Fuzzy Systems and Data Mining VIII
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220384

Affective Analysis and Visualization from Posted Text, Replies, and Images for Analysis of Buzz Factors

Reishi Amitani[a], Kazuyuki Matsumoto[b], Minoru Yoshida[b] and Kenji Kita[b]
[a] *Graduate School of Science and Technology for Innovation, Tokushima University, Japan*
[b] *Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Japan*

Abstract. In this study, we propose a method to visualize the factors that contribute to the buzz phenomenon triggered by Twitter posts. The analysis included tweets, images, and replies. Replies are after-the-fact responses posted in response to a posted tweet and therefore cannot be used to predict buzz phenomena. Therefore, they cannot be used to predict the buzz phenomena. In this study, the tweet body, images, and reply text were feature vectors, and an affective analysis model was constructed. Visualization of the relationship between the sensibility features output from this model and the number of RTs and likes (echo index), which represent the scale of the buzz, will be useful for analyzing the factors behind the popularity. Consequently, the subjective sensibility information with the most likes also tended to have a higher degree of similarity among the sensibility vectors.

Keywords. Tweet analysis, natural language processing, trend analysis

1. Introduction

In recent years, with the development of social networking services (SNS), information on various content has been shared and diffused among many users. As a result, Internet-based trends occur frequently. In particular, the rapid spread of information and the surge in popularity are sometimes referred to as “buzz”.

Corporate marketing is well-suited to social networking services, which have strong diffusion power and are attracting attention as a corporate strategy. The term “buzz marketing” is used to describe the process of expanding product awareness by strategically sending out word-of-mouth and reviews. In many successful cases, buzz marketing has led to product promotion and awareness. Famous examples include Ezaki Glico's November 11 “Pocky Day” event, and Morinaga Chocolate's “we will buy your reason for not buying Bake for a 100 yen Amazon gift certificate!” Although not directly linked to sales promotion, as in the case of the Morinaga chocolate project, there have been several cases in which the project has led to product improvements because it has increased awareness and provided an opportunity to understand consumers' honest opinions. Jindal et al. [1] and Zhang et al. [2] investigated the impact of social networking on marketing. This indicates that buzz marketing can be useful for companies if it is well-utilized. This study examined the types of tweets that are likely to cause buzz by analyzing buzz tweets. We believe that this will lead to the proposal of a buzz forecasting method that will be useful for marketing.

"Buzz tweets," which have caused a buzz phenomenon on Twitter, are often posted with only text, but many tweets are also posted with images. This is thought to be because there is a limit to the number of characters in the text posted on Twitter, and information that is difficult to express in text alone can be more easily conveyed visually by attaching images. In addition, from the perspective of analyzing the buzz phenomenon, tweets with images can be analyzed in more detail than text-only tweets because the amount of information available for analysis is greater. In addition, reply text, which is post-tweet information, was also used in the analysis. Because replies are posted after a tweet is posted as a response to that tweet, we thought that this would lead to an analysis of the factors behind the specific buzz phenomenon. In this study, we analyze the factors behind the buzz phenomenon by analyzing the affective information extracted from tweet text, images, and reply text, respectively, for tweets with images.

2. Related Studies

2.1. Classification Of Buzz Tweets

Matsumoto et al. [3] proposed a method for the binary classification of buzz and non-buzz tweets using reply data as features. However, their method collected buzz tweets from curation sites that compiled "topical tweets and interesting image tweets," and whether a tweet was buzzed or not depended largely on the subjective judgment of the collector, and no study was conducted using objective indicators, such as the number of "likes" and "RTs" (retweets). In this study, "Likes" and "RTs" are used as objective indicators to determine whether a tweet is a buzz tweet. The research method used in this study is different because it examines the factor analysis of the buzz phenomenon using the text and image content of tweets and the content of replies to these tweets. Deusser et al. [4] used Facebook data to predict buzz using SVM, AdaBoost, and random forest and compared the results. Facebook, similar to Twitter, is a type of SNS; however, the nature of the information posted differs from that of Twitter because of the nature of its real names. Twitter differs from Facebook in that it is easier for honest opinions to be posted in response to poorly reputed word of mouth.

