Information Retrieval SS 2023

Exercise 4: Machine Learning and IR, Evaluation, PageRank[1]

Prof. Goran Glavas, Benedikt Ebing, Fabian David Schmidt
Chair for Natural Language Processing

1 Machine Learning and IR

Consider the document-term matrix computed in the previous exercise and assume a binary logistic regression model has been trained to classify whether a document is about Lord of the Rings (=relevant). The learned weights correspond to $w_{Frodo} = 0.74$, $w_{Sam} = 0.986$, $w_{beast} = 0.3$, $w_{orc} = 0.625$, $w_{blue} = 0.124$, $w_0 = b = 0$.

	Frodo	Sam	beast	orc	blue
doc1	0.1249	0.2499	0	0	0
doc2	0.1249	0.1249	0.6021	0	0
doc3	0.1249	0	0	0	0.3010
doc4	0	0.1249	0	0	0.3010

1. For each of the documents in the previous task, what is the probability of each document being relevant?
2. Recall the word embedding task from the previous exercise: We want to train CBOW word embeddings over the following vocabulary:

   ["Frodo", "followed", "Sam", "into", "the", "dark", "Mordor", "Ring"]

   Our current instance consists of the center word Sam and the context words *Frodo, followed, into, the.* Our model has predicted the following prediction vector:

   $$\hat{y} = [0.40.320.120.10.050.090.120.15]$$

   What is the cross-entropy error of this example?

Evaluation

1. An IR system returns 8 relevant documents, and 10 nonrelevant documents. There are a total of 20 relevant documents in the collection. What is the precision of the system on this search, what is its recall?
2. Consider an information need for which there are 4 relevant documents in the collection. Contrast two systems run on this collection. Their top 10 results are judged for relevance as follows (the leftmost item is the top ranked search result):

System 1	RNRNN	NNNRR
System 2	NRNNR	RRNNN

[1] Exercise tasks based on "An Introduction to Information Retrieval" by Manning, Raghavan and Schütze

(a) What is the MAP of each system? Which has a higher MAP?

(b) Does this result intuitively make sense? What does it say about what is important in getting a good MAP score?

3. The following list of numbers represent the relevance scores of a returned ranked list of 20 documents retrieved in response to a query from a collection of 10,000 documents. The numbers represent a range from nonrelevant (=0) to highly relevant (=3). The top of the ranked list (the document the system thinks is most likely to be relevant) is on the left on the list. This list shows 6 relevant documents. Assume that there are 8 relevant documents in total in the collection.

   3 2 0 0 0   0 0 0 1 0   3 0 0 0 2   0 0 0 0 1

   (a) What is the precision of the system on the top 20?

   (b) What is the F1 and what is the Recall on the top 20?

   (c) Assume that these 20 documents are the complete result set of the system. What is the MAP for the query?

   Assume, now, instead, that the system returned the entire 10,000 documents in a ranked list, and these are the first 20 results returned.

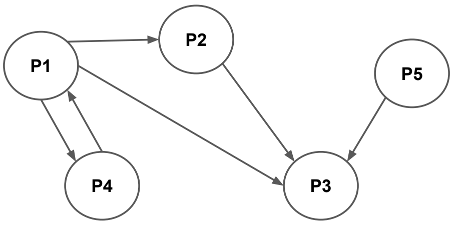
   (e) What is the largest possible MAP that this system could have?

   (f) What is the smallest possible MAP that this system could have?

   (g) What is the nDCG of the system? Assume this time that the returned documents comprises only 20 documents.

Optional: PageRank

Consider the following webgraph consisting of five websites.



Additionally we have the following content-based relevance value information (these could be for example similarity values from a vector space model) for our query:

$$r(q, P_1) = 0.43$$
$$r(q, P_2) = 0.31$$
$$r(q, P_3) = 0.05$$
$$r(q, P_4) = 0.12$$
$$r(q, P_5) = 0.84$$

Our search engine follows the random surfer model and computes the final relevance values with the PageRank-Algorithm. The probability of jumping to a random page is $1 - d = 0.1$ and correspondingly the probability of following a link is d $= 0.9$.

1. Which pages will be on top after three iterations?
2. Now consider all pages have the same initial relevance value, i.e., $r(q, P_i) = \frac{1}{5}$, $i \in \{1, ..., 5\}$. Re-compute the PageRank values. Is the result list still the same?