Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Information Technology

Academic Year 2021 – 22

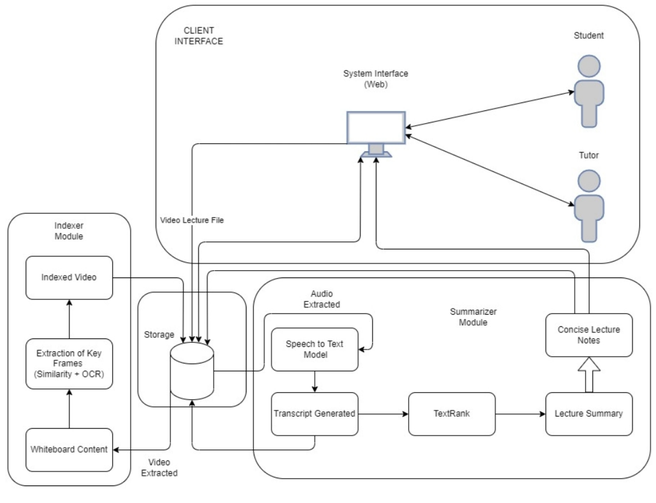
Project Title: E-Lectures Summarizer

IT Dept.: Project Guide: Dr. Abhijit Joshi; Project Members: Akanksha Mansharamani, Manav Shah, Varun Vora

Abstract: In these troubled times of pandemic, the teaching- learning process has been affected adversely at every level of education. Schools and colleges had to shift and adapt to various online platforms for achieving academic goals set by respective universities and boards, and to emulate the past trends of the academic year and the overall curriculum. Some cons of this online process are that the lectures can be long, boring, and monotonous. With the absence of a classroom environment, personal touch is lost because of which students have a shorter attention span than usual. If a student misses on a lecture or wishes to revise a concept from the recordings he/she must view the entire recording of the lecture. Moreover, the student must put in extra time and efforts into compiling notes. This is not always feasible especially when the student is on a time crunch. Thus, a system that should aid the students in studying for their exams and utilizing studying time with efficiency seems to be something that would be welcomed by open arms from students as well as the teaching community. This aid can be acquired by using AI methodologies to generate summaries of lectures in the form of concise lecture notes and time-indexed videos according to subtopics.

Introduction: The ongoing pandemic took a huge hit on the learning system as we know it. As a result, student interactions in the online classroom have declined. A UK Engagement Survey finds less than two-thirds of undergraduates feel universities ensured quality of academic experience during Covid-19. Many students struggle to keep up with college classes and have lost motivation to study. Lack of social interaction has further alleviated the confidence of learners. Via our system, we aim to eliminate all the above-mentioned problems. The proposed model will not only increase learner satisfaction but also boost the confidence of students lagging academically and help them catch up with their peers. The use of visual variety has proved to be aesthetically appealing. Thus, we plan on leveraging this to our advantage to grab the learner's attention. Our proposed system not only makes the learner process easier, but also helps student save onto study time. Students can easily view summaries before their exams instead of having to go through the trouble of frantically searching through study material and resources. This in turn, will boost academic grades of the students.

Architecture:



We have the following modules in our system:

A) User Module

The user can either be assigned the role of a student/teacher. The database has the data of all the currently registered students and teachers at the institute. Only the admin has the provision of making changes to this database, thus removing, and adding new users to the system. The tutor has the provision of either uploading the video-recordings files of the lectures conducted or providing a link of the recording. These recordings are stored in the database with a unique ID for the future reference of students/tutors which are retrieved for generating the E- lecture summary in the summarizer module and for indexing the video in the Indexing module.

B) Indexer Module

Video indexing, like the index page of a book, is a technique that enables viewers to readily access and navigate through the video's contents. The video recorded courses saved in the database will be supplied as an input to the Indexing module in our proposed architecture. Entities will then be extracted from the generated transcripts of the lecture. Once the entities have been extracted, they will be indicated with corresponding timestamps. These time-stamp entity pairs called 'bookmarks' will be displayed to the student along with a summary so that the student can directly skip to view the desired part of the lecture

C) Summarizer Module

For the summary, we will first generate transcripts from the video. The generated transcript is stored in the database. For summarizing the transcript, extractive summarization will be implemented. This section will provide a concise summary and notes of the video recording.

Algorithm:

- **TextRank for Extractive Summarization:**

TextRank is an extractive summarization technique. It is based on the concept that words which occur more frequently are significant. Hence , the sentences containing highly frequent words are important. Based on this , the algorithm assigns scores to each sentence in the text. The top-ranked sentences make it to the summary. Graph-based ranking algorithms are a way for deciding the importance of a vertex within a graph, based on global information recursively drawn from the entire graph. The basic idea implemented by a graph-based ranking model is that of voting or recommendation. When one vertex links to another one, it is basically casting a vote for that vertex. The higher the number of votes cast for a vertex, the higher the importance of that vertex.

We have to build a graph that represents the text, interconnects words or other text entities with meaningful relations. TextRank includes two NLP tasks-

1) Keyword extraction task and 2) Sentence extraction task

- **Optimal Character Recognition:**

Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten, or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo. For detecting the content of lecture slides, we use Google's Cloud Vision API which offers powerful pre-trained machine learning models through multiple REST and RPC APIs. This tool assigns labels to images and quickly classifies them into millions of predefined categories. It is best at detecting objects and faces, read printed and handwritten text, and build valuable metadata from the image dataset.