Jansen et al [5] built the Buzz Detection System (BDS). They trained a buzz detection model in BDS using approximately 120,000 posted sentences on Facebook as training data. As a result, they found that the number of times a passive user engages with a buzz tweets and the number of "likes" for a comment and found that the number of "Likes" was important as a feature for predicting buzz tweets. In the evaluation experiments, their proposed buzz detection model achieved a buzz detection rate of over 97%. In their proposed BDS, to classify buzz or non-buzz, "likes" and the past behavior of the user involved in the buzz to classify buzz or non-buzz. However, these are not information available at the time of posting, but rather information obtained after posting. While these elements are found to be necessary to improve the accuracy of buzz detection, the system does not function as a system to confirm whether or not a buzz phenomenon occurs before posting. In this study, the main features are the content of the tweet at the time of posting and the attached image, and the number of likes! and retweet counts, so the methods, objectives, and target data are different.

2.2. Personal Interest Estimation Using SNS

Masumoto et al. [6] believe that the more information available on SNSs, such as personal profiles (gender, age, and hobbies) and location information, the more accurate the interest estimation. They also believe that user interest is a factor that increases likes and RTs. Although our study differs from Masumoto et al.'s in that we did not use user information, there is a commonality in that we analyzed the relationship between likes and RTs based on the content of tweets and replies. Bhattacharya et al. [7] proposed a mechanism for inferring topics of interest to users. They showed that their method of inferring topic expertise on Twitter and transitively inferring the interests of the users who follow them is superior to using labeled latent Dirichlet allocation (L-LDA) as a topic model. Because this study uses information about the user's topic and the number of followers as clues to infer the user's interest, it is different from this study, which analyzes buzz based on the content of tweets and reply data.

2.3. Analyze trends on Social Media

Zhang et al. [8] focused on hashtags and attempted to predict their trends. Content and context, such as vocabulary and emotion, were considered predictors. This experiment showed that content and contextual factors are useful for predicting trends in a large Twitter dataset. This study aims to analyze buzz from the text, images, and replies of tweets, which is similar to their study that considers the content of tweets to be useful for predicting trends.Anusha et al. [9] proposed a method to determine whether the content is interesting by using topic modeling based on latent Dirichlet allocation (LDA) and sentiment polarity analysis based on the NLTK corpus for tweets. In this method, the interestingness of a topic is judged by the weight of the topic's spatial entropy (how much attention tweets on that topic receive) and the score of its emotional polarity (whether it receives attention in a good way). Although the method we aim to use in this study is similar to the analysis using affective information, the purpose is different because it does not use hashtags, but analyzes the factors of the buzz phenomenon using tweet content and reply data.

Ertugan [10] investigated the existence of a relationship between the effectiveness of advertising on Facebook and the benefits derived from advertising. He used statistical inference techniques to measure the benefits of the two advertisements and the degree of correlation between them when running an effective advertising campaign on Facebook. He conducted Pearson's bivariate correlation and linear regression analysis on the survey data obtained from the questionnaires to examine the benefits of Facebook advertising in terms of "customer relationship management" and "new product promotion" benefits. The results showed that Facebook is an effective advertising medium and has strong associations with respect to the benefits of "customer relationship management" and "new product promotion". In this method, statistical analysis was conducted using the results of the questionnaire, but the contents of the postings were limited to those related to some kind of advertising campaign, and the analysis was conducted only with regard to the relevance of obtaining the clear objective of benefits. Our study attempts to clarify the relationship between the content of the postings and the information diffusion effect of the tweets using a deep learning-based affective estimation model, which has different objectives and methods.

3. Proposed Method

The following is a description of the proposed method, whose overview is shown in Figure 1. The procedure is as follows:

- Step-1 Using Twitter API, collect the number of "Likes" and "RTs" given to the tweet to be analyzed, the tweet text, images attached to the tweet, and replies posted in response to the target tweet.
- Step-2 To analyze affective information, create a model to estimate affective information (affective estimation model) using the WRIME corpus of subjective and objective sentiment analysis datasets with BERT vectors as input.
- Step-3 Because affective estimation cannot be performed directly on images, a model was created to estimate text features from image features based on a dataset of image-caption-text pairs, and preprocessing was performed to convert images to text features using this model.
- Step-4 Using the affective estimation model, the affective information of the text, image, and replies of the target tweet is estimated, and the relationship between the affective information and the number of likes and RTs is visualized.
- Step-5 Based on the visualized results, we analyzed how the affective information of text, images, and replies affects the number of likes and RTs and which affective information is the factor that increases the number of likes and RTs.

