AIE425 Intelligent Recommender Systems
Assignment #1: Neighborhood CF models (user, item-based CF)
Submission Date: During Week 6 Lab (Tuesday, November 5, 2024)

1. Instructions

- The total number of pages if this file is 4 pages.
- Students complete the assignment in order.
- All students must have a GitHub account.
- Each student **MUST** prepare the homework solution and write it up as a report **BEFORE ARRIVING WEEK 6 LAB**. Your write-up may include a backdrop and theoretical description, a dataset description, a description of the experiments, results, and his or her opinions. HANDWRITTEN scans for your write-up ARE NOT ACCEPTED.
- **Each student must finish the assignment during week 6 lab and post the solution and the pre-prepared report to GitHub.**
- Failure to attend week 6 lab and complete the assignment will result in a ZERO in this assignment.
- The written report should be prepared in accordance with the guidelines in Section 5.
- The report must contain the following:
  - Assignment title page.
  - Core idea of the assignment in your own wording.
  - The solution to all steps in Section 2.3.
  - Conclusion and opinion.

2. Assignment description and requirements

2.1. Introduction

In this assignment, you will exercise some of the core topics in the Intelligent Recommender Systems especially the basic models of recommender systems, data collection using web scraper, data preparation types of ratings, neighborhood (user- and item-based) collaborative filtering algorithms.

2.2. Assignment Description

Organizations seek to tailor customers' experiences based on their preferences through suggestions. The technique is performed by developing recommender systems that allow enterprises to extract relevant material from a haystack for users while also generating insights regarding future tastes. The quality of recommendations can be improved by enhancing the current implementations of recommender systems and establishing a strong filtering mechanism on legacy data. Collaborative Filtering (CF) approaches are among the most extensively utilized recommender systems in the market today. Neighborhood collaborative filters (user- and item-based) are currently widely used recommender system algorithms.

User-item ratings are necessary to develop collaborative neighborhood filters. This process begins by collecting data such as item/product features, user behavior, contextual, demographic, and feedback (ratings) data (*see lectures 1-3*). Data sources include APIs, databases, web scrapers/Web crawlers and cookies (*see lecture 3*).

The next phase in this procedure is data preparation. In this assignment, a user-item rating matrix must be generated, where ratings are frequently specified on a scale that specifies the

particular level of like or dislike for an item. There are six types of ratings: interval-based, ordinal, binary, unary, explicit and implicit ratings (*see lecture 4*).

The third step in this procedure is to identify a peer group for the target user (in the case of user-based CFs) or a peer group for the target item (in the case of item-based CFs) using either the cosine similarity measure or the Pearson correlation coefficient (*see lecture 4 & 5*).

The final step in this process is to use the prediction algorithm to compute the rating prediction. This algorithm produces a numerical prediction indicating to what degree the user will like or dislike a certain product and the top-N list of recommended products. (*see lecture 4 & 5*).

2.3. Assignment requirements and questions:

1- Search for suitable companies in various domains that use recommender systems (*see lecture 1 & 2).*
2- List them in your report and choose one or more as the data source for the assignment.
3- Describe how the chosen company collects customer feedback and what rating type is used.
4- Prepare the collected data and take the necessary preprocessing procedures to clean it and express the feedback in the form of integer values.
5- Explain clearly the process you used to obtain and preprocess data, as well as the rating type.
6- Create your own user-item matrix and use it as the dataset for this assignment.
7- Give a complete description of the created dataset.
8- Compute the average rating and copy the results into your report under the "Assignment Results" section
9- Give a complete background/overview about user-based and item-based CF algorithms, and their detailed analytical solutions.
10- Compute the similarity using both the cosine similarity measure and the Pearson correlation coefficient to identify a peer group of users in the case of user-based CFs and a peer group of items in the case of item-based CFs.
11- Compare the results of measuring similarity using similarity measure with the results using Pearson correlation coefficient and emphasizing the pros and cons of each technique.
12- Copy the results into your report under the "Assignment Results" section.
13- Compute the rating prediction and the top-N list of recommended users/products in case of user-based and item-based CF, each case must be performed using the cosine similarity measure and the Pearson correlation coefficient.
14- Compare the results of the rating prediction and the top-N list of recommended users/products after performing step 14.
15- Copy the results into your report under the "Assignment Results" section.
16- Present, describe, compare, and evaluate the results in all cases.
17- Briefly introduce the implementation process, tools and libraries.
18- Write your own remarks about the perceived differences between user-based and item-based CF using the similarity measure and the Pearson correlation coefficient.
19- Write a conclusion that demonstrates how each strategy affected predicted accuracy.
20- Addresses any enhancement from your point of view.

