

# A Pretrained YouTuber Embeddings for Improving Sentiment Classification of YouTube Comments

Ching-Wen Hsu\*, Hsuan Liu\*, and Jheng-Long Wu\*

## Abstract

Technology is changing the way we consume information and entertainment. YouTube streaming video services provide a discussion function that allows video publishers to know what matters most to the people they want to love their brand. Through comments, video publishers can better understand the audience's thoughts and even help video publishers improve their video quality. We propose a classifier based on machine learning and BERT to automatically detect YouTuber preferences, video preferences, and excitement levels. In order to make high performance of models, we use a pretrained YouTuber embeddings to enhance performance, which is trained in advance based on roughly 175,000 pieces of videos' comments that contain YouTubers' name. YouTuber embeddings can capture some of the semantics and character of the relation between YouTubers. Experimental results show that the performances of machine learning-based models with YouTuber embeddings have improved overall accuracy and F1-score on all sentiment classifications. The result validates that YouTuber embedding training is significantly helpful when detecting audience sentiment towards YouTubers. On the contrary, BERT model cannot perfectly deal with the polarity classificational tasks when using YouTubers embeddings. However, the BERT model construction is more suitable for addressing multi-dimensional classification tasks, such as the five-labels classification task used in this task. Conclusion, the sentiment detection task on the YouTube can improve performance by the proposed multi-dimensional sentiment indicators and our solution to modify the structure on classifiers.

**Keywords:** YouTuber Embeddings, Sentiment Classification, Deep Learning, Pretrained Model

---

\* School of Big data Management, Soochow University

E-mail: 06170139@gm.scu.edu.tw; 06170114@gm.scu.edu.tw; jlwu@gm.scu.edu.tw

The author for correspondence is Jheng-Long Wu.

## 1. Introduction

Due to the rapid rise of new media, streaming platforms and video providers have increased. According to one report, 68% of people prefer watching a video rather than reading a long product manual to acquire information. People change the way of their entertainment even daily habitual. No one wants to be tied to a TV schedule, so people nowadays favor subscribing to streaming video services, such as Netflix or YouTube, to enjoy watching videos anytime and anywhere. Also, mobile phone viewers or smartphone viewers have increased astonishingly. YouTube reports that mobile video consumption is rising with an impressive rate of 100 percent every year. The large amount of data captured by video platforms provides insights for video streaming apps, and video stream services make recommendations based on audience's viewing profiles. Because of High-speed internet connectivity, more and more people have been allowed to become YouTubers and create large volumes of high-quality videos. YouTube has 16 million active users in Taiwan monthly, and nearly 93% of users have visited YouTube. It seems that YouTube has played an increasingly important role in modern life and entertainment. Therefore, we aim to analyze audience's habitual preferences on consuming information and entertainment on YouTube.

According to the audience's watching records, YouTube can create customized recommended content, which means consumers' interests have been collected and analyzed by YouTube. On the contrary, YouTubers, who upload videos to the YouTube platform, also want to check their videos' performance. YouTube has provided several analysis functions such as average view duration, browsing history, variance in audience's demographics, Etc for YouTubers to check their channel's performance. However, it lacks sentiment analysis on the audience's comments. It is verified that public views, comments, and attitudes towards many events can be analyzed through social media (Heredia *et al.*, 2016). Public reviews on Amazon were used to evaluate users' opinions and determine the audience's preference by classifying opinions into negative, positive, and neutral (Bhatt *et al.*, 2015). Therefore, we deduced that YouTube could also serve as a sentiment analysis platform because it provides an increasing number of comments.

We utilized comments to monitor YouTube viewers' emotions in the previous task by designing three sentiment indicators, YouTuber preference, video preferences, and excitement level. In this task, we are not changing sentiment indicators but aim to optimize the result of sentiment detection, hoping to get higher overall accuracy to analyze audience's feelings. We not only use comments to monitor emotions, but we also consider characteristics in YouTube channels as an additional feature. Before restarting the experiment, we trained YouTubers' correlation and established YouTuber embeddings, a critical vector in determining what characteristic YouTubers shared between each other. Also, the similarity between different channels can be calculated by placing similar YouTubers close together in the embedding space.

The social sentiment is excellent in providing a better understanding of how their audience perceives the YouTuber channel or brand. In general, sentiment analysis focuses on determining positive, negative, or neutral emotions (Cunha *et al.*, 2019). Therefore, before this task, we also conducted some experiments that used YouTube comments to identify users' positive, negative, or neutral emotions and how strong those emotions are. Unlike previous tasks, we change our method in the experiment stage. We combine comments and use our established YouTuber word embedding. Not only to capture emotions behind everything social viewers but also to measure YouTubers intimately by translating YouTubers' features into a relatively low-dimensional space. The analyzing result may help video loaders who want to identify their viewers' depth of feeling and provide a chance for YouTubers to engage with their viewers directly.

By modifying the structure of models that contain pre-trained YouTuber word embeddings as part of the sentence input, we expect a better model's performance than the previous tasks, not containing pre-trained YouTuber word embedding. Anticipate that YouTuber word embedding can provide additional information when analyzing sentiment tasks.

