Journal Report 5
9/29/19 - 10/04/19
Arnav Bansal
Computer Systems Research Lab
Period 5, Bansal

Daily Log

Monday September 2

Read through the "Synchronous: Recognize Audio" section of IBM Watson's (lengthy) Speech-to-Text API documentation (https://cloud.ibm.com/apidocs/speech-to-text?code=pythonrecognize-audio). Found this third-party website hosted on GitHub for demos using this Watson API with an interactive GUI (https://speech-to-text-demo.ng.bluemix.net/), so this site makes it easy for rather quickly obtaining results of transcripts for a given video.

Wednesday September 4

Tried to figure out how to use the Watson API directly. Looked into making HTTP requests through Python with my Watson API key. Examined the parameters of Watson's recognize method for audio transcription. There are some robust parameters that I can experiment more with later after I get a sizeable dataset for initial tests in the project. Was a bit stuck with working through Watson API and learning how to access and use it locally, so turned a bit to the API demo site for transcripts.

Friday September 6

Experimented with the Watson API demo site and obtained transcripts for 5 TED Talks for initial data set. Realized can run these on the website in parallel with multiple tabs, browsers, and/or computers, including with Arvind doing the same. For our initial purposes, this should suffice in obtaining transcripts since the more complex code for fetching them has already been implemented with this demo site. Manually browsed through and compared these transcriptions with the real TED-curated transcripts, and they still seem to be quite accurate.

Timeline

Date	Goal	Met
September 20th	Have initial results for baseline implementations of these APIs on this dataset	Yes, got transcripts from APIs (albeit with errors for laughter and applause)
September 27th	Find high-quality algorithms for laughter and applause detection in audio files, Manually approximate average lengths of each in several TED Talks	Yes, see previous page with algorithms found and summary discussion of results for lengths of laughter and applause in TED talks; Also pursued additional project goals
October 4th	Successfully set-up Watson for local implementation of curl and Python API requests for speech to text on some TED talks and compare those transcripts to the real ones manually processed by TED	Partially, still working/figuring out how to use IBM Watson's Speech to Text API in Python locally with the HTTP requests, but still managed to obtain transcripts for some TED Talks and compare them to the real ones manually processed by TED
October 11th	Among me, Arvind, and Mr. White, figure out how to use IBM Watson's API locally in Python for scaling the dataset in the near future and continue to obtain as many TED Talk transcripts as possible for initial tests of summarization in the coming weeks	
October 18th	Scale dataset using IBM Watson API to all 27 TED Talks allowed this month (500 minutes total cap), then decide when we'll need to upgrade to a higher API access level for more requests and transcripts and start video transcript summarization task of the project	

Reflection

IBM Watson is still great and the most feasible option going forward, but it's taking some time to figure out completely how to use it effectively in Python locally for when we scale the dataset. The demo website proved very fruitful and efficient, so that was nice too that I'm able to get transcripts much more simply and with similar speeds as far as processing times. I'm also able to parallelize requests via this demo site to some extent, so it may not be much slower than an implementation in Python of IBM Watson API requests, but time will tell about which of the two I proceed with, or possibly both at different upcoming stages of the project. I'm glad we're getting nearer to being able to start initial summarization tests.