Automated Generation of Multiple-Choice Questions from Video Lectures Using Python

Anupriya Kathpalia, Anushka Chattaraj, Chirag Bulchandani, Niharika Deokar

1. Abstract

In online education, video lectures are widely used to deliver course content to students. However, these lectures often lack interactivity and may not provide students with opportunities for assessment or active learning. In this project, we present an approach to automatically generate MCQs from video lectures using Pytube, Python, OpenAI's Whisper API, and Spacy. The tool extracts concepts and ideas from the lecture's transcript and uses NLP and AI techniques to generate high-quality MCQs with distractors. These MCQs are inserted at relevant points in the video, providing students with opportunities for active learning and assessment. Our approach has the potential to enhance the online learning experience by providing engaging and interactive content for learners.

2. Introduction

Online education has increased significantly since COVID-19 due to the closure of traditional educational institutions and the need for remote learning. Online education has provided a safe and effective way for students to continue their studies and maintain their academic progress. The use of technology in education has significantly increased recently, particularly in online learning environments. The use of video lectures to teach pupils content is one area that has grown quickly. While video lectures can be a useful tool for spreading knowledge, they frequently don't give students the chance for active learning or assessment.To address this gap, researchers and educators have started to explore the use of multiple-choice questions (MCQs) as a way to enhance the effectiveness of video lectures. MCQs can provide students with immediate feedback on their understanding of the material and encourage them to engage with the content actively.

2.1 Approach

In this paper, we propose a novel approach to creating MCQs for video lectures using Python. Python is a versatile and widely used programming language that offers a range of libraries and tools for data analysis and machine learning. We believe that the use of Python can facilitate the creation of high-quality MCQs that are tailored to the content of the video lectures and the learning objectives of the course.[2]

2.2 Objective

This research paper's major goal is to show that our suggested strategy is workable and efficient. In particular, we will outline the procedure for producing multiple-choice questions (MCQs) using Python, assess the calibre of the MCQs produced by our method, and go through the potential advantages and drawbacks of applying this method in online learning settings.[5]
Overall, we believe that our proposed approach has the potential to enhance the effectiveness of video lectures and improve student learning outcomes in online courses
.

3. Literature Review

	Title of paper	Year	Data	Method used in paper	Results achieved	Gaps
[1]	An Automated Multiple-Choice Question Generation using Natural Language Processing Techniques	2021	Dataset consisting of different study materials, number of words, keywords, etc. present.	Natural Language Processing Techniques, noise removal, word normalisation.	The outcome is presented in a user-friendly interface for teachers' easy accessibility using Django, a web framework for Python.	This paper is not concerned with the issue of whether multiple-choice tests are better assessment methodology than other types of tests.
[2]	SPEECH TO TEXT CONVERSION AND SENTIMENT ANALYSIS ON SPEAKER SPECIFIC DATA	2021	Inputs taken from various speakers.	NLTK, NLP methods are used like lemmatization and stemming. Sentiment Analysis is also performed.	Though the model works fine and accurately in executing text conversion on the input speech by a user, it still has some drawbacks like it cannot understand if two or more speakers speak at the same time.	Itl has some drawbacks like the system can smoothly handle only one audio input at a time and cannot understand if two or more speakers speak at the same time.
[3]	Speech to text conversion and summarization for effective understanding and documentation	2019	The speech from the source is recorded using a microphone and the feature is extracted in text format using Google Application Programming Interface (API).	The proposed method summarises the extracted text according to the rank of the sentences which can be determined through the frequency of occurrence of words.	Tmodel can be used wherever there is a requirement of summarising lengthy lectures into precise documents as the automated system will convert the speech to text and also summarise the content.	All punctuation marks in the recognized speech can help in improving the text summarization performance.
[4]	A REVIEW ON METHODS FOR SPEECH-TO-TEXT AND TEXT-TO-SPEECH CONVERSION	2020	Dataset containing text messages and video lectures.	The model for STT conversion is carried out by HMM and Neural Network as it gives the highest accuracy for STT.	The most suitable technique for STT conversion is by deploying a combination of Hidden Markov Model with Deep Neural Network, which can be implemented in Python using Google's Speech Recognition API module.	This system can be improved by considering the punctuation marks while converting speech to text. It can also be implemented in different languages like English, Hindi, etc.
[5]	Natural Language Processing for Text and Speech Processing: A Review Paper	2020	No data input as such.	NLP techniques have been discussed in this paper.	The natural language processing techniques and most useful NLP equipment kits dealing with badges and Chinese word selection have been reviewed in this paper.	Some complex concepts could have been discussed.
[6]	Natural language processing in speech understanding systems.	2019	Audio data of Telephone exchange between two people.	Template matching using dynamic programming.	Computer simulations people tend t.o speak more grammatically than when addressing a fellow human.	A multi-disciplinary approach might provide better results. .
[7]	Natural language processing: state of the art, current trends and challenge	2022	No dataset as such.	Various concepts used in NLP have been discussed along with its applications.	One objective of this paper focuses on the history, applications, and recent developments in the field of NLP. It also discusses datasets, approaches and evaluation metrics used in NLP.	There is work on regional languages, which could have been done.
[8]	Speech Recognition	2020	No dataset as such.	Acoustic-Phonetic, and Pattern Recognition along with some technologies used in speech recognition.	The model specified in the paper hopes to create this software that can instantly translate two languages with at least 90% accuracy.	The software needs to be trained to understand and recognize the voice of the dictator – transcribing voice data from more than one person is difficult.

