A Similitude Study on Weighted Pagerank and Enhanced RatioRank Algorithms Using Visit of Links of a Web Page

V. Annapoorna
Asst. Professor  
Dept of IT, ANITS

M.V. Kishore
Asst. Professor  
Dept of IT, ANITS

Abstract: World Wide Web is an Internet-wide distributed hypermedia information retrieval system which provides access to a large universe of documents also called as web pages. The web pages are formatted in markup language called HTML (Hypertext Markup Language). Each web page contains text, multimedia data as well as links to other web pages. The idea of a world of boundless information in which all items have a reference by which they can be retrieved. This means you can jump from one web page to another simply by clicking on hot spots or reference words. Web pages can be explored through their web query interfaces. To extract essential information from web, web extractors execute number of tasks based on their respective structural design. When a user enters a query to the search engine, it generally extracts a bulk number of pages in response to user’s query. To extract more resourceful web pages in the result list, various page ranking techniques are applied on the query results. Most of the ranking algorithms are either link or content oriented. Which do not think about user usage trends. In this paper, an analysis of Weighted Page Rank Algorithm (WPR) based on number of Visits of Links (VOL) of a web page and Enhanced Ratio Rank (ERR) algorithm using visit of links of a web page is performed. The paper also presents the comparison between WPR using VOL and ERR algorithms. Weighted Page Rank Algorithm based on number of Visits of Links (VOL) is being devised for web explorers, which works on the basis of weighted page rank algorithm and takes number of visits of inbound links of web pages into account. Ratio Rank is improved version of Page Rank Algorithm. Ratio Rank algorithm works on the basis of linkage organization web for calculating rank value of page. The prime idea in this is to increase the weight of in links over the weight of out links, as it is more important. So, this comparison is very useful to display most precious pages on the top of the result list on the basis of user browsing performance, which reduce the search space to a large extent.

Keywords: weighted page rank, inlinks, outlinks, Ratio Rank.

INTRODUCTION

Web plays dynamic role because it contains vast data as collection of large number of WebPages and every second new pages are added, updated and deleted in web. Retrieving efficient, relevant and meaningful information from this large source of information is very challenging job. Every search engine applies an algorithm to large number of WebPages in search results which calculates rank of every WebPages and ensure that most efficient and relevant WebPages as per query made by user appear first in search results. A page ranking mechanism called

Weighted Page Rank Algorithm based on Visits of Links (VOL) is being devised for search engines, which works on the basis of weighted page rank algorithm and takes number of visits of inbound links of web pages into account.

The original Weighted Page Rank algorithm (WPR) is an extension to the standard Page Rank algorithm. WPR takes into account the importance of both the inlinks and outlinks of the pages and distributes rank scores based on the popularity of the pages. The proposed algorithm is used to find more relevant information according to user’s query.
RANDOMZE HITS etc. As most of the ranking algorithms proposed are either link or content oriented in which consideration of user usage trends are not available. Some algorithms assign larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

We mainly focus on finding more relevant information according to user’s query. By Assigning more rank value to the outgoing links which is most visited by users and received higher popularity from number of inlinks.

**Advantages:**
- Works more Faster in rendering frequent pages based on user trends.
- Doesn’t give more rank values based on Inbounds.
- User can’t intentionally increase the rank of a page by visiting the page multiple times because the rank of the page depends on the probability of visits (not on the count of visits) on back linked pages.

**METHODOLOGY**

**Algorithm for weighted page rank using VOL:**
The original Weighted PageRank algorithm assigns larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks). The popularity from the number of inlinks and outlinks is recorded as number of inlinks and outlinks). The popularity from the outlink page gets a value proportional to its popularity (its rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

The original Weighted PageRank algorithm assigns larger rank values to more important (popular) pages. The popularity from the number of inlinks and outlinks is recorded as number of inlinks and outlinks). The popularity from the outlink page gets a value proportional to its popularity (its rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

Algorithm for weighted page rank using VOL:
The original Weighted PageRank algorithm assigns larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks). The popularity from the number of inlinks and outlinks is recorded as number of inlinks and outlinks). The popularity from the outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

Advantages:
- Works more Faster in rendering frequent pages based on user trends.
- Doesn’t give more rank values based on Inbounds.
- User can’t intentionally increase the rank of a page by visiting the page multiple times because the rank of the page depends on the probability of visits (not on the count of visits) on back linked pages.

