2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 979-8-3503-7024-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCCNT61001.2024.10726107

AUTOMATIC QUESTION GENERATION FROM YOUTUBE LECTURES USING DEEP LEARNING

Himanshu Jasuja
Department of Computer Engineering
Delhi Technological University
New Delhi, India
himanshujasuja_co20b4_24@dtu.ac.in

Ujjwal Negi
Department of Computer Engineering Delhi Technological University
New Delhi, India
ujjwalnegi_co20b7_29@dtu.ac.in

Vibhav
Department of Computer Engineering Delhi Technological University
New Delhi, India
vibhav_co20b7_41@dtu.ac.in

Ms. Gull Kaur
Department of Computer Engineering
Delhi Technological University
New Delhi, India gullkaur@dtu.ac.in

Abstract **-** In this contemporary world full of information, online lecture videos are a big fountain of knowledge. Nevertheless, quizzes have to be developed based on these videos to make evaluation of knowledge acquisition much easier. The research describes a method for generating quizzes from online teaching videos that enhances self-learning through continuous assessment. Unlike existing approaches which are resource intensive and computationally demanding, we aim at providing a Video Question Generation model that is light weight and effective. We take advantage of state-of-the-art Natural Language Processing (NLP) technology to improve our model's flexibility and allow it to be fine-tuned using T5 transformers. Our system also generates various forms of "Wh" questions such as who, when, where, what, which, why and how as well as Multiple Choice Questions (MCQs). Through this study we hope to give teachers and students alike a tool that can facilitate knowledge assessment and create an active learning environment.

Keywords - Video Question Generation, T5 Transformers, Self-learning, NLP (Natural Language Processing), Distractors

1. INTRODUCTION

The COVID-19 epidemic has affected educational systems worldwide, as we all know. Closing down of physical classrooms affected almost 1.2 billion students yet the transition to online learning was abrupt. Several e-learning websites responded by offering free access to their educational databases. The result was a significant shift towards e-learning following the outbreak that looked like a self-repairing academic setback for many people. In the period of this pandemic, it became an obligatory mode rather than an alternative course.

This transformation has speeded up the adoption of atypical learning platforms and tools that facilitate students' educational journey. The rise of online courses and self-directed remote learning makes it necessary to have an evaluation method that is continually present for self assessment. This change has produced an overwhelming digitalization of educational information in which videos are now the most important sources of knowledge on different social media platforms as well as learning platform.

In this era of increased self-learning, online lecture videos have become indispensable tool for mastering new things. However, some platforms may provide their own lecture videos and assessment tools but it can still be hard to

find relevant assessment questions from external services such as YouTube videos. Traditional learning methods too require an infinite supply of fresh questions because computer systems are inherently complicated and this had been largely dependent on human interaction.

We aim to solve this problem in our research by providing a technique that enables students to generate questions from online instructional videos. Our objective is to harmonize video material and assessment through supporting self-learning with continuous evaluation. We present a simple and effective way of generating video queries using state of the art Natural Language Processing (NLP) technologies such as T5 transformers.

This model's main aim is therefore establishing an interactive, dynamic educational ecosystem that can allow both teachers and learners easily measure knowledge acquisition in online learning environments.

2. RELATED WORK

In this work, we introduce a two-stream Composition Attention Network [5, 6]. The Action Pooling Stream and Uniform Sampling Stream methods are used in our approach to extract visual features within the two-stream setup, which improves the quality of representation. In their survey on video captioning techniques, Khurana et al [7] Assess these with regards to evaluation measures and sets of data. They underscore the necessity of video question answering (VQA) models to generate questions about images and give answers

In order to tackle these issues, Guo et al. [8] put forward a new structure for Question Generation based on selection and attention mechanisms.
This framework, which is especially designed for Generating Single Turn Questions, focuses on processing dialog history efficiently and choosing the most relevant questions from candidate questions created in each iteration of dialog history. The framework also includes a Video Question Answer model that uses reinforcement learning to improve its ability to predict replies and answer quality.

Dhavaleshwar Rao CH et al. [1, 2] developed a process flow for automatic question generation, particularly Multiple Choice Questions (MCQs). To evaluate the caliber and applicability of autonomously generated multiple-choice questions (MCQs), their work compares several approaches to question creation and assessment.

