The Impact of Social Media on Education: Analysis of Feedback on YouTube.

Sonal Lakhotia

University of Goettingen

19.07.2022

Contents

Literature References

- [1]Chei Sian Lee, Hamzah Osop, Dion Hoe-Lian Goh, Gani Kelni, Making sense of comments on YouTube educational videos: a self-directed learning perspective Link
- [2]K.M.Kavitha,Asha Shetty, Bryan Abreo, Adline D'Souza, Akarsha Kondana,Analysis and Classification of User Comments on YouTube Videos Link
- [3]Anastasiya Kotelnikova, Danil Paschenko and Elena Razova,Lexicon-based methods and BERT model for sentiment analysis of Russian text corpora Link
- [4]Doaa Mohey El-Din, Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis Link
- [5]Sayyida Tabinda Kokab, Sohail Asghar, ShehneelaNaz, Transformer-based deep learning models for the sentiment analysis of social media data Link
- [6]Twitter-roBERTa-base for Sentiment Analysis: Link

YouTube's role in learning

- YouTube is much more than a space dedicated to entertainment and recreation. It offers a powerful platform for resource sharing and eLearning. Many renowned universities such as Harvard and Stanford have made their courses available on YouTube.
- It promotes self-directed learning by enabling collaboration between the viewers and the creators at their convenience.
- The popularity of content on YouTube can be easily determined by channel rankings, the number of likes or dislikes, and the number of subscribers.
- To determine content consumption by viewers, the YouTube comments are analyzed. The sentiments of the viewers while learning or after achieving a positive result can be analyzed.

[1] Chei Sian Lee, Hamzah Osop, Dion Hoe-Lian Goh, Gani Kelni, Making sense of comments on YouTube educational videos: a self-directed learning perspective

[2]K.M.Kavitha,Asha Shetty, Bryan Abreo, Adline D'Souza, Akarsha Kondana,Analysis and Classification of User Comments on YouTube Videos

Related Work and Challenges

- Determining user perspective on YouTube content and community collaboration through sentiment analysis is discussed in [1] and [2].
- [1] uses SentiStrength to comprehend the sentiment of the comments.
- SentiStrength is a lexicon-based technique for sentiment analysis.It provides two numerical values: the first value ranges from -1 for neutral texts to -5 for particularly negative ones. The second value ranges from 1 to 5, where 1 is for neutral texts and 5 for highly positive ones.
- The analysis of 150 videos revealed that 142 videos had a higher percentage of positive attitude towards the content and learning experience.

[3] Anastasiya Kotelnikova, Danil Paschenko and Elena Razova,Lexicon-based methods and BERT model for sentiment analysis of Russian text corpora

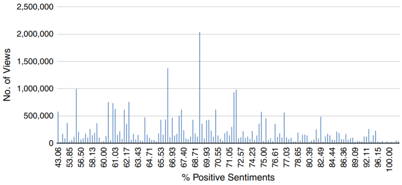


Figure: Viewership and Positive sentiments

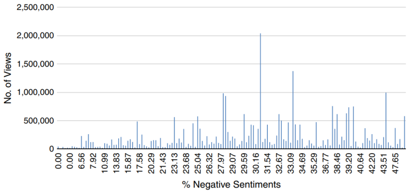


Figure: Viewership and Negative sentiments

[1] Chei Sian Lee, Hamzah Osop, Dion Hoe-Lian Goh, Gani Kelni, Making sense of comments on YouTube educational videos: a self-directed learning perspective

- [2] Classifies YouTube comments into positive, negative, relevant, and irrelevant categories. A word occurrence model(Bag Of Words Model) was used to determine the score of the word occurrence that was, the frequency of the presence or absence of these words in the test comment.
- The comments were extracted and manually classified into four classes using association lists, and the results were estimated using Precision, Recall, and Accuracy metrics.

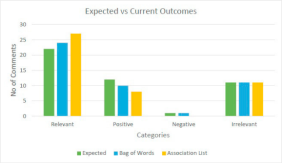


Figure: Depiction of Bag of Words Models and Association List in Comment Classification

[2] K.M.Kavitha,Asha Shetty, Bryan Abreo, Adline D'Souza, Akarsha Kondana,Analysis and Classification of User Comments on YouTube Videos

Limitations

- There are limitations to approaches used in both [1] and [2]
- Lexicon or dictionary-based approaches fail to address some vital challenges. Viewers do not always use the technical terms that are found in the existing lexicons.
- Along with traditional phrases, the viewers also use creative explanations and idiomatic expressions. They might be in different languages.
- Although the Bag-Of-Words model is the most widely used method for sentiment analysis, it has two weaknesses:
- 1) Usage of manual vocabulary in word evaluation
- 2) Emotion analysis with low accuracy due to neglecting the effects of grammar in language words and ignoring the semantics of words.

