

YouTube Transcript Summarizer To Summarize the content of YouTube.

Sourav Biswas¹, Atul Kumar Patel²

^{1,2}Student, Dept. of computer science and engineering, lovely professional university, Punjab, India

Abstract -) Watching long YouTube videos is very time-consuming and boring. Nowadays YouTube is an essential aspect of providing news and information. It is also considered a second teacher to the students, educational videos are the most viewed videos on YouTube today. In this project, we have tried to provide a quick, precise, and informative summary of a video. Many techniques are already discovered but they only provide test summarization. We have tried to get the summary of a video basically a YouTube video. For this project, we have used a hugging face transformer to summarize the content of a YouTube video along with that we have used python API to get the subtitle of a given video. After that our model will perform text summarization on it and display the summary to the user so that people can save their precious time reading the summary.

Key Words: Hugging Face Transformer, NLP, Summarization, BERT, DistilBERT

1. INTRODUCTION

Nowadays, information can be gained by various platforms and media it could be social media platforms, radio channels or new channels, etc. very huge amount of information is consumed by various users across the world. As nowadays a lot of content is consumed using YouTube. According to Buffer Library (blog), YouTube is the world's second most used social media site having 2.2 billion Monthly Active Users and Facebook is in first place having 2.8 billion Monthly Active Users. On YouTube, it is very easy to watch videos whether it is for educational purposes or entertainment purposes the user needs to find the relevant content of any genre the user wants to consume. After a successful search of the content on YouTube, the user wants to view it. A high number of users get a lot of problems while watching the content which such as network problem which restricts the user to watch the content in very low quality which leads the content to blurry and loading of the video might be sometimes frustrating also there are also a lot of clickbait videos created for earning purpose only, having lack of information in it. There are numerous situations where people click on the wrong videos because of their fancy thumbnail and titles wasting all their time gaining wrong or useless information. Students mostly these days prefer to clear all the topics which they are unaware of and also want to prepare for exams in very less time for that they fast forward the videos and try to understand the concept in very less time because of this they usually get confused in topics which they were watching. Sometimes working professionals want to know what work or terms were discussed in the meeting which was highly time-consuming in very less time they can use the idea of using summarized video text in it, so working professionals can easily save their time very effectively. Talking about text summarization, first, a summary is a brief statement or restatement of the main point's important information of the text in a shorter form [1]. The advantage of text summarization can be a decrease in time for the users to get the information precisely.

1.1 SUMMARIZATION TECHNIQUES

1.1.1 Extractive summarization

In this summarization technique, the summary is generated by collecting the most important sentences or sections of the given text also considering the most important phrases while generating the summary no new text is created, only existing sentences or texts are considered as a part of the generation of summary in the whole process [2].

1.1.2 Abstractive summarization

In this summarization technique, the main idea for using abstractive summarization is that it does not takes whole phrases or sentences from the text and concatenate them. It creates its own new sentences by selecting important points from the text, abstractive summarization technique is to generate the whole text into a meaningful summary [3].

1.2 Hugging Face Transformer

Hugging face transformer provides numerous variety of tasks such as answering the table question, classification of texts, summarization, etc. HUGGING FACE is the company that developed hugging face transformers. We have used this because it is very easy to use and convenient as well. It has many pre-trained algorithms and models and it reduces the computational cost and time [4].

Hugging Face electrical device uses the theoretic summarization approach wherever the model develops new sentences during a new kind, specifically like folks do, and produces a full distinct text that's shorter than the first.

By default hugging, face transformers use the DistilBERT technique for text summarization. DistilBERT is nothing more than a smaller version of the BERT technique developed by hugging face. Distilling BERT base is used to train DistilBERT [5].

