We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Analysis of Information Spreading by Social Media Based on Emotion and Empathy

Kazuyuki Matsumoto, Minoru Yoshida and Kenji Kita

Abstract

The number of social media users has increased exponentially in recent times, and various types of social media platforms are being introduced. While social media has become a convenient communication tool, its use has caused various social problems. Some users who cannot imagine the emotions their posts may induce in readers cause what is termed as "the flaming phenomenon." In some cases, users intentionally repeat strong remarks for self-advertisement. To identify the cause of this phenomenon, it is necessary to analyze the posted contents or the personalities of the users who cause the flaming. However, it is difficult to reach a generalized conclusion because each case varies depending on the circumstances and individual. In this chapter, we study the phenomenon of information spreading via communication on social media by conducting a detailed analysis of replies and number of retweets in Japanese, and we reveal the relation between the feedback on such posts and the emotions or empathy they result in.

Keywords: social media, information diffusion, flaming, buzz, Twitter

1. Introduction

Recently, opportunities for online communications through social media have increased. One such notable example is that of Twitter, which has a large number of users who use it to advertise products or events or simply to communicate with other users about their hobbies on a real name or handle name basis. More informal communications and business activities are preferably conducted on a real name basis via platforms such as Facebook.

Many Japanese tend use social media freely and easily without identifying themselves. This is thought to be one of the reasons why Twitter has the highest number of users compared to other social media platforms.

Some users think that they can express their opinions freely on social media platforms. Therefore, it is possible that comments about a product or service posted on the social media may not be as honest as those on consumer review sites. Thus, the high anonymity of social media results in many posts that may be less reliable with regard to their sources or contents.

Therefore, it is important to extract and analyze effective information from social media posts in terms of topics such as true/false judgments [1, 2], judgments regarding the negativity/positivity of infectious disease detection [3], and judgments of fake news [4, 5].

In this chapter, we clarify how posts can garner the attention of other users by analyzing the reply texts as they constitute feedback on the posts. Existing studies have focused on the "flaming phenomenon" on social media, proposed a method to extract the "buzz phenomenon," detected influential users for spreading information, etc. We discuss the features these studies focused on and how effective they are at detecting posts that attract other users' attention. Moreover, we analyze the sentiments (positive/negative) based on the contents of the posts and their reply texts using our proposed sensitive analysis algorithm based on natural language text. Finally, we analyze the relation between the sentiment polarity (positive/negative) of the replies and the attention level (the diffusion and popularity) using statistical processing and natural language processing approaches.

2. Related works

2.1 Social media flaming

On Twitter, the content of a post is largely and rapidly diffused using the retweet function. This manner of information diffusion on social media sometimes causes unexpected situations. When a post is critical about a specific person and other users diffuse that post with evil intent, it becomes a trigger for the flaming phenomenon. Several studies have detected and studied the flaming phenomenon [6–8].

Social media flaming is often caused by the social media users' aggressive comments or inappropriate statements. Such comments or statements can be a trigger of cyberbullying on social media.

Rösner et al. [9] published on the spread of information by word-of-mouth communication on the social web. Their study found that users' aggressive expressions tend to be used more often in anonymous environment than non-anonymous condition.

Sri Nandhini and Sheeba [10] proposed a method to detect cyberbullying on social networks. Their method can detect flaming, harassment, racism, and terrorism by using fuzzy logic and genetic algorithm.

2.2 Retweet prediction/buzz detection

Some studies have predicted the degree of information diffusion [11–13]. Recently, studies in the field of information extraction have shown the usefulness of extracting important information efficiently from large-scale web texts. If we can predict to what extent or to what scale a tweet of a certain piece of news or an event on social media would be influential, others too would be able to diffuse information on a large scale intentionally. For example, companies might be able to promote their products by posting tweets, given the great potential for diffusion.

It is widely known that the effective use of a social networking service (SNS) could minimize the cost of advertising and successfully promote a product or service. Thus, an increasing number of companies are implementing a strategy called buzz marketing to increase future sales.

Many studies have predicted users who can cause the buzz phenomenon or the period during which a buzz phenomenon tends to occur. Murakami and Suzuki [14] used information on the number of retweets to predict information diffusion. On Twitter, most information diffusions are caused by retweeting. Murakami and Suzuki modeled users' interests by analyzing the word distribution in the retweets of each user. Saito et al. [15] analyzed the transition of information diffusion behaviors using the asynchronous independent cascade model.

These methods analyzed information diffusion. However, they could not answer a fundamental question; why were these tweets diffused in the first place? If the rules governing the buzz phenomenon are clarified, many companies would benefit via increased sales. Thus, it is important to examine the feedback from users who diffused the original tweet to clarify the cause for its diffusion.

Buzz marketing aims to increase the overall demand for marketed products or services, while viral marketing disseminates information about the product or service via word-of-mouth information.

2.3 Influencer

The information diffusers mostly consist of the followers of the user, who retweet the user's tweet or bookmark it as a favorite. Therefore, when a user has more influential followers, his/her tweets have a higher probability of being diffused. We sometimes call such an influential user an influencer. Some services have analyzed influencers or the degree of their influence based on the number of their followers or the frequency of their tweets [16–18].

Liao et al. [19] proposed an influence measurement method called WeiboRank for other social media. Matsuo and Yasuda [20] analyzed the relations the users leading each community construct on Mixi, a Japanese SNS. Tsugawa and Kimura [21] proposed a method to identify influencers from sampled social networks on Facebook.

Attracting consumers' attention to products or services in a market through an influencer on a SNS, such as a mediator, is called influencer marketing. There are two types of influencers: a mega-influencer who has over 100,000 followers and a micro-influencer who has 2000–100,000 followers. Reports show that signing on such influencers to promote company products or services can make a significant difference in sales.

3. Difference between mass media and social media

Crucial differences exist between the mass media and social media. First, most of the information senders on the social media are individuals. Second, the social media provides an avenue for interactive exchange of information.

As mass media creates and broadcasts programs sponsored by enterprises/companies, it strongly reflects the sponsors' intentions. However, social media provides individual users with a place to share their honest reviews or comments and enjoys an advantage in that it can create opportunities for disseminating information about products or services in terms of the value placed on them by others (i.e., users and not the companies selling these products or services).

Recently, word-of-mouth information spreading on the social media has attracted increasing attention because it can introduce popular events or products. On the flip side, intentional diffusion of wrong information such as fake news can easily happen on the social media. First, malicious users may target an event, spreading misinformation about it, which would dupe most users except those conversant with it, or they may anonymously post fake news, pretending that the report is authentic. Else, malicious users diffuse such information using false names. Second, malicious users may target gullible users. At any rate, such activities create a negative image in the minds of some users, encouraging biased thoughts.

In the case of the mass media, broadcasting of false information causes the party to lose its credibility in the viewers' minds, and viewer ratings will decrease. In such cases, the self-purification function can alleviate matters to some extent. In the case of the social media, such problems cannot be solved easily by self-purification, because the company operating the social networking site does not check the

contents of posts to confirm their veracity before they are posted. In practical terms, it is not feasible for these enterprises to remove fake news by monitoring all posts. In the case of Facebook, the operator might remove a post if a third party reports it as offending the site's policy. Thus, the social media requires its own self-purification function. A large-sized social media enterprise may be able to devote adequate human resources to manually respond to all such reports from its users. However, this is not possible for all social media websites.

Thus, if we can calculate the reliability of the content posted online based on the feedback on social media, we might be able to protect innocent users from malicious ones. Various recent studies have focused on rumor detection or judgment of fake news on social media [22–28]. Most of these studies analyzed the contents or behaviors of the users or estimated the reliability of their remarks.

In the next section, we describe a method to analyze feedback provided for a post on social media.

4. Feedback analysis methods

Human relationships are important for a social media platform such as Twitter. By emphasizing the relationships among users, we consider whether a user can be evaluated by other users by analyzing their feedback. Next, we describe a method to analyze various elements of feedback provided via users' replies. The results of the analysis are used to investigate the relation between the information obtained from the reply text and the users' personalities estimated from the same texts. Based on this result, we analyze how the various types of feedback are related to the types of information diffusion. These findings will be useful to predict information diffusion from feedback contents. In this study, we analyze feedback tendency focusing on buzz tweets, flaming tweets, and the tweets by famous persons, using the methods described in the following subsections, namely, via the use of expressions of emotion in the Japanese language, emojis, emoticons, semantic vectors, latent Dirichlet allocation (LDA), information entropy, and personality analysis.