3. Method and significance of the assessment

- This coursework accounts for **5%** of the whole course mark and done **INDIVIDUALLY**.
- This assessment will demonstrate progression of the student's academic ability and is used to measure students' achievement intended learning outcomes (ILOs) [3.a.1, 3.a.2, 3.a.3, 3.b.1, 3.c.1, 3.c.3 and 3.d.1].
- The marks will be awarded pro-rata, depending on the offered details, evidence of academic and technical talents, professionalism of the written content, information organization, and discussion skills.
- A **40%** reduction will be applied for each individual element if evidence of thoroughness is not adequate.
- If the student appears to be performing it right but makes a mistake, they will receive a maximum of 70% for any of the separate elements.

4. Marking schema:

- 30% of the mark for the report as described in Section [1], including evidence of the presented knowledge, topic understanding, completeness of the information, analysis, dataset description, and the contribution of the student.
- 60% of the mark for the implementing of the solution, experiments, and results.
- 10% of the mark for posting the report, code, dataset, and the PLAGIARISM REPOR on GitHub.

5. Report (paper) Format:

Report must follow the following;

- be word-processed,
- a report format standard,
- use correct paragraphing, formal grammar, tenses, and spelling,
- be submitted on A4,
- The title page includes the Assignment title <AIE425 Intelligent Recommender Systems, Fall Semester 24/25> on the first line, <Assignment #1: Neighborhood CF models (user, item-based CF)> on the second line, and <Student ID, Full Name> on the third line. Use double line space, centered contents and without page numbering.
- All other pages are 1.5 line spaced,
- use 12-point Arial font size for normal text,
- use 12-point Arial font size, bolded for headings,
- have page numbers centered on the bottom of each page in the format <Page X of Y>,
- section headings are the same as the assignment parts.

Referencing and Bibliography

- The IEEE Referencing System will be used,
- A conventional and complete Bibliography/Reference List using the IEEE Referencing System will be included,
- You are allowed to use websites; however, you are not allowed to copy content or codes from websites unless they are opensource, which you must declare clearly in your report.

6. Plagiarism and Academic Honesty

- INTEGRITY and COLLABORATION: Student are encouraged to discuss issues related to the assignment with other students, but genuine collaboration on all or part of the assignment must be explicitly acknowledged, or he/she will be penalized.
- PLAGIARISM is strictly prohibited and may result in failure in this course.
- This is an exercise, so submit your own work. If you submit material that is not entirely your own, you must state this clearly in your submission.
- A PLAGIARISM REPORT is required for the submitted report. A similarity ratio of greater than 30% is not acceptable.
- Code and/or written material should NOT be shared. Even a single line of code without reference will be considered plagiarism.
- Please do not use external code unless authorized. This is easier to detect than you might think. If you must do so for any standard initialization and embedding parts, this must be referenced and properly marked/highlighted.
- ASK Teaching Staff If you have any queries or are confused whether specific activities constitute dishonesty. It's better to be safe than sorry.

7. Feedback given to students in response to assessed work.

- Specific oral feedback on the assessed components.
- The written component will be assessed directly through annotations on the page.
- Feedback for programming will also be placed on the coursework assessment sheet returned with the coursework mark.
- During contact hours, students will receive oral generic comments on every part of their assignment..
- If students need more input, they are encouraged to speak with the teaching staff.

8. Deliverables

- The whole assignment (report, references, and dataset, and code) will be submitted on GitHub during week 6 lab.
  - All files MUST be named as follow: <StudentID_StudentFirstName_FileName>.
  - Put your name, and ID at the top of each file even in programs.
- Late submission, if there is a solid reason, is permissible during the next day subject to the arrangement with the teaching staff, however you will lose 50% of the coursework mark.
- more than 48 hours delay is not accepted, and you will get **ZERO** in this assignment.

---------- END ----------