Zoomalayis Project Proposal

CS 410 - Fall 2021

Project Team:

Team Lead: Preeti Amin (netid: preetia2)
Members: Kris Gallagher (netid: kmg8) and Sunnie Wang (netid: danfeng3)

Project Description:

With remote work and education becoming more of a norm rather than the exception these days, online meetings have become an important part of life. Our project entails extracting topics from online meetings. The goal is to analyze text transcripts from meetings and extract a handful of main topics discussed during the meeting. We, specifically, plan on applying topic modeling to Zoom transcripts.

We will develop our topic modeling algorithm using a similar meeting dataset and then apply it to Zoom transcripts. AMI Corpus is one such dataset consisting of 100 hours of meeting recordings, we will only be utilizing the text transcripts from the dataset - English only.

We plan on developing the code in Python with open source libraries such as NLTK and Gensim. We will evaluate different algorithms to determine the best approach.

We plan on evaluating our algorithm manually by curating and testing a dataset of about 10-20 zoom meetings, taken from mediaspace.illinois.edu, Youtube and other resources.

Project Timeline/Workload:

Define project (10/18-10/24) - 4 hrs per team member (12 hrs total)
- Form team, define project and scope, conduct initial research on datasets, libraries etc.

Project Implementation (10/25-11/28) - 16 hrs per team member (52 hrs total)
- Finalize datasets, libraries to be used to extract topics - 8 hrs
- Compare pros and cons of various algorithms for topic modeling - 4 hrs
- Implementation of the selected algorithm in Python - 20 hrs
- Design a basic interface to input meeting data and extract topics - 4 hrs
- Project alignment meetings - 4 hrs per team member - 12 hrs total
- Curate test data from online videos (mediaspace.illinois.edu/Youtube) - 4hrs

Project Evaluation and Documentation (11/29-12/5) - 4 hrs per team member (20 hrs total)
- Test model - 10 hrs
- Refine workflow/user experience - 4 hrs
- Create project presentation and documentation - 6 hrs