4.1 Emotional expressions in the Japanese language

There are various emotional expressions in the Japanese language. Therefore, even though the analysis of positive/negative emotions may appear to be comparatively simple, an emotion dictionary providing detailed information is necessary for such an analysis.

In this chapter, we use the Japanese Appraisal Evaluation Expression Dictionary (JAD) [29], which systematically registers Japanese emotional expressions, for our analysis. When we compare a word in the reply text with that in the dictionary, the manner of processing semantically similar words poses a problem. To solve this problem, we create a database of similar expressions using word-distributed representations that were trained using Wikipedia articles (fastText [30] + Wikipedia: 300 dimensions), and thus, we increase the number of expressions that can be matched. We determine the threshold as 0.6 and list similar expressions.

Examples of Japanese emotional expressions are shown in **Table 1**.

4.2 Emoticons

Web texts are sometimes annotated with emoticons, which serve as nonverbal information. The emoticon is also called a facemark, and in Japan especially, emoticons are very popularly used on bulletin boards and in e-mails. By annotating emoticons to

Emotion polarity	Word example (Japanese)
Positive	Enchanting (*uttori*), festive (*ukareru*), enjoy (*tanoshimu*), etc.
Negative	Become irate (*ikidooru*), pent-up rage (*uppun*), feel sadness (*kanashimu*), etc.

Table 1.
Examples of Japanese emotional expression.

a sentence, the emotion of the writer can be expressed easily. Many a time, an emoticon can communicate a nuance that cannot be conveyed easily with words.

We detect the replies containing emoticons as annotations and analyze them by matching them against the emotions/meanings expressed by the emoticon. Examples of emoticons are shown in **Table 2**. The emotion categories of emoticon are joy, shame, anger, sorrow, surprise, and hate.

4.3 Emojis

Similar to emoticons, emojis are a form of nonverbal information annotated to web texts. Unlike emoticons, emojis corresponding to events, actions, as well as emotions are readily available pictographs of faces, objects, and symbols.

The varieties of available emojis differ depending on the SNS. In this study, we extract and analyze emojis used on Twitter. **Table 3** shows examples of emojis and their respective emotion categories. The emotion categories of emoji are joy, love, anger, anxiety, sorrow, surprise, and neutral.

4.4 Semantic vectors (Wiki2vec)

It is helpful to refer to word semantic dictionaries to understand the contents written in the text. However, because existing word semantic dictionaries such as a thesaurus do not include many proper nouns, they are unsuitable for analyses of social media platforms such as Twitter. Thus, in this study, we analyze a semantic vector by a unit of reply based on distributed representations called Wikipedia entity vector [31], which was created by training relations of entities based on Wikipedia articles.

We also analyze the differences among the reply text sets by acquiring sentence-distributed representation vectors based on bidirectional encoder representations from transformer (BERT) [32], which has been attracting much attention in recent

Emotion	Example of emoticon
joy (+)	ヾ (⌒▽⌒) ゞ , (o ˆ v ˆ o) , ('∀')
shame (-)	ヽ (= ´Д'=) ノ , (〃／∀＼〃) , (●艸д'o)
anger (-)	ヾ (｀□´!) ノ , (-˚ーメ) , < (｀ˆ´) >
anxiety (-)	— (￣∇￣;) , (－o－; , (￣∇￣;)
sorrow (-)	σ (／＿;) , (T ˆ T) , (´;ω;`)
surprise (0)	ε = (˚ □、˚ *) , (˚ O˚) , (◎o◎)
hate (-)	(¬＿¬) , (￣ε￣@) , (￣Д￣) ノ

Table 2.
Examples of emoticons and their emotion categories.
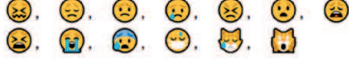
Emotion	Example of emoji
joy (+)	
love (+)	
anger (-)	
anxiety (-)	
sorrow (-)	
surprise (0)	
neutral (0)	

Table 3.
Examples of emojis and their emotion categories.

years. **Figure 1** shows the chart of BERT encoder. As preprocessing of input to the transformer encoder of BERT, the 15% of the word sequences are replaced to [MASK] tokens. The BERT tries to predict the meanings of the masked words based on non-masked words as context.

It is easy to apply BERT to the specific task by using fine-tuning. For example, in sentiment analysis, the sentiment classification model can be fine-tuned by adding the classification layer into the BERT structure.

4.5 Topic analysis via latent Dirichlet allocation

The topics of the posted contents can be analyzed by a topic modeling method. A recently proposed neural topic modeling method for topic analysis, called latent Dirichlet allocation [33], is a simple but effective approach. LDA topic modeling is often used to find topics from text data. LDA can be presented by a graphical model. **Figure 2** shows the LDA graphical model.

We analyze the differences between the sets of replies to buzz tweets, non-buzz tweets, and flaming tweets by developing a topic distribution based on LDA.
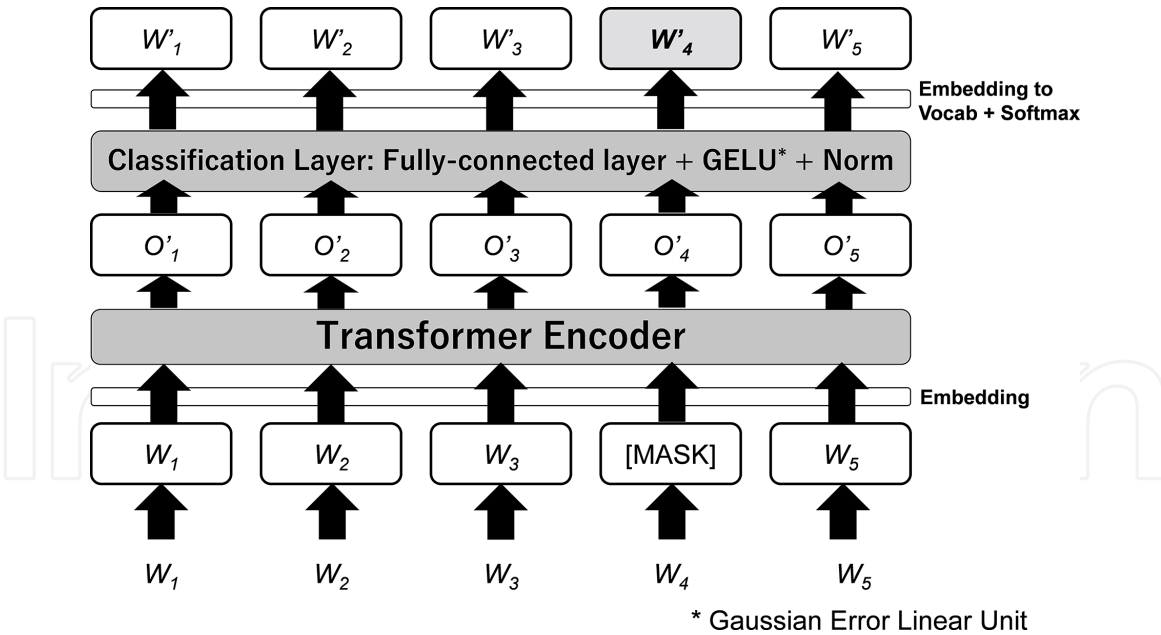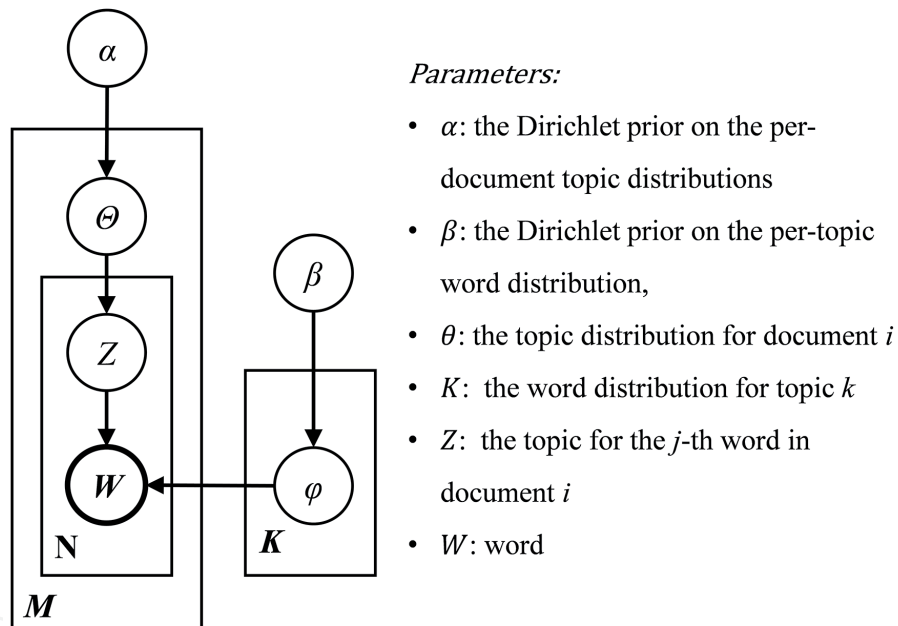
Figure 1.
BERT encoder.

