Research Discussion Assignment 1

DATA612 - Recommender Systems

William Outcault

11 June 2020

Contents

Introduction

The commercial recommender system I would like to discuss in Youtube's recommended videos. I decided on this system because of the ability for youtube to bring indivudals to the dark corners of the video streaming platform. In 2019 it was reported that 70% of Youtube's viewing time was from recommendations. Thanks to their lucrative recommender system that lures us down rabbit holes almost all of us have binged Youtube videos. That being said I was curious to research some key characteristics.

How It Works

The system is complex however from my understanding it is made up of two seperate neural networks. The first uses the user's watch history and selects videos using collaborative filtering. The second network uses logistic regression to rank the selected videos based upon features and also A/B testing.

Good or Bad Experience

To begin the "sticky" model they have built is successful in user retention. It is very difficult to diverge from Youtube and has become a source of television for some youth. Adversely they do lack in providing authoritative content upon searches relative to other recommender systems. The same searches on Google and Youtube will provide contrasting results, Google will show less low-quality or less false information results as opposed to Youtube who will display videos with a sole purpose of entertainment. The systems for each company serve different purposes therefor behaive differently.

In the past Youtube may have been deemed as a 'bad experience' because of the lack of regulations. Just recently they have tweaked their system to return fewer highly partisan channels which is good. They are working towards returning authoritative news sources towards the top of each search opposed to entertainment based content susceptible to misinformation. The site is growing away from divisive material and false information, and working towards highlighting neutral-ground material. They are continuously monitoring videos to avoid abrasive content and promoting videos for youth. Overall I would say the experience is good and getting better.

Attacks on Recommender Systems

The article https://www.washingtonpost.com/news/morning-mix/wp/2017/04/19/wisdom-of-the-crowd-imdb-users-gang-up ?utm_term=.757fa312001d was quick and to the point. The author Travis Andrews addresses collective groups of people providing low ratings for content that they have never seen before. These accounts can be referred to as bot accounts, and are made in large numbers to perform attacks. There are two types of attacks to be considered, product-push and product-nuke.

The information we can use to detect bot accounts are either the items used to construct the profile or the ratings applied to the selected items. Using this information we have to detect patterns, and detecting these patterns requires certain metrics.

- Number of Prediction-Differences: Defined for each user as the number of net prediction changes in the system after her removal from the system.
- Standard Deviation in User's Ratings: This metric represents the degree in which a rating given by a user to an item differs from her average ratings.
- Degree of Agreement with Other Users: The degree of agreement is in fact the average deviation in a user's ratings from the average rating of each item.
- Degree of Similarity with Top Neighbors: As stated by its name, this metric describes the average similarity weight with the Top-K neighbors of a user.

Using these metrics and any other certain factors available you can cluster potential bots. A larger number of prediction differences after removal of a user from the system is a clear indicator of a potential bot. For a user with a large standard deviation in ratings it might be harder to determine if they are a bot. However with a small standard deviation any particular outlier in their ratings will be easily identifiable. Falling out of line with other user ratings is tough to work with without the other metrics, however when supplemented with the other metrics this is used as a significant feature when clustering your bots. If your degree of similarity towards your top-k neighbors is substancial then you may be a bot.

Sources

https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478
https://www.infoq.com/news/2016/09/How-YouTube-Recommendation-Works/#:~:text=Another%20key%20aspect%20for
https://www.aaai.org/Papers/AAAI/2005/AAAI05-053.pdf https://dslab.epfl.ch/people/zamfir/widm2005.
pdf