CSC 4170
Web Intelligence and Social Computing

Homework Assignment #3
Date: Monday, 9 November 2009
Due Date: Monday, 23 November 2009 at 6:30 pm

`http://wiki.cse.cuhk.edu.hk/irwin.king/teaching/csc4170/2009`

1. You may form a group of no more than two persons to do the homework assignments.

2. If you decided to form a group with two persons, the final score for the assignment will be given to both persons. Moreover, since it is done by two persons, you are expected to do more than just a single person group.

3. Once the group has been formed, you should stay together throughout the class, including the class project as well.

4. Lastly, the final examination will be assessed individually.

5. Make sure you submit the electronic copy of your homework assignment to the VeriGuide system through the web as Assignment #3.

Homework Assignment

1. IR Evaluation

   Given a user's query, a search engine returned ten webpages, denoted by p_i (i is the position). After that the user gave his evaluations on the results. The evaluation is an integer from 1(most non-relevant) to 5 (most relevant). Please calculate the DCG_4, $IDCG_4$ and $NDCG_4$.

   Table 1: ground truth given by the user

positions of pages returned	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
ground truth given by user	5	2	3	4	1	5	1	2	1	2

   In real cases, users may usually evaluate on only a small scale of recommended results. For example, the evaluation from the user may be like the following table. Here, the value 0 denotes that the user did not evaluate this item. Intuitively, it is often argued that users are likely to evaluate items they are interested. In this case, how to calculate the NDCG value will be reasonable? Please give your ideas.

   Table 2: ground truth given by the user

positions of pages returned	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10
ground truth given by user	5	2	0	0	0	5	0	2	1	2

2. HITS and PageRank

HITS: The the link structure of four Web pages is shown in the following figure. The initialization of hub score and authority has been given. Please calculate the hub and authority scores of each state in the first and second iterations.

PageRank: Suppose $d = 0.85$, please calculate PageRank score of each state in the first and seconde iterations. The initiate score of each state is 0.25.

Regression: In materials of both lecture and tutorial notes, we suppose PageRank will regress at last. Please try to find relevant materials to proof this conclusion. Hints: PageRank is a special case of Markov Process, where relevant theorems can be found.

(1/4,1/4)

(1/4,1/4)

(1/4,1/4)

(1/4,1/4)

(x,y): x=hub score; y=authority score
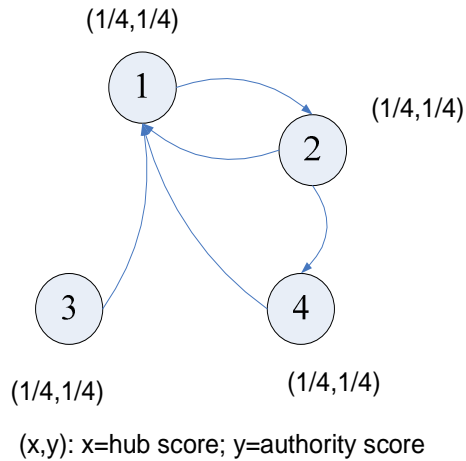
3. Similarity Calculation

Pearson Correlation Coefficient (PCC) similarity and Cosine similarity have been defined in the lecture notes. The following table shows the ratings of 2 users on 10 items. Please calculate their PCC and Cosine Similarities. (The value 0 means the user has not rated the item.)

Table 3: user-item Matrix

items	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10
user1	5	2	0	0	0	4	0	2	3	2
user2	0	1	4	7	0	2	0	0	5	1