Journal Report 2

9/8/19 - 9/15/19

Arnav Bansal
Computer Systems Research Lab
Period 5, Bansal

Daily Log

Monday September 2

Experimented with Google and Microsoft speech-to-text APIs on sample .wav file (couple megabytes). Tests proved successful with 98 percent accuracy for transcription rate. Looked for summary APIs to get an idea of next steps.

Wednesday September 4

Tried out an API Arvind found in Python that won't require payment down the line (as Google Microsoft will for larger TED talks and datasets of thousands of them): https://realpython.com/python-speech-recognition/. Achieved parallel results with Arvind on different test files, confirming it could be used. Had to overcome some hurdles regarding script helper programs in the back-end of the API. Consulted StackOverflow and tweaked some parameters to get it working. Confident can use this API going forward. Realized can also use the Google and Microsoft ones and compare results.

Friday September 6

Started thinking about how I would scale the work so far to work on the entire TED talk dataset and even with just bigger videos. Currently used the python API manually, so researched ways I could automate this, since there doesn't appear to be readily available API tokens or calls to make for automated use. Also looked more into the Mellin and discrete cosine transform algorithms from last week and thought about how I would integrate those as well for use on this dataset.

Timeline

Date	Goal	Met
August 30th	Finalize project idea, Review timeline, Make any modifications necessary, Focus on plan for upcoming weeks	Yes, Submitted Journal Report 0, Discussed plan with Mr. White
September 6th	Finish formatting dataset for initial processing and proof-of-principle model training	Yes, Have dataset with 2,461 entries of TED Talks corresponding to their transcripts and metadata
September 13th	Test various Speech/Audio to Text APIs on this dataset	Yes, Implemented Speech Recognition tests with sample .wav file with Python, Google, and Microsoft APIs.
September 20th	Have initial results for baseline implementations of these APIs on this dataset	
September 27th	Scale Python API (later in conjunction with Google and Microsoft APIs to compare results) to a small subset of the TED Talk dataset	

Reflection

It was great that we were able to find a free, Python-based API. Tests went well with small .wav files and it will now be a bit more difficult to scale this. I had to overcome some errors with parameters the Python APIs and errors in file upload sizes and lag with the Google and Microsoft APIs. After overcoming these errors, I was able to research more on how I would scale this with a helper program.

It was also good that it appears the three APIs will be compatible as far as input goes, so that makes for easy comparison for greater reliability later when using this data for summarization.