Detection of Flaming Participants Using Account Information and Stylistic Features of Posts

Aoyama, Taisei
Graduate School of Information Science and Electrical Engineering, Kyushu University

Yang, Linshuo
Graduate School of Information Science and Electrical Engineering, Kyushu University

Ikeda, Daisuke
Faculty of Information Science and Electrical Engineering, Kyushu University

https://hdl.handle.net/2324/6781035

出版情報：2022 12th International Congress on Advanced Applied Informatics IIAI-AAI 2022 Proceedings, pp.49-54, 2022-07. IEEE
バージョン：
権利関係：

Detection of Flaming Participants Using Account Information and Stylistic Features of Posts

Taisei Aoyama
Graduate School of Information Science and Electrical Engineering
Kyushu University
Fukuoka, Japan
aoyama.taisei.837@s.kyushu-u.ac.jp

Linshuo Yang
Graduate School of Information Science and Electrical Engineering
Kyushu University
Fukuoka, Japan
yang.linshuo.096@s.kyushu-u.ac.jp

Daisuke Ikeda
Faculty of Information Science and Electrical Engineering
Kyushu University
Fukuoka, Japan
ikeda.daisuke.899@m.kyushu-u.ac.jp

***Abstract*—With the rapid development of SNS in recent years, the number of SNS users has increased rapidly, and people can easily communicate interactively with an unspecified large number of people. With these changes in the information society, a phenomenon known as "flaming", in which critical comments flood SNS, has become a frequent occurrence. In recent years, various studies on flaming have been conducted, but most of them are concerned with those who receive a large number of critical comments, not on those who write critical comments, called "flaming participants". In this study, we examine the characteristics of flaming participants on Twitter by using machine learning to classify them into two groups: flaming participants and normal users. For the classification features, we use 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. The experimental results show that these features are effective in detecting Twitter flaming participants. Furthermore, we found that flaming participants use more quote tweets than the normal user, and that there are word patterns that are characteristic of flaming participants.**

***Index Terms*—flaming, Twitter, n-gram, document classification**

I. Introduction

With the rapid development of SNS in recent years, the number of SNS users has increased rapidly, and people can easily communicate interactively with an unspecified large number of people. This has increased the diffusion of information, allowing anyone to easily disseminate information to the world.

With these changes in society, the phenomenon of "flaming," in which critical comments flood in on SNS posts, has become a frequent occurrence. The number of flaming has been increasing in Japan in recent years [1]–[3]. Although there is no established definition of flaming, Yamaguchi [4] defines it as "the phenomenon of a flood of critical comments on social media about what a person has said or done," and this definition will be used in this paper.

In addition, the following terms are used for convenience in this paper.

- Flamer: A person who makes a statement that invites criticism and causes flaming.
- Contributor: A person who posts a comment in response to a statement made by a flamer.
- Flaming Participant: A contributor posting critical comments

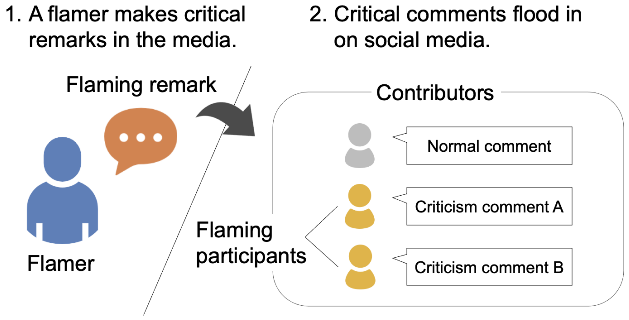

Fig. 1. Overview of concepts related to flaming

These concepts are summarized in Fig. 1.

Flaming are difficult to calm down due to their high spreading power, since critical comments continue to increase once they occur. In many cases, the flamers are eventually forced to post apologies, and even have their accounts deleted. Furthermore, the flamer will also be exposed to a large amount of slander and libel. This can be mentally taxing. In fact, there was a case in which a person was driven to suicide because of flaming [5]. In addition to the mental burden on the flamer, the problem of flaming is the atrophy of information dissemination. Flaming often involve slander and relentless attacks, with little productive discussion. Therefore, those who feel that they cannot withstand such attacks give up disseminating information. In addition, flaming have the problem of causing cyber cascade [6]. Cyber cascade is a phenomenon in which different positions on the Internet become less interchangeable. It occurs when those with middle-of-the-road opinions become fed up with the flaming and give up disseminating information, causing those with extreme opinions to exchange opinions only among those who share the same opinions. When it occurs, opinions are likely to intensify and arguments are likely to degrade.