4. Proposed Methodology

Video download: Use Pytube to download the video lecture.
Audio transcription: Use the whisper module provided by OpenAI
Text pre-processing: Use Spacy and NLTK to pre-process the text by removing stop words, punctuations, and performing lemmatization and tokenization.
Concept extraction: Use Spacy's Named Entity Recognition (NER) to extract concepts from the pre-processed text.
Question generation: Using OpenAI GPT to generate questions from the extracted concepts. The generated questions can be formatted as multiple-choice questions with distractors.
Answer key generation: Use Spacy's dependency parsing and NLTK's part-of-speech tagging to identify the correct answer from the lecture's text.
MCQ insertion: Use Pytube to insert the generated MCQs with answer options at appropriate points in the video lecture.

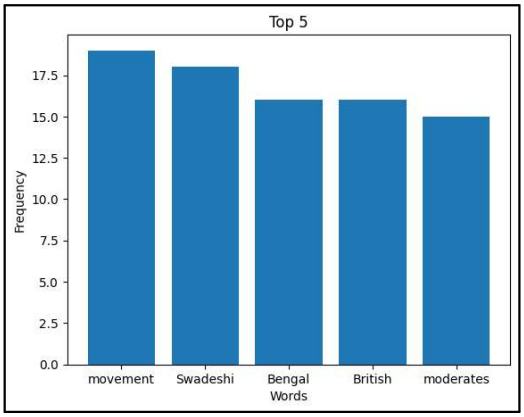

Fig 5.1 The 5 words with the largest frequency are shown here using a bar chart.

5. Dataset Description

Our dataset comprises video lectures uploaded on YouTube by various institutes and organisations. We utilised the Pytube module to download these videos. Subsequently, we generated lecture transcripts using the whisper module provided by OpenAI. These transcripts serve as the foundation for creating reports or summaries of the lectures. If a student is unable to attend a lecture, they can rely on these summaries for understanding the content. Additionally, we leverage the transcript data to generate multiple-choice questions (MCQs) that assess students' comprehension and offer insights to lecturers regarding which topics may require further clarification.

