CS 4960-6550: Introduction to Information Retrieval

Assignment 3

Instructor: Qingyao Ai *Teaching Assistant: Tao Yang*

Question 1 (10 points)

What are pointwise learning-to-rank models and pairwise learning-to-rank models (5 points)? Why should pairwise learning-to-rank models perform better than the pointwise models (5 points)?

Question 2 (12 points)

Why is feature selection important for learning-to-rank? Explain it from the perspectives of both system efficiency (6 points) and ranking effectiveness (6 points).

Question 3 (12 points)

What is the advantages and disadvantages of Latent Semantic Indexing (LSI) compared to TF-IDF? Explain your answer in terms of both retrieval effectiveness (6 points) and efficiency (6 points).

Question 4

Table 1 shows a tiny database containing the ratings (from 0 to 5) provided by 6 users for the movies A, B, C, D, E, and F.

Table 1: The ratings of different users to movie A, B, C, D, E, and F. "?" denotes "unknown".

	A	B	C	D	E	F
$Alice$	5	2	3	4	?	?
Bob	2	2	2	3	?	?
$User_1$	3	1	2	3	3	2
$User_2$	3	3	4	3	3	2
$User_3$	3	2	1	5	4	1
$User_4$	2	3	3	4	1	4

4.1 User-based Collaborative Filtering (32 points)

According to Pearson correlation, who are the two most similar users to Bob (8 points)? If we use user-based nearest neighbor collaborative filtering with Pearson correlation ($k = 2$), what are the predicted ratings of movie E and F for Alice and Bob (6 points for each)? Note that real numbers are acceptable.

4.2 Item-based Collaborative Filtering (34 points)

If we use item-based nearest neighbor collaborative filtering with Pearson correlation ($k = 2$), what are the predicted ratings of movie E and F for Alice and Bob (6 points for each)? Note that real numbers are acceptable. If we are going to recommend E or F to Bob and Alice using the predicted ratings, do user-based and item-based collaborative filtering algorithms give the same results (4 points)? If not, which algorithm might be better (6 points)?