## **2. Related Work**

Various models deal with text-based sentiment classification tasks. Machine learning-based models are used to address the text classification task (Zhang & Zheng, 2016). Other deep learning models have been used for sentiment analysis and obtained acceptable performances (Hassan & Mahmood, 2017). Recently, it has refreshed the best performance of using pre-trained language models, such as Bidirectional Encoder Representations from the Transformers (BERT) because its pre-trained method has captured linguistic structure from learning and detecting different tasks. Sun *et al.* (2019) have explored BERT pre-trained structure to deal with classification task and achieve excellent performance through the way of fine-tuning in the downstream tasks. In our previous task, we also fine-tuning BERT model to detect a multi-dimensional aspect of the audience's comments. Although the experiment results outperformed machine learning-based classifiers and even had similar outcomes in the deep learning-based classifier, it may lack task-specific knowledge and domain-related knowledge to further improve the BERT model's performance. Considering viewers may present different passion intensities through many kinds of channels they watch, so we take channels' information, which means the types and features of YouTubers, into consideration. For example, YouTubers who always share ironic videos, their viewers may reflect stronger emotions than educational videos. Peters *et al.* (2018) realized that word representations are key component in many neural language understanding models, so they introduced a new type of word representations which can deal with syntax and semantics. However, our way of dealing with complex characteristics of word use is adding YouTubers' information into each comment. We proposed pre-trained YouTuber embeddings to fully present domain-related knowledge in YouTube, so we can

confirm whether characteristic of YouTubers can improve models' comprehension. Specifically, we concatenate the original sentence embedding and YouTuber embeddings which serve as additional features when analyzing comments' emotion tendency.

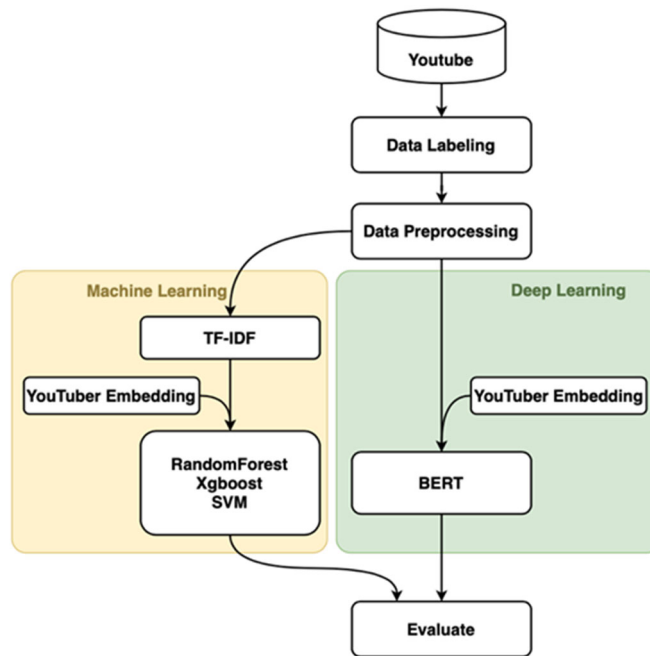
Compared with the limited dataset for training a relatedness between terms, more researchers have focused on using a pre-trained word embedding to understand semantic relatedness and similarity between terms in recent years. *Zhu et al. (2017)* show that increasing the size of datasets can identify more relations of biomedical terms even though it does not guarantee models' better precision. As a result, because of the small size of dataset, researcher often have to use pre-trained word embeddings to better capture meaningful vectors. *Rezaeinia et al. (2017)* have increased the accuracy on sentiment analysis research by using pre-trained word embeddings. Their method is experience different word representation methods, such as Part-of-Speech (POS) tagging techniques, lexicon-based approaches, and Word2Vec/Glove methods to compare their effectiveness.

Recently word embeddings methods have been widely applied in downstream models. *Aydoğan & Karci (2020)* used Word2Vec method on a large corpus of approximately 11 billion words to train word vectors, then applied to deep neural networks. The result did show that embedding method affected the rate of accuracy. Another research used pre-trained word embedding as a critical component for its downstream models. (*Miyato et al., 2017*). Cited from the above experiences, initially, we decided to utilize word vectors from the 2021 Wikipedia Chinese corpus to represent YouTuber similarity because of the large size of corpuses. However, we only focus on capturing the strong connection between each YouTuber and extracting characteristic behind YouTubers. According to our selected 25 YouTuber's channels, we select comments beneath each channel latest ten videos. Then, we filter these substantial comments by checking whether comments involve different YouTubers' names. Comments that up to standard are remained to train YouTubers embeddings. To compare whether the sentiment detection tasks can perform better by adding generating exact vectors, we propose a novel method, concatenating comments with YouTubers embeddings, to apply on classifiers.

*Usherwood & Smit (2019)* focus on comparing BERT and top classical machine learning approaches on a trinary sentiment classification task. Their task aims to verify whether BERT can perform state-of-the-art result when one only has 100-1000 labelled examples per class. As the result, BERT outperformed top classical machine learning algorithms even when training with 100 examples per class. Another research shows the superiority of BERT and support to use BERT as a default technique in NLP problems (*González-Carvajal & Garrido-Merchán 2020*). With similar task, we apply our own generated word vectors and go on the previous algorithms to determine if these approaches may represent the better result or even both BERT and machine learning-based methods are valid options.