6. Block Diagrams

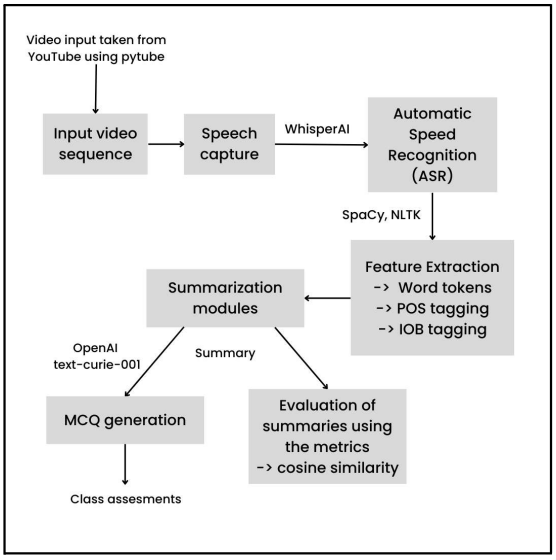

Fig 6.1 Block diagram shown here depicts the current methodology used, various Python libraries have been used like SpaVy, NLTK, etc.

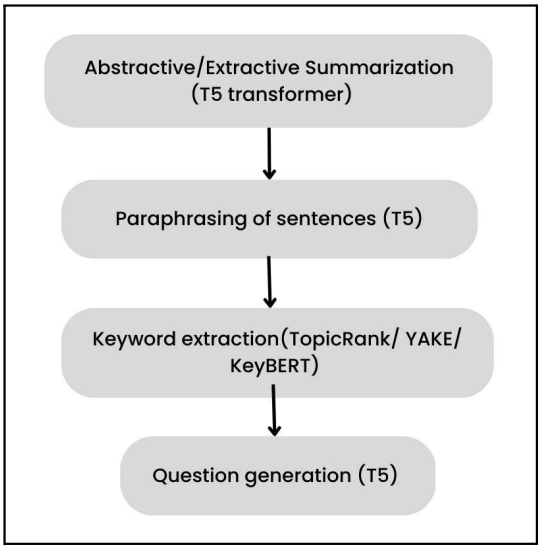

Fig 6.2 Block diagram depicting the flow after text summarization

7. Experimental Work

While abstractive methods may have lower recall, they often excel in generating more concise and coherent summaries by paraphrasing and restructuring the content. Improving recall in abstractive summarization is an ongoing research area, and techniques such as reinforcement learning, copy mechanisms, and attention mechanisms aim to address this challenge.[4]

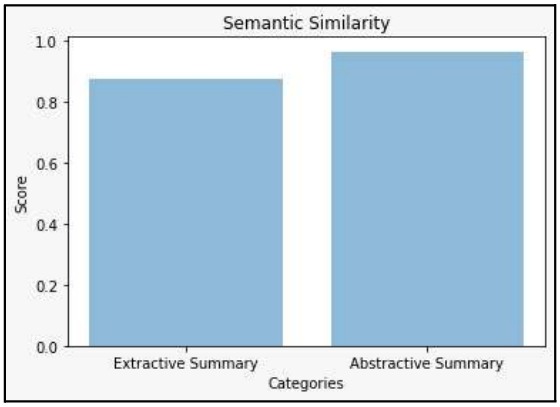

Fig 7.1 Bar graph comparison in scores of abstractive and extractive summarization

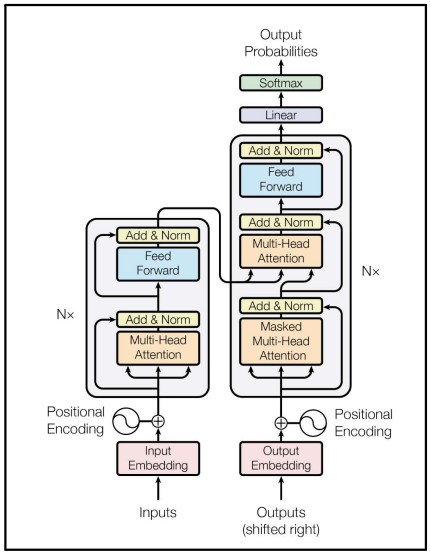

Fig 8.1 The Transformer model architecture

T5 Transformer is a state-of-the-art language model that can perform text summarization tasks. It has shown to be effective in generating high-quality abstractive summaries by using a combination of pre-training on large amounts of data and fine-tuning on specific summarization tasks.