With the increase in the number of flaming in recent years, various studies on flaming have been conducted. Yamaguchi [4] conducted a questionnaire survey of Internet monitors to test hypotheses about the actual conditions of flaming and the attributes of those involved in them. Iwasaki et al. [7] found that a feature that quantifies the public opinion of a topic is effective for predicting some Twitter flaming. Rajapaksha et al. [8] used an emotion analysis deep learning model to classify comments on official Facebook accounts of news media with respect to emotion and showed that flaming can be detected based on the number of negative comments of posts. All the studies except Yamaguchi's focused on the flamers [7], [8]. Few studies have focused on the flaming participants. Flaming is a phenomenon caused by the presence of both flamers and flaming participants. Therefore, in order to deepen our understanding of flaming, it is also important to know well about flaming participants. In addition, most of the attributes used in Yamaguchi's study are not known without conducting a survey, and few can be obtained only on platforms such as Twitter or Instagram. Therefore, in the study on SNS, it is not certain that similar conclusions can be reached, and new findings on flaming participants may be obtained.

In this study, we investigate the characteristics of Japanese flaming participants on Twitter by using machine learning to classify Twitter users into two groups: flaming participants and normal users. For the classification, we use account information and stylistic features of the postings. As a feature representing the stylistic characteristics, we use the (1, n)-gram of the part-of-speech tags, not the (1, n)-gram of words, which is often used for the (1, n)-gram. Using 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. To solve this problem, we use the (1, n)-gram of the part-of-speech tags as a feature independent of the content of the postings.

Our main contributions of this study are twofolds.

Firstly we discovered characteristics of flaming participants on Twitter. While various studies have been conducted on flaming, there are no studies that show a characteristic tendency among flaming participants in the information obtained from Twitter. In this study, we used the information obtained from Twitter to discover trends specific to flaming participants in account information and stylistic characteristics of posts. It is expected that the new findings presented in this study will be incorporated into future flaming research to further deepen our knowledge of flaming.

Secondly we constructed a dataset of account information for both flaming participants and normal users. Currently, there is no publicly available dataset on Twitter flaming. Therefore, in this study, we manually identified the flaming participants and collected the account information of the flaming participants. For normal users, we collected data in an effort to avoid bias toward a group of accounts with specific characteristics, and constructed a dataset. The authors will make the dataset publicly available in the near future. In many studies on flaming, a dataset had to be created by each individual and thus it is very costly. By releasing this dataset, it is expected to facilitate the research on flaming participants.

II. Related Work

In this section, we present previous studies on flaming and document classification. First, regarding the studies on flaming, we introduce two types of studies: flamers and flaming participants in II-A. Next, regarding the studies on document classification, we introduce a study that proposed a document classification method based on content-independent stylistic features in II-B.

A. Flaming

Iwasaki et al. [7] proposed a method to predict SBCV type flaming on Twitter. SBCV type flaming is the flaming caused by an evaluation that is out of line with public opinion. They assumed that a flaming occurs when the speaker's opinion on a topic conflicts with public opinions. They introduced a indicator representing public opinion, and classified whether the target tweet was a flaming tweet or a non-flaming tweet using a decision tree. The accuracy of the classification by the proposed method is high, indicating that the indicator is effective for the prediction of SBCV type flaming. Furthermore, the nodes of the generated decision tree do not contain Twitter-specific features (number of followers, average number of retweets, etc.), indicating that it may be applicable to other media as well.

Rajapaksha et al. [8] proposed a method for flaming detection using deep learning emotion classification. They used a deep learning model to classify the sentiment of comments on posts from three popular news media on Facebook (BBC-News, CNN, and FoxNews), and showed that flaming can be detected by referring to posts with a high number of negative comments. Word2Vec was used for word embedding, and a model consisting of three convolutional layers and one Bi-LSTM layer was used for classification. The accuracy of the classification was 85%.

In contrast, Yamaguchi [4] tested hypotheses regarding the actual situation of flaming and the attributes of flaming participants using a questionnaire survey of Internet monitors. Verification of the actual status of flaming showed that the number of flaming had increased and that only a small percentage of the respondents had written on the flaming more than once. Furthermore, the results showed that the probability of being involved in flaming was higher for those who spent more time on social networking services and those who had higher annual incomes. Although there have been few empirical studies on flaming, it made various findings on the reality of flaming and flaming participants. While Yamaguchi's study used data obtained through questionnaires, this study uses data obtained directly from Twitter.