- **SSIM for Detecting Keyframes:**

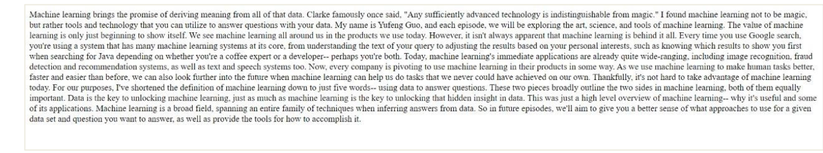
The structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric; in other words, the measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as reference.

In our algorithm, we use an SSIM threshold between consecutive keyframes to determine if a slide has content that is significantly different from the previous keyframe. These keyframes are given titles and are recorded as indexes with their timestamps in the video.

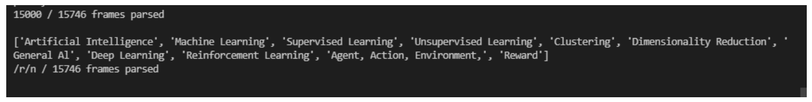
Results:

After exploring all the technologies and techniques for developing an AI-based lecture summarizer our model will attempt to make learning a fun experience. Outcomes of the application will be as follows:

• Text-based summary of the lecture based on the transcript generated


Machine learning brings the promise of deriving meaning from all of that data. Clarke famously once said, "Any sufficiently advanced technology is indistinguishable from magic." I found machine learning not to be magic, but rather tools and technology that you can utilize to answer questions with your data. My name is Yufeng Guo, and each episode, we will be exploring the art, science, and tools of machine learning. The value of machine learning is only just beginning to show itself. We see machine learning all around us in the products we use today. However, it isn't always apparent that machine learning is behind it all. Every time you use Google search, you're using a system that has many machine learning systems at its core, from understanding the text of your query to adjusting the results based on your personal interests, such as knowing which results to show you first when searching for Java depending on whether you're a coffee expert or a developer-- perhaps you're both. Today, machine learning's immediate applications are already quite wide-ranging, including image recognition, fraud detection and recommendation systems, as well as text and speech systems too. Now, every company is pivoting to use machine learning in their products in some way. As we use machine learning to make human tasks better, faster and easier than before, we can also look further into the future when machine learning can help us do tasks that we never could have achieved on our own. Thankfully, it's not hard to take advantage of machine learning today. For our purposes, I've shortened the definition of machine learning down to just five words-- using data to answer questions. These two pieces broadly outline the two sides in machine learning, both of them equally important. Data is the key to unlocking machine learning, just as much as machine learning is the key to unlocking that hidden insight in data. This was just a high level overview of machine learning-- why it's useful and some of its applications. Machine learning is a broad field, spanning an entire family of techniques when inferring answers from data. So in future episodes, we'll aim to give you a better sense of what approaches to use for a given data set and question you want to answer, as well as provide the tools for how to accomplish it.

• Detection of keyframes using snapshots of whiteboard content



• Indexed video based on bookmarks

Conclusion: Video Summarization and Indexing has been proposed and implemented using many different approaches intimating a plethora of techniques for consuming content from all parts of the video. This system needs a pipeline that bifurcates into the Summarizer and Indexer modules with disparate use of methods in each moving part. IBM Watson's speech-to-text model seems to provide the most promising results with a provision to understand multiple accents of the English language. When we arrive at the summarization task, the TextRank algorithm provides satisfactory results, in terms of the length of the summary, the organization of it and absence of any major grammatical errors. In the Indexer module, regular snapshots need to be taken using python's OpenCV library and in the next phase of implementation key frames are to be extracted with the help of similarity models to eliminate redundant frames. The visual content in these key frames is to be recognized using Google Cloud Vision's OCR tool, which provides state-of-the-art results. After detecting the whiteboard content, we use SSIM as a similarity measure to detect, label and index keyframes which are important checkpoints in a video lecture. Both modules must be integrated into the Application that will abstract the workings of this system and act as a bridge between students, teachers and the process happening behind the UI.

References:

[1] PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization, Jingqing Zhang, Yao Zhao, Mohammad Saleh, Peter J. Liu

[2] Automatic Text Summarization of Video Lectures Using Subtitles,Shruti Garg

[3] AccessMath: Indexing and Retrieving Video Segments containing Math Expressions based on Visual Similarity - Kenny Davila, Anurag Agarwal, Roger Gaborski, Richard Zanibbi, Stephanie Ludi

[4] Automated Summarization of Lecture Videos - Anusha Vimalaksha, Abhijit Prekash, Siddharth Vinay, NS Kumar - 2018 IEEE Ninth International Conference on Technology for Education

[5] Automated Detection of Handwritten Whiteboard Content in Lecture Videos for Summarization

Acknowledgment: We are extremely grateful to our project in-charge, Dr. Abhijit R Joshi and our respected Principal, Dr. Hari Vasudevan for his constant support and encouragement throughout the course of this project. Our thanks and appreciation go to the college and staff for providing us with the necessary resources and valuable suggestions. Our classmates and team members have given valuable inputs in this proposal which further gave us the inspiration to improve our project. We thank them all for their help to complete our project.