COMP 4601A
Fall 2022 - Lab #7

Objectives

The goal for this lab is to implement the item-based nearest neighbours collaborative filtering algorithm and apply it to three basic data sets.

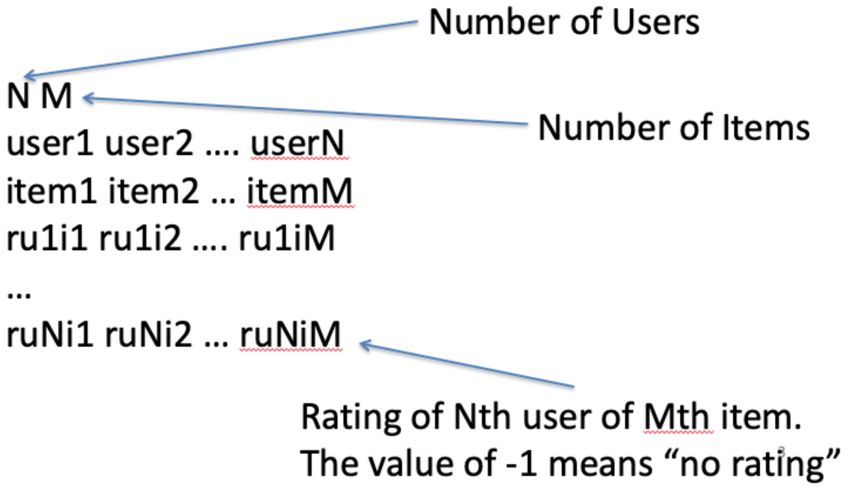
Demonstrating/Submitting

There will be several ways to receive credit for completed labs, outlined below:
1. Attend an in-person lab or online office hours and demonstrate your completed lab before the deadline. You will have to show that the goals of the lab have been completed and answer some questions about the lab and your code (see the lab reflection questions for some examples). Your grade will depend on the level of completion, as well as the quality of your design and answers. Only one partner is required for demonstration, though both are encouraged to take part. If you demonstrate your lab this way, you don’t need to submit anything on Brightspace.
2. Record a video demonstration that is ~5-10 minutes long. Ensure that your discussion in the video makes it clear that you have understood the content that the lab covers and that you demonstrate all of the required functionality. Submit a ZIP containing a copy of your code (don't include database files, etc.), your answers to the lab reflection questions, and a copy of your demonstration video (either link to a public URL in your README or include the video file directly) to Brightspace.
3. Deploy your lab to your OpenStack instance for the course and leave it running so it can be accessed by a TA for grading. Submit a ZIP containing a copy of your code (don't include database files, etc.) and your answers to the lab reflection questions to Brightspace. In your submission, include a README with the URL where your lab can be accessed by a TA. One important note - your OpenStack instance only has port 3000 open and you may have to have multiple labs running at once. I would suggest setting up your server on OpenStack to support URLs like your.server.instance.ip:3000/lab1, your.server.instance.ip:3000/lab2, etc.. I know this is straightforward with Express - I can't speak for other languages and frameworks. If you are working with a partner, only one of you should submit on Brightspace.

If you are working with a partner, only one of you should make a submission. Include a README indicating the names of both partners.

Lab Description

For this lab, you must implement the item-based nearest neighbour collaborative filtering algorithms outlined in the lecture from week #9. Your implementation must work with any of the provided text file data inputs. You should also implement your solution in a way that can be generally applied to any dataset that follows the same format. Each of these files follows the following format:



Number of Users

N M

Number of Items

user1 user2 …. userN
item1 item2 … itemM
ru1i1 ru1i2 …. ru1iM
…
ruNi1 ruNi2 … ruNiM

Rating of Nth user of Mth item.
The value of -1 means "no rating"

The first line of the file indicates the number of users and items separated by a space. The second/third lines contain each username/item separated by spaces. The remaining lines indicate the ratings of the users for the products. An entry of -1 means the user has not rated that product.

For this lab, you must demonstrate that you can run your recommender system for each text input. The output for your recommender system should be a completed matrix with all of the -1 values from the input data filled in with their predicted rating values. A simple console output is sufficient for the lab.

You should use the adjusted cosine similarity measurement for item similarity. The last two lectures have discussed the problem of selecting an appropriate neighbourhood size. For the purposes of the lab, use a neighbourhood size of 2 for all calculations by default but only consider similarity values greater than 0 (i.e., if only one similarity is greater than 0, your neighbourhood size becomes 1). Your code should be implemented in a general way that would allow modification of the neighbourhood size (i.e., create a neighbourhood size parameter). This will be important for investigating different neighbourhood sizes as part of the second assignment.