Charles Kash

Professor Joseph Puliparambil

COEN 4850 101 Intro to Intelligent Systems

September 27, 2019

Project Proposal

1) Motivation for choosing this problem.

   The motivation that I have for choosing this project is that I have always been interested in the sort of algorithm that YouTube and uses to measure how content is being viewed on their sight and what kind of videos are put on the trending tab on the front page of YouTube. I watch YouTube quite a bit for leisure and for school related topics, and more recently it has interested me in how YouTube and other media websites control what their users are able to see.

2) Framing of the problem.

   This project will take the datasets available to me listing various measurements over several months on the daily trending YouTube videos. The various measurements will be weighed against each other to see if each individual YouTube channel can be ranked, and if the ranking taken from Trending data is comparable to other ranking statistics that other websites have for the YouTube channels within the dataset, and channels that maybe should be in trending that aren't. The overall purpose is to see what kind of channels YouTube promotes through their algorithm and whether there is any sort of bias towards any particular type of content or certain groups.

3) Choice of the data set.

   The data set that will initially be focused on will be "Trending YouTube Video Statistics" that I found on Kaggle, which includes several months of data from 10 different countries, and gives each Trending videos data on publish time, tags, views, likes and dislikes, comment count, and various other measurements. I will see which measurements I can use to try to determine a ranking, and compare its accuracy with other statistic tracking sites such as Google Trends and SocialBlade.

4) Tools you will be using

   Right now as I am writing this project proposal, I don't have any plans on using any more tools than what is available with Jupyter Notebook, but as the project is in progress I may need to use various other tools to use the dataset.

5) Research involved in choosing the problem

I honestly have not put a huge amount of research directly into the creation of this exact project, but my interest in this sort of topic is great and I believe that I can garner some very interesting conclusions from this data that will make for a good report. I guess I have a lot of research by participating in the YouTube community for a long time.

Bibliography

https://www.kaggle.com/datasnaek/youtube-new