Learning to rank (Chapter 15)

Definition 1 (Learning to rank IR System)
Let's have a set of documents $D_{1..|D|}$, a set of queries $Q_{1..|Q|}$ and a set of relevance judgements $J_{1..|J|}(Q_j, D_i)$, where

$$J(Q_j, D_i) = \begin{cases} 1, & \text{if a document } D_i \text{ is } \textbf{relevant} \text{ for a query } Q_j \\ 0, & \text{if a document } D_i \text{ is } \textbf{irrelevant} \text{ for a query } Q_j \end{cases} \tag{1}$$

the objective of a learning-to-rank IR System is to (one of the following):

A) find a function $f(Q_j, D_i)$ with the property:

$$\forall J(Q_j, D_{rel}) = 1, \forall J(Q_j, D_{irrel}) = 0 : f(Q_j, D_{rel}) > f(Q_j, D_{irrel}) \tag{2}$$

B) find functions $f_q(Q_j)$, $f_d(D_I)$, $f(Q_{emb}, D_{emb})$ with the properties:

$$\forall J(Q_j, D_{rel}) = 1, \forall J(Q_j, D_{irrel}) = 0 : \\ f(f_q(Q_j), f_d(D_{rel})) > f(f_q(Q_j), f_d(D_{irrel})) \tag{3}$$

Exercise 15/1

Consider a collection of queries, documents, and judgements:

Query 1: president public speaking

Query 2: presidential elections

Doc 1: Obama speaks in Chicago

Doc 2: President has spoken this morning

Doc 3: A new president was elected

Judgement 1: J(Query 1, Doc 1) = 1

Judgement 1: J(Query 1, Doc 2) = 1

Judgement 2: J(Query 1, Doc 3) = 0

Judgement 3: J(Query 2, Doc 3) = 1

With respect to Occam's razor principle, come up with a function $f(Q_j, D_i)$ that is consistent with this data set.

Exercise 15/2

What if we change the Document 2 from previous exercise to

Doc 2: President greeted press this morning

Exercise 15/3

As the dataset grows bigger, there is a good chance that we won't come up with a function $f(Q_j, D_i)$ that will fit the dataset perfectly. How can we evaluate how well the function fits the dataset? Is it fair to evaluate this on a dataset from which the function has been inferred?

Exercise 15/4

If the quality of $f(Q_j, D_i)$ can be automatically evaluated, can we create an algorithm that will *find an optimal* f for us?

Given a fixed representation of queries and documents to be a bag of words, how can we find a $f(Q_j, D_i)$ that assigns the weights to each of the words in the representation so that the condition in the Definition 1, objective A) holds?

Discuss your ideas.

Exercise 15/5

On the other hand, if we fix the $f(Q_j, D_i)$ in the Definition 1, objective B), can how can we find the optimal representation, or *embeddings* of query $f_q(Q_j)$ and document $f_d(D_I)$ so that, the condition in Definition 1, objective B) holds?

For example, consider a case where $f(Q_{\text{emb}}, D_{\text{emb}}) = \cos(f_q(Q_j), f_d(D_I))$.

Discuss your ideas.

Exercise 15/6

What are the advantages of using the approach of a more complex objective B) (embeddings-first), as compared to objective A) (directly ranking all query-document pairs)? Discuss.
