COMP 7745/8745

Spring 2021

Project: Recommendation System for Movie Ratings

Due Date: April 23, 2021 (submit on Elearn)

Project Description

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

- Implement the item-item collaborative filtering algorithm described in class (collaborative filtering lecture) to predict ratings. You may use any language/platform for this project.
- 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. You do not need to use the names of the movies for this project.
- There is also a sample dataset containing synthetic data called exampledataset. Try to use this dataset before you try to scale up to Netflix since the Netflix data has a lot more items. Note that I have generated this data and the values here using Python, so you may get different values depending on your implementation which is fine.
- 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. You can use either of the options (cosine similarity given in the paper or correlation-coefficient which was discussed in class) to compute similarity between users.

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 for the input movie-id and print out the predicted rating. Note that, for computing the predicted rating, you can compute the neighborhood of a movie as a set of K movies that are rated by the user for which you are making the prediction. Note that you can choose any reasonable K for this. If you find it hard to

select the “nearest neighbors” you can use all neighbors, i.e., simple compute all the items that has a similarity weight that is non-zero with the item for which you are trying to make the prediction and use this to compute the predicted score.

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 can you pre-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 if you can manage to collaborate remotely, but this is completely optional. You can do this as an individual project.