In a similar setting, Patil et al. [3, 4] investigated attention modules by contrasting the outcomes of a basic encoder-decoder module. Their research focuses on assessing how effective attention is for the given job.

3. METHADOLOGY

The following section presents how questions are generated from internet instructional videos using T5 transformers and machine learning algorithms. We want to develop a lightweight, efficient video question generation model allowing for self-learning via ongoing evaluation.

3.1 REQUIRED SYSTEMS AND TOOLS

The Intel Core i5 Series processor is used, and 8 GB of installed RAM is a minimum requirement for the system. It runs on an operating system that is 64-bit. Python 3 and a Jupyter notebook are the tools utilized for implementation

3.2 SYSTEM DESIGN

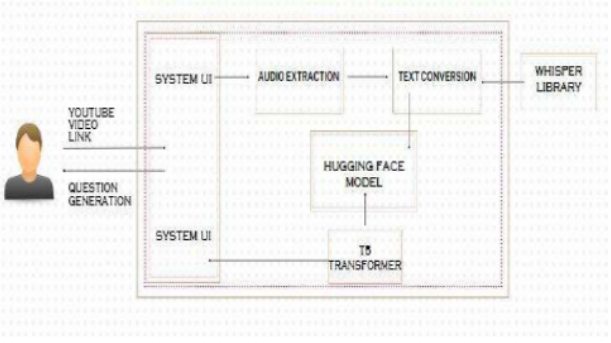

Figure-1 System Architecture

3.2.1 Video Transcription: The first thing to do is turn YouTube videos into text. By means of pytube library, we download the video from the given YouTube address. The downloaded

video will then be passed through whisper library—this has an integrated T5 transformer model for accurate transcription. It ensures that all spoken content in the video is transposed exactly as it is into a written form so that it can be scrutinized better.

3.2.2 Utilizing T5 Transformers for Summarization: Utilizing T5 Transformers for Summarization: After a movie transcript has been made, it can be summarized by using the T5 transformer model. It is a great model to use for summarizing long texts into short and informative summaries because it has been trained on large amounts of data. With the summary function, we can get a more focused version of what was in the video content that highlights key details.

3.2.3 Answer Span Extraction (Noun Phrases and Keywords): To create insightful queries, major keywords and noun-phrases are taken out from the summary text. There are two primary methods used to achieve this:

• **Noun Phrase Extraction:** Noun Phrase Extraction: In finding and selecting noun phrases from the summary text, Multipartite Rank algorithm is used in noun phrase extraction; it helps identify salient words and ideas necessary for question formation.

• **Keyword Extraction:** Although, to further extract keywords and relevant terms from the abridged information, we administer the T5 transformer model. In generating queries that are apt for the situation, this model identifies important words and phrases.

3.2.4 Generating Questions: Then, we use the transformer model T5 to formulate questions for the keywords and noun phrases that were retrieved. The get_question function has the ability to produce different types of questions like Multiple Choice Questions (MCQs) and "Wh" questions (who, when, where, what, which, why and how).

3.2.5 Distractors for MCQs: To improve multiple-choice questions, one way is to add a distractor for each correct answer. Two of the ways are wordnet based and sense-2-vec.

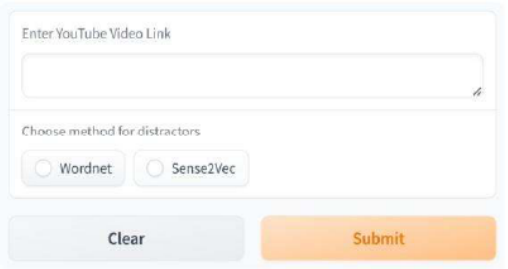

Figure-2 Methods of generating distractors

• **WordNet-Based Distractors**: This method finds correct answers by using words or concepts in the WordNet lexical database. In this manner, our distractions are far enough from the right answers yet still semantically similar.

• **Sense-2-Vec Based Distractors:** The Sense2Vec model allows us to find words and phrases that are similar or closest in meaning to the correct response. By using terms with synonymous connotations we can generate plausible distractions within an inquiry context.