B. Document Classification

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

Baba [9] proposed a method for 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 was used as the classifier. The method was applied to three tasks: citation count prediction, native language identification, and mental health prediction. The proposed method was shown to be effective in all these tasks. This method is superior in that it can classify documents without regard to their contents. In this study, we use 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 the introduction, there are many studies on the flamers [7], [8], but there are few studies on flaming participants [4]. Yamaguchi’s study [4] showed that flaming participants 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 classify them by using stylistic characteristics of posts and account information as feature values.

III. Experiment

In this study, we examine the characteristics of flaming participants on Twitter from two perspectives: their account information and stylistic features of posts. We constructed a dataset of account information and posts of both the flaming participants and normal users. Using 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.

A. Data Collection

The purpose of this study is to identify the characteristics of flaming participants and to gain new knowledge about flaming. To this end, it is first necessary to identify and collect data on flaming and flaming participants because there are no publicly available datasets specific to flaming. This section describes data collection methods.

Twitter API [10] was used to collect account information. Account information is information about the profile and settings of a Twitter account, including the user name, number of followers, and account creation date. It can be obtained using the Twitter API. The following is a specific method of collecting data of flaming participants and normal users.

1) Flaming Participants: In order to identify the flaming participants, it is first necessary to identify the cases of the flaming. First, we identified flaming cases by referring to Togetter[1]. Togetter is a website that allows users to quote tweets and publish articles on a topic. Many articles on flaming have been published, in which the accounts of flamers and replies with critical comments are introduced. Next, 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.

TABLE I
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.

- A user writing outbursts with little or no reference to the content of the flaming, and writing for the purpose of offending the flamer.
- A user who interacts with other users but does not listen to the other user’s opinion, but instead sticks to his or her own opinion, ranting and raving.

We obtained the account information of 100 flaming participants in this way. TABLE I shows examples of user classification in a flaming case based on these conditions. For example, the post “Eat your lipstick, ugly bitch.” was judged as a flaming participant because the contributor did not express his/her own opinion on the flaming and 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 flames and expressed its own opinion without using abusive words.

2) Normal Users: Next, we explain how we collect data of normal users. First, 500 Japanese language tweets were randomly obtained every hour for 24 hours. This is to avoid user bias based on the time of the tweet. Next, we obtained the account information of the user who made each tweet. Of the 12,000 accounts collected in this way, 100 were used in the experiment. At that time, accounts meeting the following conditions were excluded as unsuitable for the experiment:

- official corporate accounts,
- BOT accounts, or
- accounts whose tweets are mostly due to the auto-tweet.

B. Classification by Account Information

In this part, we explain the classification of Twitter users into flaming participants and normal users using the user’s account information. TABLE II shows the features used for classification. Follows, followers, and follower_per_follow are employed to represent the size of the community of accounts. Likes and tweets are used 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 employed to represent the influence to the post.

[1] https://togetter.com/

These features were used for classification by machine learning. Linear SVM and random forest were used for the machine learning model. 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 III shows the mean values for each indicator from the 5-fold cross validation. The hyperparameters with the best classification accuracy are n_estimators (17), max_depth (13), and C (60). The results show that the account information is effective in classifying flaming participants and normal users.

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. We tested whether there was a difference in the representative values of the 10 features used in the experiment for the two groups of users, the flaming participants and the normal users. TABLE IV 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 ones of the other four features were higher for flaming participants.

Based on these results, we checked the tweets of the flaming participants. The tweets of the flaming participants shows that they usually responded to the comments of others who had not been flamed. Furthermore, we observed that they often quoted and retweeted, expressing their own opinions. Many of the comments have a negative tone that makes 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)." In many cases, the targets of quotations and replies are tweets from politicians or news media. This is consistent with the results of previous studies [4]: people who believe it is acceptable to use strong tones of accusation on the Internet are more likely to participate in flaming than those who do not.

TABLE II
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

TABLE III
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

TABLE IV
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	*

* $p < 0.05$

From the above, it can be seen that the flaming participants use Twitter not only to post events in their daily lives as a tool to assert their opinion on controversial topics such as politics and discrimination against women. However, the content is emotional, spoken in a harsh tone, and it was found that many of them are to force their own opinions through.