[4] Doaa Mohey El-Din, Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis

Proposed Approach

- A text sentiment classification model is proposed that uses knowledge distillation and text augmentation techniques to improve the accuracy of sentiment classification in a few-shot labeling task.
- The transformer model uses the self-attention method to understand other applicable words and applies them to the one being processed. Also discussed in [5].

Models	Syntactical	Semantics	Contextual	Out of vocabulary
1-Hot encoding	[×]	[×]	[×]	[×]
BOW	[×]	[×]	[×]	[×]
TF–IDF	[×]	[×]	[×]	[×]
Word2vec	[✓]	[✓]	[×]	[×]
GloVe	[✓]	[✓]	[×]	[×]
FastText	[✓]	[×]	[×]	[✓]
BERT	[✓]	[✓]	[✓]	[✓]

Figure: [5] Analysis of Word Embedding Models

[5] Sayyida Tabinda Kokab, Sohail Asghar, ShehneelaNaz, Transformer-based deep learning models for the sentiment analysis of social media data

- BERT uses static masking while RoBERTa uses dynamic masking and is more robust. We use RoBERTa model for sentiment analysis.
- RoBERTa(Robustly Optimized BERT Pretraining Approach). It is a machine learning approach for NLP based on Transformers.
- The transformer model accounts for the words and also the context related to other words during polarity analysis
- The words in a comment or a sentence are processed sequentially and faster. There is no dependency on the previous text to identify the meaning or score of the next.
- Polarity scores of the YouTube comments would determine content consumption by the viewers and content evaluation.

Approach

- Log in to the Google Cloud console, create a new project and generate a YouTube API key which would enable the use of YouTube API to extract comments.
- Select the video for which the comments need to be analyzed. A video id enables information extraction. Make sure that the feedback section is not disabled.

Figure: A unique video Id enables information retrieval from YouTube API

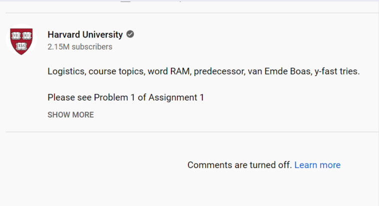


Figure: Make sure comments are not disabled

Comment Extraction

- Sentiment analysis of comments from this YouTube Video.The response consists of many other details.Extracted comments and replies from the API response.

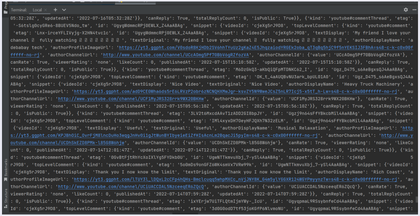


Figure: Response from YouTube API

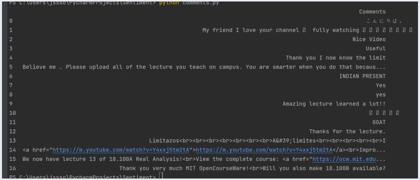
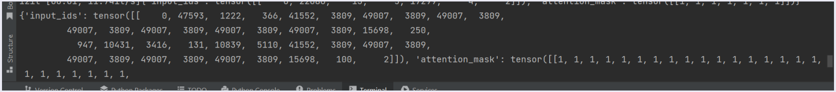


Figure: Extracted Comments for the Video

Using RoBERTa for Sentiment Analysis

- The Robustly Optimized BERT Pretraining Approach (RoBERTa) is an efficient deep learning transformer-based model that uses a GPT-2 tokenizer and has a vast vocabulary.
- The context of words in the comments are better learned, in both directions simultaneously.
- A pre-trained Twitter-roBERTa-base for Sentiment Analysis model provided by Hugging face is used to analyze the comments.
- Relevant libraries and dependencies are imported.
- The YouTube comments are encoded using the tokenizer pre-trained on the model.

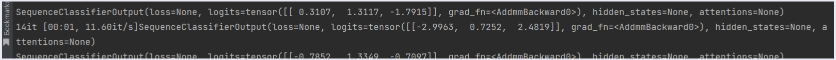
[6] Twitter-roBERTa-base for Sentiment Analysis



```
{'input_ids': tensor([[    0, 47593,  1222,   366, 41552,  3809, 49007,  3809, 49007,  3809,
         49007,  3809, 49007,  3809, 49007,  3809, 49007,  3809, 15698,   250,
           947, 10431,  3416,   131, 10839,  5110, 41552,  3809, 49007,  3809,
         49007,  3809, 49007,  3809, 49007,  3809, 15698,   100,     2]]), 'attention_mask': tensor([[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1, 1, 1, 1, 1, 1, 1,
```

