National University of Computer & Emerging Sciences, Karachi
Spring 2023
School of Computing
Midterm-I Examination [Solution]
6th March 2023, 09:00 am – 10:00 am

Course Code: CS4053 / AI4006	**Course Name:** Recommender Systems
Course Instructor: Syed Zain Ul Hassan	
Student ID:	**Section:**

Instructions:

- Return the question paper after exam.
- There are **3 questions** on **1 page** with **2 sides**.
- In case of any ambiguity, you may make assumption. But your assumption should not contradict any statement in the question paper.

Time: 60 minutes **Max Marks**: 30 Points

Question 1 (CLO: 1)	**10 points**

Provide brief answers for the following and justify:

a) Why is serendipity so important in recommender systems? How do we control serendipity in item-based Collaborative Filtering?
***Answer**: By serendipity we mean recommending items to the user such that his taste is expanded. We can control serendipity in item-based collaborative filtering by adjusting the value of hyperparameter k (as in number of items to be considered in the neighborhood).*

b) How do we handle cases where the ratings are allowed to be real-valued in Naïve Bayes Collaborative Filtering?
***Answer**: By assuming the ratings to be normally distributed and using the Gaussian or z-distribution to estimate the likelihood.*

c) Let us say that we have a photography website. The site contains both user-created and AI-generated images. Should our recommender system be giving more importance to user-created images over the AI-generated ones for recommendation? Provide a reason.
***Answer**: Although there is no definite answer to this open issue yet, one strong argument is to give preference to AI generated images as often user feedback is important in performance evaluation and improvement of the model.*

d) Can content-based recommender systems solve cold-start problem if a new item is added?
***Answer**: Yes. It can solve cold-start problem if a new item is added since this new item can also be recommended to a user based on feature similarity.*

e) If an active user has ***k*** yet unrated items in a system of ***m*** total items, how many vectors would be present in the vector space model at one time?
***Answer**: k+1 (one user vector and k item vectors).*

Question 2 (CLO: 1) **10 points**

For the given data, create user profile and item profiles. Then use these profiles to predict ratings for Item 2 and Item 4 with content-based filtering.

User	Item 1	Item 2	Item 3	Item 4
Features	*f1, f2*	*f1, f3*	*f2, f3*	*f2*
Rating	5		3	

Note: The allowed values for ratings are 1, 2, 3, 4 and 5.

***Answer**:*

Encoded user content vectors

	F1	*F2*	*F3*
I1	*1*	*1*	*0*
I3	*0*	*1*	*1*

x

Rating
5
3

	F1	*F2*	*F3*
I1	*5*	*5*	*0*
I3	*0*	*3*	*3*

Normalized user profile

	F1	*F2*	*F3*
User profile	*0.312*	*0.5*	*0.187*

Weighted item matrix and Recommendation matrix

	F1	*F2*	*F3*
I2	*1*	*0*	*1*
I4	*0*	*1*	*0*

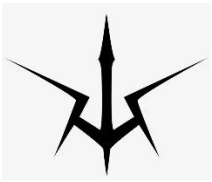
Score
0.5
0.5

Either of the items I2 or I4 can be recommended to the user.

In order to estimate the rating value, you can use adjusted cosine similarity of recommended item vector with all already rated item vectors in vector space and use the rating of item based on "closeness" of the vectors.

Question 3 (CLO: 1)	10 points

Consider the given interaction matrix for Crunchyroll website users.

	 Attack on Titan	 My Hero Academia	 Code Geass	 Vinland Saga
User 1	9	7	2	8
User 2	**?**	7	5	7
User 3	10	4	9	5
User 4	8	5	8	6

a) Find the Cosine similarity between all users.

Answer:

S(U1, U2) = 0.958

S(U1, U3) = 0.839

S(U1, U4) = 0.883

S(U2, U3) = 0.881

S(U2, U4) = 0.943

S(U3, U4) = 0.986

b) Predict the missing rating ***R***(U2, I1) using mean-centered prediction function for *k*=2.

$$R_U = \overline{r_a} + \frac{\sum_{b \in N} sim(a,b) * (r_{b,p} - \overline{r_b})}{\sum_{b \in N} sim(a,b)}$$

Closest k=2 neighbors are U1 and U4:

*R(U2, I1) = 6.33 + [((0.958 * 2.5) + (0.943 * 1.25)) / (0.958 + 0.943)]*

R(U2, I1) = 8.212 (can round off to 8)

Good luck!