Information Retrieval

Tutorial Set - II (16.09.2022)

Question 1 (TF-IDF Scoring + Relevance Feedback)

Consider the following collection of five documents and two queries:

- Doc 1: *we wish efficiency in the implementation for a particular application*
- Doc 2: *the classification methods are an application of Li's ideas*
- Doc 3: *the classification has not followed any implementation pattern*
- Doc 4: *we have to take care of the implementation time and implementation efficiency*
- Doc 5: *the efficiency is in terms of implementation methods and application methods*
- Query1: *application of classification methods*
- Query2: *efficiency in implementation of applications*

Now consider that the vocabulary is:

{*efficiency, implementation, application, classification, methods, ideas, pattern, time*}.

(a) Represent each document and Query 1 using unit normal vectors following the "lnc.ltc" scheme. Rank the 5 documents based on their relevance with query 1 (most to least) measured via the cosine similarity metric.

(b) Represent Query 2 using a unit normal vector as above. First, rank the 5 documents based on their relevance with Query 2 (most to least). Given that gold standard relevant documents are D4 and D5, find out the unit normal vectors of Query 2, modified by Rocchio's algorithm [$\alpha = 1.0$, $\beta = 0.75$, $\gamma = 0.25$] for:

(i) Relevance feedback

(ii) Pseudo-relevance feedback (Consider top 2 results to be relevant).

Then, re-rank the documents based on the modified query vectors.

Question 2 (Evaluation Metrics)

Consider a corpus of 10 documents, and two retrieval systems S1 and S2.

For each row of the table below, the documents are ranked from left to right based on relevance (most to least) with some query. For the ground truth, the relevance grade is given in brackets (4-very relevant, 0 - irrelevant).

For binary relevance, consider non-zero relevance grades as relevant, zero grade as irrelevant.

Find the following metrics for both systems: (i) Avg. Precision (ii) NDCG

GT	1(4)	3(4)	2(3)	4(2)	8(1)	9(1)	5(0)	6(0)	7(0)	10(0)
S1	1	2	3	4	5	6	7	8	9	10
S2	3	2	4	1	6	10	9	7	5	8