Youtube Video Comment Sentiment Analysis

SYRACUSE UNIVERSITY
ENGINEERING & COMPUTER SCIENCE

Contributors:

Pranay Kumar Verma (SUID-396068464)

Prateek Sahu (SUID-809311241)

Rohith Pattathil (SUID-276168593)

Nitesh Nawlani (SUID-938308230)

Sitesh Mishra (SUID-742893107)

Overview

Just after Facebook, Youtube is the social network which has the world's second largest population. By the end of January 2020, Youtube has crossed 2 Billion users. Every minute 300 hours of videos are uploaded on Youtube. As per the statistics 7 out of 10 people prefer online video platforms rather than watching live TV and Youtube is the biggest online video streaming platform. Till date about 11,000 Youtube videos have gone beyond 1 Billion views. 30 Million people visit Youtube every single day.

As per the above statistics, it is quite clear how powerful Youtube is in today's timeline. With numerous people following, it is an absolute necessity that the content of Youtube should be suitable for every single user. The only way to verify the content is by analyzing the number of views, likes/dislikes and the comments posted by the viewers. Our goal is to pick a specific Youtube channel and extract the comments from some of its latest videos and apply sentiment analysis to figure out how the recent content is being perceived by the viewers.

Significance of the Idea

Just imagine a situation where one of your friends recommend a Youtube channel where you can watch people streaming live games all over the world. But, you, being a noob who has never watched or played a single online game, don't have a single clue whether you should go ahead and watch the videos. In that case, you may need to go through the comments of that channel to find out the reality. Now there are two problems with that. The first one is some channels have millions of comments and you don't have the luxury of time to go through each and every comment. Second, we're not sure whether the comments are true or biased. There are followers who are die hard fans of the channels and there are followers who play blame games.

Along with the users, a channel owner can also use the tool to determine how the audience is reacting to their recent uploads. The tool will help them decide whether their recent content meets the expectations of their subscribers. They can get to know how the viewers are reacting to their recent uploads and based on the impact on the user base make informed decisions on the kind of videos they should be making in the future.

The project idea is based on tackling the above mentioned problems and providing the channel owner with insightful statistics. The user has a choice to give an input of a Youtube channel name. The program will select the top 50 viewed videos of that channel and extract the comments posted by all the users, this will be our dataset. Then we will go ahead with the preprocessing of the dataset which includes removing stopwords and punctuations, tokenize and lemmatize the comments, extract the useful features etc. Then we will classify the extracted comments based on their nature i.e positive, negative or neutral using sentiment analysis. Additionally, we will use different visualization methods such as word cloud, histograms, etc. to display the end result.

Implementation Details

This project will be implemented using python language and the following python packages.

Package Name	Description
Google API Client	For accessing Youtube Data API v3 to extract Videos and Comments
Pandas	For representing data in dataframe and perform manipulations
NLTK	To perform Sentiment Analysis and Preprocessing
Wordcloud	For showing most frequently used words
Matplotlib	To visualize the final outcomes
lxml	For processing XML and HTML
Request	To perform HTTP get requests.

Individual Task Distribution

Team Member	Task Description	Allotted Time Frame
Pranay Kumar Verma	Working on fetching the videos of a particular channel	03/30-04/02
Pranay Kumar Verma	Extract the most recent comments from the videos	04/03-04/06
Rohith Pattathil	Preprocessing the Dataset	04/07-04/13
Prateek Sahu	Perform Sentiment Analysis	04/14-04/20
Sitesh Mishra	Classify the comments into positive, negative or neutral	04/21-04/25
Nitesh Nawlani	Perform visualization	04/26-04/30
Prateek Sahu	Design Driver Module	05/01
Everyone	Documentation	03/30-05/01