### 3. Methodology

Figure 1 shows the proposed method for sentiment analysis and classification processes. Firstly, we collected the audience’s comments from the YouTube platform and subsequently labeled these comments according to our designed three sentiment indicators. Data preprocessing works include transferring emojis to texts and establishing a YouTube-based dictionary for tokenization. Next, all comments are converted into vectors, and YouTuber embedding is prepared to concatenate in the proper layer according to models. Finally, by the experiment stage, we evaluate the performance of each classifier in three detection tasks and discuss a comparative study.



*Figure 1. The process of the proposed sentiment analysis in this paper*

#### 3.1 Comment Collection

To cover the diversity of YouTube channels, we generated our dataset by selecting different types of YouTube channels. The composition of the selected videos’ film creation types with game 1%, education 4%, DIY with 4%, science and technology with 5%, comedy 9%, entertainment with 28%, and blog with 49%. Through these 25 selected channels, we then filter five videos from each channel that have been highly popular or controversial since 2019 because people imminently show their interest in new tread and debatable topics. Therefore, the data source contains a total of 125 videos. In this way, we collected more controversial and polarizing comments, and it becomes easier for annotators to determine the sentimental

tendency of comments. However, to avoid different accumulated numbers of comments in each video, we randomly remain 100 pieces of comments from each video. Thus, a total of 12500 pieces of comments is taken into consideration.

### 3.2 Definition of Sentiment Indicators

YouTube has provided a discussion function for audiences to express their opinion by clicking like or dislike bottom under the videos. However, positive or negative sentiment classifications cannot explain why the audience does not like the videos and what reason keeps the audience subscribing to a specific channel. There is no noticeable analysis of likes and dislikes opinion, so we design three indicators, YouTube preference, video preference, and excitement level, to investigate different aspects of the audience's comment.

- **YouTuber preference:** In the indicator of YouTuber preference, comments can roughly divide into non-relative and relative towards YouTubers. Excluding non-relative comments, the rest comments that talk about YouTubers' names or affairs can continue to dig into positive, negative, and neutral attitudes according to their comments' content.
- **Video preference:** The indicator categories are the same as YouTuber preference. Non-relative, unlike, neutral, and like are four categories used to judge Video preference. For example, comments that do not talk about video content will be labeled as non-relative comments.
- **Excitement level:** This indicator is designed into five categories, from barely excited to hyper excited. We classify the audience's speaking tone from no emotion to extreme emotion state step by step. In addition, we consider emojis a judgment in this indicator because people tend to use emojis as their comments. For example, the second level of Excited level means the audience can speak confidently and contain two types of emojis.

### 3.3 Sentiment Indicator Labeling

The main drawback of using own data sources is having to label our dataset. Therefore, the main objective is to address semantic comprehension gaps between annotators. We introduce some guidelines to properly annotate our comments. For example, watching videos before annotation is required because it might resonate powerfully with the audience's opinions. During the annotation process, we eliminate some non-relative comments, such as advertisements, comments that not using Mandarin, comments that post links to external web pages, and merely timestamps in the comments, to optimize the availability of the dataset. In the last part, we use the majority decision to filter out inconsistent labels unless each comment annotation is marked as the same point.

In Table 1, we use three methods to calculate agreement scores after labeling comments, which include Krippendorff's Alpha, Fleiss's Kappa, and Cronbach's Alpha. With

Krippendorff's Alpha method, due to the reason that values smaller than 0.667 represent as discard data, so our three indicators are shown not up to the standard. Fleiss's Kappa method stands for fair and moderate data because values between 0.21 to 0.6 are considered acceptable levels. Cronbach's Alpha method evaluates three indicators as outstanding labeling work because a value higher than 0.7 may show annotation agreement, let alone we get 0.9 on Excitement level. Therefore, two of the methods were qualified as acceptance results, and thus we provide an adequately labeled dataset to train and assess a given model.

**Table 1. Annotation agreement scores for each indicator.**

	YouTuber preference	Video preference	Excitement level
Krippendorff's Alpha	0.5829	0.4545	0.3898
Fleiss's Kappa	0.5840	0.4594	0.3928
Cronbach's Alpha	0.8520	0.7264	0.900

### 3.4 Text Preprocessing

We consider emojis as part of emotional expressions. The first step of text processing is to transfer emojis to text, so dealing with rich emojis is our priority. We transfer emojis to text by the package called "emojiswitch." Then we establish a user-defined dictionary to recognize specific words, such as YouTubers' names and the texts transferred from emojis. In this way, we go through word tokenization, and thus now we can accurately determine unique objects from a user-defined dictionary. After these two parts, we go through word tokenization using the current state-of-the-art word tokenization tool created by the Chinese Knowledge and Information Processing (CKIP) Group. This tool is available for dealing with tokenization in Mandarin. In previous task, after completing all these above steps, we can start model training and evaluating.

