A Comparative Study of Estimation of Video Viewer Emotion Using YouTube Video Comments

1st Yuki Kanno
Kogakuin University
Tokyo, JAPAN
em23019@ns.kogakuin.ac.jp

2nd Ryohei Banno
Kogakuin University
Tokyo, JAPAN
banno@computer.org

***Abstract*—YouTube is a online video sharing service that is that many people view. If we can know the emotions of video viewers, we can use them to improve usability. In this study, we propose two method for estimating the emotion of YouTube videos from the comments of each video: a BERT-based method and a rule-based method. For the former, we use BERT with fine-turning by 350 video comments to estimate the emotion. In the rule-based method, emotion values are calculated using a Japanese emotional expression dictionary. To evaluate these methods, we obtained emotion values from 100 respondents who were asked to fill out questionnaires as ground truth. The results of the evaluation using cosine similarity showed that BERT-method was able to estimate emotions with higher accuracy than the rule-based method.**

***Index Terms*—Natural Language Processing, Text Classification, Multimedia, BERT**

I. Introduction

YouTube is a large scale service with a very large social impact. In recent years, the spread of smartphones has been rapid, with the smartphone penetration rate among all households reaching as high as 88.6 percent (as of 2021) [1] . As for media, the internet user rate exceeded the TV viewer rate on average for all ages for two consecutive years. These shows that the Internet has become popular among a very wide range of age groups. Specifically, the number of active users of YouTube [1] is 2.562 billion worldwide as of January 2022 [2] . The number of users in Japan has exceeded 70 million, and it is one of the most popular video sharing service in Japan. A study shows that there were 2 billion videos on YouTube in 2016 [3] . The economic impact of video has been estimated at 350 billion yen and 100,000 jobs have been created on YouTube [4] .

Gross et al. [5] indicate that video has a great impact on people and that video is a medium suitable for emotional responses. Indeed, some videos on YouTube are entitiled with words like ”funny” and ”tear-jerking”, so there are users who use videos on YouTube to immerse themselves in specific emotions.

In this study, we focus on the emotions that user obtain from videos. If we can obtain the emotions, we can use them to improve usability, such as improving the accuracy of video recommendations, and to utilize them for business and marketing purposes. However, it is difficult to obtain emotions from the video themselves and their titles. We propose a method to estimate the emotion of YouTube video viewers the comments of each video: a BERT-based method and a rule-based method. The accuracy of the estimation output from these two methods is then compared.

II. Related Work

Sakai et al. propose the use of video comments as a means of identifying inflammatory videos on YouTube [6] . The technology used to automatically detect inflammatory videos on Nico Nico Douga[2] is applied to YouTube. The idea is based on the fact that inflammatory videos often contain negative comments. The comments added to videos are assigned an emotion value in the range of -1 to +1 using an emotion dictionary. Then, the positive/negative values of the entire comment are determined from these values. We also propose the use of a method to obtain a distributed representation of words, such as Word2Vec.

Nakazawa et al. use a BERT model trained on Wikipedia. They obtained 500 pieces of labeled training data [7] . These were used to determine the emotion value using a fine-tuned BERT model. As a comparison, they use parsing to determine the emotion value and morphological analysis to estimate the emotion polarity based on the polarity values of verbs and modals.

III. Proposed Method

The proposed method classify emotions into seven categories based on the Japanese Emotion Expression Dictionary (JIWC-Dictionary): sad, anxious, angry, disgusted, trusting, surprised, and happy [9] . The emotion that each video causes to its viewers is calculated as a seven-dimensional vector. The overall diagram of the proposed methods is shown in Fig.1.

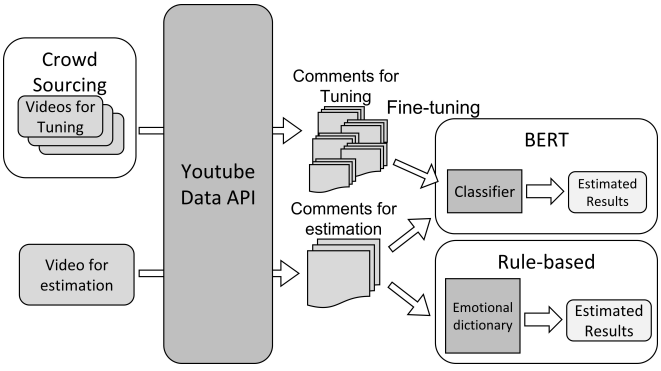

Fig. 1. Overview of the proposed method.