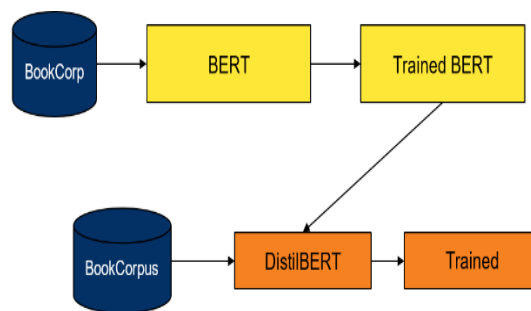


Fig 1: Hugging face using pertained DistilBERT Model from BERT

Hugging Face is used for solving NLP problems. The package provides pre-trained models that can be used for numerous NLP tasks. Transformers also supports over 100 languages. It also provides the ability to fine-tune the models with your data. Hugging Face use the Transformer library summarization pipeline to infer from the existing summarization model. If no model is provided then the pipeline will be initialized with the DistilBERT model.

Performing a report with a pre-trained electrical device may be a straightforward thanks-to-do report. An electrical device may be a machine learning design that mixes an associate encoder with a decoder and collectively learns them, permitting us to convert input sequences (e.g. phrases) into some intermediate format before we have a tendency to convert it into a human-understandable format. Transformers provide arthropod genus to simply transfer and train state-of-the-art pre-trained models. Victimization pre-trained models will scale back your reasonable prices, and carbon footprint, and prevent time from coaching a model from scratch. The models will be used across totally different modalities [6].

1. Text summarization
2. Speech recognition and
3. Image classification and audio classification

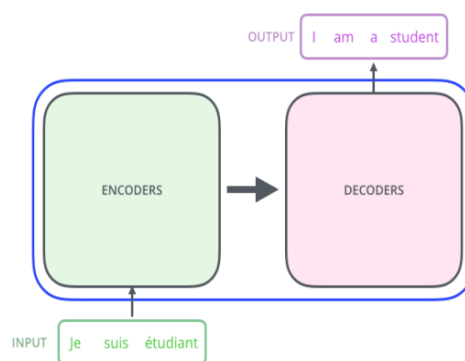


Fig 2: Architecture of Transformer

2. LITERATURE REVIEW

There are few summarization studies attempted earlier for summarization of YouTube videos. Some of them use extractive methods and some use abstractive methods, few methods are listed below for summarization.

2.1 Graph-theoretic approach

The graph-theoretic approach is an unsupervised technique. In the graph-theoretic approach, we rank the sentence and words depending on the graph. In this, we obtained the important sentences. In this approach, weighted graphs are used and sentences are represented as nodes. To connect two common nodes based on information edges are being used. After initializing the nodes sentence score is checked. Weight on the edges defines the similarity btw them [7]. The same sentence has some degree of correlation, the edges are inversely proportional to the distance between two nouns.

$$(\text{Edge-weight } (n1, n2) = 1 / (1 + \text{distance } (n1, n2)))$$

The sentence score is calculated on the summation of the relevance of all the nouns.

$$\text{Sentence score}(s) = \text{forall}(n), \text{sum}(\text{relevance}(n)) , \text{ where } n \text{ is noun.}$$

2.2 Machine Learning approach

There are many machine learning approaches for text summarization a few of them are 1. LSTM 2. BERT etc.

LSTM seq2seq:

In the LSTM model, two recurrent neural networks on either end are used first is the encoder, and the second is the decoder. The encoder takes input text as a sequence of numbers and the decoder translates the sequence into the target sequence. The advantage of using the LSTM model is that abstractive summarization is closest to human written summarization. The disadvantage of using LSTM is it is very to implement and it became very hard to process long strings.

BERT:

BERT model uses a transformer that utilizes an attention mechanism and feeds forward networks to perform text summarization. The attention mechanism computes the representation of a word by comparing it with other words in the sentence. As we have used a simple and better version of BERT. The advantage of using BERT is the BERT model is versatile and bidirectional which allows the representation of a word. The disadvantage of using BERT is summarization the quality may suffer sometimes due to the size disparity between the source and the summary [8].

