COMP 4745

Fall 2018

Project: Recommendation System for Movie Ratings

Due Date: December 1, 2018 (submit on Ecourseware)

<u>**Project Description**</u>

Here you will implement the item-item collaborative filtering algorithm (discussed in class) and apply it to a dataset containing movie ratings.

- 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 data.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 data.txt (you do not really need movie_names.txt for this project)

<u>**Submission**</u>

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

- **<u>Answering queries:</u>** We are interested in whether you would recommend a movie 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 for the input movie-id and print out the predicted rating. Note that, for computing the predicted rating, you can consider a neighborhood of K movies (choose an appropriate value of K)
- 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.

Hints:

- Pay special attention to complexity in your implementation, i.e., what measures can you compute and store in a lookup table, etc.
- Sample the dataset to create a small dataset, and try it on that dataset before scaling up your code to the full dataset

Plagiarism Notice

This is an individual project. Please <u>DO NOT</u> take copy code from the internet or from others in the class.

References

- The paper "Empirical Analysis of Predictive Algorithms for Collaborative Filtering" is a nice reference for collaborative filtering.
https://arxiv.org/ftp/arxiv/papers/1301/1301.7363.pdf