### 3.5 Training YouTuber Embedding

There are various sentiment analysis techniques, but recently, word embeddings have been widely used in sentiment classification tasks. Word2Vec and GloVe are among the most accurate and usable word embedding methods to convert words into meaningful vectors. Therefore, we trained YouTubers embedding, a dense vector representation of words that capture something about their meaning, to present meaningful vectors to understand the relationship between YouTubers.

To have the best results when using the generated embeddings, we selected ten newly released videos, due to October 2021, from 25 YouTube channels that are the primary data source in this task. The comments' contents are selected based on having YouTubers' names,

whether lead actors/actresses or supporting actors/actresses. A total of 175,000 pieces of comments remains and applied to train YouTuber correlations. As a result, we present a YouTuber embedding dictionary that stores YouTubers' names and their corresponding 300-dimensional vector. This step aims to retrieve information about the audience's perceptions of different YouTubers because YouTubers' attitudes or behavior can stand for the character of the channel. In this way, the similarity between YouTubers has been predicted and presented in a low-dimensional vector. After training YouTuber embedding, we can use this vectorial representation to replace the YouTuber variable and obtain the corresponding vector from each comment. For example, we use each comment as a key to finding which YouTuber's channel is, and the YouTuber information can continue to map with its 300-dimensional vector. The next step is to apply this embedding; the input may be comment vectors after an additional YouTuber embedding to automatically train on classification models.

### 3.6 Training Classifiers

We propose a BERT-based model via constructing an additional embedding layer before calculating the probability distributions over categorical labels. In the beginning, we did not change the input; we sum the position embeddings, word embeddings, and segmentation embeddings for each token. Then we add YouTuber embeddings to each sequence after extracting the hidden state vector. Finally, using a SoftMax classifier to determine over categorical labels. Figure 2 shows the modified structure of BERT model. We only used comments to detect audience's emotions and did not change the structure of pretrained BERT model in the previous task. This time, we still remain comments and incorporate YouTube domain knowledge by adding YouTuber embedding to detect emotion variance more precisely.

Besides the BERT-based model, machine learning-based models: RandomForest, Xgboost, and SVM, are also used as a classifier to deal with dimensional sentiment analysis tasks. We transform comments into numerical vectors using TF-IDF, greatly improving the more basic methods like word counts in text analysis with machine learning. TF-IDF gives us a way to associate each word in a document with a number that represents how relevant each word is in that document. In the previous task, the TF-IDF score was fed to algorithms. However, we add a 300-dimensional vector, which stands for Youtubers' information, after retrieving the TF-IDF score of each comment at this time. Simply put, each comment may find their corresponded YouTubes' channel at first. Then, each channel can be mapped with our pre-trained YouTube word embeddings.



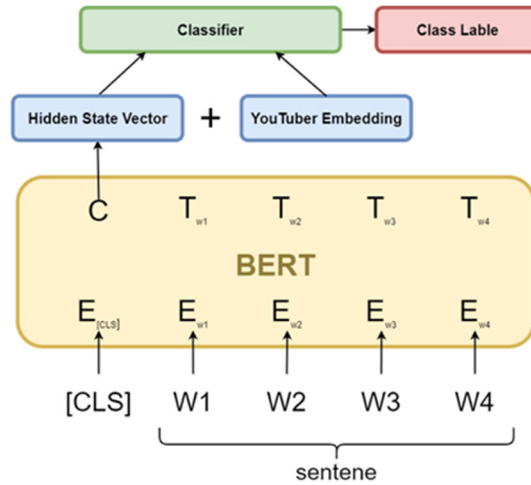


Figure 2. Structure of BERT model with YouTuber embedding.

### 3.7 Classification Tasks

Sentiment analysis is a fast-growing area and one of the well-known tasks of research in natural language processing (NLP) and text classifications. To better capture wide emotion variance on the audience’s comment, we use three sentiment indicators and five modified models to train classifiers and analyze five targets, T1 to T5 in this task. The following elaborates the meaning of five tasks for our experiment.

- **T1:** Whether comments are related to YouTubers is a binary classification task. The data sources are generated from the result of the indicator, YouTubers preferences. By rearranging the category of the labeled datasets, we merge the annotation result of unlike, neutral, and like comments into related comments. In contrast, non-related comments remain to be. This classification task is aims to discover the motivation behind watching videos. If comments are talking about YouTubers’ affairs, audience might pay attention to YouTubers.
- **T2:** Audience’s sentiment towards YouTubers is an extended issue from an indicator of YouTuber preference. Exclude non-relative comments; we extract unlike, neutral, and like comments from the annotation result. Like to dislike can serve as an indicator for YouTubers to check the followers of his or her channel. Also, YouTubers can know what attractive they own or what causes them to make a nuisance.
- **T3:** Whether comments are related to videos, also be rearranged from the indicator, video preference. We duplicate the same techniques for whether comments are related to YouTubers but present a completely different meaning. This task may explain whether the contents of the video arouse discussion or become no interest to the audience. If the topic interested to the audience, it may show more relative comments towards video.