**Algorithm for weighted page rank using VOL:**
The original Weighted PageRank algorithm assigns larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks). The popularity from the number of inlinks and outlinks is recorded as number of inlinks and outlinks). The popularity from the outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

Advantages:
- Works more Faster in rendering frequent pages based on user trends.
- Doesn’t give more rank values based on Inbounds.
- User can’t intentionally increase the rank of a page by visiting the page multiple times because the rank of the page depends on the probability of visits (not on the count of visits) on back linked pages.

**METHODOLOGY**

**Algorithm for weighted page rank using VOL:**
The original Weighted PageRank algorithm assigns larger rank values to more important (popular) pages. Each outlink page gets a value proportional to its popularity (its number of inlinks and outlinks). The popularity from the number of inlinks and outlinks is recorded as number of inlinks and outlinks). The popularity from the outlink page gets a value proportional to its popularity (its number of inlinks and outlinks).

Advantages:
- Works more Faster in rendering frequent pages based on user trends.
- Doesn’t give more rank values based on Inbounds.
- User can’t intentionally increase the rank of a page by visiting the page multiple times because the rank of the page depends on the probability of visits (not on the count of visits) on back linked pages.

**Method:**

Step 1: Finding a Website: Find a website which has rich hyperlinks.
Step 2: Building a Web Map: Then generate the web map from the selected website.
Step 3: Calculate W in (v, u): Then calculate the W in (v, u) for each node present in web graph.

Where:
- \( W_{in}^{(v,m,n)} \) is the weight of link(v,m,n) calculated based on the number of in links of page v and the number of in links of all reference pages of page v.
- \( I_n \) and \( I_p \) are the number of incoming links of page n and page p respectively.
- \( WPR_{vol}(u) \) and \( WPR_{vol}(v) \) are rank scores of page u and v respectively.
- \( Lu \) denotes number of visits of link which is pointing page u form v.
- \( TL(v) \) denotes total number of visits of all links present on v.

**Algorithm for Ratio Rank:**

Ratio Rank is improved version of Page Rank Algorithm. Ratio Rank again uses link structure or web structure mining for calculating rank value of page. That uses weight of out links, weight of in links and number of visits of links by users as the parameters for calculating ratio rank of any page.

**Input:**
- Query to be searched is the input.

**Method:**

Step 1: Finding a Website: Find a website which has rich hyperlinks.
Step 2: Building a Web Map: Then generate the web map from the selected website.
Step 3: Calculate W in (v, u): Then calculate the W in (v, u) for each node present in web graph.

Where:
- \( W_{in}^{(v,m,n)} \) is the weight of link(v,m,n) calculated based on the number of in links of page u and the number of in links of all reference pages of page u.
- \( I_n \) and \( I_p \) are the number of incoming links of page n and page p respectively.
- \( W_{out}^{(m,n)} \) is the weight of link(m,n) calculated based on the number of out links of page u and the number of in links of all reference pages of page v.
- \( O_n \) and \( O_p \) are the number of incoming links of page n and page p respectively.

**Step 4:** Apply proposed formula: Now calculate the Page Rank value of the pages.

\[ WPR_{vol}(u) = (1 - d) + d \sum_{v \in B(u)} \frac{L_v WPR_{vol}(v) W_{in}^{(v,m,n)}}{TL(v)} \]

- U represents a web page.
- B(u) is the set of pages that point to u.
- d is the dampening factor.
- WPR vol (u) and WPR vol (v) are rank scores of page u and v respectively.
- Lu denotes number of visits of link which is pointing page u form v.
- TL(v) denotes total number of visits of all links present on v.

**Ratio Rank:**

Ratio Rank is improved version of Page Rank Algorithm. Ratio Rank again uses link structure or web structure mining for calculating rank value of page. That uses weight of out links, weight of in links and number of visits of links by users as the parameters for calculating ratio rank of any page.

**Input:**
- Query to be searched is the input.

**Method:**

Step 1: Finding a Website: Find a website which has rich hyperlinks.
Step 2: Building a Web Map: Then generate the web map from the selected website.
Step 3: Calculate W in (v, u): Then calculate the W in (v, u) for each node present in web graph.

Where:
- \( W_{in}^{(v,m,n)} \) is the weight of link(v,m,n) calculated based on the number of in links of page u and the number of in links of all reference pages of page u.
- \( I_n \) and \( I_p \) are the number of incoming links of page n and page p respectively.
- \( W_{out}^{(m,n)} \) is the weight of link(m,n) calculated based on the number of out links of page u and the number of in links of all reference pages of page v.
- \( O_n \) and \( O_p \) are the number of incoming links of page n and page p respectively.

