

Programming Assignment 3

Botao Hu (botaohu), Jiayuan Ma (jiayuanm)

Completed Date: *May 17, 2013*

1 Questions

1. What was the reasoning behind giving the weights to the url, title, body, header and anchor fields for the three tasks? Were there any particular properties about the documents that allowed a higher weight to be given to one field as opposed to another?

2. What other metrics, not used in this assignment, could be used to get a better scoring function from the document?

Other metrics, like user behaviors (query independent) and other metadata of the webpage

3. In BM25F, how do B_{url} , B_{title} , B_{header} , B_{body} , B_{anchor} , λ , λ' and K_1 affect the ranking function?

4. In BM25F, why did you select a particular V_j function?

We select $V_{\text{page rank}} = \log(\lambda' + \text{page rank})$, whose plot is in Figure 1. Since this function is a monotonically increasing function and it has the property of saturation, this function can be easily and reasonably applied to the original score function.

5. For a function that includes the smallest window as one component, how does varying B and the boost function change the performance of the ranking algorithm?

B has multiplicative effects on the ranking function, where the larger B is, the more impact the smallest window component has on the ranking algorithm.

2 Parameters

The reported ndcg scores on these three tasks are in Table 1.

	NDCG
Task 1	0.8662
Task 2	0.8758
Task 3	0.8682

Table 1: NDCG scores

2.1 Task 1

```
task1_W_title = 1.0
task1_W_header = 0.5
task1_W_url = 0.1
task1_W_body = 0.3
task1_W_anchor = 2.0
```

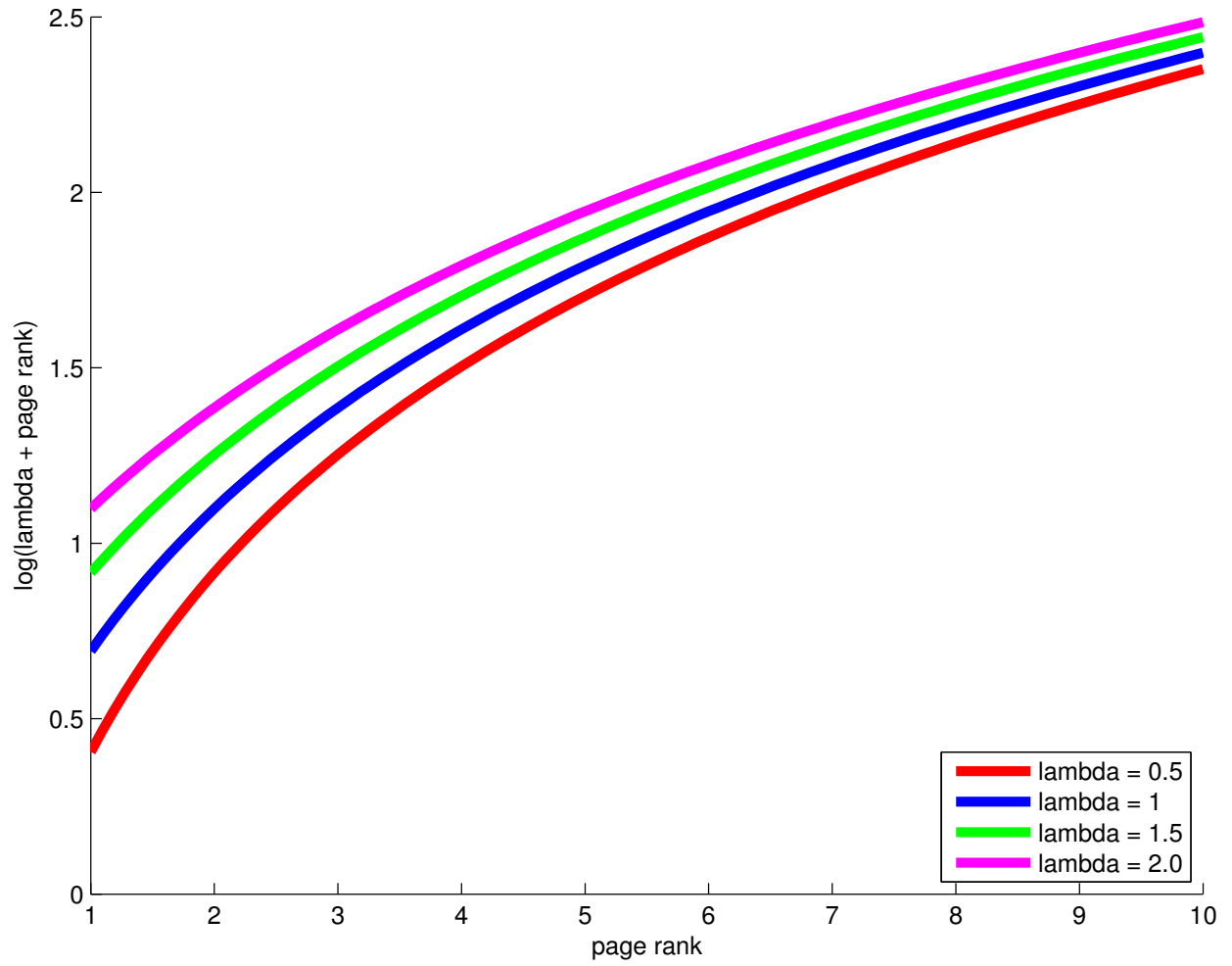


Figure 1: $V_{\text{page rank}} = \log(\lambda' + \text{page rank})$ with different λ' s

2.2 Task 2

```
task2_W_title = 1.0
task2_W_header = 0.5
task2_W_url = 0.1
task2_W_body = 0.3
task2_W_anchor = 2.0
task2_B_title = 0.3
task2_B_header = 0.5
task2_B_url = 2.0
task2_B_body = 1.0
task2_B_anchor = 0.1
task2_K_1 = 0.8
task2_lambda = 0.5
task2_lambda_prime = 0.5
```

2.3 Task 3

```
task3_W_title = 1.0
task3_W_header = 0.5
task3_W_url = 0.1
task3_W_body = 0.3
task3_W_anchor = 2.0
task3_B = 600.0
```