- **T4:** Audience's sentiment towards videos excludes non-relative comments from the indicator, video preference; The rest of the comments can deal with the audience's sentiment towards video. Even if watching the same channel, the different themes will captivate and engage different audiences. Therefore, this task may help YouTubers understand their audience's preferences within a specific channel.
- **T5:** Corresponding to the indicator of excitement level, T5 aims to analyze the audience's emotional ups and downs from barely excited to hyper excited, which can firmly confirm the degree of support from different audiences and affirm the audience's attitude towards specific issues.

## 4. Experiment

### 4.1 Dataset

Moving to the composition of annotated comments according to three indicators. We applied three indicators to five analysis tasks, so comments have also been rearranged into five datasets. When analyzing the target, whether comments are related to Youtubers, the proportion of the non-relative comments to the relative comments is three to one. It presented that audiences prefer talking about video content rather than YouTubers' affairs. At the same time, it comes out that the most significant piece of comments was labeled as like in audience's sentiment towards YouTubers, which is extracted from the above relative comments. This composition made sense because if people do not like someone, they may not notice their condition, even watching their channel. Next, relative comments in whether comments are related to video account for the majority in the task, and 60 percent of comments with a neutral attitude talked about the video's content. This proportion presented that the audience does not frequently present animosity on the YouTube platform within our selected channels. The fifth analyzing task, emotional ups and downs, revealed that the audience could express their health and happiness by commenting. The following table shows the proportion of data to our five tasks.

**Table 2. Distribution of five tasks.**

<b>Task</b>	<b>Class</b>	<b>Number</b>
T1	Non-Relative	8223 (75%)
	Non-Relative	2776 (25%)
T2	Unlike	287 (10%)
	Neutral	784 (28%)
	Like	1705 (61%)
T3	Non-Relative	1036 (10%)
	Relative	9775 (90%)
T4	Unlike	659 (7%)
	Neutral	5842 (60%)
	Like	3274 (33%)
T5	Barely excited	2788 (30%)
	Slightly excited	2478 (27%)
	Excited	2341 (25%)
	Fairly excited	1136 (12%)
	Hyper excited	471 (5%)

## 4.2 Experiment Design

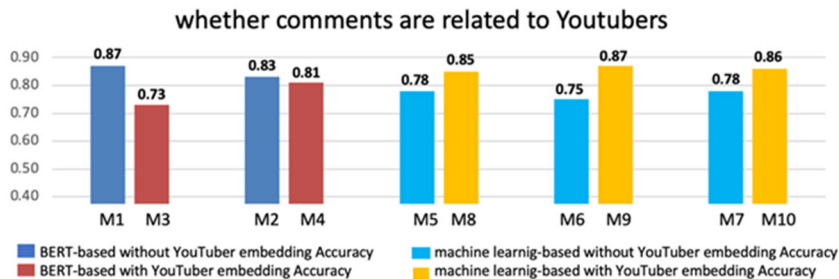
This section presents multiple models in Table 4 that we experiment with. Except for BERT models that we followed its pre-trained parameters, other models have experimented with different parameters. We configure the best parameters on each model through experiments and then apply them to analyze different aspects of sentiment tasks. Also, we use 5-fold cross-validation to ensure the performance for all models. By fixedly setting  $k=5$  to our dataset, 80% of data will be randomly selected for training and 20% for testing in each fold. In M1, M2, M3, M4, we set the number of epochs as 10 through the entire training dataset to make sure that the BERT model can have enough time to learn the pattern from social comments. After conducting experiments, we evaluate and interpret the performances of different models through the suitable metrics used for classification problems: overall accuracy. The results of social sentiment analysis are shown in the next section.

**Table 3.** There are four models use to solve three tasks.

Model	Description
M1	BERT model using <i>bert-base-multilingual-cased</i> pre-trained model
M2	BERT model using <i>distilbert-base-multilingual-cased</i> pre-trained model
M3	BERT model using <i>bert-base-multilingual-cased</i> pre-trained model + YouTuber embedding
M4	BERT model using <i>distilbert-base-multilingual-cased</i> pre-trained model + YouTuber embedding
M5	RandomForest
M6	Xgboost
M7	SVM
M8	RandomForest + YouTuber embedding
M9	Xgboost + YouTuber embedding
M10	SVM + YouTuber embedding

### 4.3 Experiment Result

Figure 3 and Figure 4 are the result of predicting the target, whether comments are related to YouTubers. The result shows that adding YouTuber embedding machine learning-based classifiers can better detect relative or non-relative comments towards YouTubers. On the contrary, after adding YouTuber embedding, the BERT model does not show better performances in the prediction result. We can also notice that M6 performed the worst in the previous task. However, it improved to become M9 and serve as the best classifier in the end.