8. Results and Discussion

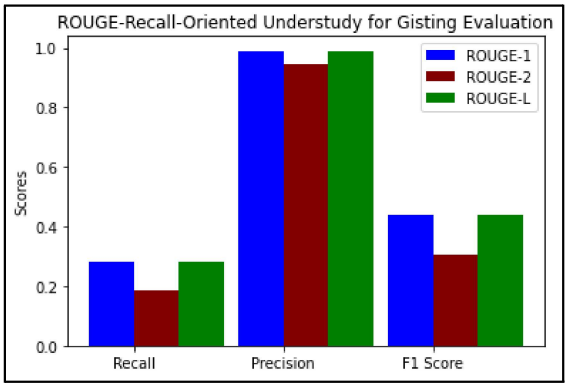

Fig 8.1 The bar graph depicts the rogue evaluation for different gisting evaluations.

ROUGE-1, ROUGE-2, and ROUGE-L are evaluation metrics commonly used in text summarization and natural language processing tasks to assess the quality of generated summaries. Here's a brief explanation of each metric:

1. ROUGE-1/ROUGE-2:

ROUGE-1/2 measures the overlap of unigrams (individual words)/two sequential words between the generated summary and the reference summary. It calculates precision, recall, and F1-score based on the count of matching unigrams. ROUGE-1 evaluates the ability of a summarization system to capture important words from the reference summary.

2. ROUGE-L:

ROUGE-L (Longest Common Subsequence) measures the longest common subsequence between the generated summary and the reference summary. It calculates precision, recall, and F1-score based on the length of the longest common subsequence and the total number of words in the reference summary.

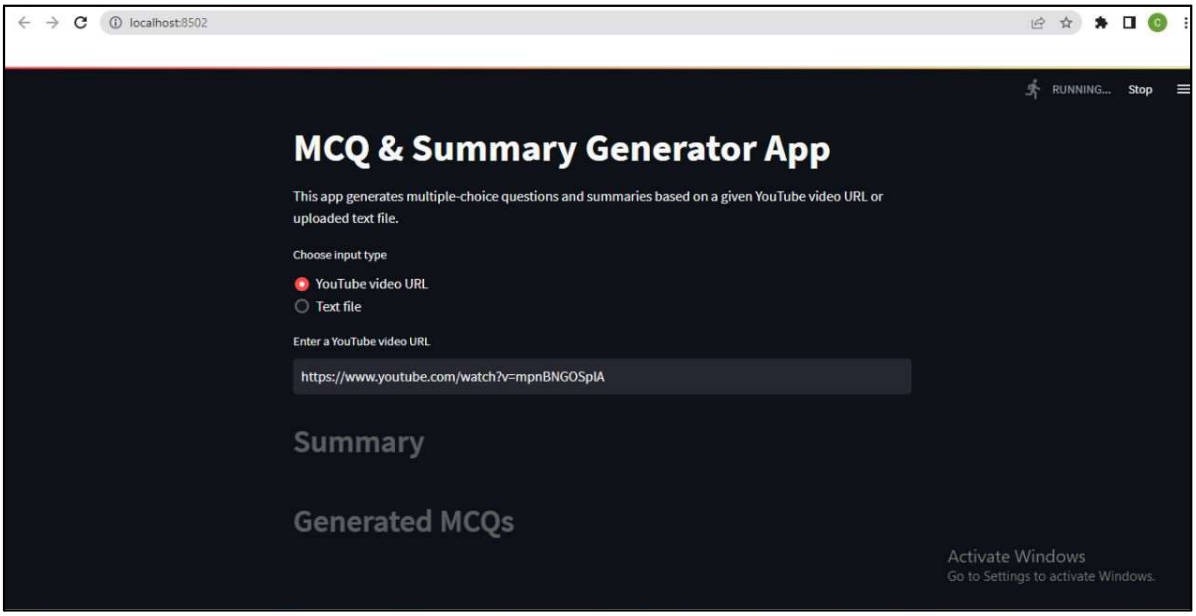

Fig 8.2 Screenshot for the GUI made on StreamLit

MCQs generated using method 1: (WordNet)