Parameters:

- α: the Dirichlet prior on the per-document topic distributions
- β: the Dirichlet prior on the per-topic word distribution,
- θ: the topic distribution for document i
- K: the word distribution for topic k
- Z: the topic for the j-th word in document i
- W: word

Figure 2.
LDA graphical model.

4.6 Information entropy

The index of the amount of information can express the richness of the written contents. We identify the differences among the features by focusing on word appearance probability and calculating the information entropy for each reply. The information entropy for each reply is calculated using Eq. (1). $p(w)$ indicates the probability of the appearance of word w in the corpus.

$$\mathrm{H(R)} = -\sum_{w \in R} p(w) \log p(w) \tag{1}$$

The higher this index becomes, the more the information is expressed in the text. In other words, if the information entropy of the reply texts of the tweets is very high, it has plenty of topics to tell about the target tweets or the related topics, and it can be said that the buzz phenomenon is being occurred in the tweet.
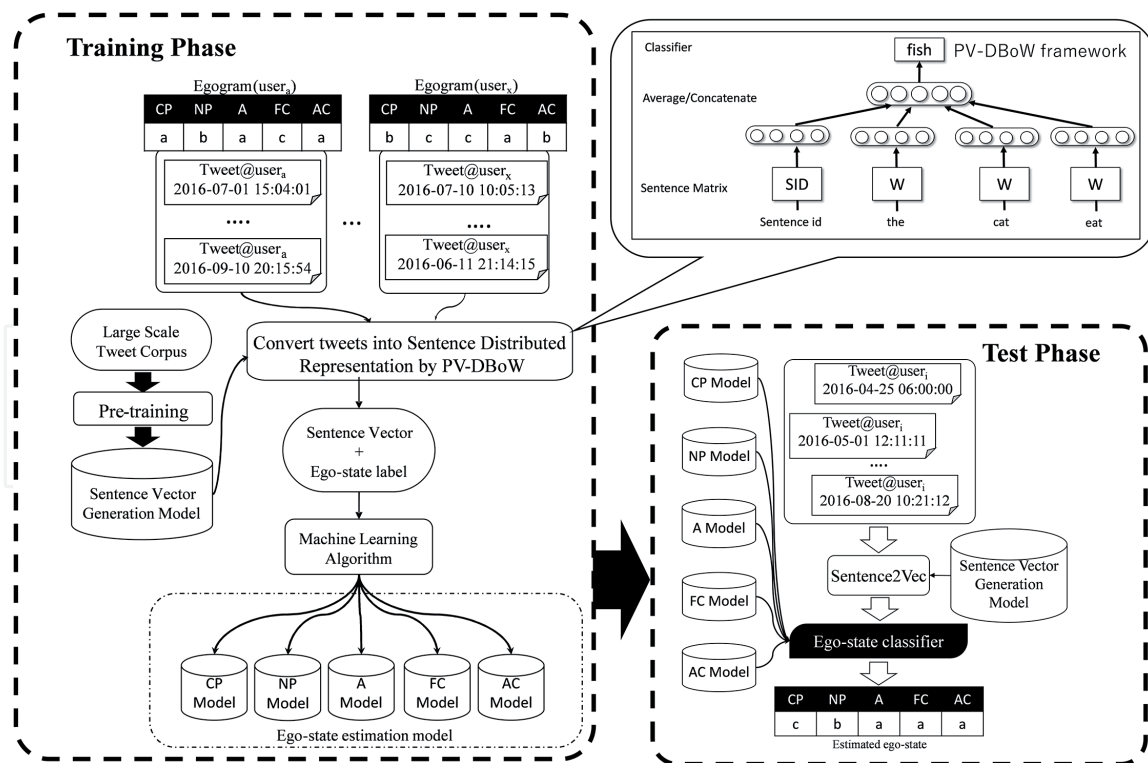
Figure 3.
System flow of personality analysis.

4.7 Personality analysis

The personality of the user who posts replies should be more or less reflected in his/her posted contents. Thus, using the personality analysis model proposed by Matsumoto et al. [34], we analyze the user's personality based on the feedback of other users who responded to his/her post.

The personality estimation model proposed by Matsumoto et al. was trained using neural networks that employ the averaged word-distributed representation vector as the input feature, and it outputs the five levels of ego state. The system flow is shown in **Figure 3**. First, we collect the tweets by the Twitter users who posted egogram assessments according to the egogram assessment website by using Twitter API. Next, we create a sentence-distributed representation generation model by using PV-DBoW algorithm which was proposed by Le and Mikolov [35]. The ego-state level classifier is created based on machine learning algorithm such as deep neural networks. We used feedforward neural networks, which have three hidden layers using sentence-distributed representation as input feature.

5. Results of the analysis

We analyze the reply information using the seven methods described in the previous section. Also, we clarify the correlations between the features obtained from the analysis results and the metadata of the posted contents by correlation analysis. First, we prepare three kinds of data: buzz tweet, flaming tweet, and non-buzz tweet.

A buzz tweet refers to diffused information, whereas a non-buzz tweet refers to diffused but neither buzz nor flaming information. A flaming tweet refers to

negatively diffused information. We collect the contents of these three types of tweets, as well as the number of retweets and the number of favorites for each. **Table 4** shows the summary of the collected target data.

	Buzz	Non-buzz	Flaming
Number of tweets	150	150	20
Number of replies	13,120	14,313	2676
Average number of RT	30,253.58	2425	4931.2
Average number of FAV	80,845.77	12,924.96	16,619.5

