WiDS - Suggestify Project

Getting Started

The objective of this project is to build a Movie Recommendation System based on Collaborative Filtering and evaluate it using RMSE method.

After reading about Recommendation Systems, their Evaluation metrics in week-1 and 2 I started my project by importing required libraries and downloading given movie, rating database. But the database given in the resources contained 20 million ratings which gave memory allocation error when I tried to make user-item interaction matrix, so I am using a smaller database from the 1st article in Objective 1.

I then made user-item interaction matrix using pivot_table() after merging movies and ratings.This matrix is used for calculating cosine similarity between users as well as between movies later in the code.

User-User Based Collaborative Filtering

I got the basic idea of what to do from the article in objective 2 and have written a function to calculate top 10 movies for a user which basically predicts ratings of unwatched movies of a given user and returns movies with highest ratings.This rating is calculated by sum of (ratings given by other user)*(similarity of other user with given user). I have then shown recommendations for user 100.

Item-Item Based Collaborative Filtering

In this we calculate similarity between movies using cosine_similarity function of sklearn library and predict the rating of unwatched movie by sum of (rating given by user to another movie)*(similarity of unwatched and another movie) and dividing it by sum of the similarities.In the file I have shown each step which I used in the function, before it to see how it works. Challenge I faced was in creating the sim_w_to_nw matrix and thinking of using matrix product.The sim_w_to_nw matrix contains not watched movies as rows and watched movies as columns. I have shown recommendations for same user 100.

Evaluating above Recommendation System

I used train-test-split in RMSE where 70% of data is used to train and remaining 30% is tested.I split the watched movies of a user into watched_test and notwatched_test. Since I am evaluating the user based recommendation system, I made a function similar to user_based where I replaced notwatched movies with notwatched_test and used same similarity matrix. Better result will be obtained by changing the similarity matrix also appropriately but I couldn't implement it due to lack of time.RMSE is calculated by taking Root Mean Square of difference of predicted and actual ratings. The function I wrote calculates RMSE of a particular user. We can average across users to get overall RMSE. I have shown RMSE for user 98.

Through this project I have learnt about recommendation systems, pandas library, basics of python and logic behind how these systems work.

By Rohit R U