C. 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 collected 1,000 posted sentences per 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. Below we describe these pre-processing. First, we cleaned the text. During this process, URLs, emoticons, and emojis are replaced with special tags. Next, we performed morphological analysis on the cleaned text. MeCab[2] [11] was used as the morphological analysis engine with mecab-ipadic-NEologd [12] as a dictionary, and each morpheme was replaced with a part-of-speech tag: [noun], [verb], etc. In cases where polarity was assigned, referring to the polarity dictionary [13], [14], we assigned p to the tag if the word was positive and n if it was negative, e.g. [n-noun]. Finally, all of the processed sentences are combined into a single document for each user, and a part-of-speech tag sequence is created for the number of users. An [EOT] tag was placed at the end of each tweet to indicate the boundary between tweets. An overview of the preprocessing for one user is shown in Fig. 2.

The TF-IDF of (1, n)-grams of part-of-speech tags is computed from the part-of-speech tag sequence obtained by preprocessing, and is used as a feature. Linear SVM was used

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

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 V shows the results when non-polarized part-of-speech tags are used, and TABLE VI shows the results when polarized part-of-speech tags are used. The value of each indicator is the mean value in the 5-fold cross validation. These also show the hyperparameter C at 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.

To discuss these results, we examined the features that significantly affected the classification. SVM computes weights for each feature during training. The larger the weight, the more characteristic the feature is of the flaming participants, and the smaller the weight, the more characteristic it is of the normal users.

First, we see features that significantly affected classification by non-polarized part-of-speech tags. TABLE VII shows a description of tag name abbreviations. TABLE VIII shows the features with the highest weights for the non-polarized models generated during cross validation with (1, 4)-gram. It can be seen that the weights of [V] [A.V.] [N] are commonly larger in the second, third, and fourth validation. The phrases that fit 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, [N] [A.V.] [P.P.] appear in common in the third, fourth, and fifth validation. Examples of these 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 some of the tweets containing such phrases express that they are questioning the other person or are dismayed about something.

Next, we see the features that significantly affected classification by polarized part-of-speech tags. TABLE IX shows the features with the highest weights for the polarized models generated during cross validation with (1, 4)-gram. The result shows that [n-N] had the highest weight in validations. Other top-ranked items included [n-N] such as [P.P.][n-N] and [n-N][P.P.]. The weight of [n-N] was about two to three times greater than the ones of the second and third places, indicating that [n-N] were given considerable importance in the classification. There are various types of words that fall under [n-N], such as “性犯罪 (meaning sex crime),” which has an incidental nature, and “ガキ (abusive use of “kid”),” which is a malicious change of an existing word. The tweets containing these words were observed to contain many of their own complaints and critical references to politics and crime.

TABLE V
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 VI
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 VII
DESCRIPTION OF TAG NAME ABBREVIATIONS

Tag name	Explain
N	Noun
V	Verb
P.P.	Postpositional Particle
A.V.	Auxiliary Verb
pre-N.A.	Pre-Noun Adjectival

TABLE VIII
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 valid.	[EOT] [N] [P.P.]	[EOT] [N]	[N] [P.P.] [N]
2nd valid.	[V] [A.V.] [N]	[N] [A.V.]	[pre-N.A.]
3rd valid.	[N] [A.V.]	[V] [A.V.] [N]	[N] [A.V.] [P.P.]
4th valid.	[A.V.] [P.P.]	[N] [A.V.] [P.P.]	[V] [A.V.] [N]
5th valid.	[N] [A.V.]	[N] [N] [N] [N]	[N] [A.V.] [P.P.]

From the above, it was found that there were characteristic word patterns in the posts of the flaming participants and that they contained a large number of negative nouns.

The non-polarized model was found to place importance on 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. The polarized model placed considerably more importance on [n-N], or 1-grams, than on the other n-grams, indicating that simply the rate of use of negative words had a significant impact on classification.

IV. CONCLUSION

In this study, we analyzed characteristics of Twitter flaming participants by classifying Twitter users into two groups: flam-

TABLE IX
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 valid.	[n-N]	[N] [N] [URL] [EOT]	[P.P.] [n-N]
2nd valid.	[n-N]	[N] [A.V.] [P.P.]	[P.P.] [n-N]
3rd valid.	[n-N]	[P.P.] [n-N]	[N] [A.V.] [P.P.]
4th valid.	[n-N]	[P.P.] [n-N]	[A.V.] [P.P.]
5th valid.	[n-N]	[P.P.] [n-N]	[n-N] [P.P.]