Table 4.
Summary of target data.
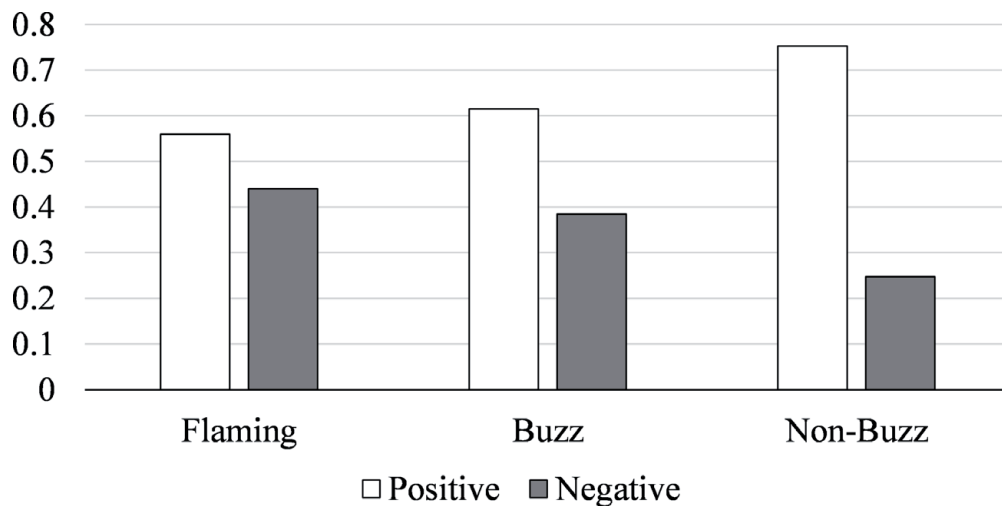
Figure 4.
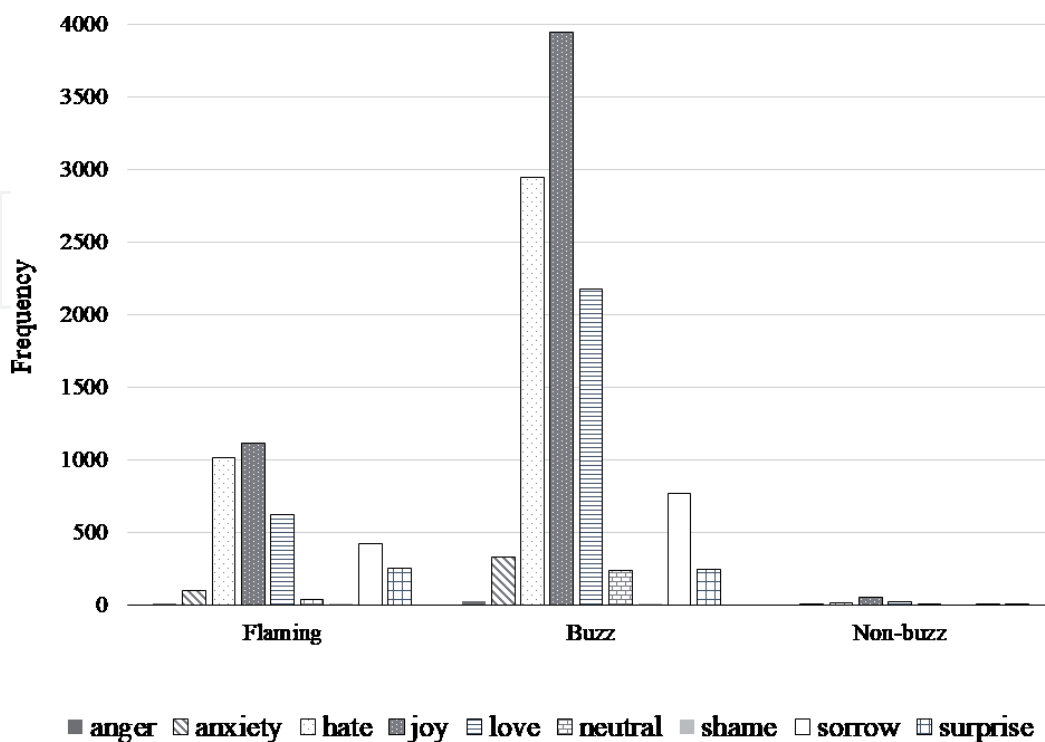*Positive/negative analysis based on the Japanese appraisal dictionary.*

Figure 5.
Emotion analysis based on the Japanese appraisal dictionary.

5.1 Result 1: analysis using words/expressions showing emotion

We conducted two types of analyses using the JAD. **Figure 4** shows the analysis result of the appearance probability of positive/negative emotion. **Figure 5** shows a more detailed analysis, which classifies the results into one of nine emotion categories. The classification shows a similar tendency between buzz and flaming tweets. In the case of non-buzz tweets, the vast majority consist of positive opinions about a famous person. In flaming tweets, negative and positive emotions are more and less frequent, respectively, than in buzz tweets, suggesting that the vast majority consists of negative opinions.

5.2 Result 2: analysis using emoticons

Figure 6 shows the average scores obtained by considering emoticons and calculating their appearance frequencies for each reply in each reply set. This figure shows evidence of similar tendencies in the usage patterns of emoticons in the buzz, non-buzz, and flaming tweets in spite of the differences in their appearance frequencies.

Almost all users using emoticons on Twitter do so to express positive emotions such as joy or love. Notably, these users use emoticons in almost every tweet, whereas the non-emoticon users almost never use them. Thus, we are of the opinion that it is not possible to obtain the correct result by analyzing feedback containing only emoticons.

5.3 Result 3: analysis using emojis

Figure 7 shows the average scores obtained by considering emotions expressed by emojis and calculating emoji frequency for each reply set. This result shows that as with the emojis, the appearance patterns are similar in each category, but differences exist with regard to the total appearance frequencies. Thus, it is possible that we might not be able to grasp the tendency of information diffusion by analyzing the feedback based on emojis alone.
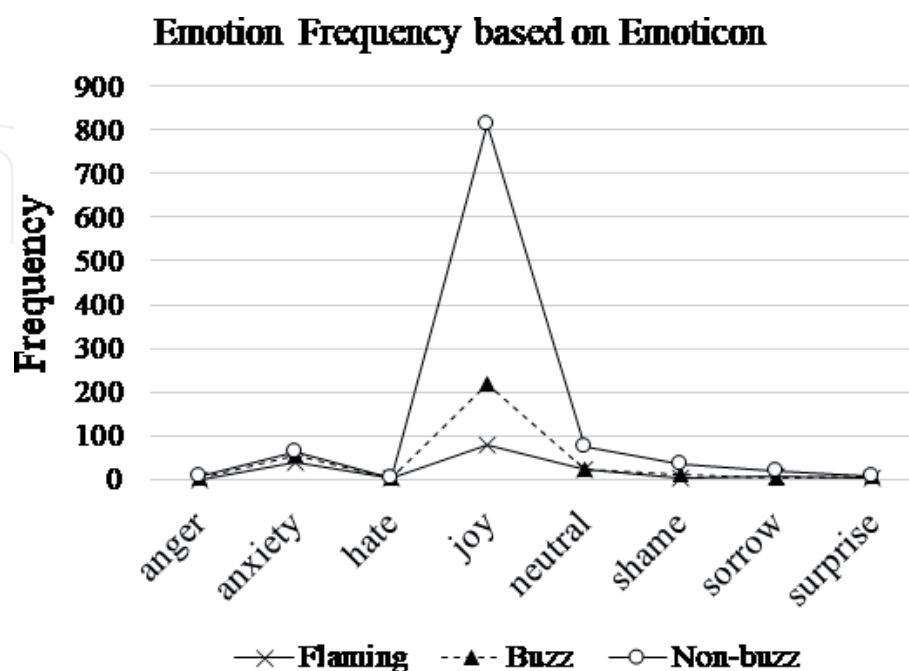

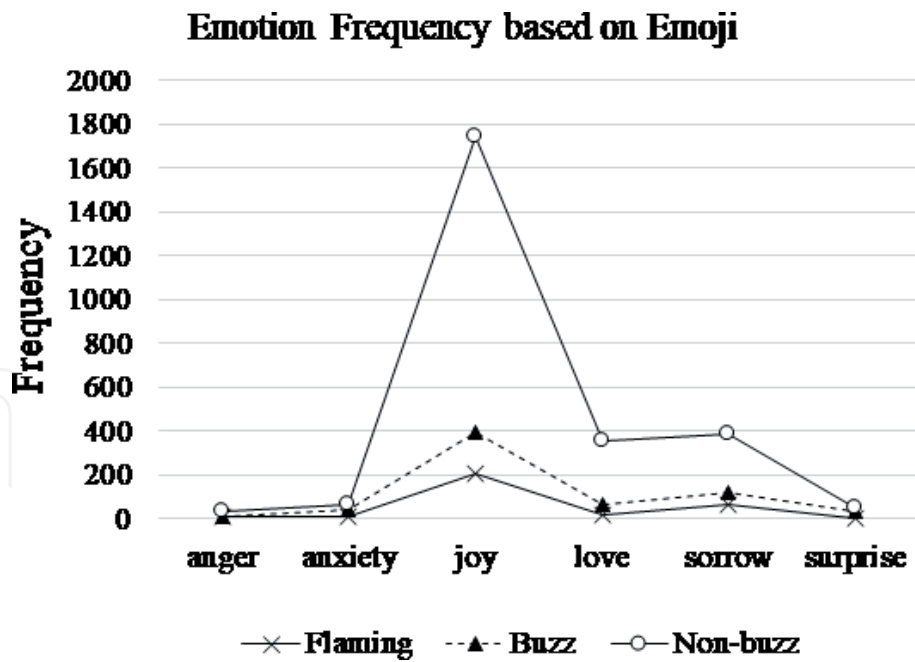**Figure 6.**
Emoticon analysis result for each reply set.