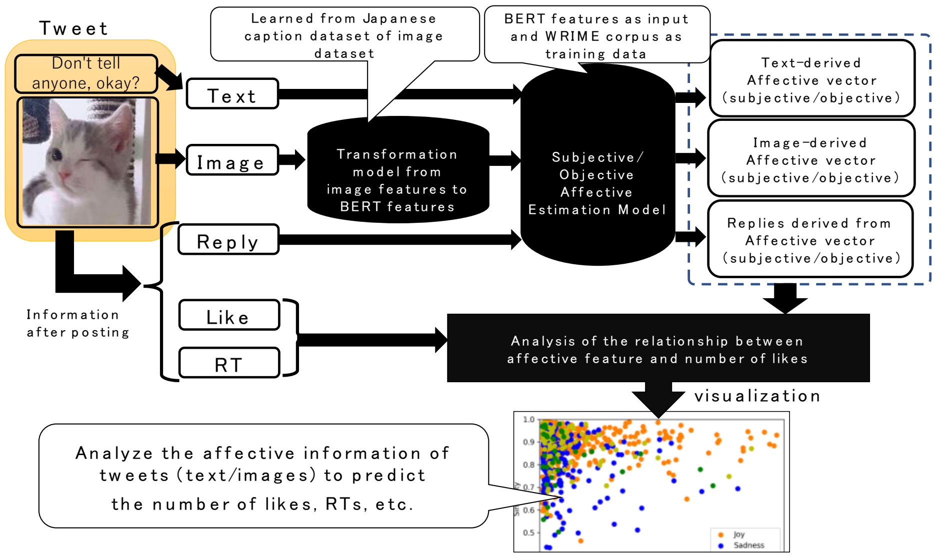

Figure 1. Overview of the proposed method.

3.1. Tweet Collection

The Twitter API [11] is an API provided by Twitter for developers. Using the Twitter API, Twitter data can be obtained without going through the official website. The Twitter API can collect various data, such as tweet IDs, searches using user IDs, and the number of likes and RTs, making it possible to efficiently collect the tweet data necessary for analysis. In this study, we collected data on tweets and replies for analysis, based on the number of likes, tweet IDs, and user IDs. Note that keywords related to a particular topic were used to narrow down the collection of tweets. Topics were selected from recently discussed news articles and entertainment-related proper nouns. We also used the keywords "cat lovers" and "dog lovers," which are often used as hashtags. Table 1 shows examples of topic keywords.

Table 1. Example of topic keywords.

Cat lover	Dog lover
COVID19 Vaccine	Uber Eats
Demon Slayer	Remote lecture
GoTo Travel	Animal Crossing New Horizons

3.2. Text Feature Extraction

In this study, we used the Japanese spoken language BERT model published by Retriva [12] to obtain a distributed representation of the text features from tweets. The Japanese spoken language BERT model has been pre-trained on the Corpus of Spoken Japanese (CSJ) [13] and is considered to have higher expressive power than conventional BERT for spoken text. In this study, we decided to use the Japanese spoken language BERT because we used Twitter tweet data, which is considered to contain a large amount of spoken language. The vector of the distributed representations obtained from this BERT model had 768 dimensions. There are three pre-trained models of BERT for spoken Japanese available, “1-6 layer-wise,” “TAPT512 60k,” and “DAPT128-TAPT512,” depending on the type of fine-tuning data used for field adaptation and the additional layers used for training.

3.3. Image Feature Extraction

As a method for extracting image affective information, image feature extraction is performed using the InceptionV3 [14] model, which is pre-trained to classify images into 1000 object categories based on the ImageNet database containing over 1 million images. These image features were converted into BERT features using the model described below, and the affective information was extracted from the BERT vector and treated as image affective information.

