

This section talks about the learning to rank problem, known as LETOR, which is where we want to learn a function that can rank order a set of items. The most common application of this problem is in information retrieval, where there might be a query q and a set of documents that are relevant to this query, such as the set of documents that contain the string q . What we want out of this situation is to sort these documents in decreasing order of relevance and showing the top k documents to the user. A standard way to measure the relevance of a document d to a query q is to use a probabilistic language model based on a bag of words model. In practice, we would need to smooth the estimated distribution, which can be done using a Dirichlet prior representing the overall frequency of each word. There are also other signals which can be used to measure relevance, such as the PageRank of a web document to measure its authoritativeness, which is derived from the web's link structure. In terms of the original example, we can also compute how often and where the query occurs in the document.

If we have binary relevance labels, we can use a standard binary classification scheme to estimate $p(y = 1|x(q, d))$. If we have ordered relevancy labels, we can use ordinal regression to predict the rating $p(y = r|x(q, d))$. In both cases, the documents can then be sorted according to this scoring metric. The overall name of this method is called the pointwise approach to LETOR, and it is widely used because of its simplicity. A flaw in this approach is that it does not take into account the location of each document in the list, so it ends up penalizing errors at the end of the list just as much as errors found at the beginning of the list, which is often not the desired behavior. There is evidence to support the assertion that people are better at judging the relative relevance of two items rather than absolute relevance, so modeling the data using a (usually

linear) scoring function can create a special kind of neural network known as RankNet, a variant of which is used in Bing's search engine.