Figure 7.
Emoji analysis result for each reply set.

5.4 Result 4: analysis using information entropy

Table 5 shows the average information entropies for the three types of tweets (buzz, non-buzz, and flaming) by calculating the amount of information diffused from the sets of replies for each tweet (i.e., word entropy calculation).

It is possible that the average entropy of the non-buzz tweets was low because the feedback contained many short sentences (such as encouraging messages or greetings containing less information). This result suggests that we can classify tweets into buzz, flaming, or non-buzz tweets based on information entropy.

5.5 Result 5: analysis using semantic vectors

The procedure for analyzing the feedback via the semantic vectors consists of the following two steps:

1. We obtain the semantic vectors (BERT, Wiki2vec, and the joint vector of BERT and Wiki2vec) for each reply.

2. For each label, we conduct clustering by k-means and plot centroid vectors in the two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) [36].

Type	Word entropy
Buzz	5.302469
Non-buzz	2.417621
Flaming	5.446601

Table 5.
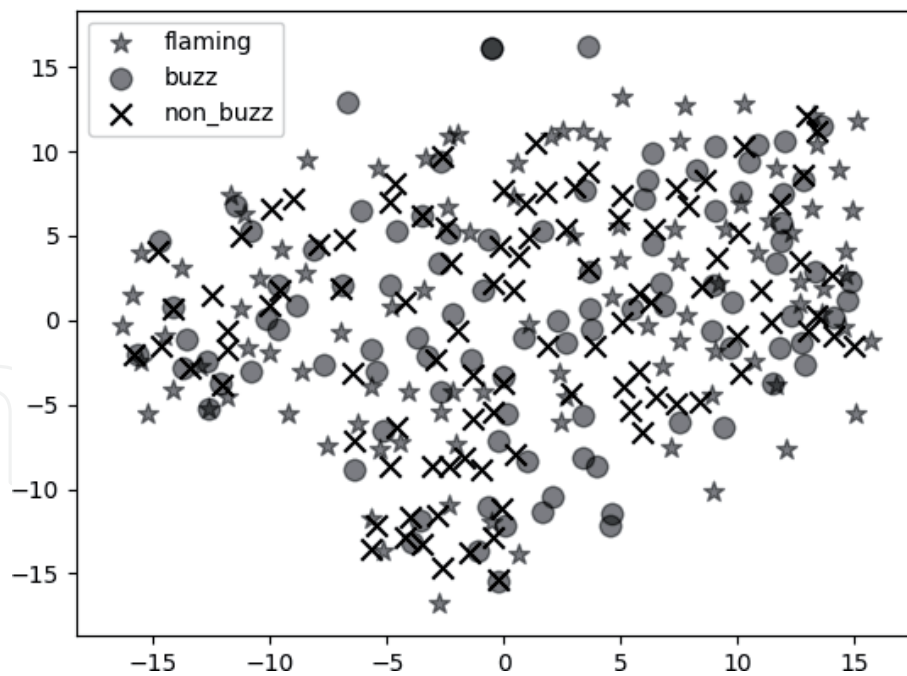*Word entropy for each type of tweet.*
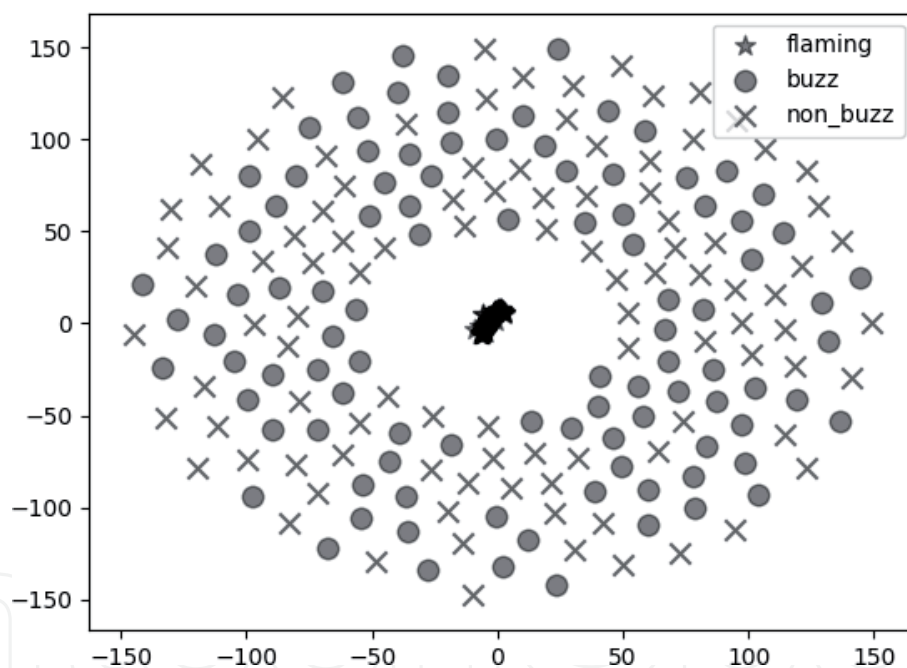
Figure 8.
Plotting by BERT vector.

Figure 9.
Plotting by Wiki2vec vector.

Figures 8–10 show the results of the visualization of each semantic vector. We use pre-trained 768 dimension vectors that are fine-tuned based on a Wikipedia article in Japanese [37]. The training corpus for Wiki2vec also comprises Wikipedia articles in Japanese, and the number of dimensions is 100. The number of dimensions for the joint vector is 868 after concatenating the 768 dimensions of the BERT vector with the 100 dimensions of the Wiki2vec vector.

When using the BERT vector, no significant differences are noted among the flaming, buzz, and non-buzz tweets. However, the visualization results of the wiki2vec and joint vectors show that replies characterizing flaming are clearly distinguishable from the other replies. We attribute this result to the fact that Wiki2vec targets the entity vector (such as proper nouns), whereas BERT is trained to identify versatile distributed representations. That is, there is a possibility that
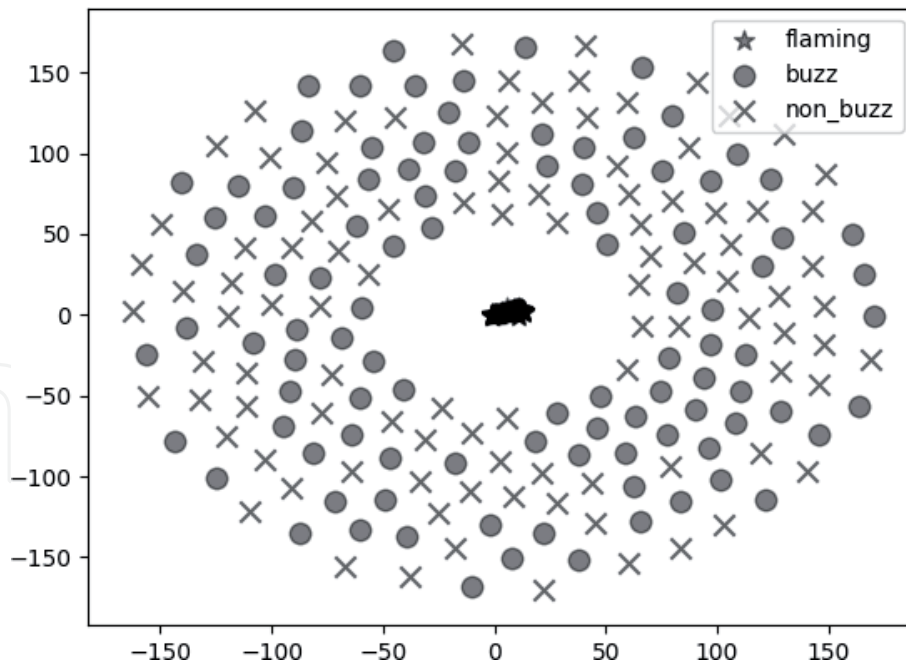
Figure 10.
Plotting by BERT+Wiki2vec vector.

Figure 11.
Word cloud made by LDA model: number of topics is 10.

specific proper nouns (such as derogatory terms for targets under attack or names of the persons concerned) were frequently used in the replies to the flaming tweets.