3.4. Conversion from Image Features to BERT Features

A neural network was used to train a model that converts image features into text features (BERT features) using the STAIR Japanese caption dataset [15] assigned to the MS-COCO dataset [16], a large image dataset. The STAIR Japanese caption dataset contains 82,783 images, which corresponds to 1:1 of the images in the MS-COCO dataset. The output of the created model had 768 dimensions. A diagram of the neural network is shown in Figure 2. The image features were 2048-dimensional feature vectors extracted using the InceptionV3 pre-trained model.

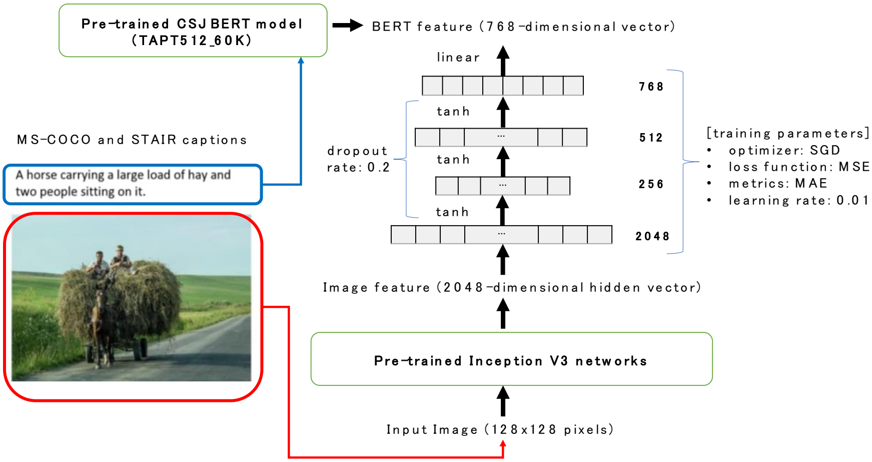

Figure 2. Conversion model from image features to BERT features.

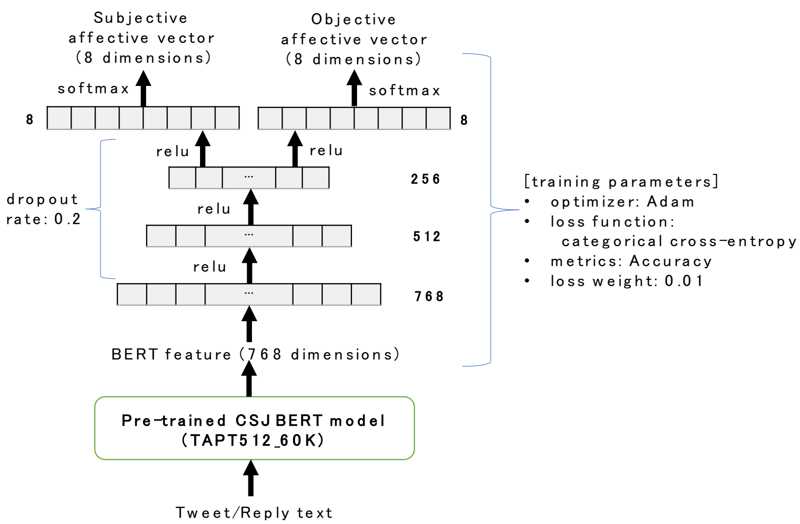

Figure 3. Affective estimation model to extract affective information from BERT features.

3.5. Extracting Affective Information

The WRIME corpus [17] (a Japanese dataset for the classification of subjective and objective emotional polarity) was used to extract the affective information. The WRIME corpus comprises text with Plutchik's eight basic emotions (joy, sadness, anticipation, surprise, anger, fear, disgust, and trust), labeled subjectively by one text author and objectively by three crowdworkers. Using this corpus, we created an affective estimation model that takes BERT vectors as input and outputs subjective and objective affective information, and extracts affective information for text, images, and replies, respectively. Figure 3 shows a diagram of the affective estimation model. The output emotion information is a vector containing probability values for each of the eight emotions (joy, sadness, anticipation, surprise, anger, fear, disgust, and trust), with subjective and objective emotions output separately. The emotion of the dimension with the largest value in this vector was the output affective label.

