Machine Learning Assignment Detailed Abstract

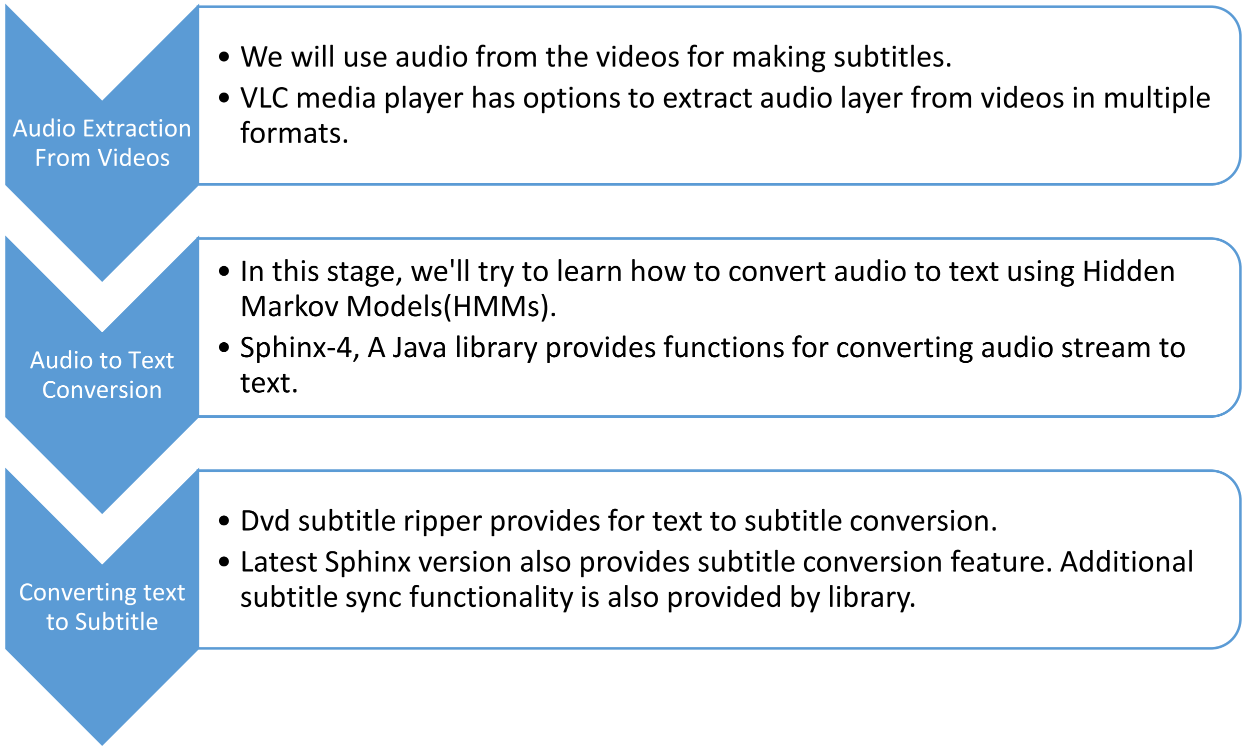
Automatic Subtitle Generation for Videos

Team members:
Nitin Labhishetty (2012A7PS038P)
Pavan Ravishankar (2012C6PS331P)
Utkarsh Pathrabe (2012A7PS034P)

Problem & Motivation:

Every time we watch a movie we have to search for subtitles, download them and add it manually to the video player. Further, if the sync and quality aren't good, we have to repeat the process. And for other videos such as lectures, conference or informational talks we seldom find subtitles. If the subtitles can be made from the audio content of the videos, it will be very convenient for the users.

Process Flow:



Audio Extraction From Videos

- We will use audio from the videos for making subtitles.
- VLC media player has options to extract audio layer from videos in multiple formats.

Audio to Text Conversion

- In this stage, we'll try to learn how to convert audio to text using Hidden Markov Models(HMMs).
- Sphinx-4, A Java library provides functions for converting audio stream to text.

Converting text to Subtitle

- Dvd subtitle ripper provides for text to subtitle conversion.
- Latest Sphinx version also provides subtitle conversion feature. Additional subtitle sync functionality is also provided by library.

Implementation Details:

Audio Extraction from Videos:

For creating subtitles/ captions for videos, we decided to use audio component from the videos due to the following reasons:

1. Video may not give information relevant to making subtitles. Also the source of speech may not be in focus or video quality maybe bad for effective analysis. Audio gives much of the relevant information for making subtitles, robust to situations like source of speech out of video frame e.g. Background narrations.
2. Speech-to-text is a widely researched domain. Primarily applied in voice command control, there are many libraries and algorithms developed for this purpose. Hence there is good support for this process.
3. Audio data is much lesser than video data in complexity, hence this is simpler data giving more effective results.

We will use VLC media player's feature to extract the audio content of all supported video formats in the form of original audio or other audio formats (such as .flac & .wav). This can be done by using the GUI or VLC command line tool.

Audio to Text Conversion

There are several libraries providing functions for Speech-to-Text conversion. We have chosen CMU Sphinx library for this purpose. Sphinx is a java library with functions for real time, audio etc. speech to text conversion. We will use StreamSpeechRecognizer for recognizing the text from audio. Using Hidden Markov Model algorithm on a dataset of audiobooks from an online digital library LibriVox, newspaper reading of Herald dataset by Centre for Speech Technology Research and collection of videos from YouTube, a classification model is created. First, the audio file is processed and converted to phonemes (phonetic sounds in English). Using this and a dataset of phoneme representations of English words (referred as acoustic file), a Hidden Markov Model is used to form the English words. The Sphinx library has an added advantage that a language model is used for creating meaningful sentences from the words.

Subtitle generation from text

Once the captions/ subtitles are formed from analyzing the audio file, it has to be converted to a supported subtitle format (like .srt) for use. Dvd subtitle ripper, Xilisoft Dvd subtitle creator can be used to create subtitles from text. We are considering looking at the latest Sphinx version that has functionality for making subtitle and syncing it with the audio for better results.

Dataset used

- Online digital library of audiobooks and their text from LibriVox.com
- Readings of the Herald newspaper
- A collection of random videos with subtitles from YouTube.com
- Acoustic data of Phoneme representation of English words by CMU (This is part of the speech recognition library, not an individual dataset)