Team name: IDK

Team members: Ming Feng Chua, Rena Pei Qi Chong, Yingxuan Wu, Kevin Li
Date: 04/14/23

This journal has to be uploaded to 1) a **dropbox folder (https://www.dropbox.com/request/oMzqzRNvauGRuF97kjyh). AND** 2) posted to your project website every two weeks by 11:59 pm on the second Friday (except the spring break). When uploading to the dropbox folder, change the file name to "Team_name_MM-DD-2023" where MM = month and DD= day of upload).

Team roles for this report (write down name):

Facilitator(s): Zijun Li, Ming Feng Chua, Rena Pei Qi Chong

Recorder(s): Zijun Li, Ming Feng Chua, Rena Pei Qi Chong

Deliverer(s): Ming Feng Chua, Rena Pei Qi Chong

Planner(s): Ming Feng Chua, Rena Pei Qi Chong

Team Contact: Ming Feng Chua

See last page for description of roles. Obviously one person can take more than one role or there can be more than one person per role or make your own roles!

0. **Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down "we are still coming up with ideas on team project".**

Our chosen topic for the project is a movie recommender system. We will be working towards a movie recommender system.

I. **What was done during the report period regarding the project:** If you want to include code include this in the Appendix. Describe what the group did (including contributions of individual team members) with regards to the group project during this report period. Give enough details so I understand what you folks have been doing over the week. Include dates of your meeting(s) and who met on these days.

After conducting further research, we have found multiple research papers supporting the view that matrix factorization could provide better performance and accuracy in terms of movie rating predictions. Thus, we have decided to attempt implementing other approaches to verify the conclusions drawn from the research papers.

We are trying to discuss the division of work to implement the other approaches to movie recommender systems and the methods to create a user-item matrix required for many of the movie recommender system approaches.

II. **What were obstacles faced if any in working on the project?** This could be technical (like not being able to implement or understand particular techniques) or time issues (midterms for other courses etc).

We are using the MovieLens 25 million dataset. As various methods require us to transform the dataset into a user-item matrix, the dataset was too large to be processed at once as it requires large memory. Additionally, the large dataset resulted in lower processing speed in our code.

We are considering changing to a smaller dataset or reducing the size of the current dataset so that we can proceed on and finish up the modeling.

III. What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines (“ I plan to finish understanding x to see if it can be implemented for our project by Wednesday etc”.)

We have delegated different movie recommender approaches, such as item-based collaborative filtering, collaborative filtering with matrix factorization and content-based filtering, to each of the team members to implement. This allows our team to compare the efficiency and accuracy of each of the methods to help us narrow down and select the best approach to use for our project. Regardless of the approach we end up with for the movie recommender system, we plan on designing a machine learning pipeline such that subsequent parts after the selection of the best approach can be easily delegated to each of the team members.

What each team member will be working on:

Rena: I will continue working on the item-based collaborative filtering, mainly using user ratings to recommend movies to similar users. I will also be reading up on matrix factorization to explore using the method for our recommendation system and assist in implementing it.

Ming Feng: I will attempt to implement and process the data to obtain the user-item matrix required for matrix factorization and item-based collaborative filtering, using the full data set which contains 25m ratings. After which, I will implement the matrix factorization through methods such as singular value decomposition. The optimal smaller dimensional matrices can then be obtained through stochastic gradient descent. After implementing this matrix factorization approach, I will design the machine learning pipeline such that work can be easily delegated amongst the team members.

Yingxuan: will try coding on modifying data sets to see if the 25m one data set is doable and further collaborate with teammates to make progress on the project building and prepare for the presentation date.

Zijun: I will try to get more resources on the coding aspect of the project and discuss with teammates about the specific tasks that are expected to be done by me in the presentation and report.

While in the biweekly document above you will describe what your team did with regards to the team project (with proper attributions of who did what in the week) there are 4 pre-defined roles. I urge you to have different people do these jobs every week so that you gain experience in each of the jobs. There can also be more than one person per job for example 2 people recording the weekly journal.

Facilitator: Manages the group for this week including setting up times for group members to meet, making sure everyone has a say in the meetings etc.

Recorder: Person in charge of recording the meetings as well as the happenings of the past two weeks and describing what was accomplished in the meeting and writing up this report.

Deliverer: Person in charge of checking the entire report and uploading the file to dropbox folder and project website, as well as the representative of the group getting in touch with the instructor.

Planner: Person in charge of what will be happening next two weeks as well as thinking about longer term goals (what more needs to be done for the project).

Team contact: Person I can email if I see any issues in the biweekly report instead of mass spamming everyone in the team.