**Figure 3.** Models' accuracy on whether comments is related to YouTubers.

Improving Sentiment Classification of YouTube Comments

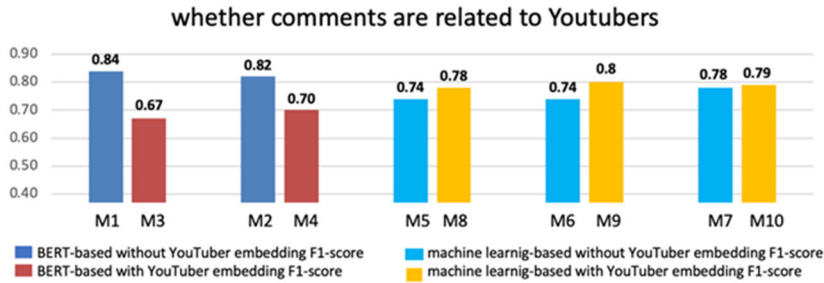


Figure 4. Models' F1-score on whether comments is related to Youtubers.

Figure 5. and Figure 6 show the result of the audience's sentiment towards YouTubers. The data in this detection task is comment about YouTubers' affairs, so we expected that adding YouTuber embedding after each comments can increase overall accuracy and F1-score. Machine learning-based classifiers proved the same result with our exception. The models 'performances have at least increased 7% in overall accuracy and 8% in F1-score. However, BERT, the variance seen from M2 to M4 surprisingly decrease.

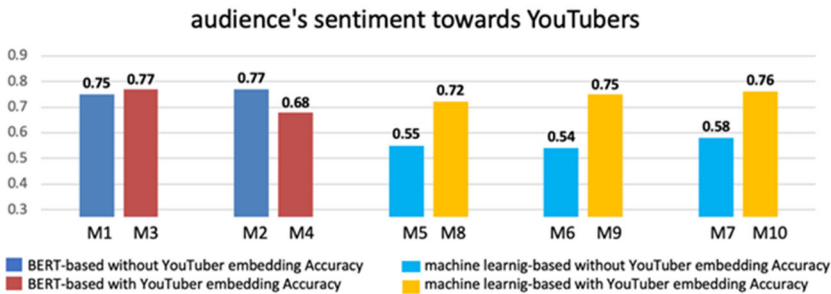


Figure 5. Models' accuracy on audience's sentiment towards Youtubers.

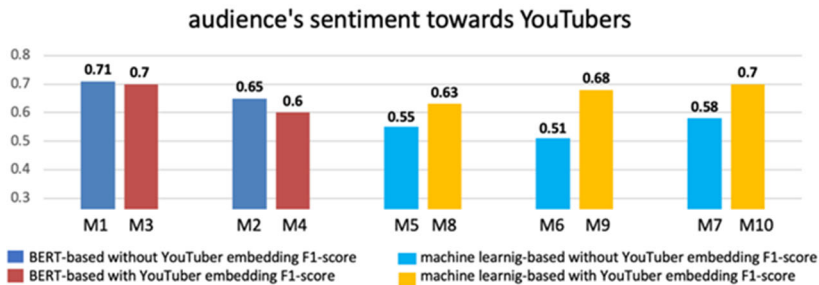
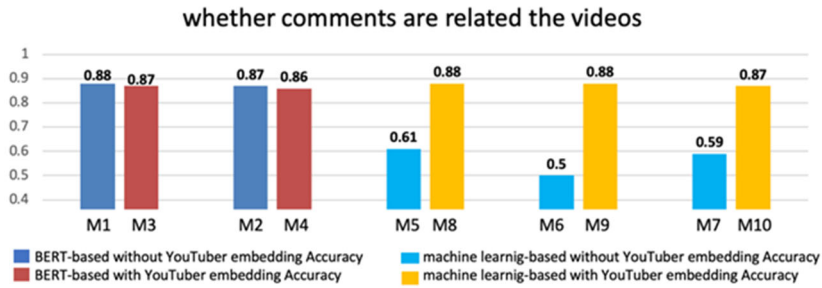


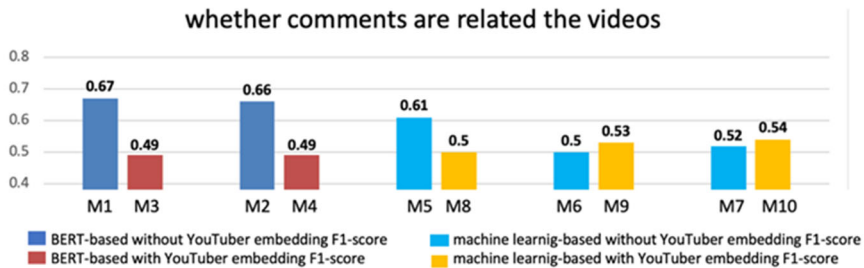
Figure 6. Models' F1-score on audience's sentiment towards Youtubers.

Figure 7 and Figure 8 are the result of predicting the target, whether comments are related to videos. We notice that overall accuracy in all models is upscale to nearly 90%. However, the

improvement in F1-score is limited, only increasing smaller than 3% or even regressing in BERT method when adding YouTuber embedding. We deduce the small amount of increment or even getting worse because YouTubers' information has little relationship with determining relative or non-relative comments towards videos.