5.6 Result 6: analysis by topic modeling

We conduct topic modeling with LDA for the numbers of topics (in this case, 10 and 20). To judge the quality of a topic model, a scale called perplexity is often used. The perplexity of the topic model is calculated for each number of topics using Eq. (2). A lower perplexity denotes a more accurate probability model. In the equation, $p(w_d)$ indicates the probability of appearance of word w in document d. N indicates the total number of words.

$$\text{perplexity}\left(W^{test}|M\right) = \exp\left(-\frac{1}{N}\sum_d \log p(w_d)\right) \quad (2)$$
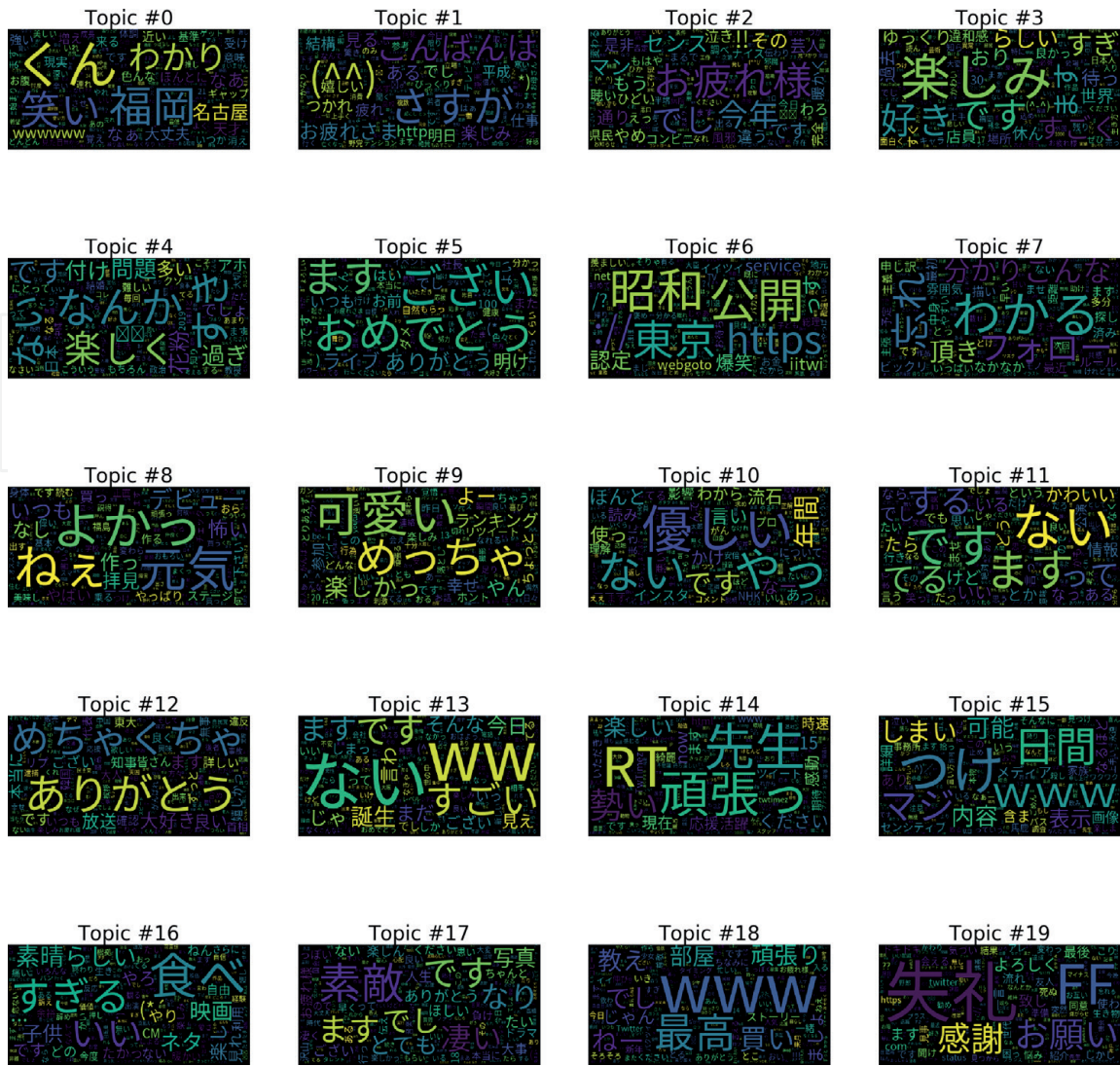
Figure 12.
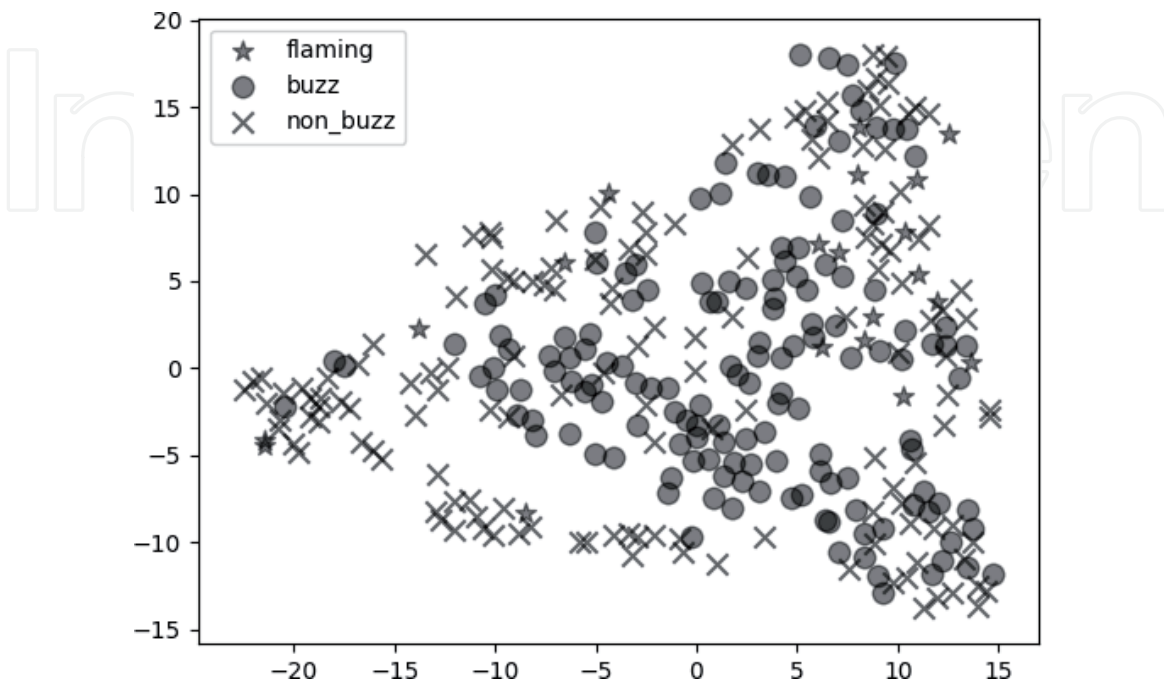*Word cloud made by LDA model: number of topics is 20.*

Figure 13.
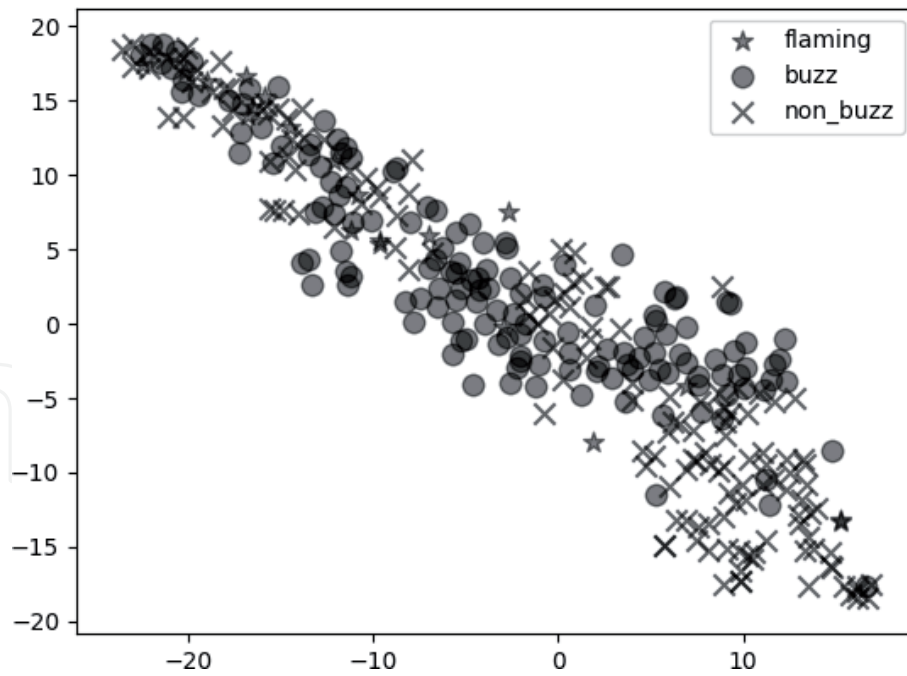*Plotting based on LDA topic vector: number of topics is 10.*

Figure 14.
Plotting based on LDA topic vector: number of topics is 20.

