DWDM PROJECT

Project 8

Implementing PageRank in Citation Network from Scratch and Ranking

- 201202036, Darpan Baheti
- 201201131, Havish Chennamaraj
- 201264197, Alok Thatikunta
- 201201127, Kamal Gurala

Page Rank

- Page Rank is a model for link analysis and ranking the nodes in a network.
- The rank value indicates an importance of a particular page.
- The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it ("incoming links").
- A page that is linked to by many pages with high PageRank receives a high rank itself.

Objective

- In this project, we implement the Page rank algorithm to rank research papers, authors and conferences, given Citation Network.
- The importance of a research paper is captured by the peer vote, which is the research paper being cited in other research papers.
- Using the scores of the research papers calculated, we formulated scores for conferences and authors and ranked them as well.
- We took into account the time factor in ranking the research papers to reduce the bias against the recent papers which get less time for being studied and consequently cited by the researchers as compared to the older papers

Work Flow and Algorithm

- Firstly, we create a database of the input citation network using sqlite.
- The algorithm for ranking the research papers based on citation network uses the two types of edges in a graph.
- Outlinks: From a given node N, link all the nodes Ni that the node N cites.
- Inlinks: To a given node N, link all the nodes Nj that cite the node N.
- These outlinks and inlinks will be used while calculating the authoritative score for each node.

Algorithm

- The algorithm for ranking the research papers consists of two parts:

   1) Creating the outlinks and the inlinks.
      - The procedure uses paper citations and creates a data structure mapping each paper to its inlinks and outlinks.

   2) Iterative PageRank algorithm to calculate the authoritative score for each node.

Algorithm for Page Ranking

1: **Initialize** Paper Rank PR to 1.0 **for each** paper R

2: **while** true
3:     $flag$ = true

4:     **for each** paper R in PR
5:         Current Score $CS = PR[R]$
6:         **if** R **in** PI
7:             Inlinks List $IL = PI[R]$
8:             New Score $NS = 0.0$
9:             **for each** inlink I **in** IL
10:                 **if** I **in** PR
11:                     $NS \mathrel{+}= PR[I]/OC[I]$
12:             $NS = (1-\theta) + \theta*NS$
13:             **if** CS **is not equal to** NS
14:                 $flag$ = false
15:             Updated Paper Rank $UPR[R] = NS$

16:     **if** $flag$ **is equal to** true
17:         **break**
18:     **copy** UPR **to** PR
19:     **clear** UPR

20: Maximum Score $MS = $ **Maximum Score in** PR
21: **for each** paper R **in** PR
22:     $PR[R] \mathrel{/}= MS$

Ranking Conferences

- The rank of a conference depends on the quality of research papers it publishes.
- Using the scores for the research papers, we will rank the various conferences by calculating a cumulative authoritative score for each conference.
- The score for a conference will essentially be the average score of all the research papers published in that conference.

Conference Score $CS = \Sigma(\text{Paper Score } PS)/(\text{ Number of Papers published in the Conference } NPC[C])$

Ranking Authors

- The rank of an author depends not only on the quality of research papers but also the conference he/she publishes his/her papers in.
- For ranking authors, thus we will use the scores of the research papers published by the author as well as the score of the conference in which the paper was published.

$$\text{Author Score} = \Sigma(\text{Paper Score } PS * \text{Conference Score } CS)/(\text{Number of Papers published by the Author } NPA[A])$$

Conclusions and Future Work

- The purpose of our work was to propose a way to identify the key research papers, conferences and authors from various fields of research.
- Time-independent algorithm for ranking, which as shown gives fairly satisfactory results.
- Project can be extended to keywords in conferences, which could help new researchers to identify the conferences of his/her interest. Similarly we can also extend this work to keywords by authors.
- Another future work can be the implementation of our work in recommendation systems. We can have a system identifying co-authors and recommending topics of interest, etc.