Automatic chapterization of Videos – Progress Report

Vibhor Jain (NetId: vibhorj2, Captain)
Nitish Jain (NetId: nitishj2)

Introduction

This is a progress report of our Automatic Chapterization of video project.

Current Status

Phases:

We had decided to divide the work in 3 major phases namely

Phase 1: Speech to Text: Using a publicly available speech to text service like Azure Cognitive service, we will convert a video to text format.

Phase 2: Preprocess: Preprocess the text removing Stop words etc. Use Glove word embedding vectors for semantic meaning of sentences. Use Azure Cognitive services API to get timestamps for each beginning sentence of a section.

Phase 3: Display the Table of Contents: Finally get the title phrase for each section, using an extraction-based method. Display section timestamps and titles on a webpage.

We are more than half way through. Below is a progress report at task level:

Work Breakdown:

S. No	Task	Status
1	End to end System design	Completed
2	Basic understanding of the Azure Cognitive service / Glove / Python web frameworks	Completed
3	Understanding various Audio / Video Formats which can be supported with Azure Cognitive Service	Completed
4	Gather and Clean Data	Completed
5	Converts Video to Audio	Completed
6	Audio to Text using Azure Cognitive APIs	In progress
7	Use Glove Word Embedding Vectors for seaming meaning of Sentences	Not started
8	Use Cognitive servicesAPI to get Timestamps for each beginning sentence of section	In progress
9	Get Title Phrase for each section	Not started
10	Display section timestamps and titles on a webpage	Not started

1) **Which tasks have been completed**? – ***Above table reflects status of our tasks***
2) **Which tasks are pending**? – ***Above table reflects status of our tasks***
3) **Are you facing any challenges?**

   We faced various challenges while implementing the project which is typical with any small to mid-sized project. Here we are listing the primary ones:

a. We spent significant time in understanding the Video / Audio codecs and containers. Finally decided to use the ffmpeg tool to convert Video to Audio as Azure cognitive services only take Audio as an input. Further, Azure cognitive services need Audio in a specified codec and specified bit rate, which took time in experimentation.
b. We now plan to scope our project to only include Audio Chapterization as Video to Audio will only add an additional step in the project which does not help in our primary goal of chapterization
c. Azure cognitive services need authentication which can be token based or subscription key based, using token based auth would have taken more time than we had planned for this task so we decided to use the Subscription key based which makes key rotations little hard but works for a non-production project.
d. With the limited time available being able to build a GUI for the project looks highly unlikely & we will create instructions to run the project instead via some command line options.
e. Testing effort of 2 hours per person looks overly optimistic.