2.3 Template based summarization:

Template-based summarization technique is also known as the user-friendly summarization technique in which the user has an authority that what should be present in the summary. In another word, we can say users create templates that contain pos tags like nouns, verbs, etc. it also contains a sequence of the sentence in which the user wants to arrange them. To generate a summary users may require many templates like that which contain pos tags. Names and dates can be added if the user wants. After this preparation, the template is said to be completed. The template-based summary contains all the required things that the user wants while generating the summary [9].

2.4 Semantic graph-based method:

The semantic graph-based method is used to show similarities between document sentences. This method used some properties of the document such as nouns, pronouns, synonyms, etc. For accurate text representation. This method has three parts first is implementing RSG second creating a linguistics graph using heuristic rules. The third and last part is to produce the summary based on RSG. Basically, RSG represents the noun and verbs in the document and edges represent the relation between them [10].

Abstraction summarization is more difficult and still much research is going on it. That is why we choose an abstractive method of summarization to make abstractive summarization easy and our main aim of this project is to provide the simplest technique of abstractive summarization so that it can help society for acquiring knowledge [11].

These are some common techniques of summarization all these methods have some advantages and disadvantages also. The common disadvantage in all the prior methods is they need to consider so many parameters for summarization and it is very difficult to understand also. Some of them required high types of machinery to perform and were very difficult to execute in the domain-specific region. For example, if we take a semantic graph-based model then the disadvantage of using this model is the drawback of this method is it can be used in a single type of document for creating an abstractive summary. In other words, we can say that its limitation is that it cannot be used in multiple documents. Instead of considering so many parameters, we should be focusing on simple parameters none of us don't have to take care about parameters and techniques we just have to add a chrome extension in which we have used the hugging face technique which will help us to get the summary of the video. The best part of hugging face transformers is that we don't need any data set, parameters, etc. Because it uses the DistilBERT technique for summarization it is the smaller and faster version of BERT. BERT is one of the most famous and best techniques for text summarization.

While working on this project we generated an easy text summarization Machine Learning model by using the Hugging Face pre-trained implementation of the BART architecture. After text summarization, we created a Flask backhand REST API to show the summary to the client. After that, we have created a chrome extension that will utilize backhand API for displaying the summary to the user. While requesting YouTube for the subtitled we Can get subtitled only for those videos whose subtitled are enabled otherwise we are not able to apply text summarization on it. I think someone can work on this. Subtitled is enable or disable we should get a summary of the text but this is not the case. In the future maybe we can get it other than this A live video summary is cannot be obtained may be in the future it will happen and also can work upon the accuracy of the summarized text as we have used abstractive summarization. Because it is a computer algorithm it cannot predict correct summarized text because when the algorithm summarized the text it erased some words from the text which are not important as it is a computer algorithm it cannot be 100 percent correct. The original text contains 4437 words after summarization we get a summary that contains 1772 words. So the accuracy of summarized text can be improved in this.

For summarization, we are requesting YouTube to provide the subtitled of the particular video we can generate an algorithm that can auto-generate the subtitled of a YouTube video without requesting YouTube for subtitled, and then we can use text summarization on it. it will be a faster method to get the summarized video on YouTube. The summarized video is in the form of text but it can be in the form of a video also. Providing results as a trim video of important points.

3. METHODOLOGY

At first, we have to use the Abstractive Summarization technique to summarize the text. In which we have used Hugging Face Transformer for summarization using a pipeline. Hugging Face Transformer uses the DistilBERT model for summarization as DistilBERT is nothing more than a smaller version of the BERT technique developed by the hugging Face company.

In this procedure firstly extracting the captions or say subtitles through python API for a particular ID of a YouTube video. In the very next step using the hugging face transformer for summarizing the obtained transcript of the YouTube video ID and for exposing the summarized version of the transcript is to create a Flask backend REST API.

So that users can get the required summarized content of a YouTube video.