4. Experiment and Results

Of the 2247 tweets obtained between January and March 2021, 811 tweets to which replies were posted were analyzed. The distribution of the number of likes for the collected tweets is shown in the histograms below (Figures 4 and 5):

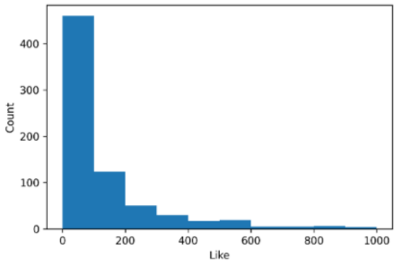

Figure 4. Distribution of likes under 1000.

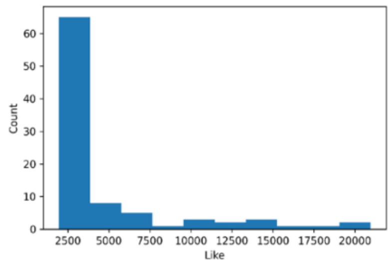

Figure 5. Distribution of more than 1000 likes

As many of the entries have a small number of likes, they are shown in two histograms. The number of likes for more than half of the tweets ranged from zero to 100, with the number of tweets decreasing as the number of likes increased.

The affective information obtained for the 811 tweets consisted of text and images, subjective and objective, for a total of four types, and the breakdown of each affective label is shown in Tables 2 and 3. The results of counting the affective information (affective labels) of the dimensions that take the maximum value of the affective vector are shown.

Table 2. Text affective information breakdown

Affective label	Subjective	Objective
Joy	509	372
Sadness	184	181
Anticipation	79	143
Surprise	39	54
Anger	0	0
Fear	0	46
Disgust	0	15
Trust	0	0

Table 3. Affective information breakdown of images

Affective label	Subjective	Objective
Joy	639	2
Sadness	15	4
Anticipation	0	36
Surprise	157	748
Anger	0	0
Fear	0	21
Disgust	0	0
Trust	0	0

Joy was the most common emotion inferred from the tweets, and the most common subjective affective information in the images, whereas surprise was the most common objective affective information.

In the subjective affective information obtained from the text and images, negative senses, such as fear, disgust, and anger, and positive senses, such as trust were not found, while in the objective affective information, negative senses, such as fear and disgust were found only slightly, confirming the difference between subjective and objective affective information.

The total number of replies obtained for the 811 tweets was 3266. Table 4 shows a breakdown of the subjective and objective affective information estimated from the responses.

Table 4. Affective information breakdown of replies

Affective label	Subjective	Objective
Joy	2278	1850
Sadness	612	642
Anticipation	251	418
Surprise	117	176
Anger	0	0
Fear	0	116
Disgust	8	64
Trust	0	0

In the responses, several negative emotions (disgust) were found in the subjective responses. Fear and disgust were observed in objective emotion, but anger, negative emotion, and trust, a positive emotion, were not estimated.

Figures 6–10 show the number of likes and RTs of the collected tweets and the results of the emotional information. It can be observed that there is some correlation between the number of likes and RTs in the collected tweets, but the number of RTs is smaller than the number of likes. A few RTs also indicated that the scale of diffusion was not high.

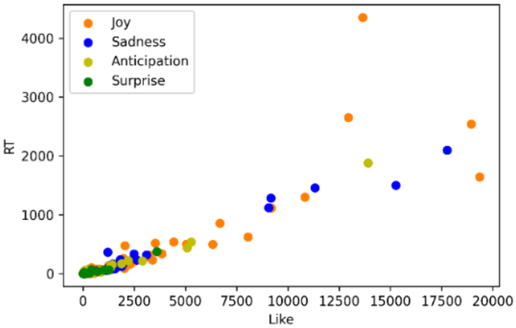

Figure 6. Distribution of subjective affective information in text.

Figure 6 shows that joy and sadness were frequently found in the subjective affective information of the text, and the tweets with many likes.

Figure 7 confirms the distribution of affective information in tweets with less than 1000 likes and shows that tweets estimated to have a surprise sensibility tend to have fewer likes.

