Journal Report 15

01/11/19 - 01/18/19

Arvind Ravipati
Computer Systems Research Lab
Period 1, White

Daily Log

Monday January 13

Continued Reading Through https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/.

Tuesday January 14

Read More about the Text Summarization Method from https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09fc70. Read About Cosine Distances and their Application in Text Summarization. Loooked into NumPy and NLTK (Natural Language Tool Kit) and their Potential Use in Future Implementation.

Thursday January 16

Began Implementing a Text Summarization Script from Scratch. Wrote Code that Split the Summary into an Array of Arrays, with Each Sentence an Element of the Array and each Word in the Sentence a Part of the Smaller Arrays. Began Writing Code that Would Construct a Graph that would Contain Similarities between Sentences. Plan to Use NumPy to Format this Graph.

Timeline

Date	Goal	Met
Dec 20	Summarize a TED Talk by Implementing a Text-Summarization API	Yes, Was Able to Summarize a TED Talk to an Extent
Jan 10	Look Into Methods of Text Summarization from Scratch	Yes, Found a Few Sites About Text Summarization from Scratch
Jan 17	Continue to Look Into Methods of Text Summarization from Scratch	Yes, Began Implementing a Text Summarization Script
Jan 24	Be Able to Summarize the Obama Speech in Less than 4 Sentences	
Jan 31	Research into Audio to Text Methods that Can Account for Punctuation	

Reflection

I was happy to finish reading through the thorough text summarization article and felt I developed a decent understanding of the material. However, I decided to pursue the method in an article I read earlier that used cosine distances between vectors of sentences to compare similarities. I hope to build a graph of these similarities and order them based from highest to lowest importance. I feel this will also be useful in controlling the number of sentences in the summary. Next week, I hope to be able to finish my implementation and then plan to look into the state of the art of audio to text transcriptions to see if there is any possible solution to detecting punctuation.