**CS598PS Project Proposal**

Amirhossein Aleyasen

Xiujia Yang

Alex Zhang

It happened a lot for us to go backward/forward a TV show to watch the parts that our favorite actor/actress acts. In this project, our purpose is to make it easier using ML methods and without manual searching and with minimal human intervention. Thus, we will try to develop a classifier to identify who is speaking in the TV shows audio. Our input will be TV shows audio files besides some sample short audio for each actor/actress. And the final output of the classifier will be all the time ranges actors/actresses talked (with start time and end time) with their name marked with the time range that the classifier detects. The evaluation is basically to find the accuracy of our classifier, in both the time range detection as well as the speaker recognition.

Our project will basically contain these three steps:

1. A model to extract the speech from the audio with some noise like background music, and laughing, and build a simple classifier to recognize the speakers for each extracted time range.
2. Improve the classifier accuracy by trying different methods and also testing it on different datasets and find the drawbacks and fix them.
3. Since providing the sample short audio for each actor/actress is not straightforward, if we have enough time, we will try to eliminate this step using unsupervised methods.