<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Feature-conscious ranking framework</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Xu, Kaijian</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Xu, K. (2007, March). Feature-conscious ranking framework. Presented at Discover URECA @ NTU poster exhibition and competition, Nanyang Technological University, Singapore.</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2007</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/9058">http://hdl.handle.net/10220/9058</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2007 The Author(s).</td>
</tr>
</tbody>
</table>
The PageRank algorithm [1] for Web search ranking models the Internet as a content-agnostic link graph, and derives the importance of individual Web pages with regard to their content or other properties. Various modifications of PageRank to take Web page features into account have been proposed [2], but most variants are intended for specific applications and cannot be generalized to handle any page feature. Hence, we propose the Feature-Conscious Ranking Framework (FeatRank) as a universal framework to incorporate any generic page feature into the page ranking computation.

Preliminaries: PageRank

Given a directed link graph with N Web pages \( \{n_1, n_2, \ldots, n_N\} \), the PageRank equation is:

\[
r = (A + \frac{1-\alpha}{N}) E r
given \alpha \in (0,1)
\]

where \( r = \{r_1, r_2, \ldots, r_N\} \) is the column vector of PageRank scores of each page, \( A = \{a_{ij}\} \) is the link adjacency matrix where \( a_{ij} = 1 \) if \( n_i \) links to \( n_j \), \( N_i \) is the number of outlinks of \( n_i \), \( \alpha \) is the probability of teleportation from \( n_i \) to other pages, and \( E = \{e_{ij}\} \) is the row vector matrix where \( e_{ij} = 1 \) if \( j \) is the page with a self-loop in the graph, \( e_{ij} = 0 \) otherwise. Usually, \( E = \gamma I \) where \( I \) is a column vector comprising all ones, giving a uniform probability of teleporting to and from any page.

The FeatRank Framework

Our work, FeatRank framework, generalizes a family of PageRank variants that augments the link graph in order to explicitly account for page features in its rank computation. Any quantifiable link-independent scalar property associated with a Web page is admissible. A self-loop with link weight \( \gamma \) corresponding to the page feature is added to each “featured” page.

Intuitively, \( \gamma \) denotes the probability that a Web surfer will remain at the same node during a random walk of the link graph. This value should be subtracted from the probabilities of following an outlink and/or random teleportation so that all the new probabilities sum up to 1 as illustrated in Figure 1. We assign a parameter \( \rho \) to denote the ratio of \( \gamma \) consumed from teleportation; the rest is consumed from teleportation.

The FeatRank equation is thus given by:

\[
L = \gamma A + (1-\gamma) E \quad \text{ (2)}
\]

where \( L = \{l_1, l_2, \ldots, l_N\} \) with \( l_i = \gamma a_{ii} + \frac{1-\gamma}{N} \) in (2),

\[
S = diag(l_1, l_2, \ldots, l_N)
\]

FeatRank Properties

Fixing \( \gamma \) to the commonly used value of 0.85, we focus on the effects of varying parameters \( \alpha, \rho \) and \( \gamma \) on featured pages and the ranking ecosystem as a whole.

We define the following metric to evaluate the rank promotion/demotion of FeatRank over PageRank:

\[
Gain_{FeatRank,PageRank} = \frac{\sum_{i=1}^{N} (R_{FeatRank}(i) - R_{PageRank}(i))}{\sum_{i=1}^{N} R_{PageRank}(i)} \quad \text{ (3)}
\]

\( R \) and \( R_{FeatRank} \) represent the ordinal ranking induced on the page by PageRank and FeatRank respectively. Note that we use a smaller evaluation page set of size M here, instead of the whole link graph. Because the average page rank change over the whole link graph will always converge to zero, it is more meaningful to measure only subsets of pages whose ranks are expected to change.

Verifying the Properties of FeatRank

To verify the properties of FeatRank, we use an add-a-look-up-to “Mozilla Firefox”, the only article in WikiMW with Wikipedia “featured article” status, and evaluate the changes in each group of pages at a defined number of hops away in the link graph, denoted by (0,1). Using 3 representative \( \rho \) values and fixing \( \alpha = 0.85 \), we vary \( \gamma \) over [0.85,0.85] to study its effect on the Gain ratio. The results (Figure 2) show that when \( \gamma = 0 \) the featured article “Mozilla Firefox” is promoted, and when \( \gamma = 0 \) it is demoted; and (2) the effect on the featured article is independent of \( \rho \). We use one configuration point \( \gamma = 0.7, \rho = 0.2 \) of FeatRank to illustrate that the effect on neighboring nodes decreases as distance from the featured page increases. Figure 4 shows that when \( \alpha \) is 0.5, the scores of nearby reachable pages will be amplified/attenuated. When \( \alpha = 0.85 \), the scores of nearby reachable pages will be amplified/attenuated. When \( \alpha = 0.85 \), the scores of nearby reachable pages will be amplified/attenuated.

Next, using 2 representative values of \( \gamma = 0.85 \) and \( \gamma = 1.0 \) for FeatRank to illustrate that the effect on neighboring nodes decreases as distance from the featured page increases. Figure 4 shows that when \( \alpha = 0.5 \), the scores of nearby reachable pages will be amplified/attenuated. When \( \alpha = 0.85 \), the scores of nearby reachable pages will be amplified/attenuated. When \( \alpha = 0.85 \), the scores of nearby reachable pages will be amplified/attenuated.

Conclusion

We conclude that FeatRank is consistent better than PageRank. However, we still need to further study the evaluation of FeatRank in identifying “good articles” given the seed set of “featured articles” by comparing its performance against PageRank using the following metric:

\[
Gain_{FeatRank,PageRank} = \frac{\sum_{i=1}^{N} (R_{FeatRank}(i) - R_{PageRank}(i))}{\sum_{i=1}^{N} R_{PageRank}(i)} \quad \text{ (3)}
\]

The evaluation conducted on all “good articles” with \( \gamma \) in [0.2, 0.8] and \( \rho \) in [0.1, 0.85] (Figure 5) showed that FeatRank is consistently better than PageRank. The best configuration is in the region with small \( \rho \) and large \( \gamma \) (Gain\text{-}PageRank=6.2633). With Gain\text{-}PageRank\text{-}GroundTruth = -22.14, FeatRank yields a dramatic 28.36% reduction in rank error compared to PageRank.

Expert evaluation on “good articles” excluding “featured articles” exhibits the same trend: the best configuration achieves Gain\text{-}PageRank\text{-}GroundTruth = -14.2341, this indicates a 21.99% reduction in rank error.