3.2.6 User Interaction and Question Visualization: Finally, Gradio was utilized so that the generated questions could be brought into an interactive user interface. Created questions are accessed by users from a YouTube video transcription which is processed into question-answer form. Distractor generation offers flexibility and customization since users can choose between WordNet and Sense2Vec techniques.

3.2.7 Assessment and Confirmation: Assessments are carried out followed by validations of our methodological approach. This include:

• **Question Quality Assessment:** Experts examine the generated questions to determine their quality, relevance and level of difficulty.

• **User Feedback*:*** User testing sessions as well as feedbacks are also held for more insight on the effectiveness and usability of generated questions for self-learning.

4. RESULTS AND DECLARATION

In Fig 3, we can see the user interface of the final product. Here we will have to just paste the youtube's lecture link and enter submit.

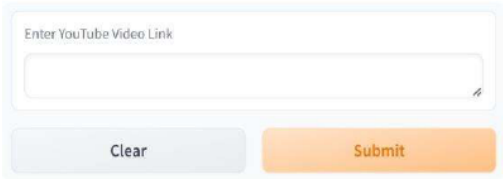

Figure-3 User Interface

Figure-4 Questions Generated

The assessment measure used for the extracted passage is that of Automatic Speech Recognition (ASR). The metrics used to assess this are: Word Error Rate (WER), Match Error Rate (MER) and Word Information Lost (WIL).

First, let's clarify some definitions.
1. Automated Speech Recognition (ASR): ASR is a technology that computers use to change spoken language into written text. It employs algorithms and models to analyze sound signals and identify speech as words.

2. Word Error Rate (WER): WER is the most common way of measuring how accurate an ASR system is. It reveals what portion of recognized output contains different words compared with reference transcript, including substitutions, deletions, insertions etc.

3. Match Error Rate (MER): Another measurement used in evaluating ASR systems which indicates the number of words in reference transcript that were not correctly recognized by an output.

4. Word Information Loss (WIL): This measures the amount of information lost during transcription relative to the original speech content itself. It tells us how many significant things are left out or changed when converting from voice to text.

The Word Information Lost (WIL) is calculated using equation 1

$$WIL = 1 - H^2 / (H+S+D)(H+S+I)$$

Let's see, calculation of MER using equation 2.

$$MER = (S+D+N) / (N = H+S+D+I) = 1 - H/N$$

And at last, calculation of WER using equation 3.

$$WER = (S+D+I) / N_1 = (S+D+I) / (H+S+D)$$

Where, D represents deletions and S represents replacements.
N denotes terms in the reference, C denotes correct words, and I stands for inserts. H represents the total number of victories.

Preprocessing involved reducing the retrieved passage to a list of ListOfWords, removing white spaces, and changing it to lowercase. Speech recognition has a minimal error rate of about 19%. Following pre-processing, the retrieved passage has an error rate of 0.29. The evaluation's findings are displayed in Table_1.

Table 1: Metric-based assessment

Metrics	MER	WIL	WER
Sample 1	0.4530	0.5430	0.2902
Sample 2	0.4789	0.5729	0.3122

Sample 3	0.5182	0.6021	0.3321
Sample 4	0.4647	0.5830	0.3245
Sample 5	0.5039	0.5572	0.3192
Sample 6	0.4939	0.5329	0.2930

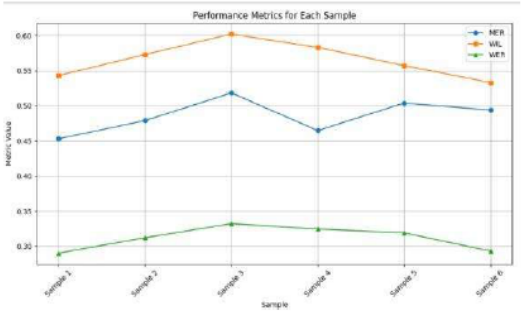

Figure-5 Graphical Representation for Table-1

It is impossible to evaluate the system's completion of a task with a specific assessment tool because doing so is left to the subjective judgement of people. Another method for evaluating model generations was human annotation. Volunteers' ratings were based on three criteria that included; Naturalness; Difficulty; and Contextual significance. The system has been reviewed by humans. Relevant, natural, and difficult questions were then marked according to intuitions by volunteers who assessed them in relation to their relevance, difficulty, and nature respectively. To gather data for this survey, forms were distributed among students, faculty members and inexperienced users (graduates); answers were recorded and analyzed later on. Questions asked are: Is the one on the right more or less natural than the one shown previously? What can you tell about these two questions? Did these questions get more difficult as you watched more videos? These results are summarized in table 2 below.