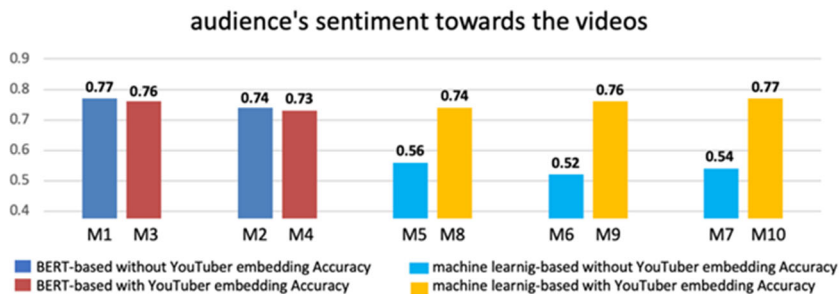


**Figure 7. Models' accuracy on whether comments is related to videos**



**Figure 8. Models' F1-score on whether comments is related to videos**

Figure 9 and Figure 10 are the result of predicting the audience's sentiment towards videos. Although data in this detection task is comments that discuss video content, the experiment result show that machine learning-based methods improved the predicted result after adding YouTuber embedding. In comparison, M4 and M4 do less well than before, decreasing from 5% to 10% and becoming the worst classifier.



**Figure 9. Models' accuracy on audience's sentiment towards videos.**

Improving Sentiment Classification of YouTube Comments

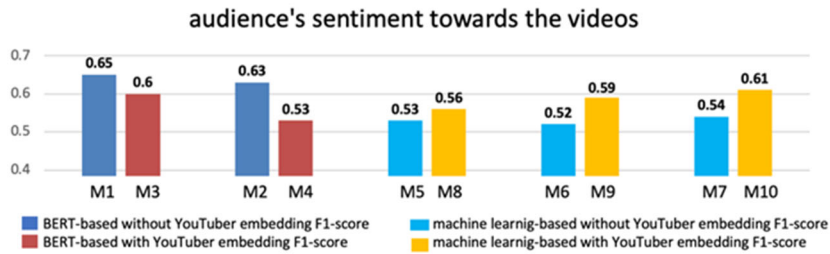


Figure 10. Models' F1-score on audience's sentiment towards videos.

Figure 11 and Figure 12 shows the result of predicting the audience's emotional ups and downs from their leaving comments. Compared with adding YouTuber embedding and without YouTuber embedding, the former method can improve model performance in machine learning-based methods. We deduce that the improvement may result from different types of YouTubers having different audiences. The more controversial YouTuber, the more excitement level may show in their audience's comments. For example, a YouTuber who prefers talking about political issues may vary their audience emotional variance than educational channels.

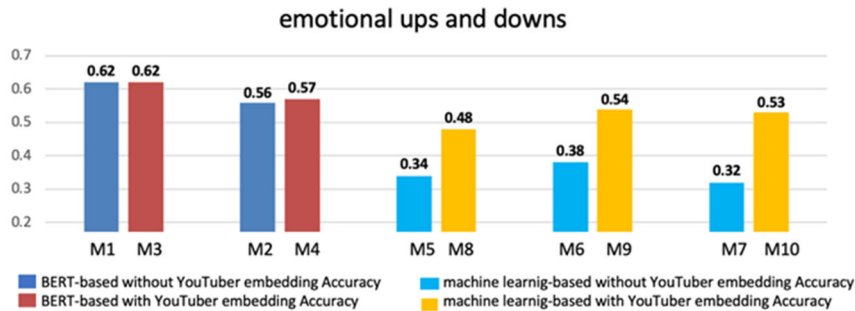


Figure 11. Models' accuracy on emotional ups and downs.

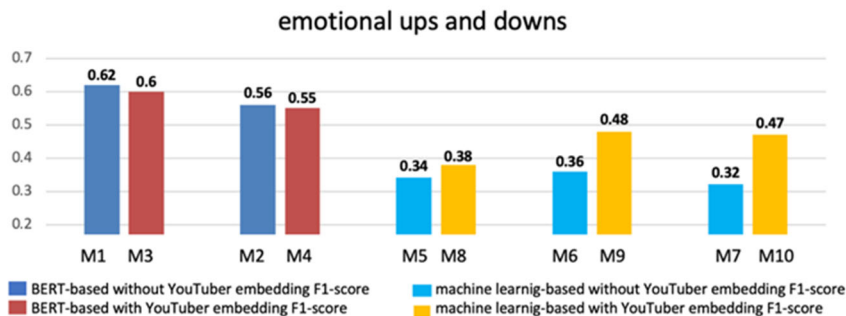


Figure 12. Models' F1-score on emotional ups and downs.

## 4.4 Discussion

In summary, there are three findings after we conducted experiments (1) Within machine learning-based models, the experiment results validate that adding YouTuber embedding is an effective way to identify audiences' emotions and depth of feeling. Also, we notice that YouTuber embedding is significantly helpful when detecting audience sentiment towards YouTubers. This result explains that we successfully trained YouTuber word embedding by using many comments with YouTubers' or guests' names who are invited on YouTuber's channel. (2) We notice that BERT neither improves the prediction score nor goes backward, a nearly ten percent decrease when predicting T1, T2, and T3. However, when predicting T5, two kinds of BERT (M3 and M4) do not regress their performance but remain top ranking. This result explains that BERT's model construction is more suitable for addressing multi-dimensional classification tasks. (3) Except for BERT models that performance well in determining audience's emotional ups and downs, BERT cannot perfectly deal with the polarity classificational tasks after adding YouTubers embedding. We also discover two characters that social media users own on the YouTube streaming platform. People prefer to discuss videos' content rather than YouTubes' affairs. In addition, people do not frequently present animosity in their comments; most people present their comments as neutral or barely excited attitudes.