[1]https://www.youtube.com

[2]https://www.nicovideo.jp

TABLE I
VIDEO FOR EVALUATION OF EMOTIONAL ESTIMATES

Movie number	Japanese title	English title
Movie 1	【衝撃】火山にゴミを捨てて処理する場合に起こること	【Shocking】 What happens when garbage is dumped into a volcano for disposal.
Movie 2	back number - 手紙 (full)	back number - Letter (full)
Movie 3	貫禄ありすぎて、父親と間違われる引きこもり生徒【ジェラードン】	A withdrawn student who is so dignified that he is mistaken for his father 【Geraldine】

A. Preprocessing of comment data

In this study, we use YouTube Data API v3[3] provided by Google Inc. to obtain comment data from videos. Using crowdsourcing, we collect 50 video IDs for each of the seven emotions from 50 people, i.e., we collect 350 video IDs in total. By using the video IDs, we can retrieve comments from the YouTube Data API. since there is an upper limit on the number of input tokens for BERT, we collected 100 comments from each video, 500, and obtained a total of 35000 comments. These texts were preprocessed to remove numbers, symbols, URLs, emoticons, etc. to make it easier to process in the BERT-based method and the rule-based method.

B. BERT-based method

In this study, we use a pre-trained model. Fine-tuning is performed by creating a category list with the same name as the file containing the training text, and reading a file with the same name as the category list and the same emotion name at the beginning of the category list from the file with the same name. After fine-tuning, seven emotion values are calculated and displayed by inputting the text file for estimation.

C. Rule-based method

The Japanese Emotion Dictionary is used. The dictionary contains seven-dimensional emotion vectors corresponding to words. The preprocessed comments are segmented using Janome (Version 0.4.2) [10]. Using the segmented comments as input, the proposed system searches for each word of the input in the Japanese emotional expression dictionary. When a word is matched, the seven-dimensional emotion vector corresponding to the word is added to the vector of the video. The same process is repeated for all morphemes in the comment, and the final calculated emotion vector is calculated as the emotion value of the video.An overview of the Rule-base method is shown in Fig.2.

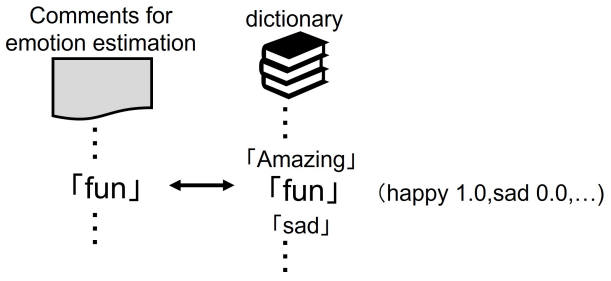

Fig. 2. Overview of the Rule-based method.

[3]https://developers.google.com/youtube/v3

IV. EXPERIMENTS AND EVALUATIONS

A. Experimental Methods

1) Data for evaluation: To obtain the ground truth of emotion estimation, we used crowdsourcing to survey 100 people about the following three videos. These videos were selected as we expected that they give different emotions to viewers.Table I shows the three videos selected.We used DeepL（Version 4.6.0.9212） for the English translation of the video titles.

A questionnaire was sent to 100 people using crowdsourcing. The respondents were asked to rate the intensity of the seven emotions obtained from the three videos for estimation on a scale of 0 to 4, where 0 is the strongest. The average of the values was used as the ground truth.

2) Fine-tuning and emotion estimation for BERT: We use a provided by the Inui Lab at Tohoku University for pre-training BERT [8]. We implemented a fine-tuned BERT model with a batch size of 32, an epoch count of 10, and 50 video comments from each emotion with 256 tokens from the beginning, and performed classification. We used 80% of the data for training, 10% for validation, and 10% for testing. Since the parameters of the model change with each fine-tuning, 10 fine-tunings were performed to calculate the average result. The emotion values of the videos for estimation were estimated for each fine-tuning, and the average of the ten estimation results was used as the final BERT emotion value.

3) Rule-based emotion estimation: Preprocessed comments were used as input for estimation. The comment was compared with the vocabulary in the Japanese emotional expression dictionary, and if a match was found, the emotion value set in the Japanese emotional expression dictionary was added as the emotion value of the comment. The final sum of the emotion values was used as the rule-based emotion value.

