MiniProject 1

CSX4207/ITX4207: Decision Support and Recommender Systems /
ITX4287: Selected Topic in Decision Support and Recommender Systems

Mini Project 1

- **Part I: A simple recommendation algorithm (9%)**

1. Download the datasets from the subfolder 'mini-project1' in Google shared drive:
    1. bookData.csv
    2. UserData.csv
    3. UserHistoricalView.csv
    4. TestUserAnswers.csv (*Used for Part II only*)
2. Select a technique discussed in the classes to create the **profile of 9 users** whose ids are in the file UserData.csv and UserHistoricalView.csv. << Store the created profiles in the file 'Part1_File1_Profile_Group[group_no].csv' with column headers.
3. Calculate **Cosine Similarity or Jaccard Coefficient (select only one) of each user profile and not-yet-read books** in bookData.csv, and then display the results (user/book similarity matrix). << Store the similarity matrix in the file 'Part1_File2_SimMatrix_Group[group_no].csv' with row and column headers added.
4. **Display the first 5 not-yet-read books for each user ID** in UserData.csv (*with the following column headers: User ID, Book's ISBN, Book's Title, similarity values*) with respect to the similarity results obtained in 3). << Store the result in the file 'Part1_File3_Recommendation_Group[group_no].csv'.

Mini Project 1 -- *Cont.*

- **Part II: A simple content based filtering algorithm (9%)**

1. Use the same datasets as given in Part I.
2. **Select a content based approach** discussed in the classes ***that is different from Part I*** to create the content based recommendation to **recommend top-10 not-yet-read books** to the users:
   - Store the **created profiles** in the file 'Part2_File1_Profile_Group[group_no].csv' with column headers.
   - Store the **similarity matrix and/or model used** in the file 'Part2_ File2_Model_Group[group_no].csv' with row and column header.
   - Display the **top 10 not-yet-read books for each user ID** in UserData.csv (*with the following column headers: User ID, Book's ISBN, Book's Title, model's calculated value, e.g., similarity result*) << Store the result in the file 'Part2_ File3_Recommendation_Group [group_no].csv'.
   - Calculate Precision (p=a/90), Recall (R=a/[no.of relevant items in the test dataset], and F-measure of your models that suggest top 10 not-yet-read books for all 9 users. << Store the result in the file 'Part2_ File4_Evaluation_Group[group_no].csv'.
3. **Grading criteria:** the scores given in Part II will be ranked with respect to the evaluated results in 2) and the effort of the algorithm(s) implemented and/or experimented.

Submission and Presentation (2%)

- Submit the code (a zip file), results and the presentation one day before the deadline (23:59 on Aug. 25 (Sec. 541) or Aug. 26 (Sec. 542).
- Every team member must present your individual contribution in class on Aug. 26 (Sec. 541) or Aug. 27 (Sec. 542). Otherwise, there is no score given.