## 5. Conclusion

This paper focuses on improving the over-all accuracy and F1-score on dimensional sentiment classification task. This time, we combine comments with YouTuber embeddings to train on the all classifiers. In machine learning-based classifiers, we use TF-IDF as sentence vectors and concatenate YouTuber Embedding in the last layer to fit in RandomForest, Xgboost, and SVM. On the contrary, we add YouTuber embeddings to the hidden state vector of BERT model. After that, we compare the above experiments' result with the previous tasks that only utilize comments as our data sources. Although BERT does not present a better prediction score on sentiment polarity problems, it perfectly deals with a multi-dimensional problem, the task of predicting the audience's excitement level. This result proves the superiority of BERT by achieving at least 10 % more in overall accuracy and F1-score than other classifiers. In comparison to the traditional machine learning classifiers, we identify that although machine learning models cannot perform as well as BERT before adding YouTuber embeddings, the performances of the machine learning-based classifiers can be dramatically improved after our proposed method which concatenating comments text with trained YouTubers embeddings to these classifier.

Analyzing the public's perception of YouTubers and the influence of their videos is a challenging task for researchers so far. Much work has been done in this paper, but it still has a long way to overcome some problems. In this research, we have emphasized the following



problems in order to make our results improve. In the future, we could explore more information on YouTube, such as combining videos' cover photo as features, to optimize multiple-dimensional sentiment analysis tasks. In this way, even if imbalanced dataset, models may identify feature represented on the picture and capture different aspects of information that cannot present in context only. In addition, with the recent emergence of deep learning, an increasing number of researchers have started to use deep neural networks to deal with sentiment analysis, we may explore deep learning techniques to automated detect the audience's preference on social media. Last but not least, others indicators, such as whether the comments contain an ironic statement or whether the comments contain an erotic statement, can be added for analyzing other aspects of the audience's comments. The latter proposed indicator may serve as a guard for children's users, and the former indicator may prevent YouTubers from getting into conflict with their fans.

## Reference

- Aydođan, M., & Karci, A. (2020). Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification. *Physica A-statistical Mechanics and Its Applications*, 541, 123288.
- Bhatt, A., Patel, A., Chheda, H., & Gawande, K. (2015). Amazon Review Classification and Sentiment Analysis. *International Journal of Computer Science and Information Technologies*, 6(6), 5107-5110.
- Cunha, A.A.L., Costa, M.C., & Pacheco, M.A.C. (2019). Sentiment Analysis of YouTube Video Comments Using Deep Neural Networks. In *Proceedings of the 18th International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, 561-570. [https://doi.org/10.1007/978-3-030-20912-4\\_51](https://doi.org/10.1007/978-3-030-20912-4_51)
- González-Carvajal, S., & Garrido-Merchán, E.C. (2020). Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv:2005.13012
- Hassan, A., & Mahmood, A. (2017). Deep learning approach for sentiment analysis of short texts. In *Proceedings of the 3rd international conference on control, automation and robotics (ICCAR 2017)*, 705-710. <https://doi.org/10.1109/ICCAR.2017.7942788>
- Heredia, B., Khoshgoftaar, T. M., Prusa, J., & Crawford, M. (2016). Cross-domain sentiment analysis: An empirical investigation. In *Proceedings of the IEEE 17th International Conference on Information Reuse and Integration (IRI 2016)*, 160-165. <https://doi.org/10.1109/IRI.2016.28>
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
- Rezaeinia, S.M., Ghodsi, A., & Rahmani, R. (2017). Improving the Accuracy of Pre-trained Word Embeddings for Sentiment Analysis. arXiv preprint arXiv:1711.08609.

- Sun, C., Qiu, X., Xu, Y. & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? In M. Sun, X. Huang, H. Ji, Z. Liu & Y. Liu (eds.), *Chinese Computational Linguistics* (p./pp. 194-206), Cham: Springer International Publishing.
- Usherwood, P., & Smit, S. (2019). Low-Shot Classification: A Comparison of Classical and Deep Transfer Machine Learning Approaches. arXiv preprint arXiv:1907.07543.
- Zhu, Y., Yan, E. & Wang, F (2017). Semantic relatedness and similarity of biomedical terms: examining the effects of recency, size, and section of biomedical publications on the performance of word2vec. *BMC Med Inform Decis Mak*, 17, 95. <https://doi.org/10.1186/s12911-017-0498-1>
- Zhang, X., & Zheng, X. (2016). Comparison of text sentiment analysis based on machine learning. In *Proceedings of the 15th international symposium on parallel and distributed computing (ISPDC 2016)*, 230-233. <https://doi.org/10.1109/ISPDC.2016.39>