B. Experimental results

For emotion vector comparison, the calculated emotion vectors were converted to unit vectors. The components of the calculated emotion vectors are shown in Tables II to IV and Figures 3 to 5 for each video, The cosine similarities between the ground truth and results of the two methods are shown in Table V. The following formula was used to calculate the cosine similarities.

$$cos(a,b) = \frac{|a||b|}{\sqrt{\sum_{i=1}^{n} a^2}\sqrt{\sum_{i=1}^{n} b^2}} \quad (i = 1, 2, 3 \dots)$$

TABLE II
UNIT VECTOR OF CALCULATED EMOTION VECTORS (MOVIE 1)

method	sad	anxious	angry	disgusted	trusting	surprised	happy
Evaluation Data	0.1487	0.2588	0.1197	0.1352	0.5465	0.6083	0.4577
BERT	0.2613	0.4541	0.3655	0.3971	0.5164	0.3850	0.1392
Rule-based	0.2013	0.4317	0.3962	0.4930	0.4282	0.3047	0.3053

TABLE III
UNIT VECTOR OF CALCULATED EMOTION VECTORS (MOVIE 2)

method	sad	anxious	angry	disgusted	trusting	surprised	happy
Evaluation Data	0.7051	0.0673	0.0972	0.0747	0.4709	0.2791	0.4285
BERT	0.6312	0.3186	0.3508	0.3593	0.2430	0.2957	0.3184
Rule-based	0.2657	0.3129	0.3492	0.4296	0.5336	0.2827	0.4004

TABLE IV
UNIT VECTOR OF CALCULATED EMOTION VECTORS (MOVIE 3)

method	sad	anxious	angry	disgusted	trusting	surprised	happy
Evaluation Data	0.0621	0.1002	0.1623	0.1599	0.2458	0.5489	0.7566
BERT	0.2308	0.2597	0.4037	0.3822	0.2633	0.3532	0.6133
Rule-based	0.2193	0.2871	0.3797	0.4466	0.5344	0.3139	0.3765

TABLE V
SIMILARITY TO DATA FOR EVALUATION

Movie	BERT	Rule-based
Movie 1	0.8364	0.8189
Movie 2	0.8638	0.7812
Movie 3	0.8929	0.7687

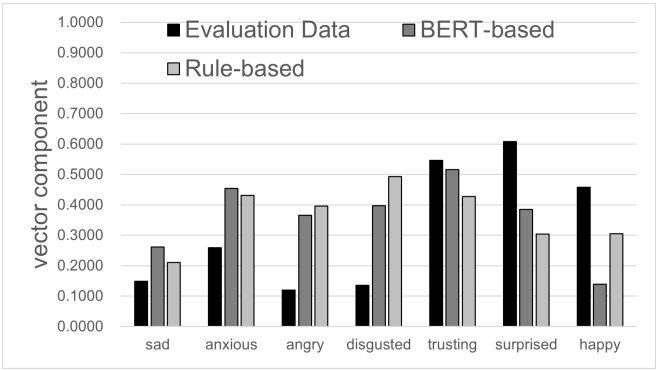

Fig. 3. Components of the emotion vector (Movie 1)

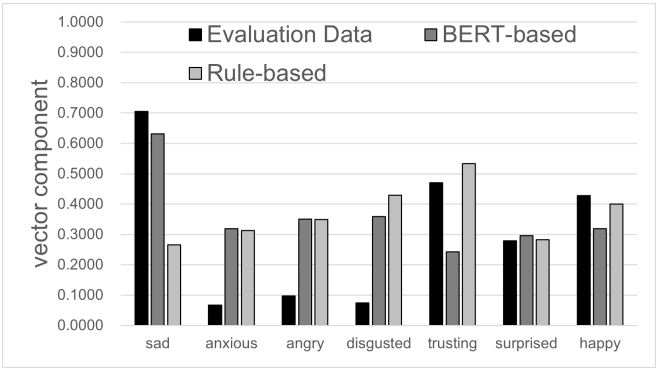

Fig. 4. Components of the emotion vector (Movie 2)

