Automatic chapterization of Videos

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

Introduction

Videos are proliferating in our everyday life, starting from pure entertainment like Netflix, Hulu to more involved videos like Khan Academy, Udemy, Pluralsight, videos are everywhere. However, despite the ubiquity of videos, navigating videos is still a difficult proposition. We have experience with documents and can easily search and navigate in documents by searching, using Table of contents, Bookmarks, Navigation hierarchies etc. However same is not yet available with videos. With the ever-increasing adoption of '*everything online life*' this has become a necessity. Nobody wants to browse through hours of videos to get to the relevant section of the video. Our goal is to Chapterize a video, i.e., automatically preparing a Table of Content for a video.

Approach

We plan to do this in the 3 major phases.

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

Preprocess: Now using the available text we would 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.

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.

Expected Outcomes

Build a chapterization of a video. Given a video produce a Table of content of the Video where every new paragraph is a Table of content.

Evaluation

The evaluation would involve watching / listening to the video and verify when new sentences begin and manually verifying if the generated Table of Content matches with the actual transitions or not.

Tools & Programming Language

We plan to use the Python programming language, with possibly a mix of C# (For website hosting / Building if time permits as a stretch goal).

Python Glove: Global Vectors for Word Representation. GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.

We also plan to use Azure Cognitive service to convert videos to Audio (i.e., Speech to Text)

Work Breakdown:

S. No	**Task**	**Hours**
1	End to end System design	2 * 2 = 4
2	Basic understanding of the Azure Cognitive service / Glove / Python web frameworks	2 * 2 = 4
1	Understanding various Audio / Video Formats which can be supported with Azure Cognitive Service	3
2	Gather and Clean Data	2
1	Converts Video to Audio	3
2	Audio to Text using Azure Cognitive APIs	5
1	Use Glove Word Embedding Vectors for seaming meaning of Sentences	5
2	Use Cognitive servicesAPI to get Timestamps for each beginning sentence of section	5
1	Get Title Phrase for each section	5
2	Display section timestamps and titles on a webpage	5
Testing		2 * 2 = 4
	TOTAL	**45**

References:

https://azure.microsoft.com/en-us/services/cognitive-services

https://nlp.stanford.edu/projects/glove/