Figure 8 shows that the distribution of objective affective information tends to show fewer likes for tweets with affective information of fear and disgust for tweets with less than 2000 likes.

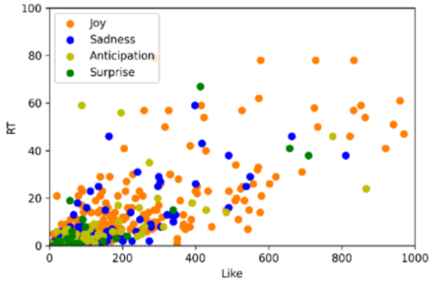

Figure 7. Distribution of subjective affective information in text.
(Tweets with less than 1000 likes)

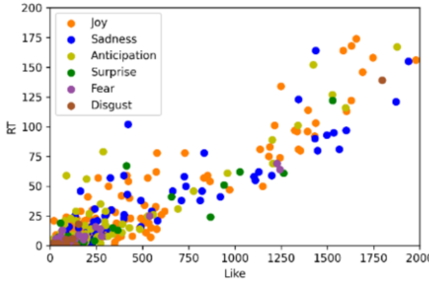

Figure 8. Distribution of objective affective information in text.
(Tweets with less than 2000 likes)

Next, we visualized the similarity between the affective vectors. Figures 9 and 10 show the visualization results. The similarity between the text affective information and image affective information vectors was calculated using cosine similarity. The results of the subjective affective analysis confirmed the trend that surprise has a higher similarity between text and images.

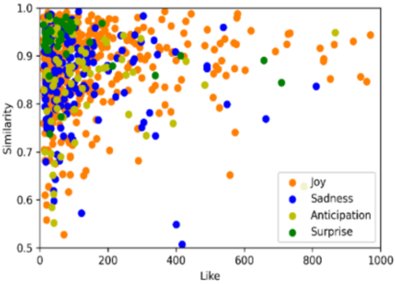

Figure 9. Affective vector similarity between text and image (subjective).

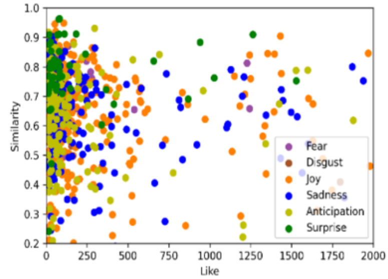

Figure 10. Affective vector similarity between text and image (objective).

Visualization of cosine similarity between text and replies based on affective information of the text. The results are shown in Figures 11 and 12. Because the number of replies varies from tweet to tweet, we used the average value of cosine similarity. The cosine similarity between text and replies varies in value. Fear and disgust have a widely distributed cosine similarity to the population, while joy tends to have high overall similarity.

An example of the collected tweets is shown in Figure 13. The text is sadness, which seems to make sense; but the image is surprise, which does not seem to capture the feature. Due to the lack of a large dataset of images with affective information, this study used a dataset of image captions and extracted affective information by converting image features to BERT vectors. However, given the nature of the dataset used, the affective information may be biased because it did not originally contain images or captions representing negative emotions.

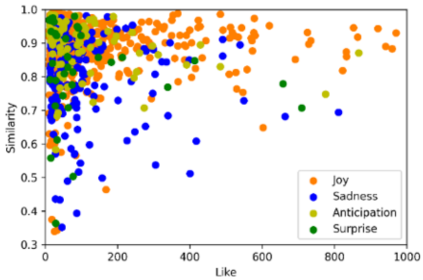

Figure 11. Affective vector similarity between text and reply (subjective).

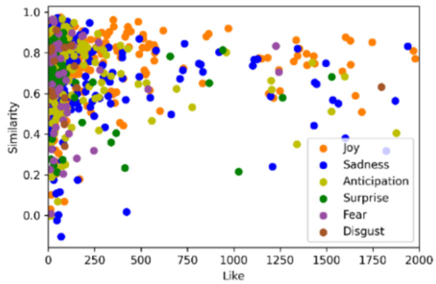

Figure 12. Affective vector similarity between text and replies (objective).

Figure 13. Examples of tweets and images.
Notes: Tweet text: He looked like he had seen something he shouldn't have.

