Journal Report 4
9/22/19 - 9/29/19
Arnav Bansal
Computer Systems Research Lab
Period 5, Bansal

Daily Log

Monday September 2

Found code for laughter detection algorithms:
(https://github.com/ideo/LaughDetection,
https://github.com/jrgillick/laughter-detection)
and papers for other methods, most involving recurrent neural networks with long short-term memory cells:
(https://labs.ideo.com/2018/06/15/how-to-build-your-own-laugh-detector/,
http://www1.icsi.berkeley.edu/ knoxm/laughter-v10.pdf,
https://pdfs.semanticscholar.org/151c/41636240c97ad6c91b09caae698329612257.pdf,
https://www.ee.columbia.edu/ dpwe/pubs/nist04-laughs.pdf,
https://www.researchgate.net/publication/221487656-Automatic-laughter-detection-using-neural-networks).

Did not find code yet for applause detection but found similar papers, which I could implement the algorithms of
(https://ieeexplore.ieee.org/document/5941827,
https://www.researchgate.net/publication/220226738-Characteristics-based-effective-applause-detection-for-meeting-speech).

Also did more work on manually estimating/timing laughter and applause moments in a handful of TED Talks and figured 3 second splits of the audio file could be viable.

Wednesday September 4

Sick again :(

Friday September 6

Corresponded with Mr. White on project, looked into IBM Watson speech-to-text, signed up for account via TJ email. The API allows for 500 minutes of transcription per month in the free version (27 TED talks that are usually 18 minutes each over the next month) and upload file size maximum is 100 MB, so this API is perfect to use. Ran a test transcription with the Al Gore TED Talk on Averting the Climate Crisis, which includes a lot of tricky/obscure words, and Watson performed quite well. It seems to think the same speaker is multiple (due to voice inflections and whatnot), but we can ignore that since TED talks are monologues. The API robustly overcomes any errors arising from laughter and applause. It uses curl for API requests and there are functionalities in Python to use the API.

Timeline

Date	Goal	Met
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	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	
October 11th	Figure out plan for potentially upgrading from free API use to scale transcription on a larger portion of our 2461 TED Talk Dataset, Implement Accordingly	

Reflection

It was a great relief that IBM Watson provide fruitful after Google and Microsoft didn't work so well and we ran into issues with the less robust Python-based/in-built Speech Recognition API. Watson has some of the most robust and state-of-the-art algorithms readily available for utilization, so that's great. It's also good that I have resources to turn to for implementing laughter and applause detection, should I need to return to that in the near future for any reason. For the time being, I'm excited to proceed with Python implementations of API calls to Watson and curating a solid dataset of TED Talk transcripts. In addition, Mr. White's suggestion for me and Arvind to now complement each other's work more should hopefully accelerate our work. Succcessful Watson Transcript example:

Voice Model:

US English broadband model (16KHz)

Keywords to spot:

IBM,admired,AI,transformations,cognitive,Artificial Intelligence,da

Detect multiple speakers

Record Audio | Upload Audio File | Play Sample 1 | Play Sample 2

Text | Word Timings and Alternatives | Keywords (3/9) | JSON

Speaker 2: Ted talks are recorded live at the Ted conference and produced with WNYC New York public radio this episode features former US vice president Al Gore Ted talks are made possible through the support of B. M. W. here's al gore.

Speaker 2: Well it's really a great honor to have the opportunity to come to the stage twice I'm extremely grateful I have been blown away by this conference thing I want to thank you all for it or the the mini.

Speaker 2: Nice comments about what I had to say the the other night and I and I say that sincerely.

Speaker 2: Partly because.

Speaker 2: I need that.

Speaker 2: Put yourselves in my position.

Speaker 2: I've seen are forced to for eight years.

Speaker 2: Now I have to take off my shoes or boots to get on an airplane.

Speaker 2: Okay one quick story to illustrate.

Speaker 2: What that's been like for me.