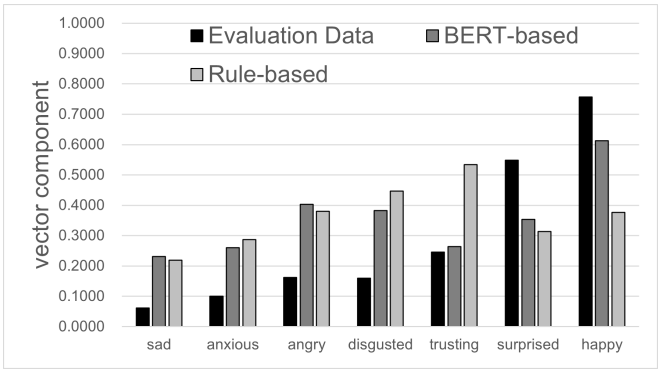

Fig. 5. Components of the emotion vector (Movie 3)

C. Discussion

In all videos, BERT calculated a higher score than the rule-based system. Since the rule-based system includes emotion values for words that are not used to express emotion, the calculated emotion values are considered to be different from the ground truth. In addition, the results of disgust and trust are high, and are far from the ground truth. The total of all the emotion values registered in the dictionary resulted in the highest disgust and the highest trust. Therefore, it is considered that disgust and trust have high values.

In the BERT estimation, fine-tuning learned the tendency of the comments for each emotion, and it is thought that it was possible to calculate scores that were closer to the evaluation data than the rule-based estimation. The tendency of comments that are common to YouTube as a whole may result in higher emotion estimates even for emotions that are not felt in the videos.

V. CONCLUSION

In this study, we proposed a BERT-based method for predicting the emotions that YouTube videos give to viewers, and compared its accuracy with that of a rule-based method.

We obtained comments from 350 videos and created a multi-label classifier by fine-tuning BERT. Using the classifier, we input comments to be estimated and output a

7-dimensional emotion vector. In the rule-based approach, a Japanese emotional expression dictionary was used, and emotion vectors were output from comments that were decomposed into morphemes for estimation.

The BERT-based emotion estimation produced a higher score than the rule-based estimation. Fine-tuning with a larger number of video comments and increasing the number of input tokens are expected to improve the accuracy.

As future work, we would like to explore accuracy improvement by increasing the number of comments for training and changing the parameters for fine-tuning. In addition, we plan to verify the accuracy of language processing methods other than BERT.

REFERENCES

[1] MIC, "Information and Communications in Japan WHITE PAPER 2022", https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2022/2022-index.html , (July.2022)

[2] Think with Google, "Why is YouTube so popular?"(in Japanese), https://ww w.thinkwithgoogle.com/intl/ja-jp/marketing-strategies/video/youtube-recap2022-1/, (accessed Dec. 20, 2022)

[3] BARRON's, Tiernan Ray, "YouTube ' s 2 Billion Videos, 197M Hours Make it an ' Immense ' Force, Says Bernstein", https://www.barrons.com/articles/youtubes-2-billio n-videos-197m-hours-make-it-an-immense-force-says-bernstein-1462978280, (accessed Jan. 8, 2023)

[4] Oxford Economics, YouTube, "YouTube Impact Report: The Economic Societal and Cultural Impact of YouTube in Japan in 2021"(in Japanese), https://www.oxfordeconomics.com/resource/a-platform-for-japane se-opportunity-assessing-the-economic-societal-and-cultura l-impact-of-youtube-in-japan-in-2021-jp/, (accessed Jan. 4, 2023)

[5] Robert W.Levenson and , J.Gross, "Emotion Elicitation Using Films.", IN COGNITION AND EMOTION, pp. 87–108, 1995.

[6] Yunosuke SAKAI, , Kanta TAKEUCHI, "A Study of Flaming Comment Detection using Text based machine learning", IPSJ SIG Technical Report Vol.2021-ICS-203 No.9(2021)

[7] Masataka Nakazawa, Katsuari Kamei, Yoichirou Maeda, and Eric W.Cooper, "An estimation of emotional polarity for simple sentence using BERT and its effectiveness", Proceedings of the 36th Symposium on Fuzzy Systems(FSS2020 Online), pp. 177–180, 2020.

[8] Tohoku University, "A BERT published by the Tohoku University", https://github.com/cl-tohoku/bert-japanese, (accessed Dec. 28, 2022)

[9] SOCIOCOM Social Computing Laboratory since 2015, "JIWC-Dictionary", "https://sociocom.naist.jp/jiwc-dictionary-en/ (accessed Dec. 12, 2022)"

[10] Tomoko Uchida,"Janome v0.4 documentation (ja)", "https://mocobeta.github.io/janome/ ", (accessed May. 26, 2023)