

Information Retrivel: Assignment 4

Shifei Chen

Exercise 9.1

The α should be 1 as the original query is still valueable in the feedback. β can be a value slightly larger than α as the relevant document is something we would like to see more in the next result, such as 1.1. $\gamma = 0$ as we do not want any non-relevant document next time.

1 Exercise 9.2

1. Users generally don't give any feedbacks, which makes it difficult to apply relevant feedback strategy. 2. The cost to provide relevant feedback for long query terms is high. 3. Web users care less about recall enhancing in general.

2 Exercise 9.3

When $\alpha = 1, \beta = \gamma = 0$, the orignal query will be identical to the revised query. But it does not always guranteen the revised query will always be closer to the relevant document. If γ is larger than β I believe the algorithm will penalize non-relevant documents more than award relevant document and pushes the query vector more away.

3 Exercise 9.4

The assumption is all of the relevant documents are clustered together and the non-relevant documents are sparsely located in the document set, hence it is much easier to find a clear centroid in the relevant docuemnts and award it than finding the centroid in the non-relevant documents. Also non-relevant documents could be differernt in many ways and the sum of their vectors could be zero in some edge cases. In that situation, the effect of penalize non-relevant docuemtns would be eliminated so to keep polarising our optimized vectors away from the non-relevant documents it is reasonable to only take 1 non-relevant document into consideration.