Step 4: Apply proposed formula: Now calculate the Page Rank value of the pages.

\[ RR(u) = (1 - d) + d \sum_{v \in B(u)} \frac{(v \# \# W_{in}^{(v,m,n)} + v \# W_{out}^{(m,n)}) \times RR(v)}{TL(v)} \]

- U represents a web page.
- RR(u) and RR(v) are ratio rank values of page u and v respectively.
d is the dampening factor.
- RR(u) and RR(v) are rank scores of page u and v respectively. Lu denotes number of visits of link which is pointing page u form v.
- TL(v) denotes total number of visits of all links present on v.
- Values of x and y is always set between 0 and 1 (Where x>y)

Algorithm for Enhanced Ratio Rank:
Enhanced-Ratio Rank also consider ratio of weight of the inlinks and weight of out links and visit counts of links by users for calculation of the rank value of particular page. It checks which ratio gives the best result i.e. which ratio of inlinks weight and out links weight helps to give better relevancy of the web pages. New Enhanced algorithm is given as follows:

$$RR(u) = (1-d) + d \sum_{v \in B} \frac{(v \cdot 0.7 \cdot W_{(v \to u)} + 0.3 \cdot W_{u \to v}) \cdot RR(v)}{TL(v)}$$

It uses same parameters as Ratio Rank equation. As in equation 70 percent of the weight of in links and the 30 percent of the weight of the out links is being used because this gives better result as compare to other ratios. By using all three parameters for computing the page rank value of WebPages and taking the best ratio of weight of in links and out links gives the better relevancy of web pages. But the problem of theme drift (some link may not give the search results about the query) still exists in this algorithm.

RESULT ANALYSIS

In this section we will describe and explain some results we got from the work we have been doing. Here we have taken a hyperlinked web graph and calculated Page Rank value of each page based on original weighted page rank algorithm and proposed algorithm i.e. based on number of visits of links (VOL).

Results for Weighted Page Rank Algorithm based on VOL:

Table 1: Ranks with varying values of ‘d’

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Systems</td>
<td>1.85</td>
</tr>
<tr>
<td>Computer Hardware</td>
<td>0.833</td>
</tr>
<tr>
<td>GUI</td>
<td>0.825</td>
</tr>
<tr>
<td>System Software</td>
<td>0.766</td>
</tr>
</tbody>
</table>

We can give a graphical representation which shows the variation between WPR (VOL) and WPR scores of different web pages of a web graph. In this we can observe that WPR (VOL) is performing better than original WPR. As the proposed algorithm calculates higher relevant score than the existing one.
Table 3: Comparison of Weighted Page Rank based on VOL and Enhanced Ratio Rank Algorithms:

<table>
<thead>
<tr>
<th>Keyword</th>
<th>RANKS</th>
<th>WPR vol</th>
<th>Enhanced RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS</td>
<td>1.85</td>
<td>2.088</td>
<td></td>
</tr>
<tr>
<td>CLI</td>
<td>1.0</td>
<td>1.198</td>
<td></td>
</tr>
<tr>
<td>MS DOS</td>
<td>1.0</td>
<td>1.085</td>
<td></td>
</tr>
<tr>
<td>GUI</td>
<td>0.575</td>
<td>1.049</td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>0.7166</td>
<td>0.915</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>0.433</td>
<td>0.631</td>
<td></td>
</tr>
</tbody>
</table>

**SAMPLE SCREENS:**

The following are the sample screens of Weighted Page Rank Algorithm based on VOL and Enhanced-Ratio Rank algorithms displaying page rank values of more relevant web pages.
CONCLUSION
The weighted page rank algorithm is discussed which is more target-oriented. This weighted page rank algorithm calculates Page Rank value of web pages based on the visits of incoming links as well as the popularity of inlinks of a web page and takes into account ratio rank factors. It is not only consider link structure it includes the users focus on a particular page, but the main problem in this concept is calculation of visits of a links for that we have given a simple concept to monitor and count the hits or visits. User generally spends a lot of time in surfing through the search results to find the relevant pages. This analysis provides more efficient results than original WPR. The ordering of pages in this way increases the relevancy of pages and thereof provides the user with quality search results.

FUTURE SCOPE
Algorithms which use link structure has mainly many challenges like emphasis on old pages, theme drift, page cheating. This work can be further be enhanced by basing it on content rather than purely link structure or web structure. The page ranking algorithm which use web structure mining doesn’t care about user’s query, Only link structure of WebPages are considered in calculation of page rank value of WebPages. On the other side the page ranking algorithm which uses web content mining take user’s query into account and doesn’t care about link structure of WebPages for calculating page rank values of the WebPages.

REFERENCES