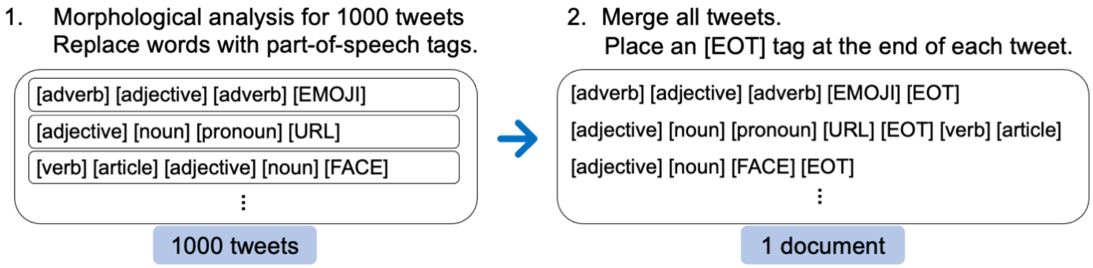
1. Morphological analysis for 1000 tweets
Replace words with part-of-speech tags.
[adverb] [adjective] [adverb] [EMOJI]
[adjective] [noun] [pronoun] [URL]
[verb] [article] [adjective] [noun] [FACE]
⋮
1000 tweets
2. Merge all tweets.
Place an [EOT] tag at the end of each tweet.
[adverb] [adjective] [adverb] [EMOJI] [EOT]
[adjective] [noun] [pronoun] [URL] [EOT] [verb] [article]
[adjective] [noun] [FACE] [EOT]
⋮
1 document

Fig. 2. Overview of pre-processing for one user

ing participants and normal users using account information and stylistic features of their posts. For classification by stylistic features, TF-IDF of (1, n)-gram of non-polarized/polarized part-of-speech tags was used.

The highest F1-score for classification by account information was 0.67, indicating that classification was possible to some extent. The results of U-tests showed that flaming participants had significantly higher rates of quoted retweets, replies, etc. than the normal users. This suggests that flaming participants tend to express their opinions directly to others on a regular basis.

Classification by stylistic features was more accurate than by account information. The highest F1-score was 0.72 for (1, n)-gram of the non-polarized part-of-speech tag and 0.78 for (1, n)-gram of the polarized part-of-speech tag. The results showed that the posts of the flaming participants showed characteristic word patterns on a daily basis, and that they contained more negative words than the normal users.

There are no previous studies on characteristics of the accounts of flaming participants. In this study, we found characteristics of flaming participants on Twitter. These characteristics include those that are unintentionally performed. These unconscious behavior cannot be detected by questionnaires. The findings of this study are expected to be useful for the development of future research on flaming and the construction of a system to prevent flaming in advance.

Two future tasks are to improve the efficiency of data collection and morphological analysis. First, in this study, we had to collect data set by ourselves, and most parts of the data collection of the flaming participants were done manually. It is desirable to automate this process to some extent by using a method to detect flaming cases such as Rajapaksha et al. [8]. Second, because Twitter posts contain many new words and are often informal in style, we found that our method did not correctly sentence structures and analyze words that were not registered in the dictionary. Therefore, future tasks are to improve the dictionary by adding new words and Twitter-specific expressions.

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 24.8the previous year! announcement of the release of the “digital crisis white paper 2022” (in Japanse), 2021. https://prtimes.jp/main/html/rd/p/000000129.000052269.html (February 2, 2022).

[4] 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.

[5] Shukan Josei Prime. One year after the suicide of underground idol tsukino noa, her mother reveals her “dying flame” and the true meaning of her suicide note (in Japanese), 2021. https://www.jprime.jp/articles/-/22049 (January 6, 2022).

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

[7] 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.

[8] 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.

[9] Takahiro Baba. Document classification using non-content features. *ISEE graduate school Kyushu University*, 2021. http://hdl.handle.net/2324/4475147.

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

[11] Kudo Taku, Yamamoto Kaoru, and Matsumoto Yuji. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 230–237, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[12] Toshinori Sato, Taiichi Hashimoto, and Manabu Okumura. Implementation of mecab-ipadic-neologd, a word segmentation dictionary, and its effective use in information retrieval (in Japanese). In *Proceedings of the Twenty-three Annual Meeting of the Association for Natural Language Processing*, pages NLP2017–B6–1. The Association for Natural Language Processing, 2017.

[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.