

COMP 7745/8745

Spring 2018

Project: Recommendation System for Movie Ratings

Due Date: April 21, 2018 (submit on Ecourseware)

Project Description

Here you will implement the collaborative filtering algorithm and apply it to a dataset containing movie ratings.

- Implement the user-user collaborative filtering algorithm described in class (slides 15,16 of the collaborative filtering lecture) for making the predictions. You may program in C, C++, C#, Java, Python (or any other programming language of your choice!)
- The dataset we will be using is a small subset of the movie ratings data from the Netflix Prize.
- The dataset description file further describes the dataset, and will help you get started. The ratings.txt contains the review ratings, and the movie_names.txt is included for your reference that has the names of the movies corresponding to the movie-ids specified in ratings.txt
- The paper “Empirical Analysis of Predictive Algorithms for Collaborative Filtering” <https://arxiv.org/ftp/arxiv/papers/1301/1301.7363.pdf> is a good reference as well. Read up to Section 2.1 of the paper (Equations 1 and 2), and you are encouraged to read further if you have time.

Submission

Provide the source code that you write along with any libraries/references you have used. Your code should give us the following option

- **Predicted Score:** We are interested in whether you would recommend a movie (that the user has not already reviewed) to a user or not. Specifically, the input given is a user-id, movie-id, and you need to compute the predicted rating for that user-id w.r.t the input movie-id and print out the predicted rating. Note that, for computing the predicted rating, you can consider all those users in the dataset who have rated at least K movies in common with the input user as a neighbor for the input user. (you can choose a reasonable K)
- **Recommendation:** Given a user-id and a year-of-release, recommend a movie for that user for that year-of-release. Specifically, rank all the predicted scores for the users on

all the movies released in that year that the user has not yet reviewed, and pick the top scored movie.

Please provide a readme file that we can use to compile your code, and run it. No GUI is needed, you can use command-line options to run your code.

Hint: Pay special attention to complexity in your implementation, i.e., what measures can you compute and store in a lookup table, etc.

Plagiarism Notice

Please DO NOT take copy code from the internet or from others in the class. You can work in groups of 2 for this project.