<table>
<thead>
<tr>
<th>Title</th>
<th>Feature-conscious ranking framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Xu, Kaijian</td>
</tr>
<tr>
<td>Citation</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>Date</td>
<td>2007</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/9058">http://hdl.handle.net/10220/9058</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2007 The Author(s).</td>
</tr>
</tbody>
</table>
Feature-Conscious Ranking Framework

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 without 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 new features. 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:

$$r = \left(\alpha A + (1 - \alpha)E\right) r$$

where $r$ is the column vector of PageRank scores of each page, $A$ is the link adjacency matrix where $A_{ij}$ indicates the presence of a link from page $n_i$ to page $n_j$, $D_i$ is the number of outlinks of $n_i$, $\alpha$ is a value between 0 and 1, $E_{ij}$ is the entry in the pure matrix where $E_{ij} = \text{probability of teleportation from } n_i \text{ to } n_j$, and $\gamma$ is a constant. Usually, $E = \gamma \cdot \left(\frac{1}{N}\right) I$ 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 must be subtracted from the probabilities of following an outlink and/or random teleportation so that all these probabilities sum up to unity as illustrated in Figure 1. We assign a parameter $\alpha$ to denote the ratio of $\gamma$ consumed from the outlinks; the rest is consumed from teleportation.

The FeatRank equation is thus given by:

$$S = \left(\alpha L + \frac{1}{N} \mathbf{1} \cdot \mathbf{1}^T\right) S$$

where $L = \alpha \left(\lambda - \gamma_{\text{out}}\right) I$, where $\gamma_{\text{out}} = \sum_{n_i} \frac{\gamma}{D_i}$ and $\gamma = \sum_{n_i} \frac{\gamma}{D_i}$ is the empirical mean of all weights.

FeatRank Properties

Fixing $\alpha$ to the commonly used value of 0.85, we focus on the effects of varying parameters $\gamma$ and $\rho$ on featured pages and the ranking ecosystem as a whole. Effects a controlled change in the ranking of featured scores. The score $r$ of a page $n_i$ varies monotonically with respect to the weight of its self loop $\gamma$:

- If $\gamma > 0$, $r$ increases monotonically with $\gamma$ (score amplification)
- If $\gamma < 0$, $r$ decreases monotonically with $\gamma$ (score attenuation)
- If $\gamma = 0$, $r$ remains unchanged.

$\rho$ controls the propagation of the score amplification/attenuation of a featured page to nearby reachable pages in the link graph. Given a featured page with a self-loop, the propagation of its score amplification/attenuation is generally inversely proportional to $\rho$ as follows:

- If $\rho = \infty$, the scores of nearby reachable pages will be attenuated/damped.
- If $\rho = 0$, the scores of nearby reachable pages will be amplified/attenuated.
- If $\rho = \infty$, $\rho = 0$, propagation, relative ordering of other pages remains unchanged.

By fine-tuning the parameters, FeatRank can be adapted to a variety of different applications. As shown in Table 1, there are altogether 7 possible operating configurations for FeatRank.

Existing Ranking Models as Special Cases of FeatRank

The generic nature of FeatRank can be demonstrated by comparing it to existing ranking models. 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 new features. Hence we propose the Feature-Conscious Ranking Framework (FeatRank) as a universal framework to incorporate any generic page feature into the page ranking computation.

Experimental Results

We conducted experiments to (1) verify the properties of FeatRank; evaluation of the parameters' effect on ranking; (2) explore actual application of FeatRank to bias ranks of featured pages and pages surrounding them.

We used two subsets of articles from the English language Wikipedia "stable release" (26 Aug 2008). The first subset, wikiWB, consists of all 159 articles under the category "Web Browsers" and is used to verify the properties of FeatRank. The second subset, wikiMW, consists of 546 articles from the category "Microsoft Windows" and is used in our pilot application of FeatRank.

We used two representative values of $\gamma$, we vary $\gamma$ in $\{0.05, 0.95\}$. Figure 3 shows that (1) the effect on surrounding pages is greatest with small $\gamma$; the effect is gradually diminished until $\gamma = \omega$. When $\gamma = \omega$, propagation occurs again with reversed trend; (2) the rank gain varies in proportion to $\gamma$.

Next, using 2 representative values of $\gamma$, we vary $\gamma$ in $\{0.5, 0.95\}$. Figure 5 shows that (1) the effect on surrounding pages is greatest with small $\gamma$; the effect is gradually diminished until $\gamma = \omega$. When $\gamma = \omega$, propagation occurs again with reversed trend; (2) the rank gain varies in proportion to $\gamma$. We use a parameter $\alpha$ to denote the ratio of $\gamma$ consumed from the outlinks; the rest is consumed from teleportation.

Exploring Real Applications for FeatRank

As a pilot study, we used FeatRank to induce a ranking of articles in the WikiMW dataset. We treat "good articles" rated "B-class" and above in Wikipedia as the ground truth, similar to ranking obtained from domain experts. Articles in the same class are assigned the same rating. We then evaluate the efficacy of FeatRank in identifying "good articles" given the seed set of "featured articles" by comparing its performance against PageRank using the following metric: $G_{\text{Gain}} = \frac{G_{\text{FeatRank}, \text{PageRank}}}{G_{\text{PageRank}}}$.

We use a parameter $\alpha$ to denote the ratio of $\gamma$ consumed from the outlinks; the rest is consumed from teleportation.

Next, using 2 representative values of $\gamma$, we vary $\gamma$ in $\{0.5, 0.95\}$. Figure 5 shows that (1) the effect on surrounding pages is greatest with small $\gamma$; the effect is gradually diminished until $\gamma = \omega$. When $\gamma = \omega$, propagation occurs again with reversed trend; (2) the rank gain varies in proportion to $\gamma$.

To verify the properties of FeatRank, we use a set of features 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 $D(n)$. Using 3 representative $\gamma$ values and fixing $\alpha = 0.85$, we vary $\gamma$ over $\{0.85, 0.85, 0.85\}$ 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 $\rho$.