5. Discussion

In this study, we estimated the affective information of text, images, and replies to tweets, and analyzed their relationship with RTs and the number of likes. A few tweets were classified as fear, disgust, anger, or trust, while others were classified as joy, sadness, anticipation, and surprise. Another potential factor is that tweets that are easily diffused do not contain many emotions, such as fear, disgust, anger, and trust, even though the performance of the affective estimation model and the training data may be problematic. We propose a method to estimate affective information from images indirectly by developing a model that converts image features into text features using data from the images and their corresponding captions as affective information cannot be estimated directly from image features in the case of affective estimation from images. However, image captions are sentences that merely describe the objects and scenes in the image and are less likely to include sensibility, which is different from the properties of spoken text sentences in the WRIME corpus.

In the analysis visualizing the similarity between vectors, those with a higher number of likes in the subjective sensibility information tended to have higher similarity between the sensibility vectors. As images are added to the text to help convey meaning, the similarity in affective vector between text and images is likely to be high. Conversely, those with low similarity in affective vector are sometimes judged to be surprising but are considered to convey less meaning, and as a result, tweets are less likely to spread.

In the analysis of the similarity of affective feature to replies, the number of replies posted differed from tweet to tweet, and the average of the sensibility vectors obtained from each reply was used in this study. Therefore, it is necessary to develop a method for analyzing affective information that also considers the number of replies. The similarity between the affective feature of the text and the reply tended to be higher for joy, suggesting that if the original tweet was positive, the reply tended to have more positive affective information.

6. Conclusion

In this study, we propose an analytical method to investigate the factors behind the buzz phenomenon on Twitter. Considering that it is difficult to analyze the buzz factor only from the text content posted on Twitter, we limited our analysis to tweets with attached images, and analyzed the relationship between the sensibility information estimated from the text and attached images, and the number of "likes," an indicator of popularity, and "RTs," an indicator of the scale of diffusion.

In the proposed method, a model for estimating affective information is trained using feature vectors extracted from text, images, and replies as inputs, and the estimated affective information is used for analysis. In addition, because this study did not focus on user attributes, it is necessary to analyze the buzz phenomenon from various angles, including user information (attributes, posting history, and follow/follower relationships), in the future, given the nature of the buzz phenomenon, in which the relationship between posters and followers has a strong relationship with the speed and scale of diffusion.

The size of the experimental dataset was small and the collection period was limited, resulting in biased data. In the future, we intend to use the Twitter API track for academic research to relax the restrictions on the collection period and number of requests and to reconsider the collection conditions. In addition, we intend to add additional analysis targets, such as the attributes of users who post tweets and the time at which tweets are posted, to clarify the characteristics affecting the number of likes and RTs.

In this study, we used affective estimation models based on BERT, but performance could be improved by using methods that further extend BERT, such as XLNet [18], BART [19], ALBERT [20], ELECTRA [21], RoBERTa [22], or T5 [23], which is trained on a larger corpus, GPT-3 [24], which is based on a larger corpus. However, fine tuning of pre-trained models is highly dependent on the suitability of the language corpus used for pre-training for the affective estimation task, as well as on the quality of the emotion corpus used for fine tuning and the bias of the class labels. For this reason, it will be necessary in the future to further fine-tune models that convert images to text features by using corpora that pair images with affective sentences, or by using models that generate images from sentences in the reverse direction, such as GLIDE [25].

Acknowledgments

This work was supported by the SCAT Research Grant 2022, JSPS Grants-in-Aid for Scientific Research JP20K12027, and JP 21K12141.

References

[1] Jindal H, Bharti M, The prediction of buzz in social internet of things, Proceeding of 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2020 Aug. 26-27, Dubai, United Arab Emirates.

[2] Yin Z, Baozhou L, Haidong Z, Can buzzing bring business? Social interactions, network centrality and sales performance: An empirical study on business-to-business communities, J Busin Res. 2020, 112: 170-189.

[3] Matsumoto K, Hada Y, Yoshida M, Kita K, Analysis of reply-tweets for buzz tweet detection, Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation (PACLIC), 2019 Sept; 13-15, Japan, pp. 138-146.

