Smulemates: A Karaoke Singing Style Recommendation System

I've grown up singing karaoke for as long as I remember – at home, at family parties, and at get-togethers with friends. It is not only one of my favorite hobbies, but also a big part of my own culture. So naturally, I decided to focus my final project on this great pastime.

Background

The aspect of karaoke that I love the most is that it brings people together. In fact, there is a whole social app around this concept, called Smule. As of 2018, this popular karaoke app had over 50 million monthly active users. It unites people from all over the world to either sing a solo, duet, or group performance. Users may also view other users' performances, and see what's popular and trending. However, what I did notice is that the app doesn't recommend specific users based on the user's particular singing style. For myself, I know that I am unable to belt out Mariah Carey notes from the 90's, so I prefer to stay within my own vocal range. So, with millions of other users on the app, how can we make it easier for users to choose who to sing with? With this project, I propose that Smule add a feature called Smulemates, a recommendation system that suggests other users with similar singing styles.

Data

As I was looking for data for my project, I came across the Stanford Digital Archive of Mobile Performances (DAMP) Dataset, consisting of several datasets of actual Smule user performances. Since the access to the files was restricted, I sent an e-mail to the alias that was specified on the website. However, after two days of not hearing back, I took it upon myself to expedite the process by trying to contact the actual owner of the dataset. First, I looked at the parent site where the dataset was being hosted – the homepage for the Stanford Center for Computer Research in Music and Acoustics. My initial step was a call to the phone number listed, but did not receive an answer. I then looked through the department's Facebook page and reached out through Facebook messenger. Although the administrator was willing to help me figure out the contact, I was redirected back to the alias I had originally e-mailed. Since I was back at square one, I decided to take a bolder step and reach out to an individual on LinkedIn who had written a research paper using the DAMP dataset. She was a former intern at Smule, and had worked on creating one of the datasets. A few hours later, I received a very kind response that provided an e-mail address to the owner of the dataset. The next morning, I e-mailed the owner directly and received access within the hour.

Feature Extraction

Dimensionality Reduction

Content-Based Recommendation System