Table 2: Human-based assessment

Metrics	Natural-ness	Difficulty	Relevance
Sample 1	3.51	2.38	4.29
Sample 2	3.87	3.02	4.35
Sample 3	4.02	3.39	3.98
Sample 4	3.78	3.20	4.56
Sample 5	4.20	2.50	2.75
Sample 6	3.45	3.90	4.05

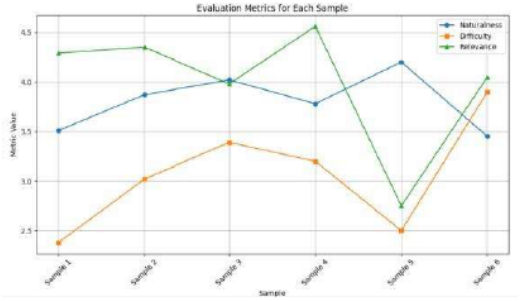

Figure-6 Graphical Representation for Table-2

5. CONCLUSION

We have introduced a new way of questions generation from online instruction videos. We generate queries, extract keywords, summarize and transcribe the videos as part of our method. T5 transformers enabled us to generate a variety of question types from transcribed and summarized video content including 'Wh' questions and Multiple Choice Questions (MCQs). To improve the quality and relevance of MCQs, we used WordNet and Sense2Vec for making distractions. Our dynamic Gradio user interface helps teachers, students enhance learning experiences and self-assessment with useful tools. There is an option for users to provide a YouTube movie link in order to create questions with distractions. This makes learning environment dynamic and interesting. In conclusion, our research enhances educational technology by providing a helpful tool for knowledge assessment as well as independent learning purposes. Future works may focus on expanding the dataset, enhancing question generation algorithms; conducting user studies for further validation etc.

REFERENCES

[1] D. R. Ch and S. K. Saha, "Automatic multiple choice question generation from text: A survey," *IEEE Transactions on Learning Technologies,* vol. 13, no. 1, pp. 14-25, 2018.

[2] L. Nie, M. Wang, Y. Gao, Z.-J. Zha, and T.-S. Chua, "Beyond text QA: multimedia answer generation by harvesting web information," *IEEE Transactions on Multimedia,* vol. 15, no. 2, pp. 426-441, 2012.

[3] C. Patil and A. Kulkarni, "Attention-based visual question generation," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2021: IEEE, pp. 82-86.

[4] C. Clark, K. Lee, M.-W. Chang, T. Kwiatkowski,

M. Collins, and K. Toutanova, "BoolQ: Exploring the surprising difficulty of natural yes/no questions," *arXiv preprint arXiv:1905.10044,* 2019.

[5] T. Yu, J. Yu, Z. Yu, and D. Tao, "Compositional attention networks with two-stream fusion for video question answering," *IEEE Transactions on Image Processing,* vol. 29, pp. 1204-1218, 2019.

[6] K. D. Muthusamy Sellamuthu, B. S. Basavaraj, L. A. Balaji, B. Mohan, and B. Ramachandran, "AGeES: automatic multiple choice question (MCQ) generation from extractive summary of video lectures using BertSum," in *International Conference on Smart Learning Environments*, 2023: Springer, pp. 22-31.

[7] K. Khurana and U. Deshpande, "Video question-answering techniques, benchmark datasets and evaluation metrics leveraging video captioning: a comprehensive survey," *IEEE Access,* vol. 9, pp. 43799-43823, 2021.

[8] Z. Guo *et al.*, "Multi-turn video question generation via reinforced multi-choice attention network," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 31, no. 5, pp. 1697-1710, 2020.