Figures 11 and **12** show the word clouds made using the word appearance probability for each topic. These figures show that many positive words appeared in the feedback, which indicates that the number of the flaming tweets is small.

Figures 13 and **14** show the topic vectors, which are compressed into the two-dimensional space by t-SNE, for each reply. Even though a number of examples exist for replies to flaming tweets, they are widely distributed.

When the number of topics is 10, multiple different clusters are generated for the buzz and non-buzz tweets. Even though similar topics are sometimes generated for both buzz and non-buzz tweets, dissimilar topics tend to be generated more often.

5.7 Result 7: analysis using personality estimation

Sets of replies are estimated for the buzz, non-buzz, and flaming tweets using neural networks. The result for the personality estimation is obtained as a vector
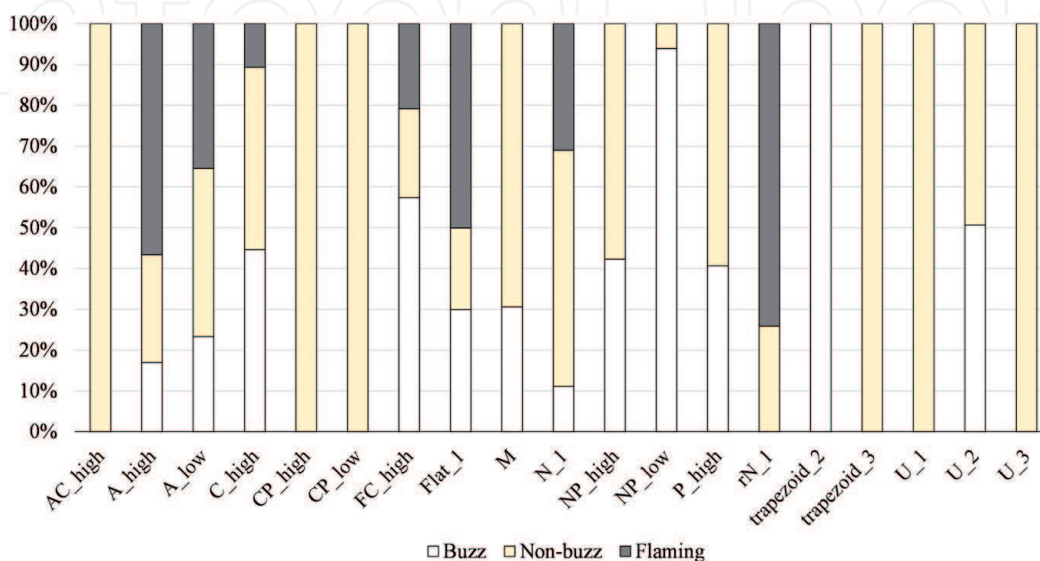

Figure 15.
Result of personality analysis.

15

that indicates five types of ego-state degrees. We call this vector the ego-state vector. The personality pattern is classified into 29 kinds as per the shape of the ego-state vector. **Figure 15** shows the proportions of the buzz, non-buzz, and flaming tweets for each personality pattern (**Table 6**).
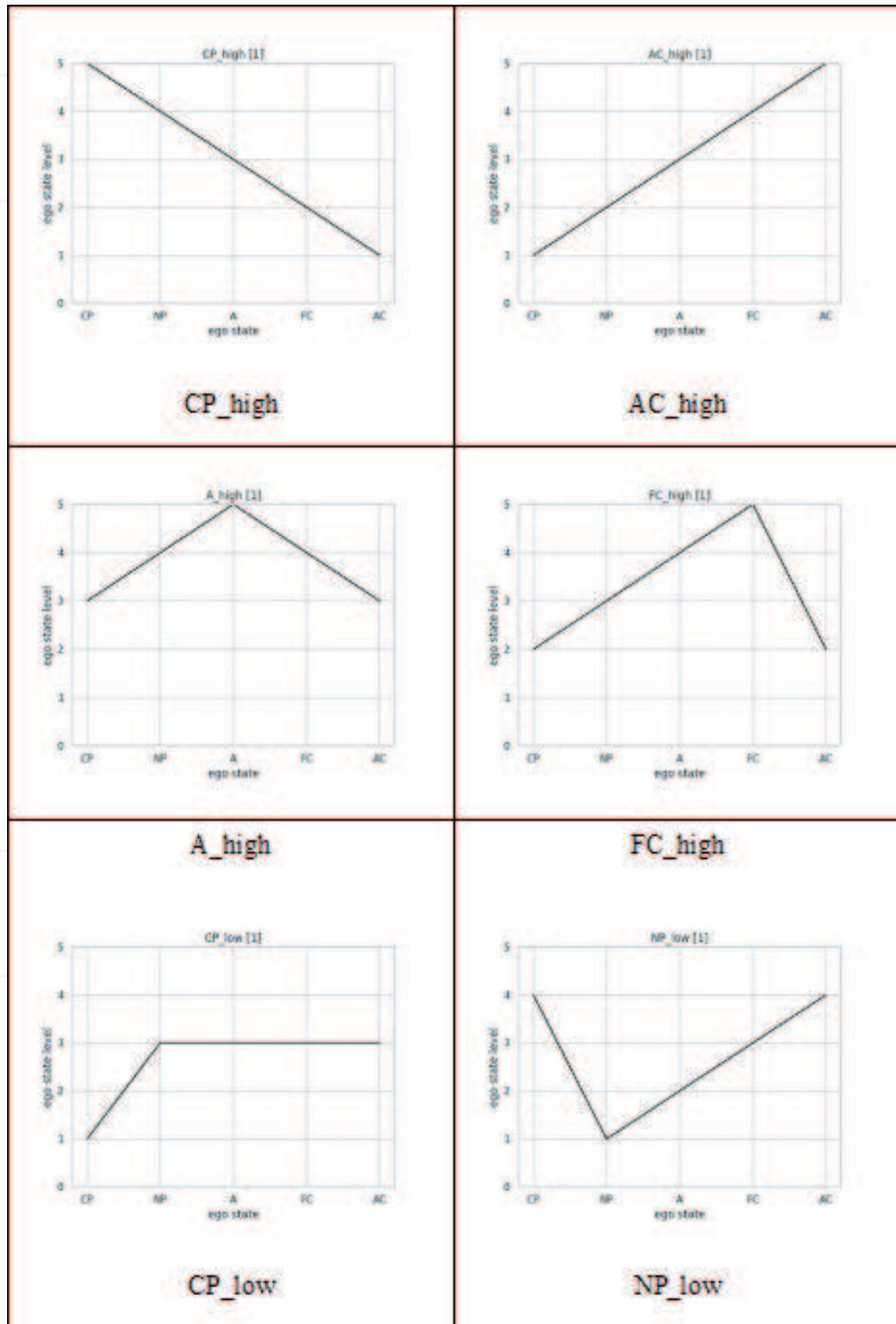

Table 6.
A part of egogram patterns.

Considerable variations in personality patterns are noted for the non-buzz tweets, while a biased tendency is observed for the buzz and flaming tweets. Notably, three non-buzz tweets and one flaming tweet are classified under rN_1, which denotes a critical and idealistic personality. In other words, these users criticized the poster of the original tweet for his/her lack of a sense of justice. The statements by such users often cause flaming.

