Question

Suppose that L is a function of n_j. Using Lee's lift model, provide a justification for RW's solution.

Answer

We claim that the RW solution can be explained by Lee's model with $L = N - n_j$:

$$\begin{aligned}\hat{P}(A_j = 1 \mid R_q = y) &= \frac{n_j + L}{N + L} \\ &= \frac{n_j + (N - n_j)}{N + (N - n_j)} \\ &= \frac{N}{2N - n_j}\end{aligned}$$

Namely, for any shared attribute, the lift given to a document known to be relevant to a query is equal to the number of documents that do not possess that attribute. The intuitive justification for

this is that query terms that are rare in the corpus are going to be the most helpful in discriminating relevant from non-relevant documents (since few documents will have these rare terms). Thus, the lift should be proportional to the rarity of the attribute in the corpus. This provides a good intuitive justification for the RW solution.