For general usage also infused this algorithm into the chrome extension so that everyone can have the access to it. Also developed chrome extension will use back-end API to display a summarized version of the text to the user.

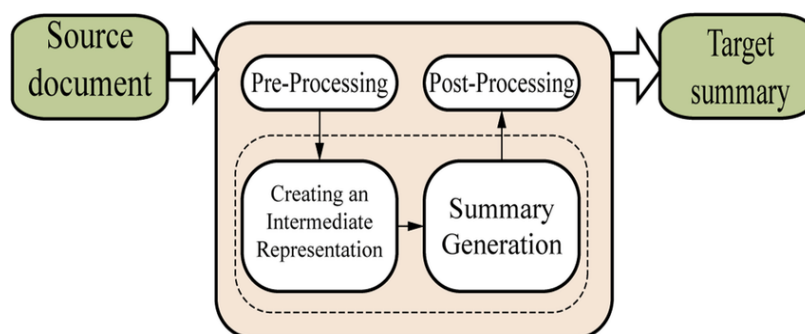
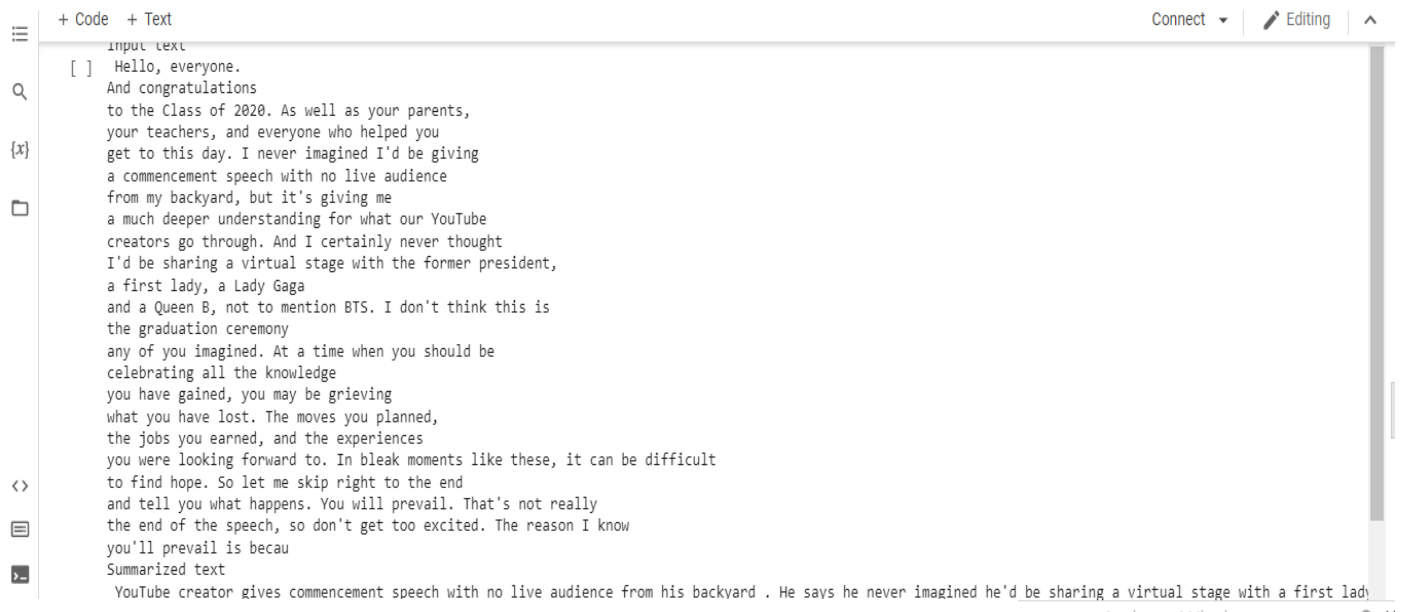


Fig 3: Architecture of Abstractive Text Summarization.

