Word-Processing with the YouTube Algorithm

Introduction:

The purpose of this project is to utilize R and run a word-processing program to simulate how YouTube runs its video recommending process as well as identify which keywords may be associated with certain genres/categories of videos to better display titles for videos to both new and old content creators to gather more views and expand their channel. By utilizing certain keywords within the title of your videos and based on keywords that are used in videos you watch, you should receive recommendations with similar words/categories. This process doesn’t seem to have been taken on by anyone and is uniquely focusing on text/word processing whereas there has been an instance which utilized running YouTubes API algorithm with videos instead of text.

Review:

Lovejoy, Chris. “I Created My Own YouTube Algorithm (to Stop Me Wasting Time).” *Chris Lovejoy*, Chris Lovejoy, 12 Nov. 2021, https://chrislovejoy.me/youtube-algorithm/.

Lovejoy uses this page to describe his process of experimenting with creating his own algorithm to determine which videos are valuable to watch and be suggested. He uses code within python to execute the metrics he collected via the YouTube API. Following this, he then uses metrics such as views per video, subscribers, comment count, etc. to filter out the top-quality videos based on searching with a specific word or set of words. He finds his algorithm is correctly identifying videos that he would enjoy and begins to run it through different mediums such as AWS Lambda. Potential next steps for Lovejoy include text/term searching which would lower the dependency on how videos will be picked up in the search. According to Lovejoy, the code is slow so having something optimized better would create a more efficient experience. Based on the results he has, there are no observable visualizations that would help to filter out results of terms or keywords and the project falls short in observing this.

In conclusion, there hasn’t been much done in trying to observe the best way to filter out the most popular/best quality videos in a similar way to how YouTube runs its algorithm and with being able to visualize the results based on certain keywords and terms for specific categories and genres. Based on research, we would be trying a new approach to an interesting problem which has been touched by few.

Link To Data: <https://www.kaggle.com/datasets/datasnaek/youtube-new?resource=download&select=USvideos.csv>