 4)-gram in classification by non-polarized part-of-speech tags

	1st	2nd	3rd
1st	[EOT] [noun] [p.p.]	[EOT] [noun]	[noun] [p.p.] [noun]
2nd	[verb] [a.v.] [noun]	[noun] [a.v.]	[pre-n.a.]
3rd	[noun] [a.v.]	[verb] [a.v.] [noun]	[noun] [a.v.] [p.p.]
4th	[a.v.] [p.p.]	[noun] [a.v.] [p.p.]	[verb] [a.v.] [noun]
5th	[noun] [a.v.]	[noun] [noun] [noun] [noun]	[noun] [a.v.] [p.p.]

result underscore the significance of [n-noun]. Other top-ranked items included [n-noun] such as [p.p.][n-noun] and [n-noun][p.p.]. The weight of [n-noun] was about two to three times greater than the ones of the second and third places, indicating that [n-noun] were considerable important in the classification. There are various types of [n-noun] words, such as "性犯罪 (meaning 'sex crime')", which conveys an incidental nature, and "ガキ (abusive use of 'kid')", which represents a malicious alteration of an existing word. The tweets containing these words were considered to express their own complaints and critical references to politics and crime.

Table 10: Features with high weights for each training of the 5-fold cross validation with (1, 4)-gram in classification by polarized part-of-speech tags

	1st	2nd	3rd
1st validation	[n-noun]	[noun] [noun] [URL] [EOT]	[p.p.] [n-noun]
2nd validation	[n-noun]	[noun] [a.v.] [p.p.]	[p.p.] [n-noun]
3rd validation	[n-noun]	[p.p.] [n-noun]	[noun] [a.v.] [p.p.]
4th validation	[n-noun]	[p.p.] [n-noun]	[a.v.] [p.p.]
5th validation	[n-noun]	[p.p.] [n-noun]	[n-noun] [p.p.]

In conclusion, it was found that there are characteristic word patterns exist in the posting of the flaming participants, and that these posts often contain a significant number of negative nouns.

In the non-polarized model, phrases commonly used to express critical opinions (e.g., "-なのに", "-した方が") and phrases that express a coercive attitude (e.g., "-んだよ", "-んだから", "-んだけど") at the end of a sentence are considered important. The polarized model placed considerably more importance on [n-noun], or 1-grams, than on the other n-grams, indicating that simply the usage rate of negative words had a significant impact on the classification.

4 Conclusion

In this study, we classified Twitter users into two groups: flaming participants and normal users based on their account information and stylistic post features. We then analyzed the traits of Twitter's flaming participants. For classification by stylistic features, TF-IDF of (1, n)-gram of non-polarized/polarized part-of-speech tags was used.

Account information classification achieved an F1-score of 0.67, indicating some level of feasibility. U-tests revealed that flaming participants tend to quote retweets and reply more frequently than normal users, suggesting they tend to express their opinions directly.

Stylistic feature classification outperformed account information with an F1-score of 0.72 for (1, n)-gram of non-polarized and 0.78 for (1, n)-gram of polarized part-of-speech tags. The results underscored that flaming participants' posts exhibited distinctive word patterns and more negative language compared to normal users.

This study has filled an important gap in research by examining the characteristics of Twitter accounts belonging to flaming participants, which hasn't been thoroughly explored before. These characteristics include unintentionally performed ones, which cannot be detected by questionnaires. The findings of this research can contribute to future research on flaming and the development of preventive systems.

To streamline future data collection and morphological analysis for future research, two key tasks are required. Firstly, manual data collection was employed in this study, but to enhance efficiency, automation methods, like those introduced by Rajapaksha et al. [10], should be explored. Secondly, Twitter's usage of new words and informal language poses challenges for accurate morphological analysis. Therefore, future tasks are needed to improve the dictionary such as adding new words and Twitter-specific expressions.

5 Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 23K28149.

References

[1] Siemple Digital Crisis Research Institute. Study Session Implementation Report "Summarizing Flaming Cases in 2020: Trends and Countermeasures for 2021 Learned from Research Data and Case Analyses" (in Japanese), 2020. `https://dcri-digitalcrisis.com/report/houkoku/studygroup201125/` (February 2, 2022).

[2] PR TIMES. Siempre Digital Crisis Institute Releases Second Annual "Digital Crisis White Paper 2021" (in Japanese), 2021. `https://prtimes.jp/main/html/rd/p/000000059.000052269.html` (February 2, 2022).

[3] PR TIMES. In 2021, there were 1,766 outbreaks of flaming, a 24.8 `https://prtimes.jp/main/html/rd/p/000000129.000052269.html` (February 2, 2022).

[4] The Cable News Network. Trans activists call J.K. Rowling essay 'devastating', 2020. https://edition.cnn.com/2020/06/11/uk/jk-rowling-trans-harry-potter-gbr-intl/index.html (June 11, 2020).

[5] Shinichi Yamaguchi. An Empirical Analysis of Actual Examples of "Flaming" and Participants' Characteristics (in Japanese). *The Journal of the Institute of Information and Communication Engineers*, 33(2):53–65, 2015.

[6] Mainichi Japan. Japan netizens, celebrities, experts blast online slander after wrestler Hana Kimura's death, 2020. https://mainichi.jp/english/articles/20200525/p2a/00m/0na/009000c (May 25, 2020).

[7] Tatsuo Tanaka and Shinichi Yamaguchi. *Research on Internet Flaming: Who stirs them up and how to deal with them? (in Japanese).* Keiso Shobo, 2016.

[8] Taisei Aoyama, Linshuo Yang, and Daisuke Ikeda. Detection of Flaming Participants Using Account Information and Stylistic Features of Posts. *the 12th International Congress on Advanced Applied Informatics (IIAI-AAI), the 14th International Conference on E-Service and Knowledge Management (ESKM 2022)*, pages 49–54, 2022.

[9] Yuki Iwasaki, Ryohei Orihara, Yuichi Sei, Hiroyuki Nakagawa, Yasuyuki Tahara, and Akihiko Ohsuga. Analysis of Flaming and Its Appliacations in CGM (in Japanese). *Transactions of the Japanese Society for Artificial Intelligence : AI*, 30(1):152–160, 2015.

[10] Praboda Rajapaksha, Reza Farahbakhsh, Noël Crespi, and Bruno Defude. Uncovering Flaming Events on News Media in Social Media. In *2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC)*, pages 1–8, Los Alamitos, CA, USA, Oct 2019. IEEE Computer Society.

[11] Takahiro Baba. *Document Classification Using Non-content Features.* PhD thesis, Kyushu University, 2021. http://hdl.handle.net/2324/4475147.

[12] Twitter API. https://developer.twitter.com/en/docs/twitter-api (January 23, 2022).

[13] Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. Collecting Evaluative Expressions for Opinion Ectraction (in Japanese). *Journal of Natural Language Processing*, 12(3):203–222, 2005.

[14] Masahiko Higashiyama, Kentaro Inui, and Yuji Matsumoto. Learning Sentiment of Nouns from Selectional Preferences of Verbs and Adjectives (in Japanese). *Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing*, pages 584–587, 2008.