However, only one buzz tweet was classified as the "trapezoid_2" type. This personality type is self-sacrificing and tends to devote himself/herself to the causes of others. Therefore, it is considered that the users used decent expressions because they were moved by the contents of the buzz tweet.

6. Discussions

The results of the analysis show that it is important to analyze feedback from multiple viewpoints. As it is difficult to gain a precise understanding of others' opinions from nonverbal expressions such as emoticons or emojis only, it is crucial to use a method that analyzes both verbal and nonverbal information. The feedback analysis method proposed in this study can successfully analyze reply texts containing words as the analysis algorithm is based on natural language text. Therefore, we believe that a sufficient amount of feedback from flaming and buzz tweets can be obtained using our method.

However, the number of replies to buzz tweets is proportional to the degree of the buzz phenomenon. This poses a problem in that we cannot collect and analyze all the replies because the number of replies is very large for certain flaming tweets. Thus, to investigate the factors affecting flaming and buzz tweets in more detail, we should analyze the correlation between the follow-follower and the contents of the reply texts.

7. Conclusions

In this chapter, we focused on information diffusion on social media and described a method to analyze feedback to specific tweets. We investigated the semantic/sensibility differences among three types of tweets—buzz, non-buzz, and flaming tweets—by analyzing replies to the posts. The results confirm the possibility of classifying the diffusion type accurately using semantic information included in the replies. However, with regard to expressions of emotion or sympathy, because we focused only on emotions that can be expressed by words, we detected many positive opinions, but significant differences were not observed.

Our results also showed that negative opinions tend to be common in the feedback to flaming tweets compared to that for buzz tweets. Flaming can cause harassment and cyberbullying and destroy personal relationship. These expressions should be detected by automatic classification systems such as flaming detectors based on artificial intelligence model, sentiment analysis model, or personality analysis model. However, buzz phenomenon and flaming phenomenon are similar in terms of information diffusion. Thus, in the future, we plan to construct an algorithm to analyze the differences between the two by considering specific conditions for each type of tweet.

Conflict of interest

The authors declare no conflict of interest.

Author details

Kazuyuki Matsumoto*, Minoru Yoshida and Kenji Kita
Tokushima University, Tokushima, Japan

*Address all correspondence to: matumoto@is.tokushima-u.ac.jp

IntechOpen

References

[1] Takahashi T, Igata N. Rumor detection on twitter. In: 2012 Joint 6th International Conference on Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS). 2012. DOI: 10.1109/SCIS-ISIS.2012.6505254

[2] Zamani S, Asadpour M, Moazzami D. Rumor detection for Persian Tweets. In: 2017 Iranian Conference on Electrical Engineering (ICEE). 2017. DOI: 10.1109/IranianCEE.2017.7985287

[3] Aramaki E, Maskawa S, Morita M, Morita M. Microblog-based infectious disease detection using document classification and infectious disease model. Journal of Natural Language Processing. 2012;**19**(5):419-435

[4] KaiShu AS, Wang S, Tang J, Liu H. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter. 2017;**19**(1):22-36

[5] Aldwairi M, Alwahedi A. Detecting fake news in social media networks. Procedia Computer Science. 2018;**141**:215-222

[6] Yoshida S, Kitazono J, Ozawa S, Sugawara T, Haga T, Nakamura S. Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection. In: 2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD); 2014. DOI: 10.1109/CIBD.2014.7011523

[7] Vinupriya A, Gomathi S. Web page personalization and link prediction using generalized inverted index and flame clustering. In: 2016 International Conference on Computer Communication and Informatics (ICCCI); 2016

[8] Shukla SSP, Nitin, Singh SP, Parande NS, Khare A, Pandey NK. Flame detector model: A prototype for detecting flames in social networking sites. In: 2012 UKSim 14th International Conference on Computer Modelling and Simulation; 2012

[9] Rösner L, Krämer NC. Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments. Social Media+Society. 2016:1-13. DOI: 10.1177/2056305116664220

[10] Sri Nandhini B, Sheeba JI. Online social network bullying detection using intelligence techniques. Procedia Computer Science. 2015;**45**:485-492

[11] Alsuwaidan L, Ykhlef M. Information diffusion predictive model using radiation transfer. IEEE Access. 2017;**5**:25946-25957

[12] Jiang J, Wen S, Yu S, Xiang Y, Zhou W. K-center: An approach on the multi-source identification of information diffusion. IEEE Transactions on Information Forensics and Security. 2015;**10**(12):2616-2626

[13] Kazama K, Imada M, Kashiwagi K. Characteristics estimation of information sources by information diffusion analysis. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; 2010. DOI: 10.1109/WI-IAT.2010.130

[14] Murakami A, Suzuki H. Predicting information diffusion in Twitter through retweeting behaviors. In: The 26th Annual Conference of the Japanese Society for Artificial Intelligence; 2012. pp. 1-4

[15] Saito K, Kimura M, Ohara K, Motoda H. Learning asynchronous-time

information diffusion models and its application to behavioral data analysis over social networks. Journal of Computer Engineering and Informatics. 2013;**1**(2):30-57

[16] User chart. Available from: http://userchart.jp/

[17] kred. Available from: https://home.kred/

[18] kuchikomi@kakaricho. Available from: https://service.hottolink.co.jp/service/kakaricho/

[19] Liao Q, Wang W, Han Y, Zhang Q. Analyzing the influential people in Sina Weibo dataset. In: 2013 IEEE Global Communications Conference (GLOBECOM); 2013. DOI: 10.1109/GLOCOM.2013.6831542

[20] Matsuo Y, Yasuda Y. How relations are built within a SNS world—Social network analysis on Mixi. Transactions of the Japanese Society for Artificial Intelligence. 2007;**22**(5):531-541

[21] Tsugawa S, Kimura K. Identifying influencers from sampled social networks. Physica A: Statistical Mechanics and its Applications. 2018;**507**(1):294-303

[22] Castillo C, Mendoza M, Poblete B. Information credibility on twitter. In: Proceedings of WWW; 2011. pp. 675-684

[23] Kwon S, Cha M, Jung K, Chen W, Wang Y. Prominent features of rumor propagation in online social media. In: Proceedings of ICDM; 2013. pp. 1103-1108

[24] Liu X, Nourbakhsh A, Li Q, Fang R, Shah S. Real-time rumor debunking on twitter. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. CIKM '15; 2015. pp. 1867-1870

[25] Ma J, Gao W, Wong K-F. Detect rumors in microblog posts using propagation structure via kernel learning. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); 2017. pp. 708-717

[26] Buntain C, Golbeck J. Automatically identifying fake news in popular Twitter threads. In: 2017 IEEE International Conference on Smart Cloud (SmartCloud); 2017. DOI: 10.1109/SmartCloud.2017.40

[27] Ajao Q, Bhowmik D, Zargari S. Fake news identification on Twitter with hybrid CNN and RNN models. In: Proceedings of the 9th International Conference on Social Media and Society (SMSociety'18); 2018. pp. 226-230

[28] Helmstetter S, Paulheim H. Weakly supervised learning for fake news detection on Twitter. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); 2018. DOI: 10.1109/asonam.2018.8508520

[29] Sano M. Reconstructing English system of attitude for the application to Japanese: An exploration for the construction of a Japanese dictionary of appraisal. In: 38th International Systemic Functional Congress; 2011

[30] Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics; 2016. Volume 2, pp. 427-431

[31] Japanese Wikipedia Entity Vector. Available from: http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/

[32] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; 2018. arXiv:1810.04805

[33] Blei DM, Lafferty DJ. Text Mining: Theory and Applications, Chapter Topic Models. UK: Taylor and Francis; 2009

[34] Matsumoto K, Tanaka S, Yoshida M, Kita K, Ren F. Ego-state estimation from short texts based on sentence distributed representation. International Journal of Advanced Intelligence (IJAI). 2017;**9**(2):145-161

[35] Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning, PMLR. 2014;**32**(2):1188-1196

[36] Laurens V, Maaten D, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008;**9**:2579-2605

[37] BERT Japanese Pretrained Model. Available from: http://nlp.ist.i.kyotou.ac.jp/index.php?BERT%E6%97%A5%E6%9C%AC%E8%AA%9EPretrained%E3%83%A2%E3%83%87%E3%83%AB