1) We must understand, friends, that this was the worst thing that Lord Kursin could have done with _____ .

a) India
b) Lebanon
c) Tajikistan
d) Tartary

More options: ['Asian country', 'Roman Empire', 'Qatar', 'Turkistan', 'Tibet', 'Kuwait', 'South Korea']

2) We are going to discuss the most unwanted thing especially for the _____ and that is partition of Bengal that took place in 1905 .

a) Indian
b) Indians
c) Arawak
d) Carib

More options: ['South American Indian']

MCQs generated using method 2: (OpenAI GPT)

1.Which of the following is true about the boycott of British manufactured goods during the Swadeshi movement?
A) It was a success
B) It was a failure
C) It was limited to specific items
D) It was not a movement

2.Which of the following is true about the Swadeshi movement?
A. The Swadeshi movement encouraged people to purchase only indigenous products.
B. The Swadeshi movement was a nonviolent protest movement.
C. The Swadeshi movement was a response to British economic policies.
D. The Swadeshi movement was motivated by patriotism.

9. Data Visualization

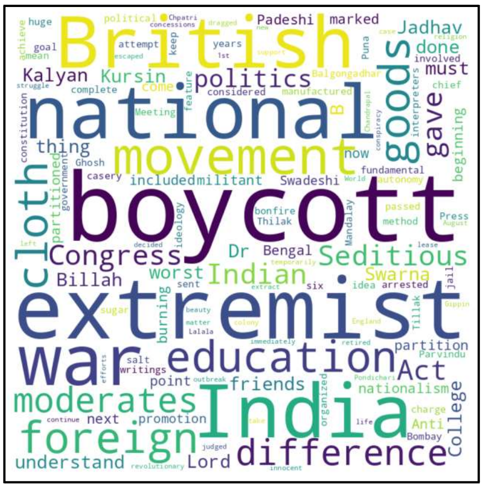

Fig 9.1 Word cloud depicting the most frequent words

10. Evaluation Matrix using Human Feedback

Video no.	URL	Domain	Video length	Total number of questions generated	Number of correct MCQs	Number of wrong MCQs generated
1	https://www.youtube.com/watch?v=tVy3dzLSMLg	Geography	04:42	23	15	8
2	https://www.youtube.com/watch?v=mpnBNGOSplA	IT	04:30	22	17	5
3	https://www.youtube.co	Sports	:8:36	57	46	11
4	ps://youtu.be/uJ60P3YjC	History	18:32	76	53	23
5	https://www.youtube.co	Language	07:46	51	40	11

11. References

[1] Nwafor, C.A. and Onyenwe, I.E., 2021. An automated multiple-choice question generation using natural language processing techniques. *arXiv preprint arXiv:2103.14757*.
[2]Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for
[3]Computational Linguistics (p. 271). Association for Computational Linguistics.
Magellan's, Rajesh Kumar M (2017 I2C2). Sentiment Analysis on Speaker Specific Speech Data
[4] Li, H., Einolghozati, A., Iyer, S., Paranjape, B., Mehdad, Y., Gupta, S., & Ghazvininejad, M. (2021). Ease: Extractive-abstractive summarization with explanations. *arXiv preprint arXiv:2105.06982*.
[5]Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal (2018). Speech to text and text to speech recognition systems-A Review, IOSR Journal of Computer Engineering (IOSR-JCE)
[6]D. Y. S. R. Ibrahim Patel, "SPEECH RECOGNITION USING HMM WITH MFCC- AN," 2010. 5. N. K. P. K. Bhupinder Singh, "Speech Recognition with Hidden Markov Model: A Review,"