Figure: Text encoding into 1's and 0's embeddings for the model to understand

- The model which is pre-trained from the Twitter-roBERTa-base for Sentiment Analysis model runs on the encoded text and returns tensors.



```
SequenceClassifierOutput(loss=None, logits=tensor([[ 0.3107,  1.3117, -1.7915]], grad_fn=<AddmmBackward0>), hidden_states=None, attentions=None)
14it [00:01, 11.60it/s]SequenceClassifierOutput(loss=None, logits=tensor([[-2.9963,  0.7252,  2.4819]], grad_fn=<AddmmBackward0>), hidden_states=None, a
ttentions=None)
```

Figure: Tensors returned after the model is run on encoded text

- The PyTorch tensors are converted to numpy arrays and softmaxed. We obtain negative, neutral and positive polarities of the texts.

[6] Twitter-roBERTa-base for Sentiment Analysis

Results

- Sentiment scores are determined from the tensors after processing.

```
100%|
                                        Comment  negative   neutral  positive senti_score
0                                        こんにちは。  0.352736  0.615987  0.031277           2
1    My friend I love your channel □  fully watching...  0.001132  0.006338   0.99253           3
2                                    Nice Video  0.006906  0.074745  0.918348           3
3                                        Useful  0.030462  0.433792  0.535747           3
4                      Thank you I now know the limit  0.020273  0.234156  0.745571           3
5    Believe me . Please upload all of the lecture ...  0.034624  0.262914  0.702461           3
6                                 INDIAN PRESENT  0.138215   0.75428  0.107504           2
7                                           Yes  0.125711  0.473165  0.401124           2
8                                           yes  0.138576  0.441277  0.420147           2
9                    Amazing lecture learned a lot!!  0.002418  0.025848  0.971734           3
10                                        □ □ □ □ □  0.012192  0.307302  0.680506           3
11                                         GOAT  0.561522  0.339878    0.0986           1
12                           Thanks for the lecture.  0.002567  0.045709  0.951724           3
13   Limitazos<br><br><br><br><br><br><br>A&#39;lim...  0.260197  0.708009  0.031794           2
14   <a href="https://m.youtube.com/watch?v=Y4xxj5t...   0.00355  0.146686  0.849764           3
15   We now have lecture 13 of 18.100A Real Analysi...  0.096057  0.800352  0.103591           2
16   Thank you very much MIT OpenCourseWare!<br>Wil...  0.000784    0.0218  0.977416           3
```

Figure: An overview of the comment, polarities and sentiment score

Results

- We visualize the sentiments of the YouTube comments from a specific MIT OpenCourseWare video. A similar pipeline can be executed to envision the course feedback and opinions of the viewers on other educational videos available on YouTube.

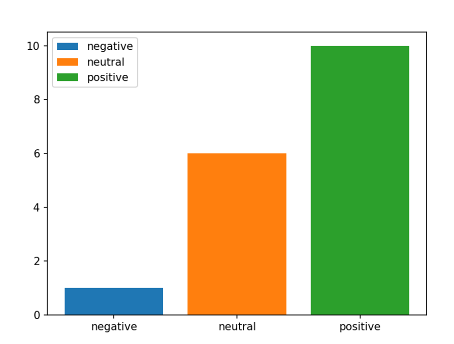


Figure: Depiction of feedback analysis of a YouTube video by MIT OpenCourseWare on Limits of Function

Limitations

- The self-attention with every other token in the input means that the processing will be in the order of O(N2) (which means it will be costly to apply transformers on long sequences).
- Transformers are extremely demanding on memory. There is a high probability that we will run out of memory or reach our run-time limit while training larger models or for longer epochs.
- The State-of-Art Transformer models can deal with fixed length text strings.
- The text is split into segments before being fed to the system as input causing context fragmentation.
- Sentiments like poetic phrases, idioms, tone detection are not analyzed due to context fragmentation.

Future Scope

- Comment classification and sentiment analysis using multi-class deep learning transformers models.
- This would help in determining if the comments and discussions foster peer learning and collaboration or express a reaction towards the content.
- Extending the emotion analysis to multiple languages and determining the precision.
- Applying the NLP pipeline to resource-poor languages.
- Named Entity Recognition with sentiment analysis to co-relate the context of the video description and the content feedback.

Summary

- Social Media has impacted education by making it available for all cost-free and at user convenience.
- YouTube promotes self-directed learning by catering to inspiration, pursuits, and authority.
- User involvement and content feedback are analyzed through the emotion analysis of the user comments.
- Most educational videos receive positive feedback from users. Few discussions might be argumentative but contribute to analyzing the topic from different perspectives.

Thank you for your attention.