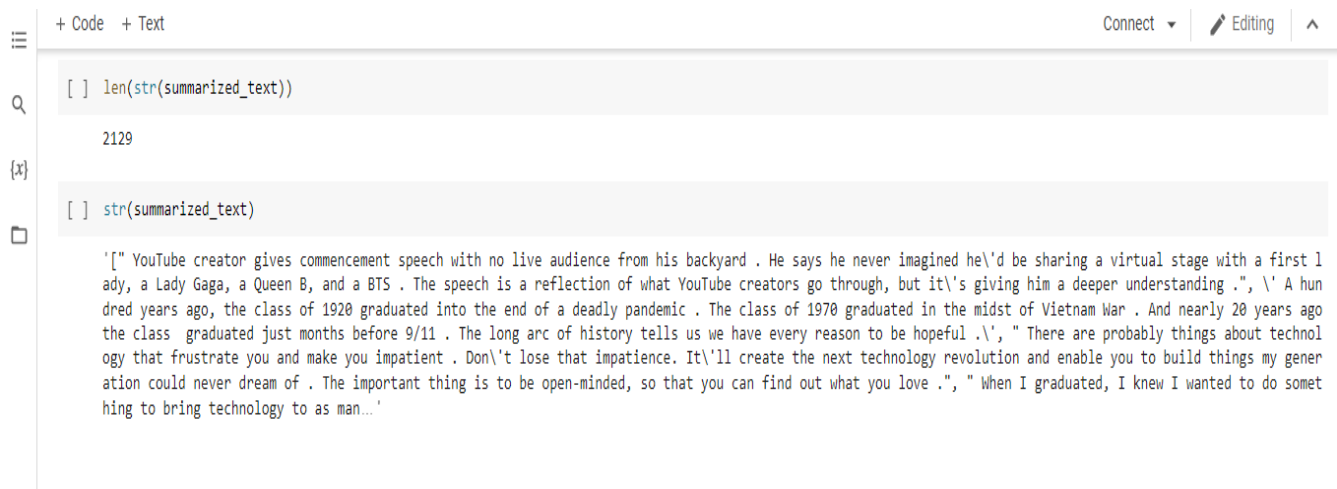
4. RESULT AND DISCUSSION

Evaluation of a summary is a very important factor in text summarization. Evaluation of summary can be used by one of the intrinsic measures or extrinsic measures. The most prominent way of evaluating the summary is the ROUGE evaluation which is a tool mostly known for its precision, also F-measure, and recall.



```
+ Code + Text
input text
[ ] Hello, everyone.
And congratulations
to the Class of 2020. As well as your parents,
your teachers, and everyone who helped you
get to this day. I never imagined I'd be giving
a commencement speech with no live audience
from my backyard, but it's giving me
a much deeper understanding for what our YouTube
creators go through. And I certainly never thought
I'd be sharing a virtual stage with the former president,
a first lady, a Lady Gaga
and a Queen B, not to mention BTS. I don't think this is
the graduation ceremony
any of you imagined. At a time when you should be
celebrating all the knowledge
you have gained, you may be grieving
what you have lost. The moves you planned,
the jobs you earned, and the experiences
you were looking forward to. In bleak moments like these, it can be difficult
to find hope. So let me skip right to the end
and tell you what happens. You will prevail. That's not really
the end of the speech, so don't get too excited. The reason I know
you'll prevail is becau
Summarized text
YouTube creator gives commencement speech with no live audience from his backyard . He says he never imagined he'd be sharing a virtual stage with a first lady
```

