Approach taken for the Solution of the Problem

By: Vatsyayan Dimri

Email: vatdimri1995@gmail.com

Problem statement

Since problem is of the type of recommendation system in which the task of the model is to predict if a particular user will like a video or dislike it or how much a particular user will rate a movie or in other words predicting the engagement score.

General Machine Learning approaches don't necessarily do well in such scenarios because they don't take into account the similarity between different users and different items (videos in our cases)

Collaborative filtering approaches usually give a much better results because they take into account how other similar users have responded to a particular content.

Method used in my Model

The type of filtering I used is Memory based in which the whole dataset is saved at the training stage and during prediction we use some type of nearest neighbor techniques to determine the rating of a video.

The filtering I used was Item based filtering because there were a lot more users than there were items and a user based approach would become computationally more complicated as the number of users increase.

Suppose we want to predict the rating for a video V1, what the algorithm does is find similar videos based on how closely they are rated by similar users. To find the similar videos we use a similarity score (or distance), if the similarity score of two items is high that means they are similar and will be rated similarly. Using the similarity score we find K Nearest Neighbor videos and then using the rating given to those K videos predict the rating of the videos V1.

The similarity score that is used in my model is cosine score which is cosine of the angle between the lines drawn by the embedding representations of items in the embedding space. Using above approach I was able to get score close to 0.46 on the test data.