[4] Deusser C, Nora J, Reubold J, Benjamin S, Oliver H, Thorsten S, Buzz in social media: detection of short-lived viral 444 phenomena. Proceedings of the Web Conference, WWW'18, pp.1443-1449, (2018).

[5] Nora Jansen, Oliver Hinz, Clemens Deusser, Thorsten Struffe, Is the Buzz on? – A Buzz Detection System for Viral Posts in Social Media, Journal of Interactive Marketing, Vol.56, Nov 2021, pp. 1-17.

[6] Masumoto C, Yamaoka H, Ooi S, Sano M, A Study on Personal Interests Estimation based on Machine Learning using SNS, Proceedings of the 6th International Conference on Communication and Information Processing, pp.74-78, November 27-29, (2020).

[7] Bhattacharya P, Zafar M, Ganguly N, Ghosh S, Gummadi K, Inferring User Interests in the Twitter Social Network, Proceedings of the 8th ACM Conference on Recommender systems, pp.357-360, 2014.

[8] Zhang P, Wang X, Li B, On Predicting Twitter Trend: Factors and Models, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp.1427-1429, 2013.

[9] Ashok A, Sanjay S, Is That Twitter Hashtag Worth Reading, Proceedings of the Third International Symposium on Women in 440 Computing and Informatics, pp.272-277, 2015.

[10] Ertugan A, Using statistical reasoning techniques to describe the relationship between Facebook advertising effectiveness and benefits gained, Procedia Computer Science, Vol.120, 2017, pp. 132-139.

[11] Twitter API, https://developer.twitter.com/en/docs/twitter-api

[12] Katumata S, Sakata H, Creation of Japanese Spoken BERT using CSJ, Proceedings of the 27th Annual Conference of the Association for Natural Language Processing, pp.805-810, (2021).

[13] Maekawa K, Overview of the Corpus of Spoken Japanese， Japanese Language Science，Vol. 15, pp.111-133, 2004-04, National Institute for Japanese Language (2004).

[14] Szegedy C, Vanhoucke V，Ioffe S，Shlens J, Re-thinking the Inception Architecture for Computer Vision, Proceedings of IEEE Conference on ComputerVision and Pattern Recognition (CVPR), pp. 2818‑2826, 2016.

[15] Yoshikawa Y,,Shigeto Y, Takeuchi A, STAIR Captions: Large-scale Japanese image caption dataset, The 23rd Annual Conference of the Association for Natural Language Processing, pp537-540, (2017)

[16] Lin T, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, C. Zitnick L, Dollár P, Microsoft COCO: Common Objects in Context, Computer Vision – ECCV 2014 pp 740–755, (2014)

[17] Suzuki H, Miyauchi Y, Akiyama K, Kajiwara T, Ninomiya T, Takemura N, Nakajima Y, Nagahara H, Japanese dataset for classification of subjective and objective emotional polarity, The 28th Annual Conference of the Association for Natural Language Processing, (2022).

[18] Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Quoc V. Le, XLNet: Generalized Autoregressive Pretraining for Language Understanding, 2020, https://arxiv.org/abs/1906.08237

[19] Lewis M, Liu Y, Goyal N, Ghazvininejad M , Mohamed A, Levy O, Stoyanov V, Zettlemoyer L, BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, 2019, https://arxiv.org/abs/1910.13461

[20] Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R, ALBERT: A Lite BERT for Self-supervised Learning of Language Representations, https://arxiv.org/abs/1909.11942

[21] Clark K, Luong M, V. Le Q, D. Manning C, ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators, 2020, https://arxiv.org/abs/2003.10555

[22] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V, RoBERTa: A Robustly Optimized BERT Pretraining Approach, 2019, https://arxiv.org/abs/1907.11692

[23] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, J. Liu P, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, 2020, https://arxiv.org/abs/1910.10683

[24] B. Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, M. Ziegler D, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D, Language Models are Few-Shot Learners, In Proceedings of NeurIPS 2020, pp. 1877-1901.

[25] Nichol A, Dhariwal P, Ramesh A, Shyam P, Mishkin P, McGrew B, Sutskever I, Chen M, GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models, In Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.