Fig 3: Required text before summarization contain 4437 words



```
+ Code + Text
[ ] len(str(summarized_text))
2129
[ ] str(summarized_text)
'[" YouTube creator gives commencement speech with no live audience from his backyard . He says he never imagined he'd be sharing a virtual stage with a first lady, a Lady Gaga, a Queen B, and a BTS . The speech is a reflection of what YouTube creators go through, but it's giving him a deeper understanding .', ' A hundred years ago, the class of 1920 graduated into the end of a deadly pandemic . The class of 1970 graduated in the midst of Vietnam War . And nearly 20 years ago the class graduated just months before 9/11 . The long arc of history tells us we have every reason to be hopeful .', " There are probably things about technology that frustrate you and make you impatient . Don't lose that impatience. It'll create the next technology revolution and enable you to build things my generation could never dream of . The important thing is to be open-minded, so that you can find out what you love .", " When I graduated, I knew I wanted to do something to bring technology to as man...'
```

Fig 4: Required text after summarization contains 1772 words

Our proposed method worked pretty well it summarized YouTube videos with ease. The original text contains 4437 words and after summarization, it only contains 1772 words. The required summary is very specific and very informative. Our model trimmed 2665 non-important words and gave us a very concise and short summary. Hope our project will satisfy user needs and save precious time.

5. Conclusion and feature Developments

In this project, we provided a systematic solution for the text summarization of a YouTube video. We have discovered the simplest way to summarize a text of a YouTube video. Concerning the summarization performance, we have used a hugging face transformer for text summarization as it is the best way to do so because it has a pre-trained model and is used for various NLP tasks. For less complexity in obtaining a summary of a YouTube video, we created a chrome extension so that adding the extension to their browser simply user can click on the summarize button for a particular YouTube video. Hope our project

solves the problem of the users as they can save their time, and effort for the lengthy YouTube videos and also provide only the important information on the topic which users can easily understand.

Feature of this project can be, providing the summary of a video in which there are no subtitles and another aspect can be, as we can see there are numerous videos available in different languages generating a summary of a non-English video and providing the summary in the English language so that there is very less language barrier.

6. REFERENCES

1. Dictionary.com. (n.d.). *Summary definition & meaning*. Dictionary.com. Retrieved March 15, 2022, from <https://www.dictionary.com/browse/summary>
2. Mehta, P., & Majumder, P. (2019). Introduction. *From Extractive to Abstractive Summarization: A Journey*, 1–9. https://doi.org/10.1007/978-981-13-8934-4_1
3. <https://www.sciencedirect.com/topics/computer-science/text-summarization>
4. Desarda, A. (2020, April 24). *Working with hugging face Transformers and TF 2.0*. Medium. Retrieved March 5, 2022, from <https://towardsdatascience.com/working-with-hugging-face-transformers-and-tf-2-0-89bf35e3555a>
5. Joshi, P. (2022, April 5). *What are transformers in machine learning*. What Are Transformers In Machine Learning. Retrieved March 15, 2022, from <https://prateekjoshi.substack.com/p/what-are-transformers-in-machine>
6. Graph-theoretic concepts. (n.d.). *A Graph-Theoretic Approach to Enterprise Network Dynamics*, 31–40. https://doi.org/10.1007/978-0-8176-4519-9_2
7. N, A., R, A., T, A., & S, S. B. (2021). An exhaustive survey on automatic text summarization using Machine Learning Approches. *Webology*, 18(05), 1184–1190. <https://doi.org/10.14704/web/v18si05/web18299>
8. Y., A., & M., A. (2018). Template based medical reports summarization. *International Journal of Computer Applications*, 179(17), 47–55. <https://doi.org/10.5120/ijca2018916301>
9. Ullah, S., & Islam, A. B. (2019). A framework for extractive text summarization using semantic graph based approach. *Proceedings of the 6th International Conference on Networking, Systems and Security - NSysS '19*. <https://doi.org/10.1145/3362966.3362971>
10. R, S. S. (2021). Abstractive summarization using Categorical Graph Network. *Revista Gestão Inovação e Tecnologias*, 11(4), 1997–2007. <https://doi.org/10.47059/revistageintec.v11i4.2249>
11. Write With Transformer. (n.d.). Retrieved March 15, 2022, from <https://transformer.huggingface.co/>